



医療統計解析のススメ

BellCurve エクセル統計による解析・多変量解析



横田幸弘

目次

● データの要約	2
● 1 標本の分析	11
● 母集団の分布	14
● 数理統計的側面(1)	30
● 数理統計的側面(2)	41
● 2 標本の分析	45
● 2 変量データの分析	55
● 3 標本以上の分析(分散分析)	61
コ克蘭の Q 検定	87
共分散分析	89
● 経時的測定データ	97
用量-反応関係の検出	103
コ克蘭・アーミテージ検定	106
● 多重比較	108
● 実験計画法	111
● ノンパラメトリック検定のまとめ	114
● 2×2 分割表の周辺	132
m×n分割表	143
尤度と最尤推定法について	150
● マンテル・ヘンツェル法	154
拡張マンテル法	158
シンプソンのパラドックス	163
● 重回帰分析	165
● ロジスティック回帰分析	170
● 生存率の分析	181
Kaplan-Meier 法	181
log-rank, 一般化 Wilcoxon 検定	185
Cox 比例ハザードモデル	189
● 効果量と必要サンプル数の計算	201
● 統計に関するいくつかの失敗例	204
● 主な参考文献	214

homepage: <http://www.shamonie.jp/>

データの要約

データを得たら？

視覚的、数量的
に要約すること

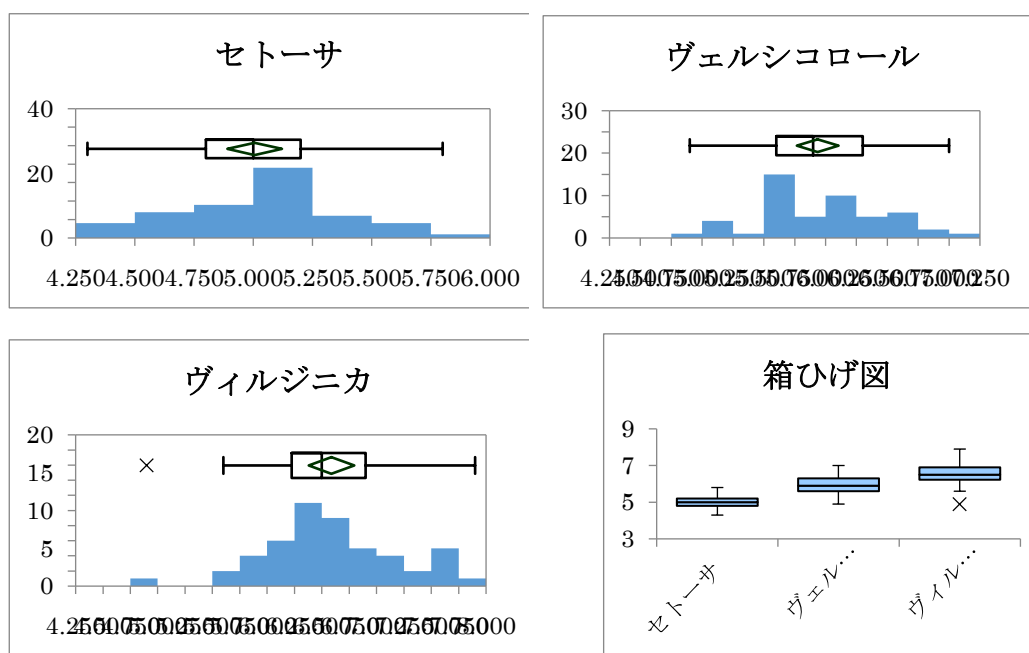
<あやめのデータ(「がくの長さ」のみ抜粋)>

あやめのデータ(がくの長さ)							
No.	セトー サ	ヴェルシコロ ル	ヴィルジニ カ	No.	セトー サ	ヴェルシコロ ル	ヴィルジニ カ
1	5.1	7.0	6.3	26	5.0	6.6	7.2
2	4.9	6.4	5.8	27	5.0	6.8	6.2
3	4.7	6.9	7.1	28	5.2	6.7	6.1
4	4.6	5.5	6.3	29	5.2	6.0	6.4
5	5.0	6.5	6.5	30	4.7	5.7	7.2
6	5.4	5.7	7.6	31	4.8	5.5	7.4
7	4.6	6.3	4.9	32	5.4	5.5	7.9
8	5.0	4.9	7.3	33	5.2	5.8	6.4
9	4.4	6.6	6.7	34	5.5	6.0	6.3
10	4.9	5.2	7.2	35	4.9	5.4	6.1
11	5.4	5.0	6.5	36	5.0	6.0	7.7
12	4.8	5.9	6.4	37	5.5	6.7	6.3
13	4.8	6.0	6.8	38	4.9	6.3	6.4
14	4.3	6.1	5.7	39	4.4	5.6	6.0
15	5.8	5.6	5.8	40	5.1	5.5	6.9
16	5.7	6.7	6.4	41	5.0	5.5	6.7
17	5.4	5.6	6.5	42	4.5	6.1	6.9
18	5.1	5.8	7.7	43	4.4	5.8	5.8
19	5.7	6.2	7.7	44	5.0	5.0	6.8
20	5.1	5.6	6.0	45	5.1	5.6	6.7
21	5.4	5.9	6.9	46	4.8	5.7	6.7
22	5.1	6.1	5.6	47	5.1	5.7	6.3
23	4.6	6.3	7.7	48	4.6	6.2	6.5
24	5.1	6.1	6.3	49	5.3	5.1	6.2
25	4.8	6.4	6.7	50	5.0	5.7	5.9

数量的要約

基本統計量			
変数	セトーサ	ヴェルシコロール	ヴィルジニカ
n	50	50	50
平均	5.0060	5.9360	6.5880
不偏分散	0.1242	0.2664	0.4043
標準偏差	0.3525	0.5162	0.6359
最小値	4.3000	4.9000	4.9000
第1四分位数	4.8000	5.6000	6.2250
中央値	5.0000	5.9000	6.5000
第3四分位数	5.2000	6.3000	6.9000
最大値	5.8000	7.0000	7.9000
四分位範囲	0.4000	0.7000	0.6750

視覚的要約



ひし形は平均値と95%信頼区間、Xは外れ値

BellCurve エクセル統計より「あやめのデータ」

Fisher, R.A. (1936). The Use of Multiple Measurements in Taxonomic Problems, *Annals of Eugenics*. 7. pp.179-188.

統計学は記述統計学と推測統計学に分けられる。

記述統計 **descriptive statistics**

推測統計 **inferential statistics**

しばらく記述統計の話です

記述統計: データが得られたとき、解析の第一歩はデータを要約し、その分布の大きな特徴を把握すること。

データの要約 (**summary**)

1) 数量的なデータの要約 (**numerical summary**)

2) 視覚的なデータの要約 (**graphical summary**)

目的は、

1. 適切な解析手法の選択のため

2. データが公正に誤りなく収集されていることを示すため

つまり視覚的、
数量的に要約

推測統計: 記述統計によって要約されたデータから母集団全体の特徴や性質を推測する。統計学の分析と呼ばれている手法はこの推測統計学のこと。観察された集団を母集団から抽出した標本とみなすのが推測統計の特徴。検定や多変量解析がそれに当たる。

☆視覚的なデータの要約 (**graphical summary**)

1) データをそのままグラフで表す方法

◆ 棒グラフ **bar graph**: 数量の大小、複数系列 OK

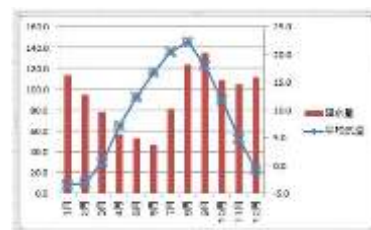
◆ 折れ線グラフ **line graph**:

経時変化→横軸は時間軸、縦軸は数量

移動平均 moving average (前後 3 日、5 年を平均しプロット=3 日、5 年移動平均など) → 長期的な傾向 (トレンド) をみる

◆ 複合グラフ **combination chart** (棒+折れ線)

◆ 円グラフ **pie graph**、帯グラフ **band graph**: 全体に対する割合の表示



2) 集計結果をグラフ、表に表す

◆ 幹葉図 **stem-and-leaf-plot**:

95、127、324、298、201、

143、17、300

幹葉表示

0 | 17 95

1 | 27 43

2 | 01 98

3 | 00 24

- ◆ 度数分布表 frequency table

度数 frequency:

全体をいくつかの階級に分けたとき、それぞれの範囲内の個数

$$\text{相対度数 relative frequency} = \frac{\text{度数}}{\text{データの総件数}}$$

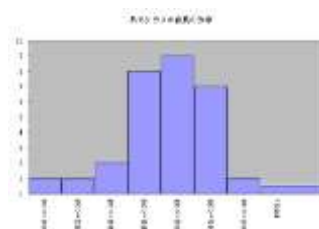
- ◆ ヒストグラム histogram:

縦軸に度数を、横軸にデータ区間(階級)をとり表す。(横軸の隙間をあけない)

データ数 N , 階級数 k のとき $k = 1 + \log_2 N$

(スタージェスの公式)

N	8	16	32	64	128	256	512	1024
k	4	5	6	7	8	9	10	11

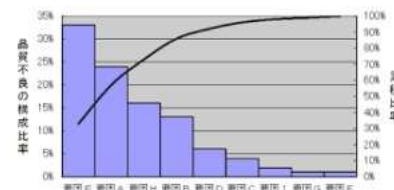


- ◆ パレート図 Pareto chart:

度数(棒グラフ) + 累積相対度数(折れ線グラフ)

項目別に層別して、出現頻度の大きさの順に

並べるとともに、累積和を示した図

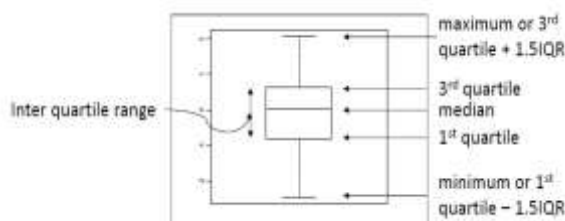


- ◆ クロス集計表 cross-tabulation table

= 分割表 (contingency table): 二変量の度数分布表

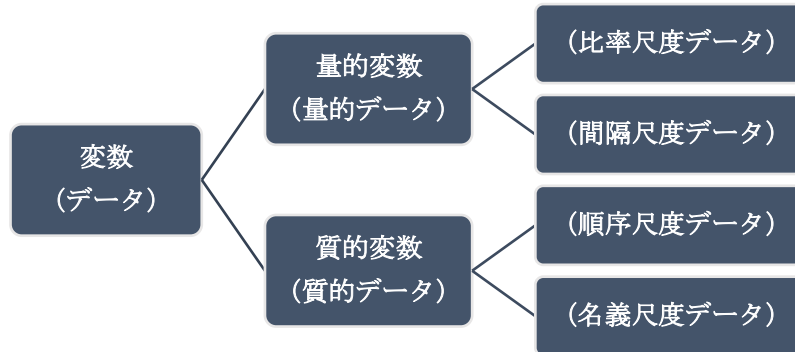
- ◆ 箱ひげ図 Box-plot ⇒ 四分位数 (p7)

5種の要約統計量: 最小値 min. または $(Q_1 - 1.5IQR)$ 、第1四分位点 1st quartile (Q_1)、中央値 median、第3四分位点 3rd quartile (Q_3)、最大値 max. または $(Q_3 + 1.5IQR)$ を表現する。①縦軸に変数値をとる。②下限が (Q_1) 、上限が (Q_3) となる "Box" を描く。③ median の位置に線を描く。④ Box の上下辺から max. または $(Q_3 + 1.5IQR)$, min. または $(Q_1 - 1.5IQR)$ まで線を引く。⑤上下線を $(Q_1 - 1.5IQR, Q_3 + 1.5IQR)$ に選択した時はそれより離れた標本は外れ値 (Outlier) として点 (●や○) で、3 IQR より離れているものを極端な外れ値として (*、×) などで表す。



☆数量的なデータの要約 (numerical summary)

データの種類



4つの尺度水準

名義尺度 **nominal scale**: **nominal data=categorical data**

区別することに意味がある ⇒性別、血液型、国籍、色、模様、都道府県
数学が好きであるとき 1、嫌いであるとき 0 を割り振る場合

*二値データ **binary data**→分割表(Row×Column 表)

順序尺度 **ordinal scale**: 間隔には意味がなく順序に意味がある

差には数値的な意味はなく中央値が意味を持つ

⇒順序(満足度、順位、等級など)、大中小、著効・有効・無効の判定
レースの着順、好き嫌いの 5 段階アンケート

間隔尺度 **interval scale**: データの間隔に意味がある

差には意味があるが、比には意味がない

平均値、中央値、最頻値、標準偏差などの統計量が意味を持つ
⇒温度、テストの点数、時刻、測定・検査結果など

比率尺度 **ratio scale**: 原点(0 という値)と比率に意味がある

比率(データの掛け算・割り算)に意味がある→データの加減乗除 OK
⇒時間(経過時間)、速度、身長、体重、血圧など

ほとんどの統計量が意味を持ち、用いることができる分析手法が多い

変数の分類: 尺度の水準による(実際に観察された変数の値がデータ)

質的変数 **qualitative variable**(binary,nominal,ordinal)

量的変数 **quantitative variable**(interval,ratio)

量的変数のもう1つの分類:

連続変数 **continuous variable** ⇒身長、体重

離散変数 **discrete variable**: 整数で数える ⇒生存数、死亡数

*統計では **variable** を変数と変量のどちらの訳で表現しても良い

基本統計量

平均 mean $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

標本分散 sample variance, s^2 $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

標本分散 sample variance, s^2 、不偏分散 unbiased variance, U^2 と区別し
 $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$, $U^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ と表す場合もある。

(標本)標準偏差 standard deviation

$$S.D = s = \sqrt{s^2}$$

偏差平方和 sum of squared deviation, sum of squares $\sum_{i=1}^n (x_i - \bar{x})^2$

標準誤差 standard error

$$S.E = \frac{s}{\sqrt{n}}$$

分位数 quantile

とは? \Rightarrow p10

最小値 minimum(value)

最大値 maximum(value)

中央値(メジアン) median

四分位数 quartile points \Rightarrow 箱ひげ図(p5)、正規確率プロット(p10)

データを小さい順に並べて、下から $1/4$ のところのデータを第1四分位数 first quartile(Q_1)、 $2/4$ のところのデータを第2四分位数(中央値と同じ)、 $3/4$ のところのデータを第3四分位数 third quartile(Q_3)という。

Q_1 、 Q_2 、 Q_3 をまとめて、四分位数 quartile、($Q_3 - Q_1$) の値を四分位範囲

第1四分位数 first quartile = 25 パーセンタイル値 top 25th percentile

第2四分位数 = median

第3四分位数 third quartile = 75 パーセンタイル値 bottom 25th percentile

四分位範囲 inter quartile range IQR, Qr $Qr = Q_3 - Q_1$

データ: 4,5,⑨,11,12,14,15,⑰,18,22 の時

全体の真ん中 median=13、前半部分の真ん中 $Q_1=9$,

後半部分の真ん中 $Q_3=17$ IQR=17-9=8

変動係数 coefficient of variation

$$CV = \frac{\sigma}{\mu} \quad CV = \frac{s}{\bar{x}} \quad \text{単位なし、別の変量でもバラつき比較可}$$

尖度 kurtosis 歪度 skewness

Kurtosis(尖度) $\alpha_4 = 3$ (正規分布) $\alpha_4 < 3$ (裾広がり) $\alpha_4 > 3$ (山が尖る)

Skewness(歪度) $\alpha_3 = 0$ (正規分布) $\alpha_3 < 0$ (山が右寄り) $\alpha_3 > 0$ (山が左寄り)

◇ 標本データのあらわし方

Mean±SDか、 Mean±SEか、 Mean± 95%CIか？

- Mean ± SE (Standard error):標準誤差= 標本平均の標準偏差
二群以上を比較するときは、平均を比較しているのでMean ± SE
Meanが効果を表すときは、その効果の範囲はSEのほうがよい
(平均値の精度を示す目的のとき使用)
- Mean ± SD (Standard deviation): SDは各データのちらばり
Mean ± SDの範囲に、データ全体の60~70%が分布している。
Meanが毒性を表すときは、その毒性のちらばりをSDで表すほうがよい
(群ごとのデータのばらつきを示す目的のとき)
- Mean± 95%CIで表す CI: confidence interval、信頼区間
Mean± 95%CI とは Mean± t(0.025) SE
nが大きければ t(0.025) =1.96
- Median (interquartile range, IQR): 中央値(Median)を中心に IQR の範囲にデータ全体の 50%が分布している。
分布が歪んでいるときは、Median (IQR) が第一選択

◇ データの標準化、分布

$$Z_i = \frac{x_i - \bar{x}}{s} \quad Z_i: Z\text{値、標準化得点}$$

平均値を0、分散を1になる分布(標準正規分布)中に各データを変換すること

$$\text{cf) 偏差値} = \frac{\text{得点} - \text{得点の平均値}}{\text{得点の標準偏差}} \times 10 + 50 = 10Z_i + 50$$

正規分布に従うデータの重要な性質

$x_1, x_2, \dots, x_i, \dots, x_n$, が正規分布 $N(\mu, \sigma^2)$ に従う時、標準化した

$$Z_i = \frac{x_i - \mu}{\sigma} \text{ は標準正規分布 } N(0, 1^2) \text{ 従う}$$

標本平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ は正規分布 $N(\mu, \frac{\sigma^2}{n})$ に従う

標本平均の標準偏差は $\frac{\sigma}{\sqrt{n}}$ であり標準誤差(standard error:SE)ともいい

$$\text{その標準化された } Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \text{ は標準正規分布 } N(0, 1^2) \text{ に従う}$$

標本平均の分布

- ある母集団から無作為抽出された標本平均はサンプルのサイズを大きくすると真の平均に近づく。(大数の法則)
- 任意の母集団(正規分布でなくてもよい)からの標本平均 \bar{X} は正規分布する。
標準化された $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ は標準正規分布 $N(0, 1^2)$ に従う(中心極限定理)

◇ 標本を得たときの考え方

データを得たら、そのデータがどのような分布に従うのかを推測する

正規分布に従うか否か(正規性の検定)

正規分布に従わなければ、標本数が多いか少ないかにより解析法を検討

サンプルが正規分布に従うか否か(正規性の検定)

Kurtosis(尖度) $\alpha_4 = 3$ (正規分布) $\alpha_4 < 3$ (裾広がり) $\alpha_4 > 3$ (山が尖る)

Skewness(歪度) $\alpha_3 = 0$ (正規分布) $\alpha_3 < 0$ (山が右寄り) $\alpha_3 > 0$ (山が左寄り)

χ^2 検定

コルモゴロフ・スミルノフ検定 (Kolmogorov-Smirnov Test):

観測データの累積率と理論的な確率分布関数を比較する検定

尖度、歪度、サンプル数を元に計算、サンプル数2000以上

シャピロ・ウィルク検定 (Shapiro-Wilk test): データが正規母集団に由来するという

帰無仮説を検定。シャピロ・ウィルクの統計量 W 値による。

サンプル数が 2000 以下

ダゴスティーノ検定 (D'Agostino's K-squared test)

尖度、歪度による検定。

サンプル数が少なくとも正規性を検出できる(20 以上)

正規性の有無:
尖度・歪度
正規確率プロット
正規性の検定 $p \geq 0.05$

*どの検定でも $p < 0.05$ では正規性なし、 $p \geq 0.05$ なら正規性がないとは言えない。

* Smirnov-Grubbs その他の棄却検定: 飛び離れた値で記載ミス、データの混入などが疑われる際、そのデータ棄却できるかどうか?を検定 \Rightarrow p13

*オムニバス検定: 歪度と尖度など、オムニバス形式に組合せ精度上昇を求める。分散分析など水準間の「どこか」に差があるといえるなど(別に多重比較が必要)もオムニバス検定と言われる。

* PC 上で数字を下記のように表すことがある。

1000 は $1.0E+3$ 、0.01 は $1.0E-2$ 、 $P=1.02191E-05$ は 0.0000102191 のこと

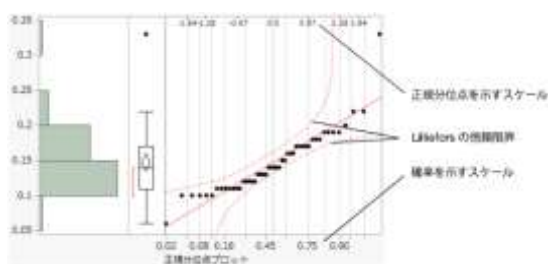
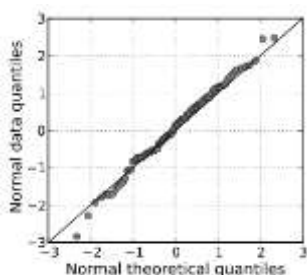
正規確率プロット normal probability plot

データの分布が正規分布に従っているかどうかを調べるためのプロット。一直線上に並んでいれば正規分布。

- ①各データの順位をつける。RANK.EQ 関数 =RANK.EQ(セル, セル範囲, 順序)
- ②順位を確率に変換 → %順位=(順位 - 1/2) / n または、順位 / n+1 など
- ③累積分布(%順位)から z 値を求める =NORM.S.INV(%順位)

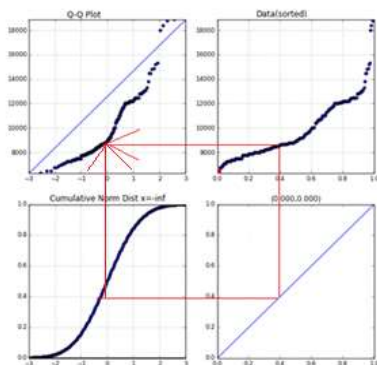
NORMINV 関数を使用すれば、累積分布(%順位)から、対応する期待値算出

- ④(縦軸、横軸)に(実データ、期待値または z 値) (実データ、%順位)など



下図は分位数を簡易的に正規分布の累積分布関数を利用してz値に変換した方法で <http://qiita.com/kenmatsu4/items/59605dc745707e8701e0> に紹介されています。

下図の右上は横軸に 180 個のあるデータを小さい順位 (1...,180) に等間隔に並べ、%順位=(順位 - 1/2) / n または、順位 / n+1 を求め最小値を第 1、180 分位数、最大値が第 180、180 分位数(=1、180 分位数)とすると 0~1 を等分した分位数は率で表した順位となる。縦軸は生データとする。左下は正規分布の累積分布関数で、縦軸 0~1 の数値(累積確率)は分位数を意味し、横軸はz値を表す。右下のような位置においた 1 辺 1 の正方形の対角線と左下図で横軸を変換し、縦軸はデータで表したものが一般的な Q-Q プロット(左上)。Q-Q プロットの"Q"とは **Quantile(分位数)**。左上の縦軸データをz変換した値とすれば標準化した正規 Q-Q プロット、また、縦軸を正規分布の累積確率、横軸を観測値の累積確率とすると正規 P-P プロットとなる。



分位数 quantile とは？四分位数 quartile、パーセンタイル percentile、などがあり、 q 分位数 (q -quantile)は分布を $1:(1-q)$ に分割する値

1標本の分析

1標本の分析:Z 検定、一標本 t 検定、二項分布

- 母平均、母分散が既知なら Z 検定

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

ある年、全国の小学3年(8歳)男子の身長は平均値 128.0 cm(標準偏差 5.44 cm)、A校3年生男子 25人の平均身長は 126.5 cmであった。低すぎか?

$$Z = \frac{126.5 - 128}{\frac{5.44}{\sqrt{25}}} = \frac{-1.5}{\frac{5.44}{5}} = \frac{-1.5}{1.088} = -1.38$$

標準正規分布表から(P=0.084、有意ではない)

Z 検定では、両側検定で Z 値の絶対値が 1.96, 片側検定で Z 値の絶対値が 1.65 以上なら有意水準 $\alpha=0.05$ での棄却域(有意)、以下なら有意差なし

*エクセル関数 =NORM.S.DIST(-1.38,TRUE)=0.083793(両側では P=0.1680)

BellCurve for Excel: 1 標本の推定と検定 → 母平均の検定 → 入力(サンプルサイズ、平均、標準偏差)、要約データ、正規分布、比較値

Z検定	
母平均の検定：正規分布	
変数	変数1
n	25
不偏分散	29.5936
標準偏差	5.4400
標準誤差	1.0880
平均	128.0000
比較値	126.5000
差	1.5000
統計量:z	1.3787
両側P値	0.1680
* : P<0.05 ** : P<0.01	
片側P値	0.0840
* : P<0.05 ** : P<0.01	

Z 値は標準正規分布図
での横軸の値

- 母平均既知、母分散が未知なら 一標本 t 検定

$$\bar{X} = \left(\frac{x_1 + x_2 + \dots + x_n}{n} \right), S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, t = \frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}}$$

ある年全国の小学3年(8歳)男子の身長は平均値128.0cmであり、A校3年生男子25人の平均身長は126.5cm(標準偏差5.44cm)であった。低すぎか？

$$t = \frac{126.5-128}{\frac{5.44}{\sqrt{25}}} = \frac{-1.5}{\frac{5.44}{5}} = \frac{-1.5}{1.088} = -1.38$$

t 分布表から自由度 25-1=24 の片側 $\alpha=0.05$ での $t_{24}(0.05)=1.711$
 $1.38 < 1.711$ より有意ではない

*エクセル関数 =T.DIST.2T(1.38,24)=0.180308(両側)

* $n > 30(100)$ の時は正規分布表をみる(片側 1.65, 両側 1.96 以上で有意)

BellCurve for Excel: 1 標本の推定と検定→母平均の検定→入力(サンプルサイズ、平均、標準偏差)、要約データ、t分布、比較値

t 検定	
母平均の検定: t分布	
変数	変数1
n	25
不偏分散	29.5936
標準偏差	5.4400
標準誤差	1.0880
平均	126.5000
比較値	128.0000
差	1.5000
統計量:t	1.3787
自由度	24
両側P値	0.1807
*: P<0.05 **: P<0.01	
片側P値	0.0904
*: P<0.05 **: P<0.01	

t 値は t 分布図での
横軸の値

- 二項分布: 比率は二項分布し数が多いと正規分布で近似できる

$$\frac{x}{n} - 1.96\sqrt{\frac{p(1-p)}{n}} < p < \frac{x}{n} + 1.96\sqrt{\frac{p(1-p)}{n}}$$

大統領選挙にA, Bの2人の人が立候補した無作為に3000人にアンケートしA候補が60%の支持を得ていた。どのくらいの信頼度か？

$$p = 0.6 \pm 1.96\sqrt{0.6 * \frac{0.4}{3000}} = 0.60 \pm 0.018$$

BellCurve for Excel: 1 標本の推定と検定→母比率の推定→入力(サンプルサイズ、比率)、要約データ、正規分布、無限母集団

母比率の推定：正規分布	
変数	変数1
n	3000
下限値	58.25%
比率	60.00%
上限値	61.75%
信頼度	95%
母集団サイズ	無限

*飛び離れた値の取り扱い: Smirnov-Grubbs の棄却検定

$$\text{統計量 } T_1 = \frac{|X_1 - \bar{X}|}{\sqrt{s^2}} = \frac{X_1 - \bar{X}}{s}$$

例)

検査結果	66, 63, 59, 57, 61, 58, 28, 60, 52, 48
------	--

平均 $\bar{X} = 55.2$ 分散 $s^2 = 118.0$ 棄却すべきかどうかの値 $X_1 = 28$

$$T_1 = \frac{|28 - 55.2|}{\sqrt{118}} = \frac{27.2}{10.9} = 2.57$$

Smirnov-Grubbs の棄却限界値(表:成書参照)データ数 $N=10$, 有意点 $\alpha=0.05$ の値 2.176 と比較し有意差あり、棄却可能

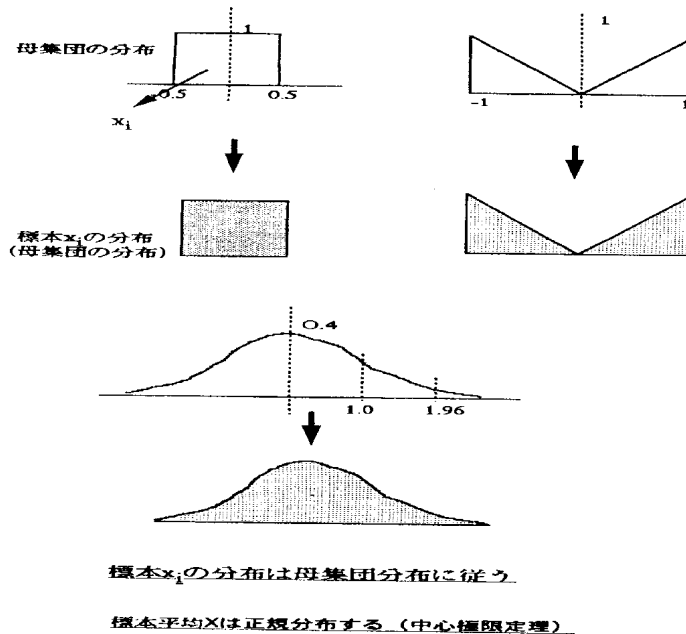
BellCurve for Excel: 基本統計・相関→外れ値検定→ Smirnov-Grubbs、最小値外れ値

外れ値検定の結果													
データ	元データ	除外済データ	外れ値	検定方法	Smirnov-Grubbs (片側検定 / 対立仮説: 最小値は外れ値である)								
No.	元データ	除外済データ	外れ値	有意水準	0.05								
1	66	66											
2	63	63		要約	元データ	除外済データ	外れ値						
3	59	59		n	10	9	1						
4	57	57											
5	61	61		検定過程	n	平均	不偏分散	標準偏差	最小値	最大値	No.	P 値	検定結果
6	58	58		1回目	10	55.2000	117.9556	10.8607	28.0000	-	7	0.0039	棄却
7	28		28	2回目	9	58.2222	29.9444	5.4722	48.0000	-	10	0.1601	保留
8	60	60											
9	52	52											
10	48	48											

棄却する場合は1回の検定のみとし、棄却理由を明記する。

母集団の分布(確率分布)について

- ◇ 各データはもとの分布(母集団:確率分布)に従う



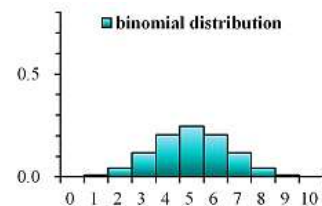
- ◇ 統計量(平均値、分散など)はその統計量の理論分布に従う

母集団:離散型確率分布 (Discrete probability distributions)

二項分布 binominal distribution

サイコロを30回(n回)振った時1の目は何回(x回)出るか

$$p(x) = {}_n C_x p^x (1-p)^{n-x} \quad , \quad \mu = np \quad , \quad \sigma^2 = np(1-p)$$

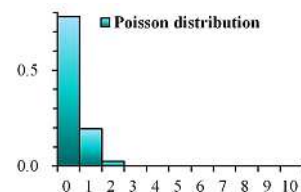


ポアソン分布 Poisson d.

一定時間内に起きる事象の回数の分布:プロイセン陸軍で1年間当たり馬に蹴られて死亡した兵士数の20年間の分布($\lambda = 0.61$ のポアソン分布)

X : 稀な現象、発生回数

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad , \quad \mu = \lambda \quad , \quad \sigma^2 = \lambda$$



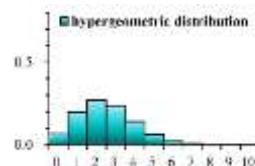
超幾何分布 hypergeometric d.

赤い玉 10 個 (A 個) と白い玉 20 個 (N-A 個) を混ぜた、計 30 個 (N 個) の中から 5 個 (M 個) の球を取りだすとき、赤い玉がちょうど 1 つ (X 個) である分布。

* Fisher 検定、log-rank 検定、一般化 Wilcoxon 検定の分布

$$p(x) = \frac{A^C x \cdot N-A^C M-x}{N^C M}$$

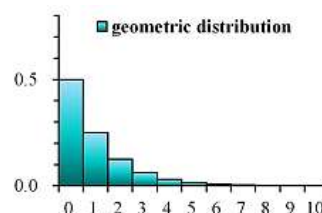
$$\mu = \frac{A \cdot M}{N} \quad , \quad \sigma^2 = \frac{AM(N-A)(N-M)}{N^2(N-1)}$$



幾何分布 geometric d.

同じコインを投げるとき何回目に初めて表がでるか

$$p(x) = p(1-p)^{x-1} \quad , \quad \mu = \frac{1}{p} \quad , \quad \sigma^2 = \frac{1-p}{p^2}$$

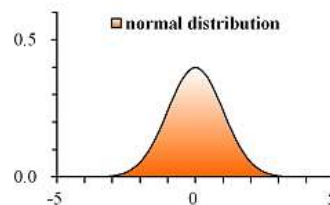


母集団:連続型確率分布 (Continuous probability distributions)

正規分布 $N(\mu, \sigma^2)$ normal d.

誤差の分布

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , \quad \text{平均 } \mu \quad , \quad \text{分散 } \sigma^2$$

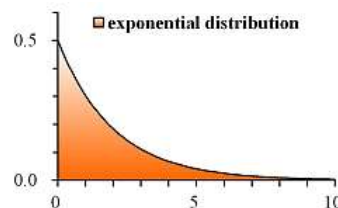


指数分布 exponential d.

ポアソン分布に従う事象の起こる時間間隔の分布

銀行の窓口で客が到着する時間間隔

$$f(x) = \lambda e^{-\lambda x} \quad , \quad \mu = \frac{1}{\lambda} \quad \sigma^2 = \frac{1}{\lambda^2}$$



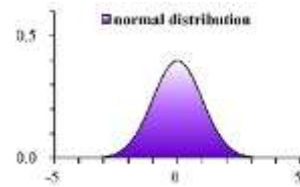
超幾何分布は Fisher 検定、log-rank 検定、
一般化 Wilcoxon 検定、Mantel-Haenszel 検
定などの理論的背景になっています。

母数推定や仮説検定などの推測統計で用いられる確率分布

標準正規分布: standard normal d. 統計量 z 値

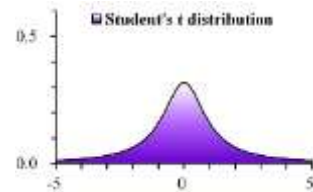
大標本、母分散が既知 の時の信頼区間の推定, 平均の検定

$$f(x) = N(0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$



t 分布: Student's t distribution 統計量 t 値

小標本、母分散が未知 の時の信頼区間の推定, 平均の検定

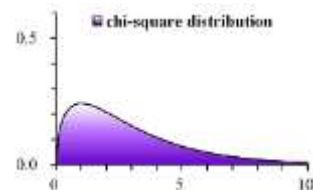


χ^2 分布 chi-square d.

χ^2 分布に従う統計量

χ^2 検定 (適合度検定・独立性検定)

$$\chi^2 = \sum_{i=1}^n z_i^2$$



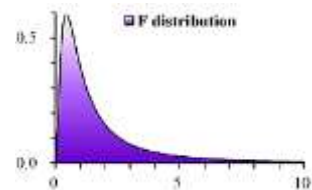
F 分布 F distribution 統計量 F 値

要因の分散/誤差の分散など

t_n : 自由度 n の t 分布、 χ_n^2 : 自由度 n の χ^2 分布、

F_n^1 : 自由度(1,n)にF分布、 $N(0,1)$: 標準正規分布のとき

$$t_n = \frac{N(0,1)}{\sqrt{\frac{\chi_n^2}{n}}} \quad F_n^1 = t_n^2 \text{ の関係あり}$$



分析の2方向

- 1) データから分布に関する確率モデル(母集団)が想定できる場合(二項分布、ポアソン分布など)や、まったく情報がない場合に確率モデルを想定し数理統計的に最尤法などを用いて証明する方法
- 2) 大数の法則、中心極限定理などの標本分布の理論を基礎に正規分布、そこから理論的に導かれる t 分布、 χ^2 分布、F 分布などを用いて証明する方法

正規分布 $N(\mu, \sigma^2)$

1. 誤差の分布
2. 正規分布から χ^2 分布・t分布・F分布が導かれる。
3. 二項分布はnが大きくなれば正規分布で近似できる
4. ある母集団から無作為抽出された標本平均はサンプルのサイズを大きくすると真の平均に近づく。「大数の法則」
5. 標本平均と真の平均との誤差は、母集団の分布がどんな分布であっても、サンプルのサイズを大きくしたとき近似的に正規分布に従う。
⇒母集団の分布が、どんな形であれ、そこから取り出した標本の標本平均は正規分布に従う。「中心極限定理」

正規分布 $N(\mu, \sigma^2)$

平均 μ と 分散 σ^2 を母数

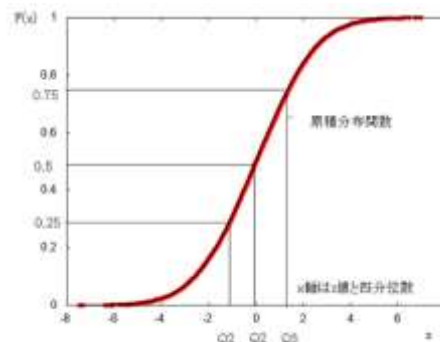
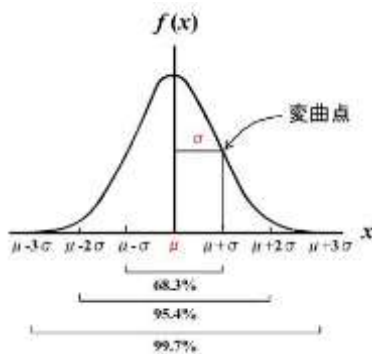
確率密度関数 $f(x)$ は $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

正規分布とデータの標準化 $z = \frac{x-\mu}{\sigma}$

正規分布データは
 $\mu \pm \sigma$ 内に 68.3%
 $\mu \pm 1.96\sigma$ 内に 95%
 含まれる

標準正規分布 $N(0, 1^2)$

確率密度関数 $f(x)$ は $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

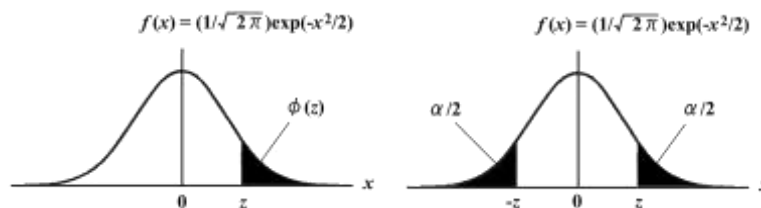


標準正規分布表

標準正規分布 $N(0,1^2)$ の確率密度関数 $f(x)$

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ において確率変数 x の値を z から ∞ まで積分した部分の面積

$\phi(z)$ の値を確率変数 z に対して記載したもの



正規分布	z	0	1	2	3	4	5	6	7	8	9
	0	0.5	0.496011	0.492022	0.488033	0.484047	0.480061	0.476078	0.472097	0.468119	0.464144
	0.1	0.460172	0.456205	0.452242	0.448283	0.44433	0.440382	0.436441	0.432505	0.428576	0.424655
	0.2	0.42074	0.416834	0.412936	0.409046	0.405165	0.401294	0.397432	0.39358	0.389739	0.385908
	0.3	0.382089	0.378281	0.374484	0.3707	0.366928	0.363169	0.359424	0.355691	0.351973	0.348268
	0.4	0.344578	0.340903	0.337243	0.333598	0.329969	0.326355	0.322758	0.319178	0.315614	0.312067
	0.5	0.308538	0.305026	0.301532	0.298056	0.294598	0.29116	0.28774	0.284339	0.280957	0.277595
	0.6	0.274253	0.270931	0.267629	0.264347	0.261086	0.257846	0.254627	0.251429	0.248252	0.245097
	0.7	0.241964	0.238852	0.235762	0.232695	0.22965	0.226627	0.223627	0.22065	0.217695	0.214764
	0.8	0.211855	0.20897	0.206108	0.203269	0.200454	0.197662	0.194894	0.19215	0.18943	0.186733
	0.9	0.18406	0.181411	0.178786	0.176186	0.173609	0.171056	0.168528	0.166023	0.163543	0.161087
σ	1	0.158655	0.156248	0.153864	0.151505	0.14917	0.146859	0.144572	0.14231	0.140071	0.137857
	1.1	0.135666	0.1335	0.131357	0.129238	0.127143	0.125072	0.123024	0.121001	0.119	0.117023
	1.2	0.11507	0.11314	0.111233	0.109349	0.107488	0.10565	0.103835	0.102042	0.100273	0.098525
	1.3	0.096801	0.095098	0.093418	0.091759	0.090123	0.088508	0.086915	0.085344	0.083793	0.082264
	1.4	0.080757	0.07927	0.077804	0.076359	0.074934	0.073529	0.072145	0.070781	0.069437	0.068112
	1.5	0.066807	0.065522	0.064256	0.063008	0.06178	0.060571	0.05938	0.058208	0.057053	0.055917
	1.6	0.054799	0.053699	0.052616	0.051551	0.050503	0.049471	0.048457	0.04746	0.046479	0.045514
	1.7	0.044565	0.043633	0.042716	0.041815	0.040929	0.040059	0.039204	0.038364	0.037538	0.036727
	1.8	0.03593	0.035148	0.034379	0.033625	0.032884	0.032157	0.031443	0.030742	0.030054	0.029379
	1.9	0.028716	0.028067	0.027429	0.026803	0.02619	0.025588	0.024998	0.024419	0.023852	0.023295
2σ	2	0.02275	0.022216	0.021692	0.021178	0.020675	0.020182	0.019699	0.019226	0.018763	0.018309

エクセル関数=NORM.S.DIST(Z,TRUEor FALSE)

χ²分布

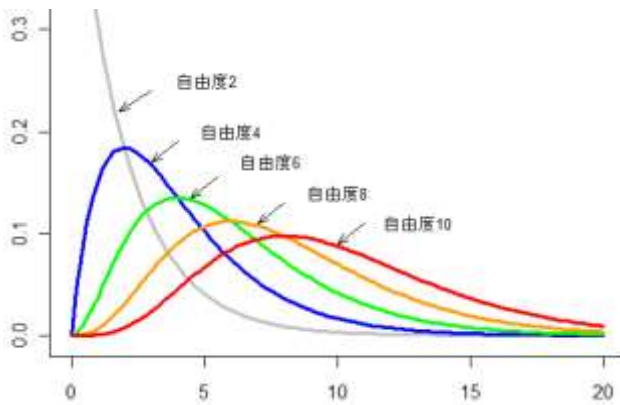
正規母集団 $N(\mu, \sigma^2)$ から抽出した n 個の独立な変数を $x_1, x_2, \dots, x_i, \dots, x_n$ とすると

$$\chi^2 = \left(\frac{x_1 - \mu}{\sigma}\right)^2 + \left(\frac{x_2 - \mu}{\sigma}\right)^2 + \dots + \left(\frac{x_n - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\chi^2 = \sum_{i=1}^n z_i^2$$

を自由度 $df = n$ の χ^2 分布という。

また、 χ^2 分布には加性があり、 $\chi_1^2, \chi_2^2, \dots, \chi_n^2$ がそれぞれ自由度 df_1, df_2, \dots, df_n の互いに独立な χ^2 分布に従うときは $\chi^2 = \chi_1^2 + \chi_2^2 + \dots + \chi_n^2$ は自由度 $df = df_1 + df_2 + \dots + df_n$ の χ^2 分布に従う。



自由度は平均値などのパラメーターを推定するたびに一つ減る

χ²適合度検定 chi-square goodness of fit test

適合度検定とは、観測度数分布が期待度数分布と同じかどうかを統計的に確かめる方法。ただし自由度 df は理論度数によって異なり、一様分布のとき、 $df = k - 1$ 、(二項 or ポアソン分布のとき、 $df = k - 2$ 、正規分布のとき、 $df = k - 3$)、 k は分類した数(セルの数)である。また一様分布とは、全ての区間を同等の値をとる分布である。

問題)サイコロを 60 回振ったとき、目の出た度数は下のようになった。このとき、目の出方に差があると考えてよいか。

	1	2	3	4	5	6	
観測度数 O_i	13	8	14	11	6	8	60
期待度数 E_i	10	10	10	10	10	10	60

帰無仮説(H₀):サイコロの目の出方に差はない

対立仮説(H₁):サイコロの目の出方に差がある

サイコロを 60 回振ったとき、「サイコロの目の出方に差はない」と考えるなら、予想される期待度数は上の表のようにそれぞれ 60/6=10 となる。

後は、各セルの観測度数と期待度数の偏りを計算して足せばよい。

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(13-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(8-10)^2}{10} = 5$$

この分布は全ての区間を同等の値をとる一様分布であるため、自由度 df は k-1=6-1=5 である。つまり、自由度 5 の χ^2 分布に従う。

χ^2 分布表より、 $\chi^2_{0.05} = 11.07$ であると分かる。 $\chi^2 < \chi^2_{0.05}$ なので、 $P > 0.05$ となり帰無仮説を棄却できない。つまり、「サイコロの目の出方に差がある」とはいえない。

*エクセル関数 =CHISQ.DIST.RT(5,5)=0.41588

BellCurve for Excel: 集計表の作成と分析 → 適合度の検定 → 実測度数、期待度数

適合度の検定								
	1	2	3	4	5	6		
観測度数O _i	13	8	14	11	6	8	60	
期待度数E _i	10	10	10	10	10	10	60	
	観測度数 O _i	期待度数E _i						
	13	10						
	8	10						
	14	10	縦型のデータ形式					
	11	10						
	6	10						
	8	10						
適合度の検定								
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01					
5.0000	5	0.4159						

χ^2 独立性の検定 chi-square test of independence

(l×m分割表)

分割表の検定

3つのサービスコースとそれぞれの満足度(5点満点)の関連をまとめたクロス集計表がある。このクロス集計表から独立性の検定を行う。有意水準は5%とする。

BellCurveエクセル統計より「サービス満足度」

観測度数						
O_{ij}	満足度1	2	3	4	5	合計
サービス1	16	23	81	82	41	243
サービス2	30	84	11	4	22	151
サービス3	93	77	34	13	8	225
合計	139	184	126	99	71	619

データのパターンが同じである期待度数は、得られた分割表の横の合計、縦の合計、総度数を用いて次のように求める。

$$\text{期待度数} = \frac{\text{横の合計} \times \text{縦の合計}}{\text{総合計}} \quad \Rightarrow \text{行1列1: } \left(\frac{243 \times 139}{619} = 54.567 \right)$$

期待度数					
E_{ij}	満足度1	2	3	4	5
サービス1	54.567	72.233	49.464	38.864	27.872
サービス2	33.908	44.885	30.737	24.150	17.320
サービス3	50.525	66.882	45.800	35.985	25.808

$$\chi^2 = \sum_{j=1}^m \sum_{i=1}^l \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\begin{aligned} &= \frac{(16-54.6)^2}{54.6} + \frac{(23-72.2)^2}{72.2} + \frac{(81-49.5)^2}{49.5} + \frac{(82-38.9)^2}{38.9} + \frac{(41-27.9)^2}{27.9} \\ &+ \frac{(30-30.9)^2}{30.9} + \frac{(84-44.9)^2}{44.9} + \frac{(11-30.7)^2}{30.7} + \frac{(4-24.2)^2}{24.2} + \frac{(22-17.3)^2}{17.3} \\ &+ \frac{(93-50.5)^2}{50.5} + \frac{(77-66.9)^2}{66.9} + \frac{(34-45.8)^2}{48.8} + \frac{(13-36)^2}{36} + \frac{(8-26.8)^2}{25.8} \\ &= 267.5 \end{aligned}$$

χ^2 値は、自由度df=(l-1)(m-1)の χ^2 分布に従うということが知られている。度数データの行(あるいは列)のパターンが同じであるかどうかに関する仮説検定は、 χ^2 値を用いることが可能

独立性の検定			
カイ二乗値	自由度	P 値	*:P<0.05 **:P<0.01
267.5	8	p < 0.001	**

自由度8, 有意水準0.01の χ^2 値は20.09, 有意水準0.001の χ^2 値は26.12

Cramer's V:2つの質的変数間の関連の強さを表す指標。クラメール(Cramer)の連関係数は、0 から 1 の値をとる。2変数の関連において、0に近いほど弱く、1に近いほど強い。

*エクセル関数 =CHISQ.DIST.RT(267.5,8)=3.33903E-53

独立性が否定された
⇒ 関連がある

BellCurve for Excel:集計表の作成と分析→独立性の検定

独立性の検定						
観測度数						
	列1	列2	列3	列4	列5	合 計
行1	16	23	81	82	41	243
行2	30	84	11	4	22	151
行3	93	77	34	13	8	225
合 計	139	184	126	99	71	619
期待度数						
	列1	列2	列3	列4	列5	
行1	54.567	72.233	49.464	38.864	27.872	
行2	33.908	44.885	30.737	24.150	17.320	
行3	50.525	66.882	45.800	35.985	25.808	
			太字：1未満 赤字：5未満			
独立性の検定						
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01			
267.5155	8	p < 0.001	**			
Cramer's V	0.4649					

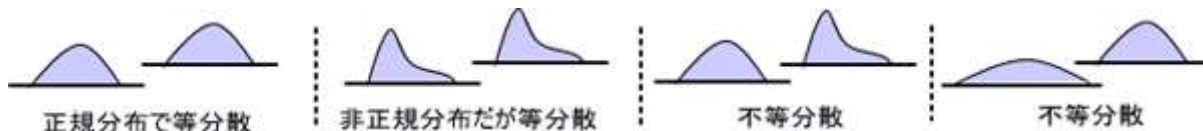
χ²分布表:たとえば自由度=1の場合は、5%水準では「3.84」、1%水準では「6.63」となる。(χ²=4.2で自由度が1の場合は、5%水準で有意差があり)

	有意確率													
df	0.99	0.98	0.95	0.9	0.8	0.7	0.5	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.00	0.00	0.00	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	5.41	6.63	10.83
2	0.02	0.04	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.61	5.99	7.82	9.21	13.82
3	0.11	0.18	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.81	9.84	11.34	16.27
4	0.30	0.43	0.71	1.06	1.65	2.19	3.36	4.88	5.99	7.78	9.49	11.67	13.28	18.47
5	0.55	0.75	1.15	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	13.39	15.09	20.52
6	0.87	1.13	1.64	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	15.03	16.81	22.46
7	1.24	1.56	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	16.62	18.48	24.32
8	1.65	2.03	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	18.17	20.09	26.12
9	2.09	2.53	3.33	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	19.68	21.67	27.88
10	2.56	3.06	3.94	4.87	6.18	7.27	9.34	11.78	13.44	15.99	18.31	21.16	23.21	29.59
11	3.05	3.61	4.57	5.58	6.99	8.15	10.34	12.90	14.63	17.28	19.68	22.62	24.72	31.26
12	3.57	4.18	5.23	6.30	7.81	9.03	11.34	14.01	15.81	18.55	21.03	24.05	26.22	32.91
13	4.11	4.77	5.89	7.04	8.63	9.93	12.34	15.12	16.98	19.81	22.36	25.47	27.69	34.53
14	4.66	5.37	6.57	7.79	9.47	10.82	13.34	16.22	18.15	21.06	23.68	26.87	29.14	36.12
15	5.23	5.98	7.26	8.55	10.31	11.72	14.34	17.32	19.31	22.31	25.00	28.26	30.58	37.70
16	5.81	6.61	7.96	9.31	11.15	12.62	15.34	18.42	20.47	23.54	26.30	29.63	32.00	39.25
17	6.41	7.26	8.67	10.09	12.00	13.53	16.34	19.51	21.61	24.77	27.59	31.00	33.41	40.79
18	7.01	7.91	9.39	10.86	12.86	14.44	17.34	20.60	22.76	25.99	28.87	32.35	34.81	42.31
19	7.63	8.57	10.12	11.65	13.72	15.35	18.34	21.69	23.90	27.20	30.14	33.69	36.19	43.82
20	8.26	9.24	10.85	12.44	14.58	16.27	19.34	22.77	25.04	28.41	31.41	35.02	37.57	45.31
21	8.90	9.91	11.59	13.24	15.44	17.18	20.34	23.86	26.17	29.62	32.67	36.34	38.93	46.80
22	9.54	10.60	12.34	14.04	16.31	18.10	21.34	24.94	27.30	30.81	33.92	37.66	40.29	48.27
23	10.20	11.29	13.09	14.85	17.19	19.02	22.34	26.02	28.43	32.01	35.17	38.97	41.64	49.73
24	10.86	11.99	13.85	15.66	18.06	19.94	23.34	27.10	29.55	33.20	36.42	40.27	42.98	51.18
25	11.52	12.70	14.61	16.47	18.94	20.87	24.34	28.17	30.68	34.38	37.65	41.57	44.31	52.62
26	12.20	13.41	15.38	17.29	19.82	21.79	25.34	29.25	31.79	35.56	38.89	42.86	45.64	54.05
27	12.88	14.13	16.15	18.11	20.70	22.72	26.34	30.32	32.91	36.74	40.11	44.14	46.96	55.48
28	13.56	14.85	16.93	18.94	21.59	23.65	27.34	31.39	34.03	37.92	41.34	45.42	48.28	56.89
29	14.26	15.57	17.71	19.77	22.48	24.58	28.34	32.46	35.14	39.09	42.56	46.69	49.59	58.30
30	14.95	16.31	18.49	20.60	23.36	25.51	29.34	33.53	36.25	40.26	43.77	47.96	50.89	59.70

エクセル関数:カイ²乗分布の右側確率の値=CHISQ.DIST.RT(X、自由度)

F 検定 F-test (等分散の検定)

「等分散とは？」 → 等しく分散していること → それぞれの群の分布の形が似ている



二標本(標本数 n_1, n_2) が等分散かどうかを調べる

帰無仮説(H_0): 「2 群間の分散に差がない(等分散である)」

対立仮説(H_1): 「2 群間の分散に差がある(等分散でない)」

最初にそれぞれの群の分散 s_1^2, s_2^2 を求め、次いで、下の式によって F 値を出す。

ただし、分子に大きい数値の方をとる

$$F = \frac{s_1^2}{s_2^2}$$

このとき、自由度は分子の自由度 $df_1 = n_1 - 1$ 、分母の自由度 $df_2 = n_2 - 1$ の F 分布に従う。F 分布表から F_α を求めることができる。

判定

$1 \leq F \leq F_\alpha$ のとき、 $P > 0.05$ となる → 帰無仮説を棄却できない → 等分散である。

$F > F_\alpha$ のとき、 $P < 0.05$ となる → 帰無仮説を棄却する → 不等分散である。

例) A クラスには 7 人、B クラスには 9 人の生徒がいる。この生徒たちに物理のテストを実施した。A クラスと B クラスの物理のテストの点は等分散かどうかを検定せよ。

A クラス	B クラス
60	49
52	40
68	52
55	37
65	55
47	38
45	45
62	
53	

	A クラス	B クラス
データ数	9	7
標準偏差	7.94	7.01
分散	63	50.5

$$F = 63/50.5 = 1.25$$

$F_{0.05}$ は分子の自由度 $df_1=9-1=8$ 、分母の自由度 $df_2=7-1=6$ の F 分布に従う。F 分布表より、 $F_{0.05}=4.15$ である。

$F = 1.25 < 4.15 = F_{0.05}$ なので、 $P > 0.05$ となり帰無仮説を棄却できない。「2 群間の分散に差がある」とは言えないため、等分散とすることができる。

*エクセル関数 =F.DIST.RT(1.25,8,6)=0.404、両側では 0.808

BellCurve for Excel: 2 標本の比較→等分散の検定

等分散性の検定					
基本統計量					
変数	n	平均	不偏分散	標準偏差	標準誤差
Aクラス	9	56.333	63.000	7.937	2.646
Bクラス	7	45.143	50.476	7.105	2.685
分散の比と効果量					
分散の比	効果量				
1.2481	1.2481				
母分散の比の95%信頼区間					
下限値	上限値				
0.2229	6.9890				
等分散性の検定 (両側検定)					
統計量:F	自由度1	自由度2	P 値	検出力($\alpha=0.05$)	
1.2481	8	6	0.8094	0.0749	

F 分布表 F distribution table

F 分布の上側確率 α ごとに数表がまとめられている。

自由度 m (分子)と自由度 n (分母)、対応する F 値(統計量)を読み取る。下側確率で見える場合には, α を下側確率として, 分子と分母の自由度を入れ替えて表中の数値を読み, その逆数を F 値とする。

F 分布表 上側確率 5% ($\alpha = 0.05$)

α 0.05	自由度 m (分子)												
	1	2	3	4	5	6	7	8	9	10	11	12	
自由度 n (分母)	1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98	243.91
	2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.405	19.413
	3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7633	8.7446
	4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9358	5.9117
	5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.7040	4.6777
	6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	4.0274	3.9999
	7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.6030	3.5747
	8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.3130	3.2839
	9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.1025	3.0729
	10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9430	2.9130
	11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.8179	2.7876
	12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.7173	2.6866
	13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6347	2.6037
	14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5655	2.5342
	15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.5068	2.4753
	20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.3100	2.2776
	25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1979	2.1649
	30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.1256	2.0921
	40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0376	2.0035
	50	4.0343	3.1826	2.7900	2.5572	2.4004	2.2864	2.1992	2.1299	2.0734	2.0261	1.9861	1.9515
	100	3.9361	3.0873	2.6955	2.4626	2.3053	2.1906	2.1025	2.0323	1.9748	1.9267	1.8857	1.8503
	200	3.8884	3.0411	2.6498	2.4168	2.2592	2.1441	2.0556	1.9849	1.9269	1.8783	1.8368	1.8008
500	3.8601	3.0138	2.6227	2.3898	2.2320	2.1167	2.0279	1.9569	1.8986	1.8496	1.8078	1.7715	
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7886	1.7522	

α 0.05	自由度 m												
	13	14	15	20	25	30	40	50	100	200	500	∞	
自由度 n													
1	244.69	245.36	245.95	248.01	249.26	250.10	251.14	251.77	253.04	253.68	254.06	254.31	
2	19.419	19.424	19.429	19.446	19.456	19.462	19.471	19.476	19.486	19.491	19.494	19.496	
3	8.7287	8.7149	8.7029	8.6602	8.6341	8.6166	8.5944	8.5810	8.5539	8.5402	8.5320	8.5264	
4	5.8911	5.8733	5.8578	5.8025	5.7687	5.7459	5.7170	5.6995	5.6641	5.6461	5.6353	5.6281	
5	4.6552	4.6358	4.6188	4.5581	4.5209	4.4957	4.4638	4.4444	4.4051	4.3851	4.3731	4.3650	
6	3.9764	3.9559	3.9381	3.8742	3.8348	3.8082	3.7743	3.7537	3.7117	3.6904	3.6775	3.6689	
7	3.5503	3.5292	3.5107	3.4445	3.4036	3.3758	3.3404	3.3189	3.2749	3.2525	3.2389	3.2298	
8	3.2590	3.2374	3.2184	3.1503	3.1081	3.0794	3.0428	3.0204	2.9747	2.9513	2.9371	2.9276	
9	3.0475	3.0255	3.0061	2.9365	2.8932	2.8637	2.8259	2.8028	2.7556	2.7313	2.7166	2.7067	
10	2.8872	2.8647	2.8450	2.7740	2.7298	2.6996	2.6609	2.6371	2.5884	2.5634	2.5481	2.5379	
11	2.7614	2.7386	2.7186	2.6464	2.6014	2.5705	2.5309	2.5066	2.4566	2.4308	2.4151	2.4045	
12	2.6602	2.6371	2.6169	2.5436	2.4977	2.4663	2.4259	2.4010	2.3498	2.3233	2.3071	2.2962	
13	2.5769	2.5536	2.5331	2.4589	2.4123	2.3803	2.3392	2.3138	2.2614	2.2343	2.2176	2.2064	
14	2.5073	2.4837	2.4630	2.3879	2.3407	2.3082	2.2664	2.2405	2.1870	2.1592	2.1422	2.1307	
15	2.4481	2.4244	2.4034	2.3275	2.2797	2.2468	2.2043	2.1780	2.1234	2.0950	2.0776	2.0658	
20	2.2495	2.2250	2.2033	2.1242	2.0739	2.0391	1.9938	1.9656	1.9066	1.8755	1.8562	1.8432	
25	2.1362	2.1111	2.0889	2.0075	1.9554	1.9192	1.8718	1.8421	1.7794	1.7460	1.7252	1.7110	
30	2.0630	2.0374	2.0148	1.9317	1.8782	1.8409	1.7918	1.7609	1.6950	1.6597	1.6375	1.6223	
40	1.9738	1.9476	1.9245	1.8389	1.7835	1.7444	1.6928	1.6600	1.5892	1.5505	1.5260	1.5089	
50	1.9214	1.8949	1.8714	1.7841	1.7273	1.6872	1.6337	1.5995	1.5249	1.4835	1.4569	1.4383	
100	1.8193	1.7919	1.7675	1.6764	1.6163	1.5733	1.5151	1.4772	1.3917	1.3416	1.3079	1.2832	
200	1.7694	1.7415	1.7166	1.6233	1.5612	1.5164	1.4551	1.4146	1.3206	1.2626	1.2211	1.1885	
500	1.7398	1.7116	1.6864	1.5916	1.5282	1.4821	1.4186	1.3762	1.2753	1.2096	1.1587	1.1132	
∞	1.7202	1.6918	1.6664	1.5705	1.5061	1.4591	1.3940	1.3501	1.2434	1.1700	1.1063	1.0000	

F 分布表 上側確率 2.5% ($\alpha = 0.025$)

$\alpha 0.025$	自由度 m											
	1	2	3	4	5	6	7	8	9	10	11	12
自由度 n												
1	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	973.03	976.71
2	38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398	39.407	39.415
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419	14.374	14.337
4	12.218	10.649	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7935	8.7512
5	10.007	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5678	6.5245
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.4098	5.3662
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.7095	4.6658
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.2434	4.1997
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.9121	3.8682
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6649	3.6209
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4737	3.4296
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.3215	3.2773
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120	3.2497	3.1975	3.1532
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093	3.1469	3.0946	3.0502
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	3.0078	2.9633
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.7209	2.6758
25	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	2.7531	2.6766	2.6135	2.5603	2.5149
30	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6513	2.5746	2.5112	2.4577	2.4120
40	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519	2.3882	2.3343	2.2882
50	5.3403	3.9749	3.3902	3.0544	2.8327	2.6736	2.5530	2.4579	2.3808	2.3168	2.2627	2.2162
100	5.1786	3.8284	3.2496	2.9166	2.6961	2.5374	2.4168	2.3215	2.2439	2.1793	2.1245	2.0773
200	5.1004	3.7578	3.1820	2.8503	2.6304	2.4720	2.3513	2.2558	2.1780	2.1130	2.0578	2.0103
500	5.0543	3.7162	3.1423	2.8114	2.5919	2.4335	2.3129	2.2172	2.1392	2.0740	2.0186	1.9708
∞	5.0239	3.6889	3.1161	2.7858	2.5665	2.4082	2.2875	2.1918	2.1136	2.0483	1.9927	1.9447

α 0.025	自由度 m												
	13	14	15	20	25	30	40	50	100	200	500	∞	
自由度 n													
1	979.84	982.53	984.87	993.10	998.08	1001.4	1005.6	1008.1	1013.2	1015.7	1017.2	1018.3	
2	39.421	39.427	39.431	39.448	39.458	39.465	39.473	39.478	39.488	39.493	39.496	39.498	
3	14.304	14.277	14.253	14.167	14.115	14.081	14.037	14.010	13.956	13.929	13.913	13.902	
4	8.7150	8.6838	8.6565	8.5599	8.5010	8.4613	8.4111	8.3808	8.3195	8.2885	8.2698	8.2573	
5	6.4876	6.4556	6.4277	6.3286	6.2679	6.2269	6.1750	6.1436	6.0800	6.0478	6.0283	6.0153	
6	5.3290	5.2968	5.2687	5.1684	5.1069	5.0652	5.0125	4.9804	4.9154	4.8824	4.8625	4.8491	
7	4.6285	4.5961	4.5678	4.4667	4.4045	4.3624	4.3089	4.2763	4.2101	4.1764	4.1560	4.1423	
8	4.1622	4.1297	4.1012	3.9995	3.9367	3.8940	3.8398	3.8067	3.7393	3.7050	3.6842	3.6702	
9	3.8306	3.7980	3.7694	3.6669	3.6035	3.5604	3.5055	3.4719	3.4034	3.3684	3.3471	3.3329	
10	3.5832	3.5504	3.5217	3.4185	3.3546	3.3110	3.2554	3.2214	3.1517	3.1161	3.0944	3.0798	
11	3.3917	3.3588	3.3299	3.2261	3.1616	3.1176	3.0613	3.0268	2.9561	2.9198	2.8977	2.8828	
12	3.2393	3.2062	3.1772	3.0728	3.0077	2.9633	2.9063	2.8714	2.7996	2.7626	2.7401	2.7249	
13	3.1150	3.0819	3.0527	2.9477	2.8821	2.8372	2.7797	2.7443	2.6715	2.6339	2.6109	2.5955	
14	3.0119	2.9786	2.9493	2.8437	2.7777	2.7324	2.6742	2.6384	2.5646	2.5264	2.5030	2.4872	
15	2.9249	2.8915	2.8621	2.7559	2.6894	2.6437	2.5850	2.5488	2.4739	2.4352	2.4114	2.3953	
20	2.6369	2.6030	2.5731	2.4645	2.3959	2.3486	2.2873	2.2493	2.1699	2.1284	2.1027	2.0853	
25	2.4756	2.4413	2.4110	2.3005	2.2303	2.1816	2.1183	2.0787	1.9955	1.9515	1.9242	1.9055	
30	2.3724	2.3378	2.3072	2.1952	2.1237	2.0739	2.0089	1.9681	1.8816	1.8354	1.8065	1.7867	
40	2.2481	2.2130	2.1819	2.0677	1.9943	1.9429	1.8752	1.8324	1.7405	1.6906	1.6590	1.6371	
50	2.1758	2.1404	2.1090	1.9933	1.9186	1.8659	1.7963	1.7520	1.6558	1.6029	1.5689	1.5452	
100	2.0363	2.0001	1.9679	1.8486	1.7705	1.7148	1.6401	1.5917	1.4833	1.4203	1.3781	1.3473	
200	1.9688	1.9322	1.8996	1.7780	1.6978	1.6403	1.5621	1.5108	1.3927	1.3204	1.2691	1.2290	
500	1.9290	1.8921	1.8592	1.7362	1.6546	1.5957	1.5151	1.4616	1.3356	1.2543	1.1918	1.1365	
∞	1.9027	1.8656	1.8326	1.7085	1.6259	1.5660	1.4835	1.4284	1.2956	1.2053	1.1277	1.0000	

エクセル関数:F 分布の上側確率(右端から確率変数までの分布の面積)

=F.DIST.RT(X, 分子の自由度、分母の自由度)

数理統計的側面(1)

この章では、標本分散 sample variance, s^2 、不偏分散 unbiased variance, U^2 と
区別し $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$, $U^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ と表す。

基本公式

期待値の公式

$$E(X) = \sum_{i=1}^n p_i x_i = \mu$$

$$E(aX) = aE(X), \quad E(aX + b) = aE(X) + b, \quad E(X + Y) = E(X) + E(Y)$$

分散の公式

$$V(aX) = a^2 V(X), \quad V(aX + b) = a^2 V(X)$$

$$V(X \pm Y) = V(X) + V(Y) \pm 2Cov(X, Y), \quad Cov(X, Y) = E(XY) - E(X)E(Y)$$

X, Y が無相関なら

$$V(X + Y) = V(X) + V(Y)$$

$$V(X) = \sigma^2 = E[(X - \mu)^2] = \sum_{i=1}^n p_i (x_i - \mu)^2$$

$$= \sum_{i=1}^n p_i x_i^2 - 2\mu \sum_{i=1}^n p_i x_i + \mu^2 \sum_{i=1}^n p_i = E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - (E(X))^2 = E(X^2) - \mu^2$$

$$\because \sum_{i=1}^n p_i = 1, \quad E(X) = \mu$$

平均値と期待値の違い

*サイコロ 10 回をふって (1, 3, 4, 5, 5, 2, 6, 4, 4, 2) のとき平均値は 3.6

$$\text{期待値 } E(X) = \left(1 \times \frac{1}{6}\right) + \left(2 \times \frac{1}{6}\right) + \left(3 \times \frac{1}{6}\right) + \left(4 \times \frac{1}{6}\right) + \left(5 \times \frac{1}{6}\right) + \left(6 \times \frac{1}{6}\right) = 3.5$$

参考) 関数の期待値

$$E\{h(X)\} = \sum_{i=1}^{\infty} h(x_i) p_i \quad \text{で定義される。}$$

$$\text{したがって } E(X^2) = \sum_{i=1}^{\infty} x_i^2 p_i$$

標本平均

期待値 μ 、分散 σ^2 の母集団からの標本 (x_1, x_2, \dots, x_n) のとき

標本平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

標本の期待値と分散

$$E(X) = \mu \rightarrow E(X_i) = \mu, V(X) = \sigma^2 \rightarrow V(X_i) = \sigma^2$$

標本平均の期待値と分散

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \sum_{i=1}^n \frac{1}{n} x_i = \mu \quad \text{または、} \quad E(\bar{X}) = \frac{1}{n} E\left(\sum_{i=1}^n x_i\right) = \frac{1}{n} n \mu = \mu$$

$$V(\bar{X}) = \frac{1}{n^2} V\left(\sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(x_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} * n \sigma^2 = \frac{\sigma^2}{n}$$

$$\text{また、} \quad V(\bar{X}) = E(\overline{X^2}) - (E(\bar{X}))^2 = \frac{\sigma^2}{n} \quad \text{より} \quad E(\overline{X^2}) = \frac{\sigma^2}{n} + \mu^2$$

標本分散 $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ の期待値の求め方

$$V(X) = \sigma^2 = E(X^2) - (E(X))^2 = E(X^2) - \mu^2 \text{より}$$

$$E(X^2) = \sigma^2 + \mu^2$$

$$E(s^2) = \frac{1}{n} E\left(\sum_{i=1}^n (x_i - \bar{x})^2\right) = \frac{1}{n} E\left[\sum_{i=1}^n \{(x_i - \mu) - (\bar{x} - \mu)\}^2\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu) \sum_{i=1}^n (x_i - \mu) + n(\bar{x} - \mu)^2\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^n (x_i - \mu)^2 - 2n(\bar{x} - \mu)^2 + n(\bar{x} - \mu)^2\right]$$

$$= \frac{1}{n} \{E[\sum_{i=1}^n (x_i - \mu)^2] - E[n(\bar{x} - \mu)^2]\}$$

$$= \frac{1}{n} \left\{n\sigma^2 - n \frac{\sigma^2}{n}\right\} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2$$

別法)

$$E(s^2) = \frac{1}{n} \{E[\sum_{i=1}^n (x_i - \mu)^2] - E[n(\bar{x} - \mu)^2]\} = E(X^2) - E(\overline{X^2})$$

$$= (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2$$

ここで標本不偏分散

$$U^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$E(U^2) = E\left(\frac{n}{n-1} s^2\right) = \frac{n}{n-1} E(s^2) = \sigma^2$$

推定量の期待値が真の母数と一致するとき、不偏推定量 unbiased estimator とい
い、 U^2 は母分散 σ^2 の不偏推定量(不偏分散)

正規分布に従うデータの重要な性質

$x_1, x_2, \dots, x_i, \dots, x_n$, が iid (independent and identically distributed) で
正規分布 $N(\mu, \sigma^2)$ に従う時、

それぞれのデータを標準化した $z_i = \frac{x_i - \mu}{\sigma}$ は標準正規分布 $N(0, 1^2)$ 従う

標本平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ は正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に従う

標本平均の標準偏差は $\frac{\sigma}{\sqrt{n}}$ であり標準誤差(standard error:SE)ともいい

その標準化された $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ は標準正規分布 $N(0, 1^2)$ に従う

χ^2 :カイ二乗の定義

$\chi^2 = \sum_{i=1}^n z_i^2$ とすると

$$\chi^2 = \left(\frac{x_1 - \mu}{\sigma}\right)^2 + \left(\frac{x_2 - \mu}{\sigma}\right)^2 + \dots + \left(\frac{x_n - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

は自由度 n の χ^2 分布に従う。

t 分布の定義:(μ を \bar{x} で置き換えたデータ平均を使うとき)

標本分散 U^2 (標本標準偏差 U)として

$U^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ とすると、 $\sum_{i=1}^n (x_i - \bar{x})^2 = (n-1) U^2$, σ^2 で割った

$\chi^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} = (n-1) \frac{U^2}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従う。

$$\frac{U^2}{\sigma^2} = \frac{1}{n-1} \left\{ \left(\frac{x_1 - \bar{x}}{\sigma}\right)^2 + \left(\frac{x_2 - \bar{x}}{\sigma}\right)^2 + \dots + \left(\frac{x_n - \bar{x}}{\sigma}\right)^2 \right\}$$

$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ は母分散 σ^2 が未知では計算できない。

そこで、 σ^2 を標本分散 U^2 (標本標準偏差 U)で代用した場合

$t = \frac{\bar{X} - \mu}{\frac{U}{\sqrt{n}}}$ は自由度 $k=n-1$ の t 分布 $t(n-1)$ に従う。

$\frac{U}{\sqrt{n}}$: 標準誤差 standard error

ここで、 $t = \frac{\bar{X} - \mu}{\frac{U}{\sqrt{n}}} = \frac{\bar{X} - \mu}{\frac{\sigma \sqrt{U^2}}{\sqrt{n} \sigma}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} * \frac{1}{\sqrt{\frac{(n-1)U^2}{\sigma^2} \frac{1}{(n-1)}}}$ と式を変形すると

$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ は $N(0,1)$ 、 $\frac{(n-1)U^2}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従う。

よって 正規分布、 χ^2 分布、 t 分布、自由度 k の関係は

$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ が $N(0,1)$ 、 $Y = \frac{(n-1)U^2}{\sigma^2}$ が自由度 $k=n-1$ の χ^2 分布に従うとき、
確率変数 $t = \frac{Z}{\sqrt{\frac{Y}{k}}}$ の従う確率分布を自由度 k の t 分布(Student's t): $t(k)$ という。
 k が大きいとき、 t 分布は $N(0,1)$ に近づき、 $k=\infty$ で標準正規分布 $N(0,1)$ に一致する。

点推定と区間推定

点推定 point estimation 母数を推定する

1) 期待値の推定

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad E(\hat{\mu}) = E(\bar{x}) = \mu$$

2) 分散の推定

$$\widehat{\sigma^2} = s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad E(s^2) = \frac{n-1}{n} \sigma^2$$

$$\widetilde{\sigma^2} = U^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad E(U^2) = \sigma^2$$

*母数の推定法には最尤法、最小二乗法などがある。

区間推定

L: 下側信頼限界、U: 上側信頼限界、

区間[L,U]: $100(1 - \alpha)\%$ 信頼区間 confidence interval

$1 - \alpha$: 信頼係数 (例、 $1 - \alpha = 0.9, 0.95, 0.99$ etc)、

有意水準(危険率) $\alpha = 0.1, 0.05, 0.01$, $\alpha/2 = 0.05, 0.025, 0.005$

覚えておくべき有意水準(危険率)

正規分布

片側で 5%(0.05)の有意水準(危険率) $\alpha = 0.05$ 両側では $\alpha/2 = 0.025$

正規母集団からの標本

1) $(x_1, x_2 \dots x_n)$ がi.i.dで $N(\mu, \sigma^2)$ に従う時(σ^2 が既知= σ^2 の時)、

標本平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ は正規分布 $N(\mu, \frac{\sigma^2}{n})$ に従うので、

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha \quad \text{となる } \frac{\alpha}{2} \text{ を正規分布表から求める。}$$

母平均 μ の $1 - \alpha$ での信頼区間は $\left[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$ となる。

2) σ^2 が未知の時

$U^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ とすると $t = \frac{\bar{x}-\mu}{\frac{U}{\sqrt{n}}}$ は自由度 $\nu=n-1$ の t 分布 $t(n-1)$ に従う

$$P\left(-t_{\frac{\alpha}{2}}(\nu) \leq \frac{\bar{X}-\mu}{\frac{U}{\sqrt{n}}} \leq t_{\frac{\alpha}{2}}(\nu)\right) = 1 - \alpha \quad \text{となる } t_{\frac{\alpha}{2}}(\nu) \text{ を } t \text{ 分布表から求める。}$$

母平均 μ の $1 - \alpha$ での信頼区間は

$$\left[\bar{X} - t_{\frac{\alpha}{2}}(\nu) \frac{U}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}}(\nu) \frac{U}{\sqrt{n}}\right] \text{ となる。}$$

自由度 ν が大きくなると、 $t_{\alpha}(\nu)$ は Z_{α} に近づき、大標本では正規分布で近似できる。

3) 母分散の信頼区間

$\frac{(n-1)U^2}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従うので

$$P\left(\chi^2_{1-\frac{\alpha}{2}}(\nu) \leq \frac{(n-1)U^2}{\sigma^2} \leq \chi^2_{\frac{\alpha}{2}}(\nu)\right) = 1 - \alpha \text{ となる}$$

$\chi^2_{1-\frac{\alpha}{2}}(\nu)$, $\chi^2_{\frac{\alpha}{2}}(\nu)$ を χ^2 分布表から求める。

$$\sigma^2 \text{ の信頼区間は } \left[\frac{(n-1)U^2}{\chi^2_{\frac{\alpha}{2}}(\nu)}, \frac{(n-1)U^2}{\chi^2_{1-\frac{\alpha}{2}}(\nu)}\right] \text{ となる。}$$

例) データ(1.5,2.6,2.3,0.6,2.0) の時 $\bar{X} = \frac{1.5+2.6+2.3+0.6+2.0}{5} = 1.8$

σ^2 が既知で $\sigma^2 = 0.7$ のとき

$$\sigma = 0.837, \frac{\sigma}{\sqrt{n}} = 0.374$$

母平均 μ の95% 信頼区間は, $\frac{\alpha}{2} = 0.025$ 、 $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

$$z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} = 1.96 \times 0.374 = 0.733 \quad \rightarrow \bar{X} \pm 0.733 \rightarrow [1.067, 2.533]$$

σ^2 が未知のとき

$$U^2 = \frac{1}{5-1} \{(1.5 - 1.8)^2 + (2.6 - 1.8)^2 + \dots + (2.0 - 1.8)^2\} = 0.615 \quad \text{で計算すると}$$

$$U = 0.784, \frac{U}{\sqrt{n}} = 0.351 \quad \text{なので}$$

$$\chi^2_{0.025}(4) = 11.1433, \chi^2_{0.975}(4) = 0.484419 \quad \text{を}\chi^2\text{分布表から読み取り}$$
$$(n-1)U^2 = 4 \times 0.615 = 2.46$$

$$\frac{(n-1)U^2}{\chi^2_{0.025}(4)} = 0.221, \frac{(n-1)U^2}{\chi^2_{0.975}(4)} = 5.078 \quad \rightarrow \sigma^2 \text{ の信頼区間は } [0.221, 5.078]$$

二項分布やポアソン分布の場合

二項分布

$$\sum_{k=0}^x {}_n C_k p_u^k (1-p_u)^{n-k} = \frac{\alpha}{2} \quad \sum_{k=x}^n {}_n C_k p_L^k (1-p_L)^{n-k} = \frac{\alpha}{2}$$

n が大きいとき正規分布で近似

$$E(x) = \mu = np \quad V(x) = np(1-p) \quad \hat{\mu} = x$$

$$\hat{p} = \frac{x}{n} \quad E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{E(x)}{n} = p \quad V(\hat{p}) = \frac{V(x)}{n^2} = \frac{(1-p)p}{n}$$

ポアソン分布

$$\sum_{k=0}^x \frac{\mu_u^k}{k!} e^{-\mu_u} = \frac{\alpha}{2} \quad \sum_{k=x}^{\infty} \frac{\mu_L^k}{k!} e^{-\mu_L} = \frac{\alpha}{2}$$

n が大きいとき正規分布で近似

$$E(x) = \mu, \quad V(x) = \mu, \quad \hat{\mu} = x$$

例) 無作為に 500 人を選ぶ。そのうち 100 人が 95%信頼区間

$$\hat{p} = \frac{100}{500} = 0.2 \quad -1.96 \leq \frac{0.2-p}{\sqrt{\frac{0.2(1-0.2)}{500}}} \leq 1.96$$

$$\rightarrow p \text{ の範囲は } 0.2 \pm 1.96 \sqrt{\frac{0.2(1-0.2)}{500}} = 0.2 \pm 0.035 \rightarrow [0.165, 0.235]$$

例) 100 年に 100 回災害発生

$$\hat{\mu} = 100 \quad -1.96 \leq \frac{100-\mu}{\sqrt{100}} \leq 1.96$$

$$\mu \text{ の範囲は } 100 \pm 1.96\sqrt{100} = 100 \pm 19.6$$

例) $\alpha = 5\text{cm}$ のとき、 \bar{x} の 95%信頼区間を 5mm にするために必要な標本の数は平均の区間推定では、標本数 n が多くなれば、信頼区間の幅も狭くなる。

$$\text{つまり、} -1.96 \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96 \text{ より、} -1.96 \frac{\sigma}{\sqrt{n}} \leq (\bar{X} - \mu) = 0.5 \leq 1.96 \frac{\sigma}{\sqrt{n}}$$

$$1.96 \frac{\sigma}{\sqrt{n}} - \left(-1.96 \frac{\sigma}{\sqrt{n}}\right) = 2 \times 1.96 \times \frac{5}{\sqrt{n}} = 0.5 \rightarrow \sqrt{n} = 39.2 \rightarrow n = 1537$$

仮説検定 hypothesis testing

統計的仮説(statistical hypothesis)の有意性の検定(test of significance)

仮説の下で期待される結果と観測結果の違いが偶然によって起こったものか否かを確率で評価する。

違いが小 \rightarrow 偶然と考えて矛盾しない(違いに意味がない)

違いが大 \rightarrow 偶然で起こる確率(p 値 p-value)は小さい(有意 significant)

帰無仮説(null hypothesis)

H_0 : 検定を行いたいものの仮説(だが、有意性を証明するためには棄却 reject される方が良い)

対立仮説(alternative hypothesis)

H_1 : H_0 が棄却された場合に採択 accept される仮説

	H_0 が真	H_0 が偽
H_0 を棄却	誤り 1	正しい
H_0 を棄却しない	正しい	誤り 2

誤り 1: 第 1 種の誤り(過誤) type I error

誤り 2: 第 2 種の誤り(過誤) type II error

例) コインを 20 回投げて 14 回表が出た場合

H_0 : コインに歪みがない → 表の確率 1/2、期待値 10
 20 回中 x 回表の確率 Bin(20, 1/2)

$$f(x) = {}_{20}C_x \left(\frac{1}{2}\right)^x \left(1 - \frac{1}{2}\right)^{20-x}$$

14 回以上表が出る確率は

$$P(x \geq 14) = \sum_{x=14}^{20} f(x) = 1 - \sum_{x=0}^{13} f(x) = 0.0577$$

H_1 : 表が出やすい(コインがゆがんでいる)

$P=0.0577$ (p 値: 0.1154)

有意水準 $\alpha = 0.1 \rightarrow H_0$ は棄却される

$\alpha = 0.05 \rightarrow H_0$ は棄却されない

仮説検定のまとめ

1 標本の分析 = 期待値 (母平均) に関する検定

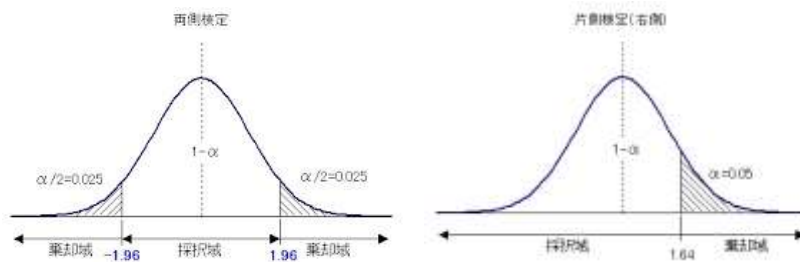
正規母集団からの標本 (x_1, x_2, \dots, x_n) が i.i.d で $N(\mu, \sigma^2)$ に従う時

標本平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ は正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$ に従う。

この標本から得られた $\mu_0 = \bar{x}$ が母平均 μ からどれくらい離れているかを比較

両側検定 $H_0: \mu = \mu_0 \quad H_1: \mu \neq \mu_0 \quad |Z| > Z_{\alpha/2}$ のときを棄却

片側検定 $H_0: \mu = \mu_0 \quad H_1: \mu > \mu_0 \quad |Z| > Z_{\alpha}$ のときを棄却



- 母平均、母分散が既知なら Z 検定

検定量 $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ が $N(0, 1)$ に従う $|Z| > Z_{\alpha/2}$ のときを棄却

例) ある年、全国の小学 3 年 (8 歳) 男子の身長は平均値 128.0 cm (標準偏差 5.44 cm)、A 校 3 年生男子 25 人の平均身長は 126.5 cm であった。低すぎか?

$$Z = \frac{126.5-128}{\frac{5.44}{\sqrt{25}}} = \frac{-1.5}{\frac{5.44}{5}} = \frac{-1.5}{1.088} = -1.38$$

検定量 $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ となる $\frac{\alpha}{2}$ を正規分布表から求める。

この場合、標準正規分布表から (P=0.084、有意ではない)

Z 検定では、両側検定で Z 値の絶対値が 1.96、片側検定で Z 値の絶対値が 1.65 以上なら有意水準 $\alpha=0.05$ での棄却域(有意)、以下なら有意差なし

$\alpha=0.01$ では Z 値の絶対値が片側検定 2.58、両側有意 2.81 以上で有意となる。

- 母平均既知、母分散が未知なら 一標本 t 検定

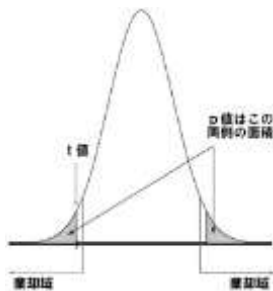
$$\bar{X} = \left(\frac{x_1 + x_2 + \dots + x_n}{n} \right), U^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

$t = \frac{\bar{X}-\mu}{\frac{U}{\sqrt{n}}}$ は自由度 n-1 の t 分布に従う。

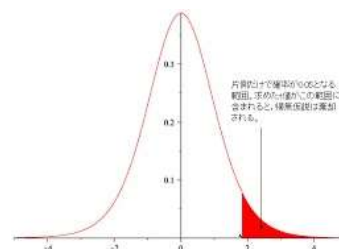
両側検定 $H_0 : \mu = \mu_0 \quad H_1 : \mu \neq \mu_0 \quad |t| > t_{\alpha/2}(n-1)$ のときを H_0 棄却

片側検定 $H_0 : \mu = \mu_0$

$H_1 : \mu > \mu_0$ または、 $\mu < \mu_0 \quad |t| > t_{\alpha}(n-1)$ のときを棄却



棄却域 $|t| > t_{\alpha}(n-1)$



例)ある年全国の小学3年(8歳)男子の身長は平均値128.0cmであり、A校3年生男子25人の平均身長は126.5cm(標準偏差5.44cm)であった。低すぎか？

$$t = \frac{126.5-128}{\frac{5.44}{\sqrt{25}}} = \frac{-1.5}{\frac{5.44}{5}} = \frac{-1.5}{1.088} = -1.38$$

t 分布表から自由度 25-1=24 の片側 $\alpha=0.05$ での $t_{24}(0.05)=1.711$

$1.38 < 1.711$ より有意ではない

[n 大の時は正規分布表をみる(片側 1.65, 両側 1.96 以上で有意)]

二項分布やポアソン分布の場合

二項分布

$$\sum_{k=0}^x {}_n C_k p_u^k (1-p_u)^{n-k} = \frac{\alpha}{2} \quad \sum_{k=x}^n {}_n C_k p_L^k (1-p_L)^{n-k} = \frac{\alpha}{2}$$

n が大きいとき正規分布で近似

$$\frac{x}{n} - 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \frac{x}{n} + 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$E(x) = \mu = np \quad V(x) = np(1-p) \quad \hat{\mu} = x$$

$$\hat{p} = \frac{x}{n} \quad E(\hat{p}) = E\left(\frac{x}{n}\right) = \frac{E(x)}{n} = p \quad V(\hat{p}) = \frac{V(x)}{n^2} = \frac{(1-p)}{n}$$

例) トランプの赤または黒を 400 回中 220 回の中

$$H_0 : p = 1/2$$

$$np = 400 \times 1/2 = 200, \quad np(1-p) = 200 \times 1/2$$

$$z = \frac{220-200}{\sqrt{100}} = 2 > 1.96 \quad \text{有意}$$

例) 大統領選挙に A, B の 2 人の人が立候補した無作為に 3000 人にアンケートし A 候補が 60% の支持を得ていた。どのくらいの信頼度か？

$$p = 0.6 \pm 1.96 \sqrt{0.6 * \frac{0.4}{3000}} = 0.60 \pm 0.018$$

ポアソン分布

$$\sum_{k=0}^x \frac{\mu_u^k}{k!} e^{-\mu_u} = \frac{\alpha}{2} \quad \sum_{k=x}^{\infty} \frac{\mu_L^k}{k!} e^{-\mu_L} = \frac{\alpha}{2}$$

n が大きいとき正規分布で近似 $E(x) = \mu$ 、 $V(x) = \mu$ 、 $\hat{\mu} = x$

例) 1 分間に平均 100 個の自然放射線、ある場所で 1 分間に 120 個観測

$H_0 : \mu = 100$ ($\rightarrow N(100, 100)$) と近似: ポアソン分布では平均と分散が等しい)

$$z = \frac{120-100}{\sqrt{100}} = 2 > 1.96 \quad (\alpha = 0.05 \text{ で有意})$$

100 分間で 10300 個観測された場合

$$z = \frac{10300-10000}{\sqrt{10000}} = 3 > 2.81 \quad (\alpha = 0.01 \text{ でも有意})$$

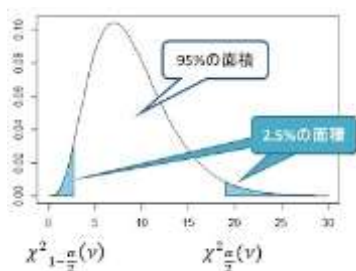
母分散に対する仮説検定

$\chi^2 = \frac{(n-1)U^2}{\sigma^2}$ は帰無仮説のもとで自由度 $n-1$ の χ^2 分布に従う。

$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$ (両側検定)

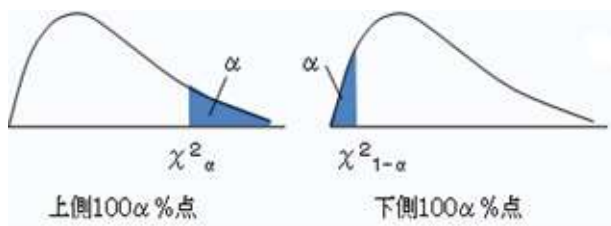
$\chi^2 < \chi^2_{1-\frac{\alpha}{2}}(n-1)$ または、 $\chi^2_{\frac{\alpha}{2}}(n-1) < \chi^2$ の時 H_0 を棄却

$\alpha = 0.05, \frac{\alpha}{2} = 0.025$ の時は下図



$H_1 : \sigma^2 > \sigma_0^2$ (片側検定) $\chi^2_{\alpha}(n-1) < \chi^2$ の時 H_0 を棄却

$H_1 : \sigma^2 < \sigma_0^2$ (片側検定) $\chi^2 < \chi^2_{1-\alpha}(n-1)$ の時 H_0 を棄却



数理統計的側面(2)

2 標本問題の基本

2つの正規分布 $N(\mu_1, \sigma_1^2)$ 、 $N(\mu_2, \sigma_2^2)$ からそれぞれ大きさ m, n の標本 $(X_1, X_2 \dots X_m)$ 、 $(Y_1, Y_2 \dots Y_n)$ を独立に抽出した場合の標本分布において

2つの母平均の差 $\mu_1 - \mu_2$ を分析するため、 $\bar{X} = \frac{1}{m}(X_1 + X_2 + \dots + X_m)$ 、

$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + \dots + Y_n)$ の差 $\bar{X} - \bar{Y}$ を調べる。 $\bar{X} - \bar{Y}$ の標本分布は μ_1 、 μ_2

だけでなく σ_1^2 、 σ_2^2 にも依存するため、

1) σ_1^2 、 σ_2^2 が既知

2) σ_1^2 、 σ_2^2 が未知だが等しい

3) σ_1^2 、 σ_2^2 が未知で等しいとは限らない

場合に分けて考える。

1) 分散が既知のとき

$V(\bar{X} - \bar{Y}) = V(\bar{X}) + V(\bar{Y}) = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}$ となり $\bar{X} - \bar{Y}$ の分布は

$N(\mu_1 - \mu_2, \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n})$ となる。

従って、 $\bar{X} - \bar{Y}$ を標準化した

$$Z = \frac{(\bar{X} - \bar{Y}) - E(\bar{X} - \bar{Y})}{\sqrt{V(\bar{X} - \bar{Y})}} = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

2) 分散は未知であるが等しいとき

$\sigma_1^2 = \sigma_2^2 = \sigma^2$ とすると、 $\bar{X} - \bar{Y}$ の分布は

$N(\mu_1 - \mu_2, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}) = N[\mu_1 - \mu_2, (\frac{1}{m} + \frac{1}{n})\sigma^2]$ に従うが、 σ^2 が未知なので、

母分散が共通の二つの標本を合併したものから、次の合併した分散

$$U^2 = \frac{\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2}{m+n-2} = \frac{(m-1)U_1^2 + (n-1)U_2^2}{m+n-2} \text{ で推定する。}$$

$$\text{ただし、} U_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2 \text{、} U_2^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2$$

この様に標本分散を定義すると、

① $(m+n-2)\frac{U^2}{\sigma^2}$ は自由度 $m+n-2$ の χ^2 分布 $\chi^2(m+n-2)$ に従う。

② U^2 と $\bar{X}-\bar{Y}$ は独立ということが知られている。

よって標準化した

$$Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{(\frac{1}{m}+\frac{1}{n})\sigma^2}} \text{ は } N(0,1) \text{ に従い、 } Z \text{ を } \sqrt{\frac{U^2}{\sigma^2}} \text{ で割った}$$

$$t = \frac{Z}{\sqrt{\frac{U^2}{\sigma^2}}} = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{(\frac{1}{m}+\frac{1}{n})U^2}} = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{U\sqrt{(\frac{1}{m}+\frac{1}{n})}}$$

は自由度 $m+n-2$ の t 分布、 $t(m+n-2)$ に従う。

3) 母分散が未知であり、等しいとは限らないとき

→ウエルチの近似法 Welch's t test

$$t = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{U_1^2}{m} + \frac{U_2^2}{n}}} \text{ は近似的に自由度が}$$

$$U = \frac{(\frac{U_1^2}{m} + \frac{U_2^2}{n})^2}{\frac{1}{m-1}(\frac{U_1^2}{m})^2 + \frac{1}{n-1}(\frac{U_2^2}{n})^2} \text{ に最も近い整数 } U^* \text{ の } t \text{ 分布、 } t(U^*) \text{ に従う。}$$

標本分散の比の標本分布: F 分布の定義

独立な確率変数 U, V がそれぞれ自由度 k_1, k_2 の χ^2 分布に従うとき、

$F = \frac{\frac{U}{k_1}}{\frac{V}{k_2}}$ が従う確率分布を自由度 (k_1, k_2) の F 分布といい、 $F(k_1, k_2)$ で表す。

U_1^2, U_2^2 が独立で、 $(m-1)\frac{U_1^2}{\sigma_1^2}$ 、 $(n-1)\frac{U_2^2}{\sigma_2^2}$ はそれぞれ自由度 $(m-1)$ 、 $(n-1)$

の χ^2 分布に従うので、 $\frac{\frac{(m-1)\frac{U_1^2}{\sigma_1^2}}{m-1}}{\frac{(n-1)\frac{U_2^2}{\sigma_2^2}}{n-1}} = \frac{\frac{U_1^2}{\sigma_1^2}}{\frac{U_2^2}{\sigma_2^2}}$ は自由度 $(m-1, n-1)$ の $F(m-1, n-1)$ に

従う。

ここで、母分散が等しい時 ($\sigma_1^2 = \sigma_2^2$)、標本分散の比 $F = \frac{U_1^2}{U_2^2}$ が F 分布に従う。

t 分布と F 分布の関係

t が自由度 k の t 分布 $t(k)$ に従うとき

$$t^2 = \frac{Z^2/1}{Y/k} \left(= \frac{\chi^2(1)}{\chi^2(k)} \right) \text{ は } F(1, k) \text{ に従う}$$

計算式の確認

平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

平均値からの偏差の合計 $\sum_{i=1}^n (x_i - \bar{x}) = 0$

偏差平方和 sum of squares の計算

$$\begin{aligned} S_{xx} &= \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \bar{x}^2 \sum_{i=1}^n 1 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} * n\bar{x} + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n} \quad \frac{(\sum x_i)^2}{n}: \text{補正項 CTx} \end{aligned}$$

標本分散 $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{S_{xx}}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$

偏差積和 sum of products (of deviation)

$$\begin{aligned} S_{xy} &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i y_i - y_i \bar{x} - x_i \bar{y} + \bar{x} * \bar{y}) \\ &= \sum_{i=1}^n x_i y_i - \bar{x} \sum_{i=1}^n y_i - \bar{y} \sum_{i=1}^n x_i + \bar{x} * \bar{y} \sum_{i=1}^n 1 \\ &= \sum_{i=1}^n x_i y_i - \bar{x} * n\bar{y} - \bar{y} * n\bar{x} + n\bar{x} * \bar{y} = \sum_{i=1}^n x_i y_i - n\bar{x} * \bar{y} \\ &= \sum_{i=1}^n x_i y_i - \frac{\sum x_i \sum y_i}{n} \quad \frac{\sum x_i \sum y_i}{n}: \text{補正項 CTxy} \end{aligned}$$

分散分析では

水準 A ₁	水準 A ₂	水準 A _i	水準 A _a	
X ₁₁	X ₂₁	...	X _{a1}	
X ₁₂	X ₂₂	...	X _{a2}	
..	
X _{1n}	X _{2n}		X _{an}	
\bar{x}_1	\bar{x}_2	\bar{x}_i	\bar{x}_a	\bar{x}

a:水準の数 a 個 n:水準でのデータの数(各水準とも n にした場合) → x_{ij} :水準 i, j 番目のデータ

$$x_{ij} = \bar{x} + (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i) \Rightarrow \bar{x} \text{を移項して}$$

$$(x_{ij} - \bar{x}) = (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i) \Rightarrow \text{両辺を二乗し各データを合計する場合、}$$

$$\text{例えば } i=1, A_1 \text{ だけで考え、} (x_{1j} - \bar{x}) = (\bar{x}_1 - \bar{x}) + (x_{1j} - \bar{x}_1)$$

平均値からの偏差の合計 $\sum_{j=1}^n (x_{1j} - \bar{x}_1) = 0$, $(\bar{x}_1 - \bar{x})$ は定数より

$$\sum_{j=1}^n (\bar{x}_1 - \bar{x}) (x_{1j} - \bar{x}_1) = (\bar{x}_1 - \bar{x}) \sum_{j=1}^n (x_{1j} - \bar{x}_1) = 0$$

$i=2, 3, \dots, a$ の時も同様なので $\sum_{i=1}^a \sum_{j=1}^n (\bar{x}_i - \bar{x}) (x_{ij} - \bar{x}_i) = 0$ よって

$$\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x})^2 = \sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

が成り立つ。

二元配置の数学モデルにおいても

	B ₁	B ₂	...	B _b
A ₁	X ₁₁₁ ... X _{11n}	X ₁₂₁ ... X _{12n}	...	X _{1b1} ... X _{1bn}
A ₂	X ₂₁₁ ... X _{21n}	X ₂₂₁ ... X _{22n}	...	X _{2b1} ... X _{2bn}
...	
A _a	X _{a11} ... X _{a1n}	X _{a21} ... X _{a2n}	...	X _{ab1} ... X _{abn}

総平均値と各水準からの偏差の合計 $\sum_{i=1}^n (\bar{x}_i - \bar{x}) = 0$, $\sum_{j=1}^a (\bar{x}_j - \bar{x}) = 0$ 等から

$$(x_{ijk} - \bar{x}) = (\bar{x}_{i*} - \bar{x}) + (\bar{x}_{*j} - \bar{x}) + (\bar{x}_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x}) + (x_{ijk} - \bar{x}_{ij})$$

の両辺を 2 乗しデータを合計すると

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x})^2 = \sum b n (\bar{x}_{i*} - \bar{x})^2 + \sum a n (\bar{x}_{*j} - \bar{x})^2 + \sum \sum n (\bar{x}_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2 + \sum \sum \sum (x_{ijk} - \bar{x}_{ij})^2 \text{ が成り立つ}$$

2 標本の分析

検定とは: 差があることを証明するために差がないと仮定し(帰無仮説 H_0)その起こる確率を計算し、 H_0 を棄却することにより差がある(対立仮説 H_1)ことを証明。

対応のあるt検定 paired t test

(関連標本t検定related samples t-test)

関連2群の差の検定 = 2群の差を観測値とする1標本の検定

⇒ 正規性のある母集団からの平均値と誤差 → 平均値 $\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i$ 、分散の

推定値 $Sd^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2$ を用いてt分布(自由度 $n-1$)を考える。= 母平均0とのt検定(関連2群の差の検定)と同じになる。

$$t = \frac{\bar{d}}{\frac{Sd}{\sqrt{n}}}$$

関連 2 群の差は
t分布そのもの

t_α は自由度 $df = n-1$ の t 分布に従い、t 分布表、から t_α の値を探す。

| t | > t_α のとき、 $P < \alpha$ となり帰無仮説を棄却する。有意差あり。

例) 血圧降下薬 X の効果を調べるため、健康な男子に薬物 X を投与して投与前と投与後の最高血圧の値を比較した。薬物 X は血圧降下の作用があるか

	血圧(前)	血圧(後)	d(=差)
A	120	106	14
B	135	113	22
C	116	98	18
D	132	114	18
E	124	107	17
F	130	120	10

$$\bar{d} = (14+22+18+18+17+10)/6 = 16.5$$

標準偏差 $Sd = 4.08$

データ数 n は 6 なので、この三つの値から t 値を計算する。

$$t = \frac{\bar{d}}{\frac{Sd}{\sqrt{n}}} = \frac{16.5}{\frac{4.08}{\sqrt{6}}} = 9.89$$

自由度 df は $n-1 = 6-1 = 5$ の t 分布に従うので $t_{0.05} = 2.571$

t 値 = $9.89 > 2.571 = t_{\alpha}$ より、 $t > t_{0.05}$ となる。つまり、 $P < 0.05$ となるので帰無仮説を棄却し、対立仮説を採用する。

*エクセル関数 =T.DIST.2T(9.89,5)=0.000180256

(参考:エクセル関数 =T.DIST.2T(2.571,5)= 0.049974635)

BellCurve for Excel: 2 標本の比較 → 母平均の差の検定 (対応あり) → パラメトリック → t 分布

母平均の差の検定 (対応あり)					
基本統計量					
変数	n	平均	不偏分散	標準偏差	標準誤差
血压(前)	6	126.167	54.567	7.387	3.016
血压(後)	6	109.667	58.667	7.659	3.127
対の差と効果量 (対の差 = 「血压(前)」 - 「血压(後)」)					
対の差		効果量			
平均	標準偏差	Cohen's d	Hedges' g	Cohen's d'	
16.5000	4.0866	2.4022	2.1929	4.0376	
差の平均の95%信頼区間 (両側検定/対立仮説:対の差 ≠ 0)					
	下限値	上限値			
t推定	12.2114	20.7886			
差の平均の検定 (両側検定/対立仮説:対の差 ≠ 0)					
	統計量:t	自由度	P 値	* : P<0.05 ** : P<0.01	
t検定	9.8901	5	p < 0.001	**	
検出力 (両側検定/対立仮説:対の差 ≠ 0)					
	α=0.05				
t検定	1.0000				

Cohen's d、Hedges' d は、効果量。SD を単位として、平均値がどれだけ離れているかの指標。d = 1 は、1SD 離れている意味 → p201

* 対応のある検定 (ノンパラメトリック)

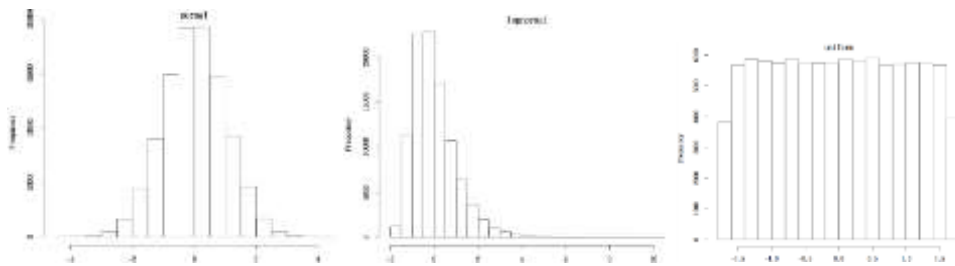
Wilcoxon Signed Rank Test → p114

独立した2標本t検定 independent samples t-test

群馬大学の青木繁伸先生がインターネットに開示している「二群の平均値(代表値)の差を検定するとき」が参考になるので要約する。

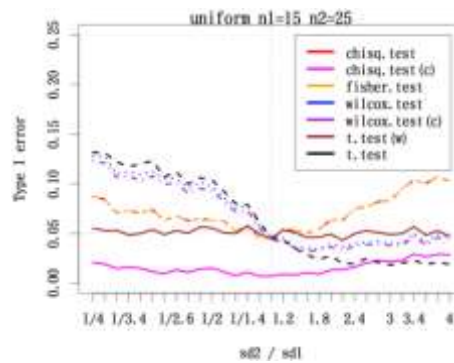
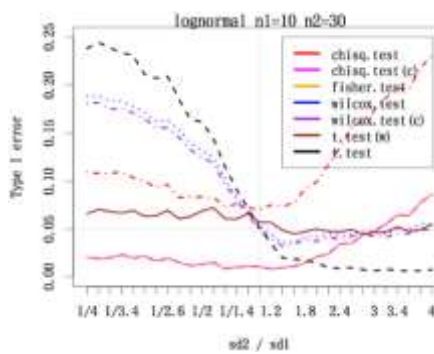
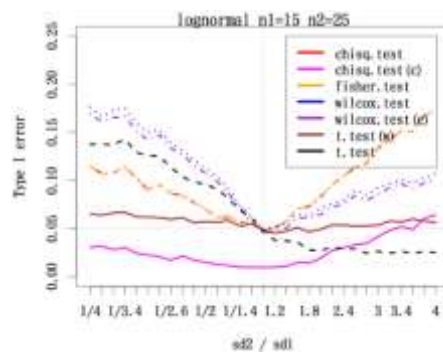
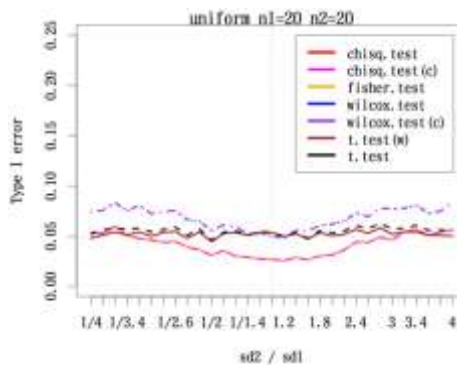
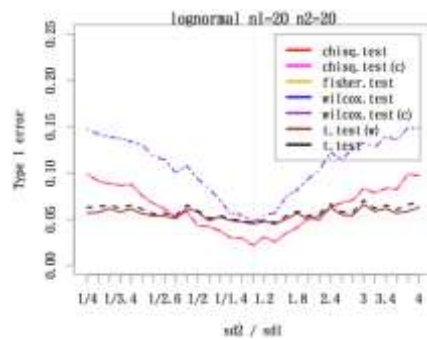
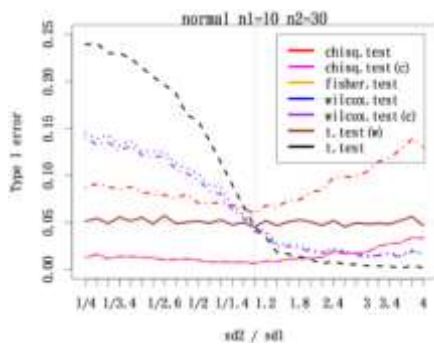
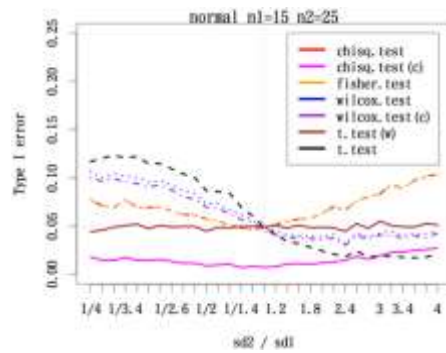
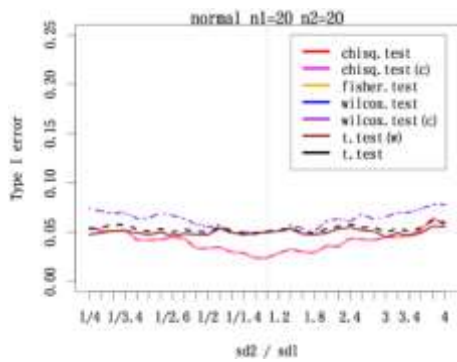
(<http://aoki2.si.gunma-u.ac.jp/lecture/BF/index.html>)

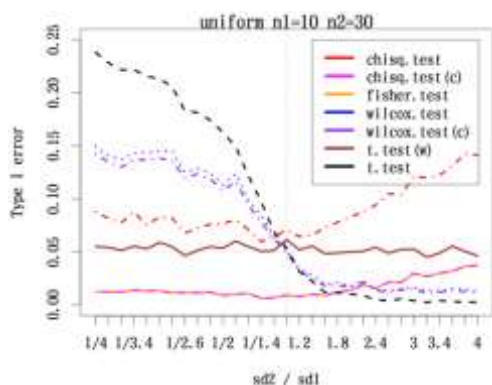
データの分布としては、正規分布のような左右対称な分布(下図左)、対数正規分布のような歪みのある分布(下図中央)、一様分布のような分布(下図右)という、3種類をあげ各種検定の第一種の過誤がどれくらいになるかを検討している。標本サイズとしては、両群が等しい場合($n_1=20, n_2=20$), 約 1:2 の場合($n_1=15, n_2=25$), 1:3 の場合($n_1=10, n_2=30$)の、3通り、分散の違いとしては、標準偏差の比が 1 ~ 4 まで 0.2 刻み、およびその逆数の 31 種類をデータの分布型と標本サイズの設定ごとに、平均値は同じで標準偏差の異なるデータを生成し検定を行うことを統計ソフトRを用いて5000回ずつ実施した。



横軸は二群の標準偏差の比, 縦軸は1000回の実験において, 有意になった割合(第一種の誤差)で、この値が 0.05 であることが期待される。

`chisq.test` は中央値検定(カイ二乗検定), `chisq.test(c)` は連続性の補正をしたもの, `fisher.test` は中央値検定に使うクロス集計表に対して Fisher の正確検定を行った結果, `wilcox.test` はマン・ホイットニーの U 検定, `wilcox.test(c)` は連続性の補正をしたもの, `t.test(w)` は等分散を仮定しない t 検定(ウェルチの方法), `t.test` は等分散を仮定する `t.test` である。





その他のサンプルサイズの場合: 条件は同じで, サンプルサイズとして $(n_1, n_2) = (5, 5), (5, 10), (5, 15), (50, 50), (50, 100), (50, 150)$ の場合についてもシミュレーションを行ったが, 結論は同じようなものであった。すなわち, どのような場合にもウェルチの検定の第一種の過誤はほぼ 0.05 で安定していた。

結論: 最初からズバリ「等分散を仮定しない t 検定」を行うのがよい。(青木先生)

$$\text{Welch's } t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

で計算される

独立した二標本の検定については上記や、正規性、等分散性を仮定しない検定、並び替え検定(≒確率化テスト、正確検定)などの考え方を参考にすると

サンプル数にかかわらず

Welch's t-test(ウェルチのt検定)でよいが、合計で20以上が望ましい。

サンプル数が少ない(合計で20未満など)場合には(下記に詳述)

Brunner-Munzel検定(小標本)または、

Mann-Whitney U test 正確検定(合計で20未満では並び替え検定による正確確率が計算される)が良いのではないかとと思われる。

並び替え検定 permutation test

参考) www.mizumot.com/files/permutation.pdf

並び替え検定はノンパラメトリックな方法に分類される。手元の標本データを母集団とみなし, そこから2群のデータ数を固定したまま無作為抽出を繰り返す。

2 群のデータ(n = 12)

Group A	Group B
20	10
40	70
30	80
40	90
40	70
50	80
M = 36.67、SD = 10.33	M = 66.67、SD = 28.75

平均値の差 = 30.00

上表のような2 群のデータが得られたとする。Group A とGroup B はランダムに割りつけられており、平均値に30.00 の差がある。この2 群はもともと同一母集団に属していたサンプルで、帰無仮説を2つの母集団は同じものであるとし、2群の間に生じた30.00 の差はグループへの割りつけによって偶然生じた差であるとみなす。その場合手元の全データ12 個をプールし、そこから各群に6 個ずつ再度割りつける。つまり、20, 40, 30, 40, 40, 50, 10, 70, 80, 90, 70, 80 (12 個の値をプールする)

ここから無作為抽出し、A とB に割りつける。例えば

A: 40, 10, 20, 80, 80, 70 (平均: 50.00) B: 40, 30, 40, 70, 50, 90 (平均: 53.33)

Group A Group B 差 = 3.33

同様にして可能な組み合わせすべてができるまで繰り返す。n = 12 で、各群に6 個のデータが割りつけられる場合のすべての組み合わせは、 $12!/6!6! = 924$ 組ある。この924 組のうち、2 群の平均値の差が30.00 かそれ以上(平均の差の絶対値で)になる組み合わせは54 あり、 $54/924 = 0.0584$ がp 値となる(両側検定)。このp 値が $p > 0.05$ であるため、「5%水準で有意差はない」という結果になる。

* Mann-Whitney'U testの正確検定は並び替え検定

Brunner-Munzel検定(ブルンナー=ムンツェル検定)

参照) 奥村晴彦先生 <http://oku.edu.mie-u.ac.jp/~okumura/stat/brunner-munzel.html>

分布が同じことは仮定せず、両群から一つずつ値を取り出したとき、どちらかが大きい確率自体が等しいという帰無仮説を検定する。2つの群の相対的な順位を比較する手法で、ノンパラメトリック検定の1つ。2群のデータを1つにまとめて順位付けし、この順位データから検定を行う。正規分布に従わなくてもよく、等分散性の仮定も必要ない。大標本では検定統計量Wが標準正規分布に従うことを利用。小標本の場合は、補正した自由度を用いたt検定による結果を利用。後者でも十分だが、Brunner-Munzelの統

計量に並べ替え検定を適用したpermutated Brunner-Munzel testもある。Mann-Whitney'U testの改良版がBrunner-Munzel検定といえる。

BellCurve for Excel: 2 標本の比較→マン=ホイットニーの U 検定、→ブルンナー=ムンツェル検定

マン=ホイットニーのU検定						
基本統計量						
変数	n	平均順位	統計量:U			
Group A	6	4.500	30			
Group B	6	8.500	6			
マン=ホイットニーのU検定						
統計数値表による検定			正規化検定		*: P<0.05 **: P<0.01	
統計量:U	両側検定	E(U)	V(U)	連続性補正	統計量:z	両側P値
6	※ 1	18	38.182	なし	1.9420	0.0521
				あり	1.8611	0.0627
正確検定						
両側P値 * : P<0.05 ** : P<0.01						
0.0584						

ブルンナー=ムンツェル検定						
基本統計量						
変数	n	平均順位				
Group A	6	4.500				
Group B	6	8.500				
ブルンナー=ムンツェル検定 (正規近似)						
統計量W	両側P値	*: P<0.05 **: P<0.01				
2.0000	0.0455	*				
ブルンナー=ムンツェル検定 (小標本)						
統計量W	自由度	両側P値	*: P<0.05 **: P<0.01			
2.0000	5.0000	0.0510				

例題) 透析患者と健常者のB-cell免疫グロブリン生成(%)を測定した結果がある。ブルンナー=ムンツェル検定により2群の値に違いがあるか検定する。

BellCurveエクセル統計より「ブルンナー=ムンツェル検定」

透析患者	健常者	透析患者	健常者
10	7	17	
13	8	14	
7	8	6	

16	6	11	
12	7	12	
12	11	7	
10	10	9	
10	8	12	
4	9	9	

BellCurve for Excel: 2 標本の比較→ブレンナー=ムンツェル検定

透析患者	健常者		ブレンナー=ムンツェル検定		
10	7	基本統計量			
13	8	変数	n	平均順位	
7	8	透析患者	18	16.19444	
16	6	健常者	9	9.611111	
12	7				
12	11	ブレンナー=ムンツェル検定 (正規近似)			
10	10	統計量W	両側P値	*P<0.05 **P<0.01	
10	8	2.550749707	0.010749	*	
4	9				
17		ブレンナー=ムンツェル検定 (小標本)			
14				*: P<0.05 **: P<0.01	
6		統計量W	自由度	両側P値	
11		2.550749707	24.04445	0.008767	**
12					
7					
9					
12					
9					

➤ マンホイットニー検定では計 20 例を超えると正確検定できない

マン=ホイットニーのU検定						
基本統計量						
変数	n	平均順位	統計量:U			
透析患者	18	16.194	41.5			
健常者	9	9.611	120.5			
マン=ホイットニーのU検定						
統計数値表による検定			正規化検定	*: P<0.05 **: P<0.01		
統計量:U	両側検定		E(U)	V(U)	連続性補正	統計量:z 両側P値
41.5	※ 1		81	373.385	なし	2.0442 0.0409 *
					あり	2.0183 0.0436 *
正確検定						
両側P値	*: P<0.05 **: P<0.01					
※ データ数の合計が20を超えるため正確検定は出力できません。						

◇ 2群の差、比などの表し方

- 検定では、 $p=0.035$ というように具体的に示すほうが好まれる。
- 最近では、差や比も信頼区間で表すほうが好まれる。

例) 2群の差: $\text{Mean} \pm 95\% \text{CI}$

この場合は95%CIに0が含まれると有意でない

例) 2群の比: オッズ比など

この場合は95%CIに1が含まれると有意でない

例	有効率(%)		P値	統計的有意	有効率の差(%)		95%信頼区間 (●=点推定)	臨床的有意
	New	Standard			点推定	95%CI		
1.	480/800=60%	416/800=52%	0.001	Yes	8%	3% to 13%		No
2.	15/25=60%	13/25=52%	0.57	No	8%	-19% to 35%		Inconclusive
3.	15/25=60%	9/25=36%	0.09	No	24%	-3% to 51%		Inconclusive
4.	240/400=60%	144/400=36%	<0.0001	Yes	24%	17% to 31%		Yes

統計的有意と臨床的有意

上表で

- 1.ではサンプル数が多いため有意差があり、 $\text{Mean} \pm 95\% \text{CI}$ で0は含まないが、差は0に近い。
- 4.ではサンプル数が多いだけでなく、 $\text{Mean} \pm 95\% \text{CI}$ でも0を含まず事前に決めた差がある。

⇒ 臨床的有意性についてはp202、204参照

2群の差と分散分析

3標本以上の分析には分散分析(analysis of variance; ANOVA)を行なう。分散分析は複数の群間の平均値の差を比較するための手法である。複数の群間の平均値を比較すると平均値差がたぐさん生じ、これらの平均値差を1つの指標にまとめる必要があり「分散」が用いられる。群間の分散が大きいということはすなわち、群の間にばらつきがある＝群間の差が大きいということになる。分散分析は、(群間分散)/(群内分散)が大きくなれば有意になりやすくなる。「群間分散が大きいこと」は、「群間の平均値差が大きいこと」を示し、「群内分散が小さいこと」は「各群内の分散が小さいこと」を示し、「有意差がある」というとき、「平均値差が大きい」ということと同時に「群内分散が小さい」ということも意味している。

実際、2群間の平均値差を分散分析にかけると、t検定と同じ結果がでる。

BellCurve for Excel: 2 標本の比較→t検定、分散分析多重比較→一元配置分散分析

2群の差と分散分析										
t検定 (→Welchの方法で比較すると)										
Aクラス		Bクラス								
	60	49								
	52	40								
	68	52								
	55	37								
	65	55								
	47	38								
	45	45								
	62									
	53									
母平均の差の検定										
基本統計量										
変数	n	平均	不偏分散	標準偏差	標準誤差					
Aクラス	9	56.333	63.000	7.937	2.646					
Bクラス	7	45.143	50.476	7.105	2.685					
等分散性の検定 (両側検定)										
統計量:F	自由度1	自由度2	P 値							
1.2481	8	6	0.8094							
平均の差	Cohen's d	Hedges' g	Cohen's d	Hedges' g						
11.1905	1.5758	1.4741	1.5222	1.4238						
母平均の差の検定 (両側検定/対立仮説: 「Aクラス」 ≠ 「Bクラス」)										
	統計量:t	自由度	P 値		*: P<0.05 **: P<0.01					
観測値	t検定	2.9250	14	0.0111 *						
	Welchの方法	2.9685	13.6534	0.0104 *						
ランク化	t検定	2.8254	14	0.0135 *						
	Welchの方法	2.8454	13.3662	0.0135 *						
一元配置分散分析で分析すると										
基本統計量										
目的変数	モデル	因子A	n	平均	標準偏差(SD)	平均+SD	標準誤差(SE)	平均-SE	平均+SE	
変数Y	因子A	Aクラス	9	56.333	7.937	48.396	64.271	2.646	53.688	58.979
		Bクラス	7	45.143	7.105	38.038	52.248	2.685	42.458	47.828

分散分析にも Welch の統計量がある

等分散性の検定										
目的変数	パートレット検定	自由度	P 値	F 値	自由度1	自由度2	P 値			
変数Y	0.0775	1	0.7807	0.1621	1	14	0.6933			
分散分析表										
因子	Type III 平方和	自由度	平均平方	F 値	P 値					
因子A	493.0804	1	493.0804	8.5556	0.0111 *					
誤差	806.8571	14	57.6327							
全体	1299.9375	15								
等分散を仮定しない検定										
手法	F 値	自由度1	自由度2	P 値		*: P<0.05 **: P<0.01				
Welch	8.8120	1	13.6534	0.0104 *						
Brown-Forsythe	8.8120	1	13.6534	0.0104 *						

参照)分散分析から Student's t test 式を導く →PDF2 p29

二変量データの分析

二変量データの分析
例)

身長	体重
157.7	51.9
163.8	47.7
155.6	45.3
159.4	57.1
163.2	57.9
163	46.3
162.4	49.9
152.5	41.4
154.9	47.3
153.2	42.7
161.5	63
160	48.8
160.3	51.9
158.1	51
155	53.3
152.2	48.4
155	53.4

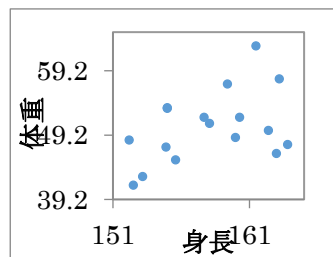
基本統計量

	身長	体重
n	17	17
平均	158.1059	50.4294
不偏分散	15.5468	30.5610
標準偏差	3.9429	5.5282
最小値	152.2000	41.4000
最大値	163.8000	63.0000
尖度	-1.4248	0.3988
歪度	-0.0428	0.5406

相関行列

	身長	体重
身長	1.0000	0.4477
体重	0.4477	1.0000

散布図(身長→体重)



- ☆ 例えば独立 2 群の t 検定は 1 変量 2 群の検定であり、1 変量 (= 要因) 3 群 (= 水準) では分散分析、2 変量のデータでは相関、単回帰分析などを行い、多変量の場合目的変数 y 、説明変数 (x_1, x_2, \dots, x_n) などと表現する。

$$\text{平均値: } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{偏差平方和: } S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$\text{不偏分散: } s_x^2 = s_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$s_y^2 = s_{yy} = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$\text{偏差積和: } S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

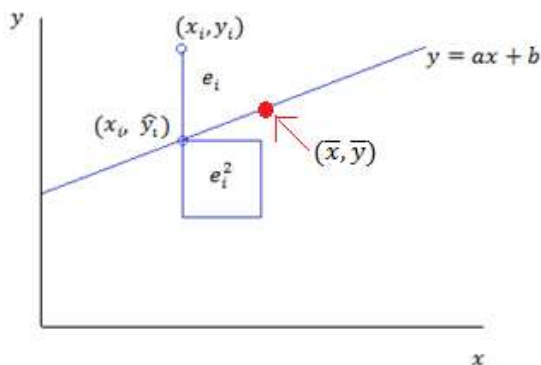
$$\text{共分散: } s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\text{相関係数: } r = \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}}$$

で定義される

単回帰分析 simple linear regression

**最小二乗法の原理を理解するには二次関数をイメージする。例えば $y = x^2 + 2x + 2$ の場合、平方完成すれば $y = (x + 1)^2 + 1$ により $x = -1$ で最小値となることがわかる。二次関数では接線の傾きが頂点で 0 になるので微分して $y' = 2x + 2$ 、 $y' = 0$ のとき同様に $x = -1$ で最小値をとることがわかる。



a(b)に関する二次関数では
a(b)で微分(偏微分)して=0
→ a,b がともに最小となるのは
→ 連立方程式の解

回帰式とし $y = ax + b$ を想定すると、ある観測値 (x_i, y_i) の時、対応する直線上の座標 (\hat{x}_i, \hat{y}_i) は x 座標が同じなので $(x_i, \hat{y}_i) = (x_i, ax_i + b)$ となる。その差 $e_i = y_i - \hat{y}_i$ つまり $e_i = y_i - (ax_i + b)$ を残差(residual)と呼び、この残差の2乗の総和 $Q = \sum e_i^2 = \sum \{y_i - (ax_i + b)\}^2$ が y 軸方向の距離だけを最小になるように、 a, b を決定する。 a は回帰係数。まず Q を a の2次関数とみて、その他は定数として偏微分、次に b の2次関数、その他は定数として偏微分すれば、**の原理から共通の最小値はその連立方程式を解けば良い。

$$\frac{\partial Q}{\partial a} = -2 \sum x_i \{y_i - (ax_i + b)\} = 0$$

$$\frac{\partial Q}{\partial b} = -2 \sum \{y_i - (ax_i + b)\} = 0$$

より

$$\sum x_i y_i - a \sum x_i^2 - b \sum x_i = 0 \quad \dots \textcircled{1}$$

$$\sum y_i - a \sum x_i - bn = 0 \quad \dots \textcircled{2}$$

これを、まず a について解くと、

$$a = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{S_{xy}}{S_{xx}}$$

②式は bn を移項し両辺を n で割れば

$$b = \frac{\sum y_i}{n} - a \frac{\sum x_i}{n} = \bar{y} - a\bar{x}$$

復習

$$\frac{d}{dx} x^n = nx^{n-1}$$

$$\frac{\partial}{\partial x} x^n y^n = y^n nx^{n-1}$$

偏微分では x について微分し y も定数とみなします

例について体重を目的変数と分析すると

	回帰係数	標準誤差	標準回帰 係数	偏相関係 数	t値	F値	P値	95%下 限	95%上 限
定数項	-48.8118	51.19412	-48.8118		-0.95346	0.909095	0.355468	-157.929	60.30591
身長	0.627688	0.323702	0.447694	0.447694	1.939094	3.760087	0.071538	-0.06227	1.317642

体重 = 0.62 × 身長 - 48.8

例について身長を目的変数と分析すると

	回帰係数	標準誤差	標準回帰 係数	偏相関係 数	t値	F値	P値	95%下 限	95%上 限
定数項	142.003	8.351149	142.003		17.00401	289.1363	3.27E-11	124.203	159.8031
体重	0.319315	0.164672	0.447694	0.447694	1.939094	3.760087	0.071538	-0.03168	0.670305

身長=0.32 体重+142

ここで身長、体重を目的変数にした時の回帰係数を掛けあわせ√すると

$\sqrt{0.62 \times 0.32} = 0.45$ となる。ここで相関係数: $r = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$ を計算すると相関行列にあ

る $0.4477 \approx 0.45$ と一致する。XとYの相関係数とは、X→Yの回帰係数とY→Xの回帰係数の(幾何)平均を意味する。

BellCurve for Excel: 多変量解析→重回帰分析(ここで2変量のみを指定すれば単回帰)

単回帰分析⇒解析ソフトは重回帰分析で														
体重を目的変数														
基本統計量														
変数	n	平均	不偏分散	標準偏差	最小値	最大値								
身長	17	158.106	15.547	3.943	152.200	163.800								
体重	17	50.429	30.561	5.528	41.400	63.000								
重回帰係数														
		決定係数												
R	修正R	R2乗	修正R2乗	ダービン=ワトソン比	AIC									
0.4477	0.3836	0.2004	0.1471	2.0418	57.3021									
回帰式の有意性 (分散分析)														
要因	平方和	自由度	平均平方	F 値	P 値									
回帰変動	98.0054	1	98.0054	3.7601	0.0715									
誤差変動	390.9699	15	26.0647											
全体変動	488.9753	16												
回帰式に含まれる変数 (偏回帰係数・信頼区間等)														
		偏回帰係数の95%信頼区間		偏回帰係数の有意性の検定		目的変数との相関		多重共線性の統計量						
変数	偏回帰係数	標準誤差	標準偏回帰係数	下限値	上限値	F 値	t 値	P 値	単相関	偏相関	トレランス	VIF		
身長	0.6277	0.3237	0.4477	-0.0623	1.3176	3.7601	1.9391	0.0715	0.4477	0.4477	1.0000	1.0000		
定数項	-48.8118	51.1941		-157.9295	60.3059	0.9091	-0.9535	0.3555						
身長を目的変数														
基本統計量														
変数	n	平均	不偏分散	標準偏差	最小値	最大値								
体重	17	50.429	30.561	5.528	41.400	63.000								
身長	17	158.106	15.547	3.943	152.200	163.800								
相関行列														
	体重	身長												
体重	1.000	0.448												
身長	0.448	1.000												
重回帰係数														
		決定係数												
R	修正R	R2乗	修正R2乗	ダービン=ワトソン比	AIC									
0.4477	0.3836	0.2004	0.1471	1.1303	45.8124									
回帰式の有意性 (分散分析)														
要因	平方和	自由度	平均平方	F 値	P 値									
回帰変動	49.8569	1	49.8569	3.7601	0.0715									
誤差変動	198.8925	15	13.2595											
全体変動	248.7494	16												
回帰式に含まれる変数 (偏回帰係数・信頼区間等)														
		偏回帰係数の95%信頼区間		偏回帰係数の有意性の検定		目的変数との相関		多重共線性の統計量						
変数	偏回帰係数	標準誤差	標準偏回帰係数	下限値	上限値	F 値	t 値	P 値	単相関	偏相関	トレランス	VIF		
体重	0.3193	0.1647	0.4477	-0.0317	0.6703	3.7601	1.9391	0.0715	0.4477	0.4477	1.0000	1.0000		
定数項	142.0030	8.3511		124.2030	159.8031	289.1363	17.0040	p < 0.001						
体重を目的{身長を目的変数2)														
		1) *2) の平方根		0.447694238		相関係数に一致								
	0.6277	0.3193												

相関係数は、
X→Yの回帰係数
とY→Xの回帰係
数の(幾何)平均を
意味する

☆ 相関係数 $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \Rightarrow r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \Rightarrow r^2 = \frac{S_{xy}}{S_{xx}} * \frac{S_{xy}}{S_{yy}}$

回帰直線は $y = \frac{S_{xy}}{S_{xx}}x + \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x} \Rightarrow y - \bar{y} = \frac{S_{xy}}{S_{xx}}(x - \bar{x})$ となる。また、

$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$ をダービン・ワトソン比 (Durbin-Watson ratio) という。

回帰分析では、異なる誤差項 (実測値と理論値の差) 間には相関がないことを仮定しているが、ダービン=ワトソン比は誤差項間に自己相関があるかないかを判別するための指標。値が 2 よりかなり小さいときは正の相関が、2 よりかなり大きいときは負の相関が、2 前後のときは相関なしと判断する。

実際のデータ (x_i, y_i) 、2 変量の平均値を (\bar{x}, \bar{y}) 、回帰直線上のデータ (x_i, \hat{y}_i)

$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ *

$SST = SSE + SSR$

全変動

$SST(\text{total sum of squares, } S_T) = s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$

残差変動 (残差平方和): \hat{y}_i で説明できなかった変動

$SSE(\text{error sum of squares, } S_e) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

回帰変動: \hat{y}_i で説明された変動

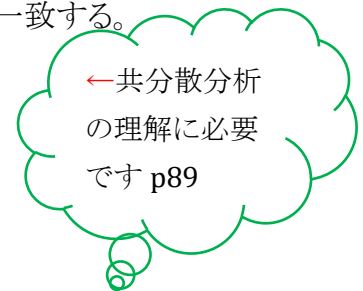
$SSR(\text{regression sum of squares, } S_R) = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$

$R^2 = \frac{SSR}{SST}$ R^2 を決定係数 (寄与率)、 R を重相関係数という。

決定係数は回帰モデルによって実データをどれくらい説明できているか (回帰分析の精度) を表す指標。相関係数 r と決定係数 R^2 は全く別の文脈で定義された量だが、最小二乗法による推定の場合、相関係数の二乗と決定係数は一致する。

ここで $r^2 = \frac{S_{xy}}{S_{xx}} * \frac{S_{xy}}{S_{yy}}$ より $R^2 = \frac{S_{xy}}{S_{xx}} * \frac{S_{xy}}{S_{yy}}$ であり

$R^2 = \frac{SSR}{SST}$ 、 $SST = s_{yy}$ より $\Rightarrow SSR = \frac{S_{xy}}{S_{xx}} * \frac{S_{xy}}{S_{yy}} * s_{yy} = \frac{S_{xy}^2}{S_{xx}}$



* については $\hat{y}_i = \bar{y} + a(x_i - \bar{x})$, $a = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$ を考慮すると $\sum_{i=1}^n (y_i - \bar{y})^2 =$

$\sum_{i=1}^n \{(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})\}^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i) * (\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ より

$\sum_{i=1}^n (y_i - \hat{y}_i) * (\hat{y}_i - \bar{y}) = 0$ を証明すればよい。 $\sum_{i=1}^n (y_i - \hat{y}_i) * (\hat{y}_i - \bar{y}) = \sum_{i=1}^n (y_i - \bar{y} - a(x_i - \bar{x})) * (\bar{y} + a(x_i - \bar{x}) - \bar{y}) = a \{ \sum_{i=1}^n (y_i - \bar{y}) * (x_i - \bar{x}) - a \sum_{i=1}^n (x_i - \bar{x})^2 \} = 0$

相関 correlation

1) Pearson の積率相関係数 Pearson (product-moment) correlation coefficient

$$r = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$$

データが正規分布するときに使用できる。

共分散は 2 変数の平均値と観測値の差の積であり、観測値の分散が大きいほど共分散も大きくなる。異なる指標間で比較できるように各変数の標準偏差で割る。

相関係数の解釈	
相関係数の絶対値	解釈
0.0～0.2	ほとんど相関関係がない
0.2～0.4	やや相関関係がある
0.4～0.7	かなり相関関係がある
0.7～1.0	強い相関関係がある

x, y 相互間の結びつきの強さを示すには、相関係数 r より、それを 2 乗した r^2 を用いるほうが良い。(決定係数 coefficient of determination)

2) 順位相関係数 rank correlation coefficient: Kendall と Spearman

Kendall では対応する 2 つの順位の一致率から算出

Spearman では順位の差から算出

3) 偏相関係数 partial correlation coefficient

偽相関とは変数 A と変数 B が高い相関関係にあるとき、本来的には変数 A と変数 Y のみが相関関係にあるのにも関わらず、変数 B と変数 Y の相関係数も高い値を示す現象を指す。例えば、身長と IQ は恐らく正の相関を示すが、これは実際には年齢と IQ が相関しているだけで、身長が直接 IQ と何か関わりがあるわけではない。見かけ上の相関、疑似相関を危惧してひとつの変数の影響を除いた他の二つの変数の相関係数を求める手法。

x, y, z の変数があり、x, y の相関係数を r_{xy} 、x と z の相関係数を r_{xz} 、y と z の相関係数を r_{yz} としたとき、z を制御変数とした偏相関係数 $r_{xy.z}$ は以下の式になる。 $-1 \leq r \leq 1$ の範囲をとり、絶対値が 1 に近いほど関係は強いことを示す。

$$r_{xy.z} = \frac{r_{xy} - r_{yz}r_{zx}}{\sqrt{1 - r_{yz}^2}\sqrt{1 - r_{zx}^2}}$$

3標本以上の分析

3標本以上の分析に2標本の検定を繰り返すことは検定の多重性といわれ、判定が甘くなる。例えば有意水準5%のt検定を4回繰り返し、すべての2群でt検定を行なうと結果として約20%で検定した場合での有意差となる。個々の仮説の有意水準 α_i 、最終的な有意水準 α とし有意水準 α_i で検定をh回繰り返せば

$$1 - \alpha = \prod_{i=1}^h (1 - \alpha_i) \quad \alpha_i \text{が一定なら}\Pi \text{は}(1 - \alpha_i) \text{を}h \text{回かけることだから}$$

$$= (1 - \alpha_i)^h$$

$$\alpha = 1 - (1 - \alpha_i)^h, \alpha_i \text{が小さい時は}$$

$$\alpha \approx 1 - (1 - h\alpha_i) = h\alpha_i$$

つまり最終的な有意水準はh回検定するとき各検定の有意水準のh倍になる。

検定の多重性の問題点

一元配置分散分析(One Way ANOVA)

ANOVA: analysis of variance

2標本では1変量2群の分析となる。例えば、3標本の一元配置分散分析では1変量3群となり、変量は要因、群は水準と表現し1要因3水準と表現する。

基本的な構造(ある1つの要因がA-Fまでの水準で測定)

水準 A(A ₁)	水準 B(A ₂)	..	水準(A _q)
a ₁	b ₁		f ₁
a ₂	b ₂		f ₂
..	..		f ₃
a _i
	b _j		..
			f _k

ただし、ここではデータ数を同じにして説明

水準 A ₁	水準 A ₂	..	水準 A _a
a ₁₁	a ₂₁		a _{a1}
a ₁₂	a ₂₂		a _{a2}
..
a _{1n}	a _{2n}		a _{an}

全変動 S_T = 水準間変動(級間変動) S_A + 水準内変動(誤差変動) S_E

$$\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x})^2 = \sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

データの構造

$x_{ij} = \bar{x} + (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i)$ であるから
 各データ(x_{ij}) は平均(\bar{x}), 主効果($\bar{x}_i - \bar{x}$), 誤差($x_{ij} - \bar{x}_i$) の和に変形できる
 ($x_{ij} = \mu + \alpha_i + \varepsilon_{ij}$)

⇒ \bar{x} を移項して

$$(x_{ij} - \bar{x}) = (\bar{x}_i - \bar{x}) + (x_{ij} - \bar{x}_i) \Rightarrow \text{両辺を二乗し各データを合計すると}$$

a:水準の数 a 個 n:水準でのデータの数(各水準とも n にした場合)のとき

$$\sum_{i=1}^a \sum_{j=1}^n (\bar{x}_i - \bar{x})(x_{ij} - \bar{x}_i) = 0 \text{ なので (} \rightarrow \text{ p32 参照)}$$

$$\sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x})^2 = \sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^a \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

が成り立つ。

ここで自由度は水準間変動では a-1、全変動では an-1 (= N(総数)-1)、

誤差変動では an-a(= a(n-1)= N-a)より

一元配置分散分析表

a:水準の数 a 個 n:水準でのデータの数(→各水準とも n にした場合)

変動の要因	偏差平方和	自由度	不偏分散	F_0
水準間変動	S_A	a-1	$V_A = \frac{S_A}{a-1}$	$\frac{V_A}{V_E}$
水準内変動	S_E	a(n-1)	$V_E = \frac{S_E}{a(n-1)}$	

$$F_0 = \frac{V_A}{V_E} \geq F_{(a-1, a(n-1))}(\alpha)$$

ならば、有意水準 α で仮設 H_0 を棄却する

例題) A(=水準A₁),B(=水準A₂),C(=水準A₃)各メーカーの製品比較のため100個入り製品をメーカーごとに4箱ずつ不良品の数を調べた(計12箱1200個)。

例解統計入門(鈴木義一郎p163)より

メーカー不良品の数計

メーカー	1箱ごとの不良品の数				計	
水準 A ₁ (A)	12	8	10	6	36(平均 9 = \bar{x}_1)	96 (平均 96/12=8 = \bar{x})
水準 A ₂ (B)	14	10	6	6	36(平均 9 = \bar{x}_2)	
水準 A ₃ (C)	4	6	4	10	24(平均 6 = \bar{x}_3)	

元のデータ			
12	8	10	6
14	10	6	6
4	6	4	10

(i:行、j:列)

(各データ x_{ij} - 全体の平均 \bar{x}) = ばらつき①				メーカーのばらつきの平均
4	0	2	-2	1
6	2	-2	-2	1
-4	-2	-4	2	-2

(ばらつき① - メーカーによるばらつきの平均) = 変動誤差 ②			
4-1=3	0-1=-1	2-1=1	-2-1=-3
6-1=5	2-1=1	-2-1=-3	-2-1=-3
-4-(-2)=-2	-2-(-2)=0	-4-(-2)=-2	2-(-2)=4

データの構成

x_{ij}				\bar{x}	$\bar{x}_i - \bar{x}$	$x_{ij} - \bar{x}_i$				
12	8	10	6	= 8	+	1,1,1,1				
14	10	6	6			1,1,1,1				
4	6	4	10			-2,-2,-2,-2				
						+	3	-1	1	-3
							5	1	-3	-3
							-2	0	-2	4

$x_{ij} - \bar{x}$				$\bar{x}_i - \bar{x}$	$x_{ij} - \bar{x}_i$				
4	0	2	-2	=	+	1,1,1,1			
6	2	-2	-2			1,1,1,1			
-4	-2	-4	2			-2,-2,-2,-2			
						3	-1	1	-3
						5	1	-3	-3
						-2	0	-2	4

(総平均とデータとの差) = (総平均と水準A_iの平均との差) + (水準A_iとデータとの差)

$(x_{ij} - \bar{x})^2$					$(\bar{x}_i - \bar{x})^2$					$(x_{ij} - \bar{x}_i)^2$			
16	0	4	4	=	1,1,1,1				+	9	1	1	9
36	4	4	4		1,1,1,1					25	1	9	9
16	4	16	4		4,4,4,4					4	0	4	16
112					24					88			

$$\text{全変動 } S_T = \text{水準間変動 } S_A + \text{水準内変動 } S_E$$

級間変動
誤差(残差)変動

一元配置分散分析表

変動の要因	偏差平方和	自由度	不偏分散	F_0
水準間変動	24	3-1=2	$V_A = \frac{24}{2} = 12$	$\frac{V_A}{V_E} = \frac{12}{9.78} = 1.23$
水準内変動	88	3(4-1)=9	$V_E = \frac{88}{9} = 9.78$	

$$F_0 = \frac{V_A}{V_E} \geq F_{(a-1, a(n-1))}(\alpha)$$

ならば、有意水準 α で仮設 H_0 を棄却する

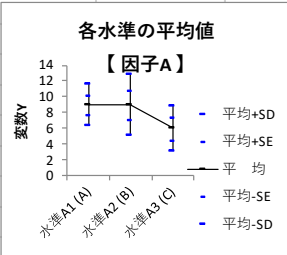
この例では $F_0 = 1.23 \leq F_{(2,9)}(0.05) = 4.2$ (有意ではない)

エクセル関数:=F.DIST.RT(1.23,2,9)=0.337101

BellCurve for Excel: 分散分析多重比較→一元配置分散分析→多重比較の選択

メーカー	1箱ごとの不良品の数			
水準A1 (A)	12	8	10	6
水準A2 (B)	14	10	6	6
水準A3 (C)	4	6	4	10
	水準A1 (A)	水準A2 (B)	水準A3 (C)	
	12	14	4	列3水準に並べ替え
	8	10	6	
	10	6	4	
	6	6	10	

分析結果

一元配置分散分析										
基本統計量										
目的変数	モデル	因子A	n	平均	標準偏差(平均-SD	平均+SD	標準誤差(平均-SE	平均+SE
変数Y	因子A	水準A1 (A)	4	9.000	2.582	6.418	11.582	1.291	7.709	10.291
		水準A2 (B)	4	9.000	3.830	5.170	12.830	1.915	7.085	10.915
		水準A3 (C)	4	6.000	2.828	3.172	8.828	1.414	4.586	7.414
										
等分散性の検定										
バートレット検定				ルビーン検定						
目的変数	カイ二乗値	自由度	P 値	F 値	自由度1	自由度2	P 値			
変数Y	0.4656	2	0.7923	0.6000	2	9	0.5694			

分散分析表										
因子	Type III 平方和	自由度	平均平方	F 値	P 値	*: P<0.05 **: P<0.01				
因子A	24.0000	2	12.0000	1.2273	0.3378					
誤差	88.0000	9	9.7778							
全体	112.0000	11								
等分散を仮定しない検定										
手法	F 値	自由度1	自由度2	P 値	*: P<0.05 **: P<0.01					
Welch	1.2845	2	5.8686	0.3446						
Brown-Forsythe	1.2273	2	7.9780	0.3430						
多重比較検定										
因子	目的変数	手法	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値
因子A	変数Y	Bonferroni	水準A1 (A)	水準A2 (B)	9.0000	9.0000	0.0000	2.2111	0.0000	1.0000
			水準A1 (A)	水準A3 (C)	9.0000	6.0000	3.0000	2.2111	1.3568	0.6237
			水準A2 (B)	水準A3 (C)	9.0000	6.0000	3.0000	2.2111	1.3568	0.6237
		Tukey-Kramer	水準A1 (A)	水準A2 (B)	9.0000	9.0000	0.0000	2.2111	0.0000	1.0000
			水準A1 (A)	水準A3 (C)	9.0000	6.0000	3.0000	2.2111	1.3568	0.4017
			水準A2 (B)	水準A3 (C)	9.0000	6.0000	3.0000	2.2111	1.3568	0.4017

- 一元配置分散分析の前提条件:各水準の分散は等しい
 - ⇒ 本来、等分散の検定を事前に行ってから分散分析に入る
- 等分散の検定: Bartlett検定、Levene検定
(PCで解析するときチェックを入れる。以下の計算が行われる。)

参照) <http://kusuri-jouhou.com/statistics/bartlett.html>

Bartlett検定(元のデータが正規分布していることが前提)

3群以上の標本を得たとき、それぞれの母分散が等しいことを検定する。

帰無仮説 H_0 : 各群の母分散は等しい $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$

対立仮説 H_1 : 少なくとも一群 (i, j) の母分散は等しくない $\sigma_i \neq \sigma_j$

群の数: k 、 i 群の不偏分散: s_i^2 、 i 群のサンプル数: N_i 、総サンプル数: N

検定式:

$$T = \frac{A}{C} \quad A = (N - k) \ln \frac{U}{N - k} - \sum_{i=1}^k (N_i - 1) \ln s_i^2$$

$$U = \sum_{i=1}^k (N_i - 1) s_i^2 \quad C = 1 + \frac{1}{3(k-1)} \left\{ \sum_{i=1}^k \left(\frac{1}{N_i - 1} \right) - \frac{1}{N - k} \right\}$$

$\frac{U}{N-k}$ は各群の不偏分散に対する重み付き平均

検定: 検定統計量 T は、自由度が $(k-1)$ の χ^2 分布に従う。

有意確率の判定

$$T > \chi_{(\alpha, k-1)}^2 \quad \chi_{(\alpha, k-1)}^2 : \text{自由度が } (k-1) \text{ の } \chi^2 \text{ 分布の上側確率値}$$

\Rightarrow 帰無仮説を否定できる (分散は均一ではない)

$$T \leq \chi_{(\alpha, k-1)}^2 \quad \Rightarrow \text{帰無仮説を否定出来ない (分散は均一)}$$

例

	測定値	データ数	分散
A1	○○○	3	10
A2	○○○○○	5	19
A3	○○○○	4	8

帰無仮説(H_0): 各群の分散は均一である、対立仮説(H_1): 各群の分散は均一でない

$$U = (3 - 1) * 10 + (5 - 1) * 19 + (4 - 1) * 8 = 120$$

$$\frac{U}{N-k} = \frac{(3-1)*10+(5-1)*19+(4-1)*8}{12-3} = 13.3$$

$$A = (12 - 3) \ln 13.3 - \{ (3 - 1) \ln 10 + (5 - 1) \ln 19 + (4 - 1) \ln 8 \} = 0.6686$$

$$C = 1 + \frac{1}{3(3-1)} \left\{ \frac{1}{3-1} + \frac{1}{5-1} + \frac{1}{4-1} - \frac{1}{12-3} \right\} = 1.20$$

$$T = \frac{0.6686}{1.2} = 0.5572$$

自由度 $df=k-1=3-1=2$ の χ^2 分布に従うので、 χ^2 分布表より $\chi^2_{0.05}=5.99$ となる。

χ^2 値 $< \chi^2_{0.05}$ なので、 $P > 0.05$ となり帰無仮説を棄却できない。つまり、「各群の分散は均一でない」とはいえない。

Levene検定(ルビーン検定) (元のデータが非正規でも可)

F検定と同じ、分散の検定、外れ値に強い特徴あり。偏差(測定値-平均値)の絶対値を観測値とみなして「平均値の差のt検定」をする。F検定は偏差(測定値-平均値)の2乗の和を使い、Leveneは偏差の絶対値の和を使う。F検定は外れ値があると、それが2乗で効いてくるので有意となりやすい欠点がある。Leveneの検定は外れ値の影響を受け難いという利点がある。この方法は3組以上の比較にも拡張される。

Levene検定統計量 W は

$$W = \frac{(N-k) \sum_{i=1}^k n_i (\bar{z}_i - \bar{z}_{..})^2}{(k-1) \sum_{i=1}^k \sum_{j=1}^{n_i} (z_{ij} - \bar{z}_i)^2} \text{と定義}$$

$$z_{ij} = |y_{ij} - \bar{y}_i| \quad \text{ただし、} \bar{y}_i \text{ は } i \text{ 番目の小標本の平均値}$$

$W > F(\alpha, k-1, N-k) \Rightarrow$ 帰無仮説を否定できる(分散は均一ではない)

$W \leq F(\alpha, k-1, N-k) \Rightarrow$ 帰無仮説を否定出来ない(分散は均一)

➤ Bartlett検定、Levene検定で「分散は均一でない」と判断されたとき

Kruskal-Wallis検定(ノンパラメトリック法) p128

Bartlett検定で「各群の分散は均一である」と判断されれば一元配置分散分析法をすればよいが、「各群の分散は均一でない」と判断されたならKruskal-Wallis検定をする。

帰無仮説(H_0):「各群で差がない」と仮定する。

対立仮説(H_1):「各群で差がある」と仮定する。

Kruskal-Wallis検定では、まずすべてのデータを小さいものから1番→2番→……→N番と順位を付ける。この順位を各群ごとに足して R_i を求める。 R_i を求めたら、下の計算式によって統計量 H を計算する。(N=データの総数、k=群数、 n_i =群内のデータ数)

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

確率を求めるとき、条件によって異なる求め方が必要。

• $k=3$ かつ $N \leq 17$ のとき → Kruskal-Wallis検定表から求める

$P \geq \alpha$ のとき帰無仮説を棄却できない。 $P < \alpha$ のとき帰無仮説を棄却する。

• 上の条件以外 → H は近似的に自由度 $df=k-1$ の χ^2 分布に従う → χ^2 分布表

$\chi^2 \leq \chi^2_{\alpha}$ のとき、 $P \geq \alpha$ となり帰無仮説を棄却できない、

$\chi^2 > \chi^2_{\alpha}$ のとき、 $P < \alpha$ となり帰無仮説を棄却する

Kruskal-Wallis検定表										
n1	n2	n3	P<0.05	P<0.01		n1	n2	n3	P<0.05	P<0.01
2	2	2	-	-		3	3	3	5.6	7.2
2	2	3	4.717	-		3	3	4	5.791	6.746
2	2	4	5.333	-		3	3	5	5.649	7.079
2	2	5	5.16	6.533		3	3	6	5.615	7.41
2	2	6	5.436	6.655		3	3	7	5.62	7.228
2	2	7	5.143	7		3	3	8	5.617	7.35
2	2	8	5.356	6.664		3	3	9	5.589	7.422
2	2	9	5.26	6.897		3	3	10	5.588	7.372
2	2	10	5.12	6.537		3	3	11	5.583	7.418
2	2	11	5.164	6.766		3	4	4	5.599	7.144
2	2	12	5.173	6.761		3	4	5	5.656	7.445
2	2	13	5.199	6.792		3	4	6	5.61	7.5
2	3	3	5.361	-		3	4	7	5.623	7.55
2	3	4	5.444	6.444		3	4	8	5.623	7.585
2	3	5	5.251	6.909		3	4	9	5.652	7.614
2	3	6	5.349	6.97		3	4	10	5.661	7.617
2	3	7	5.357	6.839		3	5	5	5.706	7.578
2	3	8	5.316	7.022		3	5	6	5.602	7.591
2	3	9	5.34	7.006		3	5	7	5.607	7.697
2	3	10	5.362	7.042		3	5	8	5.614	7.706
2	3	11	5.274	7.094		3	5	9	5.67	7.733
2	3	12	5.35	7.1134		3	6	6	5.625	7.725
2	4	4	5.455	7.036		3	6	7	5.689	7.756
2	4	5	5.273	7.205		3	6	8	5.678	7.796
2	4	6	5.34	7.34		3	7	7	5.688	7.81
2	4	7	5.376	7.321		4	4	4	5.692	7.654
2	4	8	5.393	7.35		4	4	5	5.657	7.76
2	4	9	5.4	7.364		4	4	6	5.681	7.795
2	4	10	5.345	7.357		4	4	7	5.65	7.814
2	4	11	5.365	7.396		4	4	8	5.779	7.853
2	5	5	5.339	7.339		4	4	9	5.704	7.91
2	5	6	5.339	7.376		4	5	5	5.666	7.823
2	5	7	5.393	7.45		4	5	6	5.661	7.936
2	5	8	5.415	7.44		4	5	7	5.733	7.931
2	5	9	5.396	7.447		4	5	8	5.718	7.992
2	5	10	5.42	7.514		4	6	6	5.724	8
2	6	6	5.41	7.467		4	6	7	5.706	8.039
2	6	7	5.357	7.491		5	5	5	5.78	8
2	6	8	5.404	7.522		5	5	6	5.729	8.028
2	6	9	5.392	7.566		5	5	7	5.708	8.108
2	7	7	5.398	7.491		5	6	6	5.765	8.124
2	7	8	5.403	7.571						

メーカー不良品を Kruskal-Wallis 検定で手書きで分析すると

12	8	10	6	⇒	順位に変換				R_i	R_i^2
14	10	6	6		11	7	9	4.5	31.5	992.3
4	6	4	10		12	9	4.5	4.5	30	900
					1.5	4.5	1.5	9	16.5	272.3

$$H = \frac{12}{12(12+1)} \left(\frac{992.3}{4} + \frac{900}{4} + \frac{272.3}{4} \right) - 3(12+1) = \frac{541.2}{13} - 39 = 41.6 - 39 = 2.6 < 5.692$$

メーカー不良品をKruskal-Wallis検定(PC)で分析すると

BellCurve for Excel: ノンパラメトリック検定→クラスカル=ウォリス検定と多重比較→多重比較の選択

クラスカル=ウォリス検定と多重比較					
水 準	水準A1 (A)	水準A2 (B)	水準A3 (C)		
n	4	4	4		
平均順位	7.88	7.50	4.13		
クラスカル=ウォリス検定					
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01		
2.7703	2	0.2503			
多重比較 : Scheffe					
水準1	水準2	カイ二乗値	P 値	* : P<0.05 ** : P<0.01	
水準A1 (A)	水準A2 (B)	0.0228	0.9886		
水準A1 (A)	水準A3 (C)	2.2832	0.3193		
水準A2 (B)	水準A3 (C)	1.8494	0.3967		
多重比較 : Shirley-Williams					
対立仮説	水準1	水準2	統計量	棄却値(5%)	棄却値(2.5%)
対照群 ≦	水準A1 (A)	水準A3 (C)	-0.8311	1.716	2.015
	水準A1 (A)	水準A2 (B)	-0.1488	1.645	1.960
対照群 ≧	水準A1 (A)	水準A3 (C)	1.5110	1.716	2.015
	水準A1 (A)	水準A2 (B)	0.1488	1.645	1.960

追記(ノンパラメトリックを使用しない考え方)

独立2群間の検定でWelchのt検定が勧められたと同様、母分散の異なる(可能性のある)3群以上の平均値の差には、予備検定なしで等分散性を仮定しない一元配置分散分析(Welchの方法(ANOVAの結果表示として指定))を行うことが推奨される。

三群の平均値(代表値)の差を検定するとき

(<http://aoki2.si.gunma-u.ac.jp/lecture/BF/sankouzu.html>)

二元配置分散分析(Two Way ANOVA)

二元配置分散分析では要因 A は a 水準、要因 B は b 水準、 A_x, B_y での繰り返し数を (等しくなくても良いが) 今回は等しく n の分析とする。

二元配置の数学モデル

	B ₁	B ₂	...	B _b	計	平均
A ₁	X ₁₁₁ ... X _{11n}	X ₁₂₁ ... X _{12n}	...	X _{1b1} ... X _{1bn}		
A ₂	X ₂₁₁ ... X _{21n}	X ₂₂₁ ... X _{22n}	...	X _{2b1} ... X _{2bn}		
...			
A _a	X _{a11} ... X _{a1n}	X _{a21} ... X _{a2n}	...	X _{ab1} ... X _{abn}		
計						
平均						

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

α_i : 水準 A_i の主効果、 β_j : 水準 B_j の主効果
 $(\alpha\beta)_{ij}$: (A_i, B_j) の交互作用、 ε_{ij} : 誤差項

(各データ) = (全体の平均) + (列の効果) + (行の効果) + (交互作用の効果) + (誤差)

データの構造

$$x_{ijk} = \bar{x} + (\bar{x}_{i*} - \bar{x}) + (\bar{x}_{*j} - \bar{x}) + (\bar{x}_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x}) + (x_{ijk} - \bar{x}_{ij})$$

と式を変形した時、 \bar{x} を移項して

$$(x_{ijk} - \bar{x}) = (\bar{x}_{i*} - \bar{x}) + (\bar{x}_{*j} - \bar{x}) + (\bar{x}_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x}) + (x_{ijk} - \bar{x}_{ij})$$

ここで両辺を 2 乗しデータを合計すると

全変動 S_T = 水準 A_i 間変動 S_A + 水準 B_j 間変動 S_B + 交互作用 $S_{A \times B}$ + 水準内変動 S_E

$$\sum \sum \sum (x_{ijk} - \bar{x})^2 = \sum b n (\bar{x}_{i*} - \bar{x})^2 + \sum a n (\bar{x}_{*j} - \bar{x})^2 + \sum \sum n (\bar{x}_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2 + \sum \sum \sum (x_{ijk} - \bar{x}_{ij})^2$$

が成り立つ

\bar{x} : 総平均

水準 A_i の主効果 $(\bar{x}_{i\cdot} - \bar{x})$: (データのある行の平均) - (総平均)

水準 B_j の主効果 $(\bar{x}_{\cdot j} - \bar{x})$: (データのある列の平均) - (総平均)

交互作用 $(\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}) = (\bar{x}_{ij} - (\bar{x}_{i\cdot} - \bar{x}) - (\bar{x}_{\cdot j} - \bar{x}) - \bar{x})$:

(少ブロック (A_i, B_j) の n 個の平均)

-(データのある行の平均) - (データのある列の平均) + (総平均)

誤差 $(x_{ijk} - \bar{x}_{ij})$:

(データの値) - (少ブロック (A_i, B_j) の n 個の平均)

言い換えると

(個々の値) = (全体の平均) + (列の効果) + (行の効果) + (交互作用の効果) + (誤差)

\Rightarrow (個々の値) - (全体の平均) = (列の効果) + (行の効果) + (交互作用の効果) + (誤差)

(行の効果) = (行の平均) - (全体の平均)

(列の効果) = (列の平均) - (全体の平均)

(交互作用の効果) = (小ブロックの平均) - (全体平均) - (列の効果) - (行の効果)

(誤差) = (個々の値) - (全体の平均) - (列の効果) - (行の効果) - (交互作用の効果)

変動の要因	偏差平方和	自由度	不偏分散	F_0
水準 A_i 間変動	S_A	$a-1$	$V_A = \frac{S_A}{a-1}$	$\frac{V_A}{V_E}$
水準 B_j 間変動	S_B	$b-1$	$V_B = \frac{S_B}{b-1}$	$\frac{V_B}{V_E}$
交互作用 $A \times B$ の変動	$S_{A \times B}$	$(a-1)(b-1)$	$V_{A \times B} = \frac{S_{A \times B}}{(a-1)(b-1)}$	$\frac{V_{A \times B}}{V_E}$
水準内変動	S_E	$ab(n-1)$	$V_E = \frac{S_E}{ab(n-1)}$	

仮説 $H_A: \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$ (水準 A_i の主効果は等しい)

仮説 $H_B: \beta_1 = \beta_2 = \dots = \beta_b = 0$ (水準 B_j の主効果は等しい)

仮説 $H_{A \times B}: (\alpha\beta)_{11} = \dots = (\alpha\beta)_{ab} = 0$ (交互作用はない)

1) 交互作用がない場合 $\Rightarrow \frac{V_{A \times B}}{V_E} \leq F_{0(a-1)(b-1), ab(n-1)}(0.05)$ の時

主効果の判定 (仮説 H_A 、仮説 H_B の検定) を行う

2) 交互作用がある場合 (相乗効果または相殺効果がある = 目的の一つ)

$$\Rightarrow \frac{V_{A \times B}}{V_E} \geq F_{0(a-1)(b-1), ab(n-1)}(0.05) \text{ の時}$$

仮説 $H_{A \times B}$ が捨てられ交互作用があれば主効果の検定はしない。

代わりに組 $(A_1B_1), (A_1B_2), \dots, (A_2B_1), \dots, (A_aB_1), \dots, (A_aB_b)$ を 1 つの因子とした一元配置分散分析を行なう。

例)ある薬剤について 100mg、600mg、2400mg と 3 つの異なる用量で、3 時間から 24 時間まで時間を変えて表皮細胞の分裂の割合を測定した。用量の違い、時間の違いによって、分裂の割合が異なるか、有意水準 5% で分散分析する。

(要因 A は時間(水準 3, 6, 12, 24 時間)、要因 B は用量)

BellCurve エクセル統計より

「薬剤による表皮細胞分裂の割合」石村 貞夫. (1994). 『すぐわかる統計処理』. 東京図書.

薬剤による表皮細胞分裂の割合			
	100mg	600mg	2400mg
3 時間	13.2	16.1	9.1
	15.7	15.7	10.3
	11.9	15.1	8.2
6 時間	22.8	24.5	11.9
	25.7	21.2	14.3
	18.5	24.2	13.7
12 時間	21.8	26.9	15.1
	26.3	31.3	13.6
	32.1	33.8	16.2
24 時間	25.7	30.1	15.2
	28.8	33.8	17.3
	29.5	29.6	14.8

手順 1) 平均 (\bar{x}_{ij} と \bar{x}_i 、 \bar{x}_j 、 \bar{x}) を計算

	100mg	600mg	2400mg	平均
3 時間	13.6	15.63	9.2	12.81
6 時間	22.33	23.3	13.3	19.64
12 時間	26.73	28.83	14.97	23.51
24 時間	28	31.17	15.77	24.98
平均	22.665	24.7325	13.31	20.24

行、列の平均、小ブロックの平均 \bar{x}_{ij} ■
総平均を求めることからスタート。

	B ₁	B ₂	B ₃	平均
A ₁	\bar{x}_{11}	\bar{x}_{12}		$\bar{x}_{1.}$
A ₂		\bar{x}_{23}		$\bar{x}_{2.}$
A ₃	\bar{x}_{31}		\bar{x}_{33}	$\bar{x}_{3.}$
A ₄		\bar{x}_{42}		$\bar{x}_{4.}$
平均	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$	\bar{x}

手順 2) $x_{ijk} - \bar{x}$

	100mg	600mg	2400mg
3 時間	-7.04	-4.14	-11.14
	-4.54	-4.54	-9.94
	-8.34	-5.14	-12.04
6 時間	2.56	4.26	-8.34
	5.46	0.96	-5.94
	-1.74	3.96	-6.54
12 時間	1.56	6.66	-5.14
	6.06	11.06	-6.64
	11.86	8.06	-4.04
24 時間	5.46	9.86	-5.04
	8.56	13.56	-2.94
	9.26	9.36	-5.44

各データから総平均を引く。

全変動 $S_T = \sum \sum \sum (x_{ijk} - \bar{x})^2$

	100mg	600mg	2400mg	合計
3 時間	49.5616	17.1396	124.0996	190.8008
	20.6116	20.6116	98.8036	140.0268
	69.5556	26.4196	144.9616	240.9368
6 時間	6.5536	18.1476	69.5556	94.2568
	29.8116	0.9216	35.2836	66.0168
	3.0276	15.6816	42.7716	61.4808
12 時間	2.4336	44.3556	26.4196	73.2088
	36.7236	122.3236	44.0896	203.1368
	140.6596	64.9636	16.3216	221.9448
24 時間	29.8116	97.2196	25.4016	152.4328

各データから総平均を引いた値を二乗する。合計すれば全変動
→変動は偏差平方(和)でみる

	73.2736	183.8736	8.6436	265.7908
	85.7476	87.6096	29.5936	202.9508
合計	547.7712	699.2672	665.9452	1912.984

行(列)平均から総平均を引いた値(効果)の二乗の総和は行(列)間変動

手順 3) 水準 A_i 間変動 $S_A = \sum b n (\bar{x}_{i*} - \bar{x})^2$ $\{(行の平均)-(全体の平均)\}^2$

3 時間	55.2049	$\leftarrow (12.81-20.24)^2$
6 時間	0.36	$\leftarrow (19.64-20.24)^2$
12 時間	10.6929	$\leftarrow (23.51-20.24)^2$
24 時間	22.4676	$\leftarrow (24.98-20.24)^2$
	88.7254	
×水準 B の数×n	798.5286	

*データの数は $a \times b \times n (4 \times 3 \times 3)$ のうち上記は $b \times n (\times 3 \times 3)$ が合計となる。

手順 4) 水準 B_j 間変動 $S_B = \sum a n (\bar{x}_{*j} - \bar{x})^2$ $\downarrow \{(列の平均)-(全体の平均)\}^2$

100mg	600mg	2400mg	
$\downarrow (22.67-20.24)^2$	$\downarrow (24.73-20.24)^2$	$\downarrow (13.31-20.24)^2$	
5.880625	20.18256	48.0249	74.08808
×水準 A の数×n			889.057

*データの数は $a \times b \times n (4 \times 3 \times 3)$ のうち上記は $a \times n (\times 4 \times 3)$ が合計となる。

手順 5) $(\bar{x}_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})$

	100mg	600mg	2400mg
3 時間	-1.635	-1.6725	3.32
6 時間	0.265	-0.8325	0.59
12 時間	0.795	0.8275	-1.61
24 時間	0.595	1.6975	-2.28

交互作用 $S_{A \times B} = \sum \sum n (\bar{x}_{ij} - \bar{x}_{i*} - \bar{x}_{*j} + \bar{x})^2$

	100mg	600mg	2400mg	
3 時間	2.673225	2.797256	11.0224	16.49288
6 時間	0.070225	0.693056	0.3481	1.111381
12 時間	0.632025	0.684756	2.5921	3.908881
24 時間	0.354025	2.881506	5.1984	8.433931
				29.94708
×n(=×3)				89.84123

←交互作用とは小ブロックの平均から総平均、行効果($\bar{x}_{i*} - \bar{x}$)、列効果($\bar{x}_{*j} - \bar{x}$)を引いた値の二乗の総和

*データの数は $a \times b \times n$ ($4 \times 3 \times 3$) のうち上記は $\times n$ ($= \times 3$) が合計となる。

手順 6) 誤差 ($x_{ijk} - \bar{x}_{ij}$)

	100mg	600mg	2400mg
3 時間	-0.4	0.47	-0.1
	2.1	0.07	1.1
	-1.7	-0.53	-1
6 時間	0.47	1.2	-1.4
	3.37	-2.1	1
	-3.83	0.9	0.4
12 時間	-4.93	-1.93	0.13
	-0.43	2.47	-1.37
	5.37	-0.53	1.23
24 時間	-2.3	-1.07	-0.57
	0.8	2.63	1.53
	1.5	-1.57	-0.97

← 誤差はすべてのデータについて属する小ブロックの平均 \bar{x}_{ij} との差であり、それぞれを二乗し合計した総和が誤差変動、または水準内変動と呼ばれる

水準内変動 $S_E \sum \sum \sum (x_{ijk} - \bar{x}_{ij})^2$

	100mg	600mg	2400mg	
3 時間	0.16	0.2209	0.01	
	4.41	0.0049	1.21	
	2.89	0.2809	1	
6 時間	0.2209	1.44	1.96	
	11.3569	4.41	1	
	14.6689	0.81	0.16	
12 時間	24.3049	3.7249	0.0169	
	0.1849	6.1009	1.8769	
	28.8369	0.2809	1.5129	
24 時間	5.29	1.1449	0.3249	
	0.64	6.9169	2.3409	
	2.25	2.4649	0.9409	
	95.2134	27.8001	12.3534	135.3669

分散分析表

変動の要因	偏差平方和	自由度	不偏分散	F_0
水準 A_i 間変動	798.5286	4-1 =3	$V_A = \frac{798.5286}{3} = 266.16$	$\frac{266.16}{5.640} = 47.19^{**}$
水準 B_j 間変動	889.057	3-1 =2	$V_B = \frac{889.057}{2} = 444.54$	$\frac{444.54}{5.640} = 78.82^{**}$
交互作用 $A \times B$ の変動	89.84123	(4-1)(3-1) =6	$V_{A \times B} = \frac{89.84123}{6} = 14.975$	$\frac{14.975}{5.640} = 2.66^*$
水準内変動	135.3669	4*3(3-1) =24	$V_E = \frac{135.3669}{24} = 5.640$	

F.DIST.RT(266.16,3,24)=1.51479E-18

F.DIST.RT(444.54,2,24)=1.0875E-19

F.DIST.RT(2.66,6,24)=0.040173817

↑ 主効果(行間変動、列間変動)、交互作用とも有意→主効果の多重比較は無意味→交互作用を分析

BellCurve for Excel: 分散分析多重比較→二元配置分散分析

二元配置分散分析													
基本統計量													
目的変数	モデル	因子(列)	因子(行)	n	平均	標準偏差(SD)	平均+SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE		
変数Y	因子(列)	100mg		12	22.667	6.586	16.081	29.253	1.901	20.765	24.568		
		600mg		12	25.192	6.902	18.290	32.093	1.992	23.199	27.184		
		2400mg		12	13.308	2.850	10.458	16.159	0.823	12.486	14.131		
			因子(行)		3時間	9	12.811	3.063	9.748	15.874	1.021	11.790	13.832
					6時間	9	19.644	5.227	14.418	24.871	1.742	17.902	21.387
					12時間	9	24.122	7.759	16.364	31.881	2.586	21.536	26.708
				24時間	9	24.978	7.238	17.739	32.216	2.413	22.565	27.391	
		因子(列) * 因子(行)	100mg	3時間	3	13.600	1.931	11.669	15.531	1.115	12.485	14.715	
			100mg	6時間	3	22.333	3.623	18.711	25.956	2.092	20.242	24.425	
			100mg	12時間	3	26.733	5.164	21.570	31.897	2.981	23.752	29.715	
			100mg	24時間	3	28.000	2.022	25.978	30.022	1.168	26.832	29.168	
			600mg	3時間	3	15.633	0.503	15.130	16.137	0.291	15.343	15.924	
			600mg	6時間	3	23.300	1.825	21.475	25.125	1.054	22.246	24.354	
			600mg	12時間	3	30.667	3.493	27.173	34.160	2.017	28.650	32.684	
			600mg	24時間	3	31.167	2.294	28.872	33.461	1.325	29.842	32.491	
			2400mg	3時間	3	9.200	1.054	8.146	10.254	0.608	8.592	9.808	
			2400mg	6時間	3	13.300	1.249	12.051	14.549	0.721	12.579	14.021	
			2400mg	12時間	3	14.967	1.305	13.662	16.272	0.754	14.213	15.720	
			2400mg	24時間	3	15.767	1.343	14.424	17.110	0.775	14.991	16.542	

各水準の平均値
【因子(列)】

各水準の平均値
【因子(行)】

各水準の平均値
【因子(列) * 因子(行)】

等分散性の検定						
	ルビーン検定					
目的変数	F 値	自由度1	自由度2	P 値		
変数Y	1.7651	11	24	0.1183		
分散分析表						
因子	TypeIII平方和	自由度	平均平方	F 値	P 値	* : P<0.05 ** : P<0.01
因子(列)	940.6706	2	470.3353	75.4212	p < 0.001	**
因子(行)	836.7533	3	278.9178	44.7262	p < 0.001	**
因子(列) * 因子(行)	104.0050	6	17.3342	2.7796	0.0339	*
誤差	149.6667	24	6.2361			
全体	2031.0956	35				

ここで交互作用が有意なので主効果の多重比較は無意味⇒一元配置で分析
(下の表を自分で作成)

各小ブロック(A_i, B_j)の組を一つの水準と考える。
ここでは各水準データ数 3 で 12 水準の一元配置分散分析をする

交互作用が有意なので⇒一元配置で分析											
100mg-3	100mg-6	100mg-12	100mg-24	600mg-3	600mg-6	600mg-12	600mg-24	2400mg-3	2400mg-6	2400mg-12	2400mg-24
13.2	22.8	21.8	25.7	16.1	24.5	26.9	30.1	9.1	11.9	15.1	15.2
15.7	25.7	26.3	28.8	15.7	21.2	31.3	33.8	10.3	14.3	13.6	17.3
11.9	18.5	32.1	29.5	15.1	24.2	33.8	29.6	8.2	13.7	16.2	14.8
(一元配置分散分析用に表の組み換え)											

↑ 交互作用が有意なときは各小ブロック(A_i, B_j)の
組を一つの水準と考えた一元配置分散分析する

BellCurve for Excel: 分散分析多重比較→一元配置分散分析

一元配置分散分析						
基本統計量						
目的変数	モデル	因子A	n	平均	標準偏差(SD)	標準誤差(SE)
変数Y	因子A	100mg-3	3	13.600	1.931	1.115
		100mg-6	3	22.333	3.623	2.092
		100mg-12	3	26.733	5.164	2.981
		100mg-24	3	28.000	2.022	1.168
		600mg-3	3	15.633	0.503	0.291
		600mg-6	3	23.300	1.825	1.054
		600mg-12	3	30.667	3.493	2.017
		600mg-24	3	31.167	2.294	1.325
		2400mg-3	3	9.200	1.054	0.608
		2400mg-6	3	13.300	1.249	0.721
		2400mg-12	3	14.967	1.305	0.754
		2400mg-24	3	15.767	1.343	0.775

等分散性の検定							
	バートレット検定			ルビーン検定			
目的変数	カイニ乗値	自由度	P 値	F 値	自由度1	自由度2	P 値
変数Y	13.0752	11	0.2884	1.7651	11	24	0.1183
分散分析表							
因子	TypeIII平方和	自由度	平均平方	F 値	P 値	*: P<0.05 **: P<0.01	
因子A	1881.4289	11	171.0390	27.4272	p < 0.001	**	
誤差	149.6667	24	6.2361				
全体	2031.0956	35					
等分散を仮定しない検定							
手法	F 値	自由度1	自由度2	P 値	*: P<0.05 **: P<0.01		
Welch	26.7945	11	9.2654	p < 0.001	**		
Brown-Forsythe	27.4272	11	10.0775	p < 0.001	**		

多重比較検定

Tukey-Kramer	100mg-3	100mg-6	13.6000	22.3333	8.7333	2.0390	4.2832	0.0108	*
	100mg-3	100mg-12	13.6000	26.7333	13.1333	2.0390	6.4412	p < 0.001	**
	100mg-3	100mg-24	13.6000	28.0000	14.4000	2.0390	7.0624	p < 0.001	**
	100mg-3	600mg-3	13.6000	15.6333	2.0333	2.0390	0.9972	0.9963	
	100mg-3	600mg-6	13.6000	23.3000	9.7000	2.0390	4.7573	0.0035	**
	100mg-3	600mg-12	13.6000	30.6667	17.0667	2.0390	8.3702	p < 0.001	**
	100mg-3	600mg-24	13.6000	31.1667	17.5667	2.0390	8.6155	p < 0.001	**
	100mg-3	2400mg-3	13.6000	9.2000	4.4000	2.0390	2.1580	0.5910	
	100mg-3	2400mg-6	13.6000	13.3000	0.3000	2.0390	0.1471	1.0000	

	100mg-3	2400mg-1	13.6000	14.9667	1.3667	2.0390	0.6703	0.9999	
	100mg-3	2400mg-2	13.6000	15.7667	2.1667	2.0390	1.0626	0.9937	
	100mg-6	100mg-12	22.3333	26.7333	4.4000	2.0390	2.1580	0.5910	
	100mg-6	100mg-24	22.3333	28.0000	5.6667	2.0390	2.7792	0.2492	
	100mg-6	600mg-3	22.3333	15.6333	6.7000	2.0390	3.2860	0.0974	
	100mg-6	600mg-6	22.3333	23.3000	0.9667	2.0390	0.4741	1.0000	
	100mg-6	600mg-12	22.3333	30.6667	8.3333	2.0390	4.0870	0.0171	*
	100mg-6	600mg-24	22.3333	31.1667	8.8333	2.0390	4.3322	0.0097	**
	100mg-6	2400mg-3	22.3333	9.2000	13.1333	2.0390	6.4412	p < 0.001	**
	100mg-6	2400mg-6	22.3333	13.3000	9.0333	2.0390	4.4303	0.0077	**
	100mg-6	2400mg-1	22.3333	14.9667	7.3667	2.0390	3.6129	0.0492	*
	100mg-6	2400mg-2	22.3333	15.7667	6.5667	2.0390	3.2206	0.1109	
	100mg-12	100mg-24	26.7333	28.0000	1.2667	2.0390	0.6212	1.0000	
	100mg-12	600mg-3	26.7333	15.6333	11.1000	2.0390	5.4439	p < 0.001	**
	100mg-12	600mg-6	26.7333	23.3000	3.4333	2.0390	1.6839	0.8587	
	100mg-12	600mg-12	26.7333	30.6667	3.9333	2.0390	1.9291	0.7316	
	100mg-12	600mg-24	26.7333	31.1667	4.4333	2.0390	2.1743	0.5806	
	100mg-12	2400mg-3	26.7333	9.2000	17.5333	2.0390	8.5991	p < 0.001	**
	100mg-12	2400mg-6	26.7333	13.3000	13.4333	2.0390	6.5883	p < 0.001	**
	100mg-12	2400mg-1	26.7333	14.9667	11.7667	2.0390	5.7709	p < 0.001	**
	100mg-12	2400mg-2	26.7333	15.7667	10.9667	2.0390	5.3785	p < 0.001	**
	100mg-24	600mg-3	28.0000	15.6333	12.3667	2.0390	6.0651	p < 0.001	**
	100mg-24	600mg-6	28.0000	23.3000	4.7000	2.0390	2.3051	0.4990	
	100mg-24	600mg-12	28.0000	30.6667	2.6667	2.0390	1.3078	0.9699	
	100mg-24	600mg-24	28.0000	31.1667	3.1667	2.0390	1.5531	0.9096	
	100mg-24	2400mg-3	28.0000	9.2000	18.8000	2.0390	9.2203	p < 0.001	**
	100mg-24	2400mg-6	28.0000	13.3000	14.7000	2.0390	7.2095	p < 0.001	**
	100mg-24	2400mg-1	28.0000	14.9667	13.0333	2.0390	6.3921	p < 0.001	**
	100mg-24	2400mg-2	28.0000	15.7667	12.2333	2.0390	5.9998	p < 0.001	**
	600mg-3	600mg-6	15.6333	23.3000	7.6667	2.0390	3.7601	0.0357	*
	600mg-3	600mg-12	15.6333	30.6667	15.0333	2.0390	7.3730	p < 0.001	**
	600mg-3	600mg-24	15.6333	31.1667	15.5333	2.0390	7.6182	p < 0.001	**
	600mg-3	2400mg-3	15.6333	9.2000	6.4333	2.0390	3.1552	0.1261	
	600mg-3	2400mg-6	15.6333	13.3000	2.3333	2.0390	1.1444	0.9887	
	600mg-3	2400mg-1	15.6333	14.9667	0.6667	2.0390	0.3270	1.0000	
	600mg-3	2400mg-2	15.6333	15.7667	0.1333	2.0390	0.0654	1.0000	
	600mg-6	600mg-12	23.3000	30.6667	7.3667	2.0390	3.6129	0.0492	*
	600mg-6	600mg-24	23.3000	31.1667	7.8667	2.0390	3.8582	0.0287	*
	600mg-6	2400mg-3	23.3000	9.2000	14.1000	2.0390	6.9152	p < 0.001	**
	600mg-6	2400mg-6	23.3000	13.3000	10.0000	2.0390	4.9044	0.0025	**
	600mg-6	2400mg-1	23.3000	14.9667	8.3333	2.0390	4.0870	0.0171	*
	600mg-6	2400mg-2	23.3000	15.7667	7.5333	2.0390	3.6947	0.0412	*
	600mg-12	600mg-24	30.6667	31.1667	0.5000	2.0390	0.2452	1.0000	
	600mg-12	2400mg-3	30.6667	9.2000	21.4667	2.0390	10.5282	p < 0.001	**
	600mg-12	2400mg-6	30.6667	13.3000	17.3667	2.0390	8.5174	p < 0.001	**
	600mg-12	2400mg-1	30.6667	14.9667	15.7000	2.0390	7.7000	p < 0.001	**
	600mg-12	2400mg-2	30.6667	15.7667	14.9000	2.0390	7.3076	p < 0.001	**
	600mg-24	2400mg-3	31.1667	9.2000	21.9667	2.0390	10.7734	p < 0.001	**
	600mg-24	2400mg-6	31.1667	13.3000	17.8667	2.0390	8.7626	p < 0.001	**
	600mg-24	2400mg-1	31.1667	14.9667	16.2000	2.0390	7.9452	p < 0.001	**
	600mg-24	2400mg-2	31.1667	15.7667	15.4000	2.0390	7.5528	p < 0.001	**
	2400mg-3	2400mg-6	9.2000	13.3000	4.1000	2.0390	2.0108	0.6827	
	2400mg-3	2400mg-1	9.2000	14.9667	5.7667	2.0390	2.8282	0.2293	
	2400mg-3	2400mg-2	9.2000	15.7667	6.5667	2.0390	3.2206	0.1109	
	2400mg-6	2400mg-1	13.3000	14.9667	1.6667	2.0390	0.8174	0.9994	
	2400mg-6	2400mg-2	13.3000	15.7667	2.4667	2.0390	1.2098	0.9828	
	2400mg-1	2400mg-2	14.9667	15.7667	0.8000	2.0390	0.3924	1.0000	

例 2) 教科と組別の2つの要因

(<http://www.geisya.or.jp/~mwm48961/statistics/bunsan2.htm> より一部改変)

教科と組別の2つの要因があつて、それらの組合せ(1組の英語など)ごとに繰り返し試験が行われた。教科の要因は英語・数学の2つの値をとり水準は2, 組別の要因は1組・2組の2つの値をとり水準は2となる。各組合せについて、各4回の試験が行われているので、「繰り返し数」は4となる。

- 要因 A は2つの要因(2変量)のように見えるが1要因であり、教科(1要因)の中の2水準、英語、数学ととらえる。要因 B は組(2つの水準、1組、2組)

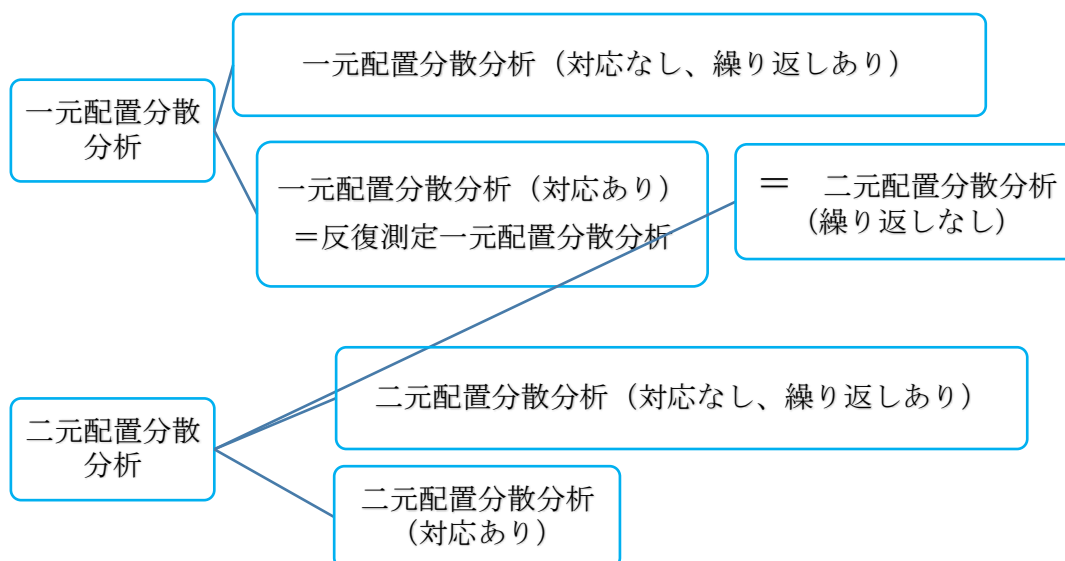
教科	1組	2組
英語	56.4	55.1
	52.3	53.4
	50.2	63.2
	45.2	56.3
数学	58.6	50.6
	67.8	55.3
	60.2	45.8

BellCurve for Excel: 分散分析多重比較→二元配置分散分析

二元配置分散分析											
基本統計量											
目的変数	モデル	因子(列)	因子(行)	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
変数Y	因子(列)	1組		8	55.813	6.840	48.972	62.653	2.418	53.394	58.231
		2組		8	54.363	4.962	49.400	59.325	1.754	52.608	56.117
	因子(行)		英語	8	54.013	5.238	48.775	59.250	1.852	52.161	55.864
			数学	8	56.163	6.522	49.640	62.685	2.306	53.856	58.469
	因子(列) * 因子(行)	1組	英語	4	51.025	4.659	46.366	55.684	2.330	48.695	53.355
			数学	4	60.600	5.133	55.467	65.733	2.566	58.034	63.166
		2組	英語	4	57.000	4.301	52.699	61.301	2.151	54.849	59.151
			数学	4	51.725	4.518	47.207	56.243	2.259	49.466	53.984
等分散性の検定											
	ルビーン検定										
目的変数	F 値	自由度1	自由度2	P 値							
変数Y	0.0314	3	12	0.9921							
分散分析表											
因子	Type III 平方和	自由度	平均平方	F 値	P 値	*: P<0.05 **: P<0.01					
因子(列)	8.4100	1	8.4100	0.3868	0.5456						
因子(行)	18.4900	1	18.4900	0.8505	0.3746						
因子(列) * 因子(行)	220.5225	1	220.5225	10.1430	0.0079**						
誤差	260.8950	12	21.7413								
全体	508.3175	15									
交互作用が有意なので下記は無意味 ⇒ 一元配置分散分析を行う											

ここでも交互作用が有意、(A_i, B_j)の組を一つの水準として一元配置分散分析する

◇ ANOVA(一元配置、二元配置)の解析法の選択に注意



対応あり(反復測定)とは:

1 要因の各データが a 回繰り返し使用(水準数 a)されるということ。

一元配置分散分析(対応あり)

- = 反復測定一元配置分散分析 Repeated measures ANOVA
- = ブロック内での反復のない乱塊法 randomized block design
- = 二元配置分散分析(繰り返しなし) two-way ANOVA without replication

例題)5頭のブタすべてに3種の成長ホルモン刺激剤を与え、各刺激剤ごとに体重増加率を調べる。成長ホルモン刺激剤によって体重の増加率が異なるか、有意水準5%で分散分析を行う。また、どのホルモン剤との間に有意差があるかについて多重比較も行って確認する。

BellCurve エクセル統計より「成長ホルモン刺激剤の用量によるブタの体重の違い」

ブタ	A1	A2	A3
1	9.48	9.24	8.66
2	9.52	9.95	8.50
3	9.32	9.20	8.76
4	9.98	9.68	9.11
5	10.00	9.94	9.75

BellCurve for Excel: 分散分析多重比較→一元配置分散分析(対応あり)

一元配置分散分析(対応あり)

一元配置分散分析 (対応あり)											
基本統計量											
目的変数	モデル	因子A	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE	
変数Y	因子A	A1	5	9.660	0.310	9.350	9.970	0.139	9.521	9.799	
		A2	5	9.602	0.365	9.237	9.967	0.163	9.439	9.765	
		A3	5	8.956	0.497	8.459	9.453	0.222	8.734	9.178	
<div style="display: flex; align-items: center;">  <div style="margin-left: 20px;"> <p>Mauchly の球面性検定⇒p100</p> <p>このデータでは P 値は 0.3752 となり、イプシロンで修正した自由度を用いる必要はない。</p> </div> </div>											
Mauchlyの球面性検定											
被験者内効果	Mauchly's W	カイ二乗値	自由度	P 値	Epsilon	下限値	Greenhouse-Geisser	Huynh-Feldt-Lecoutre	Chi-Muller		
因子A	0.5202	1.9604	2	0.3752	0.5000		0.6758	0.8982	0.5000		
分散分析表											
因子			Type III 平方和	自由度	平均平方	F 値	P 値	*: P<0.05 ** : P<0.01			
被験者内因子	因子A	球面性仮定	1.5272	2	0.7636	11.3944	0.0046 **				
		Greenhouse-Geisser	1.5272	1.3516	1.1299	11.3944	0.0146 *				
		Huynh-Feldt-Lecoutre	1.5272	1.7965	0.8501	11.3944	0.0065 **				
		Chi-Muller	1.5272	1	1.5272	11.3944	0.0279 *				
		下限値	1.5272	1	1.5272	11.3944	0.0279 *				
		誤差 (因子A)	球面性仮定	0.5361	8	0.0670					
被験者間因子	ブタ (誤差)	Greenhouse-Geisser	0.5361	5.4063	0.0992						
		Huynh-Feldt-Lecoutre	0.5361	7.1859	0.0746						
		Chi-Muller	0.5361	4	0.1340						
		下限値	0.5361	4	0.1340						
		全体		1.3717	4	0.3429					
					3.4350	14					
多重比較検定											
因子	目的変数	手法	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値	
因子A	変数Y	Sidak	A1	A2	9.6600	9.6020	0.0580	0.1637	0.3543	0.9808	
			A1	A3	9.6600	8.9560	0.7040	0.1637	4.2999	0.0078 **	
			A2	A3	9.6020	8.9560	0.6460	0.1637	3.9457	0.0127 *	

球面性仮定による分散分析の結果 P 値が **0.0046** となり、有意な結果。成長ホルモン刺激剤の用量によってブタの体重増加率は異なる。↑

二元配置分散分析(繰返しなし)でも同じ結果

二元配置分散分析											
基本統計量											
目的変数	モデル	因子(列)	因子(行)	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
変数Y	因子(列)	A1		5	9.660	0.310	9.350	9.970	0.139	9.521	9.799
		A2		5	9.602	0.365	9.237	9.967	0.163	9.439	9.765
		A3		5	8.956	0.497	8.459	9.453	0.222	8.734	9.178
	因子(行)		1	3	9.127	0.422	8.705	9.548	0.243	8.883	9.370
			2	3	9.323	0.745	8.579	10.068	0.430	8.893	9.753
			3	3	9.093	0.295	8.798	9.388	0.170	8.923	9.264
		4	3	9.590	0.442	9.148	10.032	0.255	9.335	9.845	
		5	3	9.897	0.131	9.766	10.027	0.075	9.821	9.972	

多重比較検定										* : P<0.05 ** : P<0.01	
因子	目的変数	手法	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値	
因子(列)	変数Y	Bonferroni	A1	A2	9.6600	9.6020	0.0580	0.1637	0.3543	1.0000	
			A1	A3	9.6600	8.9560	0.7040	0.1637	4.2999	0.0078	**
			A2	A3	9.6020	8.9560	0.6460	0.1637	3.9457	0.0128	*
		Sidak	A1	A2	9.6600	9.6020	0.0580	0.1637	0.3543	0.9808	
			A1	A3	9.6600	8.9560	0.7040	0.1637	4.2999	0.0078	**
			A2	A3	9.6020	8.9560	0.6460	0.1637	3.9457	0.0127	*
		Tukey-Kramer	A1	A2	9.6600	9.6020	0.0580	0.1637	0.3543	0.9337	
			A1	A3	9.6600	8.9560	0.7040	0.1637	4.2999	0.0065	**
			A2	A3	9.6020	8.9560	0.6460	0.1637	3.9457	0.0106	*

二元配置分散分析(対応あり) two-way repeated measures ANOVA

(*手計算は困難→PCでの解析とその解釈が中心になります。)

例題) 2種類の薬剤AとBの効果を知るため、AとBをそれぞれ6匹のラットに投与し、投与開始から3週経過時点まで1週間毎にある検査を行い検査値の経時的変化のデータを得た。このデータより投与薬剤の違いを被験者間因子、時間経過を被験者内因子として二元配置分散分析を行い、薬剤の効果を検討。

BellCurve エクセル統計より「2種類の薬剤の効果のある検査値でみる実験」

2種類の薬剤の効果のある検査値でみる実験					
個体 No.	投与薬剤	スタート	1週	2週	3週
1	A	7.5	8.6	6.9	0.8
2	A	10.6	11.7	8.8	1.6
3	A	12.4	13.0	11.0	5.6
4	A	11.5	12.6	11.1	7.5
5	A	8.3	8.9	6.8	0.5
6	A	9.2	10.1	8.6	3.8
7	B	13.3	13.3	12.9	11.1
8	B	10.7	10.8	10.7	9.3
9	B	12.5	12.7	12.0	10.1
10	B	8.4	8.7	8.1	5.7
11	B	9.4	9.6	8.0	3.8
12	B	11.3	11.7	10.0	8.5

対応ありとは、同じ動物、人などに経時的に繰り返すこと

個体差による誤差が減り、対応なしより検出力がよい

BellCurve for Excel: 分散分析多重比較→二元配置分散分析(対応あり)

二元配置分散分析 (対応あり)												
基本統計量												
目的変数	モデル	投与薬剤	被験者内因子	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE	
変数Y	投与薬剤	A		24	8.225	3.608	4.617	11.833	0.736	7.489	8.961	
		B		24	10.108	2.349	7.759	12.457	0.480	9.629	10.588	
		被験者内因子		スタート	12	10.425	1.864	8.561	12.289	0.538	9.887	10.963
			1週	12	10.975	1.758	9.217	12.733	0.507	10.468	11.482	
			2週	12	9.575	1.996	7.579	11.571	0.576	8.999	10.151	
			3週	12	5.692	3.664	2.028	9.355	1.058	4.634	6.749	
		投与薬剤 * 被験者内因子	A	スタート	6	9.917	1.903	8.014	11.820	0.777	9.140	10.694
			A	1週	6	10.817	1.888	8.928	12.705	0.771	10.046	11.588
			A	2週	6	8.867	1.884	6.982	10.751	0.769	8.097	9.636
			A	3週	6	3.300	2.834	0.466	6.134	1.157	2.143	4.457
			B	スタート	6	10.933	1.845	9.089	12.778	0.753	10.180	11.686
			B	1週	6	11.133	1.781	9.353	12.914	0.727	10.406	11.860
	B	2週	6	10.283	2.001	8.282	12.285	0.817	9.466	11.100		
	B	3週	6	8.083	2.787	5.296	10.871	1.138	6.945	9.221		

被験者間因子 = 投与薬剤
被験者内因子 = 時間の経過

等分散性の検定											
バートレット検定				ルビーン検定							
被験者内因子	カイ二乗値	自由度	P 値	F 値	自由度1	自由度2	P 値				
スタート	0.0044	1	0.9470	0.0871	1	10	0.7740				
1週	0.0157	1	0.9004	0.1777	1	10	0.6822				
2週	0.0165	1	0.8978	0.0488	1	10	0.8295				
3週	0.0013	1	0.9718	0.0222	1	10	0.8846				
Mauchlyの球面性検定											
Epsilon											
被験者内効果	Mauchly's W	カイ二乗値	自由度	P 値	下限値	Greenhouse-Geisser	Huynh-Feldt-Lecoutre	Chi-Muller			
被験者内因子	0.0110	39.3316	5	p < 0.001	0.3333	0.3693	0.3818	0.3405			
分散分析表											
因子	被験者内因子	TypeIII平方和	自由度	平均平方	F 値	P 値	*: P<0.05 ** : P<0.01				
被験者内因子	球面性仮定	205.1500	3	68.3833	113.6565	p < 0.001 **					
	Greenhouse-	205.1500	1.1078	185.1848	113.6565	p < 0.001 **					
	Huynh-Feldt	205.1500	1.1455	179.0933	113.6565	p < 0.001 **					
	Chi-Muller	205.1500	1.0214	200.8600	113.6565	p < 0.001 **					
	下限値	205.1500	1	205.1500	113.6565	p < 0.001 **					
投与薬剤 * 被験者内因子	球面性仮定	35.5000	3	11.8333	19.6676	p < 0.001 **					
	Greenhouse-	35.5000	1.1078	32.0451	19.6676	p < 0.001 **					
	Huynh-Feldt	35.5000	1.1455	30.9910	19.6676	p < 0.001 **					
	Chi-Muller	35.5000	1.0214	34.7576	19.6676	0.0012 **					
	下限値	35.5000	1	35.5000	19.6676	0.0013 **					
誤差 (被験者内因子)	球面性仮定	18.0500	30	0.6017							
	Greenhouse-	18.0500	11.0781	1.6293							
	Huynh-Feldt	18.0500	11.4549	1.5757							
	Chi-Muller	18.0500	10.2136	1.7673							
	下限値	18.0500	10	1.8050							
被験者間因子	投与薬剤	42.5633	1	42.5633	2.5404	0.1420					
	被験者因子 (誤差)	167.5433	10	16.7543							
全体		468.8067	47								

Mauchlyの球面性検定の項目で、p<0.05であれば「Greenhouse-Geisser」→p110

分散分析表では、「Greenhouse-Geisser」のP値を読む

分散分析表
時間経過(被験者内因子)の主効果が有意、投与薬剤との交互作用も有意、交互作用があるということは、投与薬剤によって経時的変化が異なることを示している。投与薬剤(被験者間因子)については有意な主効果なし

「被験者内因子」の各水準における「投与薬剤」の平均			
目的変数	変数Y	投与薬剤	
《平均》		A	B
被験者内因子	スタート	9.917	10.933
	1週	10.817	11.133
	2週	8.867	10.283
	3週	3.300	8.083
《n》		投与薬剤	
		A	B
被験者内因子	スタート	6	6
	1週	6	6
	2週	6	6
	3週	6	6
	セル内繰り返し数		
	最小	最大	
		6	6

時間経過別の投与薬剤間の平均値比較では3週目の差が最も大きく、時間経過別の投与薬剤の単純主効果検定の結果をみても、有意な差になっているのは3週目のみ

「被験者内因子」の各水準における「投与薬剤」の単純主効果の検定								
目的変数	被験者内因子	因子	平方和	自由度	平均平方和	F 値	P 値	* : P<0.05 ** : P<0.01
変数Y	スタート	投与薬剤	3.1008	1	3.1008		0.6683	0.4296
		誤差		12	4.6398			
	1週	投与薬剤	0.3008	1	0.3008		0.0648	0.8033
		誤差		12	4.6398			
	2週	投与薬剤	6.0208	1	6.0208		1.2976	0.2769
		誤差		12	4.6398			
	3週	投与薬剤	68.6408	1	68.6408		14.7938	0.0023 **
		誤差		12	4.6398			

「被験者内因子」の各水準における「投与薬剤」の多重比較検定											
目的変数	手法	被験者内因子水準1		水準2	平均1	平均2	差	標準誤差	統計量	P 値	
変数Y	Tukey	スタート	A	B	9.9167	10.9333		1.0167	0.6218	1.6350	0.1280
		1週	A	B	10.8167	11.1333		0.3167	0.6218	0.5093	0.6198
		2週	A	B	8.8667	10.2833		1.4167	0.6218	2.2783	0.0418 *
		3週	A	B	3.3000	8.0833		4.7833	0.6218	7.6925	p < 0.001 **

投与薬剤別の経過時間による平均値比較:A 剤の方が経過時間による変化が大きい。
A 剤、B 剤、個別に時間経過の主効果を検定すると、いずれでも有意な効果が見られる。

「投与薬剤」の各水準における「被験者内因子」の単純主効果の検定								
目的変数	投与薬剤	因子	平方和	自由度	平均平方和	F 値	P 値	* : P<0.05 ** : P<0.01
変数Y	A	被験者内因子	205.4750	3	68.4917	113.8366		p < 0.001 **
		誤差	18.0500	30	0.6017			
	B	被験者内因子	35.1750	3	11.7250	19.4875		p < 0.001 **
		誤差	18.0500	30	0.6017			

投与薬剤別に経過時間による差を多重比較検定
 A 剤では、どの経過時間同士の差も有意
 B 剤では、3 週目と他の時間経過の差だけが有意↓

「投与薬剤」の各水準における「被験者内因子」の多重比較検定								* : P<0.05 ** : P<0.01		
目的変数	手法	投与薬剤	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値
変数Y	Tukey	A	スタート	1週	9.9167	10.8167	0.9000	0.3167	2.8421	0.0379 *
			スタート	2週	9.9167	8.8667	1.0500	0.3167	3.3158	0.0122 *
			スタート	3週	9.9167	3.3000	6.6167	0.3167	20.8947	p < 0.001 **
			1週	2週	10.8167	8.8667	1.9500	0.3167	6.1579	p < 0.001 **
			1週	3週	10.8167	3.3000	7.5167	0.3167	23.7368	p < 0.001 **
			2週	3週	8.8667	3.3000	5.5667	0.3167	17.5789	p < 0.001 **
		B	スタート	1週	10.9333	11.1333	0.2000	0.3167	0.6316	0.9210
			スタート	2週	10.9333	10.2833	0.6500	0.3167	2.0526	0.1919
			スタート	3週	10.9333	8.0833	2.8500	0.3167	9.0000	p < 0.001 **
			1週	2週	11.1333	10.2833	0.8500	0.3167	2.6842	0.0540
			1週	3週	11.1333	8.0833	3.0500	0.3167	9.6316	p < 0.001 **
			2週	3週	10.2833	8.0833	2.2000	0.3167	6.9474	p < 0.001 **

- 二元配置分散分析(対応なし)の例題:行は時間の4水準、列は用量で3水準(繰返し数、3)となるが、この場合の実験動物はカエルであり、その都度解剖するため、繰返し使用することはできない。

	100mg	600mg	2400mg
3時間	13.2	16.1	9.1
	15.7	15.7	10.3
6時間	11.9	15.1	8.2
	22.8	24.5	11.9
12時間	25.7	21.2	14.3
	18.5	24.2	13.7
24時間	21.8	26.9	15.1
	26.3	31.3	13.6
	32.1	33.8	16.2
	25.7	30.1	15.2
	28.8	33.8	17.3
	29.5	29.6	14.8

- 二元配置分散分析(対応あり)の例題:行は薬剤でABの2水準、列は時間の4水準(繰返し数6)となる。使用する動物はラットで、経時的に同じラットからデータをとる。

個体No.	投与薬剤	スタート	1週	2週	3週
1	A	7.5	8.6	6.9	0.8
2	A	10.6	11.7	8.8	1.6
3	A	12.4	13.0	11.0	5.6
4	A	11.5	12.6	11.1	7.5
5	A	8.3	8.9	6.8	0.5
6	A	9.2	10.1	8.6	3.8
7	B	13.3	13.3	12.9	11.1
8	B	10.7	10.8	10.7	9.3
9	B	12.5	12.7	12.0	10.1
10	B	8.4	8.7	8.1	5.7
11	B	9.4	9.6	8.0	3.8
12	B	11.3	11.7	10.0	8.5

- 二元配置分散分析(対応あり)の分析では、二元配置分散分析(対応なし)で有意差を計算するときの分母(誤差)から更に各時間で使用される実験動物による個体変動の部分が除去され分母(誤差)が小さくなり有意差がでやすくなる。(同じデータでも一元配置分散分析⇒二元配置分散分析(対応なし)⇒二元配置分散分析(対応あり)の順に有意差がでやすくなる。対応のあるデータかどうかを見極めることが大切)

参考:タイプ I~IV の平方和(名前は SAS 起源、PC で分析時指定する場合あり)

分散分析は各主効果と交互作用を独立して推定できる場合(直交性あり)に平方和を分解するもの。直交性が崩れている場合でも変動要因の平方和を計算する方法として考案されたものがタイプ I~IV の平方和。釣り合い型データではどれでも同じ。非釣り合い型で直交したデータの場合、タイプ I とタイプ II は一致する。タイプ III は交互作用で調整するので一致しない。非直交なデータであっても、交互作用がないモデルの場合は、タイプ II, タイプ III, タイプ IV は一致。交互作用に関しては、タイプ II は主効果の計算において交互作用を無視。タイプ III は交互作用の大きさを「バランスよく」算定し主効果の検討において交互作用を考慮。セルサイズに関しては、タイプ II では各セルのデータ数で重みをつけた加重平均が周辺平均として計算され、主効果が検討される。タイプ III では、セルサイズで等しく重みづけた平均が計算される。空のセルがある時タイプ IV を利用する。要因間に因果的な順序が想定される場合などは、タイプ I を使う。それ以外はタイプ II かタイプ III を使うことが多い。主効果に主たる関心がある場合はタイプ II, 交互作用に関心がある場合はタイプ III ということが多い。

⇒参考としたインターネット ofmind.net/doc/anova-note

コクランの Q 検定

対応のある 3 つ以上の 2 値変数(1/0 データの変数)をもとに、すべての変数で比率に差がないかを検定する。変数がバイナリの場合の Friedman の検定の特殊例として提示され、マクネマー検定を 3 変数以上に拡張した検定手法ともいえる。また、名義尺度における繰り返しのない二元配置分散分析に相当する。

R:行 row, C:列 column

	A1	A2	...	Ai	Ak		
1	0or1	0or1		0or1	0or1	1行計 R ₁	R ₁ ²
2	0or1	0or1		0or1	0or1	2行計 R ₂	R ₂ ²
...	
j	0or1	0or1		0or1	0or1	j行計 R _j	R _j ²
n	0or1	0or1		0or1	0or1	n 行計 R _n	R _n ²
						Σ R _j	Σ R _j ²
	1列計 C ₁	2 列計 C ₂		i 列計 C _i	k 列計 C _k	Σ C _i	
	C ₁ ²	C ₂ ²		C _i ²	C _k ²	Σ C _i ²	

統計量

$$Q = \frac{(k - 1) * \{k * \sum C_i^2 - (\sum C_i)^2\}}{(k * \sum R_j - \sum R_j^2)}$$

Q は自由度 $k - 1$ のカイ二乗分布に従う。

$Q > \chi^2_{(k-1,0.05)}$ のとき有意差あり。

例題) 10 人の被験者から、5,000Hz、10,000Hz、20,000Hz の 3 つの周波数で聴覚検査を行った結果がある。この結果から、周波数によって聴こえた人の比率に差がないかを、コ克兰の Q 検定を用い検定する。有意水準は 5%に設定。

BellCurve エクセル統計より「コ克兰の Q 検定」

聴覚検査の結果				0:聴こえない
被験者	5000Hz	10000Hz	20000Hz	1:聴こえる

	5000Hz	10000Hz	20000Hz		R _j	R _j ²
	0	0	0		0	0
	0	0	0		0	0
	0	0	0		0	0
	1	0	0		1	1
	1	1	0		2	4
	1	1	0		2	4
	1	1	0		2	4
	1	1	1		3	9
	1	1	1		3	9
	1	1	1		3	9
					16	40
C _i	7	6	3	16		
C _i ²	49	36	9	94		

Q 値を計算し、エクセルで確率を求めた

$$Q = \frac{(3-1) \cdot \{3 \cdot 94 - 16^2\}}{(3 \cdot 16 - 40)} = \frac{52}{8} = 6.5$$

エクセル関数: カイ二乗分布右側確率: CHISQ.DIST.RT(6.5,2) = 0.038774208 有意差あり

BellCurve for Excel: ノンパラメトリック検定 → コ克兰の Q 検定

コ克兰のQ検定									
度数	5000Hz	10000Hz	20000Hz						
値: 0	3	4	7						
値: 1	7	6	3						
コ克兰のQ検定									
カイ二乗値	自由度	P 値	*: P<0.05 ** : P<0.01						
6.5000	2	0.0388	*						
度数の変化									
聴こえる人 (値が1の度数) が周波数が20,000Hzでは3人と全体の3割に下がっています。									
コ克兰のQ検定									
P値は0.0388でP<0.05となり、周波数により聴こえる人の比率は異なると言えます。									

共分散分析(ANCOVA)

ANCOVA: analysis of covariance

例えばやせている人ほど血圧が高いという結果が出た時、実は、やせている人ほどお年寄りが多かったなどということがある。この様なとき、年齢の影響を調整した共分散分析(重回帰分析の特殊型)により、調整平均(調整最小2乗平均)を計算すべき。

例題)

ある化合物の合成反応の量を4つの温度条件のもとで測定する。この合成反応は原材料に含まれる特定物質の含有率とも関連があると考えられるため、含有率を事前に測定し、共分散分析により実験結果を分析する。回帰直線の有意性の検定と共分散分析を行う。

BellCurve エクセル統計より「合成反応における化合物の収量」

温度	収量	含有率
60°C	64.8	12.5
60°C	58.1	6.6
60°C	64.8	11.8
60°C	61.5	10.1
60°C	58.3	5.5
65°C	62	9.9
65°C	62.3	9.6
65°C	61.3	9.7
65°C	62.2	8.2
65°C	64.2	10.2
70°C	64.3	11
70°C	65.2	11.5
70°C	64.9	10.4
70°C	62.9	9.3
75°C	60.3	7.7
75°C	60.3	7.8
75°C	60.9	7.3
75°C	61.2	8.1
75°C	64.1	11.7

のデータが得られた時

要因配置に基づくデータ表
データの型

	水準 A ₁ (データ数 n ₁) (変量、共変量)		水準 A ₂ (データ数 n ₂) (変量、共変量)		水準 A _a (データ数 n _a) (変量、共変量)	
	収量	含有率	収量	含有率	収量	含有率
	y ₁₁	x ₁₁	y ₂₁	x ₂₁	y _{a1}	x _{a1}
	y ₁₂	x ₁₂	y ₂₂	x ₂₂	y _{a2}	x _{a2}
	y _{1n₁}	x _{1n₁}	y _{2n₂}	x _{2n₂}	y _{ana}	x _{ana}
y 計						
x 計						

手順 1-1) 共分散分析用データ

	60°C n ₁ =5		65°C n ₂ =5		70°C n ₃ =4		75°C n ₄ =5		
	収量	含有率	収量	含有率	収量	含有率	収量	含有率	
	y ₁	x ₁	y ₂	x ₂	y ₃	x ₃	y ₄	x ₄	
	64.8	12.5	62	9.9	64.3	11	60.3	7.7	
	58.1	6.6	62.3	9.6	65.2	11.5	60.3	7.8	
	64.8	11.8	61.3	9.7	64.9	10.4	60.9	7.3	
	61.5	10.1	62.2	8.2	62.9	9.3	61.2	8.1	
	58.3	5.5	64.2	10.2			64.1	11.7	
y 計	307.5		312		257.3		306.8		1183.6
x 計		46.5		47.6		42.2		42.6	178.9

手順 1-2) y^2, x^2, xy を計算する。

	60°C			65°C			70°C			75°C			
	収量	含有率		収量	含有率		収量	含有率		収量	含有率		
	y ₁ ²	x ₁ ²	y ₁ x ₁	y ₂ ²	x ₂ ²	y ₂ x ₂	y ₃ ²	x ₃ ²	y ₃ x ₃	y ₄ ²	x ₄ ²	y ₄ x ₄	
	4199.04	156.25	810	3844	98.01	613.8	4134.49	121	707.3	3636.09	59.29	464.31	
	3375.61	43.56	383.46	3881.29	92.16	598.08	4251.04	132.25	749.8	3636.09	60.84	470.34	
	4199.04	139.24	764.64	3757.69	94.09	594.61	4212.01	108.16	674.96	3708.81	53.29	444.57	

	3782.25	102.01	621.15	3868.84	67.24	510.04	3956.41	86.49	584.97	3745.44	65.61	495.72	
	3398.89	30.25	320.65	4121.64	104.04	654.84				4108.81	136.89	749.97	
合計 y ²	18954.83			19473.46			16553.95			18835.24			73817.48
合計 x ²		471.31			455.54			447.9			375.92		1750.67
合計 yx			2899.9			2971.37			2717.03			2624.91	11213.21

手順 1-3) 補正項、全変動、水準間変動、水準内変動を計算する。

補正項	$N = n_1 + n_2 + n_3 + n_4$	$N = 19$
$CT_y = (y \text{ の合計})^2 / N$	$CT_y = (1183.6)^2 / 19 =$	73732.05053
$CT_x = (x \text{ の合計})^2 / N$	$CT_x = (178.9)^2 / 19 =$	1684.484737
$CT_{yx} = (y \text{ の合計})(x \text{ の合計}) / N$	$CT_{yx} = (1183.6)(178.9) / 19 =$	11144.52842

全変動

$ST_y = (y^2 \text{ の合計}) - CT_y$	$ST_y = (73817.48) - 73732.1 =$	85.42947368
$ST_x = (x^2 \text{ の合計}) - CT_x$	$ST_x = (1750.7) - 1684.5 =$	66.18526316
$ST_{yx} = (yx \text{ の合計}) - CT_{yx}$	$ST_{yx} = (11213.2) - 11144.5 =$	68.68157895

水準間変動

$SA_y = (y_1 \text{ の合計})^2 / n_1 + (y_2 \text{ の合計})^2 / n_2 + (y_3 \text{ の合計})^2 / n_3 + (y_4 \text{ の合計})^2 / n_4 - CT_y$	
$SA_y = (307.5)^2 / 5 + (312)^2 / 5 + (257.3)^2 / 4 + (306.8)^2 / 5 - 73732.1 =$	24.06997368
$SA_x = (x_1 \text{ の合計})^2 / n_1 + (x_2 \text{ の合計})^2 / n_2 + (x_3 \text{ の合計})^2 / n_3 + (x_4 \text{ の合計})^2 / n_4 - CT_x$	
$SA_x = (46.5)^2 / 5 + (47.6)^2 / 5 + (42.2)^2 / 4 + (42.6)^2 / 5 - 1684.5 =$	9.279263158
$SA_{yx} = (y_1 \text{ の合計})(x_1 \text{ の合計}) / n_1 + (y_2 \text{ の合計})(x_2 \text{ の合計}) / n_2 + (y_3 \text{ の合計})(x_3 \text{ の合計}) / n_3$	
$+ (y_4 \text{ の合計})(x_4 \text{ の合計}) / n_4 - CT_{yx}$	
$SA_{yx} = (307.5)(46.5) / 5 + (312)(47.6) / 5 + (257.3)(42.2) / 4$	
$+ (306.8)(42.6) / 5 - 11144.5 =$	13.91257895

水準内変動

$SE_y = ST_y - SA_y$	$SE_y = 85.42947 - 24.1 =$	61.3595
$SE_x = ST_x - SA_x$	$SE_x = 66.2 - 9.3 =$	56.906
$SE_{yx} = ST_{yx} - SA_{yx}$	$SE_{yx} = 68.7 - 13.9 =$	54.769

手順 2) 回帰直線の平行性の検定

解析ソフトでは「温度 * 含有率」の交互作用が有意でなければ温度ごとの回帰直線は平行と仮定 ⇒ 下記は回帰の変動からの計算

2-1) ここで改めて、それぞれの水準の単回帰分析を行なう

回帰平方和と誤差変動が計算される各 $SR = (S_{xy})^2 / S_{xx}$

	60°C n ₁ =5		65°C n ₂ =5		70°C n ₃ =4		75°C n ₄ =5		
	収量	含有率	収量	含有率	収量	含有率	収量	含有率	
	y ₁	x ₁	y ₂	x ₂	y ₃	x ₃	y ₄	x ₄	
	64.8	12.5	62	9.9	64.3	11	60.3	7.7	
	58.1	6.6	62.3	9.6	65.2	11.5	60.3	7.8	
	64.8	11.8	61.3	9.7	64.9	10.4	60.9	7.3	
	61.5	10.1	62.2	8.2	62.9	9.3	61.2	8.1	
	58.3	5.5	64.2	10.2			64.1	11.7	
y 計	307.5		312		257.3		306.8		1183.6
x 計		46.5		47.6		42.2		42.6	178.9

回帰平方和 SR1 SR2 SR3 SR4 各 SR を単純に合計: SBx

41.5 0.53 2.4 9.3 53.73

誤差変動 2.097177046 4.125284757 0.776115242 0.70539636 7.7040

ここで仮説 H₀: すべての傾きが等しい(平行) H₁: 一つでも平行ではない

参考: 単回帰分析の項目から)

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \Rightarrow r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \Rightarrow r^2 = \frac{S_{xy}}{S_{xx}} * \frac{S_{xy}}{S_{yy}}$$

回帰直線は

$$y = \frac{S_{xy}}{S_{xx}}x + \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x} \Rightarrow y - \bar{y} = \frac{S_{xy}}{S_{xx}}(x - \bar{x}) \text{ となる。}$$

実際のデータ(x_i, y_i)、2変量の平均値を (x̄, ȳ)、回帰直線上のデータ(x_i, ŷ_i)

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SST = SSE + SSR$$

全変動

$$SST(\text{total sum of squares, } S_T) = S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

残差変動(残差平方和): ŷ_iで説明できなかった変動

$$SSE(\text{error sum of squares, } S_e) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

回帰変動: ŷ_iで説明された変動

$$SSR(\text{regression sum of squares, } S_R) = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$R^2 = \frac{SSR}{SST} \quad R^2 \text{を決定係数(寄与率)、} R \text{を重相関係数という。}$$

決定係数は回帰モデルによって実データをどれくらい説明できているか(回帰分析の精度)を表す指標。相関係数 r と決定係数 R^2 は全く別の文脈で定義された量だが、最小二乗法による推定の場合、相関係数の二乗と決定係数は一致する。

$$\text{ここで } r^2 = \frac{s_{xy}}{s_{xx}} * \frac{s_{xy}}{s_{yy}} \text{ より } R^2 = \frac{s_{xy}}{s_{xx}} * \frac{s_{xy}}{s_{yy}} \text{ であり } R^2 = \frac{SSR}{SST} \text{、} SST = s_{yy} \text{ より}$$

$$SSR = SR = \frac{s_{xy}}{s_{xx}} * \frac{s_{xy}}{s_{yy}} * s_{yy} = \frac{s_{xy}^2}{s_{xx}}$$

2-2) 非平行性の変動の計算

共分散分析での変動:

群差 A + 回帰の総変動 B (= 共通回帰 B1 + 非平行性 B2) + 残差の合計 C

非平行性の変動は、交互作用がないときの変動を意味する

SBx: 各回帰平方和の合計 (⇒ B に相当)

(SEyx)²/SEx: 全体の水準内変動からの計算値

⇒ 共通回帰での回帰平方和 (⇒ B1 に相当)

SNp: 共通回帰で説明がつかない回帰変動 (⇒ B2 に相当)

SNp = SBx - (SEyx)²/SEx

$$= 1.018$$

2-3) 残差変動の計算

$$SE' = SEy - \sum_{r=1}^4 SR = SEy - SBx = (STy - SAy) - SBx$$

SE': 残差変動、平方和 (⇒ C に相当)

	平方和	自由度	平均平方	F
非平行性の変動	SNp	水準の数-1		
		(a-1)	VNp = SNp / (a-1)	Vnp / VE'
	1.018	4-1=3	0.339333	0.489241
残差変動	SE'	(データ数) - 2 (水準の数)	VE' = SE' / (N-2a)	(平行性は否定されない)
		(N-2a)		
	7.6295	19-2*4=11	0.693591	

$F_{11}^3 = 3.59 (F \text{分布表}) > 0.489$ なので仮説 H_0 は否定されない
(平行性は否定されない)

3) 回帰の有意性の検定

回帰直線の傾きが 0 でないこと

H_0 : 傾きが 0 H_1 : 傾きが 0 ではない

共通の回帰係数 $\beta = SE_{yx}/SE_x$					
	0.962447				
であるが					
回帰の有意性の検定	を行なう				
		平方和	自由度	平均平方	F1,N-a-1
回帰による変動		$(SE_{yx})^2/SE_x$	1	$VR=(SE_{yx})^2/SE_x$	VR/VE
		52.71225	1	52.71225	
水準内変動		$SE_y - (SE_{yx})^2/SE_x$	N-a-1	$VE = \{SE_y - (SE_{yx})^2/SE_x\} / (N-a-1)$	
		8.647249	19-4-1=14	0.617661	85.34177
					p<0.001**
					(回帰直線の傾きは0ではない)

よって、回帰係数の傾きは 0 ではない

4) 水準間の差の検定 (1 元配置の共分散分析)

$$S_A = (ST_y - (ST_{yx})^2/ST_x)^* - (SE_y - (SE_{yx})^2/SE_x)^{**}$$

	平方和	自由度	平均平方	
水準間変動	SA	a-1	VA=SA/a-1	
	5.510176	3	1.836725	
水準内変動	$SE_y - (SE_{yx})^2/SE_x$	N-a-1	$VE = \{SE_y - (SE_{yx})^2/SE_x\} / (N-a-1)$	
	8.647249	14	0.617661	$F = 1.837/0.618 = 2.97$
			有意差なし	$< F_{14}^3 = 3.34 (F$ 分布表)

2) 3) より回帰係数の傾きが 0 でなく、各水準の傾きが平行であるから、水準間の差は共通の傾きで補正された時の y 切片の差に相当する。(すべてのデータの変動から回帰による変動を差し引いた)*から(水準内変動から水準内変動での回帰変動を差し引いた)**値を引いた SA は補正された水準間変動に相当する。

$$F.DIST.RT(2.97, 3, 14) = 0.068037661$$

結論: 含有率には違いがあるが、温度(主効果)有意差なし

↑上記は石村貞夫先生の「分散分析のはなし」共分散分析を参考に手計算した結果。

↓下記は BellCurve for Excel での結果。

BellCurve for Excel: 分散分析多重比較→共分散分析

共分散分析							
要因配置に基づくデータ表							
60°C		65°C		70°C		75°C	
収量	含有率	収量	含有率	収量	含有率	収量	含有率
64.8	12.5	62	9.9	64.3	11	60.3	7.7
58.1	6.6	62.3	9.6	65.2	11.5	60.3	7.8
64.8	11.8	61.3	9.7	64.9	10.4	60.9	7.3
61.5	10.1	62.2	8.2	62.9	9.3	61.2	8.1
58.3	5.5	64.2	10.2			64.1	11.7
データ形式は下記にする							
合成反応における化合物の収量							
温度	収量	含有率	目的変数：収量				
60°C	64.8	12.5	固定因子：温度				
60°C	58.1	6.6	共変量：含有率				
60°C	64.8	11.8	モデル：				
60°C	61.5	10.1	typeIII平方和				
60°C	58.3	5.5	すべての因子（温度、含有率）指定				
65°C	62	9.9	指定する因子（温度、含有率指定）⇒交互作用（モデルに追加）				
65°C	62.3	9.6					
65°C	61.3	9.7					
65°C	62.2	8.2					
65°C	64.2	10.2					
70°C	64.3	11					
70°C	65.2	11.5					
70°C	64.9	10.4					
70°C	62.9	9.3					
75°C	60.3	7.7					
75°C	60.3	7.8					
75°C	60.9	7.3					
75°C	61.2	8.1					
75°C	64.1	11.7					

共分散分析										
基本統計量										
目的変数	モデル	温度	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
収量	温度	60°C	5	61.500	3.301	58.199	64.801	1.476	60.024	62.976
		65°C	5	62.400	1.079	61.321	63.479	0.483	61.917	62.883
		70°C	4	64.325	1.021	63.304	65.346	0.511	63.814	64.836
		75°C	5	61.360	1.581	59.779	62.941	0.707	60.653	62.067
共変量										
目的変数	モデル	温度	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
含有率	温度	60°C	5	9.300	3.117	6.183	12.417	1.394	7.906	10.694
		65°C	5	9.520	0.773	8.747	10.293	0.346	9.174	9.866
		70°C	4	10.550	0.947	9.603	11.497	0.473	10.077	11.023
		75°C	5	8.520	1.801	6.719	10.321	0.805	7.715	9.325
等分散性の検定										
ルビーン検定										
目的変数	F 値	自由度1	自由度2	P 値						
収量	2.9111	3	15	0.0689						
回帰の有意性の検定										
	平方和	自由度	平均平方	F 値	P 値	*: P<0.05 **: P<0.01				
回帰による	52.7123	1	52.7123	85.3418	p < 0.001**					
回帰からの(水準内変動)	8.6472	14	0.6177							
分散分析表										
因子	TypeIII 平方和	自由度	平均平方	F 値	P 値	*: P<0.05 **: P<0.01				
温度	5.5102	3	1.8367	2.9737	0.0678					
含有率	52.7123	1	52.7123	85.3418	p < 0.001**					
誤差 (水準内変動)	8.6472	14	0.6177							
全体	85.4295	18								

含有率には
違いがある
が、温度(主
効果)有意
差なし

最小二乗平均 LSMEAN										
目的変数	モデル	温度	n	最小二乗平均	標準誤差	95%信頼区間				
収量	温度	60°C	5	61.611	0.352	60.857	62.366			
		65°C	5	62.300	0.352	61.546	63.054			
		70°C	4	63.233	0.410	62.353	64.113			
		75°C	5	62.222	0.364	61.442	63.002			
多重比較検定										
因子	目的変数	手法	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値
温度	収量	Bonferroni	60°C	65°C	61.6114	62.2997	0.6883	0.4976	1.3832	1.0000
			60°C	70°C	61.6114	63.2334	1.6219	0.5431	2.9867	0.0588
			60°C	75°C	61.6114	62.2221	0.6107	0.5037	1.2126	1.0000
			65°C	70°C	62.2997	63.2334	0.9337	0.5380	1.7354	0.6277
			65°C	75°C	62.2997	62.2221	0.0776	0.5079	0.1527	1.0000
			70°C	75°C	63.2334	62.2221	1.0112	0.5680	1.7802	0.5805
		Tukey-Kramer	60°C	65°C	61.6114	62.2997	0.6883	0.4976	1.3832	0.5292
			60°C	70°C	61.6114	63.2334	1.6219	0.5431	2.9867	0.0431*
			60°C	75°C	61.6114	62.2221	0.6107	0.5037	1.2126	0.6294
			65°C	70°C	62.2997	63.2334	0.9337	0.5380	1.7354	0.3431
			65°C	75°C	62.2997	62.2221	0.0776	0.5079	0.1527	0.9987
			70°C	75°C	63.2334	62.2221	1.0112	0.5680	1.7802	0.3226

経時的測定データ (longitudinal data)

同じ対象から繰り返しデータを測定する反復測定データ repeated measurement data のうち、経時的に測定 of 順序を変更できないものを経時的測定データ longitudinal data という。経時的な場合、年齢などが変化するため、処置を行う前の値(baseline 値) が変化し単純な処理では系統的な誤差が生じる可能性がある。適切な統計解析法が明確にされていないために幾つかの方法が提起されている。

- 各時点における群間比較

二群の比較であれば、Welch's t-test もしくは Mann-Whitney U test、三群以上であれば一元配置分散分析(One-way ANOVA) か Kruskal-Wallis test を用いる。baseline からの差、比以外に検定の多重性の問題が生じるため Bonferroni の多重比較(k 回の検定を同時に行うとき、同時有意水準を α にするためには、個々の検定の有意水準を α/k とする。)を使用するが、時点ごとの群間比較は経時測定データの特徴を生かしていないため、baseline において高い値をとるサンプルはそれ以降の時点においても高い値をとる傾向があるなどの情報を生かしていない点で時点ごとの群間比較は「切れ味の悪い」検定になる。

- 反復測定分散分析(Repeated measures ANOVA)

反復測定分散分析法は対応のある t 検定を 3 群以上に拡張した方法と考えることができる。

- 投与前値を共変量とみなした解析(共分散分析)
- 多変量分散分析(MANOVA)なども使用される。
- 線形混合効果モデルは専門家の間で推奨されている。

反復測定分散分析

一元配置分散分析(対応あり)

=反復測定一元配置分散分析=二元配置分散分析(繰り返しなし)

例題)5 頭のブタすべてに 3 種の成長ホルモン刺激剤を与え、各刺激剤ごとに体重増加率を調べる。成長ホルモン刺激剤によって体重の増加率が異なるか、有意水準 5% で分散分析を行う。また、どのホルモン剤との間に有意差があるかについて多重比較も行って確認する。(再掲) →P81

ブタ	A1	A2	A3
1	9.48	9.24	8.66
2	9.52	9.95	8.50
3	9.32	9.20	8.76
4	9.98	9.68	9.11
5	10.00	9.94	9.75

BellCurve エクセル統計より「成長ホルモン刺激剤の用量によるブタの体重の違い」

対応のある二元配置分散分析(対応あり)

例題) 2種類の薬剤の効果をおある検査値でみる実験(再掲) → P83

個体 No.	投与薬剤	スタート	1 週	2 週	3 週
1	A	7.5	8.6	6.9	0.8
2	A	10.6	11.7	8.8	1.6
3	A	12.4	13	11	5.6
4	A	11.5	12.6	11.1	7.5
5	A	8.3	8.9	6.8	0.5
6	A	9.2	10.1	8.6	3.8
7	B	13.3	13.3	12.9	11.1
8	B	10.7	10.8	10.7	9.3
9	B	12.5	12.7	12	10.1
10	B	8.4	8.7	8.1	5.7
11	B	9.4	9.6	8	3.8
12	B	11.3	11.7	10	8.5

フリードマン検定 (Friedman test) ノンパラメトリック

1) 一元配置分散分析(対応あり)反復測定に相当するノンパラメトリック検定。

典型的な例は医療データにおける「投与前」→「1分後」→「10分後」→「20分後」といった経時的な変化を調べる場合。

フリードマン検定は、量的データをランクデータ(順序データ)に変換して計算する。したがって、フリードマン検定を利用した場合の量的変数の代表値は「平均値」ではなく「中央値」(厳密には平均ランク)になる。

帰無仮説(H₀):水準間の平均ランクに差はない

(「投与前」=「1分後」=「10分後」=「20分後」)

対立仮説(H₁):水準間の平均ランクに差がある

(「投与前」≠「1分後」≠「10分後」≠「20分後」)

帰無仮説が5%水準で棄却されれば、「水準間の平均ランクに差がある(=有意差あり)」と判定される。フリードマン検定で有意差を認めた場合に、多重比較を行うことになる(v⇒ P108)。

2) 繰り返しのない二元配置分散分析に相当するノンパラメトリック検定でもある。

Bartlett 検定で「各群の分散が等しくない」と判定されたとき Friedman 検定を用いる。Friedman 検定では順位の付け方が2通りある。

帰無仮説(H₀):「各群で差がない」、対立仮説(H₁):「各群で差がある」

行(要因B)を検定する場合

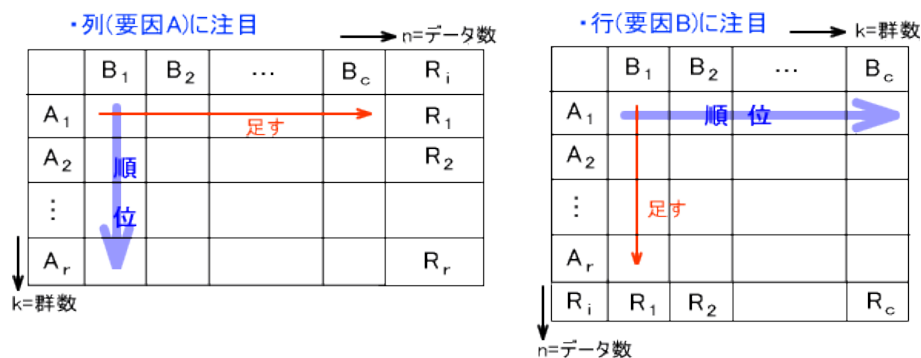
列(要因A)の各群ごとに順位をつける。その後、順位を足して R_iを求める。

(行数を k,列数を nとする)

列(要因A)を検定する場合

行(要因B)の各群ごとに順位をつける。その後、順位を足して R_iを求める。

(行数を n,列数を kとする)



R_i を求め、下の式によって統計量 χ^2_r を計算する。

$$\chi^2_r = \frac{12}{nk(k+1)} \sum_{i=1}^k R_i^2 - 3n(k+1)$$

確率を求めるとき、群数 k によって求め方が異なる。

- $k \leq 4$ のとき → Friedman 検定表から求める
 $P \geq \alpha$ のとき帰無仮説を棄却できない。
 $P < \alpha$ のとき帰無仮説を棄却する。
- $k > 4$ のとき → χ^2_r は近似的に自由度 $df=k-1$ の χ^2 分布に従う
 $\chi^2 \leq \chi^2_\alpha$ のとき、 $P \geq \alpha$ となり帰無仮説を棄却できない
 $\chi^2 > \chi^2_\alpha$ のとき、 $P < \alpha$ となり帰無仮説を棄却する

Friedman 検定表

3群(k=3)			4群(k=4)		
ブロック数(n)	P<0.05	P<0.01	ブロック数(n)	P<0.05	
3	6	-	2	6	
4	6.5	8	3	7.4	9
5	6.4	8.4	4	7.8	9.6
6	7	9	5	7.8	9.96
7	7.14	8.86	6	7.6	10.2
8	6.25	9	7	7.8	10.4
9	6.22	9.56	8	7.65	10.4
10	6.2	9.6	9	7.8	10.87
11	6.55	9.46	10	7.8	10.8
12	6.17	9.5	11	7.91	11.07
13	6	9.39	12	7.9	11.1
14	6.14	9	13	7.99	11.1
15	6.4	8.93	14	7.89	11.1
			15	8.04	11.2

例題)

マウスに薬剤 X を投与し、血糖値を観察した。投与前後で血糖値が変化したといえるかどうか検定する。ただし、「各時間での分散は均一でない」とする。

	投与前	投与直前	1 時間後	2 時間後	3 時間後
a	162	60	88	143	160
b	160	50	112	136	156
c	150	100	135	149	152
d	155	92	85	115	148

帰無仮説(H₀):投与前後で血糖値に変化がない。

対立仮説(H₁):投与前後で血糖値に変化がある。

投与前後での血糖値の変化を見るので行(時間)に注目する。まず、列(マウス)ごとに数値の小さい順に順位をつける。そして、順位を群ごとに足して R_iを求める。

	投与前	投与直前	1時間後	2時間後	3時間後
a	5	1	2	3	4
b	5	1	2	3	4
c	4	1	2	3	5
d	5	2	1	3	4
R _i	19	5	7	12	17

(行に注目するので、n=4,k=5となる)

R_iを求めたら、下の式によって統計量 χ^2_r を求める。

$$\chi^2_r = \frac{12}{nk(k+1)} \sum_{i=1}^k R_i^2 - 3n(k+1)$$

$$= \frac{12}{4 \cdot 5(5+1)} (19^2 + 5^2 + 7^2 + 12^2 + 17^2) - 3 \cdot 4(5+1) = 14.8$$

k>4なので、計算した χ^2_r 値は近似的に自由度 df=k-1=5-1=4 の χ^2 分布に従う。 χ^2 分布表より、 $\chi^2_{0.05}=9.488$ である。

$\chi^2_r=14.8 > 9.488=\chi^2_{0.05}$ なので、P<0.05 となり帰無仮説を棄却できる。つまり、「投与前後で血糖値に変化がある」ということができる。

BellCurve for Excel: [ノンパラメトリック検定→フリードマン検定](#)

フリードマン検定						
順位変換前の表						
		投与前	投与直前	1時間後	2時間後	3時間後
	a	162	60	88	143	160
	b	160	50	112	136	156
	c	150	100	135	149	152
	d	155	92	85	115	148
順位変換後の表						
		投与前	投与直前	1時間後	2時間後	3時間後
	a	5	1	2	3	4
	b	5	1	2	3	4
	c	4	1	2	3	5
	d	5	2	1	3	4
パソコンでの検定では変換前後のどちらでも結果は同じ						

フリードマン検定						
ケースの要約						
	n	%				
有効ケース	4	100.0%				
不明ケース	0	0.0%				
全 体	4	100.0%				
	投与前	投与直前	1時間後	2時間後	3時間後	
平均順位	4.7500	1.2500	1.7500	3.0000	4.2500	
フリードマン検定						
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01		← 有意差あり	
14.8000	4	0.0051	**			
ケンドールの一致係数						
0.9250						
Scheffeの対比較					* : P<0.05 ** : P<0.01	
水準1	水準2	平均順位1	平均順位2	カイ二乗値	自由度	P 値
投与前	投与直前	4.7500	1.2500	9.8000	4	0.0439 *
投与前	1時間後	4.7500	1.7500	7.2000	4	0.1257
投与前	2時間後	4.7500	3.0000	2.4500	4	0.6536
投与前	3時間後	4.7500	4.2500	0.2000	4	0.9953
投与直前	1時間後	1.2500	1.7500	0.2000	4	0.9953
投与直前	2時間後	1.2500	3.0000	2.4500	4	0.6536
投与直前	3時間後	1.2500	4.2500	7.2000	4	0.1257
1時間後	2時間後	1.7500	3.0000	1.2500	4	0.8698
1時間後	3時間後	1.7500	4.2500	5.0000	4	0.2873
2時間後	3時間後	3.0000	4.2500	1.2500	4	0.8698

多重比較(for non-parametric data)→P109

用量-反応関係の検出

用量、程度が増加するに従ってある反応が増加していく(直線的、非直線的かは別に
して)ときの検定

反応が計量的なときは

- 1) 回帰分析
- 2) Jonckheere の順位和検定

反応が計数的なときは

- 1) Mantel-extension 法
- 2) Cochran-Armitage 検定などが行われる

ヨンクヒール・タプストラの検定 (Terpstra-) Jonckheere test

参照) kasuya.ecology1.org/stats/jonck01.html

Jonckheere の検定は、順序制約のある仮説用のノンパラメトリックな検定で、3 つ以上の処理があつて、目的変数がだんだん大きくなるという傾向的变化があるという内容を検定したいときに使用。(次第に小さくなる場合には、処理に付ける番号を反対にする。)目的変数が、第 1 の処理 < 第 2 の処理 < 第 3 の処理 < ... (以下略) といった傾向的变化をするかどうかを問題にする。

Jonckheere の検定の検定統計量は、異なる 2 つの処理間での、目的変数の大小が対立仮説にあっているデータの組み合わせの個数。

たとえば、3 つの処理があつて、それぞれ目的変数の値が以下のようなとする。処理の番号が大きいほうが目的変数の値が大きいという傾向的变化を対立仮説とする。

3 処理の場合の組み合わせは、処理 1vs 処理 2、処理 2vs 処理 3、処理 1vs 処理 3 の 3 つ。一般にそれぞれの組み合わせで、2 つの異なる処理のデータの大小を比べると、処理のサンプル数を n_i と n_j とすると、比較の個数は $n_i \times n_j$ である。

第 1 の処理 5.4 6.7

第 2 の処理 7.9 6.5 6.9

第 3 の処理 8.6 7.8 8.2

上の例では、

処理 1vs 処理 2 サンプル数が 2 と 3 なので、6 つの比較がある。

処理 2vs 処理 3 サンプル数が 3 と 3 なので、9 つの比較がある。

処理 1vs 処理 3 サンプル数が 2 と 3 なので、6 つの比較がある。

処理番号の大きい方から数える、

処理 1vs 処理 2 第 2 の処理が大きいものは (6.5vs6.7 以外) 5 つなので、5

処理 2vs 処理 3 第 3 の処理が大きいものは 6.5vs6.7 以外のすべてなので、8

処理 1vs 処理 3 第 3 の処理は 6 つすべてで大きいので、6
 検定統計量は 3 つを合計して $5+8+6=19$ となる。

もう1つ別の例、やはり3処理で、処理の番号が大きいほうが目的変数の値が大きい
 という傾向的变化を対立仮説とする。同一値のときは 0.5 とする。

第 1 の処理 3 8 5 7 7
 第 2 の処理 7 5 8 8 5
 第 3 の処理 8 1 0 9 1 1 9

処理 1vs 処理 2 第 2 の処理の各データからみて、 $3+1.5+4.5+4.5+1.5=15$
 処理 2vs 処理 3 第 3 の処理の各データからみて、 $4+5+5+5+5=24$
 処理 1vs 処理 3 第 3 の処理の各データからみて、 $4.5+5+5+5+5=24.5$
 検定統計量は合計して $15+24+24.5=63.5$ となる。

データを 2 つずつ比べて、その大小の個数を数えて統計量を構成するところは、U
 検定や Kendall の順位相関係数 τ とよく似ている。Jonckheere の検定の検定統計
 量の帰無仮説のもとでの期待値は、各処理のサンプル数を n_i 、全体でのサンプル数
 つまり $\sum n_i$ を N として、 $(N^2 - \sum (n_i)^2) / 4$ である。

例題) 薬物濃度を 10ppm から 10000ppm まで上がるにつれて、ラットの赤血球数が
 徐々に下がっていく傾向にあるのか、Jonckheere=タプストラ検定により検定。有意水
 準は 5%。

異なる薬物濃度に対するラットの赤血球数			
10ppm	100ppm	1000ppm	10000ppm
8.06	7.97	7.66	8
8.27	7.66	7.71	7.89
8.45	8.05	7.88	7.79
8.51	8.3	8.05	7.91
8.14	8.03	7.8	7.4

BellCurve for Excel: ノンパラメトリック検定→ヨンクヒール=タブストラ検定

ヨンクヒール=タブストラ検定				
異なる薬物濃度に対するラットの赤血球数				
10ppm	100ppm	1000ppm	10000ppm	
8.06	7.97	7.66	8	
8.27	7.66	7.71	7.89	
8.45	8.05	7.88	7.79	
8.51	8.3	8.05	7.91	
8.14	8.03	7.8	7.4	
ヨンクヒール=タブストラ検定				
サンプルサイズ				
水準	10ppm	100ppm	1000ppm	10000ppm
n	5	5	5	5
増加傾向 (列の水準のデータが行の水準のデータよりも大きい組合せ)				
	100ppm	1000ppm	10000ppm	
10ppm	3	0	0	
100ppm		6	5	
1000ppm			14	
総数	28			
減少傾向 (列の水準のデータが行の水準のデータよりも小さい組合せ)				
	100ppm	1000ppm	10000ppm	
10ppm	22	25	25	
100ppm		17	20	
1000ppm			11	
総数	120			
ヨンクヒール=タブストラ検定				
対立仮説	組合せ	統計量:z	P 値	* : P<0.05 ** : P<0.01
増加傾向	28	-3.1628	0.9992	
減少傾向	120	3.0282	0.0012	**
減少傾向				
各濃度とも5匹のラットを測定しているので、2つの濃度間では25の組み合わせが生じる。				
1000ppmと10000ppmの組み合わせを除き、減少する組み合わせが多数を占める。				
ヨンクヒール=タブストラ検定				
減少傾向のP値は0.0012でP<0.05となり、薬物濃度が上がるにつれ、赤血球数は単調に減少していると言える。				

← 減少傾向
に有意差あり

コクラン・アーミテージ検定 Cochran-Armitage test

薬剤の投与量など順序性が仮定できるカテゴリ間で、2値の応答割合の傾向をみる。

統計量

$$Z = \chi_{CA} = \frac{\sum_{j=1}^r X_j d_j - p \sum_{j=1}^r n_j d_j}{\sqrt{p(1-p) \left\{ \sum_{j=1}^r n_j d_j^2 - \frac{1}{N} (\sum_{j=1}^r n_j d_j)^2 \right\}}}$$

$$N = \sum_{j=1}^r n_j$$

$$p = \frac{\sum_{j=1}^r X_j}{N} \quad (\text{片側}), \text{両側} 2p$$

表を見ながらデータを
当てはめれば意外と
簡単に計算できる

例)ある薬剤について各群30例からなる4用量群の結果、用量反応関係はあるか。

有意水準5%で両側検定せよ。

群	用量(mg/kg/day)	効果あり	効果なし	計
G1	0	9	21	30
G2	10	10	20	30
G3	20	8	22	30
G4	30	15	15	30

群	n_i	X_i	d_i	$X_i d_i$	$n_i d_i$	$n_i d_i^2$
G1	30	9	0	-	-	-
G2	30	10	1	10	30	30
G3	30	8	2	16	60	120
G4	30	15	3	45	90	270
Σ	120	42		71	180	420

d_i は用量の順序付け (0,10,20,30でもよいが0,1,2,3としてもよい)

$$p = \frac{42}{120} = 0.35$$

$$\sum_{j=1}^r X_j d_j = 71$$

$$\sum_{j=1}^r n_j d_j = 180$$

$$\sum_{j=1}^r n_j d_j^2 = 420$$

$$Z = \chi_{CA} = \frac{71 - 0.35 * 180}{\sqrt{0.35(1 - 0.35) \left\{ 420 - \frac{1}{120} * 180^2 \right\}}} = 1.37$$

エクセル関数NORM.S.DIST(1.37,TRUE)= 0.914574	P=0.085	両側 p 値 =0.170
---	---------	------------------

有意な用量反応関係なし

BellCurve for Excel: 集計表の作成と分析→コクラン=アーミテージ検定

群	用量(mg/kg/day)	効果あり	効果なし	計
G1	0	9	21	30
G2	10	10	20	30
G3	20	8	22	30
G4	30	15	15	30

コクラン=アーミテージ検定			
観測度数			
用量(mg/kg/day)	効果あり	効果なし	
0	9	21	
10	10	20	
20	8	22	
30	15	15	
コクラン=アーミテージ検定 * : P<0.05 ** : P<0.01			
統計量:z	1.3695		
両側P値	0.1709		
片側P値	0.0854		
Cramer's V	0.1882		

検定の結果は有意でなく、投与量が増えても割合に傾向がないという帰無仮説は棄却されない。(有意な用量反応関係なし)

多重比較Multiple comparison

多重比較の2通りの立場

事前比較 (A priori comparisons):

結果を分析する前に、理論的な背景などにより、比較する平均値についての仮説がある場合は、ANOVA 抜きで、多重比較検定を行う。

事後比較 (Post hoc comparisons):

比較する平均値についての明確な仮説がない場合は、ANOVA で比較対象を決めて、多重比較を行う。

どの群間に差があるのか(2群間のt検定の繰り返しではだめ！)

個々の仮定の有意水準 α_i 、最終的な有意水準 α

$$1 - \alpha = \prod_{i=1}^h (1 - \alpha_i) \quad \alpha_i \text{が一定なら}\Pi \text{は}(1 - \alpha_i) \text{を}h \text{回かけることだから}$$

$$= (1 - \alpha_i)^h \rightarrow \alpha = 1 - (1 - \alpha_i)^h \quad \alpha_i \text{が小さい時、}\alpha \approx 1 - (1 - h\alpha_i) = h\alpha_i$$

h回検定するとき各検定の水準を α/h とすれば全体の有意水準を α に保てる。

例えば有意水準5%のt検定を4回繰り返しすべての2群でt検定を行なうと結果として20%で検定した場合での有意差となる。有意水準を $5/4=1.25\%$ で検定する。(Bonferroni(ボンフェローニ)法)

多重比較法の種類(for parametric data)

対比較とは、「全群総当りの検定」

対照との比較とは「例えばプラセボとの比較」

対比とは「いくつかの投与群をグループ化して比較」、傾向性を知りたいという場合

	比較	ANOVA	各群の n	正規分布	分散
Dunnet法	対照との比較 (同時)	不要	制限なし	正規分布	等分散
	Tukey-kramer法などと比べて、有意差が出やすい。 独立多群間で比較する。(関連多群のデータには適応困難)				
Williams法	対照との比較	不要	一致	正規分布	等分散
	母平均に単調増加、または単調減少があるとき使用。Dunnet法より検出力がよくなる。				
Tukey-Kramer法	対比較(同時)	必要	制限なし	正規分布	等分散
Games-Howell の方法	Tukey法は、各群のデータ数(n)が一致する必要があるが、Tukey-Kramer法は一致しなくても良い。有意差の検出力に優れる。				

Bonferoni法 (=Dunnの方法) Dunn-Sidak 修正 t 検定	対比較(非同時)	不要	制限なし	制限なし	制限なし
	多重比較検定としてほとんどの場合に使用できる。検定全体の有意水準を検定数で割った値を有意水準とする検定法。群が多い場合などは、有意差が出にくい。5群以上では、使用しない方がよい。				
Scheffe法	同時対比	必要	制限なし	正規分布	等分散
	n個の各群におけるすべての対比の中で、有意なものをさがす検定。適用範囲は広く頑強があるが、有意差は出にくい。比較したい興味のある群が限定されている場合は、Dunnett法やTukey-Kramer法などを使う方がよい。				

多重比較法の種類(for non-parametric data)

- 1) 各群の観測値は形状が同一の分布からの無作為標本
- 2) 各標本は互いに独立か

どのタイプの多重比較か選択重要

	比較	ANOVA	各群の n	正規分布	分散
Steel法	対照との比較 (同時)	不要	制限なし	制限なし	制限なし
	対照群と各群を比較する順位を用いた多重比較検定。Dunnett法のノンパラ版。				
Shirley-Williams法	対照との比較	不要	一致	制限なし	制限なし
	単調増加、または単調減少があるとき使用。Williams法のノンパラ版。				
Steel-Dwass法	対比較(同時)	不要	制限なし	制限なし	制限なし
	すべての2群同士を同時に比較する順位を用いた多重比較検定。検出力が高く、有意差が出やすい。Tukey-Kramer法のノンパラ版。				
Bonferoni法	検定統計量に制限が無し				
	有意水準を補正する多重検定。ほとんどの場合に使用できノンパラメトリックな手法としても使用できる。				
Scheffe法	ノンパラメトリック検定のクラスカルウォリス検定、フリードマン検定でも利用できる。				

- Fisher's PLSD法、Student-Newman-Keuls法第1種の過誤を起こす可能性があるのであまり使わない方がよいといわれる。
- Bonferoni法は簡便だが検定数が増えると有意差が出にくい。
- パラメトリックな場合、すべての対比較をする Tukey法は検出力がよい。
- コントロール群と実験群との比較には Dunnett法が有意差が出やすい。

対応のある分散分析での注意点

難しいのでよくわからないが要は以下の通り。

参考) 井関龍太のページ <http://riseki.php.xdomain.jp/index.php?FrontPage>

Mauchly(モークリー)の球面性検定

対応のある分散分析の場合、被験者内因子の分散が等しいかどうかを確認するため Mauchly の球面性検定を行う(球面性の仮定が棄却されないことを示したい)。球面性の仮定(sphericity assumption)とは、すべての被験者内因子間の差の分散が等しいことを仮定するもので、分散分析を適切に行うための必要十分条件となる。データのばらつき方が条件間で等しければ“球”に近づくというイメージで理解する。

Mauchly の検定において $p \geq 0.05$ 以上ならば、球面性の仮定は棄却されず、 $p < 0.05$ の場合、球面性は仮定できない。球面性の仮定が成り立たない場合は、被験者内因子の検定の有意確率が小さくなるので ϵ (イプシロン)という値によって自由度の調整を行った分散分析をする。 ϵ は 0~1(1 を越える場合は 1 と見なす)を取る値で、0 に近いほど球面性の仮定から離れていることを示している。 ϵ を F 検定の際の自由度にかけることによってデータを F 分布に近似させ P 値を求めるが、0 に近いほど F 検定に対して厳しい制限をかけることになる。 ϵ の算出には、「Greenhouse-Geisser」「Huynh-Feldt-Lecoutre」「Chi-Muller」「下限値」の 4 つの手法があり、Greenhouse-Geisser は、最も一般的に使用される調整値で、Greenhouse-Geisser の ϵ が 0.75 に近い、あるいは大きい場合 Huynh-Feldt 修正を行う。下限値は、 ϵ の取りうる値の理論的下限値で上記手法では修正を行う事ができない場合、 ϵ の下限修正を使用する。下限値は、PC がまだそれほど普及していなかった時代、 ϵ を計算する手間を省くために使われ、最も厳しい調整を行う下限値を使ってもまだ有意なら確実に有意と判断する。

BellCurve for Excel の解析結果の解釈: Mauchly の球面性検定の項目で、 $p \geq 0.05$ であれば球面性仮定は棄却されず、そのまま分散分析表の球面性仮定の横のデータを読む。Mauchly の球面性検定の項目で、 $p < 0.05$ であれば自由度に掛けるいくつかの ϵ 値「Greenhouse-Geisser」「Huynh-Feldt-Lecoutre」「Chi-Muller」「下限値」に着目し、分散分析表の「Greenhouse-Geisser」「Huynh-Feldt-Lecoutre」「Chi-Muller」「下限値」の横のデータから結果を読む。

分散分析表での「Greenhouse-Geisser」「Huynh-Feldt-Lecoutre」「Chi-Muller」「下限値」のデータは調節した対応のある分散分析の F 値、p 値を示している。

実験計画法

R.A.Fisherによる

1. 無作為化(randomization): 確率分布を想定できる
2. 繰り返し(replication): 誤差分散の評価を可能にする
3. 局所管理(local control): 偶然誤差を小さくし実験の精度の向上

$$F_0 = \frac{V_A}{V_E} \geq F_{(a-1, N-a)}(\alpha) \text{ で検定した場合、}$$

V_E = 偶然誤差 + 個体差 + 温度差 + 慣れの誤差 + ...この中で、大きな影響がある要因を取り上げ、その影響を除くことが重要

実験計画において

ある因子A(ある薬剤)をA₁(5 μ g/ml), A₂(10 μ g/ml), A₃(15 μ g/ml)の3水準に分けて各3匹ずつのラットに一日のうち10a.m, 1p.m, 4p.mの3回の処理を3日間行なう。A₁, A₂, A₃にラットをrandom allocationしたとしても

A ₁	A ₁	A ₁	A ₂	A ₂	A ₂	A ₃	A ₃	A ₃
10	1	4	10	1	4	10	1	4
一日目			二日目			三日目		

このような実験では、慣れ、学習効果などの系統誤差がはいる。

完全無作為化 completely randomized design (=ブロック因子なし)

全部で9回の処理の順番を無作為化

A ₂	A ₃	A ₂	A ₁	A ₁	A ₃	A ₁	A ₂	A ₃
10	1	4	10	1	4	10	1	4
一日目			二日目			三日目		

⇒ 一元配置分散分析で解析する

この実験では一日目にA₂が2回処理されるなど日による影響が平等でない。

乱塊法 randomized block design

randomized block design (=ブロック因子1個)

一日に3回の処理をブロック化し無作為化(局所管理)

A ₂	A ₃	A ₁	A ₁	A ₃	A ₂	A ₁	A ₂	A ₃
10	1	4	10	1	4	10	1	4
一日目			二日目			三日目		

⇒ 効果、ブロック(日間変動)の2要因を二元配置分散分析で解析する

このような実験でも日内変動が無視できないこともある。

例題) 4種類の肥料に対し、その効果の差を推定するための実験計画を考える。実験を5回繰り返す場合に、乱塊法による実験計画表を作成する。

肥料の種類
肥料A
肥料B
肥料C
肥料D

乱塊法による実験計画表。

乱塊法では、繰り返しの中で処理がランダム化されており、一回の実験の中で同じ処理が出てくることはない。

[BellCurve for Excel: 分散分析多重比較→乱塊法](#)

乱塊法				
1回目	2回目	3回目	4回目	5回目
肥料B	肥料D	肥料B	肥料C	肥料B
肥料A	肥料C	肥料A	肥料A	肥料A
肥料C	肥料A	肥料D	肥料B	肥料D
肥料D	肥料B	肥料C	肥料D	肥料C

無作為抽出法による実験計画

無作為抽出法では、実験全体の中でランダム化されており、一回の実験の中で同じ処理が複数回出てくることもある。

無作為抽出法				
1回目	2回目	3回目	4回目	5回目
肥料D	肥料A	肥料D	肥料C	肥料A
肥料C	肥料B	肥料C	肥料B	肥料C
肥料B	肥料D	肥料B	肥料A	肥料A
肥料C	肥料A	肥料B	肥料D	肥料D

ラテン方格法 Latin square design (=ブロック因子2個)

各時間に処理が一回ずつになるようにする

A ₁	A ₂	A ₃	A ₂	A ₃	A ₁	A ₃	A ₁	A ₂
10	1	4	10	1	4	10	1	4
一日目			二日目			三日目		

⇒ 要因が効果、日間変動、日内変動の3要因を三元配置分散分析で解析
 一般のソフトでの解析が難しい → PDD3 BugsXLA(WinBUGS)参照

- 例えば、要因数7つでそれぞれの要因の水準数が3つの場合の実験(調査)をしようとする、要因と水準の組み合わせの数は2187通り(3×3×3×3×3×3×3)となるため、2187回の実験をしなくてはならない。実験回数を減らしても同じ結果を得られるようにできないかと1920年代にイギリスのR.A.フィッシャーによって考えだされた方法が実験計画法の始まり。直交表やラテン方格の計画行列といった組み合わせ表を使うことで実験回数を減らすことが可能となる。下記の実験の場合、直交表(L18)を使えば2187回の実験が18回の実験回数に減らすことができるなどあるが難しいので省略。

候補1

要因	第1水準	第2水準	第3水準	第4水準
a	a1	a2	a3	a4
b	b1	b2	b3	b4
c	c1	c2	c3	c4

候補2

要因	第1水準	第2水準	第3水準	第4水準
a	a1	a2		
b	b1	b2		
c	c1	c2	c3	c4
d	d1	d2	d3	d4

ブロック化の意義

- 主効果/誤差分散をはっきりさせる。
- ブロック/誤差分散で有意とでも、むしろブロック化に意義あり。
(ブロック化しなければ主効果/誤差分散のうち分母が大きくなっていた。)
- 2因子に交互作用がないことが前提条件
(2因子が圧と温度などでは、交互作用の可能性があるが、1因子が薬剤、1因子(=ブロック)が日や処理順を変えるなどは交互作用はない。)

ノンパラメトリック検定のまとめ

ウィルコクソンの符号付順位検定 Wilcoxon Signed Ranks Test

対応のある2つの標本について、それぞれのデータ対の差の順にもとづいて検定する。変数が順序尺度、もしくは、正規性があるか不明で間隔・比例尺度の場合に使うことができ、データ対の順位がわかる場合は、符号検定よりも効率が良い。

検定統計量の算出

- 2つの標本のデータの各組を差 $d_i = A_i - B_i$ の絶対値を求める。
(差が0の組は、この後の手続きから除外し、標本数 n を、差が0でない組の数とする)。それぞれの差の絶対値 $|d_i|$ に対応する組の数をもとに、差の絶対値の小さいほうから順位をつける
同一順位の場合は、次のように扱う(平均順位)
2位が2つある場合: 2位と3位の中間 $(2+3)/2=2.5$ 位
4位が3つある場合: 4位と5位と6位の中間 $(4+5+6)/3=5$ 位
- 差 d_i が正の値の順位の和を $T+$ 、負の値の順位の和を $T-$ とし $T+$ と $T-$ の小さい方の値を T とする。
 - データ数 n が $n \leq 25$ のときは Wilcoxon の検定表の n に対応する T_α を求める。
 $T < T_\alpha$ のとき帰無仮説を棄却する。 有意差あり。
 $T \geq T_\alpha$ のとき帰無仮説を棄却できない。
 - データ数 n が $n > 25$ のときは下の式から平均値と標準偏差を出して z 値を出す。この式によって T の分布は近似的に正規分布する。

$$\text{平均値 } U_T = \frac{n(n+1)}{4}$$

$$\text{標準偏差 } \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

$$z = \frac{T - U_T}{\sigma_T} \quad \text{標準正規分布表から確率 } P \text{ を求める。}$$

例) 開腹手術をした患者 10 名について、手術 1 週間後と 2 週間後で ATP 放出量に変化があったかウィルコクソンの符号付き順位検定により検定。

開腹手術を施行した婦人科患者 10 名の手術 1 週間後と 2 週間後のコラーゲン添加時血小板ATP放出量

No.	1週間後	2週間後	差
1	27.1	38.1	-11.0
2	20.0	19.6	0.4
3	28.6	44.4	-15.8
4	6.6	0.8	5.8
5	10.5	17.9	-7.4
6	6.9	6.9	0.0
7	20.8	22.8	-2.0
8	15.8	27.6	-11.8
9	15.4	17.1	-1.7
10	24.1	32.6	-8.5

BellCurve エクセル統計より「開腹手術を施行した婦人科患者 10 名の手術 1 週間後と 2 週間後のコラーゲン添加時血小板ATP放出量古川 俊之, 丹後 俊郎. (1993). 『新版 医学への統計学』. 朝倉書店.

BellCurve for Excel: 2 標本の比較→ウィルコクソンの符号付き順位検定

ウィルコクソンの符号付き順位検定			
基本統計量			
変数	n		
1週間後	10		
2週間後	10		
ケースの要約 (d=「1週間後」-「2週間後」)			
	n	順位和	
d=0の対	1		
d<0の対	7	40	
d>0の対	2	5	
合計	10		
ウィルコクソンの符号付き順位検定			
統計数値表による検定		正規化検定	
統計量	両側検定	統計量:z	両側P値 * : P<0.05 ** : P<0.01
	5 5%有意	2.0732	0.0382 *
10例のうち差が0の1例を除いたn=9,T=5			
検定表でn=9,T=5の時両側5%のP=0.039 (有意)			

例題(小標本)

A 大学と B 大学は今年の 12 種目のインカレでの順位はそれぞれ下のようになった。この結果から、どちらかの大学の方が上回っているといえるかどうかを検定しなさい。

種目	a	b	c	d	e	f	g	h	i	j	k	l
A 大学	3	16	4	3	4	7	1	2	3	7	10	4
B 大学	5	8	7	3	5	2	7	11	10	2	3	14
競技順位の差(A-B)	-2	8	-3	0	-1	5	-6	-9	-7	5	7	-10

帰無仮説:A 大学と B 大学の順位には差がない。

対立仮説:どちらかの大学の順位が、全体的に上回っている。

まず、符号を無視して差 d を小さい方から順に並べて順位をつける。

その後、d の符号(+と-)によって順位を分ける。

差 d	-1	-2	-3	5	5	-6	7	-7	8	-9	-10
順位	1	2	3	4.5	4.5	6	7.5	7.5	9	10	11

+	4.5	4.5	7.5	9							
-	1	2	3	6	7.5	10	11				

＋の符号の数の数が少ないので、＋の符号の順位を足して T 値とする。

$$T \text{ 値} = 4.5 + 4.5 + 7.5 + 9 = 25.5$$

差が 0 のデータは除くので、データ数 $n = 12 - 1 = 11$ となる。

$n \leq 25$ なので、Wilcoxon 検定表から $T_{0.05}$ を求める。表を見ると $n = 11$ のとき、 $P < 0.05$ の T の有意点は 10 であり、計算された T 値は両側確率 $P < 0.05$ となる T の有意点よりも大きい。よって、帰無仮説を棄却することができず、「どちらかの大学の順位が、全体的に上回っている」ということができない。

インターネット <http://kusuri-jouhou.com/statistics/wilcoxon.html> より一部改変

BellCurve for Excel: 2 標本の比較→ウィルコクソンの符号付き順位検定(データは列に変換)

ウィルコクソンの符号付き順位検定			
変数	n		
A大学	12		
B大学	12		
ケースの要約 (d = 「A大学」 - 「B大学」)			
	n	順位和	
d=0の対	1		
d<0の対	7	40.5	
d>0の対	4	25.5	
合計	12		
ウィルコクソンの符号付き順位検定			
統計数値表による検定		正規化検定	
統計量	両側検定	統計量:z	両側P値
	25.5 -	0.666831	0.50488
n=11の時 $T \leq 10$ 、 $P = 0.04199$ は有意、今回 $T = 25.5 > 10$			

データ数 n が $n \leq 25$ の場合
 $T \geq T_{\alpha}$ のとき帰無仮説を棄却できない。

ウィルコクソンの符号付順位和検定の統計表								
N	両側検定 $\alpha=0.05$		両側検定 $\alpha=0.01$		片側検定 $\alpha=0.05$		片側検定 $\alpha=0.01$	
5	---	---	---	---	0	0.03125	---	---
6	0	0.03125	---	---	2	0.04688	---	---
7	2	0.04688	---	---	3	0.03906	0	0.00781
8	3	0.03906	0	0.00781	5	0.03906	1	0.00781
9	5	0.03906	1	0.00781	8	0.04883	3	0.00977
10	8	0.04883	3	0.00977	10	0.04199	5	0.00977
11	10	0.04199	5	0.00977	13	0.04150	7	0.00928
12	13	0.04248	7	0.00928	17	0.04614	9	0.00806
13	17	0.04785	9	0.00806	21	0.04712	12	0.00854
14	21	0.04944	12	0.00854	25	0.04529	15	0.00830
15	25	0.04791	15	0.00836	30	0.04730	19	0.00903
16	29	0.04431	19	0.00919	35	0.04672	23	0.00912
17	34	0.04477	23	0.00934	41	0.04919	27	0.00871
18	40	0.04828	27	0.00896	47	0.04937	32	0.00912
19	46	0.04937	32	0.00945	53	0.04776	37	0.00904
20	52	0.04844	37	0.00944	60	0.04865	43	0.00962
21	58	0.04599	42	0.00902	67	0.04790	49	0.00974
22	65	0.04616	48	0.00927	75	0.04921	55	0.00948
23	73	0.04844	54	0.00915	83	0.04899	62	0.00978
24	81	0.04906	61	0.00958	91	0.04755	69	0.00972
25	89	0.04826	68	0.00964	100	0.04787	76	0.00937
26	98	0.04935	75	0.00938	110	0.04967	84	0.00949
27	107	0.04910	83	0.00961	119	0.04769	92	0.00933
28	116	0.04775	91	0.00954	130	0.04964	101	0.00956
29	126	0.04803	100	0.00988	140	0.04817	110	0.00952
30	137	0.04971	109	0.00993	151	0.04805	120	0.00983
31	147	0.04789	118	0.00974	163	0.04909	130	0.00989
32	159	0.04977	128	0.00991	175	0.04919	140	0.00972
33	170	0.04843	138	0.00984	187	0.04848	151	0.00987
34	182	0.04838	148	0.00957	200	0.04882	162	0.00981
35	195	0.04946	159	0.00963	213	0.04839	173	0.00956
36	208	0.04964	171	0.00996	227	0.04890	185	0.00959
37	221	0.04904	182	0.00962	241	0.04868	198	0.00986

38	235	0.04947	194	0.00956	256	0.04932	211	0.00994
39	249	0.04915	207	0.00976	271	0.04928	224	0.00986
40	264	0.04976	220	0.00978	286	0.04862	238	0.00999
41	279	0.04966	233	0.00964	302	0.04875	252	0.00997
42	294	0.04892	247	0.00973	319	0.04960	266	0.00980
43	310	0.04904	261	0.00966	336	0.04983	281	0.00984
44	327	0.04994	276	0.00980	353	0.04950	296	0.00974
45	343	0.04885	291	0.00979	371	0.04984	312	0.00982
46	361	0.04987	307	0.00997	389	0.04966	328	0.00977
47	378	0.04901	322	0.00969	407	0.04899	345	0.00989
48	396	0.04889	339	0.00992	426	0.04896	362	0.00988
49	415	0.04943	355	0.00971	446	0.04949	379	0.00976
50	434	0.04945	373	0.00996	466	0.04954	397	0.00979

例題(大標本)

A 大学と B 大学は毎年スポーツ種目で競い合っている。今年の 30 種目のインカレでの順位はそれぞれ下のようになった。この結果から、どちらかの大学の方が上回っているといえるかどうかを検定しなさい。

種目	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
A 大学	3	16	4	2	4	7	1	2	3	7	10	4	3	3	5
B 大学	5	8	7	2	5	2	7	11	10	2	3	14	2	15	13
競技順位の差 (A-B)	-2	8	-3	0	-1	5	-6	-9	-7	5	7	-10	1	-12	-8
種目	p	q	r	s	t	u	v	w	x	y	z	α	β	γ	δ
A 大学	3	18	2	3	7	1	3	1	8	9	5	13	4	3	13
B 大学	6	6	15	6	3	17	2	5	2	11	2	2	1	9	2
競技順位の差 (A-B)	-3	12	-13	-3	4	-16	1	-4	6	-2	3	11	3	-6	11

インターネット <http://kusuri-jouhou.com/statistics/wilcoxon.html> より一部改変

BellCurve for Excel: 2 標本の比較→ウィルコクソンの符号付き順位検定

ウィルコクソンの符号付き順位検定															
種目	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
A大学	3	16	4	2	4	7	1	2	3	7	10	4	3	3	5
B大学	5	8	7	2	5	2	7	11	10	2	3	14	2	15	13
競技順位 (A-B)	-2	8	-3	0	-1	5	-6	-9	-7	5	7	-10	1	-12	-8
種目	p	q	r	s	t	u	v	w	x	y	z	α	β	γ	δ
A大学	3	18	2	3	7	1	3	1	8	9	5	13	4	3	13
B大学	6	6	15	6	3	17	2	5	2	11	2	2	1	9	2
競技順位 (A-B)	-3	12	-13	-3	4	-16	1	-4	6	-2	3	11	3	-6	11
	A大学	B大学													
	3	5													
	16	8													
	4	7													
	2	2													
	4	5													
	7	2													
	1	7													
	2	11													
	3	10													
	7	2													
	10	3													
	4	14													
	3	2													
	3	15													
	5	13													
	3	6													
	18	6													
	2	15													
	3	6													
	7	3													
	1	17													
	3	2													
	1	5													
	8	2													
	9	11													
	5	2													
	13	2													
	4	1													
	3	9													
	13	2													

ウィルコクソンの符号付き順位検定								
基本統計量								
変数	n							
A大学	30							
B大学	30							
ケースの要約 (d=「A大学」-「B大学」)								
	n	順位和						
d=0の対	1							
d<0の対	16	246						
d>0の対	13	189						
合計	30							
ウィルコクソンの符号付き順位検定								
統計数値表による検定			正規化検定					
統計量	両側検定		統計量	両側P値	* : P<0.05 ** : P<0.01			
	189 -		0.6163	0.5377				

有意差なし

中央値検定(メディアン検定) Median Test

例題:「表のような観察値が得られた。両群の母代表値(母中央値)に差があるかどうか検定せよ。」

二群の観察値					
第1群の観察値	1.2	1.5	1.8	2.6	
第2群の観察値	1.3	1.9	2.9	3.1	3.9

検定手順:

1. 前提

帰無仮説 H_0 :「2群の母代表値(母中央値)に差はない」

対立仮説 H_1 :「2群の母代表値(母中央値)に差がある」

有意水準 α で両側検定を行う(片側検定も定義できる)

2. 2群のデータをこみにして、中央値を求める。

例題では、データを小さい順に並べると 1.2, 1.3, 1.5, 1.8, 1.9, 2.6, 2.9, 3.1, 3.9 となるので、中央値は 1.9 である。

3. 各群のデータを中央値以上と中央値未満に分け、以下の分割表を作る。

2×2 分割表			
	中央値以上	中央値未満	合計
第 1 群	a	b	a+b
第 2 群	c	d	c+d
合計	a+c	b+d	n

例題では

2×2 分割表			
	中央値以上	中央値未満	合計
第 1 群	1	3	4
第 2 群	4	1	5
合計	5	4	9

4. 得られた分割表に対して、独立性の検定を行う。すなわち、群と中央値より大きいか小さいかが独立であるかどうかを検定するわけである。
 例題では、独立性の検定の結果、有意確率は $P=0.0989$ となり、帰無仮説が採択される。

<http://www.weblio.jp>より引用

BellCurve for Excel: ノンパラメトリック検定→中央値検定

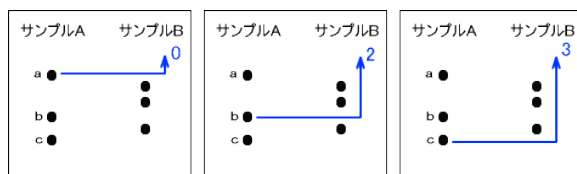
中央値検定				
	第 1 群の観察値	第 2 群の観察値		
	1.2	1.3		
	1.5	1.9		
	1.8	2.9		
	2.6	3.1		
		3.9		
基本統計量				
変 数	n	中央値		
第 1 群の観察値	4	1.650		
第 2 群の観察値	5	2.900		
	中央値未満	中央値以上	合 計	
第 1 群の観察値	3	1	4	
第 2 群の観察値	1	4	5	
合 計	4	5	9	
全体の中央値=1.9				
中央値検定				* : P<0.05 ** : P<0.01
手 法	カイ二乗値	自由度	P 値	有意差なし
独立性の検定	2.723	1	0.0989	
Yatesの補正	0.951	1	0.3296	
Fisherの直接確率 (両側P値)			0.2063	
Fisherの直接確率 (片側P値)			0.1667	

マン・ホイットニーのU検定 Mann-Whitney U Test

A) 独立2群の差を検定するときのノンパラメトリックな方法。

統計量Uの求め方

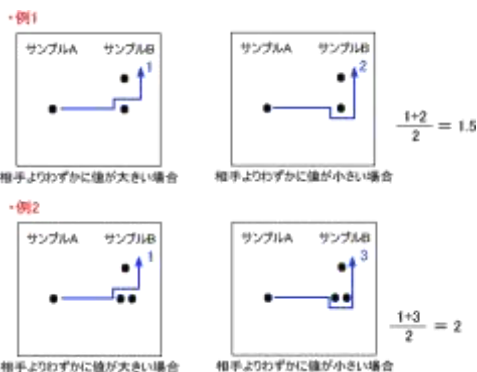
M統計量Uは、片方の群のそれぞれの点に注目して、個々の点よりも大きい他方のデータ数を数える。



上の図では、統計量Uは $2 + 3 = 5$ となる。

(ただし、サンプルBに注目して統計量Uは $1 + 1 + 2 = 4$ でもよい。)

もし、両群に同じ値があるなら「相手よりわずかに値が大きい場合」と「相手よりわずかに値が小さい場合」を想定してその平均値をとる。



帰無仮説(H_0): 「2群間に差がない」

対立仮説(H_1): 「2群間に差がある」

統計量Uは

1) $n_1 \leq 20$ かつ $n_2 \leq 20$ のとき、 $n_1 n_2 = U + U'$ が成立し Mann-Whitney検定表の数値Aとすると(ここでも $n_1 n_2 = A + A'$ が成立し) $U < U'$, A (= 下側確率) $<$ A' (= 上側確率)とした場合

$U < A$ または $A' < U'$ であれば有意

2) n_1 または n_2 の一方が20より大きいとき、下の式から平均値と標準偏差を出してz値を出す。この式によってUの分布は近似的に正規分布する。

$$\text{平均値} \quad \mu_U = \frac{n_1 n_2}{2}$$

$$\text{標準偏差} \quad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

$$z = \frac{U - \mu_U}{\sigma_U} \quad \text{標準正規分布表から確率 } P \text{ を求める。}$$

例題(小標本)

大腸がんを患っている患者と胃がんを患っている患者に鎮痛薬Xを投与した。このとき、薬がどれくらい効き目があったかを医師にアンケートをとって、20段階で評価してもらった。大腸がんと胃がんで鎮痛薬Xに効果の違いがあるといつてよいか検定しなさい。

大腸がん	20	18	15	13	10	6		
胃がん	17	16	12	9	8	6	4	2

統計量Uを大腸がんの群に注目して求める。

$$U \text{ 値} = 2 + 2 + 3 + 5.5 = 12.5 (U' = 6 \times 8 - 12.2 = 35.8)$$

Mann-Whitney検定表で、 $n_1 = 6$ 、 $n_2 = 8$ のとき下側有意点は8(上側確率 $6 \times 8 - 8 = 40$)であり、計算したU値は8よりも大きい($35.8 < 40$ より)ので、 $P \geq 0.05$ となり帰無仮説を棄却できない。「大腸がんとうがんで鎮痛薬Xに効果の違いがある」とはいえない。

(統計量Uまで計算され、検定表からの読み取りは自分ですする必要あり。ただし、同時に示される正確検定などの結果は重要。これにより判断できる)

BellCurve for Excel: ノンパラメトリック検定 → マン=ホイットニー検定 (行 → 列データに変換して)

マン=ホイットニーのU検定						
基本統計量						
変数	n	平均順位	統計量:U			
大腸がん	6	9.417	12.5			
胃がん	8	6.063	35.5			
有意差なし						
マン=ホイットニーのU検定						
統計数値表による検定			正規化検定		* : P<0.05 ** : P<0.01	
統計量:U	両側検定	E(U)	V(U)	連続性補正	統計量:z	両側P値
12.5	※ 1	24	59.868	なし	1.4863	0.1372
				あり	1.4217	0.1551
正確検定						
両側P値	* : P<0.05 ** : P<0.01					
0.1505						
※1 同順位があるため統計数値表を利用できません。						

Mann-Whitney検定表(両側確率)

$P < 0.05$

Mann-Whitney検定表																	
$n_2 =$	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$n_1 = 1$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	0	0	0	0	1	1	1	1	1	2	2	2	2
3	-	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5		2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6			5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7				8	10	12	14	16	18	20	22	24	26	28	30	32	34
8					13	15	17	19	22	24	26	28	31	34	36	38	41
9						17	20	23	26	28	31	34	37	39	42	45	48
10							23	26	29	33	36	39	42	45	48	52	55
11								30	33	37	40	44	47	51	55	58	62
12									37	41	45	49	53	57	61	65	69
13										45	50	54	59	63	67	72	76
14											55	59	64	67	74	78	83
15												64	70	75	80	85	90
16													75	81	86	92	98
17														87	93	99	105
18															99	106	112
19																113	119
20																	127

例題(大標本)

A大学の野球部とサッカー部に10種類の体力テストを5段階で評価した。スコアを個人別に求めると下図になった。体力テストのスコアに差があるといつてよいか。

野球部	50	49	49	47	45	45	43	43	43	41
サッカー部	48	48	46	44	42	42	42	41	41	36
野球部に注目				2	3	3	4	4	4	8

野球部	38	37	36	35	34	33	33	31	31	30		
サッカー部	34	32	30	30	29	28	28	28	26	25	24	22
野球部に注目	9	9	9.5	10	10.5	11	11	12	12	13		

統計量Uを野球部に注目して求める。

$$U \text{ 値} = 2 + 3 + 3 + 4 + 4 + 4 + 8 + 9 + 9 + 9.5 + 10 + 10.5 + 11 + 11 + 12 + 12 + 13 = 135$$

近似的に正規分布となる。

$$\text{平均値 } \mu_U = \frac{20 \times 22}{2} = 220$$

$$\text{標準偏差 } \sigma_U = \sqrt{\frac{20 \times 22(20 + 22 + 1)}{12}} = 39.7$$

$$z = \frac{135 - 220}{39.7} = -2.14 < -1.96$$

有意差あり(野球部とサッカー部で体力テストのスコアに差がある)

BellCurve for Excel: ノンパラメトリック検定→マン=ホイットニー検定(行→列データに変換して)

マン=ホイットニーのU検定							
基本統計量							
変数	n	平均順位	統計量:U				
野球部	20	25.750	135				
サッカー部	22	17.636	305				
マン=ホイットニーのU検定							
統計数値表による検定			正規化検定		* : P<0.05 ** : P<0.01		
統計量:U	両側検定	E(U)	V(U)	連続性補正	統計量:z	両側P値	
135	※ 1	220	1573.217	なし	2.1430	0.0321	*
				あり	2.1304	0.0331	*
正確検定							
両側P値	* : P<0.05 ** : P<0.01						
※ データ数の合計が20を超えるため正確検定は出力できません。							
※1 同順位があるため統計数値表を利用できません。							

B)表形式で段階的評価(例えば 1.著効、2.有効、3. 不変、4. 悪化)がある場合

Mann-Whitney 検定で評価することができる。

例)新薬 A と標準薬 B の 2 種類の薬剤を投与し 1.著効、2.有効、3. 不変、4. 悪化の 4 群に分けて評価した。

	著効	有効	不変	悪化
A 群	5	10	6	2
B 群	2	5	10	4

この表を χ^2 検定で分析するのは間違い。段階的評価があるので MH 検定

マン=ホイットニーのU検定						
基本統計量						
変数	n	平均順位	統計量:U			
A群	23	18.957	323			
B群	21	26.381	160			
マン=ホイットニーのU検定						
統計数値表による検定			正規化検定		* : P<0.05 ** : P<0.01	
統計量:U	両側検定	E(U)	V(U)	連続性補正	統計量:z	両側P値
160	※ 1	241.5	1641.358	なし	2.0117	0.0443 *
				あり	1.9993	0.0456 *
正確検定						
両側P値 * : P<0.05 ** : P<0.01						
※ データ数の合計が20を超えるため正確検定は出力できません。						
※1 同順位があるため統計数値表を利用できません。						
 独立性の検定=この場合は間違い (MHを選択) 						
観測度数						
	著効	有効	不変	悪化	合計	
A群	5	10	6	2	23	
B群	2	5	10	4	21	
合計	7	15	16	6	44	
期待度数						
	著効	有効	不変	悪化		
A群	3.659	7.841	8.364	3.136		
B群	3.341	7.159	7.636	2.864		
太字 : 1未満 赤字 : 5未満						
独立性の検定						
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01			
4.5375	3	0.2090				
Cramer's \	0.3211					

並べ替え検定 permutation test ⇒P49

平均値差の正しい p 値を求める方法は、並べ替え検定(permutation test)や確率化検定(randomization test)と呼ばれる。並べ替え検定はノンパラメトリックな方法に分類される。手元の標本データを母集団と見なし、そこから無作為抽出を繰り返すことにより全データを並び替え、得られた全ての場合において、元データが得られる確率を求める。並べ替え検定では、この2群はもともと同じ母集団に属していたサンプルで、母集団を代表している値であると考え(帰無仮説は「2つの母集団は同じものである」)。手元のデータから各群に再度割りつけるという作業を行い繰り返す。可能な組み合わせすべてができるまで繰り返す
マン・ホイットニーの U 検定(20例以下、正確検定)では、正確確率が計算でき並べ替え検定と同じ p 値が算出される。

ブルンナー・ムンツェル検定 Brunner-Munzel Test ⇒P50

分布が同じことは仮定せず、両群から一つずつ値を取り出したとき、どちらかが大きい確率自体が等しいという帰無仮説を検定する。2つの群の相対的な順位を比較する手法で、ノンパラメトリック検定の1つ。2群のデータを1つにまとめて順位付けし、この順位データから検定を行う。正規分布に従わなくてもよく、等分散性の仮定も必要ない。大標本では検定統計量Wが標準正規分布に従うことを利用。小標本の場合は、補正した自由度を用いたt検定による結果を利用。Mann-Whitney'U testの改良版。

クラスカル・ウォリス検定 Kruskal-Wallis検定 ⇒P67

Bartlett検定で「各群の分散は均一である」と判断されれば一元配置分散分析法をする。「各群の分散は均一でない」と判断されればKruskal-Wallis検定をする。

帰無仮説(H₀):「各群で差がない」、対立仮説(H₁):「各群で差がある」

Kruskal-Wallis検定では、まずすべてのデータを小さいものから1番→2番→.....→N番と順位を付ける。この順位を各群ごとに足してR_iを求める。次いで、下の計算式によって統計量Hを計算する。(N=データの総数、k=群数)

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

k=3かつN≤17のとき → Kruskal-Wallis検定表から求める

P≥αのとき帰無仮説を棄却できない。

P<αのとき帰無仮説を棄却する。

上の条件以外するとき → Hは近似的に自由度df=k-1のX²分布に従う

X²≤X²_αのとき、P≥αとなり帰無仮説を棄却できない

X²>X²_αのとき、P<αとなり帰無仮説を棄却する

例)あるホルモンの活性を調べると、次のような結果を得た。群間に差があるかどうかを検定せよ。

					データ数	平均値	分散
A群	21	23	19	25	4	22	5
B群	10	6	14		3	10	10.7
C群	25	31	29	23	4	27	10
D群	14	16	21		3	17	8.7

Bartlett検定で「各群の分散は均一である」と判断されているが、ここでは「各群の分散は均一でない」として検定を進める。

帰無仮説(H₀):ホルモン活性に差がない。対立仮説(H₁):ホルモン活性に差がある。
データを小さい順で1から14番まで順位をつける。そのときの順位は下のようになる。

					R _i
A群	7.5	9.5	6	11.5	34.5
B群	2	1	3.5		6.5
C群	11.5	14	13	9.5	48
D群	3.5	5	7.5		16

$$N = 4 + 3 + 4 + 3 = 14$$

順位を足してR_iを求め、次の式によって統計量Hを計算する。

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1) = \frac{12}{14(14+1)} \left(\frac{34.5^2}{4} + \frac{6.5^2}{3} + \frac{48^2}{4} + \frac{16^2}{3} \right) - 3 * (14 + 1)$$

この場合、k=3かつN≤17ではないのでX²_αを求める。X²_{0.05}=7.815であると分かる。このときの自由度はdf=k-1=4-1=3である。

X² = 10.595 > 7.815 = X²_{0.05} より、P < 0.05となり帰無仮説は棄却される。よって、「ホルモン活性に差がある」といえる。

BellCurve for Excel: ノンパラメトリック検定→クラスカル=ウォリス検定と多重比較

クラスカル=ウォリス検定と多重比較				
水準	A群	B群	C群	D群
n	4	3	4	3
平均順位	8.63	2.17	12.00	5.33
クラスカル=ウォリス検定				
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01	
10.6928	3	0.0135	* 有意差あり	
多重比較 : Scheffe				
水準1	水準2	カイ二乗値	P 値	* : P<0.05 ** : P<0.01
A群	B群	4.1221	0.2486	
A群	C群	1.3133	0.7260	
A群	D群	1.0708	0.7841	BC間で有意差あり
B群	C群	9.5561	0.0227 *	
B群	D群	0.8671	0.8333	
C群	D群	4.3924	0.2221	

◇ ノンパラメトリック法に基づく多重比較法

ノンパラメトリック法に基づく多重比較法を適用させるにあたり、注意しなければならない点を以下に挙げる。

1. 比較対照とする標本の母集団分布はどのような分布でも良いが、各標本の母集団分布は同一でなければならない。
2. 各標本の不等分散性が存在する場合でも、母集団分布が不明とは限らない。母集団が正規分布に従って、各標本が等分散とならないこともある。
3. 標本の大きさが小さいとき(特に、大きさ $n < 10$ のとき)、ノンパラメトリック法の適用が困難となる(大標本近似のため)。また、各標本の大きさが著しく異なるときにも誤った判定をすることがある。
4. 同順位が多いとき、同順位補正が必要。

ノンパラメトリック法が適するのは、外れ値(outlier)が存在するときや、変数の測定尺度が順序尺度のときである。

以下に代表的な手順を述べる。

Steel-Dwass の方法: **Tukey** の方法のノンパラメトリック版である。従って、全ての対比較を同時に検定するための多重比較法となる。

Steel の方法: パラメトリック法の **Dunnnett** の方法と同様の適用である。対照群と他の2つ以上の処理群を対比較するとき用いる方法である。

Dunnの方法: 平均値の代わりに平均順位を用いて検定する方法。Dunnett の方法と同様に 1つの対照群と 2つ以上の処理群の対比較のみを同時に検定する手順もある。米国 National Institutes of Healthの National Toxicology Program 毒性癌原性試験の報告書で常用されている方法である。

Shirley-Williamsの方法: Williamsの方法のノンパラメトリック版

なお, Bonferroni の方法は検定統計量に制限が無いことから, ノンパラメトリックな手法としても考えることができる。

フリードマン検定 Friedman Test →P99

Bartlett 検定で「各群の分散が等しくない」と判定されたら一元配置分散分析(対応あり)、二元配置分散分析法(繰り返しなし)の代わりに、Friedman 検定を用いる。

2×2 分割表の周辺

2×2 表での分析

検査の信頼性 (感度、特異性、陽性・陰性適中率、陽性・陰性尤度比、診断・誤診断のオッズ比)

臨床研究の分析 (オッズ比、相対危険度、寄与危険度、NNT など)

検定 (χ^2 、フィッシャー、コーエンの κ 係数、マクネマー)

他の関連性測定(ϕ 係数、クラメル V、ユールの関連係数 Q)

		検査		計
		陽性	陰性	
疾患	あり	a	b	a+b
		陽性	偽陰性	有病者
	なし	c	d	c+d
		偽陽性	陰性	無病者
		a+c	b+d	a+b+c+d
	計	陽性者	陰性者	総数(N)

	検査陽性	検査陰性	
有病	真陽性率 = 感度 $1-\beta$	偽陰性率 = $(1-\text{感度})$ β	有病率
無病	偽陽性率 = $(1-\text{特異度})$ α	真陰性率 = 特異度 $1-\alpha$	$1-\text{有病率}$
計	陽性率	陰性率	全体

$$\frac{a}{a+b} = \frac{\text{検査+}}{\text{疾患+}} = \text{感度(sensitivity)} \quad : \text{有病者で検査が陽性になる確率}$$

$$\frac{d}{c+d} = \frac{\text{検査-}}{\text{疾患-}} = \text{特異度(specificity)} \quad : \text{無病者で検査が陰性になる確率}$$

$$\frac{c}{c+d} = \frac{\text{検査+}}{\text{疾患-}} = \text{偽陽性率(false positive value)}$$

α (第 I 種過誤率)

$$\frac{b}{a+b} = \frac{\text{検査-}}{\text{疾患+}} = \text{偽陰性率(false negative value)}$$

β (第 II 種過誤率)

$$\frac{a}{a+c} = \frac{\text{疾患+}}{\text{検査+}} = \text{陽性適中率(positive predictive value :PPV)}$$

$$\frac{d}{b+d} = \frac{\text{疾患-}}{\text{検査-}} = \text{陰性適中率(negative predictive value :NPV)}$$

$$\frac{a+b}{N} = \text{有病率(prevalence rate)}$$

$$\frac{a+d}{N} = \text{感度} \times \text{有病率} + \text{特異度} \times (1 - \text{有病率}) = \text{正診率(accuracy)}$$

検査の信頼性

$$\frac{a+c}{N} = \frac{a}{N} + \frac{c}{N} = \frac{a}{a+b} \cdot \frac{a+b}{N} + \frac{c}{c+d} \cdot \frac{c+d}{N} =$$

$$= \text{感度} \times \text{有病率} + (1 - \text{特異度}) \times (1 - \text{有病率})$$

$$= (\text{検査}) \text{陽性率} (\text{positive ratio})$$

尤度比: 陽性または陰性における、同じ判定内での有病者と無病者の比

陽性尤度比(likelihood ratio of positive result :LR+)

$$\frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{\text{感度}}{(1 - \text{特異度})} = \frac{\frac{\text{陽性}}{\text{有病者}}}{\frac{\text{偽陽性}}{\text{無病者}}}$$

LR+: 臨床的には、ある検査結果が“陽性”の場合に
“診断確定”の可能性が検査前の何倍になったかを示す
→ 検査後(事後)オッズ = 検査前(事前)オッズ × LR+

陰性尤度比(likelihood ratio of negative result :LR-)

$$\frac{\frac{b}{a+b}}{\frac{d}{c+d}} = \frac{(1 - \text{感度})}{\text{特異度}} = \frac{\frac{\text{偽陰性}}{\text{有病者}}}{\frac{\text{陰性}}{\text{無病者}}}$$

検査後確率(=陽性適中率 PPV) = $\frac{\text{検査後オッズ}}{1 + \text{検査後オッズ}}$

* 尤度: 仮説が成り立つとしたらどの程度の確率でデータが得られるか(≒確率)

オッズ: 起きる確率と起きない確率の比

$$\frac{\text{有病者}}{\text{無病者}} = \frac{a+b}{c+d} = \frac{\text{有病率}}{1 - \text{有病率}} = \text{事前オッズ}$$

$$\frac{a}{c} = \frac{a(c+d)}{c(a+b)} \cdot \frac{a+b}{c+d} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} \cdot \frac{a+b}{c+d} = \text{陽性尤度比} \times \text{事前オッズ}$$

= 陽性の事後オッズ

$$\frac{b}{d} = \frac{b(c+d)}{d(a+b)} \cdot \frac{a+b}{c+d} = \frac{\frac{b}{a+b}}{\frac{d}{c+d}} \cdot \frac{a+b}{c+d} = \text{陰性尤度比} \times \text{事前オッズ}$$

= 陰性の事後オッズ

$$\frac{\frac{a}{c}}{\frac{b}{d}} = \frac{\text{陽性の事後オッズ}}{\text{陰性の事後オッズ}} = \text{オッズ比}$$

$$\Rightarrow \text{オッズ比} = \frac{\text{陽性尤度比}}{\text{陰性尤度比}} = \frac{\text{感度} \times \text{特異度}}{(1 - \text{感度})(1 - \text{特異度})}$$

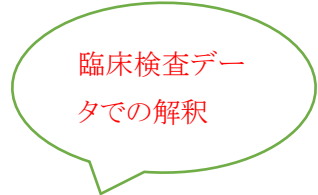
$$\text{検査前オッズ(事前オッズ)} = \frac{\text{有病者}}{\text{無病者}} = \frac{\text{有病率}}{1-\text{有病率}} = \frac{a+b}{c+d} = \frac{\text{検査前確率}}{1-\text{検査前確率}}$$

→ 検査前確率 = 「有病率」、(1 - 検査前確率) = 「疾病なしの確率」に相当する

$$\text{検査後オッズ(事後オッズ)} \quad (\text{陽性の場合}) = \frac{\text{陽性で疾病あり(a)}}{\text{陽性で疾病なし(c)}} = \frac{\text{検査後確率}}{1-\text{検査後確率}}$$

$$\rightarrow \text{検査後確率(=陽性適中率 PPV)} = \frac{\text{検査後オッズ}}{1+\text{検査後オッズ}}$$

ベイズの定理 → p140 検査後オッズ = 検査前オッズ × 尤度比



検査後確率 = 陽性適中率 (positive predictive value : PPV)

$$\begin{aligned} &= \frac{a}{a+c} = \frac{\frac{a}{(a+b)} \cdot \frac{(a+b)}{N}}{\frac{(a+c)}{N}} = \frac{\text{感度} \times \text{有病率}}{\text{検査陽性率}} = \frac{\text{感度} \times \text{有病率}}{\text{感度} \times \text{有病率} + (1-\text{特異度}) \times (1-\text{有病率})} = \frac{1}{1 + \frac{(1-\text{特異度}) \times (1-\text{有病率})}{\text{感度} \times \text{有病率}}} \\ &= \frac{1}{1 + \frac{1}{\text{LR}+} \times \frac{(1-\text{有病率})}{\text{有病率}}} = \frac{1}{1 + \frac{1}{\text{LR}+} \times \frac{(1-\text{検査前確率})}{\text{検査前確率}}} = \frac{1}{1 + \frac{1}{\text{陽性尤度比}} \times \frac{1}{\text{検査前オッズ}}} = \frac{1}{1 + \frac{1}{\text{検査後オッズ}}} = \frac{\text{検査後オッズ}}{1 + \text{検査後オッズ}} \end{aligned}$$

* 検査の信頼度は感度、特異度よりも陽性(または陰性)適中率で判断する

例)

1.) 検査前確率(pre-test probability)が 20%、尤度比が 4 の時の検査後確率は?
検査前オッズ = 20/80 = 1/4 = 0.25 ベイズの定理より、0.25 × 4 = 1 (検査後オッズ)

$$\text{検査後確率} = \text{陽性適中率} = \frac{\text{検査後オッズ}}{1+\text{検査後オッズ}} = \frac{1}{1+(1+1)} = 1/2 = 0.5 = 50\%$$

「検査陽性であった場合に本当に疾病に罹患している確率が 50%である」の意味。

2) 感度 90%、特異度 95%、検査前確率 0.2、この時の検査後確率は?

$$\text{陽性尤度比} = 90\% / (100\% - 95\%) = 90/5 = 18$$

$$\text{検査前オッズ} = 20\% / (100\% - 20\%) = 20/80 = 1/4 = 0.25$$

$$\text{検査後オッズ} = 0.25 \times 18 = 4.5$$

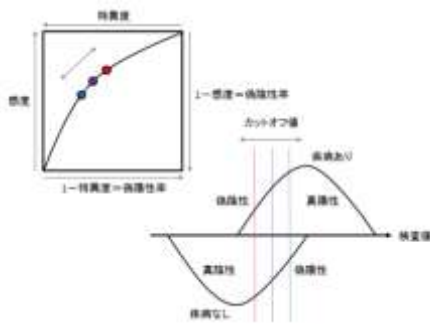
$$\text{検査後確率} = \text{陽性適中率} = 4.5 / (1 + 4.5) = 4.5 / 5.5 = 0.818 = 81.8\%$$

「検査陽性であった場合に本当に疾病に罹患している確率が 81.8%である」の意味。

* カットオフ値について

参照) <http://note.chiebukuro.yahoo.co.jp/detail/n187856>

下図は、ROC 曲線 (Receiver Operating Characteristic curve: 受信者動作特性曲線) と呼び、縦軸 (上向き) が感度、横軸 (右向き) が偽陽性率 (= 1 - 特異度) である。カットオフ値を上げた時は左下へ、カットオフ値を下げた時は右上へプロットされ、

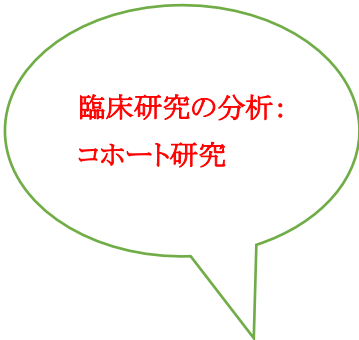


全体としては左上に凸の曲線を描く。この曲線は、感度を上げると特異度が下がる関係にある。ROC 曲線のグラフをたどると、感度と特異度のバランスの取れたカットオフ値を探ることができる。左下(感度=0, 偽陽性率=0)の点から、グラフをたどって右上(感度=1, 偽陽性率=1)の点へ進むと、最初は感度の上昇が急で、偽陽性率はあまり上昇し

ないが、接線の傾きが 1 をまたぐ変曲点を超えると、感度の上昇が鈍くなり、偽陽性率が急に上昇(特異度が急に低下)する。感度と特異度のバランスが取れたカットオフ値は、「接線の傾き=感度/偽陽性率=感度/(1-特異度)=尤度比」が 1 となる左上の点となる。検査が優秀かどうかは、「感度と特異度をともに上げられる」つまり、「より左へ強い曲率を持った曲線」かどうかで決まる。曲率の強い曲線の程度を「AUC (Area Under the Curve: 曲線下面積)」で表し、AUC の大きな検査が精度の高い検査になる。AUC=1 の時には左上の角点となり、AUC=0.5 の時には左下の角点から右上の角点にかけて右上がりの直線になる。この曲線の接線の傾き=感度/(1-特異度)を「尤度比」と呼ぶ。

相対危険度、相対リスク (relative risk) =リスク比 (risk ratio) : RR

	疾病あり	疾病なし	
暴露あり	a	b	a+b
暴露なし	c	d	c+d
	a+c	b+d	N



$$RR = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

相対危険度は暴露群の発生率(リスク)を非暴露群の発生率(リスク)で割ることにより求めることができ、暴露因子があると何倍危険か(暴露因子と疾病発生との関連の強さ)を示す指標となる。主にコホート研究で用いられる。

寄与危険度、寄与リスク (Attributable Risk)

=リスク差=絶対危険度、絶対リスク (Absolute Risk) : AR

暴露群の疾患発生率から非暴露群の疾患発生率を引いたもの。

$$AR = \frac{a}{a+b} - \frac{c}{c+d}$$

AR が 0.05⇒「暴露因子のために 100 人に 5 人疾病が発生している=100 人に介入すれば 5 人は予防できる。」と解釈できる。

「寄与危険割合(寄与リスクパーセント)AR%」

曝露因子により発生したと考えられる者"が"曝露群"を占める割合

$$AR\% = \frac{\frac{a}{a+b} - \frac{c}{c+d}}{\frac{a}{a+b}} \quad \text{リスク比を用いて} \quad AR\% = \frac{RR-1}{RR} \quad \text{でも計算できる。}$$

RR が 1 を上回る場合に用いられ、RR が 1 を下回る場合は「絶対リスク減少」

治療必要数 (number needed to treat:NNT):

$$NNT = \frac{1}{AR}$$

「対照となる治療ないし自然経過に加えて、その治療の 1 例の効果を観察するためには、その治療を何人の患者に用いなければならないか」という指標。1 人の患者を救うためにどれだけの費用と薬が必要なのかがより明確になる。NNT がマイナスのときは対照群より成績が悪いことを示す。

オッズ、オッズ比 →p133

$$\frac{\frac{a}{c}}{\frac{b}{d}} = \frac{\text{陽性の事後オッズ}}{\text{陰性の事後オッズ}} = \text{オッズ比}$$



case-control study では「リスク比 (risk ratio)」を計算できないため、「オッズ比 (odds ratio)」で代用。オッズ比は、ロジスティック回帰モデルでも使用される。

χ² 検定

独立性の検定 (ピアソン)

2×2 表では自由度 1、α=0.05 の棄却値 3.841

計算値 > 3.841 → H₀: 変数は独立であるは棄却され、H₁: 関係ありを採択



	肺癌診断	肺癌診断なし	合計
現在喫煙	60 (33.16)	300 (326.84)	360
現在喫煙なし	10 (36.84)	390 (363.16)	400
計	70	690	760

*(期待値) ex. 33.16=360×70/760

$$\chi^2 = \sum_{i=1, j=1}^{i=R, j=C} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2 = (60-33.06)^2/33.16 + (300-326.84)^2/326.84 + (10-36.84)^2/36.84 + (390-363.16)^2/363.16 = 45.45 > 3.841$$

データが離散的な場合イエーツの補正を使用することがある

$$\chi^2 = \sum_{i=1, j=1}^{i=R, j=C} \frac{(|O_{ij} - E_{ij}| - 0.5)^2}{E_{ij}}$$

χ^2 から計算できる相関統計量

φ係数: 二値変数の関連度尺度でピアソンの相関係数とほぼ同じ解釈でよい

$$\phi = \sqrt{\frac{\chi^2}{n}} \quad \text{ここでは } \phi = \sqrt{\frac{45.45}{760}} = 0.24$$

クラメル(Cramer)のV

(2×2より大きい表でのφの拡張、2×2ではφと同じ)

$$V = \sqrt{\frac{\chi^2}{n(k-1)}} \quad k \text{ は列数または行数の少ないほう}$$

Fisherの正確確率

	Event+	Event-	
A あたり玉	a X	b	a+b A
Non-A はずれ玉	c M-X	d	c+d N-A
計	a+c M	b+d	N N

- 薬物使用者(A) a+b 人、未使用者(Non-A) c+d 人、合計 N 人の中で、心臓死(event+)が a+c 人あり、うち薬物使用者(A) a 人であった。
- あたり玉 A 個、はずれ玉 N-A 個、合計 N 個の玉の入った箱がある。この箱から M 個取り出すとき、M 個の中であたり球 X のでる確率は？あたり球の個数 X は超幾何分布に従う。(M 個の玉を一つずつ取り出し箱に戻さない)

$$p = \frac{A^X \cdot N-A^{M-X}}{N^M} \rightarrow \text{変形すると } p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!a!b!c!d!}$$

Fisherの正確確率: $p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{N!a!b!c!d!}$ と同値である

例題 1)新薬を投与した 9 人と、対照薬を投与した 6 人について、副作用の有無を調べクロス集計表にまとめた。新薬の方が副作用が出やすいかをフィッシャーの直接確率を求めて、有意水準 5%で検定する。また、 χ^2 検定、リスク比なども求める。

薬効調査

	副作用		
	あり	なし	症例数
新薬	7	2	9
対照薬	1	5	6
計	8	7	15

フィッシャーの直接確率検定	
両側 P 値	0.0406*
片側 P 値	0.0350*
Cramer's V	0.6001
Yule's Q	0.8919

独立性の検定		
期待度数		
	あり	なし
新薬	4.800	4.200
対照薬	3.200	2.800

	カイ二乗値	自由度	P 値
補正なし	5.4018	1	0.0201*
Yates の補正	3.2254	1	0.0725
Cramer's V	0.6001		
Yule's Q	0.8919		

リスク比とオッズ比			
リスク			
A:新薬,あり	0.7778	B:新薬,なし	0.2222
C:対照薬,あり	0.1667	D:対照薬,なし	0.8333

			95%信頼区間	
			下限値	上限値
		値		
リスク差	A-C	0.6111	-	-
リスク比	A/C	4.6667	0.7539	28.8873
オッズ比	AD/BC	17.5000	1.2233	250.3572

例題 2) 例題 1) 同様に下表を分析せよ。

	肺癌診断	肺癌診断なし	合計
現在喫煙	60 (33.16)	300 (326.84)	360
現在喫煙なし	10 (36.84)	390 (363.16)	400
計	70	690	760

独立性の検定		
期待度数		
	肺癌診断	肺癌診断なし
現在喫煙	33.158	326.842
現在喫煙なし	36.842	363.158

	カイ二乗値	自由度	P 値
補正なし	45.4741	1	$p < 0.001^{**}$
Yates の補正	43.7958	1	$p < 0.001^{**}$
Cramer's V	0.2446		
Yule's Q	0.7727		

フィッシャーの直接確率検定	
両側 P 値	$p < 0.001^{**}$
片側 P 値	$p < 0.001^{**}$
Cramer's V	0.2446
Yule's Q	0.7727

リスク比とオッズ比			
リスク			
A:現在喫煙,肺癌診断	0.1667	B:現在喫煙,肺癌診断なし	0.8333
C:現在喫煙なし,肺癌診断	0.0250	D:現在喫煙なし,肺癌診断なし	0.9750

		95%信頼区間		
		値	下限値	上限値
リスク差	A-C	0.1417	-	-
リスク比	A/C	6.6667	3.4659	12.8232
オッズ比	AD/BC	7.8000	3.9273	15.4915

コーエンの κ 係数 (kappa coefficient) : 一致の尺度

	+	-
+	a	b
-	c	d



実測一致率 $P_o = (a+d)/(a+b+c+d)$

期待一致率 $P_e = (a+c)(a+b)/(a+b+c+d)^2 + (b+d)(c+d)/(a+b+c+d)^2$

$$K = \frac{P_o - P_e}{1 - P_e}$$

2 種類の方法で、複数の被検者が改善したかどうかを、「改善」「悪化」で判断した一致度を検討したとき、次のような結果となったとする。

		方法 A		
		改善	悪化	合計
方法 B	改善	59	6	65
	悪化	11	166	177
	合計	70	172	242

一般的には、2 つの方法の結果が一致した数の和を総数で除した

$(59 + 166) / 242 = 0.94$ が一致率となるが、この一致率は「偶然による一致率(期待一致率)」を勘案していない点で、「見かけ上の一致率(実測一致率)」といわれる。「偶然による一致率」は、「方法 A, B が偶然に『改善』という判断が一致する確率」と「方法 A, B が偶然に『悪化』という判断が一致する確率」の和で表される。方法 A, B が偶然に「改善」という判断が一致する確率:、方法 A が「改善」と判断する確率と方法 B が「改善」と判断する確率の積で求められる。 $(70 / 242) \times (65 / 242) = 0.08$

同様に、方法 A, B が偶然に「悪化」という判断が一致する確率は、

$(172 / 242) \times (177 / 242) = 0.52$

「偶然による一致率(期待一致率)」は、 $0.08 + 0.52 = 0.60$ となる。

$$K = \frac{P_o - P_e}{1 - P_e} = \frac{0.94 - 0.60}{1 - 0.60} = 0.85$$

κ 係数は、完全一致で 1、偶然の一致で 0、偶然より一致度が低いとき負の値となり、係数が高ければ高いほど一致度が高いと判定。

一般的な判断基準

0 ~ 0.40 : 低い一致 (poor agreement)

0.41 ~ 0.60 : 中等度の一致 (moderate)

0.61 ~ 0.80 : かなりの一致 (good to fair)

0.81 ~ : 高い一致 (excellent)

BellCurve for Excel: 集計表の作成と分析 → カッパ係数

カッパ係数							
観測度数							
	改善	悪化	合 計				
改善	59	6	65				
悪化	11	166	177				
合 計	70	172	242				
期待度数							
	改善	悪化					
改善	18.802	46.198					
悪化	51.198	125.802					
			95%信頼区間		* : P<0.05 ** : P<0.01		
重み付け	カッパ係数	標準誤差	下限値	上限値	統計量:z	両側P値	
なし	0.8255	0.0406	0.7458	0.9051	12.8580	p < 0.001	**
1次の重み	0.8255	0.0406	0.7458	0.9051	12.8580	p < 0.001	**
2次の重み	0.8255	0.0406	0.7458	0.9051	12.8580	p < 0.001	**
重み表	一致率 0.8255						
1次の重み				2次の重み			
	1.00	0.00			1.00	0.00	
改善	0.00	1.00		改善	0.00	1.00	

マクネマー(McNemar)検定

データが対(対応がある)の時の χ^2 検定

	CM 後死刑に賛成	CM 後死刑に反対	合計
CM 前死刑に賛成	15	25	40
CM 前死刑に反対	10	20	30
計	25	45	70

a	b
c	d

マクネマー検定は不一致率の分布に基づく $\chi^2 = \frac{(b-c)^2}{b+c}$

Yates の連続補正を使用する場合は $\chi^2 = \frac{(|b-c|-1)^2}{b+c}$

$$\chi^2 = \frac{(25-10)^2}{25+10} = \frac{225}{35} = 6.43 > 3.841, \text{ または } \chi^2 = \frac{(|25-10|-1)^2}{25+10} = \frac{196}{35} = 5.6 > 3.841$$

CM 前後で人々の意見が変わらなかった (H_0) は棄却された

BellCurve for Excel: 集計表の作成と分析 → カップ係数

マクネマー検定			
観測度数			
	CM後死刑に賛成	CM後死刑に反対	合計
CM前死刑に賛成	15	25	40
CM前死刑に反対	10	20	30
合計	25	45	70
マクネマー検定			
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01
5.6000	1	0.0180 *	

不一致率の χ^2 値が大きい(有意) ⇒
意見が変わらなかったという帰無仮説 (H_0)
は棄却された

m×n分割表(RC表)

分割表の構成により検定方法が異なる

エクセル統計(柳井久江著)p261-276より

1) 行・列ともに分類において順序関係がない場合

⇒ χ^2 独立性の検定

大学生男女 100 人に各自の社交性についてのアンケート調査を行い、兄弟構成をもとに集計して次のような結果が得られた。社交性と兄弟構成に関連があるかどうか危険率5%で検定する。

	一人子	長子	末子	中間の子
社交的	7	6	8	9
普通	8	13	13	22
非社交的	1	2	7	4

BellCurve for Excel: 集計表の作成と分析→独立性の検定

独立性の検定					
観測度数					
	一人子	長子	末子	中間の子	合計
社交的	7	6	8	9	30
普通	8	13	13	22	56
非社交的	1	2	7	4	14
合計	16	21	28	35	100
期待度数	一人子	長子	末子	中間の子	
社交的	4.800	6.300	8.400	10.500	
普通	8.960	11.760	15.680	19.600	
非社交的	2.240	2.940	3.920	4.900	
			太字：1未満 赤字：5未満		
独立性の検定					
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01		
5.8138	6	0.4444			
Cramer's V	0.1705				

有意差なし: 独立性
あり=関連なし

マン=ホイットニーのU検定							
基本統計量							
変数	n	平均順位	統計量:U				
新薬 A	30	33.133	371				
対照薬 B	30	27.867	529				
マン=ホイットニーのU検定							
統計数値表による検定			正規化検定			* : P<0.05 ** : P<0.01	
統計量:U	両側検定	E(U)	V(U)	連続性補正	統計量:z	両側P値	
371	※ 1	450	4252.119	なし	1.2115	0.2257	
				あり	1.2038	0.2287	
正確検定							
両側P値 * : P<0.05 ** : P<0.01							
※ データ数の合計が20を超えるため正確検定は出力できません。							
※1 同順位があるため統計数値表を利用できません。							

有意差なし

3) 行・列の一方に順序関係があり分類数 3 以上の場合

⇒Kruskal-Wallis test

男性の健康診断で血液型と飲酒傾向を調べたところ次のようなデータが得られ、集計すると次の分割表にまとめられた。血液型と飲酒傾向に関連があるかどうか危険率 5%で検定する。

	飲まない	時々飲む	よく飲む	毎日飲む
A	13	11	6	4
AB	3	5	2	0
B	5	3	8	4
O	12	11	4	3

飲酒傾向を4段階で表し、次のように数値を割り当てる。

- 1 飲まない
- 2 時々飲む(月8日以内)
- 3 よく飲む(週4日以内)
- 4 毎日飲む(週5日以上)

BellCurve for Excel: ノンパラメトリック検定→クララルカル=ウォリス検定と多重比較

(表を3列のデータ形式に変換)

行列の一方に順序（分類数3以上）					
	飲まない1	時々飲む2	よく飲む3	毎日飲む4	
A	13	11	6	4	
A B	3	5	2	0	
B	5	3	8	4	
O	12	11	4	3	
	A	AB	B	O	
	1	1	1	1	1
	1	1	1	1	1
	1	1	1	1	1
	1	2	1	1	1
	1	2	1	1	1
	1	2	2	1	1
	1	2	2	1	1
	1	2	2	1	1
	1	3	3	1	1
	1	3	3	1	1
	1		3	1	1
	1		3	2	2
	2		3	2	2
	2		3	2	2
	2		3	2	2
	2		4	2	2
	2		4	2	2
	2		4	2	2
	2			2	2
	2			2	2
	2			2	2
	2			3	3
	3			3	3
	3			3	3
	3			3	3
	3			4	4
	3			4	4
	3			4	4
	4				
	4				
	4				
	4				

クラスカル=ウォリス検定と多重比較				
水準	A	AB	B	O
n	34	10	20	30
平均順位	45.63	44.05	58.73	43.28
クラスカル=ウォリス検定				
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01	
4.8434	3	0.1836	有意差なし	
多重比較 : Scheffe				
水準1	水準2	カイ二乗値	P 値	* : P<0.05 ** : P<0.01
A	AB	0.0285	0.9987	
A	B	3.1768	0.3652	
A	O	0.1294	0.9881	
AB	B	2.1129	0.5493	
AB	O	0.0065	0.9999	
B	O	4.2110	0.2396	

4) 行・列二方向とも順序関係がある場合

⇒Spearman's correlation coefficient by rank test

(スピアマンの順位相関係数)

男性の健康診断の身長と体重から肥満度を BMI(Body Mass Index)を計算し、肥満度を次の4段階で評価して、20歳代、30歳代、40歳代の年齢別に集計し分割表にまとめた。年齢と肥満度の関連性を危険率5%で検定する。

	痩せ	普通	過体重	肥満
20歳代	3	24	6	2
30歳代	5	20	7	2
40歳代	3	12	6	6

痩せ	BMI<20	-1
普通	20<= BMI <24	0
過体重	24<= BMI <26.4	1
肥満	26.4<=BMI	2

BellCurve for Excel: 基本統計・相関→スピアマンの順位相関行列

(表の行列にそれぞれ順位をつけ2列のデータ形式に変換)

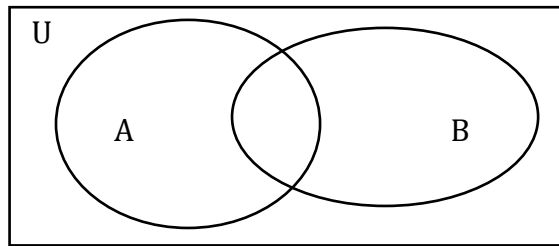
スピアマンの順位相関行列				
	年代	肥満度		
年代	1.0000	0.1441	相関係数 0.1441	
肥満度	0.1441	1.0000		
順位相関係数の検定 [上三角：P値/下三角：*,P<0.05 **,P<0.01]				
	年代	肥満度		
年代	-	0.1614	相関係数に有意性なし	
肥満度		-		
サンプルサイズ				
	年代	肥満度		
年代	96	96		
肥満度	96	96		

m×n分割表(RC表)では、行、列での
順序関係の有無に注意を払うこと

参考) 尤度と最尤推定法について

条件付き確率とベイズの定理

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$



$P(A)$: 事象 A が起きる確率 $= \frac{A}{U}$

$P(B|A)$: 事象 A が起きたもとで事象 B が起きる確率 $= \frac{(A \cap B)}{A}$ とすると

$\frac{(A \cap B)}{U} = \frac{A}{U} \cdot \frac{(A \cap B)}{A}$ つまり「A も B も起きる確率」= 「A が起きる確率」× 「A が起きたも

とで B が起きる(A の中での B の) 確率」

⇒ 「 $P(A \cap B) = P(A)P(B|A)$ 」

同様に B 側から考えると $\frac{(A \cap B)}{U} = \frac{B}{U} \cdot \frac{(A \cap B)}{B}$

⇒ $P(A \cap B) = P(B)P(A|B)$

よって恒等式 $P(A)P(B|A) = P(B)P(A|B) \Rightarrow P(B|A) = \frac{P(A|B)P(B)}{P(A)}$

$P(B)$: 事前確率 (prior probability)

$P(B|A)$: 事後確率 (posterior probability)

もともと B が起きる確率は $P(B)$ であったが A が起きるとい情報後(起きた後)は $P(B|A)$ になる。事象の確率という概念を採用し、必ずしも頻度に基づかない確率を尤度とする。

$P(A|B)$: 尤度(likelihood)

事象 A に関するデータが得られた時、それを反映し尤度 $P(A|B)$ の乗算によって、事象 B の確率は事前確率から事後確率に更新される。

→ PDF3 参照

最尤推定法 method of maximum likelihood estimation

与えられたデータからそれが従う確率分布の母数を点推定する方法。尤度最大化によって最尤推定値を計算する。そのモデルで定義される「尤度 (ゆうど)」を最大化させる母数 (parameter) 推定値を計算する → これが最尤推定値 (maximum likelihood estimate; MLE)

幾つかのデータ x_1, x_2, \dots, x_n が得られた時、データを散布図として作成すれば、個々のデータが分布する母関数の形が推定できる。この時母関数 $f(x)$ とすると、各データは同時に $f(x_1), f(x_2), \dots, f(x_n)$ を満たす (積になる) ことになる。ここで尤度関数 $L(\theta)$ (likelihood function)、ただし θ はその関数の母数。

$L(\theta) = \prod_1^n f(x_1) \cdot f(x_2) \cdots \cdot f(x_n)$ と定義され、これから最尤推定量を求める。

例えば、散布図やヒストグラムからそのデータがポアソン分布と考えられた時、データを x_1, x_2, \dots, x_n 、とすると、

Poisson 分布の確率密度関数は、 $p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$ であるから、

尤度関数 (データの得られる確率) は、 $\lambda = \theta$ (推定値) として、

$$L(\theta) = \prod_1^n p(x_n) = \prod_{i=1}^n e^{-\theta} \frac{\theta^{x_i}}{x_i!}$$

対数尤度関数は、

$$\log L(\theta) = (x_1 + x_2 + \cdots + x_n) * \log(\theta) - n\theta + \log(1/(x_1! * x_2! * \cdots * x_n!))$$

この最大値を求めるため、 θ で微分して、 $=0$ とする方程式をたてると、

$$\frac{\partial}{\partial \theta} \log L(\theta) = \frac{(x_1 + x_2 + \cdots + x_n)}{\theta} - n = 0$$

よって、 $\theta = \frac{(x_1 + x_2 + \cdots + x_n)}{n}$ となる。

(正規分布の場合)

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

尤度関数はあえて θ に置き換えず母数を推定値として

$$L(\mu, \sigma^2; x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu)^2\right]$$

$$\log L(\mu, \sigma^2; x) = \log\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu)^2\right]\right)$$

$$\begin{aligned}
& \log \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (x_i - \mu)^2 \right] \right) \\
&= \log \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} + \log \prod_{i=1}^n \exp \left[-\frac{1}{2\sigma^2} (x_i - \mu)^2 \right] \\
&= \log \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} + \log \prod_{i=1}^n e^{\left[\frac{-1}{2\sigma^2} (x_i - \mu)^2 \right]} \\
&= \log \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} + \log \prod_{i=1}^n e^{\left[\frac{-1}{2\sigma^2} \right]} e^{(x_i - \mu)^2} \\
&= -\frac{1}{2} \log \prod_{i=1}^n (2\pi\sigma^2) - \frac{1}{2\sigma^2} \log \prod_{i=1}^n e^{(x_i - \mu)^2} \\
&= -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2
\end{aligned}$$

パラメータを平均値としたので、 μ で微分すると μ に対して右辺第一項と第二項は定数だから微分すると 0 になる。よって尤度方程式からなくなり、第三項のみを考えると、

$$\frac{\partial}{\partial \mu} \log L(\hat{\mu}, \hat{\sigma}^2) = \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu}) = 0$$

つまりこの式が成立するのは、 $\hat{\mu} = \bar{x}$ のとき、つまり標本集団の平均値のパラメータは母集団の平均値に一致することになり最尤推定値は平均値に一致した。分散に関しては同様に σ^2 で微分すると以下。

$$\begin{aligned}
\frac{\partial}{\partial \sigma^2} \log L(\hat{\mu}, \hat{\sigma}^2) &= -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\mu})^2 = 0 \\
\hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2
\end{aligned}$$

交絡因子 Confounding factor

<例>

1. 血圧の高い人は(年齢が高く)給料が高い
2. ある先生の手術は(病期のよいひとだけするので)再発が少ない

交絡因子の制御方法

研究のデザインと実施時に調整

1. 対象集団の限定 restriction :性、年齢、嗜好、地域などによって調査対象を限定する。
2. マッチング matching :主に症例対照研究やコホート研究などで行われる。比較する2群において交絡因子になり得る年齢、性などの分布を比較し条件を同等にする。個人マッチングと群マッチングがある。
3. 無作為化 randomization 介入研究で行われる。介入群と非介入群を完全に無作為に割り付けることで、既知・未知の交絡因子すべてが両群で均等となる。

解析段階での交絡因子の調整

1. 層化 stratification :交絡因子の層(カテゴリー)ごとに解析する。
例)10歳ごとに年代を区切って解析するなど
2. 標準化 standardization :年齢構成の異なる集団間で比較しなければならないときに各層の構成割合を調整した上で、合計して観察する。
3. 多変量解析 multivariate analysis :結果(疾病や健康事象)を「従属変数」曝露を「独立変数」として各要因を統計モデルに組み込んで要因間の影響を取り除き、各要因の単独での影響を評価する方法。

交絡因子の調整

- 1) 研究のデザインと実施時に調整(無作為化など)
- 2) 解析時点で調整(疫学調査など)

説明変数、交絡因子ともに計量データ:

共分散分析(analysis of covariance ANCOVA)

説明変数、交絡因子ともに計数データ:

Mantel-Haenszel 法(2×2分割表に対し)

Mantel-extension 法(2×k分割表に対し)

目的変数が計数データ:

ロジスティック回帰など

層別 2×2 分割表に対するマンテル・ヘンツェル法 (Mantel-Haenszel test)

交絡因子により層別を行い、各層でのオッズ比を計算し、各層のオッズ比は均一であることを前提に要約オッズ比(共通オッズ比の推定量)を計算し Breslow-Day 検定によりその均一性を証明し共通オッズ比と確定させる。

マンテル・ヘンツェル共通オッズ比 OR_{MH} = 調整相対危険度 RR_{MH}

	第 1 階層			第 2 階層						第 k 階層		
	患者	対照	計	患者	対照	計	患者	対照	計	患者	対照	計
要因あり	a ₁	b ₁	M ₁₁	a ₂	b ₂	M ₁₂	⋯	⋯	⋯	a _k	b _k	M _{1k}
要因なし	c ₁	d ₁	M ₀₁	c ₂	d ₂	M ₀₂	⋯	⋯	⋯	c _k	d _k	M _{0k}
計	N ₁₁	N ₀₁	T ₁	N ₁₂	N ₀₂	T ₂	⋯	⋯	⋯	N _{1k}	N _{0k}	T _k

$$OR_{MH}(RR_{MH}) = \frac{\sum_{i=1}^k \frac{a_i d_i}{n_i}}{\sum_{i=1}^k \frac{b_i c_i}{n_i}}$$

を計算し(厳密には要約オッズ比)

有意性検定

$H_0: OR_{MH}(RR_{MH}) = 1$ $H_1: OR_{MH}(RR_{MH}) \neq 1$ については

$$\chi_{MH} = \frac{\sum_{i=1}^k a_i - \sum_{i=1}^k \frac{M_{1i} N_{1i}}{T_i}}{\sqrt{\sum_{i=1}^k \frac{M_{1i} M_{0i} N_{1i} N_{0i}}{T_i^2 (T_i - 1)}}$$

が標準正規分布に従うことを利用する

$OR_{MH}(RR_{MH})$ の $100(1-\alpha)\%$ 信頼区間は

上限 $OR_{MH}^{(1+Z(\frac{\alpha}{2})/\chi_{MH})}$ 、下限 $OR_{MH}^{(1-Z(\frac{\alpha}{2})/\chi_{MH})}$ で計算できる

	40-49 歳			50-59 歳			60 歳以上			全体		
	患者	対照	計	患者	対照	計	患者	対照	計	患者	対照	計
5 カップ以上/日 飲む	7	19	26	11	14	25	10	15	25	28	48	76
飲まない	2	16	18	5	19	24	4	21	25	11	56	67
計	9	35	44	16	33	49	14	36	50	39	104	143
各年齢層での オッズ比(≒RR)	2.95			2.99			3.5					

丹後俊郎：新版医学への統計学 p 247 より

共通オッズ比 OR_{MH} (調整相対危険度 RR_{MH})

$$OR_{MH} (RR_{MH}) = \frac{\frac{7 \times 16}{44} + \frac{11 \times 19}{49} + \frac{10 \times 21}{50}}{\frac{19 \times 2}{44} + \frac{14 \times 5}{49} + \frac{15 \times 4}{50}} = 3.15$$

$$\chi_{MH} = \frac{28 - \left(\frac{26 \times 9}{44} + \frac{25 \times 16}{49} + \frac{25 \times 14}{50} \right)}{\sqrt{\frac{26 \times 18 \times 9 \times 35}{44^2 \times 43} + \frac{25 \times 24 \times 16 \times 33}{49^2 \times 48} + \frac{25 \times 25 \times 14 \times 36}{50^2 \times 49}}} = 2.82 > Z(\alpha=0.01/2) = 2.58$$

実際に表を見ながら
数字を入れると計算
の仕方は意外と単純

また

$$OR_{MH} \text{ の } 100(1-\alpha)\% \text{ 信頼区間の上限} = 3.15^{1 + \frac{1.96}{2.82}} = 6.99$$

$$OR_{MH} \text{ の } 100(1-\alpha)\% \text{ 信頼区間の下限} = 3.15^{1 - \frac{1.96}{2.82}} = 1.42$$

BellCurve for Excel での解析では下記の形式にする

		患者	対照
40-49 歳	5 カップ以上/日飲む	7	19
	飲まない	2	16
50-59 歳	5 カップ以上/日飲む	11	14
	飲まない	5	19
60 歳以上	5 カップ以上/日飲む	10	15
	飲まない	4	21

BellCurve for Excel: 集計表の作成と分析 → コクラン=マンテル=ヘンツェル検定

(コクラン=)マンテル=ヘンツェル検定												
	40-49歳			50-59歳			60歳以上			全体		
	患者	対照	計	患者	対照	計	患者	対照	計	患者	対照	計
5カップ以上/日飲む	7	19	26	11	14	25	10	15	25	28	48	76
飲まない	2	16	18	5	19	24	4	21	25	11	56	67
計	9	35	44	16	33	49	14	36	50	39	104	143
			患者	対照								
40-49歳	5カップ以上/日飲む		7	19								
	飲まない		2	16								
50-59歳	5カップ以上/日飲む		11	14								
	飲まない		5	19								
60歳以上	5カップ以上/日飲む		10	15								
	飲まない		4	21								

左表にh整理し を選択

コクラン=マンテル=ヘンツェル検定									
観測度数				層別の独立性の検定					
		患者	対照			自由度	値	P 値	
40-49歳	5カップ以上/日飲む	7	19	40-49歳	Pearsonのカイ二乗	1	1.6344	0.2011	
	飲まない	2	16		尤度比カイ二乗	1	1.7367	0.1876	
					MHのカイ二乗	1	1.5973	0.2063	
					Fisherの直接確率			0.2704	
					Cramer's V			0.1927	
		患者	対照			自由度	値	P 値	
50-59歳	5カップ以上/日飲む	11	14	50-59歳	Pearsonのカイ二乗	1	2.9884	0.0839	
	飲まない	5	19		尤度比カイ二乗	1	3.0460	0.0809	
					MHのカイ二乗	1	2.9274	0.0871	
					Fisherの直接確率			0.1284	
					Cramer's V			0.2470	
		患者	対照			自由度	値	P 値	
60歳以上	5カップ以上/日飲む	10	15	60歳以上	Pearsonのカイ二乗	1	3.5714	0.0588	
	飲まない	4	21		尤度比カイ二乗	1	3.6613	0.0557	
					MHのカイ二乗	1	3.5000	0.0614	
					Fisherの直接確率			0.1137	
					Cramer's V			0.2673	

↑ 層別の独立性の検定(手順①: χ^2 、Fisher の検定などで $p \geq 0.05$ を確認)
 年齢で層別したそれぞれの結果について、4種類の独立性の検定を行い、いずれの年齢層においても、独立性は棄却されず、コーヒーと膵臓癌に関連はみられない。

層別のリスク比・オッズ比				
		値	95%信頼区間	
			下限値	上限値
40-49歳	ケースコントロール(オッズ比)	2.9474	0.5350	16.2372
	コーホート(列1のリスク比)	2.4231	0.5672	10.3507
	コーホート(列2のリスク比)	0.8221	0.6184	1.0930
			95%信頼区間	
50-59歳	ケースコントロール(オッズ比)	2.9857	0.8447	10.5530
	コーホート(列1のリスク比)	2.1120	0.8616	5.1768
	コーホート(列2のリスク比)	0.7074	0.4725	1.0590
			95%信頼区間	
60歳以上	ケースコントロール(オッズ比)	3.5000	0.9206	13.3065
	コーホート(列1のリスク比)	2.5000	0.9029	6.9221
	コーホート(列2のリスク比)	0.7143	0.4969	1.0268

↑ 層別のリスク比・オッズ比(手順②:各層のオッズ比の95%CIが1を含むこと確認)
 検定の結果いずれも有意ではなかったため、オッズ比の95%信頼区間も、1を含んだ結果になっている。

オッズ比の等質性検定		自由度	値	P 値	*: P<0.05 **: P<0.01
Breslow-Day		2	0.0367	0.9818	
条件付き独立性の検定		自由度	値	P 値	*: P<0.05 **: P<0.01
CMHの相関統計量		1	7.9729	0.0047	**

↑ オッズ比の等質性の検定 (Breslow-Day)

(手順③: 要約オッズ比を計算し Breslow-Day 検定するが手計算は困難、Breslow-Day 検定で等質性が証明されれば要約オッズ比 = 共通オッズ比と確定)

↑ 条件付き独立性の検定

(手順④: 条件付き独立検定で p 値 < 0.05 であれば層別した因子は関連あり = 共通オッズ比が有効と判断)

オッズ比の等質性の検定 (Breslow-Day の検定): 各層におけるオッズ比は全て等しいという帰無仮説を検定、P 値は 0.9818 であり、オッズ比は均一
各年齢層で調整を行ったクロス集計表において条件付き独立性の検定 P 値は 0.0047 であり独立性は棄却される。

年齢調節後の独立性は棄却 = 関連がある

調整済みリスク比・オッズ比推定		値	95%信頼区間	
			下限値	上限値
ケースコントロール(共通オッズ比)	MHの推定値	3.1530	1.4043	7.0788
	ロジット推定量	3.1554	1.4064	7.0795
コーホート(列1の共通リスク比)	MHの推定値	2.3115	1.2521	4.2673
	ロジット推定量	2.2992	1.2486	4.2338
コーホート(列2の共通リスク比)	MHの推定値	0.7464	0.6090	0.9148
	ロジット推定量	0.7617	0.6262	0.9265

共通オッズ比 ↑

Breslow-Day 検定

多層の 2 行 × 2 列クロス集計から、層別要因の影響を調整したオッズ比を推定する際に、オッズ比が全ての層で等質かどうかを検定する方法。 $OR_{MH}(RR_{MH})$ は厳密には共通オッズ比の推定量で要約オッズ比という。要約オッズ比を利用し検定を行い(計算は複雑、PC で計算された結果をみて) Breslow-Day の p 値 ≥ 0.05 のときオッズ比の均一性は否定できない(=Mantel-Haenszel 検定の妥当性が証明され共通オッズ比と判断できる。)

条件付き独立性の検定 (ペイズの定理を基礎に数学的証明が必要、結果の判断ができれば十分)

*層別をしないで解析すると関連が見られるが、層別をして解析すると関連が見られなくなるような現象はシンプソンのパラドックス(条件付き独立)。 ⇒ p163

2 つの変数間の関係について、3 つめの変数で層別した場合に関係が見られなくなるような場合、最初の 2 つの変数は条件付き独立であるという。3 つめの変数で層別された 2 x 2 表について、条件付き独立であることを帰無仮説とした検定で p 値 < 0.05 であれば条件付き独立性が棄却され(=層別した因子は関連あり) 関連づけることで共通オッズ比が有効と判断できる。

コクラン・アーミテージ検定と拡張マンテル検定

コーヒーを飲む習慣のパターンと膵臓癌との患者-対照群調査を解析するとき

		患者	対照
40代	0杯	2	16
	1-2杯	9	47
	3-4杯	9	24
	5杯以上	7	19
50代	0杯	5	19
	1-2杯	21	50
	3-4杯	22	25
	5杯以上	11	14
60歳以上	0杯	4	21
	1-2杯	22	55
	3-4杯	22	31
	5杯以上	10	15

丹後俊郎：新版医学への統計学 p 250 より

Mantel-haenszel 法は各層を 2×2 分割表に作り直し、複数の分割表を統合できる

		患者	対照
40-49歳	5カップ以上/日飲む	7	19
	飲まない	2	16
50-59歳	5カップ以上/日飲む	11	14
	飲まない	5	19
60歳以上	5カップ以上/日飲む	10	15
	飲まない	4	21

Cochran-Armitage test は1層のみを取り上げ、各群に傾向性があるか否かをみる

		患者	対照
40代	0杯	2	16
	1-2杯	9	47
	3-4杯	9	24
	5杯以上	7	19

拡張マンテル法 (Mantel-haenszel-trend 分析) は多層のままで傾向性を分析する

コクラン・アーミテージ検定 Cochran-Armitage (trend) test

因子への暴露水準 j	得点 m_j	患者 r_j	対照	オッズ比	計 n_j
1	0	r_1	$n_1 - r_1$	1.00	n_1
2	1	r_2	$n_2 - r_2$	OR_2	n_2
...	...				
j	$j-1$				
a	$a-1$	r_a	$n_a - r_a$	OR_a	n_a
計		r	$T-r$		T

a 個の暴露水準の順に自然な順序があり、水準 j ($j=1\sim a$) に含まれる個体数を n_j 、反応した個体数 r_j 、暴露水準 j の暴露水準1に対するオッズ比 OR_j が暴露水準が上がるごとに増加するという仮説

片側検定の場合 $H_0: OR_1=1=OR_2\cdots=OR_a$

$H_1: OR_1=1 \leq OR_2 \leq \cdots \leq OR_a$ (上昇傾向の場合) を検定する。

各水準 j ($j=1\sim a$) に得点 m_j を与え、統計量を計算する。

反応の合計得点 $O = \sum_{j=1}^a r_j m_j$ 、期待値 $E = \frac{r \sum_{j=1}^a n_j m_j}{T}$ 、

分散 $V = \frac{r(T-r)}{T^2(T-1)} * \{T(\sum_{j=1}^a n_j m_j^2) - (\sum_{j=1}^a n_j m_j)^2\}$ を計算して

$Z = \frac{O-E}{\sqrt{V}}$ が標準正規分布に従う または、 $Z^2 = \frac{(O-E)^2}{V}$ が $\chi_1^2(\alpha)$ に従うことにより検定

★ 拡張マンテル法 (Mantel-haenszel-trend 分析) で1層のみと考えた分析法に相当する。得点の与え方を0から開始し、 $m_j = j - 1$ が計算しやすくなる。他に、群の順位

R_j を利用して $m_j = \frac{R_j}{T}$ とする方法、 ($R_a = n_1 + n_2 + \cdots + n_{a-1} + \frac{n_a+1}{2}$) や、もともと

の特性値 m_j があればそれを利用などがある。 → p106参照

上記の40代のみデータを上げる。 $m_j = j - 1$ とする

水準 j	コーヒー	患者 r_j	対照	計 n_j	得点 m_j	$n_j \times m_j$	$n_j \times m_j^2$	$r_j \times m_j$
1	0杯	2	16	18	0	0	0	0
2	1-2杯	9	47	56	1	56	56	9
3	3-4杯	9	24	33	2	66	132	18
4	5杯以上	7	19	26	3	78	234	21
合計		27= r	106	133= T		200= $\sum_{j=1}^a n_j m_j$	422	48= O

$$O=48 \quad E = \frac{27}{133} * 200 = 40.6 \quad V = \frac{27*106}{133^2*132} (133*422 - 200^2) = 19.8$$

$$Z^2 = \frac{(48-40.6)^2}{19.8} = 2.77 < 3.84 (= \chi_1^2(0.05))、用量反応関係は有意ではない$$

BellCurve for Excel: 集計表の作成と分析→コクラン=アーミテージ検定

コクラン=アーミテージ検定			
観測度数			
	患者	対照	
0	2	16	
1	9	47	
2	9	24	
3	7	19	
コクラン=アーミテージ検定 * : P<0.05 ** : P<0.01			
統計量:z	1.7221		
両側P値	0.0850		
片側P値	0.0425 *		
Cramer's V	0.1564		

z 値 1.722 (< 1.96)
 $(z^2 = \chi_1^2 = 2.97 < 3.84)$
 は得点の与え方により、上の計算値と少し値が違うが、同じ解析結果、両側 p 値では有意差なし

拡張マンテル検定で多層の分析 (Mantel-haenszel-trend 分析)

多層 (k層) のうちの一層 G_i ($i=1\sim k$) での表 (水準内データ数 a)

層 G_i	患者 r_{ij}	非患者	計 n_{ij}	得点 m_{ij}	$n_{ij} * m_{ij}$	$n_{ij} * (m_{ij})^2$	$r_{ij} * m_{ij}$
G_{i1}	r_{i1}	$n_{i1} - r_{i1}$	n_{i1}	m_{i1}	$n_{i1} * m_{i1}$	$n_{i1} * (m_{i1})^2$	$r_{i1} * m_{i1}$
G_{i2}	r_{i2}	$n_{i2} - r_{i2}$	n_{i2}	m_{i2}	$n_{i2} * m_{i2}$	$n_{i2} * (m_{i2})^2$	$r_{i2} * m_{i2}$
G_{i3}	r_{i3}	$n_{i3} - r_{i3}$	n_{i3}	m_{i3}	$n_{i3} * m_{i3}$	$n_{i3} * (m_{i3})^2$	$r_{i3} * m_{i3}$
..				..			
G_{ij}	r_{ij}	$n_{ij} - r_{ij}$	n_{ij}	m_{ij}	$n_{ij} * m_{ij}$	$n_{ij} * (m_{ij})^2$	$r_{ij} * (m_{ij})$
..				..			
G_{ia}	r_{ia}	$n_{ia} - r_{ia}$	n_{ia}	m_{ia}	$n_{ia} * m_{ia}$	$n_{ia} * (m_{ia})^2$	$r_{ia} * (m_{ia})$
計	r_i	$T_i - r_i$	T_i		$\sum (n_{ij} * m_{ij})$	$\sum (n_{ij} * (m_{ij})^2)$	$\sum (r_{ij} * m_{ij})$

層 G_i での反応の合計得点 $O_i = \sum_{j=1}^{(i)} r_j m_j$ 、期待値 $E_i = \frac{r_i \sum_{j=1}^{(i)} n_j m_j}{T_i}$

分散 $V_i = \frac{r_i(T_i - r_i)}{T_i^2(T_i - 1)} * \left\{ T_i \left(\sum_{j=1}^{(i)} n_j m_j^2 \right) - \left(\sum_{j=1}^{(i)} n_j m_j \right)^2 \right\}$ とすると

統計量: 暴露水準が a 個ある ($j=1\sim a$)、k 個の層 $G_1 \sim G_k$ において

$$Z = \chi_{ME} = \frac{(O_1 - E_1) + (O_2 - E_2) + \dots + (O_k - E_k)}{\sqrt{V_1 + V_2 + \dots + V_k}} \quad \dots \text{式(*) を計算する}$$

片側検定有意

H_0 を棄却 (上昇傾向) : $Z > Z(\alpha/2)$ 、 H_0 を棄却 (下降傾向) : $Z < -Z(\alpha/2)$

両側検定では $Z^2 > \chi_1^2(\alpha)$ と同等 (有意)

BellCurve for Excel: 集計表の作成と分析→拡張マンテル検定

拡張マンテル検定				
コーヒーを飲む習慣のパターンと膵臓癌との患者-対照群調査				
		患者	対照	計
40代	0杯	2	16	18
	1-2杯	9	47	56
	3-4杯	9	24	33
	5杯以上	7	19	26
50代	0杯	5	19	24
	1-2杯	21	50	71
	3-4杯	22	25	47
	5杯以上	11	14	25
60歳以上	0杯	4	21	25
	1-2杯	22	55	77
	3-4杯	22	31	53
	5杯以上	10	15	25
	計	144	336	480
[1層あたりの水準数] に「4」				

拡張マンテル検定				層別の独立性の検定				
				* : P<0.05 ** : P<0.01				
観測度数		患者	対照		自由度	値	P 値	
40代	0杯	2	16	40代	カイ二乗	3	3.2548	0.3540
	1-2杯	9	47		尤度比カイ二乗	3	3.3127	0.3459
	3-4杯	9	24		MHのカイ二乗	1	2.9435	0.0862
	5杯以上	7	19		Cramer's V		0.1564	
		患者	対照		自由度	値	P 値	
50代	0杯	5	19	50代	カイ二乗	3	6.7687	0.0796
	1-2杯	21	50		尤度比カイ二乗	3	6.8676	0.0762
	3-4杯	22	25		MHのカイ二乗	1	5.9573	0.0147 *
	5杯以上	11	14		Cramer's V		0.2013	
		患者	対照		自由度	値	P 値	
60歳以上	0杯	4	21	60歳以上	カイ二乗	3	6.2680	0.0993
	1-2杯	22	55		尤度比カイ二乗	3	6.5665	0.0871
	3-4杯	22	31		MHのカイ二乗	1	5.5893	0.0181 *
	5杯以上	10	15		Cramer's V		0.1866	
				条件付き独立性の検定				
					自由度	値	P 値	
					CMHの相関統計量	1	14.4792	p < 0.001 **

層別の独立性の検定：年齢で層別したそれぞれの結果について、3種類の独立性の検定を行う。40代は独立性が棄却されず、コーヒーを飲む量と膵臓癌には関連が認められなかったが、50代と60歳以上では、MHのカイ二乗で独立性は棄却された。
 条件付き独立性の検定：年齢の影響を調整して独立性の検定を行った結果 P 値 = 0.0001 となり P < 0.05。したがって、年齢を調整すると独立性が棄却され、コーヒーを飲む量と膵臓癌の有無には関連が認められる。

→ CMH 相関統計量 $\chi_{extended-MH}^2 = \left\{ \frac{(\sum_i S_{xy}^{(i)})^2}{\sum_i (S_{xx}^{(i)} S_{yy}^{(i)})} \right\} Z = \chi_{ME} = \frac{(O_1 - E_1) + (O_2 - E_2) + \dots + (O_k - E_k)}{\sqrt{V_1 + V_2 + \dots + V_k}}$ に関連あり

CMH(Cochran-Mantel-Haenszel)統計量について

p150表の単一分割表Group i (i群)において

x軸:j方向の分布、y軸:患者1、対照0で相関を考えると

$$\sum_j x_{ij} = \sum_j m_{ij} n_{ij} \quad , \quad \sum_j x_{ij}^2 = \sum_j m_{ij}^2 n_{ij}$$

$$\sum_j y_{ij} = \sum_j (0 * (n_{ij} - r_{ij}) + 1 * r_{ij}) = r_i$$

$$\sum_j y_{ij}^2 = \sum_j (0^2 * (n_{ij} - r_{ij}) + 1^2 * r_{ij}) = r_i$$

$$\sum_i x_{ij} y_{ij} = \sum_i (0 * m_{ij} (n_{ij} - r_{ij}) + 1 * m_{ij} r_{ij}) = \sum_j m_{ij} r_{ij}$$

$$S_{xx}^{(i)} = \sum_j x_{ij}^2 - \frac{(\sum_j x_{ij})^2}{T_i} = \frac{T_i \sum_j m_{ij}^2 n_{ij} - (\sum_j m_{ij} n_{ij})^2}{T_i}$$

$$S_{yy}^{(i)} = \sum_j y_{ij}^2 - \frac{(\sum_j y_{ij})^2}{T_i} = \frac{r_i(T_i - r_i)}{T_i}$$

$$S_{xy}^{(i)} = \sum_j x_{ij} y_{ij} - \frac{\sum_j x_{ij} \sum_j y_{ij}}{T_i} = \frac{T_i \sum_j m_{ij} r_{ij} - r_i \sum_j m_{ij} n_{ij}}{T_i}$$

複数分割表から得られるCMH相関統計量: $Z^2 = \chi_{extended-MH}^2$ は

$$\chi_{extended-MH}^2 = \left\{ \frac{(\sum_i S_{xy}^{(i)})^2}{\sum_i \left(\frac{1}{T_i} S_{xx}^{(i)} S_{yy}^{(i)} \right)} \right\} \text{ で与えられる}$$

ここで分割表が1つの時Tの代わりにT-1を採用した場合がコクランアーミテージ検定

$$\chi_{extended-MH}^2 = \left\{ \frac{(S_{xy})^2}{\frac{1}{T-1} S_{xx} S_{yy}} \right\} = (T-1)R^2 = \chi_{Cochran-Armitage}^2 \quad R: \text{相関係数}$$

$$\chi_{extended-MH}^2 = \left\{ \frac{(S_{xy})^2}{\frac{1}{T} S_{xx} S_{yy}} \right\} = TR^2 = \chi_{trend}^2 \quad \text{統計量となる}$$

SAS における CMH(Cochran-Mantel-Haenszel)統計量について

- カテゴリーに順序がない RC 表での CMH 統計量:CMH Type3
 χ^2 独立性の検定の結果にほぼ一致する。
- 一方のカテゴリーに順序がある RC 表での CMH 統計量:CMH Type2
「用量・反応関係」を調べるとして得点として 1,2,...,k と自然数を割り当てて Mantel extension 法,Cochran-Armitage 法により検定した統計量。得点として与えるスコアは、カテゴリ変数が文字型の場合には自然数、数値型の場合にはその変数の値。自然数ではなくウィルコクソン・スコアを割り当てた CMH 統計量 Type2 は、マンホイットニー、クラスカル・ワリス検定と等価になる。
- 両方のカテゴリーに順序がある RC 表での CMH 統計量:CMH Type1
行変数・列変数ともにウィルコクソン・スコアを割り当てる。CMH 統計量 Type1 は、周辺スコアを与えたときのピアソンの相関係数の自乗×(N-1)に一致する。

シンプソンのパラドックス (Simpson's Paradox)

新薬の旧薬に対する生存率の優位性を考えてみる。

社会学者の研究メモ <http://jtsutsui.hatenablog.com/entry/20100125/1264353773>

	生存	死亡
新薬	100	900
旧薬	500	600

(表の一部改変)

BellCurve for Excel: 集計表の作成と分析→リスク比とオッズ比

リスク比とオッズ比				
観測度数	生存	死亡	合計	
新薬	100	900	1000	
旧薬	500	600	1100	
合計	600	1500	2100	
リスク比とオッズ比				
リスク				
A:新薬,生存	0.1000	B:新薬,死亡	0.9000	
C:旧薬,生存	0.4545	D:旧薬,死亡	0.5455	
			95%信頼区間	
		値	下限値	上限値
リスク差	A-C	-0.3545	-	-
リスク比	A/C	0.2200	0.1807	0.2679
オッズ	新薬の生存オッズ = 100/900 = 0.11、旧薬の生存オッズ = 500/600 = 0.83			
オッズ比	AD/BC = 0.11/0.83	0.1333	0.1051	0.1692

これだけみると、旧薬生存のオッズは 0.83 (0.5448/0.5455)、新薬生存のオッズは 0.11 (0.100/0.900)、オッズ比 0.13 で新薬の方が圧倒的に死亡しやすい。ここで第 3 の変数である性別を導入し、層別分割表を書いて以下とする

		生存	死亡
女	新薬	88	890
	旧薬	5	100
男	新薬	12	10
	旧薬	495	500

BellCurve for Excel: 集計表の作成と分析→コクラン=マンテルヘンツェル

コクラン=マンテル=ヘンツェル検定				層別の独立性の検定			
観測度数	生存	死亡		自由度	値	P 値	
女	新薬	88	890	Pearsonのカイ二乗	1	2.1675	0.1410
	旧薬	5	100	尤度比カイ二乗	1	2.5099	0.1131
				MHのカイ二乗	1	2.1655	0.1411
				Fisherの直接確率			0.1968
				Cramer's V		0.0447	
男	新薬	12	10	Pearsonのカイ二乗	1	0.1981	0.6563
	旧薬	495	500	尤度比カイ二乗	1	0.1983	0.6561
				MHのカイ二乗	1	0.1979	0.6564
				Fisherの直接確率			0.6735
				Cramer's V		0.0140	

層別のリスク比・オッズ比				
		95%信頼区間		
女		値	下限値	上限値
	ケースコントロール(オッズ比)	1.9775	0.7846	4.9845
	コーホート(列1のリスク比)	1.8896	0.7851	4.5479
	コーホート(列2のリスク比)	0.9555	0.9116	1.0016
		95%信頼区間		
男		値	下限値	上限値
	ケースコントロール(オッズ比)	1.2121	0.5189	2.8313
	コーホート(列1のリスク比)	1.0964	0.7449	1.6138
	コーホート(列2のリスク比)	0.9045	0.5699	1.4356

オッズ比の等質性検定		自由度	値	P 値	
	Breslow-Day		1	0.5964	
条件付き独立性の検定		自由度	値	P 値	
	CMHの相関統計量		1	1.9861	
調整済みリスク比・オッズ比推定					
		95%信頼区間			
			値	下限値	上限値
	ケースコントロール(共通オッズ比)	MHの推定値	1.5625	0.8458	2.8866
		ロジット推定量	1.5161	0.8114	2.8326
	コーホート(列1の共通リスク比)	MHの推定値	1.3317	0.8958	1.9795
		ロジット推定量	1.1977	0.8408	1.7060
	コーホート(列2の共通リスク比)	MHの推定値	0.9501	0.8911	1.0129
		ロジット推定量	0.9550	0.9113	1.0008

女の新薬生存のオッズ比は 1.98、男のオッズ比は 1.21 となり共通オッズ比は 1.5625 (0.846~2.887)となる。男女の区別なくデータを見たときは新薬が不利だったが、男女別にすると新薬が有利になるため、パラドックスと呼ばれている。男女で薬の効き方の方向は同じ(プラス)だが、全体的に死亡率のレベルが異なり(旧薬でも新薬でも男性の方が生存率が高い)、そして女性は新薬被験者が多く、男性は旧薬被験者が多く、死亡率の高い女性が新薬を多くとっていたため、調整変数である性別を考慮した場合と無視した場合とで、逆の有意な差が見られる(シンプソンのパラドックス)。解決するためには層別に分けコクラン・マンテルヘンツェル法(または、ロジスティック回帰分析)で分析。⇒p154, 170

重回帰分析 multiple regression analysis

**最小二乗法の原理を理解するには二次関数をイメージする。例えば
 $y = x^2 + 2x + 2$ の場合、平方完成すれば $y = (x + 1)^2 + 1$ により $x = -1$ で最小値となることがわかる。二次関数では接線の傾きが頂点で 0 になるので微分して $y' = 2x + 2$ 、 $y' = 0$ のとき同様に $x = -1$ で最小値をとることがわかる

重回帰分析においても係数の決め方は上記の原理のもと単回帰での方法と同じ。ある観測値 $(x_{i1}, x_{i2}, \dots, x_{in}, y_i)$ の時、つまり実際のデータが y_i のとき回帰式での予測値は $b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_nx_{in}$ 、その差 $d = y_i - (b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_nx_{in})$ を残差 (residual) と呼び、この残差の2乗の総和 $Q = \sum d^2 = \sum \{y_i - (b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_nx_{in})\}^2$ が最小になるように、 $b_0, b_1, b_2, \dots, b_n$ を決定する。つまり $Q = \sum \{y_i - (b_0 + b_1x_{i1} + b_2x_{i2} + \dots + b_nx_{in})\}^2$ を $b_0, b_1, b_2, \dots, b_n$ で偏微分し $= 0$ とおいた $n+1$ 個の連立方程式を解く。

→ 重回帰式 $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$

例)

球速	遠投	懸垂	握力
120	62	71	80
130	90	32	79
95	53	20	32
141	88	28	68
128	79	45	50
118	88	30	70
108	83	22	56
87	53	40	38
120	92	15	62
110	70	10	43
100	70	10	78
127	82	31	76
112	87	13	47
130	63	18	65
99	79	15	72
102	69	22	50
130	77	28	48
110	78	17	39
120	68	33	43
128	81	42	52

解析(係数の決め方)の基本原理
 は最小二乗法、偏微分して連立
 方程式

BellCurve for Excel: を選択し多変量解析→重回帰分析

(目的変数→球速、説明変数→遠投、懸垂、握力指定)

基本統計量						
変数	n	平均	不偏分散	標準偏差	最小値	最大値
遠投	20	75.600	136.779	11.695	53.000	92.000
懸垂	20	27.100	214.937	14.661	10.000	71.000
握力	20	57.400	232.779	15.257	32.000	80.000
球速	20	115.750	202.513	14.231	87.000	141.000
相関行列						
	遠投	懸垂	握力	球速		
遠投	1.000	-0.187	0.394	0.539		
懸垂	-0.187	1.000	0.191	0.291		
握力	0.394	0.191	1.000	0.374		
球速	0.539	0.291	0.374	1.000		

目的変数は球速。説明変数間で握力と遠投に相関あり。2つのうち球速と相関がより強いのは遠投→変数選択に握力は不要か？(多重共線性)

ここで全変数を投入して式を求めてみる

全変数								
回帰式の精度								
ステップ	R	修正R	R2乗	修正R2乗	ダービン=ワトソン比	AIC	変数選択の結果	
ステップ0	0.0000	-	0.0000	-	2.3744	-	投入	除去
ステップ1	0.6739	0.5931	0.4541	0.3517	2.3115	101.0841	(すべて)	(なし)
回帰式の有意性 (分散分析)								
ステップ	要因	平方和	自由度	平均平方	F 値	P 値		
ステップ0	回帰変動	0.0000	0	-	-	-		
	誤差変動	3847.7500	19	202.5132				
	全体変動	3847.7500	19					
ステップ1	回帰変動	1747.2326	3	582.4109	4.4363	0.0189		
	誤差変動	2100.5174	16	131.2823				
	全体変動	3847.7500	19					

R²を決定係数(寄与率)といい、当てはまりの尺度、Rは重相関係数

回帰式に含まれる変数 (偏回帰係数・信頼区間等)									
ステップ	変数	偏回帰係数	標準誤差	標準偏回帰係	下限値	上限値	F 値	t 値	P 値
ステップ0	定数項	115.7500	3.1821		109.0898	122.4102	1323.1795	36.3755	p < 0.001
ステップ1	遠投	0.7112	0.2556	0.5844	0.1694	1.2529	7.7433	2.7827	0.0133
	懸垂	0.3764	0.1909	0.3877	-0.0283	0.7810	3.8869	1.9715	0.0662
	握力	0.0648	0.1960	0.0695	-0.3508	0.4804	0.1092	0.3305	0.7453
	定数項	48.0687	19.0678		7.6468	88.4907	6.3551	2.5209	0.0227

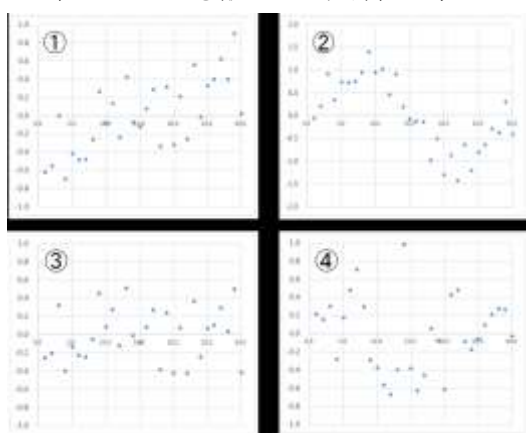
$$y = 0.711(\text{遠投}) + 0.376(\text{懸垂}) + 0.065(\text{握力}) + 48.069(\text{修正 R2 乗} = 0.352)$$

- * 「遠投」以外の変数が全て固定された状態で「遠投」が 1m 増加した時の「球速」の予測値の増加量は偏回帰係数である「0.7112」
- * 重回帰式の当てはまりのよさ→自由度調整済み決定係数(修正 R2 乗)

- * 偏回帰係数の値は偏回帰係数の大小を表すもの。
- * 偏回帰係数の P 値は個々の偏回帰係数の有意性を見るための指標
- * 重相関係数 $R(=\sqrt{R^2})$ は、実際に観測された目的変数の値と、重回帰式をあてはめて計算した推定値(理論値)との相関係数

参考) 例えば、4 つのデータに対して重回帰分析を行ったところ、次のような残差プロットが得られた。このうち、分析が適切に行われていると考えられるものは？
 重回帰式における残差プロットでは、下記の 2 点に着目する。

- 1) 段々上がっていく、下がっていくなどの傾向性がないか
- 2) どの点でも散らばり具合は等しいか



- ×: 残差が負から正になっている傾向がみられる。①
- ×: 周期的に変化している傾向がみられる。②
- : 特に傾向もなく一様に散らばっているのが正しい。③
- ×: 分散が中ほどで大きくなっている。④

変数どうしが「線形結合」しているものがある場合、ある変数が別の変数から 100%説明できてしまうことを意味し、このような変数を含む場合には、重回帰分析を行うことができない。エクセル統計では「線形結合している変数を除いて分析する」にチェックを入れると、線形結合している変数を自動的に除いた上で解析を行う。

ついで**変数増減法**により変数を選択した場合

変数選択の方法	投入基準P値	除去基準P値	投入、除去基準 0.20				変数選択の結果	
増減法	0.200	0.200	重相関係数	決定係数	ダービン=ワトソン比	AIC	投入	除去
ステップ	R	修正R	R2乗	修正R2乗	ダービン=ワトソン比	AIC		
ステップ0	0.0000	-	0.0000	-	2.3744	-	遠投	
ステップ1	0.5392	0.5013	0.2907	0.2513	2.2394	102.3202	懸垂	
ステップ2	0.6711	0.6210	0.4504	0.3857	2.3584	99.2202	(なし)	

ダービン・ワトソン比:
 誤差間相関なし→p59

ステップ	要因	平方和	自由度	平均平方	F 値	P 値
ステップ0	回帰変動	0.0000	0	-	-	-
	誤差変動	3847.7500	19	202.5132		
	全体変動	3847.7500	19			
ステップ1	回帰変動	1118.6028	1	1118.6028	7.3777	0.0142
	誤差変動	2729.1472	18	151.6193		
	全体変動	3847.7500	19			
ステップ2	回帰変動	1732.8923	2	866.4462	6.9648	0.0062
	誤差変動	2114.8577	17	124.4034		
	全体変動	3847.7500	19			

ステップ	変数	偏回帰係数	標準誤差	標準偏回帰	偏回帰係数の95%信頼区間				偏回帰係数の有意性の検定			目的変数との相関		多重共線性の統計量		
					下限値	上限値	F 値	t 値	P 値	単相関	偏相関	トレランス	VIF			
ステップ0	定数項	115.7500	3.1821		109.0898	122.4102	1323.1795	36.3755	P < 0.001							
ステップ1	遠投	0.6561	0.2415	0.5392	0.1486	1.1635	7.3777	2.7162	0.0142	0.5392	0.5392	1.0000	1.0000			
	定数項	66.1510	18.4669		27.3534	104.9485	12.8317	3.5821	0.0021							
ステップ2	遠投	0.7488	0.2227	0.6154	0.2789	1.2187	11.3015	3.3618	0.0037	0.5392	0.6319	0.9649	1.0364			
	懸垂	0.3948	0.1777	0.4068	0.0200	0.7697	4.9379	2.2221	0.0401	0.2915	0.4744	0.9649	1.0364			
	定数項	48.4425	18.5289		9.3500	87.5350	6.8353	2.6144	0.0181							

$$y = 0.749(\text{遠投}) + 0.395(\text{懸垂}) + 48.442(\text{修正 R2 乗} = 0.386)$$

$$y = 0.615(\text{遠投}) + 0.407(\text{懸垂}) \quad (\text{標準化偏回帰係数})$$

AIC 最小モデルを選択

変数増減法では握力が除去されている。後の式のほうが修正 R2 乗は大きく当てはまりが良い。これは相関係数で遠投と握力に 0.394 の関係があり、独立変数間に相関(線形関係あり)があるため回帰係数に不安定さが生まれた(多重共線性 multi-collinearity)ためである。VIF(variance inflation factors) は多重共線性の指標、10 以下は問題なし。(10 以上では多重共線性の弊害あり。許容度 Tolerance = 1/VIF の関係)

原則) 1) 変数の数は少ないほうが良い。

2) いきなり多変量解析せず単回帰分析を行い、相関関係を確認し相関の強い2つ以上の変数を同時に説明変数に含めないこと。

3) **AIC(赤池の information criteria)が最小のモデルを選択**する。→p174
また、偏回帰係数の有意性の検定を行い帰無仮説「偏回帰係数は 0 である」の有意確率 P 値が「除去基準 P 値」(0.20)より大きい場合、その説明変数が回帰式から除去する。

◇ **標準化偏回帰係数**どの変数の影響が強いかは偏回帰係数では単位が異なるため無意味。全ての変数を平均=0、標準偏差=1 と標準化した標準化偏回帰係数により判断する。

- ◇ 決定係数、修正 R2 乗により式で何%説明できるかわかる。
- ◇ 変数選択: Stepwise 法を使用、変数を一つずつ追加していく変数増加法、一つずつ減らしていく変数減少法、除去し、投入を繰り返す変数増減法あり。
- ◇ 説明変数 x は一般に、連続量や 2 値変数だが x に名義尺度を使いたいときはダミー変数を作る。

ダミー変数の作り方

名義尺度や順序尺度などのカテゴリカル変数を、重回帰分析や多重ロジスティック回帰分析などの説明変数として利用するため、カテゴリごとに 0、1 の 2 値をとるダミー変数に変換する。すべてのカテゴリをダミー変数化して多変量解析の説明変数とすると、線形結合を起こし分析できなくなる。最後のカテゴリをダミー変数化せず、(カテゴリ数-1)個のダミー変数に変換する。

BellCurve for Excel: ユーティリティ→ダミー変数への変換

喫煙	年齢	性別	お酒
喫煙者	20~24才	男性	好き
喫煙者	20~24才	女性	普通
喫煙者	20~24才	男性	好き
喫煙者	20~24才	女性	嫌い
喫煙者	25~29才	男性	普通
喫煙者	25~29才	女性	好き
非喫煙者	25~29才	男性	嫌い
喫煙者	30~39才	男性	普通
喫煙者	30~39才	女性	普通
喫煙者	30~39才	男性	普通
非喫煙者	30~39才	女性	好き
非喫煙者	30~39才	女性	嫌い
喫煙者	30~39才	女性	嫌い
喫煙者	40~49才	男性	好き
喫煙者	40~49才	女性	好き
喫煙者	40~49才	男性	普通
非喫煙者	40~49才	女性	嫌い
非喫煙者	40~49才	男性	嫌い
非喫煙者	40~49才	女性	普通
喫煙者	50才以上	男性	好き
非喫煙者	50才以上	女性	嫌い
喫煙者	50才以上	男性	普通
非喫煙者	50才以上	女性	嫌い
非喫煙者	50才以上	女性	普通
非喫煙者	50才以上	男性	普通

喫煙 1:喫煙者 0:非喫煙者 →1 列で処理
 年齢: (20~24 才) (25~29 才) (30~39 才) (40~49 才) →相当する 4 つの列を作り 1 (50 才以上) →列は作らずすべて 0
 性別 女性 1 男性 0 →1 列
 お酒 嫌い 1、好き 1 →2 列
 普通 2 列とも 0

例えば季節ダミーでは変数 X_k : 春 1、他 0、
 X_{k+1} : 夏 1、他 0、 X_{k+2} : 秋 1、他 0 →3 列

No.	喫煙 喫煙:喫煙者	年齢 年齢:20~24才	年齢:25~29才	年齢:30~39才	年齢:40~49才	性別 性別:女性	お酒 お酒:嫌い	お酒:好き
1	1	1	0	0	0	0	0	1
2	1	1	0	0	0	1	0	0
3	1	1	0	0	0	0	0	1
4	1	1	0	0	0	1	1	0
5	1	0	1	0	0	0	0	0
6	1	0	1	0	0	1	0	1
7	0	0	1	0	0	0	1	0
8	1	0	0	1	0	0	0	0
9	1	0	0	1	0	1	1	0
10	1	0	0	1	0	0	0	0
11	0	0	0	1	0	1	0	1
12	0	0	0	1	0	1	1	0
13	1	0	0	1	0	1	1	0
14	1	0	0	0	1	0	0	1
15	1	0	0	0	1	1	0	1
16	1	0	0	0	1	0	0	0
17	0	0	0	0	1	1	1	0
18	0	0	0	0	1	0	1	0
19	0	0	0	0	1	1	0	0
20	1	0	0	0	0	0	0	1
21	0	0	0	0	0	1	1	0
22	1	0	0	0	0	0	0	0
23	0	0	0	0	0	1	1	0
24	0	0	0	0	0	1	0	0
25	0	0	0	0	0	0	0	0

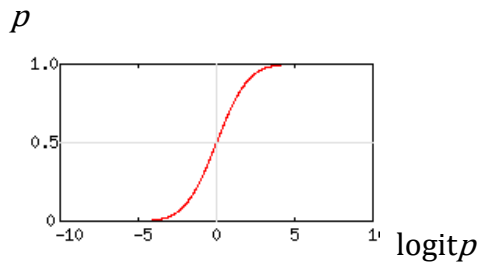
ロジスティック回帰分析 logistic regression analysis

あることが起こる確率を p とすると、起こらない確率は $(1 - p)$

$\frac{p}{1-p}$ はオッズ(odds)と呼ばれている

ここで、 $\lambda = \text{logit}p = \log_e\left(\frac{p}{1-p}\right)$ という変換(ロジット、対数オッズ)を考える。

下図の関係になり $\text{logit}p$ がどんな値をとっても $0 < p < 1$ で



これを p について解くと $p = \frac{\exp(\lambda)}{1 + \exp(\lambda)} = \frac{1}{1 + \exp(-\lambda)}$

ある集団(予後因子のない)で起きる確率 $p_{x_i=0}$ 、ある集団(予後因子のある)で起きる確率 $p_{x_i=1}$ とすると

$\left(\frac{p_{x_i=1}}{1-p_{x_i=1}}\right) / \left(\frac{p_{x_i=0}}{1-p_{x_i=0}}\right)$ はオッズ比であり相対危険度 relative risk にほぼ等しい。

ここで $\text{logit}p$ が予後因子を表す変数 X_i の一次式(β は係数)で表現されたものをロジスティックモデルという。

$$\text{logit}p = \log_e\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

$$\Leftrightarrow \left(\frac{p}{1-p}\right) = \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)$$

予後因子 X_1 が 0, 1 の 2 値のみをもつ場合(他の予後因子は変化なし)

$$\text{logit}p_{x_1=0} = \alpha + \beta_2 X_2 + \cdots + \beta_n X_n$$

$$\text{logit}p_{x_1=1} = \alpha + \beta_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

$$\text{よって } \text{logit}p_{x_1=1} - \text{logit}p_{x_1=0} = \beta_1$$

$$\log_e \left(\frac{p_{x_i=1}}{1-p_{x_i=1}} \right) / \left(\frac{p_{x_i=0}}{1-p_{x_i=0}} \right) = \beta_1 \text{ より } \left(\frac{p_{x_i=1}}{1-p_{x_i=1}} \right) / \left(\frac{p_{x_i=0}}{1-p_{x_i=0}} \right) = \exp(\beta_1)$$

オッズ比(相対危険度)は $\exp(\beta_1)$ となる。

その 95%信頼区間(95%CI)は $\exp(\beta_1 \pm 1.96 \times SE)$ となる。

ロジスティック回帰分析(二項ロジスティック回帰):

二項、順序、多項ロジスティック回帰などがあるが二項ロジスティック回帰のみ扱う。

従属変数が 2 値変数の際に用い確率のデータを扱うことができる。

2 値変数(満足・不満足、成功・失敗、発症・非発症、賛成・反対など)

→ある事象の生起の有無(0と1のデータと考えることが多い)

がんの発症確率や生存率などの"確率"について回帰分析を用いて考えたいとき

目的変数が比率データの時 等

説明変数は離散量や連続量を扱うこともできる。

$$\text{logit}p_{x_1=i_0} = \alpha + \beta_1 i_0 + \beta_2 X_2 + \dots + \beta_n X_n$$

$$\text{logit}p_{x_1=i} = \alpha + \beta_1 i + \beta_2 X_2 + \dots + \beta_n X_n$$

$$\text{よって} \text{logit}p_{x_1=i} - \text{logit}p_{i_0} = \beta_1(i - i_0), \quad \left(\frac{p_{x_1=i}}{1-p_{x_1=i}}\right) / \left(\frac{p_{x_1=i_0}}{1-p_{x_1=i_0}}\right) = \exp\{\beta_1(i - i_0)\}$$

その 95%信頼区間(95%CI)は $\exp\{(\beta_1 \pm 1.96 \times SE) \times (i - i_0)\}$ となる。

ロジスティック回帰分析と線形回帰分析の比較

- 正規線形モデルでの回帰分析は 従属変数が量的変数、従属変数が正規分布に従うかどうか。→ 回帰分析はデータの分布に正規分布が仮定されている。
- ロジスティック回帰分析→ 従属変数が質的変数、説明変数は連続値でも名義変数でも可。独立変数の尺度, 分布型に対して厳密な仮定をおいていない。係数としてオッズ比を求めることができ、各対象者の事象の起こる確率を求められる。

ロジスティック回帰分析の注意点

i) 誤差の少ない信頼のおける多数のデータに適用したか?

目安: 例数 \geq (変数の数 $\times 10$) または (変数の数の 2 乗) の大きい方
疾患の発症例数と非発症例数はできるだけ同じくらいが理想

ii) ロジスティック回帰分析に組み込んだ項目が適切か? (多重共線性の有無)

→ 説明変数同志で強い相関のあるものは同時に組み入れない。

iii) 組み込んだ項目はリスクファクターだけか? 診断指標に相当するものはないか?

iv) ロジスティック回帰式が実質科学的に納得できるか?

v) ロジットは確率が 0 または 1 になる時は計算できない(完全分離)

説明変数によって疾患の発症と非発症が完全に決まってしまう時は計算不可能例)

y	0	0	0	0	0	1	1	1	1	1
x	1.2	2.8	3.9	4.5	5.5	6.3	7.1	8.2	9.4	9.9

BellCurve for Excel: 多変量解析→二項ロジスティック回帰分析

変数選択過程										
回帰式の精度										
			決定係数						変数選択の結果	
ステップ	-2対数尤度	AIC	R2乗	Cox-Snell	Nagelkerke	相関係数	誤判別率	投入	除去	
ステップ0	969.9431	-	-	-	-	-	48.71%	(すべて)		
ステップ1	509.1524	513.1524	0.4751	0.4823	0.6431	0.7410	15.86%		(なし)	

変数選択結果											
回帰式の精度											
		決定係数									
-2対数尤度	AIC	R2乗	Cox-Snell	Nagelkerke	相関係数	誤判別率					
509.1524	513.1524	0.4751	0.4823	0.6431	0.7410	15.86%					
回帰式の有意性											
ステップ		尤度比	自由度	P 値							
ステップ1	ステップ	460.7907	1	p < 0.001							
	モデル	460.7907	1	p < 0.001							
回帰式に含まれる変数 (偏回帰係数・信頼区間等)											
			偏回帰係数の95%信頼区間			オッズ比の95%信頼区間		偏回帰係数の有意性検定			
変 数	偏回帰係数	標準誤差	標準偏回帰係数	下限値	上限値	オッズ比	下限値	上限値	Wald	自由度	P 値
衝撃力	1.2278	0.0857	2.4557	1.0599	1.3958	3.4139	2.8861	4.0381	205.3703	1	p < 0.001
定数項	-5.0248	0.3673		-5.7447	-4.3049	0.0066	0.0032	0.0135	187.1488	1	p < 0.001

$$\text{logit}(y) = -5.025 + 1.228x, y = \frac{1}{1 + \exp[-(-5.025 + 1.228x)]}$$

参考) 通常の 2 変量の回帰分析に比較した場合

変数選択の過程で計算される -2 対数尤度 969.9431 …… 全変動に相当

回帰式の有意性での尤度比(カイ 2 乗値) 460.7907 …… 回帰変動に相当

回帰式の精度 -2 対数尤度 509.1524 …… 残差変動に相当

回帰係数の有意性は Wald の統計量 205.37(p<0.001) :有意

寄与率 R²(McFadden) 0.475, Cox & Snell 0.482, Nagelkerke 0.6431

関連事項:

-2 対数尤度 (尤度比検定)

尤度比検定は、2 つのモデルの適合性を比較する検定。すべてのパラメータが自由な無制約モデルと少ないパラメータを持ち制約を受けるモデルにおいて帰無仮説と対立仮説における尤度の比(λ)を求め、その対数値に-2 をかけた値、統計量-2log λ は、無制約モデルと制約モデルの持つ自由なパラメータの数の差に等しい自由度を持つカイ二乗分布に従う。モデルに定数項のみを含む場合の-2 対数尤度と、モデルに定数項と説明変数を含む場合の-2 対数尤度の差を用いて、尤度比検定を行う。

AIC(赤池の information criteria):統計モデルの予測の良さの選択規準

$-2\log L+2k$ (または、 $-2MLL+2k$)

L:最大尤度、MLL: 最大対数尤度 k: パラメータ数

①比較したい色々な曲線(例えば各種成長曲線、正規分布とその他の分布などに適用しモデル選択)②次数を上げてその測定データとの適合度を高めることができるが、同種の別のデータには合わなくなる(過適合問題、Overfitting)を解決。

⇒ AIC 最小のモデルを選択

AIC は統計モデルのあてはまりの良さ(goodness of fit)ではなく、予測の良さ(goodness of prediction)を重視するモデル選択規準

重回帰分析、ロジスティック回帰分析などでは、変数選択の過程で全変数採用、変数除去された場合などで、投入、除去する基準 p 値を(例えば 0.200)と決めて AIC の減少をみながら決める。

ワルド検定(Wald test):係数の有意性の検定

係数(重み)をゼロにしてもよい、という帰無仮説に対して計算される p 値の水準を評価する。p 値が小さいほど「ゼロにしてもよい」という仮説が棄却される可能性が高くなり、逆に当該指標が有効である可能性が示されることになる。

⇒ β_i の推定値を $\hat{\beta}_i$ とすると、 β_i は正規分布からのデータであり Z 検定できる。

$z = W_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)}$ が 1.96 より大きければ 5%水準で有意。(または、zを 2 乗し

たものは自由度1の χ^2 分布に従う。2 乗したものが 3.84 より大きければ有意。)

Wald 信頼区間

オッズ比の信頼区間:95%信頼区間同一の調査, 同一の計算方法を用いた場合, 推定した信頼区間の中に 100 回中 95 回入る。ロジスティック回帰分析では, Wald 信頼区間 $\exp(\beta \pm 1.96 \times \text{標準誤差})$: β は回帰係数。この 95%信頼区間に 1 を含まなければ, その要因は 5%水準で有意, 1 を含めば 5%水準で有意ではない

モデルの適合度の評価:Hosmer-Lemeshow 検定

標本サンプルを 10 分の 1 に分けて, 各グループのモデルの良さを検討する。観測値と予測値の適合を評価するため, χ^2 検定を行う。帰無仮説:「観測値=期待値」(ロジスティック回帰モデルはデータに適合する)、対立仮説:「観測値 \neq 期待値」(ロジスティック回帰モデルはデータに適合しない) → $p < 0.05$ ならばモデルが適合していないことになる。注意点、ロジスティック回帰分析にも多重共線性の問題が存在する。

回帰式に、相関の高い変数を組み合わせていれたときに、回帰式の当てはまりが悪くなる場合が存在する→ 確認手段として、相関の高い変数の組み合わせが存在しないかを確認(多重共線性)。

変数選択法

① 強制投入法

複数の説明変数を同時にモデルに投入する

独立変数の重要性の順序などの仮説がない場合に使用

② 変数増加法

指定した独立変数のうち従属変数に最も強く関連している変数が選ばれ、以後順番に相関の強い変数が選ばれる

③ 変数減少法

指定した独立変数のうち従属変数に対して最も関連が弱く有意でない変数から順番に削除されていく。

例題

統計 WEB の閲覧経験の有無と、統計学の試験勉強の時間を説明変数として、統計学の試験の合否を予測する二項ロジスティックモデルを作成

統計学試験の結果			
No.	閲覧経験	勉強時間	試験結果
1	1	24	1
2	1	18	1
3	0	15	1
4	1	16	1
5	0	10	1
6	1	26	1
7	1	2	1
8	0	24	1
9	1	18	1
10	1	22	1
11	1	3	1
12	1	6	0
13	0	15	0
14	0	12	0
15	1	6	0

16	0	6	0
17	1	12	0
18	0	12	0
19	1	18	0
20	1	3	0
21	1	8	0
22	0	9	0
23	0	12	0
24	0	6	0
25	0	8	0
26	1	12	0

変数表

試験結果	合格	1
	不合格	0
閲覧経験	あり	1
	なし	0

解析結果のみ拡大

変数	偏回帰係数	オッズ比	オッズ比の95%信頼区間		偏回帰係数の有意性検定			
			下限値	上限値	Wald	自由度	P 値	
閲覧経験	1.0843	2.9572	0.4583	19.0813	1.2991	1	0.2544	
勉強時間	0.1662	1.1809	1.0150	1.3738	4.6328	1	0.0314	*
定数項	-3.0374	0.0480	0.0044	0.5255	6.1838	1	0.0129	*

BellCurve for Excel: 閲覧経験勉強時間試験結果を選択し多変量解析→二項ロジスティック回帰分析

二項ロジスティック回帰分析							
目的変数の要約							
試験結果	n	%					
試験結果=0	15	57.69%					
試験結果=1	11	42.31%					
全 体	26	100.00%					
基本統計量							
目的変数	変 数	n	平 均	不偏分散	標準偏差	最小値	最大値
全 体	閲覧経験	26	0.577	0.254	0.504	0.000	1.000
	勉強時間	26	12.423	46.894	6.848	2.000	26.000
試験結果=0	閲覧経験	15	0.467	0.267	0.516	0.000	1.000
	勉強時間	15	9.667	16.381	4.047	3.000	18.000
試験結果=1	閲覧経験	11	0.727	0.218	0.467	0.000	1.000
	勉強時間	11	16.182	67.364	8.208	2.000	26.000
相関行列							
目的変数		閲覧経験	勉強時間				
全 体	閲覧経験	1.000	0.089				
	勉強時間	0.089	1.000				
試験結果=0	閲覧経験	1.000	-0.091				
	勉強時間	-0.091	1.000				
試験結果=1	閲覧経験	1.000	-0.012				
	勉強時間	-0.012	1.000				
線形結合している変数							
なし							

変数選択の方法:全変数											
回帰式の精度											
	決定係数								変数選択の結果		
-2対数尤度	AIC	R2乗	Cox-Snell	Nagelkerke	相関係数	誤判別率	投入	除去			
27.6181	33.6181	0.2204	0.2594	0.3487	0.5309	26.92%					
回帰式の有意性											
ステップ		尤度比	自由度	P 値							
ステップ1	ステップ	7.8078	2	0.0202							
	モデル	7.8078	2	0.0202							
回帰式に含まれる変数 (偏回帰係数・信頼区間等)											
*: P<0.05 **: P<0.01											
		偏回帰係数の95%信頼区間				オッズ比の95%信頼区間		偏回帰係数の有意性検定			
変 数	偏回帰係数	標準誤差	標準偏回帰係数	下限値	上限値	オッズ比	下限値	上限値	Wald	自由度	P 値
閲覧経験	1.0843	0.9513	0.5357	-0.7802	2.9487	2.9572	0.4583	19.0813	1.2991	1	0.2544
勉強時間	0.1662	0.0772	1.1163	0.0149	0.3176	1.1809	1.0150	1.3738	4.6328	1	0.0314*
定数項	-3.0374	1.2214		-5.4313	-0.6434	0.0480	0.0044	0.5255	6.1838	1	0.0129*

偏回帰係数 β_1 の時オッズ比は $\exp(\beta_1)$ となる。
オッズ比の95%信頼区間に1を含まなければ有意。

例題2) ロジスティック・Cox回帰入門(高橋善弥太著)よりP4サンプルデータ
劇症肝炎40例

予後:生存0, 死亡1、ウイルス:A=1 B=2、治療:治療群1, 対照群0

予後	ウイルス	年齢	治療
0	1	38	1
0	1	36	1
0	1	46	1
0	1	47	1
0	1	45	1
0	1	45	1
0	1	42	1
0	1	40	1
0	1	46	1
0	1	40	1
0	2	30	1
0	2	45	1
1	1	55	1
1	1	58	1
1	2	38	1
1	2	50	1
1	2	55	1
1	2	45	1
1	2	52	1
0	1	42	0
0	1	41	0
0	1	52	0
0	2	50	0
0	2	36	0
1	1	45	0
1	1	50	0
1	1	63	0
1	2	54	0
1	2	60	0
1	2	51	0
1	2	48	0
1	2	56	0
1	2	58	0
1	2	53	0
1	2	47	0
1	2	58	0
1	2	60	0

1	2	53	0
1	2	57	0
1	2	53	0

二項ロジスティック回帰分析							
目的変数の要約							
予後	n	%					
予後=0	17	42.50%					
予後=1	23	57.50%					
全 体	40	100.00%					
基本統計量							
目的変数	変 数	n	平 均	不偏分散	標準偏差	最小値	最大値
全 体	ウイルス	40	1.550	0.254	0.504	1.000	2.000
	年齢	40	48.500	59.641	7.723	30.000	63.000
	治療	40	0.475	0.256	0.506	0.000	1.000
予後=0	ウイルス	17	1.235	0.191	0.437	1.000	2.000
	年齢	17	42.412	30.382	5.512	30.000	52.000
	治療	17	0.706	0.221	0.470	0.000	1.000
予後=1	ウイルス	23	1.783	0.178	0.422	1.000	2.000
	年齢	23	53.000	33.818	5.815	38.000	63.000
	治療	23	0.304	0.221	0.470	0.000	1.000
相関行列							
目的変数		ウイルス	年齢	治療			
全 体	ウイルス	1.000	0.277	-0.347			
	年齢	0.277	1.000	-0.450			
	治療	-0.347	-0.450	1.000			
予後=0	ウイルス	1.000	-0.224	-0.251			
	年齢	-0.224	1.000	-0.216			
	治療	-0.251	-0.216	1.000			
予後=1	ウイルス	1.000	-0.111	-0.110			
	年齢	-0.111	1.000	-0.299			
	治療	-0.110	-0.299	1.000			
線形結合している変数							
なし							

参考)

R²乗 (寄与率) : McFaddenのR²乗とも言う。0から1までの値をとり、1に近いほど精度が高い。
 Cox-Snell R²乗: R²乗は値が小さくなりすぎるため、この点を修正したもの。最大値は1にならない。
 Nagelkerke R²乗: Cox-Snell R²乗の最大値が1にならない点を修正したもの。

変数選択の方法	投入基準P値	除去基準P値							
増減法	0.200	0.200							
変数選択結果									
回帰式の精度									
		決定係数						変数選択の結果	
-2対数尤度	AIC	R2乗	Cox-Snell	Nagelkerke	相関係数	誤判別率	投入	除去	
23.5740	29.5740	0.5678	0.5390	0.7242	0.7865	15.00%			
回帰式の有意性									
ステップ		尤度比	自由度	P 値					
ステップ2	ステップ	6.8212	1	0.0090					
	モデル	30.9744	2	p < 0.001					

回帰式に含まれる変数 (偏回帰係数・信頼区間等)											
			偏回帰係数の95%信頼区間			オッズ比の95%信頼区間		偏回帰係数の有意性検定			
変 数	偏回帰係数	標準誤差	標準偏回帰	下限値	上限値	オッズ比	下限値	上限値	Wald	自由度	P 値
ウイルス	2.6260	1.0971	1.3064	0.4757	4.7763	13.8187	1.6092	118.6675	5.7292	1	0.0167
年齢	0.2881	0.0959	2.1966	0.1002	0.4759	1.3338	1.1054	1.6095	9.0307	1	0.0027
定数項	-17.3152	5.0863		-27.2841	-7.3463	0.0000	0.0000	0.0006	11.5892	1	p < 0.001

年齢については連続量でオッズ比は(1歳に対して)1.3338
 年齢10歳の加齢に対する値、オッズ比にするには
 1.3338を10乗し = **17.8242470377305** となる

← 5歳、10歳毎などの年齢に対するオッズ比の求め方

生存率の分析

生存時間解析(survival time analysis)

観察開始時から, ある疾患の発症や死亡, 入院といったイベントの発生までの時間を記録し, 生存時間に影響を与える要因を検討解析する。

イベント (event) = エンドポイント(end point) = 結果 (outcome)

解析対象の時間: 生存時間 (survival time = failure time)

説明する変数: 説明変数(explanatory variable) = 独立変数 (independent variable) = 共変量(covariate) = 危険因子(risk factor) = 予後因子 (prognostic factor)

打ち切り例(censored case)とは

- 1) 追跡不能例(withdraw、lost to follow-up)
- 2) 研究期間終了時の生存例(censoring)

打ち切りが生じた場合, 生存時間そのものは不明となるが, 「観察期間中にイベントが起これなかった」という情報は手に入ることになる。

Kaplan-Meier法

(product-limit推定法)

Kaplan-Meier法を手書きする。S(t):時間tにおける生存率

S(t)を時間tまで生存時間が短い順に並び替える。

打ち切り例には1を代入し, r番目の症例での累積生存率は

$$S(t) = \frac{(n-1)}{(n-1+1)} \times \frac{(n-2)}{(n-2+1)} \times \dots \times \frac{(n-r)}{(n-r+1)} = \prod_{i=1}^r \frac{(n-i)}{(n-r+1)}$$

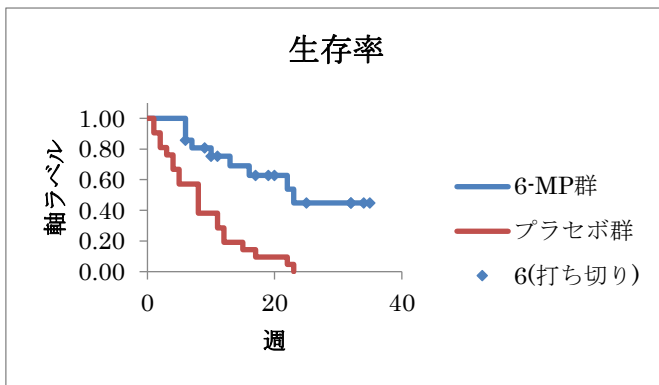
群	週	状態	群	週	状態
6-MP 群	6	1	プラセボ群	1	1
6-MP 群	6	1	プラセボ群	1	1
6-MP 群	6	1	プラセボ群	2	1
6-MP 群	7	1	プラセボ群	2	1
6-MP 群	10	1	プラセボ群	3	1
6-MP 群	13	1	プラセボ群	4	1
6-MP 群	16	1	プラセボ群	4	1
6-MP 群	22	1	プラセボ群	5	1
6-MP 群	23	1	プラセボ群	5	1
6-MP 群	6	0	プラセボ群	8	1
6-MP 群	9	0	プラセボ群	8	1
6-MP 群	10	0	プラセボ群	8	1
6-MP 群	11	0	プラセボ群	8	1
6-MP 群	17	0	プラセボ群	11	1
6-MP 群	19	0	プラセボ群	11	1
6-MP 群	20	0	プラセボ群	12	1
6-MP 群	25	0	プラセボ群	12	1
6-MP 群	32	0	プラセボ群	15	1
6-MP 群	32	0	プラセボ群	17	1

6-MP 群	34	0	プラセボ群	22	1
6-MP 群	35	0	プラセボ群	23	1

生存率表		21 人からスタート				
系列	週	死亡	打ち切り	$\frac{(n-r)}{(n-r+1)}$		累積生存率
6-MP 群	6	3	1	$\frac{21}{21} \cdot \frac{20}{21} \cdot \frac{19}{20} \cdot \frac{18}{19} \cdot \frac{17}{17}$		0.8571
	7	1	0	$\frac{16}{17}$		0.8067
	9	0	1	$\frac{15}{15}$		0.8067
	10	1	1	$\frac{14}{15} \cdot \frac{13}{13}$		0.7529
	11	0	1	$\frac{12}{12}$		0.7529
	13	1	0	$\frac{11}{12}$		0.6902
	16	1	0	$\frac{10}{11}$		0.6275
	17	0	1	$\frac{9}{9}$		0.6275
	19	0	1	$\frac{8}{8}$		0.6275
	20	0	1	$\frac{7}{7}$		0.6275
	22	1	0	$\frac{6}{7}$		0.5378
	23	1	0	$\frac{5}{6}$		0.4482
	25	0	1	$\frac{4}{4}$		0.4482
	32	0	2	$\frac{2}{2}$		0.4482
	34	0	1	$\frac{1}{1}$		0.4482
35	0	1	$\frac{1}{1}$		0.4482	
系列	週	死亡	打ち切り	残数		累積生存率
プラセボ群	1	2	0		21	0.9048
	2	2	0		19	0.8095
	3	1	0		17	0.7619
	4	2	0		16	0.6667
	5	2	0		14	0.5714
	8	4	0		12	0.3810
	11	2	0		8	0.2857
	12	2	0		6	0.1905
	15	1	0		4	0.1429
	17	1	0		3	0.0952
	22	1	0		2	0.0476
23	1	0		1	0.0000	

死亡: 前の分子から 1 を引き分子とし、前の分母で割り、分子は次の分母となる。
打ち切り: 追跡不能例では前の分子から 1 を引き分子とし、この時、分母も 1 減るため分子と同数になるので比が 1 となる

丹後俊郎：新版医学への統計学 p 206 より



縦軸の 0.5 から横軸に平行に線を引けば、グラフとの交点が MST

BellCurve for Excel: 生存分析・ハザード分析→カプラン=マイヤー法

カプラン=マイヤー法による生存分析			
ケースの要約			
系 列	死 亡	打ち切り	全 体
6-MP群	9	12	21
プラセボ群	21	0	21

MST

生存時間の平均値と中央値								
系 列	平均値	標準誤差	95%信頼区間		中央値	標準誤差	95%信頼区間	
			下限値	上限値			下限値	上限値
6-MP群	23.2874	2.8275	17.7457	28.8291	23.0000	5.2554	12.6997	33.3003
プラセボ群	8.6667	1.4114	5.9004	11.4330	8.0000	1.6690	4.7287	11.2713

生存率曲線の差の検定					
手 法		カイニ乗値	自由度	P 値	* : P<0.05 ** : P<0.01
ログランク検定	Peto-Peto	15.2329	1	p < 0.001	**
	Cochran-Mantel-Haenszel	16.7929	1	p < 0.001	**
一般化Wilcoxon検定	Gehan-Breslow	13.4579	1	p < 0.001	**
	Peto-Prentice	14.0841	1	p < 0.001	**

log-rank test・
Generalized
Wilcoxon test ともに
有意

生存率表									
系 列	週	死 亡	打ち切り	残 数	累積生存率	標準誤差	95%信頼区間 (変換なし)		
							下限値	上限値	
6-MP群	6	3	1	21	0.8571	0.0764	0.7075	1.0000	
	7	1	0	17	0.8067	0.0869	0.6363	0.9771	
	9	0	1	16	0.8067	0.0869	0.6363	0.9771	
	10	1	1	15	0.7529	0.0963	0.5641	0.9418	
	11	0	1	13	0.7529	0.0963	0.5641	0.9418	
	13	1	0	12	0.6902	0.1068	0.4808	0.8995	
	16	1	0	11	0.6275	0.1141	0.4039	0.8510	
	17	0	1	10	0.6275	0.1141	0.4039	0.8510	
	19	0	1	9	0.6275	0.1141	0.4039	0.8510	
	20	0	1	8	0.6275	0.1141	0.4039	0.8510	
	22	1	0	7	0.5378	0.1282	0.2865	0.7891	
	23	1	0	6	0.4482	0.1346	0.1844	0.7120	
	25	0	1	5	0.4482	0.1346	0.1844	0.7120	
	32	0	2	4	0.4482	0.1346	0.1844	0.7120	
	34	0	1	2	0.4482	0.1346	0.1844	0.7120	
35	0	1	1	0.4482	0.1346	0.1844	0.7120		
プラセボ群	1	2	0	21	0.9048	0.0641	0.7792	1.0000	
	2	2	0	19	0.8095	0.0857	0.6416	0.9775	
	3	1	0	17	0.7619	0.0929	0.5797	0.9441	
	4	2	0	16	0.6667	0.1029	0.4650	0.8683	
	5	2	0	14	0.5714	0.1080	0.3598	0.7831	
	8	4	0	12	0.3810	0.1060	0.1733	0.5887	
	11	2	0	8	0.2857	0.0986	0.0925	0.4789	
	12	2	0	6	0.1905	0.0857	0.0225	0.3584	
	15	1	0	4	0.1429	0.0764	0.0000	0.2925	
	17	1	0	3	0.0952	0.0641	0.0000	0.2208	
	22	1	0	2	0.0476	0.0465	0.0000	0.1387	
	23	1	0	1	0.0000	0.0000	0.0000	0.0000	

グラフ用データ					
週	6-MP群	6(打ち切り)	週	プラセボ群	1(打ち切り)
0	1.0000		0	1.0000	
6	1.0000		1	1.0000	
6	0.8571	0.8571	1	0.9048	
7	0.857142857		2	0.904761905	
7	0.8067		2	0.8095	
9	0.8067	0.8067	3	0.80952381	
10	0.806722689		3	0.7619	
10	0.7529	0.7529	4	0.761904762	
11	0.7529	0.7529	4	0.6667	
13	0.752941176		5	0.666666667	
13	0.6902		5	0.5714	
16	0.690196078		8	0.571428571	
16	0.6275		8	0.3810	
17	0.6275	0.6275	11	0.380952381	
19	0.6275	0.6275	11	0.2857	
20	0.6275	0.6275	12	0.285714286	
22	0.62745098		12	0.1905	
22	0.5378		15	0.19047619	
23	0.537815126		15	0.1429	
23	0.4482		17	0.142857143	
25	0.4482	0.4482	17	0.0952	
32	0.4482	0.4482	22	0.095238095	
32	0.4482	0.4482	22	0.0476	
34	0.4482	0.4482	23	0.047619048	
35	0.4482	0.4482	23	0.0000	

Kaplan-Meier法の欠点

- * 単変量解析であり、交絡因子の影響を受けやすいため、年齢、病期などで層別する必要がある。
- * 症例数がすくなくと層別した場合解析が困難となる
- * このような場合はCoxの比例ハザードモデルがよい。
- * カプラン・マイヤーで作成できるグラフは視覚的に2群に差があるかどうかは把握しやすいが、有意差を示すかどうかはログランク検定(log-rank test)、一般化ウィルコクソン検定(Generalized Wilcoxon test)などの検定を用いる必要がある。

ログランク検定 (log-rank test) と 一般化ウイルコクソン検定 (Generalized Wilcoxon test)

ログランク検定 (log-rank test) : Peto-Peto による方法と Cochran-Mantel-Haenszel による方法の 2 種類ある。Peto-Peto による方法は計算方法が簡便だが、Cochran-Mantel-Haenszel による方法に比べて少し有意になりにくい性質がある。単にログランク検定という場合、一般に Cochran-Mantel-Haenszel による方法のことを指す。

一般化ウイルコクソン検定 (Generalized Wilcoxon test) : Gehan-Breslow による方法と Peto-Prentice による方法の 2 種がある。Gehan-Breslow による方法は、初期に非常に多くの打ち切りがある場合、最初の数例の死亡でほとんど結果が決まってしまう。また、打ち切りのパターンに依存して大きく結果が異なる欠点がある。これらの欠点を改善したものが Peto-Prentice による方法となる。打ち切りがない場合、Gehan-Breslow による方法と Peto-Prentice による方法の結果は等しくなる。

例) 投与群は 1: 実薬群、2: プラセボ群とする。

土居正明氏 www012.upp.so-net.ne.jp/doi/.../logrank/Logrank_genWil.pdf より

投与群	被験者番号	生存時間またはセンサーまでの時間	死亡 (1) か センサー (0) か
1	1	5	0
1	2	7	0
1	3	8	1
1	4	12	0
2	5	2	0
2	6	3	1
2	7	4	1
2	8	10	0

データに対して、log-rank test・Generalized Wilcoxon test の両方を考える。

$S_1(t)$: 実薬群の生存関数、 $S_2(t)$: プラセボ群の生存関数 (両側検定で)

帰無仮説 H_0 : 2 つの生存曲線に差はない $S_1(t) = S_2(t)$

対立仮説 H_1 : 2 つの生存曲線に差がある $S_1(t) \neq S_2(t)$

生存関数の形そのものは何でもよく、この 2 つの検定はデータの従う分布に依存しないノンパラメトリックな検定。

死亡発生ごとのクロス表

- 1) クロス表の作成 log-rank test・Generalized Wilcoxon test 共に、死亡が発生するごとに 2×2 クロス表を作成する。

t=3	死亡例数	生存例数	合計
実薬群	0	4	4
プラセボ群	1	2	3
合計	1	6	7

t=4	死亡例数	生存例数	合計
実薬群	0	4	4
プラセボ群	1	1	2
合計	1	5	6

t=8	死亡例数	生存例数	合計
実薬群	1	1	2
プラセボ群	0	1	1
合計	1	2	3

censored case は、発生時点では生存関数に影響を与えないので、発生時点でのクロス表を作成する必要なし。censored case の影響が出るのは、次に死亡が発生した時点。

- 2) 死亡例数の期待値・分散の計算、次に、実薬群の死亡例数の期待値・分散を計算する。なお、ここで重要なポイントは各投与群の合計例数・両群合わせた死亡例数・両群合わせた生存例数は全て固定された定数を考えるということ。つまり、

時点t	死亡例数	生存例数	合計
実薬群	d_t		M_t (固定)
プラセボ群			$N_t - M_t$ (固定)
合計	n_t (固定)	$N_t - n_t$ (固定)	N_t (固定)

とする。このとき、確率変数 d_t は超幾何分布に従う。以下、

N_t : 時点 t での、両群合わせた at risk number (「その時点の直前での対象例数」を指す。つまり、「その時点での死亡やセンサーの症例を含めた例数」)

d_t : 時点 t での実薬群の死亡例数

e_t : 時点 t での実薬群の死亡例数の期待値

v_t : 時点 t での実薬群の死亡例数の分散

とし、超幾何平均の平均・分散を用いると

$$e_t = E[d_t] = \frac{n_t M_t}{N_t}$$

$$v_t = V[d_t] = \frac{M_t n_t (N_t - M_t) (N_t - n_t)}{N_t^2 (N_t - 1)}$$

これらを時点ごとに計算

t = 3 の場合、

$$N_3 = 7, d_3 = 0, e_3 = 1 \cdot 4 / 7 = 4/7, V_3 = 4 \cdot 3 \cdot 1 \cdot 6 / 7^2 \cdot (7-1) = 12/49$$

t = 4 の場合

$$N_4 = 6, d_4 = 0, e_4 = 1 \cdot 4 / 6 = 4/6, V_4 = 4 \cdot 2 \cdot 1 \cdot 5 / 6^2 \cdot (6-1) = 8/36$$

t = 8 の場合

$$N_8 = 3, d_8 = 1, e_8 = 1 \cdot 2 / 3 = 2/3, V_8 = 2 \cdot 1 \cdot 1 \cdot 2 / 3^2 \cdot (3-1) = 2/9$$

表にまとめると以下の通り。

t	N _t	d _t	e _t	v _t
3	7	0	4/7	12/49
4	6	0	4/6	8/36
8	3	1	2/3	2/9

2×2 のクロス表が 3 つでき、log-rank test・Generalized Wilcoxon test 共に、この 3 つのクロス表を合わせて、両群の生存関数に違いがあるのかどうかを検討する方法。(時点ごとの死亡発生例数 d₃, d₄, d₈ は全て独立と仮定。)

log-rank test

ここで、各時点の死亡例数について平等に足し算をした

$$d_{\log} = d_3 + d_4 + d_8$$

を用いて構成されるのが log-rank test。

$$d_{\log} = 0 + 0 + 1 = 1$$

であり、期待値・分散は独立性から両方とも足し算で求まり、

$$E[d_{\log}] = E[d_3 + d_4 + d_8] = E[d_3] + E[d_4] + E[d_8]$$

$$= e_3 + e_4 + e_8 = 4/7 + 4/6 + 2/3 = 12/21 + 14/21 + 14/21 = 40/21 = 1.9048$$

$$V[d_{\log}] = V[d_3 + d_4 + d_8] = V[d_3] + V[d_4] + V[d_8] \quad (* d_3, d_4, d_8 \text{ は独立})$$

$$= 12/49 + 2/9 + 2/9 = 304/411 = 0.6893$$

となる。これを用いて検定統計量

$$\chi^2_{\log} = (d_{\log} - E[d_{\log}])^2 / V[d_{\log}]$$

$$= (1 - 1.9048)^2 / 0.6893 \approx 1.1877$$

を求める。2 群の場合、帰無仮説のもとで χ^2_{\log} が漸近的に自由度 1 の χ^2 分布に従うことを用いて検定を行う。つまり、有意水準 5% の両側検定の場合、

$\chi^2_{\log} > \chi^2(1, 0.95)$ のときに帰無仮説が棄却される。今回は $\chi^2_{\log} = 1.1877$, $\chi^2(1, 0.95) = 3.8415$ より、帰無仮説は棄却されない。

log-rank test・Generalized Wilcoxon test の違いと使い分け

時点ごとに d₃, d₄, d₈ の重みを変更するのが Generalized Wilcoxon test。「最初は例数が多いから信頼性の高い値になるが、後になると例数が減るから信頼性の低い値になる」と考える。例えば、100 例生存していた中から 20 例死亡すると、「大体 2 割」とある程度自信を持って言えそうだが、5 例中 1 例死亡しても「大体 2 割？」と信頼性が低下する。そのため、例数の多く残っている前の方の時点に大きい重みを、例数の少なくなる後の方の時点に小さい重みをつける。結果として、「時間が経てば経つほど、群間差が開いてくる」タイプの薬剤に対しては Generalized Wilcoxon test よりも log-rank test の方が有意差が付きやすい。一方、「結局はほぼ全員が死亡するのだが、生存時間が延びる」タイプの薬剤では、Generalized Wilcoxon test の方が差が付きやすくなる。

Generalized Wilcoxon test の構成

Generalized Wilcoxon test では、重みとして「両群合わせた at risk number」を用いる。log-rank test の d_{log} の代わりに $d_{\text{Wil}} = N_3d_3 + N_4d_4 + N_8d_8$ を用いて検定を行なう。

今、 $d_{wil} = 7 \cdot 0 + 6 \cdot 0 + 3 \cdot 1 = 3$ であり、期待値・分散は
 $E[d_{wil}] = E[N_3d_3 + N_4d_4 + N_8d_8] = N_3E[d_3] + N_4E[d_4] + N_8E[d_8]$
 $= N_3e_3 + N_4e_4 + N_8e_8 = 7 \cdot 7 \cdot 4/7 + 6 \cdot 4/6 + 3 \cdot 2/3 = 4 + 4 + 2 = 10$
 $V[d_{wil}] = V[N_3d_3 + N_4d_4 + N_8d_8] = N_3^2 V[d_3] + N_4^2 V[d_4] + N_8^2 V[d_8]$
 (* d_3, d_4, d_8 は独立)
 $= N_3^2 v_3 + N_4^2 v_4 + N_8^2 v_8$
 $= 7^2 \cdot 12/49 + 6^2 \cdot 8/36 + 3^2 \cdot 2/9 = 12 + 8 + 2 = 22$ となる。

これより、検定統計量は

$\chi^2_{wil} = (d_{wil} - E[d_{wil}])^2 / V[d_{wil}] = (3-10)^2 / 22 = 2.2273$ となる。

log-rank testと同様、帰無仮説のもとで χ^2_{wil} は漸近的に自由度 1 の χ^2 分布に従うので、有意水準 5%の両側検定は $\chi^2_{wil} > \chi^2(1, 0.95)$ のときに帰無仮説を棄却すればよい。今回のデータでは、 $\chi^2_{wil} = 2.2273$, $\chi^2(1, 0.95) = 3.8415$ より、帰無仮説は棄却されない。

まとめ

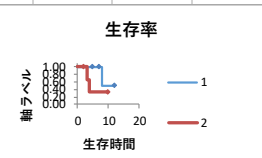
log-rank test・Generalized Wilcoxon test の両方とも生存関数を比較するノンパラメトリック検定である。時点の情報は、死亡の発生時点での例数しか用いていない(死亡もしくはセンサーの発生時間そのもの、発生時間の間隔等は無視している)。log-rank test は、どの時点での死亡の発生も同等と考える。Generalized Wilcoxon test は「最初の方が例数が多いから信頼性が高い。後の方は例数が少ないから信頼性が低い」として、時点ごとの発生例数に対して、その時点の両群合わせた at risk number で重みづけをする。

BellCurve for Excel: 生存分析・ハザード分析→カプラン=マイヤー法

被験者番号	投与群	生存時間	死亡 (1)	センサー (0)
1	1	1	5	0
2	1	1	7	0
3	1	1	8	1
4	1	1	12	0
5	2	2	2	0
6	2	2	3	1
7	2	2	4	1
8	2	2	10	0

計算と同値
 $\chi^2_{log} = 1.1875$
 $\chi^2_{wil} = 2.2273$

カプラン=マイヤー法による生存分析									
ケースの要約									
系 列	死 亡	打ち切り	全 体						
1	1	3	4						
2	2	2	4						
生存時間の平均値と中央値									
系 列	平均値	標準誤差	95%信頼区間		中央値	標準誤差	95%信頼区間		
1	10.0000	1.4142	7.2282	12.7718	8.0000	0.0000	8.0000	8.0000	
2	5.6667	1.7847	2.1687	9.1646	4.0000	0.8165	2.3997	5.6003	
生存率曲線の差の検定									
手 法		カイ二乗値	自由度	P 値	*: P<0.05 **: P<0.01				
ログランク検定	Peto-Peto	1.1772	1	0.2779					
	Cochran-Mantel-Haenszel	1.1875	1	0.2758					
一般化Wilcoxon検定	Gehan-Breslow	2.2273	1	0.1356					
	Peto-Prentice	1.7245	1	0.1891					



Cox 比例ハザードモデル

臨床研究では対象はヒトであり、各個体で年齢、性別、食習慣、治療の有無などまちまちである。これを解析するために、Kaplan-Meier法を利用し、複数の予後規定因子とハザードを関連づけるSemi-parametric modelが比例ハザードモデルである。

$S(t)$:生存率(生存時間)関数

T :生存期間(死亡までの時間)を示す確率変数

$$P(T \geq t) = S(t)$$

$F(t)$: T の分布関数=死亡関数(死亡が起こるまでの時間の累積分布関数)

$$F(t) = 1 - S(t)$$

$f(t)$: T の確率密度関数

$$f(t) = \frac{d}{dt} F(t) = \frac{d}{dt} (1 - S(t)) = -\frac{d}{dt} S(t) \quad \text{と定義する。}$$

逆に

$$F(t) = p(T \leq t) = \int_{-\infty}^t f(x) dx = \int_0^t f(x) dx$$

$$S(t) = p(T > t) = 1 - F(t) = \int_t^{\infty} f(x) dx$$

$$\frac{d}{dt} S(t) = \frac{d}{dt} [1 - F(t)] = -f(t) \quad \Rightarrow f(t) = -\frac{d}{dt} S(t) \quad \text{と表すこともできる。}$$

「 t の直前まで生存したという条件で、人が続く Δt の期間に死亡する確率(条件付き確率)」 $P(t \leq T < t + \Delta t | T \geq t) \Rightarrow$ この部分は観察可能なデータ

これをを Δt で割った値

$\frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$ は単位時間当たりの平均死亡率を表す。

(イメージとして年齢別死亡率 \equiv ハザード)

$\Delta t \rightarrow 0$ としたときの極限を $h(t)$ とすると

(事象 A が起きたもとで事象 B が起きる確率 $P(B|A) = \frac{P(A \cap B)}{P(A)}$ より \Rightarrow 参考p140)

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{p(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{p\{(t \leq T < t + \Delta t) \cap (T \geq t)\}}{p(T \geq t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \frac{p(t \leq T < t + \Delta t)}{p(T \geq t)} = \frac{1}{p(T \geq t)} \lim_{\Delta t \rightarrow 0} \frac{p(t \leq T < t + \Delta t)}{\Delta t} \\ &= \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t} = \frac{1}{S(t)} \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \\ &= \frac{1}{S(t)} \frac{d}{dt} F(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log\{S(t)\} \\ &\Rightarrow \log S(t) = -\int h(t) dt \quad (\dots \text{ 計算式1}) \end{aligned}$$

$h(t)$:ハザード関数 (hazard function)

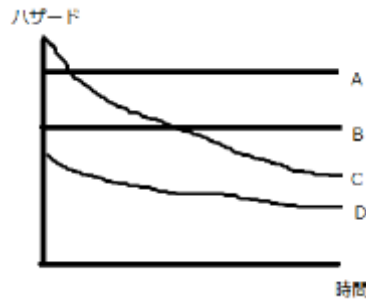
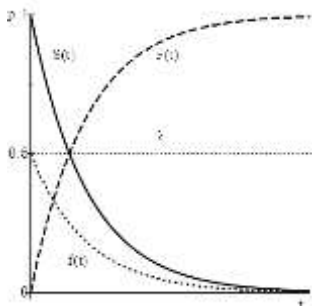
「 t 時点直前まで死亡(イベント)が起こらなかったという条件の下で、 t 時点でイベントが起こる瞬間死亡率」を表す。 $h(t)$:確率ではないので1より大にもなり得る。

$$h(t) = \frac{f(t)}{S(t)} = -\frac{\frac{dS(t)}{dt}}{S(t)}$$

モデルとして $S(t) = \exp(-\lambda t)$ (ポアソンモデル、標的モデル) を考える。

鉄砲を構えた的に向かって100人の軍隊が突撃するとき、1時間当たりの敵の弾の命中率を50%とすると、一時間後には50人死亡し50人生存する。2時間後には25人死亡し25人が生存する。瞬間死亡率は常に50% (=hazard)

$$\text{ハザード関数 } h(t) = -\frac{\frac{dS(t)}{dt}}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda = 0.5 \text{ (一定)}$$



この場合はハザード関数は定数となる。

ハザード比(hazard ratio): 2群のハザードの比

各群のハザードは年齢別死亡率と同様、時間 t の関数であり t と共に変化する。上図 A と B 関係ではハザードは定数なので、例えば $h_A(t) = 0.5$ $h_B(t) = 0.2$ とすればハザード比 $= 0.5 / 0.2 = 2.5$ となる。

C と D では時間 t と共にハザードが減少するが、その比(ハザード比) $\frac{h_C(t)}{h_D(t)}$ が一定であれば時間 t の関数にならず、その値を特定できる。

Cox 比例ハザードモデルは

$$h(t, X) = h_0(t) \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n) \text{ で定義される。}$$

前提条件)

- ① ハザード比(hazard ratio)が各時点で一定(proportionality)、
 - ② 共変量のみ依存し特定の時間には依存せず、共変量是对数をとると線形になる(対数線形性)
- ①、②のとき上図 A と B 関係だけではなく C と D の様な関係にも適用できる。

ここで

$h(t, X)$: 比較する群の瞬間死亡率(hazard)
時間 t においては生存し、 $t+dt$ に死亡する割合

$h_0(t)$ (baseline hazard 関数) : 時間 t に依存する部分、どのような関数でも良く、パラメーターとして定義されていない。対照群の個体の中での平均的な個体の瞬間

死亡率(baseline hazard)で、説明変数 $X_1, X_2 \dots X_n$ が全て 0 となったときのハザード関数のこと

$\exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)$ 相対危険度関数(relative risk function): 共変量のみ依存し、特定の時間には依存しない部分。パラメーターとしてモデル化されている。このモデルは Semi-parametric model と呼ばれる。

* データはハザードであり、 $h(t) = \frac{f(t)}{S(t)} = -\frac{\frac{dS(t)}{dt}}{S(t)}$ であるので、kaplan-Meier

推定量を求め代入すればハザードの推定値を得ることができる理屈。

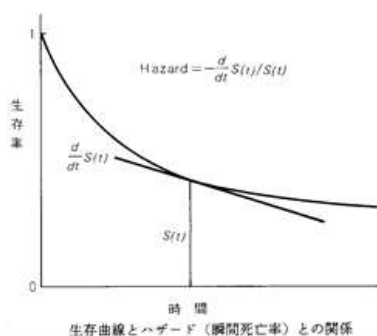
* 2 群のハザード比を取ると $h_0(t)$ は約分されて式からなくなる。

($h_0(t)$ は推定する必要がない)

解析は手計算では困難なので解析ソフトによるが、解釈するためには幾つかの知識を整理する必要がある。

$S(t)$ 累積生存率(生存率): 開始時 1 にした場合の時間 t での生存者の割合、とすると

$h(t) = -\frac{\frac{dS(t)}{dt}}{S(t)}$: 時間 t の累積生存率を 1 とした時の生存曲線の傾きの絶対値に相当する。



Cox 比例ハザードモデルは、治療法や曝露要因のイベント発生への影響を検討する方法の 1 つ。治療法や年齢、性別などの各共変量が、ハザード関数 $h(t, X)$ に影響をおよぼしているか否かを検討できる。式中の β が共変量に対応する偏回帰係数、 x が共変量を表し、ハザード分析では、偏回帰係数 β を指数変換したハザード比に注目する。偏回帰係数が 5% の有意水準で有意な場合、ハザード比の 95% 信頼区間は 1 を含まない。ハザード比は相対リスク (RR) に一致し、1 より小さいときはリスクを下げ、1 より大きいときはリスクを上げる。

- 統計ソフトを利用し変数選択をする際、「変数増減法」、「変数減増法」、「変数増加法」、「変数減少法」の 4 種類の方法がある。変数選択を利用することで、指定された有意水準の元でハザード関数に与える影響が大きい共変量が自動的に選択される。

生存率曲線では各共変量の最小値を設定した「値 A」と最大値を設定した「値 B」の 2 つの生存率曲線を 1 つのグラフで出力される。

- モデル診断のための **log-log (ログマイナスログ)プロット**:
上記のハザード関数が真のハザード関数であるかどうかを確認するために、log-log (ログマイナスログ)プロットを用いる。**log-log** プロットは横軸が t 、縦軸が $\log\{-\log S(t)\}$ のグラフで、グラフ内の 2 本の曲線が交わることがなく、上下に平行移動した状態であれば、上式を真のハザード関数とみなせる。各ケースについてハザード比を求め、ハザード比の中央値を基準に全体を 2 群に分けて 2 本の曲線が描かれる。

上記計算式 1 $\log S(t) = -\int h(t)dt$ に

$h(t|x_1=1) = h_0(t)\exp\beta_1$ $h(t|x_1=0) = h_0(t)$ を代入すると

$\log S(t|x_1=1) = -\int h_0(t)\exp\beta_1 dt = -\exp\beta_1 \int h_0(t)dt$

$\log S(t|x_1=0) = -\int h_0(t)dt$

マイナス符号入れ替えてをさらに log をとると

$\log(-\log S(t|x_1=1)) = \log(\int h_0(t)dt) + \beta_1$

$\log(-\log S(t|x_1=0)) = \log(\int h_0(t)dt)$

Kaplan-Meier 法で $S(t|x_1=1)$ 、 $S(t|x_1=0)$ を計算しプロットすれば平行でその差は β_1 となる。

2 群を比較する場合

前提) 両群のハザードは、共変量 x_1 の値を除いて等しい

例えば実薬群 $x_1 = 1$ 、プラセボ群 $x_1 = 0$

この時

$$h(t|x_1 = 1) = h_0(t)\exp(\beta_1 * 1 + \beta_2 x_2 + \dots + \beta_n x_n)$$

$$h(t|x_1 = 0) = h_0(t)\exp(\beta_1 * 0 + \beta_2 x_2 + \dots + \beta_n x_n)$$

$$\frac{h(t|x_1 = 1)}{h(t|x_1 = 0)} = \frac{h_0(t)\exp(\beta_1 * 1 + \beta_2 x_2 + \dots + \beta_n x_n)}{h_0(t)\exp(\beta_1 * 0 + \beta_2 x_2 + \dots + \beta_n x_n)} = \exp(\beta_1)$$

x_1 がダミー変数 (0 or 1) である場合

$$h(t) = h_0(t) \times \exp\{\beta_1 \times (x_1 = \text{薬剤})\}$$

Cox 回帰分析の結果、モデルが以下の様に推定されたとする

$$h(t) = h_0(t) \times \exp\{2.0 \times \text{薬剤}\} \quad (\text{薬剤 A} \rightarrow x_1 = 1, \text{薬剤 B} \rightarrow x_1 = 0 \text{ とする})$$

$$\text{薬剤 A のハザード: } h_A(t) = h_0(t) \times \exp\{2.0 \times 1\} = h_0(t) \times \exp(2.0)$$

$$\text{薬剤 B のハザード: } h_B(t) = h_0(t) \times \exp\{2.0 \times 0\} = h_0(t)$$

$$\text{ハザード比: } h_A(t) / h_B(t) = \{h_0(t) \times \exp(2.0)\} \div h_0(t) = \exp(2.0)$$

$h_0(t)$ の値を推定しなくても、「薬剤 B に対する薬剤 A のハザード比」が $\exp(2.0) = 7.4$ と決まり、「ハザード比」は 7.4 (薬剤 B に対する薬剤 A の改善割合は 7.4 倍) と解釈することが出来る。

薬剤間のハザード比に対する「p」の意味: H_0 : 薬剤間のハザード比が 1 かどうかの検定結果は「p:0.07」となったとすると、5% よりも大きいので「帰無仮説が間違っているとはいえない」と結論付ける

x₁ が離散量の場合

例えば x₁ が病期で I 期～IV期である場合

$$x_1 = k (=1, 2, 3 \text{ or } 4)$$

$h_k(t) / h_1(t) = \exp\{\beta_1 \times (k-1)\}$ は k=1 を基準にした k(2, 3, or 4) 期のハザード比となる。

x₁ が連続量の場合

例えば WBC 5000, 1000 でのハザードを計算する場合

$$\frac{h(t|x_1 = 5000)}{h(t|x_1 = 1000)} = \frac{h_0(t)\exp(\beta_1 \cdot 5000 + \beta_2 x_2 + \dots + \beta_n x_n)}{h_0(t)\exp(\beta_1 \cdot 1000 + \beta_2 x_2 + \dots + \beta_n x_n)} = \exp(\beta_1 * (5000 - 1000)) \text{ となる。}$$

別例 (A 群、B 群で x₁ = 罹病期間のとき) で

$$h(t) = h_0(t) \times \exp\{\beta_1 \times \text{罹病期間(年)}\} \text{ において}$$

$$h(t) = h_0(t) \times \exp\{-0.207 \times \text{罹病期間}\} \text{ となったとすると}$$

罹病期間が 1 年 (1 単位) 増えた時の対数ハザード比 = -0.207

罹病期間が 1 年 (1 単位) 増えた時のハザード比 = $\exp(-0.207) = 0.812$

罹病期間が 5 年 (5 単位) 増えた時の対数ハザード比 = $-0.207 \times 5 = -1.03$

罹病期間が 5 年 (5 単位) 増えた時のハザード比 = $\exp(-1.03) = 0.357$

(罹病期間が 5 年増えた時のハザード比は「 $\exp(-0.207) \times 5$ 」ではなく

「 $\exp(-0.207 \times 5)$ 」となる)

- x_2, \dots, x_n についても同様に β_2, \dots, β_n を推定するが (部分尤度法という方法を用いた複雑な計算なので) コンピューターに任せることになる。

例題) ヘロイン中毒患者へメタドン投与治療を行ったデータがある。「治療状況」は 1 が治療からの脱落を、0 が打ち切りを示している。このデータについて Cox 比例ハザードモデルを行い、「メタドン治療から脱落する(治療を放棄する)リスク」についての解析を行う。

ヘロイン中毒患者へメタドン投与治療

治療状況	治療期間	治療施設	服役有無	メタドンの投与量	治療状況	治療期間	治療施設	服役有無	メタドンの投与量
1	428	1	0	50	0	613	1	1	60
1	275	1	1	55	1	192	1	1	80
1	262	1	0	55	0	405	1	0	80
1	183	1	0	30	1	667	1	0	50
1	259	1	1	65	0	905	1	0	80
1	714	1	0	55	1	247	1	0	70
1	438	1	1	65	1	821	1	0	80
0	796	1	1	60	1	821	1	1	75
1	892	1	0	50	0	517	1	0	45
1	393	1	1	65	0	346	1	1	60
0	161	1	1	80	1	294	1	0	65
1	836	1	1	60	1	244	1	1	60
1	523	1	0	55	1	95	1	1	60

1	612	1	0	70	1	376	1	1	55
1	212	1	1	60	1	212	1	0	40
1	399	1	1	60	1	96	1	0	70
1	771	1	1	75	1	532	1	0	80
1	514	1	1	80	1	522	1	1	70
1	512	1	0	80	1	679	1	0	35
1	624	1	1	80	0	408	1	0	50
1	209	1	1	60	0	840	1	0	80
1	341	1	1	60	0	148	1	1	65
1	299	1	0	55	1	168	1	0	65
0	826	1	0	80	1	489	1	0	80
1	262	1	1	65	0	541	1	0	80
0	566	1	1	45	1	205	1	0	50
1	368	1	1	55	0	475	1	1	75
1	302	1	1	50	1	237	1	0	45
0	602	1	0	60	1	517	1	0	70
1	652	1	0	80	1	749	1	0	70
1	293	1	0	65	1	150	1	1	80
0	564	1	0	60	1	465	1	0	65
1	394	1	1	55	1	708	2	1	60
1	755	1	1	65	0	713	2	0	50
1	591	1	0	55	0	146	2	0	50
0	787	1	0	80	1	450	2	0	55
1	739	1	0	60	0	555	2	0	80
1	550	1	1	60	1	460	2	0	50
1	837	1	0	60	0	53	2	1	60
1	612	1	0	65	1	122	2	1	60
0	581	1	0	70	1	35	2	1	40
1	523	1	0	60	0	532	2	0	70
1	504	1	1	60	0	684	2	0	65
1	785	1	1	80	0	769	2	1	70
1	774	1	1	65	0	591	2	0	70
1	560	1	0	65	0	769	2	1	40
1	160	1	0	35	0	609	2	1	100
1	482	1	0	30	0	932	2	1	80
1	518	1	0	65	0	932	2	1	80
1	683	1	0	50	0	587	2	0	110
1	147	1	0	65	1	26	2	0	40
1	563	1	1	70	0	72	2	1	40
1	646	1	1	60	0	641	2	0	70
1	899	1	0	60	0	367	2	0	70
1	857	1	0	60	0	613	1	1	60
1	180	1	1	70	1	192	1	1	80
1	452	1	0	60					
1	760	1	0	60					
1	496	1	0	65					

1	258	1	1	40					
1	181	1	1	60					
1	386	1	0	60					
0	439	1	0	80					
0	563	1	0	75					
1	337	1	0	65					

BellCurve for Excel: 生存分析・ハザード分析→Cox比例ハザードモデル

アイテム	カテゴリー	コード	
治療状況	治療脱落		1
	打ち切り		0
服薬有無	あり		1
	なし		0
治療期間			⇒時間
治療状況 = censored or not			⇒状態
治療施設	服薬有無		⇒共変量
メタドンの投与量			

Cox比例ハザードモデル							
状態の要約							
治療状況	n	%					
治療状況=0	36	30.25%					
治療状況=1	83	69.75%					
全 体	119	100.00%					
基本統計量							
治療状況	変 数	n	平 均	不偏分散	標準偏差	最小値	最大値
全 体	治療期間	119	485.849	55863.452	236.355	26.000	932.000
	治療施設	119	1.185	0.152	0.390	1.000	2.000
	服薬有無	119	0.420	0.246	0.496	0.000	1.000
	メタドンの投与量	119	63.151	181.723	13.480	30.000	110.000
治療状況=0	治療期間	36	556.528	55548.771	235.688	53.000	932.000
	治療施設	36	1.444	0.254	0.504	1.000	2.000
	服薬有無	36	0.389	0.244	0.494	0.000	1.000
	メタドンの投与量	36	68.611	242.302	15.566	40.000	110.000
治療状況=1	治療期間	83	455.193	53534.621	231.375	26.000	899.000
	治療施設	83	1.072	0.068	0.261	1.000	2.000
	服薬有無	83	0.434	0.249	0.499	0.000	1.000
	メタドンの投与量	83	60.783	139.318	11.803	30.000	80.000

回帰式に含まれる共変量(変数選択過程)増減法により変数選択を行った結果、回帰式へ投入された共変量は各ステップで除去基準の P 値 0.200 より小さかったため、いずれの共変量も回帰式から除去されず。

回帰式に含まれない共変量(変数選択過程)増減法により変数選択を行った結果、回帰式に含まれない共変量のうち、投入基準の P 値 0.200 より小さい共変量が順次各ステップで回帰式へ投入された。

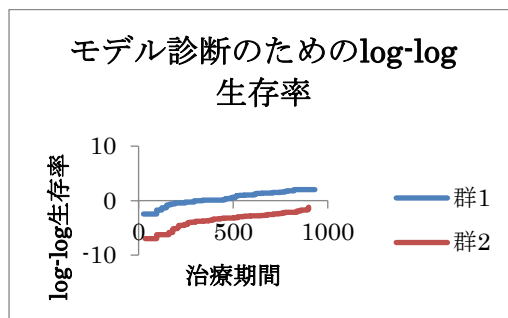
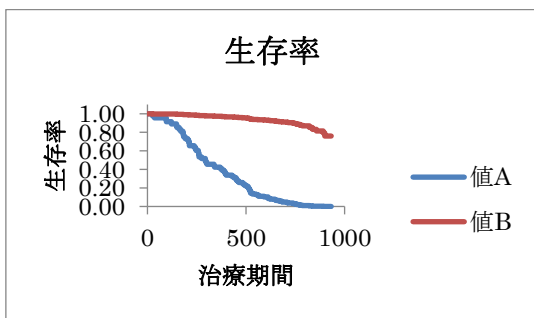
変数選択の方法	投入基準P値	除去基準P値								
増減法	0.200	0.200								
変数選択過程										
モデル										
			全体(スコア検定)			前のステップからの変化			変数選択の結果	
ステップ	-2対数尤度	AIC	カイ二乗値	自由度	P 値	カイ二乗値	自由度	P 値	投入	除去
ステップ0	654.5647	654.5647	-			-			メタドンの投与量	
ステップ1	639.6233	641.6233	14.8756	1	p < 0.001	14.9414	1	p < 0.001	治療施設	
ステップ2	631.3990	635.3990	20.0297	2	p < 0.001	8.2242	1	0.0041	服役有無	
ステップ3	627.5840	633.5840	23.1159	3	p < 0.001	3.8150	1	0.0508	(なし)	

回帰式に含まれる共変量			Wald検定			ハザード比		95%信頼区間	
ステップ	共変量	係数	標準誤差	カイ二乗値	自由度	P 値	Exp(係数)	下限値	上限値
ステップ0	(なし)								
ステップ1	メタドンの投与量	-0.0334	0.0086	15.0125	1	p < 0.001	0.9672	0.9510	0.9837
ステップ2	治療施設	-1.0525	0.4259	6.1068	1	0.0135	0.3490	0.1515	0.8043
	メタドンの投与量	-0.0320	0.0089	12.8053	1	p < 0.001	0.9686	0.9518	0.9857
ステップ3	治療施設	-1.1739	0.4317	7.3953	1	0.0065	0.3092	0.1327	0.7205
	服役有無	0.4538	0.2300	3.8929	1	0.0485	1.5743	1.0030	2.4709
	メタドンの投与量	-0.0345	0.0092	13.9727	1	p < 0.001	0.9661	0.9488	0.9837

回帰式に含まれない共変量	スコア検定			
ステップ	共変量	カイ二乗値	自由度	P 値
ステップ0	治療施設	8.5309	1	0.0035
	服役有無	0.9351	1	0.3335
	メタドンの投与量	14.8756	1	p < 0.001
ステップ1	治療施設	6.6785	1	0.0098
	服役有無	1.9680	1	0.1607
ステップ2	服役有無	3.9533	1	0.0468
ステップ3	(なし)			

「治療施設」は1より2の方が、「服役歴」がはある(=1)よりない(=0)の方が、「メタドンの投与量」は多いほうが治療が継続されるという結果。いずれの変数も P<0.05、ハザード比の95%信頼区間に1を含まない。

変数選択結果										
モデル	全体(スコア検定)			前のステップからの変化						
	-2対数尤度	AIC	カイ二乗値	自由度	P 値	カイ二乗値	自由度	P 値		
	627.5840	633.5840	23.1159	3	p < 0.001	3.8150	1	0.0508		
回帰式に含まれる共変量			Wald検定			*: P<0.05		ハザード比		95%信頼区間
共変量	係数	標準誤差	カイ二乗値	自由度	P 値	** : P<0.01	Exp(係数)	下限値	上限値	
治療施設	-1.1739	0.4317	7.3953	1	0.0065**		0.3092	0.1327	0.7205	
服役有無	0.4538	0.2300	3.8929	1	0.0485*		1.5743	1.0030	2.4709	
メタドンの投与量	-0.0345	0.0092	13.9727	1	p < 0.001**		0.9661	0.9488	0.9837	



Acute leukemia (多変量解析による臨床研究、浜島信之著より一部改変)

AMLのデータ					
(多変量解析による臨床研究、浜島信之著より一部改変)					
性	年齢	治療法	生存期間	生死	
0	62	0	2	1	
1	57	0	19	0	
0	49	1	5	1	
0	72	1	0	1	
1	63	0	3	0	
0	51	0	17	0	
1	68	1	2	1	
0	40	1	3	0	
1	38	0	15	0	
1	67	1	14	0	
0	81	1	12	1	
0	54	0	13	1	
0	57	0	11	0	
1	63	0	11	0	
0	48	1	10	0	
1	35	1	8	1	
0	62	1	8	0	
0	59	0	3	0	
1	75	0	7	1	
0	71	1	15	1	
0	60	0	6	0	
0	77	1	2	1	
1	42	0	5	0	
0	54	1	4	0	
1	63	0	3	1	
0	59	1	3	0	
0	72	1	1	1	
1	44	0	4	0	
0	67	0	2	0	
0	50	0	1	0	

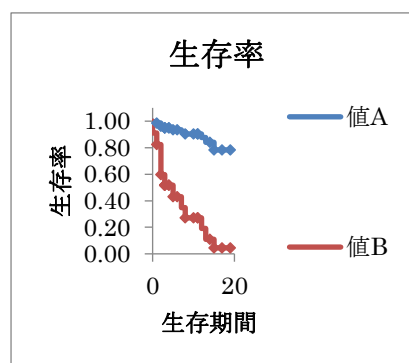
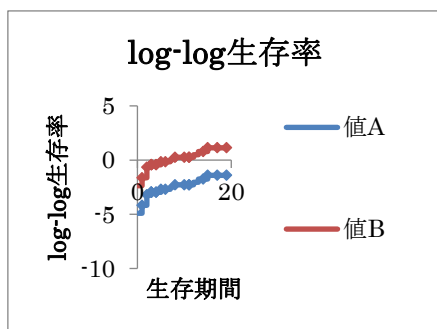
BellCurve for Excel: 生存分析・ハザード分析→Cox比例ハザードモデル

Cox比例ハザードモデル							
ケースの要約							
	n	%					
有効ケース	30	100.00%					
時間のみ不明	0	0.00%					
状態のみ不明	0	0.00%					
共変量のみ不明	0	0.00%					
時間のみ有効	0	0.00%					
状態のみ有効	0	0.00%					
共変量のみ有効	0	0.00%					
いずれも不明	0	0.00%					
全 体	30	100.00%					
状態の要約							
生死	n	%					
生死=0	18	60.00%					
生死=1	12	40.00%					
全 体	30	100.00%					
基本統計量							
生死	変 数	n	平 均	不偏分散	標準偏差	最小値	最大値
全 体	生存期間	30	6.967	29.413	5.423	0.000	19.000
	性	30	0.367	0.240	0.490	0.000	1.000
	年齢	30	58.667	142.575	11.940	35.000	81.000
	治療法	30	0.467	0.257	0.507	0.000	1.000
生死=0	生存期間	18	7.722	31.624	5.624	1.000	19.000
	性	18	0.389	0.252	0.502	0.000	1.000
	年齢	18	54.500	83.559	9.141	38.000	67.000
	治療法	18	0.333	0.235	0.485	0.000	1.000
生死=1	生存期間	12	5.833	26.333	5.132	0.000	15.000
	性	12	0.333	0.242	0.492	0.000	1.000
	年齢	12	64.917	175.720	13.256	35.000	81.000
	治療法	12	0.667	0.242	0.492	0.000	1.000

変数選択の方法										
全変数										
モデル										
		全体 (スコア検定)				前のステップからの変化				
-2対数尤度	AIC	カイ二乗値	自由度	P 値	カイ二乗値	自由度	P 値			
59.8629	65.8629	5.8204	3	0.1207	5.6129	3	0.1320			
回帰式に含まれる共変量										
			Wald検定				* : P<0.05	ハザード比	95%信頼区間	
共変量	係 数	標準誤差	カイ二乗値	自由度	P 値	** : P<0.01	Exp(係数)	下限値	上限値	
性	0.0165	0.6574	0.0006	1	0.9800		1.0166	0.2803	3.6873	
年齢	0.0433	0.0281	2.3678	1	0.1239		1.0442	0.9882	1.1034	
治療法	0.6194	0.7054	0.7711	1	0.3799		1.8578	0.4662	7.4028	

変数選択の方法	投入基準P値	除去基準P値								
増減法	0.200	0.200								
変数選択過程										
モデル										
			全体 (スコア検定)			前のステップからの変化			変数選択の結果	
ステップ	-2対数尤度	AIC	カイ二乗値	自由度	P 値	カイ二乗値	自由度	P 値	投入	除去
ステップ0	65.4758	65.4758	-			-			年齢	
ステップ1	60.7017	62.7017	4.5918		1	0.0321	4.7742	1	0.0289	(なし)
回帰式に含まれる共変量										
			Wald検定			ハザード比			95%信頼区間	
ステップ	共変量	係 数	標準誤差	カイ二乗値	自由度	P 値	Exp(係数)	下限値	上限値	
ステップ0	(なし)									
ステップ1	年齢	0.0552	0.0264	4.3856		1	0.0362	1.0568	1.0035	1.1128
回帰式に含まれない共変量										
		スコア検定								
ステップ	共変量	カイ二乗値	自由度	P 値						
ステップ0	性	0.4507	1	0.5020						
	年齢	4.5918	1	0.0321						
	治療法	2.9303	1	0.0869						
ステップ1	性	0.0591	1	0.8079						
	治療法	0.8381	1	0.3599						

変数選択結果										
モデル										
		全体 (スコア検定)			前のステップからの変化					
-2対数尤度	AIC	カイ二乗値	自由度	P 値	カイ二乗値	自由度	P 値			
60.7017	62.7017	4.5918		1	0.0321	4.7742	1	0.0289		
回帰式に含まれる共変量										
		Wald検定			*: P<0.05			ハザード比	95%信頼区間	
共変量	係 数	標準誤差	カイ二乗値	自由度	P 値	** : P<0.01	Exp(係数)	下限値	上限値	
年齢	0.0552	0.0264	4.3856		1	0.0362*	1.0568	1.0035	1.1128	



統計を意識して論文を読むとき(研究の位置づけの概略)

医学論文				統計的指標
原著 Regular papers, original articles	臨床研究 Clinical investigations and reports	介入研究 Intervention study	(ア) 対照のない臨床試験 (イ) 対照のある臨床試験(比較試験) 1) 歴史的対照のある比較試験 2) 同時対照(並行群)のある比較試験 a. 無作為化(ランダム化)しない比較試験 b. 無作為化比較試験 Randomized controlled trials(RCT) (1) 非盲検無作為化比較試験 (2) 盲検無作為化比較試験 ①単盲検無作為化比較試験 ②二重盲検無作為化比較試験	
		観察研究 Observational study	横断研究 Cross-sectional study	
	縦断研究 Longitudinal study		(前向き)コホート研究 Prospective cohort study	incidence(発生率) relative risk (risk ratio, R.R)
			ケース・コントロール研究 Case control study	Odds ratio (R.R の近似値)
	基礎研究 Basic science reports	動物実験 Animal experiments	実験計画法 Design of experiments	
解析) 「結果」側の変数が、 量的変数(比例尺度・間隔尺度)の場合:t 検定・回帰分析・分散分析・共分散分析 名義尺度の場合: χ^2 (カイ自乗)検定・Fisher の正確確率検定・Mantel-Haenszel 検定・ロジスティック回帰 順序尺度の場合:Wilcoxon 検定・Mantel-Haenszel 検定・比例オッズモデル 生存時間の場合:Logrank 検定・一般化 Wilcoxon 検定・Cox 回帰(比例ハザードモデル) 等が使用される				

効果量と必要サンプル数の計算

効果量(effect size)

例えば p 値について、サンプル・サイズが大きくなればなるほど、実質的な差がなかった場合でも、 p 値は小さくなり、統計的に有意であるという結果が得られやすくなるという問題がある。。そこで、サンプル・サイズによって変化しない、標準化された指標である効果量(effect size)が役立つ。「2 グループの平均値の差を標準化した効果量」である Cohen's d は、 t 検定のような平均値の差を比較するときに使用する。

サンプル・サイズ、有意水準(α)、検定力($1-\beta$)、効果量の 4 つは、他の 3 つが決まれば残りの 1 つが決まるという関係である。推奨される検定力($1-\beta$)は 0.8, 有意水準(α)は 0.05 と決まっているので、統計的検定を用いる研究を計画する際に実際に考慮しなければならないのは、サンプル・サイズと効果量になる。

効果量には 70 以上の種類がある : d 族: 群間差についての効果量、Cohen の d 、Hedges の g 、Glass の Δ (デルタ) : r 族: 変数間の関係の大きさの効果量、相関係数 r 、分散分析の効果量 η^2 (イータ二乗)など。

効果量の評価					
検定	指標	効果量の基準			補足
		小	中	大	
差の検定(t 検定)	r	0.1	0.3	0.5	対応のある t 検定、2 標本の差の検定で同一値. r と d の 2 種類ある
	d	0.2	0.5	0.8	
分散分析	η^2	0.01	0.06	0.14	多重比較法は差の検定(t 検定)の r を利用
相関	r	0.1	0.3	0.5	相関係数そのままである
カイ二乗検定(2×2 分割表)	ϕ	0.1	0.3	0.5	連関係数である
カイ二乗検定(上記以外の分割表)	Cramer の V	0.1	0.3	0.5	連関係数である
差の検定(ノンパラメトリック検定)	r	0.1	0.3	0.5	マンホイットニー検定、ウィルコクソン検定、クラスカルワリス検定、フリードマン検定で求められる検定統計量 Z を $Z = r\sqrt{n}$ として求める
重回帰分析	R^2	0.02	0.13	0.26	決定係数

水本 篤, 竹内 理, "研究論文における効果量の報告のために—基礎概念と注意点—", 『英語教育研究』31(2008), 57-66 より引用

その他解析結果によく登場する効果量

ユールの Q(Yule's Q)

2 行×2 列のクロス集計表の各セルの値を以下のように a, b, c, d を用いて表したとき、下式により Q を求めることができる

	B_1	B_2	計
A_1	a	b	r_1

$$Q = \frac{ad - bc}{ad + bc} \quad \begin{array}{cccc} A_2 & c & d & r_2 \\ \text{計} & c_1 & c_2 & n \end{array}$$

$-1 \leq Q \leq 1$ の値をとり、絶対値が 1 に近いほど関連が強い。

2 群の比較(2 標本の差)をするときの必要サンプル数の計算

- 1) 検出力 $1 - \beta \Rightarrow 0.8 \Rightarrow$ この時の Z 値 0.84
- 2) 有意水準 $\alpha \Rightarrow 0.05 \Rightarrow$ この時の Z 値 1.96
- 3) 分散 σ^2

Cohen の d (2 群のサンプルサイズが異なる場合)

$$\sigma = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}}$$

Hedges の g の場合

$$\sigma = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

- 4) 標準化平均差 SMD (= 効果サイズ ES): $ES = \frac{\delta}{\sigma}$ (検出したい差 δ)
- 5) 1 群当たりの症例数 n

Lehr's formula (片側検定の場合)

$$n = 2 * (Z_\alpha + Z_\beta)^2 * \frac{\sigma^2}{\delta^2} = 15.68 \left(\frac{\sigma}{\delta}\right)^2 \doteq 16 \div \left(\frac{\delta}{\sigma}\right)^2 = \frac{16}{ES^2}$$

ある研究論文に掲載された結果(結果を鵜呑みにしないことの例)

グループ	人数	平均点	標準偏差	対応のない t 検定
処置群	11	78.91	8.42	P=0.000
対照群	11	76.82	7.41	

この場合共通の標準偏差を

$$\sigma = \sqrt{\frac{11 * 8.42^2 + 11 * 7.41^2}{22}} = 7.93$$

$\delta = 78.91 - 76.82 = 2.09$ とした時

$ES = \frac{2.09}{7.93} = 0.26$ であるから、有意差 0.05, 検定力 0.8 でのサンプル数は各群

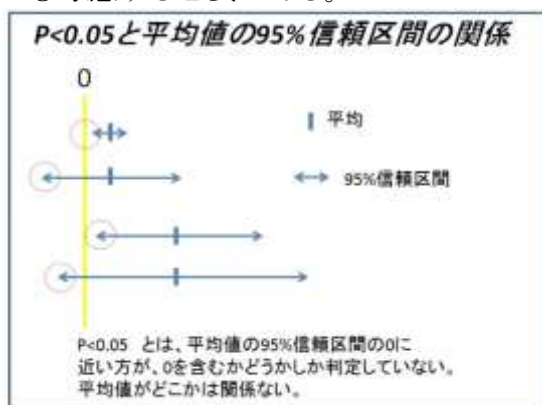
$n = \frac{16}{0.26^2} = 236$ 必要なので上記 p 値には疑問が残る。

実際 Welch's t を計算すると t 値 0.62、自由度 24、 $p = 0.32$

臨床的有意性 clinical significance.

RCT などの臨床研究が実施され、介入群と非介入群との間で、エンドポイントとして設定されたイベント発生率の違いが偶然のばらつき範囲内に収まっているという帰無仮説を検定するとき、帰無仮説とは、もともと“等しい”という仮定なので、平均値の 95%

信頼区間が差の場合は0を、比の場合は1を含むかどうかだけが問題となる。平均値がどこかは関係ない。 その結果は1) p 値が 0.05 より小さいからといって、その結果が臨床的に重要であるということにはならないし、逆に p 値が 0.05 より大きいからといって、その介入には効果がないということにもならない。2) サンプル数によって結果が変わり、標本数さえ大きければ、たとえ臨床的には役に立たないほど小さな違いであっても、統計学的には有意差があると判定され得る。3) 偶然有意差が出た可能性なども考慮する必要がある。



p 値の大きさではなく、むしろ介入の効果の大きさ (effect size) が臨床的に意味のあるものかどうかを検討せねばならない。また標本数が少ないため、本来は臨床的に意義があるのに統計学的に有意差が認められていないケースも多いため、臨床研究の計画段階で、必要な標本数を推定する必要がある。 α 、 β エラー、effectsize の推定値、の3つが重要である (上記: 2 標本の差の場合の必要サンプル数)。たとえば、対照群で 40% の患者にイベント発生が予測される疾患に対して、新しい治療法の効果を調べる際、その適用にてイベント発生率が 30% 以下になれば、臨床的に意義があると考える場合、 α エラー 5%、 β エラー 20% の設定では各群 380 人の被験者が必要と計算される。実際、サンプル数の計算は複雑で比の場合、相対リスクを推定する (Cox 比例ハザードモデルやロジスティック回帰分析等) の場合、解析ソフトの利用が必要 (BellCurve for Excel: ユーティリティ→サンプルサイズの計算)。

臨床的有意性の規準は事前に表明しておくべきであり、“臨床的有意性” という用語は、事前に定義される場合を除き使用することは推奨されない。

統計に関するいくつかの失敗例

1. バイアス(偏り):1936年のアメリカ大統領選挙(世界大恐慌の中)

民主党:フランクリン・ルーズベルト、共和党:アルフレッド・ランドン

リテラリー・ダイジェスト:(これに先立つ5回の大統領選挙において予測を外したことがない)200万人以上を対象から回収した調査結果を基に共和党のランドン候補が57%の得票を得て当選することを予想。

アメリカ世論研究所(ジョージ・ギャラップが率いる):3000という少ない対象者からの回答を基にルーズベルト候補が54%の得票を得て当選することを予想。

結果:ルーズベルト候補が60%の得票を得て勝利

理由:リテラリー・ダイジェストは、自誌の購読者(大恐慌の最中になお雑誌購読を続けられる裕福な人たち)を対象に、自動車保有者と電話利用者の名簿を使って1千万人もの対象者に郵便を送り、返送された2百万以上の回答を積み上げた。アメリカ世論研究所は、母集団全体(この場合は大統領選挙の投票権を持つ人の全体)を「収入中間層・都市居住者・女性」「収入下位層・農村部居住者・男性」のように互いに重ならないグループに分け、それぞれのグループに対して決まった割合で対象を抽出(割り当て法)。⇒ 標本のバイアス(偏り)は解析時には修正できないため正確な解析方法を用いても間違った結果になる。

2. 統計的有意と臨床的有意

各2000人で2種の薬の有効率を調べた。

基準薬剤Aの有効率は20% 400/2000

新薬Bの有効率は25% 500/2000であった。新薬は臨床的に有効か?

BellCurve for Excel: 2標本の比較 → 母比率の差の検定

母比率の差の検定				
基本統計量				
変数	n	比率		
薬剤A	2000	20.00%		
薬剤B	2000	25.00%		
母比率の差の95%信頼区間				
比率の差	下限値	上限値		
-5.00%	-7.58%	-2.42%		
母比率の差の検定				
統計量:z	両側P値	* : P<0.05 **	片側P値	* : P<0.05 ** : P<0.01
-3.7864	p < 0.001	**	p < 0.001	**

P<0.001、統計的に有意差あり

しかし、新薬は臨床的にも有効といえるか。

統計学的検定では、同じ差でもサンプル数を増やせば有意差がでる。統計的解釈を行う上では、サンプル数、検出力、臨床的有意性などを考えてから結論を出す

ことが重要である。

もし求める臨床的な差が 5% で満足であればサンプル数は下図に計算される。

→ 参照 PDF2 p5

	B の有効率	P1		0.25
	A の有効率	P2		0.2
		P	$p = \frac{p_1 + p_2}{2}$	0.225
		R	$R = \sqrt{2p(1-p)}$	0.59055059
		S	$s = \sqrt{p_1(1-p_1) + p_2(1-p_2)}$	0.58949131
	1群のサンプルサイズ	n	$n = \left(\frac{Z_{\alpha/2}R + Z_{\beta}S}{\delta} \right)^2$	1092.50326
	比率の差	δ		0.05
有意水準 α (両側) 5%、検出力 β 80% では、 $Z_{(\alpha/2)}=1.96, Z_{\beta}=0.84$				

有効率の差が 5% の時、臨床試験に必要なサンプル数は一群に約 1100 人であり、2000 人は必要なかったことになる (費用、労力の無駄)。

もしその差を 10% 欲しいとして $\alpha = 0.05$ 、 $1 - \beta = 0.8$ であればサンプル数は

	B の有効率	P1		0.3
	A の有効率	P2		0.2
		P	$p = \frac{p_1 + p_2}{2}$	0.25
		R	$R = \sqrt{2p(1-p)}$	0.61237244
		S	$s = \sqrt{p_1(1-p_1) + p_2(1-p_2)}$	0.60827625
	1群のサンプルサイズ	n	$n = \left(\frac{Z_{\alpha/2}R + Z_{\beta}S}{\delta} \right)^2$	292.821238
	比率の差	δ		0.1
有意水準 α (両側) 5%、検出力 β 80% では、 $Z_{(\alpha/2)}=1.96, Z_{\beta}=0.84$				

つまり、臨床的に 10% の差があれば新薬は有効と考えた場合、サンプル数を両側検定で 1 標本あたり 300 例前後 (計 600 例程度) のデータを集め検定すべきである。その結果、差が 5% であった場合

母比率の差の検定					
基本統計量					
変数	n	比率			
薬剤 A	300	20.00%			
薬剤 B	300	25.00%			
母比率の差の95%信頼区間					
比率の差	下限値	上限値			
-5.00%	-11.67%	1.67%			
母比率の差の検定					
統計量:z	両側P値	* : P<0.05	片側P値	* : P<0.05	** : P<0.01
-1.4665	0.1425		0.0713		

で有意差なしと判定される。

☆ 臨床的有意とは、臨床試験施行者が有意と判断した有効率の差を適切なサンプル数によって証明した場合にいうことができる。

*100 種類の薬があり 10 種類が本当に効き、90 種類は効かないとする(真に有効であるものの割合を基準率(ここでは 0.1)という)。そしてそれぞれの薬に対して個別に、 $\alpha = 0.05$, $1 - \beta = 0.8$ で基準薬との臨床試験が行われ検定したとする。10 種類の有効な薬のうち 8 種類は検出される($1 - \beta = 0.8$)。残りの 90 種類のうち 5 種類は効果のないに効果ありの判定される($\alpha = 0.05$)。つまり、13 種類が効果ありと判定される。効果あると判定された薬のうち本当に効くのは 8/13 (62%)、逆に効くと判定された薬の中では 38%が偽陽性ということになる。このような検定で P=0.05 で検定したので、13 種類が基準薬と同等または有効でないとの判断が間違えるのは 100 回に 5 回しかない、と表現することは誤り。特に基準率が低い実験(例えば 0.001)では初めから有効な薬は 1000 に一つなので、100 の実験に含まれていない可能性は高く、全く含まれていない実験を行っても有効薬は 5 種類選ばれてくる。逆に 10000 の実験が行われればそのうちの 10 種類の有効な薬から 8 種を選び出すとき、9990 の中から 500 の薬が有効と判定され、有効と判定された 500/508 (98.4%)は偽陽性ということになる。

3. 分割表の解析は間違えやすい

下図のようなデータを分析した際どのように考えるか

	poor	fair	average	good	excellent
A	1	2	3	4	5
B	5	4	3	2	1

① χ^2 検定を行えば [BellCurve for Excel: 集計表の作成と検定](#) →独立性の検定

独立性の検定							
観測度数							
	列1	列2	列3	列4	列5	合計	
行1	1	2	3	4	5	15	
行2	5	4	3	2	1	15	
合計	6	6	6	6	6	30	
期待度数							
	列1	列2	列3	列4	列5		
行1	3.000	3.000	3.000	3.000	3.000		
行2	3.000	3.000	3.000	3.000	3.000		
独立性の検定							
カイ二乗値	自由度	P 値	* : P<0.05 ** : P<0.01				
6.6667	4	0.1546					
Cramer's V	0.4714						

② 単純なマンホイットニー検定は集計ではなく、一つ一つがデータの判断となるから、A,Bは並び替えただけの全く同じデータ

マン=ホイットニーのU検定							
基本統計量							
変数	n	平均順位	統計量:U				
A	5	5.500	12.5				
B	5	5.500	12.5				
マン=ホイットニーのU検定							
統計数値表による検定			正規化検定		* : P<0.05 ** : P<0.01		
統計量:U	両側検定		E(U)	V(U)	連続性補正統計量:z	両側P値	
12.5	※ 1		12.5	22.222	なし	0.0000	1.0000
正確検定							
両側P値	* : P<0.05 ** : P<0.01						
1.0000							

③ Poor、fair、average、good、excellent に順序を付け順に 1, 2, 3, 4, 5 とすると表は例えば poor では A は 1 という量が 1 つ、B では 5 つとなり次の表をマンホイットニー検定することになる。

A	B
1	1
2	1
2	1
3	1
3	1
3	2
4	2
4	2
4	2
4	3
5	3
5	3
5	4
5	4
5	5

マン=ホイットニーのU検定						
基本統計量						
変数	n	平均順位	統計量:U			
A	15	19.500	52.5			
B	15	11.500	172.5			
マン=ホイットニーのU検定						
統計数値表による検定			正規化検定		* : P<0.05 ** : P<0.01	
統計量:U	両側検定	E(U)	V(U)	連続性補正	統計量:z	両側P値
52.5	※ 1	112.5	558.621	なし	2.5386	0.0111 *
				あり	2.5174	0.0118 *

この場合は有意差あり。

データ収集の状況、データの意味を考えて判断する必要があるが、ほとんどの場合③が正解となる。BellCurve for Excel: 2標本の比較 →マンホイットニーのU検定

4.交絡因子(共変量)に注意

A、Bの2地区で血圧に差があるか？

地区	年齢	血圧	地区	年齢	血圧
A	44	190	B	45	221
A	36	140	B	58	197
A	57	270	B	66	125
A	39	126	B	62	235
A	32	98	B	64	253
A	41	212	B	50	167
A	46	161	B	54	260
A	41	181	B	66	234
A	40	115	B	50	172
A	46	135	B	35	168
A	41	198	B	61	220
A	62	217	B	57	191
A	44	150	B	43	127
A	30	64	B	68	329
A	44	151	B	48	141
A	52	189	B	66	284
A	49	219	B	44	225
A	48	181	B	59	215
A	40	203	B	59	197
A	48	229	B	51	127
A	33	139	B	52	207

両地区の血圧のt検定、マンホイットニー検定を行うと有意差あり

BellCurve for Excel: 2標本の比較 →母平均の差の検定、マンホイットニーのU検定

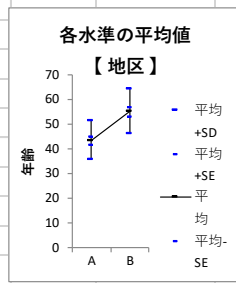
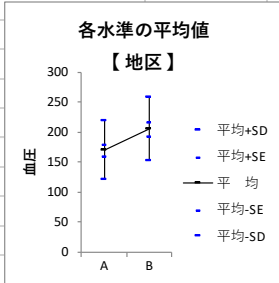
母平均の差の検定					
基本統計量					
変数	n	平均	不偏分散	標準偏差	標準誤差
A地区	21	169.905	2401.990	49.010	10.695
B地区	21	204.524	2827.862	53.178	11.604
等分散性の検定 (両側検定)					
統計量:F	自由度1	自由度2	P 値		
1.1773	20	20	0.7187		
平均の差と効果量 (平均の差 = 「A地区」 - 「B地区」)					
効果量 (観測値)			効果量 (ランク化)		
平均の差	Cohen's d	Hedges' g	Cohen's d	Hedges' g	
34.6190	0.6937	0.6770	0.6761	0.6598	
母平均の差の95%信頼区間 (両側検定/対立仮説: 「A地区」 ≠ 「B地区」)					
下限値		上限値			
t推定	2.7244	66.5137			
Welchの方法	2.7178	66.5203			
母平均の差の検定 (両側検定/対立仮説: 「A地区」 ≠ 「B地区」)					
		統計量:t	自由度	P 値	
観測値	t検定	2.1937	40	0.0341 *	
	Welchの方法	2.1937	39.7365	0.0342 *	
ランク化	t検定	2.1379	40	0.0387 *	
	Welchの方法	2.1379	39.7969	0.0387 *	
検出力 (両側検定/対立仮説: 「A地区」 ≠ 「B地区」)					
		α=0.05			
観測値	t検定	0.5718			
	Welchの方法	0.5717			
ランク化	t検定	0.5503			
	Welchの方法	0.5502			

マン=ホイットニーのU検定							
基本統計量							
変数	n	平均順位	統計量:U				
A地区	21	17.619	302				
B地区	21	25.381	139				
マン=ホイットニーのU検定							
統計数値表による検定			正規化検定				
統計量:U	両側検定	E(U)	V(U)	連続性補正	統計量:z	両側P値	* : P<0.05 ** : P<0.01
139 ※ 1		220.5	1579.866	なし	2.0504	0.0403 *	
				あり	2.0379	0.0416 *	

この場合、年齢と血圧に相関関係があり、B地区に高齢者が多い(解析略)
そこで、共分散分析を行うと血圧に有意差を認めない(年齢に有意差)。

[BellCurve for Excel: 分散分析・多重比較 → 共分散分析](#)

共分散分析										
基本統計量										
目的変数	モデル	地区	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
血圧	地区	A	21	169.905	49.010	120.895	218.915	10.695	159.210	180.600
		B	21	204.524	53.178	151.346	257.701	11.604	192.919	216.128



説明変数(血圧)、交絡因子(年齢)ともに計量データでは共分散分析をおこなう。

共変量										
共変量	モデル	地区	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
年齢	地区	A	21	43.476	7.827	35.649	51.303	1.708	41.768	45.184
		B	21	55.143	9.035	46.108	64.178	1.972	53.171	57.114
等分散性の検定										
ルビーン検定										
目的変数	F 値	自由度1	自由度2	P 値						
血圧	0.6909	1	40	0.4108						
回帰の有意性の検定										
	平方和	自由度	平均平方	F 値	P 値	*: P<0.05 **: P<0.01				
回帰による	39967.7133	1	39967.7133	24.1182	p < 0.001	**				
回帰からの分散分析表	64629.3344	39	1657.1624							
因子	TypeIII 平方和	自由度	平均平方	F 値	P 値	*: P<0.05 **: P<0.01				
地区	568.3409	1	568.3409	0.3430	0.5615					
年齢	39967.7133	1	39967.7133	24.1182	p < 0.001	**				
誤差	64629.3344	39	1657.1624							
全体	117181.0714	41								

交絡因子の調整

事前に割り付けを無作為化(RCT など)

無作為化が困難な場合解析時点で調整(疫学調査など)

説明変数、交絡因子ともに計量データ:

共分散分析(analysis of covariance ANCOVA)

説明変数、交絡因子ともに計数データ:

Mantel-Haenszel 法(2×2分割表に対し)

Mantel-extension 法(2×k分割表に対し)

目的変数が計数データ:

ロジスティック回帰など

交絡を調整する上での注意点

週刊日本医事新法 J-CLEAR 通信(54): 交絡因子の調整という魔物より

「原因と結果との間にある中間変数は交絡因子ではない」

高血圧患者を生活習慣改善群と降圧薬投与群にわけ、どちらの予後が良好かをみる観察試験を行ったとしよう。生データでは降圧薬投与群にイベントが多くなる。それは治療前血圧が著明に高い患者に降圧薬が処方され、また治療前血圧が著明に高い患者にイベントが多い、という関連があるからである。そのため、この観察試験から正しい結果を得るには、血圧という交絡因子の調整を行わねばならない。多くの場合、この

変数を選ぶときに誤りが生じる。それは、治療後血圧を交絡因子としてしまう誤りである。治療後血圧は治療の結果でもあり、かつイベントの原因でもある。したがって、中間変数であるから交絡因子ではない。これを交絡因子として調整することは、治療後血圧を等しくして 2 群間のイベントの差をみることになる。この場合、治療前血圧は降圧薬投与群のほうが高くなる。その結果、「降圧薬投与群にイベントが多い」という奇妙な結果となる。正しい交絡因子は、治療前血圧である。そこで治療前の血圧で調整、すなわち治療前血圧が等しいという条件にすると、「降圧薬投与群のほうが血圧はより低下し、イベントが抑制される」という正しい結果が得られる。しかし、観察研究では既に治療されている症例を対象にするため、治療前血圧が不明なことが多い。したがって、治療後血圧を採用してしまうというエラーが生じやすい。横断研究では特にそうである。ロジスティック回帰では、アウトカムの死亡者、生存者のいずれか少ないほうの数を 10 で割った数までを説明変数として加えることができる。

重要な交絡因子は必ず研究前に調査し、モデルに加えて補正しなければならない交絡因子の数を見積もり、それに十分対応できるサンプル数を集めることが重要です。10 個の交絡をどうしても補正したいのであれば、少なくとも「 $10 \times 10 = 100$ 」のイベントが集まるようにサンプルを設定する必要がある。

5. 交互作用(相乗効果または相殺効果)

A 薬、B 薬の効果、A 薬、B 薬併用 (A+B) による交互作用を調べたい。

浜田知久馬著、学会・論文発表のための統計学 p30 より(一部改変)

対照群	A	B	A+B
187	168	195	129
160	170	207	138
197	181	178	159
232	176	137	149
218	182	176	155
186	154	144	130
168	191	123	113
156	155	138	126
188	159	156	94
187	150	153	120

BellCurve for Excel: 分散分析・多重比較 → 一元配置分散分析

一元配置分散分析										
基本統計量										
目的変数	モデル	因子A	n	平均	標準偏差(SD)	平均-SD	平均+SD	標準誤差(SE)	平均-SE	平均+SE
変数Y	因子A	対照群	10	187.900	23.877	164.023	211.777	7.550	180.350	195.450
		A	10	168.600	13.858	154.742	182.458	4.382	164.218	172.982
		B	10	160.700	27.309	133.391	188.009	8.636	152.064	169.336
		A+B	10	131.300	19.934	111.366	151.234	6.304	124.996	137.604

各水準の平均値
【因子A】

等分散性の検定										
バートレット検定					ルビーン検定					
目的変数	カイ二乗値	自由度	P 値	F 値	自由度1	自由度2	P 値			
変数Y	3.9703	3	0.2647	1.3939	3	36	0.2604			

分散分析表										
因子	Type III 平方和	自由度	平均平方	F 値	P 値	*: P<0.05 **: P<0.01				
因子A	16584.8750	3	5528.2917	11.6063	p < 0.001 **					
誤差	17147.5000	36	476.3194							
全体	33732.3750	39								

等分散性を仮定しない検定										
手法	F 値	自由度1	自由度2	P 値	*: P<0.05 **: P<0.01					
Welch	11.7798	3	19.3347	p < 0.001 **						
Brown-Forsythe	11.6063	3	30.3638	p < 0.001 **						

多重比較検定												
因子	目的変数	手法	対立仮説	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値	
因子A	変数Y	Dunnett	対照群 ≠	対照群	A	187.9000	168.6000	19.3000	9.7603	1.9774	0.1366	
				対照群	B	187.9000	160.7000	27.2000	9.7603	2.7868	0.0227 *	
				対照群	A+B	187.9000	131.3000	56.6000	9.7603	5.7990	p < 0.001 **	
				対照群 <	対照群	A	187.9000	168.6000	19.3000	9.7603	1.9774	0.9978
				対照群	B	187.9000	160.7000	27.2000	9.7603	2.7868	0.9999	
				対照群	A+B	187.9000	131.3000	56.6000	9.7603	5.7990	1.0000	
				対照群 >	対照群	A	187.9000	168.6000	19.3000	9.7603	1.9774	0.0684
				対照群	B	187.9000	160.7000	27.2000	9.7603	2.7868	0.0114 *	
				対照群	A+B	187.9000	131.3000	56.6000	9.7603	5.7990	p < 0.001 **	

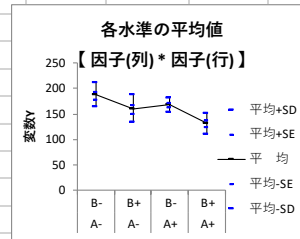
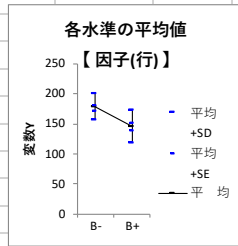
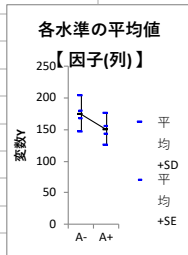
一元配置分散分析と多重比較で有意差があったとき、相乗作用という言葉は使用できない。

データを二元配置分散分析の形式に整える。

	A-	A+
B-	187	168
	160	170
	197	181
	232	176
	218	182
	186	154
	168	191
	156	155
	188	159
	187	150
B+	195	129
	207	138
	178	159
	137	149
	176	155
	144	130
	123	113
	138	126
	156	94
	153	120

BellCurve for Excel: 分散分析・多重比較 → 二元配置分散分析

二元配置分散分析											
基本統計量											
目的変数	モデル	因子(列)	因子(行)	n	平均	標準偏差(平均-SD)	平均+SD	標準誤差(SE)	平均-SE	平均+SE	
変数Y	因子(列)	A-		20	174.300	28.601	145.699	202.901	6.395	167.905	180.695
		A+		20	149.950	25.403	124.547	175.353	5.680	144.270	155.630
	因子(行)		B-	20	178.250	21.425	156.825	199.675	4.791	173.459	183.041
			B+	20	146.000	27.730	118.270	173.730	6.201	139.799	152.201
	因子(列) * 因子(行)	A-	B-	10	187.900	23.877	164.023	211.777	7.550	180.350	195.450
		A-	B+	10	160.700	27.309	133.391	188.009	8.636	152.064	169.336
		A+	B-	10	168.600	13.858	154.742	182.458	4.382	164.218	172.982
		A+	B+	10	131.300	19.934	111.366	151.234	6.304	124.996	137.604



等分散性の検定						
ルビーン検定						
目的変数	F 値	自由度1	自由度2	P 値		
変数Y	1.3939	3	36	0.2604		
分散分析表						
因子	TypeIII平方和	自由度	平均平方	F 値	P 値	* : P<0.05 ** : P<0.01
因子(列)	5929.2250	1	5929.2250	12.4480	0.0012	**
因子(行)	10400.6250	1	10400.6250	21.8354	p < 0.001	**
因子(列) * 因子(行)	255.0250	1	255.0250	0.5354	0.4691	
誤差	17147.5000	36	476.3194			
全体	33732.3750	39				

相乗効果(相殺効果)あり
 というためには二元配置
 分散分析で交互作用が
 有意であることが必要。

「因子(行)」の各水準における「因子(列)」の単純主効果の検定								
目的変数	因子(行)	因子	平方和	自由度	平均平方和	F 値	P 値	* : P<0.05 ** : P<0.01
変数Y	B-	因子(列)	1862.4500	1	1862.4500	3.9101	0.0557	
		誤差	17147.5000	36	476.3194			
	B+	因子(列)	4321.8000	1	4321.8000	9.0733	0.0047	**
		誤差	17147.5000	36	476.3194			

「因子(行)」の各水準における「因子(列)」の多重比較検定										
目的変数	手法	因子(行)	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値
変数Y	Tukey	B-	A-	A+	187.9000	168.6000	19.3000	9.7603	1.9774	0.0557
			B+	A-	A+	160.7000	131.3000	29.4000	9.7603	3.0122

「因子(列)」の各水準における「因子(行)」の単純主効果の検定								
目的変数	因子(列)	因子	平方和	自由度	平均平方和	F 値	P 値	* : P<0.05 ** : P<0.01
変数Y	A-	因子(行)	3699.2000	1	3699.2000	7.7662	0.0084	**
		誤差	17147.5000	36	476.3194			
	A+	因子(行)	6956.4500	1	6956.4500	14.6046	p < 0.001	**
		誤差	17147.5000	36	476.3194			

「因子(列)」の各水準における「因子(行)」の多重比較検定										
目的変数	手法	因子(列)	水準1	水準2	平均1	平均2	差	標準誤差	統計量	P 値
変数Y	Tukey	A-	B-	B+	187.9000	160.7000	27.2000	9.7603	2.7868	0.0084 **
			A+	B-	B+	168.6000	131.3000	37.3000	9.7603	3.8216

因子(列)*因子(行)が有意ではない=交互作用(相乗効果)なし⇒ 相加効果と判断

主な参考文献

- 石村貞夫(1993):分散分析のはなし. 東京図書
丹後俊郎(1993):新版医学への統計学. 朝倉書店
浜島信之(1993):多変量解析による臨床研究 第二版. 名古屋大学出版会
高橋善弥太(1995):一医者のための— ロジスティック・Cox 回帰入門
宮原英夫・丹波俊郎(1995):医学統計学ハンドブック. 朝倉書店
浜田知久馬(2000):学会・論文発表のための統計学. 真興交易医書出版部
柳川堯(2012):離散多変量データの解析. 共立出版株式会社
中村剛(2013):医学統計学シリーズ、Cox 比例ハザードモデル. 朝倉書店
柳井久江(2015):4steps エクセル統計、オーエムエス出版
宮田敏(2015):医学統計勉強会、東北大学大学院医学系研究科
EBM 開発学寄附講座
BellCurve エクセル統計 統計 WEB 解析事例、統計用語集より
インターネットから例題、解説を利用(ウェブサイト表示)