

医学統計勉強会

第3回 連続変数の比較

— 二標本問題・分散分析 —

帝京大学臨床研究センター（TARC）・帝京大学大学院公衆衛生学研究科 共催

帝京大学大学院公衆衛生学研究科

宮田 敏

平均・中央値の差の検定

二標本問題：二つのグループの平均値の差の検定

帰無仮説 $H_0: \mu_1 = \mu_2$

対立仮説 $H_1: \mu_1 \neq \mu_2$

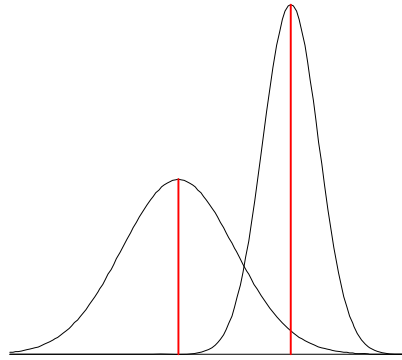
Welch's t-test（ウェルチのt検定）：二群のデータがそれぞれ**正規分布**に従う。**不等分散**を仮定する。

⇔ Mean ± SE, Mean ± SD に対応。

Mann-Whitney test, Wilcoxon's rank sum test：二群のデータは任意の**同じ形の分布**に従う。二群の分散は**等分散**になる。⇔ Median (IQR) に対応。分布の形が同じで中央値が等しければ、平均も等しくなる。

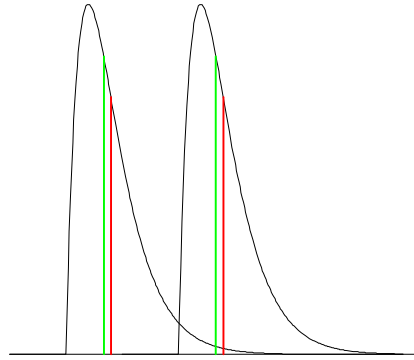
検定をしたら、必ず **p 値を明記**する。

T検定に適した比較



- 二群とも正規分布に従う
- 分散は異なってもよい

MW検定に適した比較



- 正規分布に従わなくてもよい
- 分布の形状は同じ。分散も同じ。

データの分布が正規分布に従わず、分散も等しくない場合、取りあえず元データを**対数変換**するなどして、**等分散**に近づける。それでもだめなら、専門家にご相談ください。

2023/9/27

医学統計勉強会 第3回 連続変数の比較

3

```
> # 二標本問題 ##
> x1 <- c(8, 17, 9, 5, 14, 13, 13, 3, 9, 0)
> x2 <- c(19, 13, 13, 18, 21, 18, 11, 21, 20, 11)
>
> boxplot(x1, x2)
>
> t.test(x1, x2)# Welch's t test

Welch Two Sample t-test

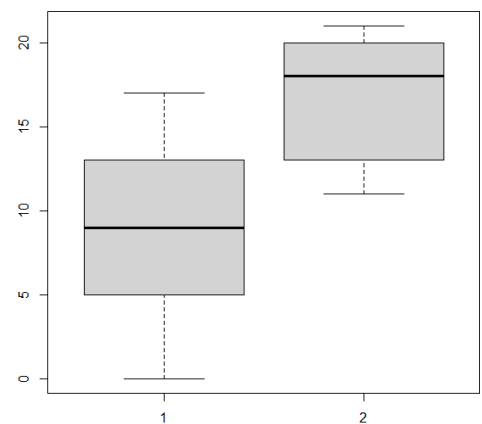
data: x1 and x2
t = -3.4953, df = 16.829, p-value = 0.002807
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.870201 -2.929799
sample estimates:
mean of x mean of y
 9.1      16.5

> wilcox.test(x1, x2)# Mann-Whitney test, Wilcoxon's rank sum test

Wilcoxon rank sum test with continuity correction

data: x1 and x2
W = 14, p-value = 0.00698
alternative hypothesis: true location shift is not equal to 0

警告メッセージ:
wilcox.test.default(x1, x2) で:
タイがあるため、正確な p 値を計算することができません
```



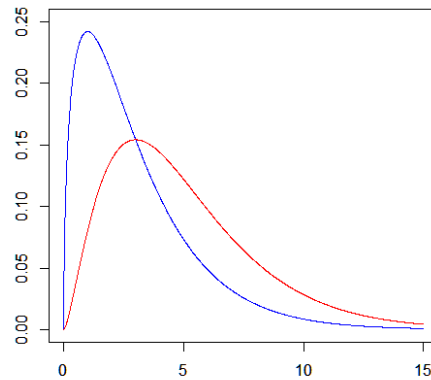
2023/9/27

医学統計勉強会 第3回 連続変数の比較

4

比較に困る二標本問題

医学・生物学データには、値の範囲が正の値に限られるものがある。そのようなデータの場合、分布が左右非対称（≠正規分布）である一方で、分散も異なる場合がある。



問題点は二つ。

- 非正規性 ⇒ サンプル数を増やすことで、**中心極限定理による正規近似**を用いて t 検定に持ち込む。
- 不等分散性 ⇒ ノンパラメトリックな枠組みでは、対処できない。やはりサンプル数を増やして、不等分散を許容する **Welch' s t-test** に持ち込む。

一般に20程度以上のサンプルが必要であるとされている。サンプルがそれ以下になるときは、統計家にご相談下さい。

分布の正規性の確認

標本分布の正規性の確認は、適切なモデルを選択する上で重要。

Definition: n 個の標本を大きさ順に並べたとき、 i 番目に小さな標本は $[100(i - 0.5)/n]$ **標本パーセント点 (sample percentile)** であるという。

例えば標本が、正規分布など特定の確率分布から抽出されたとする。このとき、その特定の分布の理論上の $[100(i - 0.5)/n]$ パーセント点は標本パーセント点の近くにあるはずである。

そうであるならば、理論上のパーセント点と標本パーセント点をプロットすれば、直線状に並ぶはず。

QQ-norm plot (Normal probability plot)

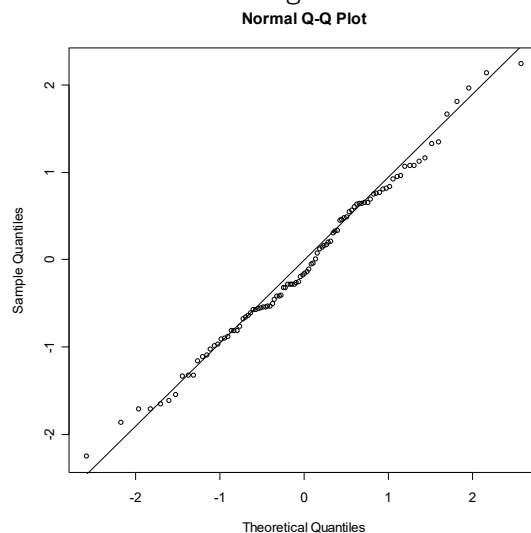
Definition: n 個の標本が得られたとき、標準正規分布 $N(0, 1)$ の $[100(i - 0.5)/n]$ パーセント点と i 番目に小さな観測値 = $[100(i - 0.5)/n]$ 標本パーセント点のプロットを **QQ-norm plot** という。

もし標本が正規分布から得られたならば、標本の観測値(を標準化したものは、理論的な $N(0, 1)$ のパーセント点の近くにあるはずだから、QQ-norm plotは**直線**の近くにプロットされるはずである。

データのgraphical summaryでは、**QQ-norm plotによる正規性の確認**も行う。

R-code of QQ-norm plot:

```
> x <- rnorm(100)      ## sample data ##  
  
> qqnorm(x)  
> qqline(x)           ## add a straight line ##
```



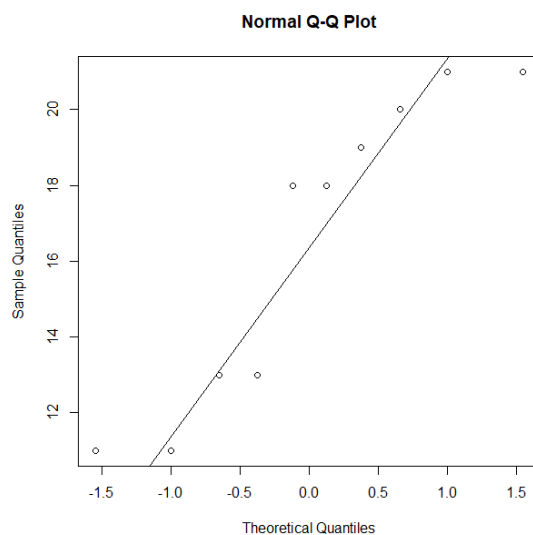
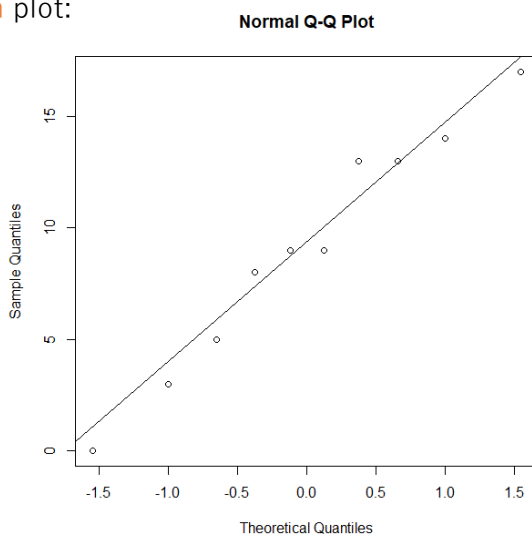
2023/9/27

医学統計勉強会 第3回 連続変数の比較

9

R-code of QQ-norm plot:

```
> # QQ-norm plot #  
>  
> qqnorm(x1)  
> qqline(x1)  
>  
> qqnorm(x2)  
> qqline(x2)
```



QQ-norm plot はどちらも直線状にプロットされており, x_1, x_2 ともに正規分布に従うと考えて矛盾はない.

2023/9/27

医学統計勉強会 第3回 連続変数の比較

10

Shapiro-Wilk normality test (シャピロ=ウィルクの正規性検定)

「データは正規分布に従う」という帰無仮説に対して、「正規分布には従わない」という対立仮説を検定するものである。正規性を帰無仮説とするので、帰無仮説が棄却されなければ正規性が示されたとする少し変わった検定である。

```
> shapiro.test(x1)

Shapiro-Wilk normality test

data:  x1
W = 0.96708, p-value = 0.8625
```

$P > 0.05$ であるから、帰無仮説は棄却されない
⇔ 正規性の仮定は、否定されない。

三群以上の比較：

帰無仮説 $H_0: \mu_1 = \dots = \mu_k$, k : グループの数

対立仮説 H_1 : 少なくとも一つの母平均が他から異なる

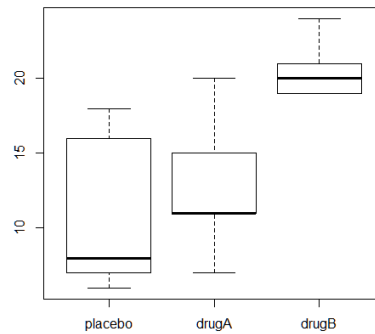
分散分析 (Analysis of Variance, ANOVA)：各群のデータがそれぞれ**正規分布**に従う。**等分散**を仮定する。

Kruskal-Wallis rank sum test：各群のデータは任意の**同じ形の分布**に従う。当然各群の分散は**等分散**になる。No-parametric test

データの分布が正規分布に従わず、分散も等しくない場合。やはり**対数変換**などで、**等分散**に近づける。

Box-Cox変換：分散の安定化と正規性の向上を同時に達成する変換。詳細は、ご相談ください。

サンプルデータ



このサンプルデータの場合、各群の分散は明らかに異なっている。また、placebo群やdrugA群は分布が左右非対称であり、正規性の仮定も疑わしい。その意味で、分散分析の仮定が満たされているか疑問があるが、ともあれ分散分析を行ってみる。

- One-way ANOVA : $p = 0.0119$
- One-way ANOVA (not assuming equal variances) : $p = 0.01127$
- Kruskal-Wallis rank sum test : $p = 0.0182$

一元配置分散分析 (constant variance)

```
> # One-way ANOVA #
> fit.aov <- aov(z ~ group, data=Data)
> summary(fit.aov)
          Df Sum Sq Mean Sq F value Pr(>F)
group      2  260.4   130.20   6.565 0.0119 *
Residuals 12  238.0    19.83
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

一元配置分散分析 (unequal variance)

```
> # One-way ANOVA (not assuming equal variances) #
> oneway.test(z ~ group, data=Data)

One-way analysis of means (not assuming equal variances)

data:  z and group
F = 9.5017, num df = 2.0000, denom df = 6.6401, p-value = 0.01127
```

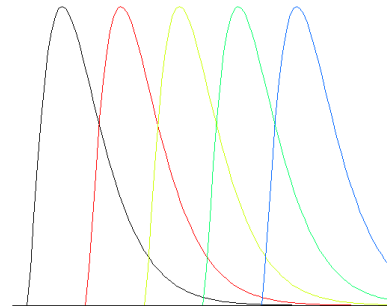
Kruskal-Wallis rank sum test (non-parametric analog of one-way analysis of variance)

```
> # Kruskal-Wallis rank sum test #
> kruskal.test(z ~ group, data=Data)

Kruskal-Wallis rank sum test

data: z by group
Kruskal-Wallis chi-squared = 8.0122, df = 2, p-value = 0.0182
```

クラスカル-ウォリス検定では、各群の分布は location が異なったとしても同じ形の分布に従う必要がある。



モデル診断 (model diagnostics)

分散分析にはいくつかのモデルの仮定が必要。

- **独立性** サンプルは、互いに独立である。
- **正規性** サンプルは、正規分布に従う。
- **等分散性** サンプルの分散は、一定である。

$$X_{ij} \sim N(\mu_i, \sigma^2), \quad i = 1, \dots, I, j = 1, \dots, n_i$$

$\mu_i = E(X_{ij})$: expected value of i -th group,
 $\sigma^2 > 0$: common variance of error terms

$$X_{ij} = \mu_i + \epsilon_{ij}, \quad \epsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$

分散分析のモデルの仮定は、誤差項 ϵ_{ij} の仮定に書き換えられる。

$$\text{error terms} \quad \epsilon_{ij} = X_{ij} - \mu_i$$

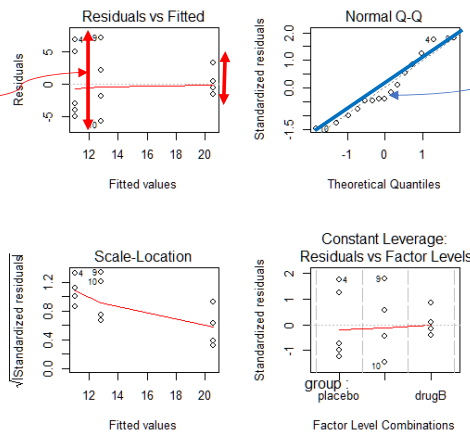
- 独立性 誤差項 ϵ_{ij} は、互いに独立である。
- 正規性 誤差項 ϵ_{ij} は、正規分布に従う。
- 等分散性 誤差項 ϵ_{ij} の分散は、 σ^2 で一定である。

誤差項は観察できないので、以下の**残差 (residuals)** で置き換える。

$$\text{residuals} \quad e_{ij} = X_{ij} - \bar{X}_i : \bar{X}_i : \text{sample mean of } i\text{-th group}$$

- 残差の**正規性**を確認するには、残差の **QQ-norm plot** を描いて**プロットが直線上にあるかどうか**を検討する。
- **等分散性**を確認するには、**残差と各群の標本平均のプロット (residuals vs. fitted value plot)** を描く。もし各群の分散が一定であれば、それぞれの群の残差はほぼ同じ範囲に分布するはずである。各群の残差を確認し、**残差の範囲に大きな違いが無ければ**、**等分散性の仮定は満たされている**と考える。

残差の範囲が異なる



ほぼ直線

サンプルデータの場合、QQ-norm plotはほぼ直線上にあるので、正規性の仮定には問題がなさそうである。

しかし、residuals vs. fitted value plotを診ると、残差の範囲には大きな違いがある。等分散性の仮定は疑わしい。

モデルの修正:分散の安定化と正規性の改善

モデル診断の後、model assumptionのいくつかを満たされない場合がある。

- Independence
- Constant variance
- Normality

独立性の仮定が満たされない場合、モデルの修正は通常困難。Responseを分類する、新たな要因を見つけてくることを考える。

分散安定化のための変換:

•対数変換 $y' = \log(y)$

•指数変換 $y' = y^\lambda$

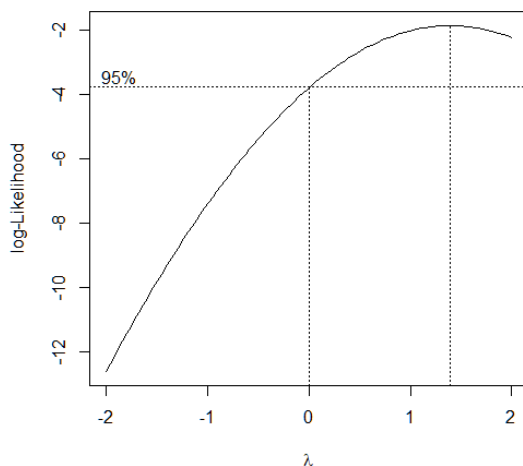
•分散の安定化と正規性の改善を同時に達成する変換

Box-Cox transformation

$$y^{(\lambda)} = \begin{cases} (y^\lambda - 1)/\lambda & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases}$$

where $y = \text{Days} + \alpha, (\alpha = 1)$

Box-Cox変換はパラメター λ で特徴付けられ、その適合度(fitness)は対数尤度(log-likelihood)と呼ばれる尺度で測られる。



```
> bxcx <- boxcox(fit.aov)
> lambda <- bxcx$lambda[which.max(bxcx$y)]
>
> fit.aov2 <- aov((z^lambda - 1)/lambda ~ group, data=Data)
> summary(fit.aov2)
          Df Sum Sq Mean Sq F value Pr(>F)
group      2   2154  1076.9    7.139 0.00907 **
Residuals 12   1810   150.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

多重比較 (Multiple Comparison)

分散分析の帰無仮説 $H_0: \mu_1 = \dots = \mu_k$ が棄却されたとき、どの μ_i が他から有意に異なるかが知りたい。

Tukey's HSD (Honestly Significant Difference) : すべての対比 ($\mu_i - \mu_j$) についての検定を同時に行う。可能な対比の組み合わせは、 $k(k-1)/2$ 通り。

Dunnettの方法 : グループの一つがコントロール群である時、コントロール群と他の($k-1$)の対照群との比較を同時に行う。

Williamsの方法 : 対立仮説 $H_1: \mu_1 \leq \dots \leq \mu_k$ (あるいはその逆) を検定する。

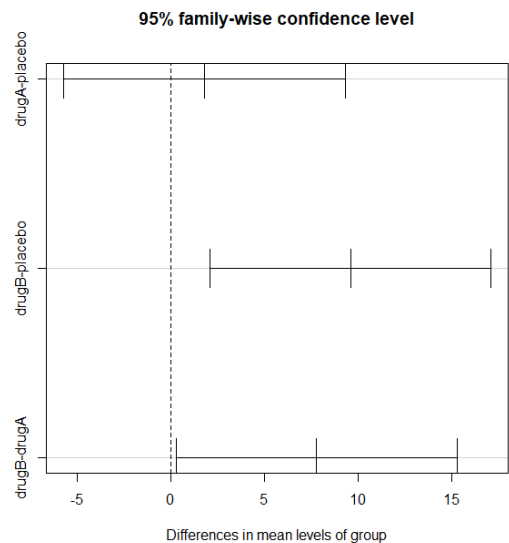
Tukey の方法 (Tukey HSD, honestly significant difference)

```
> # Tukey's method #
> library(MASS)
> fit.mc <- TukeyHSD(fit.aov)
> print(fit.mc)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = z ~ group, data = Data)

$group
      diff      lwr      upr    p adj
drugA-placebo  1.8 -5.714351  9.314351 0.8018504
drugB-placebo  9.6  2.0856489 17.114351 0.0133182
drugB-drugA    7.8  0.2856489 15.314351 0.0418080

> plot(fit.mc)
```



Dunnett の方法 (Dunnett post hoc test)

```
> # Dunnett's method (Dunnett post hoc test) #
> #install.packages("multcomp")# 最初の一回のみ実行
> library(multcomp)
> fit.dunnett <- glht(fit.aov, linfct = mcp(group = "Dunnett"))
> confint(fit.dunnett, level=0.95)

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Dunnett Contrasts

Fit: aov(formula = z ~ group, data = Data)
Quantile = 2.5026
95% family-wise confidence level

Linear Hypotheses:
      Estimate lwr      upr
drugA - placebo == 0  1.8000 -5.2488  8.8488
drugB - placebo == 0  9.6000  2.5512 16.6488

> summary(fit.dunnett)

Simultaneous Tests for General Linear Hypotheses

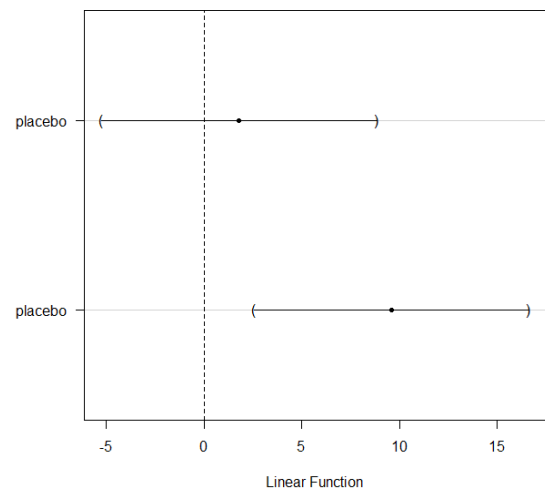
Multiple Comparisons of Means: Dunnett Contrasts

Fit: aov(formula = z ~ group, data = Data)

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
drugA - placebo == 0  1.800    2.817   0.639 0.75279
drugB - placebo == 0  9.600    2.817   3.408 0.00963 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

> plot(fit.dunnett)
.
```

95% family-wise confidence level



2023/9/27

医学統計勉強会 第3回 連続変数の比較

25

Williams の方法 (Williams method)

```
> # William's method #
> fit.williams <- glht(fit.aov, linfct = mcp(group = "Williams"))
> confint(fit.williams, level=0.95)

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Williams Contrasts

Fit: aov(formula = z ~ group, data = Data)
Quantile = 2.3856
95% family-wise confidence level

Linear Hypotheses:
      Estimate lwr      upr
C 1 == 0  9.6000  2.8807 16.3193
C 2 == 0  5.7000 -0.1191 11.5191

> summary(fit.williams)

Simultaneous Tests for General Linear Hypotheses

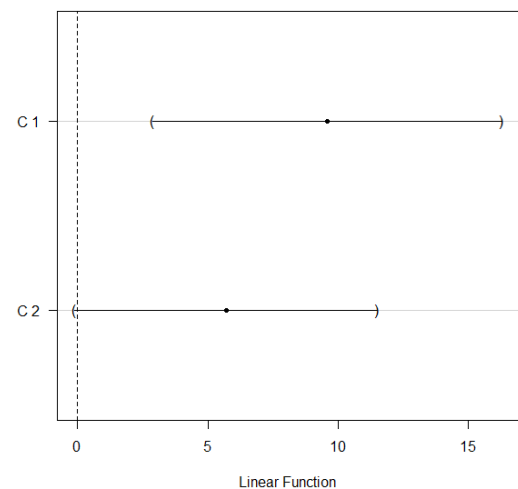
Multiple Comparisons of Means: Williams Contrasts

Fit: aov(formula = z ~ group, data = Data)

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
C 1 == 0  9.600    2.817   3.408 0.00787 **
C 2 == 0  5.700    2.439   2.337 0.05443 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

> plot(fit.williams)
.
```

95% family-wise confidence level



2023/9/27

医学統計勉強会 第3回 連続変数の比較

26

多重比較 (Multiple Comparison)

前項の方法は、すべて正規性の仮定を必要とするパラメトリックな方法。正規性を必要としない、ノンパラメトリックな方法も存在する。

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

	$H_1 : \mu_1 \neq \mu_2, \dots, \mu_i \neq \mu_j, \dots, \mu_1 \neq \mu_k$	$H_1 : \mu_1 \neq \mu_2, \mu_1 \neq \mu_3, \dots, \mu_1 \neq \mu_k$	$H_1 : \mu_1 \leq \dots \leq \mu_k$
パラメトリック検定	Tukeyの方法	Dunnettの方法	Williamsの方法
ノンパラメトリック検定	Steel-Dwassの方法	Steelの方法	Shirley-Williamsの方法

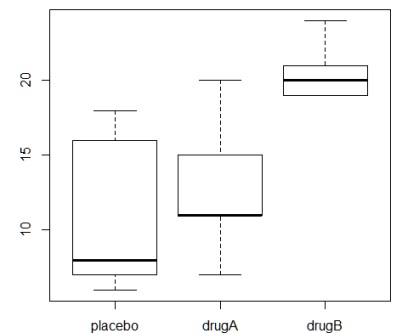
Steel-Dwass's method

```
> attach(Data)
> z
[1] 8 16 7 18 6 11 15 11 20 7 19 20 21 19 24
> group
[1] placebo placebo placebo placebo placebo drugA drugA drugA drugA
[10] drugA drugB drugB drugB drugB drugB
Levels: placebo drugA drugB
>
> # Steel-Dwass's method #
> pSDCFlig(z, group, method="Monte Carlo")
Ties are present, so p-values are based on conditional null distribution.
Group sizes: 5 5 5
Using the Monte Carlo (with 10000 Iterations) method:

For treatments placebo - drugA, the Dwass, Steel, Critchlow-Fligner W Statistic is 0.8917.
The smallest experimentwise error rate leading to rejection is 0.8325 .

For treatments placebo - drugB, the Dwass, Steel, Critchlow-Fligner W Statistic is 3.704.
The smallest experimentwise error rate leading to rejection is 0.0236 .

For treatments drugA - drugB, the Dwass, Steel, Critchlow-Fligner W Statistic is 2.9814.
The smallest experimentwise error rate leading to rejection is 0.0747 .
```



参考文献：

- 石村 貞夫, 石村 光資郎 (著) 「入門はじめての分散分析と多重比較」東京図書 (2008/01) ISBN-13: 978-4489020292

統計解析ソフトRについて

- 舟尾暢男 (著) 「The R Tips 第3版: データ解析環境Rの基本技・グラフィックス活用集」オーム社 (2016/10/13) ISBN-13: 978-4274219580
- <http://cse.naro.affrc.go.jp/takezawa/r-tips/r.html>