B04

假設檢定 & 變異數分析

吳漢銘

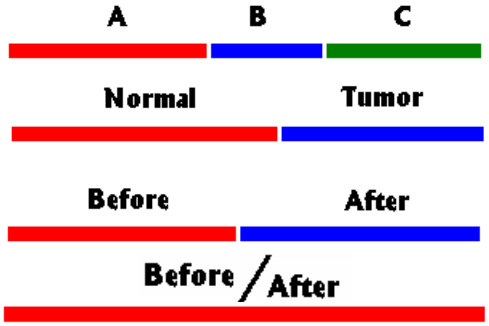
國立政治大學 統計學系

http://www.hmwu.idv.tw

本章大綱

- 簡介統計假設檢定 (Hypothesis Testing)
- 倍數變化 (Fold-Change)
- 平均數檢定 (t檢定)
  - 單樣本、成對樣本、雙樣本
- 單因子變異數分析 (One-way Analysis of Variance, ANOVA)
- 無母數檢定 (Non-parametric Tests)
  - Sign Test，Wilcoxon Signed-Rank Test (paired), Mann-Whitney Test, Kruskal-Wallis Test
- 事後比較檢定 (Post Hoc Tests)
  - Student-Newman-Keuls (SNK) Test, Tukey's HSD Test
- Chi-squared Test, Permutation Test

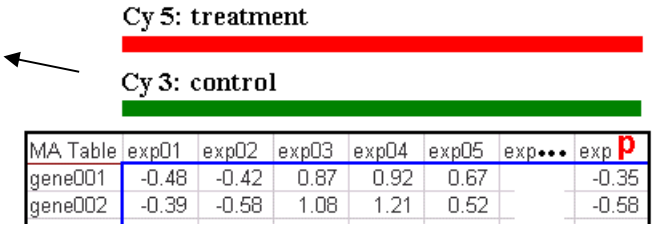
假設檢定的應用:
Finding Differentially Expressed Genes



→ More than two samples

→ Two-sample (independent)

→ Paired-sample (dependent)



MA Table	exp01	exp02	exp03	exp04	exp05	exp•••	exp p
gene001	-0.48	-0.42	0.87	0.92	0.67		-0.35
gene002	-0.39	-0.58	1.08	1.21	0.52		-0.58

MA Table	exp01	exp02	exp03	exp04	exp05	exp•••	exp p	p-values	
gene001	-0.48	-0.42	0.87	0.92	0.67		-0.35	0.067	
gene002	-0.39	-0.58	1.08	1.21	0.52		-0.58	0.052	
gene003	0.87	0.25	-0.17	0.18	-0.13		-0.13	0.013	*
gene004	1.57	1.03	1.22	0.31	0.16		-1.02	0.016	*
gene005	-1.15	-0.86	1.21	1.62	1.12		-0.44	0.112	
gene006	0.04	-0.12	0.31	0.16	0.17		0.08	0.017	*
gene007	2.95	0.45	-0.40	-0.66	-0.59		-0.76	0.059	
gene008	-1.22	-0.74	1.34	1.50	0.63		-0.55	0.063	
gene009	-0.73	-1.06	-0.79	-0.02	0.16		0.03	0.516	
gene010	-0.58	-0.40	0.13	0.58	-0.09		-0.45	0.009	*
gene011	-0.50	-0.42	0.66	1.05	0.68		0.01	0.068	
gene012	-0.86	-0.29	0.42	0.46	0.30		-0.63	0.030	*
gene013	-0.16	0.29	0.17	-0.28	-0.02		-0.04	0.002	*
gene014	-0.36	-0.03	-0.03	-0.08	-0.23		-0.21	0.423	
gene015	-0.72	-0.85	0.54	1.04	0.84		-0.64	0.084	
gene016	-0.78	-0.52	0.26	0.20	0.48		0.27	0.048	
gene017	0.60	-0.55	0.41	0.45	0.18		-1.02	0.018	*
gene018	-0.20	-0.67	0.13	0.10	0.38		0.05	0.538	
gene019	-2.29	-0.64	0.77	1.60	0.53		-0.38	0.053	
gene020	-1.46	-0.76	1.08	1.50	0.74		-0.70	0.074	
gene021	-0.57	0.42	1.03	1.35	0.64		-0.40	0.764	
gene022	-0.11	0.13	0.41	0.60	0.23		0.19	0.423	
gene•••									
gene n	-1.79	0.94	2.13	1.75	0.23		-0.66	0.723	

Microarray Data Matrix

gene001	-0.48	-0.42	0.87	0.92	0.67		-0.35

ranking →

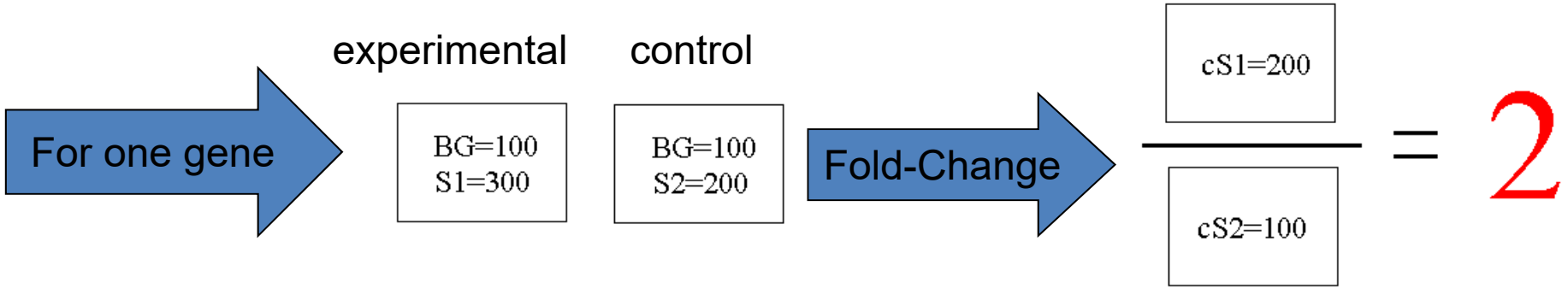
p-values or Statistics

0.002	*
0.009	*
0.013	*
0.016	*
0.017	*
0.018	*
0.030	*
0.048	
0.052	
0.059	
0.063	
0.067	
0.068	
0.074	
0.084	
0.112	
0.423	
0.423	
0.516	
0.538	
0.723	
0.764	
..	

fix number

above some level

Fold-Change Method: Compare Two Sample Means



1) Calculate fold-change.

2) Rank the genes.

3) Select genes.

```
> exp.m <- apply(df[, index.exp], 1, mean)
> ctl.m <- apply(df[, index.ctl], 1, mean)
> plot(exp.m, ctl.m)
> abline(a=0, b=1)
> fc <- exp.m/ctl.m
> no.genes <- 50
> sort(fc, decreasing = TRUE)[1:no.genes]
```

<dno't run>

Fold-Change Method

Method 1: Select genes based on Numbers

- average differential expression > **FC.**

Problems:

- **FC** is an arbitrary threshold.
- **FC** does not take into account individuals and sample size.

Example:

- s2 (200) close to BG (100), the difference could represent noise.
- credible: a gene is regulated 2-fold with 10000, 5000 units.

Method 2: Select genes based on %

- Choose 5% of genes that have the largest expression ratios.

Problems:

- Possible that no genes have statistically significantly different gene expression.

Hypothesis Testing (1)

Hypothesis Test

a procedure for determining if an assertion about a characteristic of a population is reasonable.

Example

"average price of a gallon of regular unleaded gas in Massachusetts is $2.5"

Is this statement true?

- find out every gas station.
- find out a small number of randomly chosen stations.

Sample average price was $2.2.

- Is this 30 cent difference a result of chance variability, or
- is the original assertion incorrect?

Hypothesis Testing (2)

null hypothesis:

- H_0: $\mu = 2.5$. (the average price of a gallon of gas is \$2.5)
- H_0: $\mu_A - \mu_B = \mu_0$.

alternative hypothesis:

- H_a: $\mu > 2.5$. (gas prices were actually higher)
- H_a: $\mu < 2.5$.
- H_a: μ != 2.5.

significance level (alpha):

- Decide in advance.
- Alpha = 0.05: the probability of incorrectly rejecting the null hypothesis when it is actually true is 5%.

Hypothesis Testing (3)

Biological Question → ***Statistical Formulation***

H_0: No differential expressed.

H_0: no difference in the mean gene expression in the group tested.

H_0: The gene will have equal means across every group.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ (…= μ_n)

- **A p-value=0.05 indicates that you would have only a 5% chance of drawing the sample being tested if the null hypothesis was actually true.**
- **The p-value is the smallest level of significance at which a null hypothesis may be rejected**

H_0: no differential expressed.

- The test is significant
  = Reject **H_0**
- False Positive
  = (Reject **H_0** | **H_0** true)
  = concluding that a gene is differentially expressed when in fact it is not.

The *p*-values

p-values

- probability of **false positives** (Reject $\mathbf{H_0}$ | $\mathbf{H_0}$ true).
- probability of observing your data under the assumption that the null hypothesis is true.
- *p-value* = 0.03: only a 3% chance of drawing the sample if the null hypothesis was true.

Decision Rule

- Reject H_0 if *p-value* is less than alpha.
- $P < 0.05$ commonly used. (Reject $\mathbf{H_0}$, the test is significant)
- The lower the *p-value*, the more significant.

p-value 的定義是：在已知(現有)的抽樣樣本下，能棄卻 H_0(虛無假設)的最小顯著水準。

p-value：若(前提) H_0 為真，則 test statistic 出現的可能性。(若p-value越小，表示抽樣樣本越(極端)不可能出現，因此推翻前提，拒絕H_0)。

p-value：以現有的抽樣所進行的推論，可能犯 type I error 的機率。(若p-value越小，表示拒絕H_0不太可能錯，因此拒絕H_0)。

林澤民，看電影學統計：p值的陷阱
(The Pitfalls of p-Values)
http://blog.udn.com/nilnimest/84404190
社會科學論叢2016年10月第十卷第二期
社會科學前沿課題論壇

"只要是使用正確的意義，p-value並沒有問題，只是不要去誤用它。不要只是著重在統計顯著性，因為model對錯的機率跟p-value不一樣。要使用p-value作檢定，要把它跟α來做比較，所以問題不只是p-value，而是α。界定了α之後，才知道結果是不是顯著。當得到一個顯著的結果以後，必須再來衡量偽陽性反機率的問題，也就是model後設機率的問題，這就不是p-value可以告訴你的。"

Type of Errors

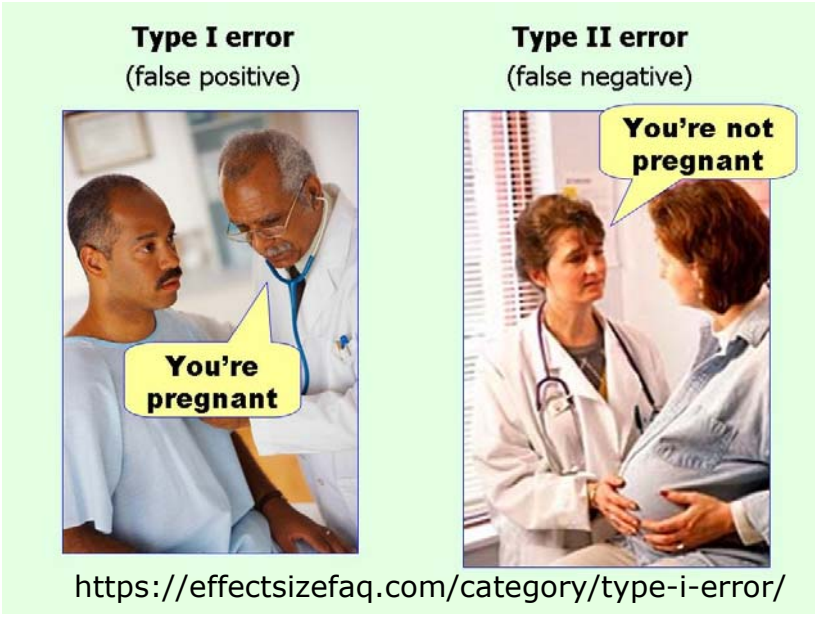
Type I Error (alpha)

calling genes as differentially expressed when they are NOT (when you see things that are not there.)

Type II Error

NOT calling genes as differentially expressed when they ARE (when you dont see things that are there)

Hypothesis Testing		Truth	
		H_0	H_1
Decision	Reject H_0	Type I Error (alpha) (false positive)	Right Decision (true positive)
	Don't Reject H_0	Right Decision	Type II Error (beta)



https://effectsizefaq.com/category/type-i-error/

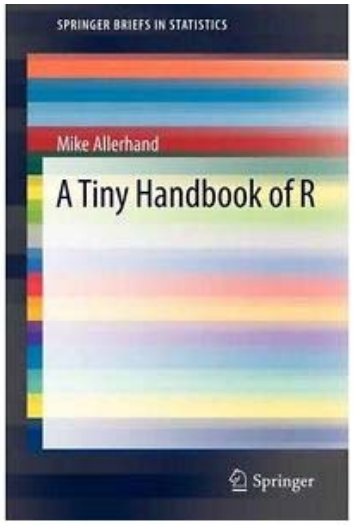
H_0: Not Pregnant

$$\text{Power} = 1 - \beta.$$

The Hypothesis Tests in Base R

The hypothesis tests provided in the base installation include[1]:

Hypothesis tests	
t.test	one and two-sample t tests
wilcox.test	one and two sample Wilcoxon tests
var.test	one and two sample F-tests of variance
cor.test	Correlation coefficient and p-value (Pearson's, Spearman's, or Kendall's)
binom.test	Sign test of a binomial sample
prop.test	Binomial test for comparing two proportions
chisq.test	Chi-squared test for count data
fisher.test	Fisher's exact test for count data
friedman.test	Friedman's rank sum test
kruskal.test	Kruskal–Wallis rank sum test
ks.test	1 or 2-sample Kolmogorov–Smirnov tests



平均數檢定 in R

Hypothesis Testing	One Sample	Two Samples		> two Groups
	-	Paired data	Unpaired data	Complex data
Parametric (variance equal)	t-test t.test(x, mu = 0)	t-test t.test(x, y, var.equal = TRUE) t.test(x, y, paired = TRUE, var.equal = TRUE)	t-test t.test(x, y, var.equal = TRUE)	One-Way Analysis of Variance (ANOVA) aov(x~g, data) oneway.test(x~g, data, var.equal = TRUE)
Parametric (variance not equal)		Welch t-test t.test(x, y) t.test(x, y, paired = TRUE)	Welch t-test t.test(x, y)	Welch ANOVA oneway.test(x~g, data)
Non-Parametric (無母數檢定)	Wilcoxon Signed-Rank Test wilcox.test(x, mu = 0)	Wilcoxon Signed-Rank Test wilcox.test(x, y) wilcox.test(x, y, paired = TRUE)	Wilcoxon Rank-Sum Test (Mann-Whitney U Test) wilcox.test(x, y)	Kruskal-Wallis Test kruskal.test(x, g)

pairwise.t.test {stats}: Calculate pairwise comparisons between group levels with corrections for multiple testing

TukeyHSD {stats}: Compute Tukey Honest Significant Differences

Steps of Hypothesis Testing

1. Determine the null and alternative hypothesis, using mathematical expressions if applicable.
2. Select a significance level (alpha).
3. Take a random sample from the population of interest.
4. Calculate a test statistic from the sample that provides information about the null hypothesis.
5. Decision

Hypothesis Testing:

two-sided z-test & p-value

H_0: $\mu = 35$ null hypothesis

H_1: $\mu \neq 35$ alternative hypothesis ($\mu > 35$; $\mu < 35$) one-sided

α signifcant level : =0.05

test statistic $z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

Reject Ho if $|z| > z_{0.05}$

$$H_0 : \mu = m$$
$$H_1 : \mu \neq m$$
$$\alpha = P_{H_0}(|Z| > z_{\alpha/2})$$

Sample Data: =33.6

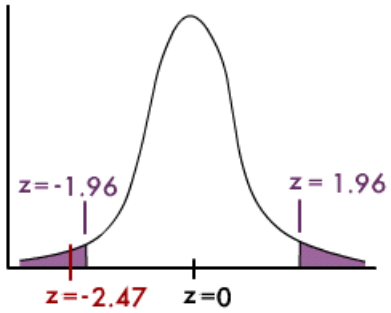
test statistic: z=-2.47

$(1-\alpha)100\%$ Confidence Interval:

$$P(z_{\alpha/2} < Z < z_{1-\alpha/2}) = 1 - \alpha$$

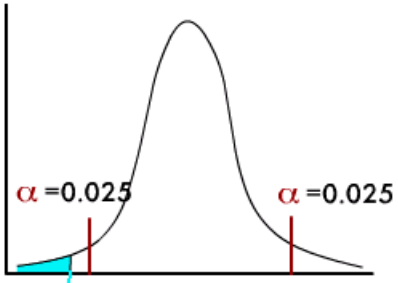
p-value $= P_{H_0}(|Z| > z_0)$, $z_0 = \frac{\bar{X} - m}{\sigma/\sqrt{n}}$

The Classical Approach



Conclusion: since the z value of the test statistic (-2.47) is less than the critical value of z= -1.96, we reject the null hypothesis.

The *P*-Value Approach



P -value = 0.0068 times 2 (for a 2-sided test) = 0.0136

Conclusion: since the *P* -value of 0.0136 is less than the significance level of α=0.05, we reject the null hypothesis.

檢定力 (Statistical Power)

- **Question**: What if I do a t-test on a pair of samples and **fail to reject the null hypothesis**--does this mean that there is no significant difference?
- **Answer**: Maybe yes, maybe no.

- For two-sample t-test, power is the probability of rejecting the hypothesis that the means are equal when they are in fact not equal.
  $P(RH_0 \mid \text{not } H_0)$ = Power = 1 -P(Type-II error)
- The power of the test depends upon the sample size, the magnitudes of the variances, the alpha level, and the actual difference between the two population means.
- Usually you would only consider the power of a test when you failed to reject the null hypothesis.
- High power is desirable (0.7 to 1.0): High power means that there is a high probability of rejecting the null hypothesis when the null hypothesis is false.

See also: `power.t.test {stats}`: Power calculations for one and two sample t tests.

One Sample t-test

Assumption: the variable is normally distributed.

One sample t-test

$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$ (two-tailed).
μ: population mean.
α: significant level (e.g., 0.05).
Test Statistic:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}, \quad t_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

$\bar{X}$: sample mean.

S: sample standard deviation.

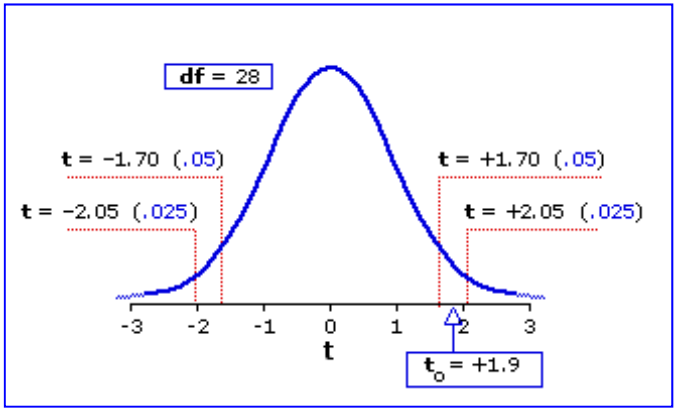
n: number of observations in the sample.

- Reject H_0 if $|t_0| > t_{\alpha/2,n-1}$.
- Power $= 1 - \beta$.
- $(1-\alpha)100\%$ Confidence Interval for μ: $\bar{X} - t_{\alpha/2}S/\sqrt{n} \leq \mu < \bar{X} + t_{\alpha/2}S/\sqrt{n}$
- $p\text{-}value = P_{H_0}(|\mathbf{T}| > t_0)$, $\mathbf{T} \sim t_{n-1}$.

Question

- whether a gene is differentially expressed for a condition with respect to baseline expression?
- H_0: μ=0 (log ratio)

MA Table	exp01	exp02	exp03	exp04	exp05	exp•••	exp p
gene001	-0.48	-0.42	0.87	0.92	0.67		-0.35
gene002	-0.39	-0.58	1.08	1.21	0.52		-0.58
gene003	0.87	0.25	-0.17	0.18	-0.13		-0.13



Two Sample t-test

Paired Sample t-test

$H_0 : \mu_d = \mu_0$
$H_1 : \mu_d \neq \mu_0$ (two-tailed).
μ_d: mean of population differences.
α: significant level (e.g., 0.05).
Test Statistic:

$$T_d = \frac{\bar{d} - \mu_d}{S_d/\sqrt{n}}, \quad t_d = \frac{\bar{d} - \mu_0}{S_d/\sqrt{n}}$$

$\bar{d}$: average of sample differences.

S_d: standard deviation of sample difference

n: number of pairs.

- Reject H_0 if $|t_d| > t_{\alpha/2,n-1}$.
- Power $= 1 - \beta$.
- $(1-\alpha)100\%$ Confidence Interval for μ_d:
  $\bar{d} - t_{\alpha/2} S/\sqrt{n} \leq \mu_d < \bar{d} + t_{\alpha/2} S/\sqrt{n}$
- $p\text{-}value = P_{H_0}(|\mathbf{T}| > t_d)$, $\mathbf{T} \sim t_{n-1}$.

Two Sample t-test (Unpaired)

$H_0 : \mu_x - \mu_y = \mu_0$
$H_0 : \mu_x - \mu_y \neq \mu_0$
α: significant level (e.g., 0.05).
Test Statistic:

$$t_0 = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}}$$

for homogeneous variances:
$df = n + m - 2$

for heterogeneous variances:
adjusted df

Reject H_0 if $|t_0| > t_{\alpha/2,\, df}$

Assumptions of t-test

Be Normal

- paired t-test,
  the distribution of the subtracted data that must be normal.
- unpaired t-test,
  the distribution of both data sets must be normal.

How to Detect Normality

- **Plots**: Histogram, Density Plot, QQplot,…
- **Test for Normality**: Jarque-Bera test, Lilliefors test, Kolmogorov-Smirnov test.

Homogeneous

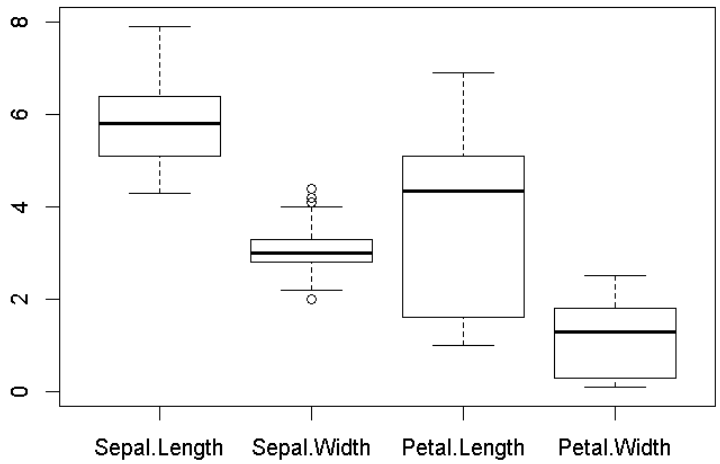
- the variances of the two population are equal.
- Test for equality of the two variances: Variance ratio F-test.

t.test {stats}: Student's t-Test

Description: Performs one and two sample t-tests on vectors of data.

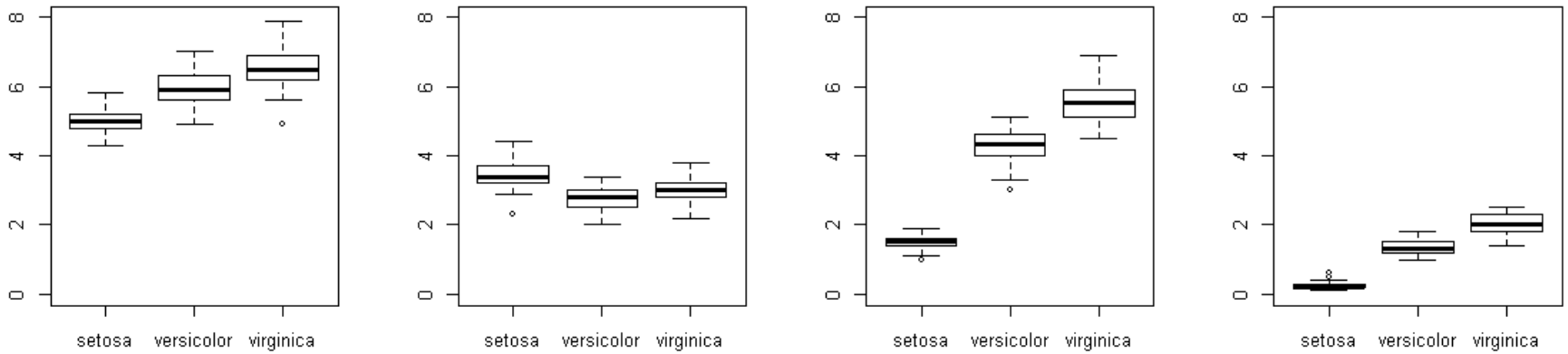
Usage:
```
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)
```

```
> x <- iris$Sepal.Length
> y <- iris$Petal.Length
> alpha <- 0.05
> (vt <- (var.test(x, y)$p.value <= alpha))
[1] TRUE
> t.test(x, y, var.equal = !vt )

        Welch Two Sample t-test

data:  x and y
t = 13.098, df = 211.54, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.771500 2.399166
sample estimates:
mean of x mean of y
 5.843333  3.758000
```



var.test {stats}: Performs an F test to compare the variances of two samples from normal populations.
H_0: the ratio of the variances of the populations from which x and y were drawn, or in the data to which the linear models x and y were fitted, is equal to ratio=1.

Using t.test(x ~ g)




```
> myData <- data.frame(value = iris$Sepal.Width[-(1:50)],
+  group <- iris[-(1:50), 5])
> alpha <- 0.05
> (bt <- bartlett.test(value ~ group, data=myData)$p.value <= alpha)
[1] FALSE
> t.test(value ~ group, data=myData, var.equal=!bt)
        Two Sample t-test

data:  value by group
t = -3.2058, df = 98, p-value = 0.001819
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.33028246 -0.07771754
sample estimates:
mean in group versicolor  mean in group virginica
                   2.770                    2.974
```

Test Homogeneity of Variances

- **`var.test {stats}`**: an F test to compare the variances of two samples from normal populations.
- **`bartlett.test {stats}`**: a parametric test of the null that the variances in each of the groups (samples) are the same.
- **`ansari.test {stats}:`** Ansari-Bradley two-sample test for a difference in scale parameters. (testing for equal variance for non-normal samples)
- **`mood.test {stats}:`** another rank-based two-sample test for a difference in scale parameters.
- **`fligner.test {stats}`**: Fligner-Killeen (median) is a rank-based (nonparametric) k-sample test for homogeneity of variances.
- **`leveneTest {car}`**: Levene's test for homeogeneity of variance across groups.

- NOTE: Fligner-Killeen's and Levene's tests are two ways to test the ANOVA assumption of "equal variances in the population" before conducting the ANOVA test.
- Levene's is widely used and is typically the default in SPSS.

Other t-Statistics

B-statistic

Lonnstedt and Speed, Statistica Sinica 2002: parametric empirical Bayes approach.

- B-statistic is an estimate of the posterior log-odds that each gene is DE.
- B-statistic is equivalent for the purpose of ranking genes to the penalized t-statistic $t = \frac{\bar{M}}{\sqrt{(a+s^2)/n}}$, where a is estimated from the mean and standard deviation of the sample variances s^2.

$$M_{gj}|\mu_g, \sigma_g \sim N(\mu_g, \sigma_g^2)$$

$$B_g = \log \frac{P(\mu_g \neq 0|M_{gj})}{P(\mu_g = 0|M_{gj})}$$

Lonnstedt, I. and Speed, T.P. Replicated microarray data. *Statistica Sinica* , 12: 31-46, 2002

Penalized t-statistic

Tusher et al (2001, PNAS, SAM)
Efron et al (2001, JASA)

$$t = \frac{\bar{M}}{(a+s)/\sqrt{n}}$$

General Penalized t-statistic

(Lonnstedt et al 2001)

$$t = \frac{b}{s^* \times SE}$$

multiple regression model

Penalized two-sample t-statistic

$$t = \frac{\bar{M}_A - \bar{M}_B}{s^* \times \sqrt{1/n_A + 1/n_B}}, \quad \text{where } s^* = \sqrt{a + s^2}$$

Robust General Penalized t-statistic

單因子變異數分析 (One-Way ANOVA)

- ANOVA can be considered to be a generalization of the *t*-test, when
  - compare more than two groups (e.g., *drug 1*, *drug 2*, and *placebo*), or
  - compare groups created by more than one independent variable while controlling for the separate influence of each of them (e.g., *Gender*, *type of Drug*, and *size of Dose*).

- One-way ANOVA compares groups using one parameter.
- ANOVA can test the following:
  - Are all the means from more than two populations equal?
  - Are all the means from more than two treatments on one population equal?
  - (This is equivalent to asking whether the treatments have any overall effect.)

One-Way ANOVA

- Assumptions
  - The subjects are sampled randomly.
  - The groups are independent.
  - The population variances are homogenous.
  - The population distribution is normal in shape.

- As with t-tests, violation of homogeneity is particularly a problem when we have quite different sample sizes.

- Homogeneity of variance test
  - Bartlett's test (1937)
  - Levene's test (Levene 1960)
  - O'Brien (1979)
  - ...

ANOVA Table

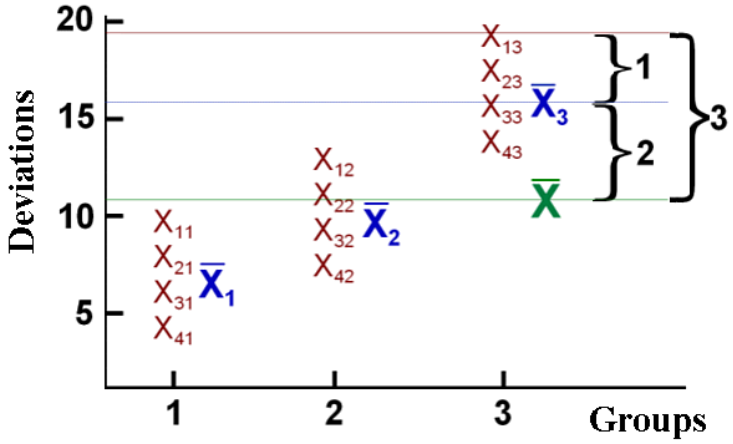
Groups

1	2	...	j	...	k
X_{11}	X_{12}	$\cdots$	X_{1j}	$\cdots$	X_{1k}
X_{21}	X_{22}	$\cdots$	X_{2j}	$\cdots$	X_{2k}
			$\cdots$		
X_{i1}	X_{i2}	$\cdots$	X_{ij}	$\cdots$	X_{ik}
$\vdots$	$X_{n_2 2}$	$\cdots$	$\vdots$	$\cdots$	$X_{n_k k}$
$X_{n_1 1}$			$X_{n_i j}$		

$$T_j = \sum_{i=1}^{n_j} X_{ij} \qquad \bar{X}_j = \frac{T_j}{n_j}$$

$$T = \sum_{j=1}^{k} T_j \qquad \bar{X} = \frac{T}{N}$$

$$S^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j} \frac{(X_{ij} - \bar{X})^2}{N-1}$$



$$(X_{ij} - \bar{X}) = (X_{ij} - \bar{X}_j) + (\bar{X}_j - \bar{X})$$

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

$$X_{ij} = \mu_j + \epsilon_{ij} \qquad i = 1, \cdots, n_j \quad j = 1, \cdots, k$$

$$\epsilon_{ij} \sim N(0, \sigma^2)$$

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - \bar{X})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}[(X_{ij} - \bar{X}_j) + (\bar{X}_j - \bar{X})]^2$$

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - \bar{X})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_j)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{X}_j - \bar{X})^2$$

ANOVA Table

Source	SS	df	MS	F	p
Between	SS_B	$p-1$	MS_B	MS_B/MS_W	< 0.05
Within	SS_W	$N-p$	MS_W		
Total	SS_T	$N-1$			

$$SS_{Total} = SS_{Within} + SS_{Between}$$

$$F = \frac{MS_{Between}}{MS_{Within}}$$

Reject H_0, if $F_{obs} > F_{\{\alpha, k-1, N-k\}}$

Welch ANOVA

Welch's F Test

- Use when the sample sizes are unequal.
- Use when the sample sizes are equal but small.

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

$$X_{ij} = \mu_j + \epsilon_{ij}$$

$$\epsilon_{ij} \sim N(0, \sigma_j^2)$$

$$i = 1, \cdots, n_j$$

$$j = 1, \cdots, k$$

$$s_j^2 = \frac{\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_j)^2}{n_j - 1}$$

$$w_j = \frac{n_j}{s_j^2}$$

$$\bar{X}' = \frac{\sum_{j=1}^{k} w_j \bar{X}_j}{\sum_{j=1}^{k} w_j}$$

$$F' = \frac{\frac{\sum_{j=1}^{k} w_j (\bar{X}_j - \bar{X}')^2}{k-1}}{1 + \frac{2(k-2)}{k^2-1} \sum_{j=1}^{k} \left(\frac{1}{n_j - 1}\right)\left(1 - \frac{w_j}{\sum_{j=1}^{k} w_j}\right)^2}$$

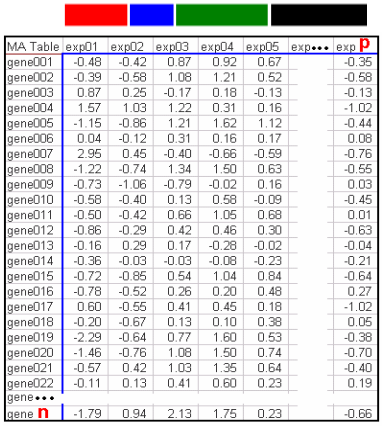
$$df' = \frac{k^2 - 1}{3 \sum_{j=1}^{k} \left(\frac{1}{n_j - 1}\right)\left(1 - \frac{w_j}{\sum_{j=1}^{k} w_j}\right)^2}$$

Reject H_0, if $F'_{obs} > F_{\{\alpha, k-1, df'\}}$

Small Round Blue Cell Tumors (SRBCT) Dataset

cDNA Microarrays

- ***#Samples:*** 63 兒童小圓藍細胞腫瘤
  four types of SRBCT of childhood:
  - Neuroblastoma (NB) (12),
  - Non-Hodgkin lymphoma (NHL) (8),
  - Rhabdomyosarcoma (RMS) (20)
  - Ewing tumours (EWS) (23).
- ***#Genes***: 6567 genes



MA Table	exp01	exp02	exp03	exp04	exp05	exp•••	exp p
gene001	-0.48	-0.42	0.87	0.92	0.67		-0.35
gene002	-0.39	-0.58	1.08	1.21	0.52		-0.58
gene003	0.87	0.25	-0.17	0.18	-0.13		-0.13
gene004	1.57	1.03	1.22	0.31	0.16		-1.02
gene005	-1.15	-0.86	1.21	1.62	1.12		-0.44
gene006	0.04	-0.12	0.31	0.16	0.17		0.08
gene007	2.95	0.45	-0.40	-0.66	-0.59		-0.76
gene008	-1.22	-0.74	1.34	1.50	0.63		-0.55
gene009	-0.73	-1.06	-0.79	-0.02	0.16		0.03
gene010	-0.58	-0.40	0.13	0.58	-0.09		-0.45
gene011	-0.50	-0.42	0.66	1.05	0.68		0.01
gene012	-0.86	-0.29	0.42	0.46	0.30		-0.63
gene013	-0.16	0.29	0.17	-0.28	-0.02		-0.04
gene014	-0.36	-0.03	-0.03	-0.08	-0.23		-0.21
gene015	-0.72	-0.85	0.54	1.04	0.84		-0.64
gene016	-0.78	-0.52	0.26	0.20	0.48		0.27
gene017	0.60	-0.55	0.41	0.45	0.18		-1.02
gene018	-0.20	-0.67	0.13	0.10	0.38		0.05
gene019	-2.29	-0.64	0.77	1.60	0.53		-0.38
gene020	-1.46	-0.76	1.08	1.50	0.74		-0.70
gene021	-0.57	0.42	1.03	1.35	0.64		-0.40
gene022	-0.11	0.13	0.41	0.60	0.23		0.19
gene•••							
gene n	-1.79	0.94	2.13	1.75	0.23		-0.66

6567 x 63

Interests:

- To identify genes that are differentially expressed in one or more of these four groups.

More on SRBCT:
http://www.thedoctorsdoctor.com/diseases/small_round_blue_cell_tumor.htm

Khan J, Wei J, Ringner M, Saal L, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu C, Peterson C and Meltzer P. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Medicine 2001, 7:673-679
Stanford Microarray Database

Apply ANOVA to SRBCT data

- **khan {made4}**: Microarray gene expression dataset from Khan et al., 2001. Subset of 306 genes.
- http://svitsrv25.epfl.ch/R-doc/library/made4/html/khan.html
- Khan contains gene expression profiles of four types of small round blue cell tumours of childhood (SRBCT) published by Khan et al. (2001). It also contains further gene annotation retrieved from SOURCE at http://source.stanford.edu/.

```
> source("https://bioconductor.org/biocLite.R")
> biocLite("made4")
> library(made4)
> data(khan)
>
> # some EDA works should be done  before ANOVA
> # get the p-value from a anova table
> Anova.pvalues <- function(x){
+   x <- unlist(x)
+   SRBCT.aov.obj <- aov(x ~ khan$train.classes)
+   SRBCT.aov.info <- unlist(summary(SRBCT.aov.obj))
+   SRBCT.aov.info["Pr(>F)1"]
+ }
> # perform anova for each gene
> SRBCT.aov.p <- apply(khan$train, 1, Anova.pvalues)
```

```
if (!requireNamespace("BiocManager", quietly = TRUE))
    install.packages("BiocManager")

BiocManager::install("made4")
```

Apply ANOVA to SRBCT data

```
> # select the top 5 DE genes
> order.p <- order(SRBCT.aov.p)
> ranked.genes <- data.frame(pvalues=SRBCT.aov.p[order.p],
+                            ann=khan$annotation[order.p, ])
> top5.gene.row.loc <- rownames(ranked.genes[1:5,  ])
> # summarize the top5 genes
> summary(t(khan$train[top5.gene.row.loc, ]))
     770394           236282           812105           183337           814526
 Min.   :0.0669   Min.   :0.0364   Min.   :0.1011   Min.   :0.0223   Min.   :0.1804
 1st Qu.:0.3370   1st Qu.:0.1557   1st Qu.:0.3250   1st Qu.:0.1273   1st Qu.:0.4294
 Median :0.6057   Median :0.2412   Median :0.7183   Median :0.2701   Median :0.6677
 Mean   :1.5508   Mean   :0.3398   Mean   :1.1619   Mean   :0.5013   Mean   :0.9640
 3rd Qu.:2.8176   3rd Qu.:0.3563   3rd Qu.:1.5543   3rd Qu.:0.5104   3rd Qu.:1.3620
 Max.   :5.2958   Max.   :1.3896   Max.   :5.9451   Max.   :3.7478   Max.   :3.5809
```

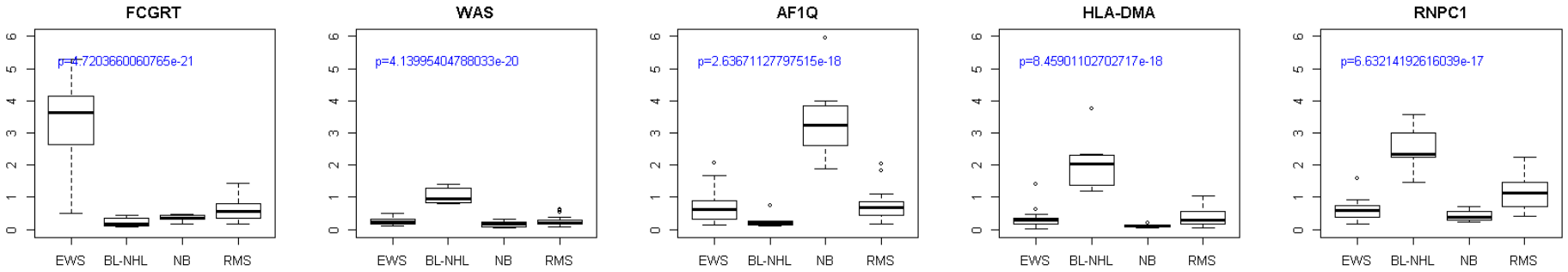
```
> # draw the side-by-side boxplot for top5 DE genes
> par(mfrow=c(1, 5), mai=c(0.3, 0.4, 0.3, 0.3))
> # get the location of xleft, xright, ybottom, ytop.
> usr <- par("usr")
> myplot <- function(gene){
+   # use unlist to convert "data.frame[1xp]" to "numeric"
+   boxplot(unlist(khan$train[gene, ]) ~ khan$train.classes,
+           ylim=c(0, 6), main=ranked.genes[gene, 4])
+   text(2, usr[4]-1, labels=paste("p=", ranked.genes[gene, 1],
+        sep=""), col="blue")
+   ranked.genes[gene,]
+ }
```

(重要技巧) 利用Key (gene.row.loc) 去連結多組資料(train, annotation)。

Apply ANOVA to SRBCT data

```
> # print the top5 DE genes info
> do.call(rbind, lapply(top5.gene.row.loc, myplot))
> # lappay returns "list" and use rbind to convert it to "data.frame"
> # Try sapply?
```

```
> do.call(rbind, lapply(top.gene.row.loc, myplot))
            pvalues ann.CloneID ann.UGCluster ann.Symbol ann.LLID ann.UGRepAcc ann.LLRepProtAcc ann.Chromosome  ann.Cytoband
770394 4.720366e-21      770394     Hs.111903      FCGRT     2217     AK074734        NP_004098             19       19q13.3
236282 4.139954e-20      236282       Hs.2157        WAS     7454     BM455138        NP_000368              X Xp11.4-p11.21
812105 2.636711e-18      812105      Hs.75823       AF1Q    10962     BC022448        NP_006809              1          1q21
183337 8.459011e-18      183337     Hs.351279    HLA-DMA     3108     AK055186        NP_006111          6;10;5          6p21.3
814526 6.632142e-17      814526     Hs.236361      RNPC1    55544    NM_017495        NP_906270             20      20q13.31
```



課堂練習: 試用Kruskal-Wallis Test重覆上述分析。

Non-parametric Statistics

- Do not assume that the data is normally distributed.
- Two good reasons to use non-parametric statistic for microarray data.
  - *Microarray data is noisy:*
    - many sources of variability and outliers are frequent.
    - Non-parametric methods are robust to outliers and noisy data.
  - *Microarray data analysis is high throughput:*
    - Need to check the normality of every gene.
    - Those genes with outliers or which were not normally distributed would then need a different analysis.
    - It makes more sense to apply a test that is distribution free and thus can be applied to all genes in a single pass.

- Sign Test: **`SIGN.test {BSDA}`**
- Wilcoxon Rank Sum (Mann-Whitney U Test) Test: **`wilcox.test {stats}, wilcox_test {coin}`**
- Wilcoxon Signed Rank Tests : **`wilcox.test {stats}, wilcox_test {coin}`**
- Kruskal-Wallis Rank Sum Test: **`kruskal.test {stats}`**

Sign Test

- Given *n* pairs of data, the sign test tests the hypothesis that the median of the differences in the pairs is zero.

- The test statistic is the number of positive differences.

- If the null hypothesis is true, then the numbers of positive and negative differences should be approximately the same.

- In fact, the number of positive differences will have a Binomial distribution with parameters *n* and *p*.

Pair	Before	After	Sign
1	89	73	+
2	83	77	+
3	80	58	+
4	72	77	−
5	77	70	+
6	74	62	+
7	69	67	+
8	65	68	−
9	60	44	+
10	55	50	+
11	54	46	+
12	50	38	+
13	42	47	−
14	48	40	+
15	44	43	+
16	38	29	+
17	36	25	+

The Sign Test:

when $n_1 = n_2 \leq 50$

$H_0 : P = Q = \frac{1}{2}$

$H_1 : P \neq Q \neq \frac{1}{2}$

$T = \#"+"$

At $\alpha = 0.01$, two-tailed test,
reject H_0 if $T \geq 14$ when $N = 17$.
(Binomial Probability)

$\#"+" = 14$
$\#"-" = 3$

The obtained T=14 is equal to the critical value, so we reject H_0.

Wilcoxon Signed-Rank Test (paired)

- Null hypothesis: the population median from which both samples were drawn is the same.

- The sum of the ranks for the "positive" (up-regulated) values is calculated and compared against a precomputed table to a p-value.
  - Sorting the absolute values of the differences from smallest to largest.
  - Assigning ranks to the absolute values.
  - Find the sum of the ranks of the positive differences.
- If the null hypothesis is true, the sum of the ranks of the positive differences should be about the same as the sum of the ranks of the negative differences.

Pair	Before	After	Diff.	Rank
1	89	73	16	15.5
2	83	77	6	7
3	80	58	22	17
4	72	77	−5	5
5	77	70	7	8
6	74	62	12	13.5
7	69	67	2	2
8	65	68	−3	3
9	60	44	16	15.5
10	55	50	5	5
11	54	46	8	9.5
12	50	38	12	13.5
13	42	47	−5	5
14	48	40	8	9.5
15	44	43	1	1
16	38	29	9	11
17	36	25	11	12

The Wilcoxon signed-rank Test:

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$

$T = \min\{\sum_+ \text{Rank}, \sum_- \text{Rank}\}$

At $\alpha = 0.01$, two-tailed test,
reject H_0 if $T \neq 23$ when $N = 17$.
(Table)

(The zero difference is ignored when assigning ranks. $N_{new} = N_{old} - \#\{ties\}$)

$T = \min\{\sum_+ \text{Rank} = 140, \sum_- \text{Rank} = 13\}$
$= 13$

The obtained T=13 is less than the critical value 23, so we reject H_0.

Mann-Whitney Test

(Wilcoxon Rank-Sum Test, unpaired)

- The data from the two groups are combined and given ranks. (1 for the largest, 2 for the second largest,...)

- The ranks for the larger group are summed and that number is compared against a precomputed table to a p-value.

Group		Rank	
G_1	G_2	G_1	G_2
26	16	3	11
22	10	4	17
19	8	7.5	19
21	13	5.5	13.5
14	19	12	7.5
18	11	9	15.5
29	7	2	20
17	13	10	13.5
11	9	15.5	18
34	21	1	5.5
$n_1 = 10$	$n_2 = 10$	$R_1 = 69.5$	$R_2 = 104.5$

The Mann-Whitney U Test:

$H_0 : F_1 = F_2$

$H_1 : F_1 \neq F_2$

$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

or

$U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$

$R_i = \sum_i \text{Rank}$

At $\alpha = 0.05$, two-tailed test for $n_1 = 10, n_2 = 10$, reject H_0 if $U \leq 23$ or $U' \geq 77$ (Table)

U: the number of times that a score from Group 1 is lower in rank than a score from Group 2.

$U = 85.5, \quad U' = 14.5$

The obtained $U = 85.5$ is less than the critical value 77, so we reject H_0.

Kruskal-Wallis Test

- The Kruskal Wallis test can be applied in the one factor ANOVA case. It is a non-parametric test for the situation where the ANOVA normality assumptions may not apply.
- Each of the n_j should be at least 5 for the approximation to be valid.

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

$$H_1 : \mu_i \neq \mu_j \quad \text{for at least one set of } i \text{ and } j$$

$$W = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

$$W \sim \chi^2_{k-1} \text{ under } H_0$$

Reject H_0 if $W > CHIPPF(\alpha, k-1)$, the chi-square percent point function

Groups

1	2	...	j	...	k
X_{11}	X_{12}	$\cdots$	X_{1j}	$\cdots$	X_{1k}
X_{21}	X_{22}	$\cdots$	X_{2j}	$\cdots$	X_{2k}
			$\cdots$		
X_{i1}	X_{i2}	$\cdots$	X_{ij}	$\cdots$	X_{ik}
$\vdots$	$X_{n_2 2}$	$\cdots$	$\vdots$	$\cdots$	$X_{n_k k}$
$X_{n_1 1}$			$X_{n_i j}$		

Rank Data

1	2	...	j	...	k
R_{11}	R_{12}	$\cdots$	R_{1j}	$\cdots$	R_{1k}
R_{21}	R_{22}	$\cdots$	R_{2j}	$\cdots$	R_{2k}
			$\cdots$		
R_{i1}	R_{i2}	$\cdots$	R_{ij}	$\cdots$	R_{ik}
$\vdots$	$R_{n_2 2}$	$\cdots$	$\vdots$	$\cdots$	$R_{n_k k}$
$R_{n_1 1}$			$R_{n_i j}$		

$$F(x) = P(X \leq x) = P(X \leq G(\alpha)) = \alpha$$

$$x = G(\alpha) = G(F(x))$$

The percent point function (ppf) is the inverse of the cumulative distribution function.

Parametric vs. Non-Parametric Test

Parametric Tests

- Assume that the data follows a certain distribution (normal distribution).
- Assuming equal variances and Unequal variances.
- More powerful.
- Not appropriate for data with outliers.

Non-Parametric Tests

- When certain assumptions about the underlying population are questionable (e.g. normality).
- Does not assume normal distribution
- No variance assumption
- Ranks the order of raw/normalized data across conditions for analyses
- Decrease effects of outliers (Robust)
- Not recommended if there is less than 5 replicates per group
- Needs a high number of replicates
- Less powerful

t-test	Non-parametric
Easy	Easy
Powerful	Robust
Widely Implemented	widely implemented
Not appropriate for data with outliers	Less powerful

事後檢定 (Post Hoc Tests)
Student-Newman-Keuls (SNK) Test

assuming
equal sample sizes and
homogeneity of variance

Group	A	B	C	D
Mean	2	3	7	8

alpha = 0.01
$n = 5$
$df = 16$

$$\sqrt{\frac{MSE}{n}} = \sqrt{\frac{.5}{5}} = 0.316$$

snk {mutoss}: {Unified Multiple Testing Procedures}
snk.test {GAD}: {Analysis of variance from general principles}
SNK.test {agricolae}: {Statistical Procedures for Agricultural Research}

"r" is the number of means spanned by a given comparison.

r, df, alpha → studentized range statistic q

1. $r = 4$, $q_{.01} = 5.19$

A vs D: $q = \frac{8-2}{0.316} = 18.99$, $p < 0.01$

2. $r = 3$, $q_{.01} = 4.79$

a. A vs C: $q = \frac{7-2}{0.316} = 15.82$, $p < 0.01$

b. B vs D: $q = \frac{8-3}{0.316} = 15.82$, $p < 0.01$

3. $r = 2$, $q_{.01} = 4.13$

a. A vs B: $q = \frac{3-2}{.316} = 3.16$, $p > 0.01$

b. B vs C: $q = \frac{7-3}{.316} = 12.66$, $p < 0.01$

c. C vs D: $q = \frac{8-7}{.316} = 3.16$, $p > 0.01$

Tukey's HSD Test

- To test all pairwise comparisons among means using the Tukey Honestly Significant Difference, calculate HSD for each pair of means using the following formula: $\frac{M_1 - M_2}{\sqrt{MS_w\left(\frac{1}{n}\right)}}$

  (1) Mi – Mj is the difference between the pair of means.
  (2) MS_w is the Mean Square Within, and n is the number in the group or treatment.

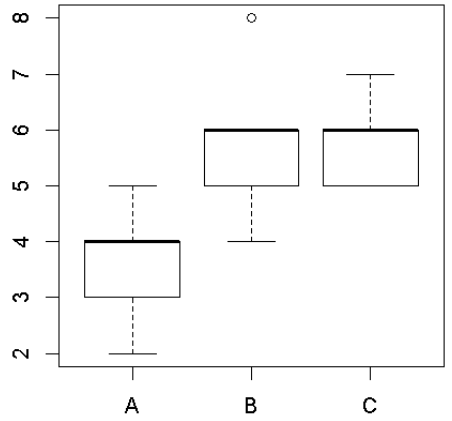
- **Steps:**
  - Step 1: Perform the ANOVA test. Assuming your F value is significant, you can run the post hoc test.
  - Step 2: Choose two means from the ANOVA output.
  - Step 3: Calculate the HSD statistic for the Tukey test using the formula.
  - Step 4: Find the score in Tukey's critical value table.
  - Step 5: Compare the score you calculated in Step 3 with the tabulated value you found in Step 4. If the calculated value from Step 3 is bigger than the critical value from the critical value table, the two means are significantly different.
- **Assumptions for the test**
  - Observations are independent within and among groups.
  - The groups for each mean in the test are normally distributed.
  - There is equal within-group variance across the groups associated with each mean in the test (homogeneity of variance).
- **Tukey's test and SNK test**
  - All alpha's in Tukey's test are compared to the same critical value.
  - All alpha's in SKN test are compared to a different critical value.
  - This test is more conservative (less powerful) than the SNK test.

範例: ANOVA + Post Hoc Test

- A drug company tested three formulations of a pain relief medicine for migraine headache sufferers. For the experiment 27 volunteers were selected and 9 were randomly assigned to one of three drug formulations. The subjects were instructed to take the drug during their next migraine headache episode and to report their pain on a scale of 1 to 10 (10 being most pain).

Drug A:	4	5	4	3	2	4	3	4	4
Drug B:	6	8	4	5	4	6	5	8	6
Drug C:	6	7	6	6	7	5	6	5	5

```
> pain <- c(4, 5, 4, 3, 2, 4, 3, 4, 4, 6, 8, 4, 5,
+  4, 6, 5, 8, 6, 6, 7, 6, 6, 7, 5, 6, 5, 5)
> drug <- c(rep("A", 9), rep("B", 9), rep("C", 9))
> migraine <- data.frame(pain, drug)
> plot(pain ~ drug, data=migraine)
> migraine.aov <- aov(pain ~ drug, data=migraine)
> summary(migraine.aov)
            Df Sum Sq Mean Sq F value   Pr(>F)
drug         2  28.22  14.111   11.91 0.000256 ***
Residuals   24  28.44   1.185
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # reject the null hypothesis of equal means for all three drug groups
```




```
> kruskal.test(pain ~ drug, data=migraine)
        Kruskal-Wallis rank sum test
data:  pain by drug
Kruskal-Wallis chi-squared = 14.395, df = 2, p-value = 0.0007483
```

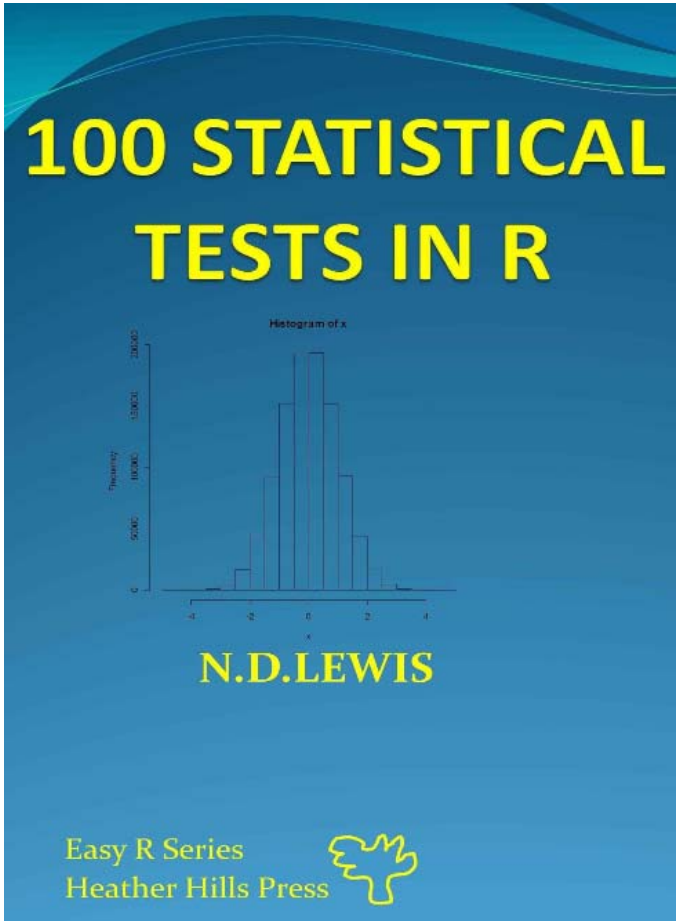
Pairwise Comparisons

```
> pairwise.t.test(pain, drug, p.adjust="bonferroni")

        Pairwise comparisons using t tests with pooled SD

data:  pain and drug

  A       B
B 0.00119 -
C 0.00068 1.00000

P value adjustment method: bonferroni
>
> TukeyHSD(migraine.aov)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = pain ~ drug, data = migraine)

$drug
          diff        lwr      upr     p adj
B-A 2.1111111  0.8295028 3.392719 0.0011107
C-A 2.2222222  0.9406139 3.503831 0.0006453
C-B 0.1111111 -1.1704972 1.392719 0.9745173
>
> # conclude that the mean pain is significantly different for drug A
```

課堂練習:

試用SNK Test 和 Tukey's HSD Test分析SRBCT data。

卡方檢定: chisq.test



N.D Lewis, 100 Statistical Tests in R, Publisher: CreateSpace Independent Publishing Platform (April 15, 2013)

卡方檢定: chisq.test

- 適合度檢定(test of goodness of fit): 檢定資料是否符合某個比例關係或某個機率分佈
- 齊一性檢定(test of homogeneity): 檢定幾個不同類別中的比例關係是否一致
- 獨立性檢定(test of independence): 檢定兩個分類變數之間是否互相獨立。

chisq.test {stats}: Pearson's Chi-squared Test for Count Data

Description:

chisq.test performs chi-squared contingency table tests and goodness-of-fit tests.

Usage:

```
chisq.test(x, y = NULL, correct = TRUE, p =
rep(1/length(x), length(x)), rescale.p = FALSE,
simulate.p.value = FALSE, B = 2000)
```

Chi-Square Test for Independence

- H_0: In the population, the two categorical variables are independent.
- H_a: In the population, two categorical variables are dependent.

For testing independence in $I \times J$ contingency tables

$$H_0: \pi_{ij} = \pi_{i+}\pi_{+j} \quad \text{for all } i \text{ and } j$$

$\mu_{ij} = n\pi_{ij} = n\pi_{i+}\pi_{+j}$ as the expected frequency.

estimated expected frequencies.

$$\hat{\mu}_{ij} = np_{i+}p_{+j} = n\left(\frac{n_{i+}}{n}\right)\left(\frac{n_{+j}}{n}\right) = \frac{n_{i+}n_{+j}}{n}$$

The *Pearson chi-squared statistic* for testing H_0 is

$$X^2 = \sum \frac{(n_{ij} - \mu_{ij})^2}{\mu_{ij}}$$

The X^2 statistic has approximately a chi-squared distribution, for large n. **(WHY?)**

Table 2.5. Cross Classification of Party Identification by Gender

	Party Identification			
Gender	Democrat	Independent	Republican	Total
Females	762 (703.7)	327 (319.6)	468 (533.7)	1557
Males	484 (542.3)	239 (246.4)	477 (411.3)	1200
Total	1246	566	945	2757

Note: Estimated expected frequencies for hypothesis of independence in parentheses. Data from 2000 General Social Survey.

```
> M <- as.table(rbind(c(762, 327, 468),
                      c(484, 239, 477)))
> dimnames(M) <- list(gender = c("F", "M"),
+                     party = c("Democrat",
                               "Independent",
                               "Republican"))
> M
      party
gender Democrat Independent Republican
     F      762         327        468
     M      484         239        477
> (res <- chisq.test(M))
        Pearson's Chi-squared test

data:  M
X-squared = 30.07, df = 2, p-value = 2.954e-07
```

Permutation Test (randomization or re-randomization tests)

- The permutation test is a test where the null-hypothesis allows to reduce the inference to a randomization problem.
- The outcome data are analyzed many times (once for each acceptable assignment that could have been possible under H_0) and then compared with the observed result, without dependence on additional distributional or model-based assumptions.
- **Perform a permutation test (general):**
  1. Analyze the problem, choice of null-hypothesis
  2. Choice of test statistic **T**
  3. Calculate the value of the test statistic for the observed data: $\mathbf{t_{obs}}$
  4. Apply the randomization principle and look at all possible permutations, this gives the distribution of the test statistic **T** under H_0.
  5. Calculation of p-value:

$$p = \mathrm{P}(\mathrm{T} \geq \mathrm{t_{obs}} \mid \mathrm{H_0}) \approx \frac{\#\{\mathrm{t}^* \geq \mathrm{t_{obs}}\}}{\#\ \text{permutations}}$$

Ref: Mansmann, U. (2002), Practical microarray analysis: resampling and the Bootstraap.Heidelberg.

Permutation Test

Coexpression of genes

H_0: Gene 1 and Gene 2 are not correlated.

Test statistic T:

Pearson (or Spearman) correlation coefficient, calculate t_{obs}

Randomization: Under H_0 it is possible to permute the values observed for Gene 2. There are n! possibilities.

p-value: $p = P(T \geq t_{obs} \mid H_0) \approx \dfrac{\#\{ T^* \geq t_{obs} \}}{n!}$

Data

Gene1	Gene2
g_1^1	g_1^2
⋮	⋮
g_n^1	g_n^2

↓

$g_{(1)}^1$	$g_{(1)}^2$
⋮	⋮
$g_{(n)}^1$	$g_{(n)}^2$

Random Permutation for group labels

Gene 1	Gene 2	Group	→	Group
1.4482	1.0709	1		2
0.4850	0.9324	1		1
1.1331	1.2379	1		4
		⋮		⋮
0.8015	0.6765	2		1
		⋮		⋮
1.3726	1.2373	3		4
		⋮		⋮
1.1030	1.735	4		2
0.5148	1.0015	4		3

The permutation test allows determining the statistical significance of the score for every gene.

See also: the **coin** package and the **lmPerm** package:
coin: Conditional Inference Procedures in a Permutation Test Framework
lmPerm: Permutation Tests for Linear Models