Multisample Inference

Outline

1. One-Way ANOVA--- Fixed Effects
   ① One-Way ANOVA
   ② Post-hoc Test After One-Way ANOVA
2. Two-Way ANOVA
3. The Kruskal-Wallis Test
4. One-Way ANOVA--- Random Eeffects
5. Meta Analysis
6. Mixed Model

One-Way ANOVA

Qualitative Factor

Fixed Effects

Example: Pulmonary disease

- A topic of public-health interest is whether *passive smoking* (exposure among nonsmokers to cigarette smoke in the atmosphere) has a measurable effect on pulmonary (肺部) health.

  - Endpoint: Forced Mid-expiratory flow (FEF) in L/s
  - 6 groups:
    - nonsmokers (NS),
    - passive smokers (PS),
    - noninhaling smokers (NI),
    - light smokers (LS),
    - moderate smokers (MS),
    - heavy smokers (HS)

FEF data for smoking and nonsmoking males

Group name	Mean FEF	SD FEF	n
NS	3.78	0.79	200
PS	3.30	0.77	200
NI	3.32	0.86	50
LS	3.23	0.78	200
MS	2.73	0.81	200
HS	2.59	0.82	200

One-Way ANOVA--- Fixed Effects

- Can be viewed in two different ways:
  - Extension of “two-sample” t-test to more than two groups.
  - Extension of simple linear regression to case where X is qualitative. => multiple regression

One-way ANOVA model

- k groups and n_i observations in the ith group
- y_{ij}: the jth observation in the ith group

$$y_{ij} = \mu + \alpha_i + e_{ij}, e_{ij} \sim N(o, \sigma^2).$$

- Constraint needed for "identifiability"
  - $\sum_{i=1}^{k} \alpha_i = 0$ (we use this approach here)
  - $\alpha_k = 0$

Interpretation

$$y_{ij} = \mu + \alpha_i + e_{ij}, e_{ij} \sim N(o, \sigma^2).$$

- μ: the underlying mean of all groups taken together.
- α_i: the difference between the mean of the ith group and the overall mean.
- e_{ij}: random error about the mean $\mu + \alpha_i$ for an individual observation from the ith group.

Multiple linear regression

- Recall: One-way ANOVA

$$y_{ij} = \mu + \alpha_i + e_{ij}, e_{ij} \sim N(o, \sigma^2).$$

- This is equivalent to:

$$y = \alpha + \sum_{j=1}^{k-1} \beta_j x_j + e$$

- where $x_j = \mathrm{I}(\text{subject is in group } (j+1))$ is dummy variable. And group 1 is referred to as the reference group.

- β_j: the difference between the average value of *y* for subjects in group (j+1) vs. the average value of *y* for subjects in the reference group.

. One-way ANOVA(fixed effect) Multiple linear regression

Overall comparison

- $H_0: \alpha_1 = \cdots = \alpha_k = 0$
- H_1: at least one of the $\alpha_j \neq 0$
- *F* test
  - Between SS
  - Within SS

- $H_0: \beta_1 = \cdots = \beta_k = 0$
- H_1: at least one of the $\beta_j \neq 0$
- *F* test
  - Regression SS
  - Residual SS

The F statistic and p-values are the same

Testing in One-way ANOVA

1. Simplest question: is there any group (main) effect?

$$H_0: \alpha_1 = \cdots = \alpha_k = 0?$$

2. Other questions: is the effect the same in groups 1 and 2?

$$H_0: \alpha_1 = \alpha_2?$$

Partitioning the variance

- Let $\bar{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$, and $\bar{\bar{y}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}}{n}$, $n = \sum_{i=1}^{k} n_i$.
- $y_{ij} - \bar{\bar{y}} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{\bar{y}})$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{\bar{y}})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_i - \bar{\bar{y}})^2$$

Total SS = Within SS + Between SS

- In shorthand:

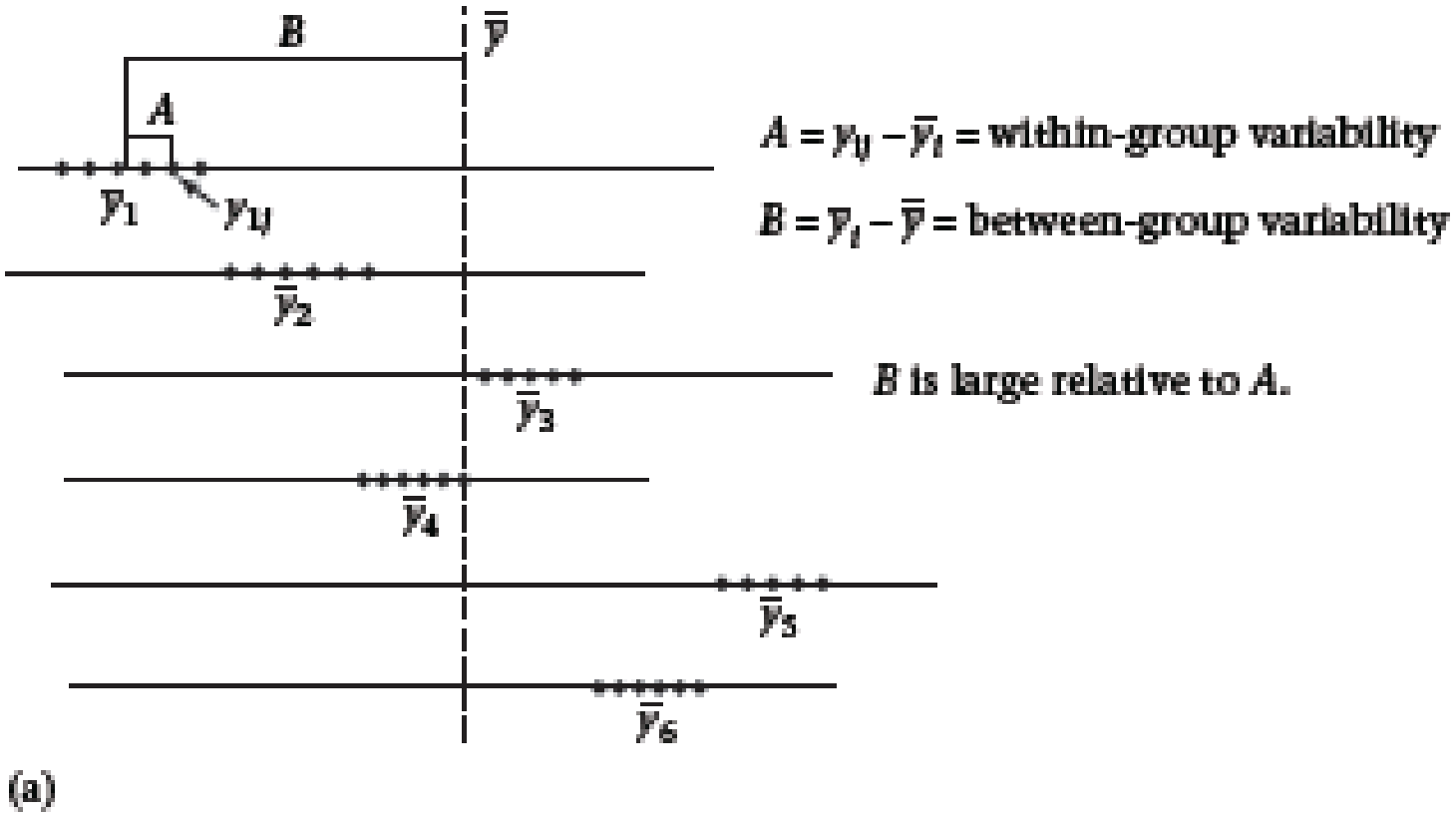
SSTO = SSW + SSB

$\sim\sigma^2\chi^2(n-1)$ $\sim\sigma^2\chi^2(n-\mathrm{k})$ $\sim\sigma^2\chi^2(\mathrm{k}-1,\lambda)$

- $MSB = \hat{\sigma}_B^2 = \frac{SSB}{k-1}$, $MSW = \hat{\sigma}_W^2 = \frac{SSW}{n-k}$

One-way ANOVA: the picture

- Comparison of between-group and within-group variability



Back to the hypothesis testing

- Is there any group (main) effect? (overall comparison)

$$H_0: \ \alpha_1 = \cdots = \alpha_r = 0$$
$$H_1: \text{at least one } \alpha_i \neq 0$$

- The null hypothesis specifies a global relationship between the means.

- Our test statistic is

$$F = \frac{MSB}{MSW} = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_W^2} = \frac{\text{variance between the groups}}{\text{variance within the groups}}$$

The F-statistic

- Under H_0

$$F \sim F_{k-1,n-k}$$

- Thus
  - Reject H_0 if $F > F_{\alpha,k-1,n-k}$
  - Fail to reject H_0 if $F \leq F_{\alpha,k-1,n-k}$

- $F_{\alpha,k-1,n-k}$ is the value on the $F_{k-1,n-k}$ distribution that, used as a cutoff, gives an area in the upper tail $= \alpha$

- And the exact p-value is given by the area to the right of *F* under an the $F_{k-1,n-k}$ distribution $= \Pr(F_{k-1,n-k} > \mathrm{F})$

A one-way ANOVA table

The results from the ANOVA are typically displayed in an ANOVA table, as in the following table:

Source of variation	SS	df	MS	F statistic	p-value
Between	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}-\bar{\bar{y}})^2 = \text{SSB}$	$k-1$	$\frac{\text{SSB}}{k-1} = \text{MSB}$	$\frac{\text{MSB}}{\text{MSW}}$	$\Pr(F_{k-1,n-k} > \text{F})$
Within	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2 = \text{SSW}$	$n-k$	$\frac{\text{SSW}}{k-1} = \text{MSW}$		
Total	SST = SSB + SSW	$n-1$			

Example: ANOVA for Pulmonary disease

Outcomes: Mid-expiratory flow (FEF) in L/s

Group name	Mean FEF	SD FEF	n
NS	3.78	0.79	200
PS	3.30	0.77	200
NI	3.32	0.86	50
LS	3.23	0.78	200
MS	2.73	0.81	200
HS	2.59	0.82	200

Pulmonary disease
ANOVA result

- We obtain the following result from Pulmonary disease example

	SS	df	MS	F statistic	P-value
Between	184.38	5	36.875	58.0	<0.001
Within	663.87	1044	0.636		
Total	848.25				

- Conclusions
  - Since $F = 58.0 > F_{0.001,5,1044} = 4.14$, we reject H_0
  - There is a significant difference in the mean FEF among the groups(P-value<0.001).

Testing in One-way ANOVA

1. Simplest question: is there any group (main) effect?

$$H_0: \alpha_1 = \cdots = \alpha_k = 0?$$

2. Other questions: is the effect the same in groups 1 and 2?

$$H_0: \alpha_1 = \alpha_2?$$

Which groups are different?

- We might proceed to make individual comparisons. E.g., is the effect the same in groups 1 and 2?

$$H_0: \alpha_1 = \alpha_2?$$

- Conduct two-sample t-tests to test for a difference in means for each pair of groups (assuming equal variance):

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Recall: $s = \hat{\sigma}_W^2 = MSW$ is the 'pooled' sample estimate of the common variance

The t-statistic

- Under H_0

$$t \sim t_{n-k}$$

- Thus
  - Reject H_0 if $\mathrm{t} > \mathrm{t}_{\alpha/2,\, n-k}$ or $\mathrm{t} < \mathrm{t}_{1-\alpha/2,\, n-k}$
  - Fail to reject H_0 if $\mathrm{t}_{1-\alpha/2,\, n-k} \leq \mathrm{t} \leq \mathrm{t}_{\alpha/2,n-k}$

- And the exact p-value is given by

$$p = \begin{cases} 2 \times \Pr(t_{n-k} < t)\,, & \text{if } t < 0 \\ 2 \times \Pr(t_{n-k} > t)\,, & \text{if } t \geq 0 \end{cases}$$

- This test is often referred to as the *least significant difference* (LSD) method

Multiple Comparisons

- Performing individual comparisons require multiple hypothesis tests

- If $\alpha = 0.05$ for each comparison, there is a 5% chance that each comparison will falsely be called significant
- Overall, the probability of Type I error is elevated above 5%

- For n independent tests, the probability of making a Type I error at least once is $1 - (1 - 0.05)^n$
  - Example: For n = 10 tests, the probability of at least one Type I error is 0.40 !

- Question: How can we address this multiple comparisons issue?

Multiple Testing

- Differentially expressed genes detection

	Case				Control			
Gene 1	X_{11}	X_{12}	⋯	X_{1n_1}	Y_{11}	Y_{12}	⋯	Y_{1n_1}
2	X_{11}	X_{11}	⋯	X_{11}	Y_{11}	Y_{11}	⋯	Y_{11}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- Different genes may have different gene expression levels
- For each gene, two-sample t-test …

  ⇨ Multiple testing problem

FWER

- Type I error, $\alpha = 0.05$ in the ordinary hypothesis testing

- Multiple hypothesis testing:

$$H_{i0}: \mu_{i1} = \mu_{i2}, i = 1,2, \dots, n$$

- Family wise type I error rate (FWER)

FWER= $P(\bigcup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n}(1 - \alpha_i)$

A_i: type I error occurs in the i-th hypothesis testing

Bonferroni Adjustment

- A possible correction for multiple comparisons
- Test each hypothesis at level $\alpha^* = \alpha / \binom{k}{2}$

FWER

- Basic idea: ensures overall Type I error rate does not exceed α = 0.05
  - For example, for our Pulmonary disease, $k = 6$, given $\alpha = .05$, thus $\alpha^* = .05/\binom{6}{2} = 0.0033$
- However, this adjustment may be too conservative

False-Discovery Rate

- Control of FWER may result great loss in detection power

$\Rightarrow$An alternative approach based on the *false-discovery rate* (FDR)

	Null hypothesis is true (H_0)	Alternative hypothesis is true (H_A)	Total
Test is declared significant	V	S	R
Test is declared non-significant	U	T	$m - R$
Total	m_0	$m - m_0$	m

- Basic idea: attempts to control the proportion of false positive results among reported statistically significant results.

FDR vs. FWER

	Null hypothesis is true (H_0)	Alternative hypothesis is true (H_A)	Total
Test is declared significant	V	S	R
Test is declared non-significant	U	T	$m - R$
Total	m_0	$m - m_0$	m

- $Q = \begin{cases} V/R, & R > 0 \\ 0, & R = 0 \end{cases}$
- FDR=E(Q)=E(V/R|R>0)P(R>0)
- FWER=P(V≥1)
- If m=m_0, FDR=E(1|R>0)P(R>0)=P(R>0)=P(V>0)=FWER;
- If m>m_0, $0 < \mathrm{V/R} \leq 1 \Rightarrow \mathrm{V/R} \leq I(V > 0) \Rightarrow$ FDR ≤FWER
- Controlling FDR may gain power comparing to controlling

FDR Testing Procedure (BH, 1995)

- Procedure:
  1. For a given α, find the largest k such that $p_k \leq \frac{k}{m}\alpha$,
  2. Reject the null hypothesis for all H_i for $i = 1,2,\ldots,k$.

  which ensures that $\mathrm{E(Q)} \leq \frac{m_0}{m}\alpha \leq \alpha$

- R function: p.adjust

BH Procedure in textbook

① Suppose we have conducted *k* separate tests with *p*-values = $p_1, \ldots, p_k$.

② For convenience we will renumber the tests so that $p_1 \leq p_2 \leq \ldots \leq p_k$.

③ Define $q_i = kp_i/i$, $i = 1, \ldots, k$, where i = rank of the *p*-values among the *k* tests.

④ Let FDR_i = false-discovery rate for the ith test be defined by $\min(q_i, \ldots, q_k)$.

⑤ Find the largest *i* such that $\mathrm{FDR}_i < \mathrm{FDR}_0$ = critical level for the FDR (usually .05).

⑥ Reject H_0 for the hypotheses 1, . . . , *i*, and accept H_0 for the remaining hypotheses.

FDR related concepts

- $Fdr = \frac{E(V)}{E(R)}, pFDR = E(\frac{V}{R}|R > 0)$
- Finding the truly alternative hypotheses is regarded as identifying a small component in a mixture, i.e.,

$$p \sim \pi_0 f_0 + (1 - \pi_0) f_1,$$

where $p|H_0 \sim f_0$ and $p|H_1 \sim f_1$, and $\pi_0 \cong 1$

- Empirical Bayes method

  (ref: Large-scale inference, Bradley Efron, 2010)

- $FDR(x) = \frac{\pi_0 F_0(x)}{F(x)}$=prob. of being null/prob. of being rejected
- $local\ fdr(x) = \frac{\pi_0 f_0(x)}{f(x)} = P(H_0|x)$

  prob. of being null given x/1

Demonstrating example: I

- A subsample of 520 cases of cardiovascular disease (CVD) and 1100 controls was obtained among men in a prospective cohort study.
- Baseline blood samples were analyzed for 50 candidate single nucleotide polymorphisms (SNPs).
- The association of each SNP with CVD was assessed using contingency-table methods.
- A chi-square test for trend was run for each SNP. This yielded 50 separate *p*-values. => next slide

Demonstrating example: II

Use of the FDR approach to analyzing the CVD data

	SNP	Naïve *p*-value	Bonferroni *p*-value	q_i	FDR_i
1	gene30	<.0001	.0035	.0035	.0035
2	gene20	.011	.54	.28	.16
3	gene48	.017	.86	.28	.16
4	gene50	.017	.87	.22	.16
5	gene4	.018	.92	.18	.16
6	gene40	.019	.94	.16	.16
7	gene7	.026	1.00	.18	.18
8	gene14	.034	1.00	.21	.21
9	gene26	.042	1.00	.23	.23
10	gene47	.048	1.00	.24	.24

Case study: LEAD

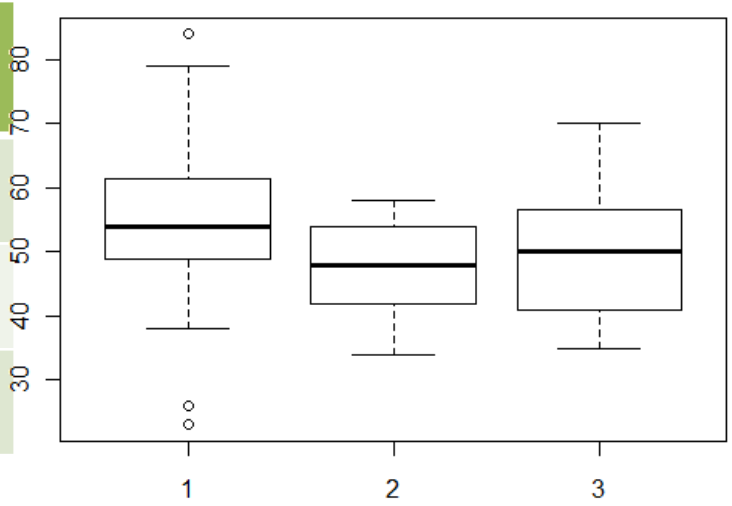
- we analyzed the difference in mean finger-wrist tapping score (MAXFWT) by three lead-exposure groups according to the variable LEAD_TYPE:
  - If LEAD_GRP = 1, then the child had normal blood-lead levels (<40 µg/100 mL) in both 1972 and 1973 (control group).
  - If LEAD_GRP = 2, then the child had elevated blood-lead levels (≥40 µg/100 mL) in 1973 (the currently exposed group).
  - If LEAD_GRP = 3, then the child had elevated blood-lead levels in 1972 and normal blood-lead levels in 1973 (the previously exposed group).

First look at the data

Descriptive statistics by group

Group (g)	N	N(no missing)	Mean	Std Dev
1	77	63	55.09524	10.93487
2	22	17	47.58824	7.08042
3	21	15	49.40000	10.19664

Box-plot by group



boxplot(maxfwt~Lead_type)

R codes: I
LEAD

```
> summary(fit <- aov(maxfwt~g))
            Df Sum Sq Mean Sq F value  Pr(>F)
g            2  966.8  483.40  4.5985 0.01249 *
Residuals 92 9671.1 105.12
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’
1
25 observations deleted due to missingness
> pairwise.t.test(maxfwt,Lead_type,p.adjust.method =
"none") # t test

Pairwise comparisons using t tests with pooled SD

data: maxfwt and Lead_type

   1      2
2 0.0087 -
3 0.0563 0.6191
```

R codes: II

LEAD

```
> pairwise.t.test(maxfwt,Lead_type,p.adjust.method = "bonferroni")
# Bonferroni correction

Pairwise comparisons using t tests with pooled SD

data: maxfwt and Lead_type

  1     2
2 0.026 -
3 0.169 1.000
P value adjustment method: bonferroni
> pairwise.t.test(maxfwt,Lead_type,p.adjust.method = "BH")
# FDR: False Discovery rate control

Pairwise comparisons using t tests with pooled SD

data: maxfwt and Lead_type

  1     2
2 0.026 -
3 0.084 0.619
P value adjustment method: BH
```

Two-Way ANOVA

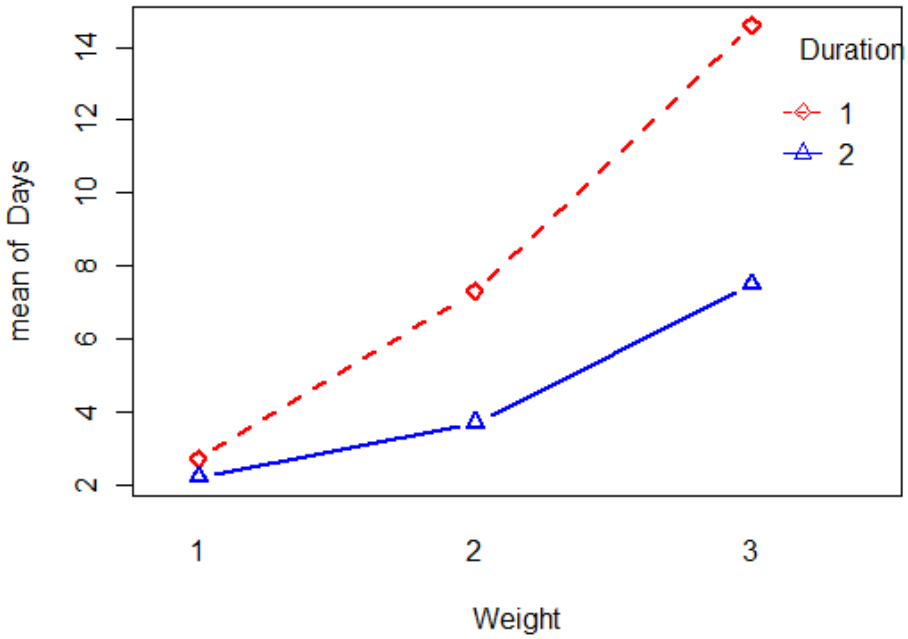
Two-Way ANOVA

- We have learned the fixed-effects one-way ANOVA, in which groups were defined by only one variable, e.g., cigarette smoking.

- In some instances, the groups being considered can be classified by two different variables and thus can be arranged in the form of an $R \times C$ contingency table.

- We would like to be able to look at the effects of each variable after controlling for the effects of the other variable.
- => the two-way ANOVA.

Example: Kidney



- Recovery time depends on weight gain between treatments and duration of treatment.
- Two levels of duration, three levels of weight gain.

Definition

- **Main effect**
  - effect of one treatment variable considered in isolation (ignoring other variables in the study)
- **Interaction effect**
  - the effect of one variable depends on the level of the other variable.
  - For example:
    - SV males have mean SBP levels that are 10 mm Hg lower than those of normal males, whereas SV females have mean SBP levels identical to those of normal females.

    => an interaction effect between sex and dietary group

Two-way ANOVA model

- General model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

- where $e_{ij} \sim N(o, \sigma^2)$.
- Constraint needed for "identifiability"
  - $\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{c} \beta_j = 0$
  - $\sum_{i=1}^{r} \gamma_{ij} = 0,$ for all j
  - $\sum_{j=1}^{c} \gamma_{ij} = 0,$ for all i
- Using dummy variables, this is still a multiple regression problem.

Interpretation

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

- μ: the underlying mean of all groups taken together.
- α_i: the effect of row group
- β_j: the effect of column group
- γ_{ij}: the interaction effect between row group and column group.
- e_{ijk}: random error about the mean $\mu + \alpha_i + \beta_j + \gamma_{ij}$ for an individual observation from the ith group.

Testing in Two-way ANOVA

1. Are there main effects for the grouping variables?

$$H_0: \alpha_1 = \cdots = \alpha_r = 0$$
$$H_0: \beta_1 = \cdots = \beta_c = 0$$

2. Are there interaction effects?

$$H_0: \gamma_{ij} = 0, 1 \leq i \leq r, 1 \leq j \leq c.$$

Two-way ANOVA table

Source of variation	SS($n_{ij} = n$)	df
Row (A)	$nc\sum_{i=1}^{r}(\bar{y}_{i..} - \bar{y}_{...})^2 = \text{SSA}$	r−1
Column(B)	$\text{nr}\sum_{j=1}^{c}(\bar{y}_{.j.} - \bar{y}_{...})^2 = \text{SSB}$	$c-1$
Interaction(AB)	$n\sum_{i=1}^{r}\sum_{j=1}(\bar{y}_{ij.} - \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...})^2 = \text{SSAB}$	$(r-1)(c-1)$
Error	$\sum_{i=1}^{r}\sum_{j=1}(y_{ijk} - \bar{y}_{ij.})^2 = \text{SSE}$	$(n-1)\text{rc}$
Total	SST = SSA + SSB + SSAB + SSE	$nc-1$

F-tests for two-way ANOVA

- $MS = SS/df$

- F-test

$$F_A = \frac{MSA}{SSE} \sim F(r-1, (n-1)rc)$$

$$F_B = \frac{MSB}{SSE} \sim F(c-1, (n-1)rc)$$

$$F_{AB} = \frac{MSAB}{SSE} \sim F((r-1)(c-1)$$

R codes:
Kidney

```
> kidney.lm <- lm(Days ~ Duration * Weight)
> summary(kidney.lm)

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.700      1.702   1.586 0.1186
Duration2           -0.500      2.407  -0.208 0.8362
Weight2              4.600      2.407   1.911 0.0613 .
Weight3             11.900      2.407   4.943 7.84e-06 ***
Duration2:Weight2   -3.100      3.405  -0.911 0.3666
Duration2:Weight3   -6.600      3.405  -1.939 0.0578 .

Residual standard error: 5.383 on 54 degrees of freedom
Multiple R-squared: 0.408, Adjusted R-squared: 0.3532
F-statistic: 7.444 on 5 and 54 DF, p-value: 2.262e-05
```

```
> print(anova(kidney.lm))
Analysis of Variance Table

Response: Days
                Df  Sum Sq Mean Sq  F value    Pr(>F)
Duration         1  209.07  209.07   7.2147  0.009587 **
Weight           2  760.43  380.22  13.1210  2.269e-05 ***
Duration:Weight  2  109.03   54.52   1.8813  0.162240
Residuals       54 1564.80   28.98
> # Another way to look at interactions
> anova(lm(Days ~ Duration + Weight), kidney.lm)
Analysis of Variance Table

Model 1: Days ~ Duration + Weight
Model 2: Days ~ Duration * Weight
 Res.Df     RSS  Df Sum of Sq       F    Pr(>F)
1     56  1673.8
2     54  1564.8   2    109.03  1.8813   0.1622
```

The Kruskal-Wallis Test

a nonparametric alternative to the one-way ANOVA

Example: Ophthalmology

- A study was conducted to compare the anti-inflammatory effects of four different drugs in albino rabbits after administration of arachidonic acid .

- For each animal in a group, one of the four drugs was administered to one eye and a saline solution was administered to the other eye.
- Ten minutes later arachidonic acid was administered to both eyes. Both eyes were evaluated every 15 minutes thereafter for lid closure.

- assessment—a lid-closure score: 0 = eye completely open, 3 = eye completely closed, and 1, 2 = intermediate states.

Outcomes:
Ophthalmology

The measure of effectiveness (x) is the change in lid-closure score (from baseline to follow-up) in the treated eye minus the change in lid-closure score in the saline eye.

Ocular anti-inflammatory effects of four drugs on lid closure after administration of arachidonic acid

Rabbit Number	Indomethacin		Aspirin		Piroxicam		BW755C	
	Score[a]	Rank	Score	Rank	Score	Rank	Score	Rank
1	+ 2	13.5	+ 1	9.0	+ 3	20.0	+ 1	9.0
2	+ 3	20.0	+ 3	20.0	+ 1	9.0	0	4.0
3	+ 3	20.0	+ 1	9.0	+ 2	13.5	0	4.0
4	+ 3	20.0	+ 2	13.5	+ 1	9.0	0	4.0
5	+ 3	20.0	+ 2	13.5	+ 3	20.0	0	4.0
6	0	4.0	+ 3	20.0	+ 3	20.0	− 1	1.0

[a](Lid-closure score at baseline − lid-closure score at 15 minutes)$_{\text{drug eye}}$ − (lid-closure score at baseline − lid-closure score at 15 minutes)$_{\text{saline eye}}$

The Kruskal-Wallis test

- In the last example, the score is ordinal variable => a nonparametric method

- generalize the Wilcoxon rank-sum test to enable us to compare more than two samples.

- Based on ranks
  - If the average ranks are close to each other, then H_0, that the treatments are equally effective, is accepted.
  - If the average ranks are far apart, then H_0 is rejected and we conclude at least some of the treatments are different.

The Kruskal-Wallis test statistic

- If there are no ties,

$$H = H^* = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(\bar{R}_i - \bar{\bar{R}}\right)^2$$

- If there are ties,

$$H = \frac{H^*}{1 - \frac{1}{N^3 - N} \sum_{j=1}^{g} (t_j^3 - t_j)}$$

- $N = \sum_i n_i$ and R_i: the rank sum for the ith group.
- t_j: the number of observations (i.e., the frequency) with the same value in the *j*th cluster of tied observations
- g is the number of tied groups.

Contd.

- Applicable only if $\min(n_i) \geq 5$. (Otherwise, either the sample should be combined with another sample or special small-sample tables should be used.)

- For a level α test,
  - If $H > \chi^2_{\alpha,k-1}$, then reject H_0
  - If $H \leq \chi^2_{\alpha,k-1}$, then accept H_0

- To assess statistical significance, the *p*-value is approximate by

$$p = \Pr\left(\chi^2_{k-1} > H\right)$$

Dunn Procedure

- Multiple comparison under the Kruskal-Wallis test
- To compare the *i*th and *j*th treatment groups

$$z = \frac{\bar{R}_i - \bar{R}_j}{\sqrt{\frac{N(N+1)}{12} \times \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}}$$

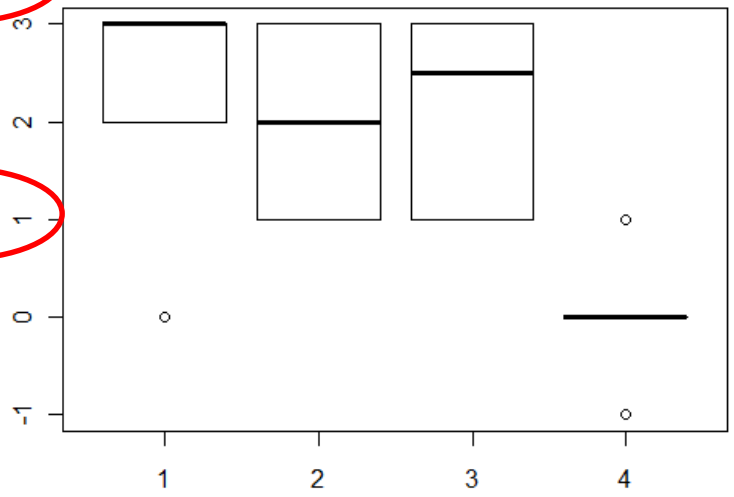
- For a two-sided level α test, denote $\alpha^* = \frac{\alpha}{k(k-1)}$
  - If $|z| \leq z_{\alpha^*}$, then accept H_0
  - If $|z| > z_{\alpha^*}$, then reject H_0

Back to:
Ophthalmology

```
> kruskal.test(ocular,g) # Kruskal-Wallis rank sum
test

Kruskal-Wallis rank sum test

data: ocular and g
Kruskal-Wallis chi-squared = 11.8041, df = 3, p-
value = 0.008085

> boxplot(ocular ~ g)
> library("pgirmess")
> kruskalmc(ocular,g)

Multiple comparison test after Kruskal-Wallis
p.value: 0.05
```

Comparisons

	obs.dif	critical.dif	difference
1-2	2.083333	10.77064	FALSE
1-3	1.000000	10.77064	FA
1-4	11.916667	10.77064	TR
2-3	1.083333	10.77064	FA
2-4	9.833333	10.77064	FA
3-4	10.916667	10.77064	TR



Box-plot by groups