

Comparison Among Several Sample

Professor: Hammou El Barmi
Columbia University

- The one-way analysis of variance (ANOVA) is a generalization of the two sample t-test ($k \geq 2$)
- Assume the populations of interest have the following (unknown) population means and variances

| | population 1 | population 2 | ... | population k |
|----------|--------------|--------------|-----|--------------|
| mean | μ_1 | μ_2 | ... | μ_k |
| variance | σ_1^2 | σ_2^2 | ... | σ_k^2 |

- Goal: test whether $\mu_1 = \mu_2 = \dots = \mu_k$
- We will compare these means without assuming any parametric relationships (regression does assume such a relationship).

Example:

- Suppose we have five medical treatments and ten subjects on each treatment.
- Goal: Compare the treatments in terms of their effectiveness
- If there were two treatments, what would we use?
- We will compare means among treatment groups.
- In the context of ANOVA, we say these five treatment make one factor with five levels and each level represents a treatment.

- To answer this question, random samples from each of the k -populations (each population corresponds to a level of the factor) leading to

| | sample 1 | sample 2 | ... | sample k |
|-----------------|-----------------------------------|-----------------------------------|-----|-----------------------------------|
| size | n_1 | n_2 | ... | n_k |
| sample | $Y_{11}, Y_{12}, \dots, Y_{1n_1}$ | $Y_{21}, Y_{22}, \dots, Y_{2n_2}$ | ... | $Y_{k1}, Y_{k2}, \dots, Y_{kn_k}$ |
| sample mean | $\bar{Y}_{1\bullet}$ | $\bar{Y}_{2\bullet}$ | ... | $\bar{Y}_{k\bullet}$ |
| sample variance | s_1^2 | s_2^2 | ... | s_k^2 |

- The sample means are $\bar{Y}_{1\bullet}, \bar{Y}_{2\bullet}, \dots, \bar{Y}_{k\bullet}$ and the average response over all the samples is

$$\bar{Y}_{\bullet\bullet} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} Y_{ij}}{\sum_{i=1}^k n_i} = \frac{\sum_{i=1}^k n_i \bar{Y}_{i\bullet}}{n}$$

where

$$n = \sum_{i=1}^k n_i.$$

To carry out the comparison, we assume that level of the factor (i.e. at each treatment), there is a probability distribution for the response and that

- Each probability distribution is normal
- These probability distribution have the same variances. That is,
$$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$
- The responses for each treatment is a random sample from the corresponding distribution
- These samples are independent

- An F test is used to test $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ against $H_a : \text{Not } H_0$ (that is at least two means are not equal)
- The assumptions needed for the test are analogous to the pooled two sample t-test
- The F-test is computed from the ANOVA table which breaks the spread in the combined data SST (Total Sum of Squares) into two components (or sums of squares): within sum of squares (SSE) and the between sums of square (SSR)

$$SST = SSE + SSR$$

- The Between SS (often called the model Sum of Squares) measures the spread between the sample means

$$SSR = \sum_{i=1}^k n_i (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2$$

- The within SS (often called Error Sum of Squares) is

$$SSE = \sum_{i=1}^k \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\bullet})^2$$

- Each SS has its own degrees of freedom (df)

$$df(SST) = n - 1 \quad df(SSR) = k - 1 \quad \text{and} \quad df(SSE) = n - k$$

- it is always the case that

$$df(SST) = df(SSR) + df(SSE)$$

- The mean square error for each source of variation is the corresponding SS divided by its df , that is,

$$MSR = \frac{SSR}{k - 1} \quad \text{and} \quad MSE = \frac{SSE}{n - k}$$

The sums of squares and their dfs are neatly arranged into called the ANOVA table

| Source | df | SS | MS | F |
|------------------------|-----|-----|-------------------|-----------|
| Model (Between Groups) | k-1 | SSR | $MSR = SSR/(k-1)$ | MSB/MSE |
| Error (Within Groups) | n-k | SSE | $MSE = SSE/(n-k)$ | |
| Between Groups (Model) | n-1 | SST | | |

- The decision on whether to reject $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$ is based on the

$$F = \frac{MSR}{MSE}$$

- We have $E(MSE) = \sigma^2$ and

$$E(MSR) = \sigma^2 + \frac{\sum_{i=1}^k n_i (\mu_i - \tilde{\mu}_{\bullet})^2}{k-1}$$

where

$$\tilde{\mu}_{\bullet} = \frac{\sum_{i=1}^k n_i \mu_i}{n}.$$

Therefore when H_0 is true

$$\frac{E(MSR)}{E(MSE)} = 1$$

- Large values of F indicate large variability among the sample means relative to the spread of the data within the samples. That is, large values of F suggest that H_0 is false
- We reject H_0 if $F > F(\alpha, k - 1, n - k)$ or if $p\text{-value} < \alpha$.
- For $k = 2$, the F test is equivalent to the pooled two-sample t -test

- During cooking, doughnuts absorb fat in various amounts.
- A scientist wished to learn whether the amount absorbed depends on the type of fat.
- For each of 4 fats, 6 batches of 24 doughnuts were prepared. The data are grams of fat absorbed by batch.
- Let μ_i = population mean of fat i absorbed per batch of 24 doughnuts.
- The Scientist wishes to test $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ against $H_a : \text{Not } H_0$.

| fat 1 | fat 2 | fat 3 | fat 4 |
|-------|-------|-------|-------|
| 264 | 278 | 275 | 255 |
| 272 | 291 | 286 | 266 |
| 268 | 297 | 278 | 249 |
| 277 | 282 | 271 | 264 |
| 290 | 285 | 263 | 270 |
| 276 | 277 | 276 | 268 |

Example

```
> fat<-c(rep("fat1",6),rep("fat2",6),rep("fat3",6),rep("fat4",6))
> amount<-c(264,272,268,277,290,276,278,291,297,282,285,277,275,286,278,271,
263,276,255,266,249,264,270,268)
> data<-data.frame(fat,amount)
> summary(data[,2][data[,1]=='fat1'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
264.0  269.0   274.0   274.5  276.8   290.0

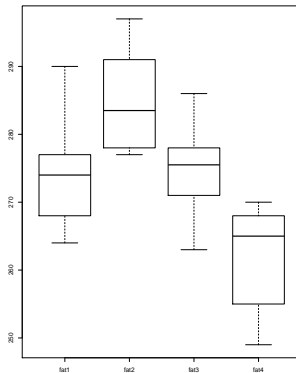
> summary(data[,2][data[,1]=='fat2'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
277.0  279.0   283.5   285.0  289.5   297.0

> summary(data[,2][data[,1]=='fat3'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
263.0  272.0   275.5   274.8  277.5   286.0

> summary(data[,2][data[,1]=='fat4'])
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
249.0  257.2   265.0   262.0  267.5   270.0
```

```
> boxplot(data[,2] data[,1])
```

Figure: Histogram and Box Plots



```
> summary(fit)
              Df Sum Sq Mean Sq F value Pr(>F)
data[, 1]      3   1596    531.8   7.948 0.0011 **
Residuals     20   1338     66.9
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$SSR = 1596, SSE = 1328, MSR = 531.8, MSE = 66.9, F = 7.95$$

If we take $\alpha = 0.05$, we have $F(1 - 0.05, 2, 20) = 3.098391$. since $7.95 > 3.098391$ we reject H_0

Also $p\text{-value} = 0.0011 < 0.05$ we reject H_0

- The ANOVA F-test checks whether all the population means are equal.
- Multiple comparisons are often used as a follow up to a significant ANOVA F-test to determine which population means are different.
- We will discuss Fisher's, Bonferroni's and Tukey's methods for comparing all pairs of means

Fisher's least significant difference method (LSD) is a two step process

- (1) Carry out the ANOVA F-test of $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$. If H_0 is not rejected stop and conclude that there insufficient evidence to claim differences among the population means. If H_0 is rejected, go to step 2
- (2) Compare each pair of means using a pooled two sample t-test at the alpha level using $s_{pooled} = \sqrt{MSE}$ from the ANOVA table and $df = df(SSE)$, that is test $H_0 : \mu_i = \mu_j$ against $H_a : \mu_i \neq \mu_j$ for all pair (i, j) using

$$t = \frac{\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}}{\sqrt{MSE} \sqrt{1/n_i + 1/n_j}}$$

and reject H_0 if $|t| > t_{n-k}(\alpha/2)$. of equivalently if

$$|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}| > t_{n-k}(\alpha/2) \sqrt{MSE} \sqrt{1/n_i + 1/n_j}$$

- (3) The minimum absolute difference between $\bar{Y}_{i\bullet}$ and $\bar{Y}_{j\bullet}$ need to reject H_0 is the LSD, the quantity on the right hand side of the equation above
- (4) If $n_1 = n_2 = \dots = n_k$

$$LSD = t_{n-k}(\alpha/2) \sqrt{MSE} \sqrt{2/n_1}$$

- ① In our example $s_{pooled} = \sqrt{MSE} = \sqrt{67} = 8.18$, $n - k = 20$ and if $\alpha = 0.05$, $t_{20}(0.025) = 2.086$. Since $n_1 = n_2 = n_3 = n_4 = 6$,

$$LSD = 2.086 \times 8.18 \times \sqrt{2/6} = 9.85.$$

- ② Any two sample means that differ by at least 9.85 in magnitude are significantly different at 5%.
- ③ One way to get Fisher comparisons in R uses `pairwise.t.test()` with `p.adj.sut.method`.
- ④ The resulting summary of multiple comparisons is in terms of p-values for all pairwise two sample t-tests using the pooled standard deviation from the ANOVA using `pool.sd=TRUE`.


```
> pairwise.t.test(data[,2],data[,1],pool.sd=TRUE,p.adjust.method="none" )
```

Pairwise comparisons using t tests with pooled SD

data: data[, 2] and data[, 1]

| | fat1 | fat2 | fat3 |
|------|-------|---------|-------|
| fat2 | 0.038 | - | - |
| fat3 | 0.944 | 0.044 | - |
| fat4 | 0.015 | 9.3e-05 | 0.013 |

P value adjustment method: none

Multiple Comparisons

There are $c = 4(4 - 1)/2 = 6$ comparisons of two fats

| Comparison | Absolute difference in means | Exceeds LSD | p-value |
|------------|------------------------------|-------------|----------------------|
| 1 versus 2 | 10.50 | Yes | 0.038 |
| 1 versus 3 | 0.33 | No | 0.944 |
| 1 versus 4 | 12.50 | Yes | 0.015 |
| 2 versus 3 | 10.17 | Yes | 0.044 |
| 2 versus 4 | 23.00 | Yes | 9.3×10^{-5} |
| 3 versus 4 | 12.83 | Yes | 0.013 |

There are three groups here $\{4\}$, $\{1, 3\}$ and $\{2\}$

- If the F-test indicates that a factor is significant, then any pair of means that differ by at least LSD are considered to be different.
- This is the least conservative of all the procedures, because no adjustment is made for multiple comparisons (so when doing lots of comparisons this makes Type I errors likely)
- The Bonferroni method controls the FER by reducing the individual comparison rate
- The FER is guaranteed to be no larger than a pre-specified amount say α by setting the individual error rate for each of the $k(k-1)/2$ comparisons of interest equal to

$$\alpha = \frac{\alpha}{k(k-1)/2}$$

- To implement the Bonferroni adjustment in R use `p.adjust.method="bonf"`

```
> pairwise.t.test(data[,2],data[,1],pool.sd=TRUE,p.adjust.method="bonf" )
```

Pairwise comparisons using t tests with pooled SD

data: data[, 2] and data[, 1]

| | fat1 | fat2 | fat3 |
|------|---------|---------|---------|
| fat2 | 0.22733 | - | - |
| fat3 | 1.00000 | 0.26241 | - |
| fat4 | 0.09286 | 0.00056 | 0.07960 |

P value adjustment method: bonferroni

- The LSD and Bonferroni methods comprise the ends of the spectrum of multiple comparisons methods
- Among multiple comparisons procedure, the LSD method is the most likely to find differences whether real or due to variation while Bonferroni is often the most conservative method
- The Bonferroni method is conservative but tends to work well when the number of comparisons is small, say 4 or less
- For $r > 4$, Bonferroni starts to get much more conservative than necessary

- Another multiple comparisons procedure is Tukey's method (a.k.a. Tukey's Honest Significance Test). The function `TukeyHSD()` creates a set of confidence intervals on the differences between means with the specified family-wise probability of coverage.
- The general form is `TukeyHSD(fit, conf.level = 0.95)`. Here `fit` is a fitted model object (e.g., an `aov.fit`) and `conf.level` is the confidence level.
- Tukey's method is designed for equal sample sizes but can be used for different sample sizes too.
- The method rejects the equality of a pair of means based on the studentized range distribution. To implement this method at α , reject $H_0 : \mu_i = \mu_j$ when

$$|\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}| > \frac{q(1 - \alpha, k, n - k)}{\sqrt{2}} \sqrt{MSE} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

where $q(1 - \alpha, k, n - k)$ is the α th level critical value of the studentized range distribution

```
> TukeyHSD(fit)
  Tukey multiple comparisons of means
    95\% family-wise confidence level
```

```
Fit: aov(formula = data[, 2] ~ data[, 1], data = data)
```

```
$'data[, 1]'
```

| | diff | lwr | upr | p adj |
|-----------|-------------|------------|------------|-----------|
| fat2-fat1 | 10.5000000 | -2.719028 | 23.7190277 | 0.1510591 |
| fat3-fat1 | 0.3333333 | -12.885694 | 13.5523611 | 0.9998693 |
| fat4-fat1 | -12.5000000 | -25.719028 | 0.7190277 | 0.0679493 |
| fat3-fat2 | -10.1666667 | -23.385694 | 3.0523611 | 0.1709831 |
| fat4-fat2 | -23.0000000 | -36.219028 | -9.7809723 | 0.0004978 |
| fat4-fat3 | -12.8333333 | -26.052361 | 0.3856944 | 0.0590077 |

- We discuss three different parametrizations for describing the variation in means.
- Which of these methods is used depends on what we want the resulting parameters to mean and on the nature of constraints we may wish to impose on the model
- Method 1: Factor effects method (center point method):

- let

$$\mu_{\bullet} = \frac{1}{k} \sum_{i=1}^k \mu_i \quad \text{and} \quad \alpha_i = \mu_i - \mu_{\bullet} \quad (\Rightarrow \sum_{i=1}^k \alpha_i = 0)$$

- ANOVA model : $\mu_i = \mu_{\bullet} + \alpha_i, i = 1, 2, \dots, k$
- Regression Model

$$Y_{ij} = \mu_{\bullet} + \alpha_i + \epsilon_{ij}$$

with

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_{k-1} \\ \mu_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & -1 & -1 & \dots & -1 \end{bmatrix} \begin{bmatrix} \mu_{\bullet} \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{k-1} \end{bmatrix} = X_1 \beta$$

- Interesting hypotheses: $H_0 : \mu_1 = \mu_2 = \dots = \mu_k \Leftrightarrow H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_{k-1} = 0$
or $H_0 : C\beta = \mathbf{0}$ where $C = [\mathbf{0}, I_{k-1}]$

In our example, to carry out the analysis we use the following

```
> a<-gl(4,6)    #this creates the level (4 levels repeated 6 times each)

> lm(data[,2]~ a, contrasts = list(a = "contr.sum"))    # this fit the model
Call:
lm(formula = data[, 2] ~ a, contrasts = list(a = "contr.sum"))
```

Coefficients:

| (Intercept) | a1 | a2 | a3 |
|-------------|--------|---------|--------|
| 274.0833 | 0.4167 | 10.9167 | 0.7500 |

```
> summary(lm(data[,2]~ a, contrasts = list(a = "contr.sum")))
```

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-------------|----------|------------|---------|-------------|
| (Intercept) | 274.0833 | 1.6698 | 164.143 | < 2e-16 *** |
| a1 | 0.4167 | 2.8922 | 0.144 | 0.88689 |
| a2 | 10.9167 | 2.8922 | 3.775 | 0.00119 ** |
| a3 | 0.7500 | 2.8922 | 0.259 | 0.79804 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.18 on 20 degrees of freedom

Multiple R-squared: 0.5438, Adjusted R-squared: 0.4754

F-statistic: 7.948 on 3 and 20 DF, p-value: 0.001104

The fitted model is

$$\hat{Y}_{ij} = 274.0833 + \hat{\alpha}_i$$

where $\hat{\alpha}_1 = 0.4167$, $\hat{\alpha}_2 = 10.9167$, $\hat{\alpha}_3 = 0.7500$ and $\hat{\alpha}_4 = -\hat{\alpha}_1 - \hat{\alpha}_2 - \hat{\alpha}_3 = -12.0834$

Reference cell method

- Define $\mu^* \equiv \mu_1$ (reference cell) and $\alpha_i^* = \mu_i - \mu^*$ ($\alpha_1^* = 0$ by definition).
- ANOVA model: $\mu_i = \mu^* + \alpha_i^*, i = 1, 2, \dots, k$.
- Regression model:

$$Y_{ij} = \mu^* + \alpha_i + \epsilon_{ij}$$

with

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu^* \\ \alpha_2^* \\ \alpha_3^* \\ \vdots \\ \alpha_k^* \end{bmatrix} = X_2 \beta$$

- Interesting hypotheses:

$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \Leftrightarrow H_0 : \alpha_2^* = \alpha_3^* = \dots = \alpha_k^* = 0$ or $H_0 : C\beta = \mathbf{0}$ where $C = [\mathbf{0}, I_{k-1}]$

```
> lm(data[,2]~factor(data[,1]))
```

Call:

```
lm(formula = data[, 2] ~ factor(data[, 1]))
```

Coefficients:

| | | |
|-----------------------|-----------------------|-----------------------|
| (Intercept) | factor(data[, 1])fat2 | factor(data[, 1])fat3 |
| 274.5000 | 10.5000 | 0.3333 |
| factor(data[, 1])fat4 | | |
| -12.5000 | | |

The reference mean here is the mean of fat1. The model is

$$\hat{Y} = \begin{cases} 274.5, & \text{if fat 1} \\ 274.5 + 10.5 = 285 & \text{if fat 2} \\ 274.5 + 0.3 = 274.83 & \text{if fat 3} \\ 274.5 - 12.5 = 222 & \text{if fat 4} \end{cases}$$

```
> summary(lm(data[,2]~factor(data[,1])))
```

Call:

```
lm(formula = data[, 2] ~ factor(data[, 1]))
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|----------|---------|--------|--------|---------|
| -13.0000 | -6.6250 | 0.6667 | 4.5000 | 15.5000 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-----------------------|----------|------------|---------|------------|
| (Intercept) | 274.5000 | 3.3396 | 82.196 | <2e-16 *** |
| factor(data[, 1])fat2 | 10.5000 | 4.7229 | 2.223 | 0.0379 * |
| factor(data[, 1])fat3 | 0.3333 | 4.7229 | 0.071 | 0.9444 |
| factor(data[, 1])fat4 | -12.5000 | 4.7229 | -2.647 | 0.0155 * |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.18 on 20 degrees of freedom

Multiple R-squared: 0.5438, Adjusted R-squared: 0.4754

F-statistic: 7.948 on 3 and 20 DF, p-value: 0.001104

Cell mean method (here the cell means are the parameters)

- ANOVA model: $\mu_i = \mu_i$.
- Regression model:

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

with

$$\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_k \end{bmatrix} = X_3 \beta$$

- Interesting hypotheses: $H_0 : \mu_1 = \mu_2 = \dots = \mu_k \Leftrightarrow H_0 := \alpha_3^* = \dots = \alpha_k^* = 0$ or $H_0 : C\beta = \mathbf{0}$ where $C = I_k$

```
> lm(data[,2]~factor(data[,1])-1)
```

Call:

```
lm(formula = data[, 2] ~ factor(data[, 1]) - 1)
```

Coefficients:

| | | |
|-----------------------|-----------------------|-----------------------|
| factor(data[, 1])fat1 | factor(data[, 1])fat2 | factor(data[, 1])fat3 |
| 274.5 | 285.0 | 274.8 |
| factor(data[, 1])fat4 | | |
| 262.0 | | |

Regression Techniques for ANOVA

```
> summary(lm(data[,2]~factor(data[,1])-1))
```

Call:

```
lm(formula = data[, 2] ~ factor(data[, 1]) - 1)
```

Residuals:

| Min | 1Q | Median | 3Q | Max |
|----------|---------|--------|--------|---------|
| -13.0000 | -6.6250 | 0.6667 | 4.5000 | 15.5000 |

Coefficients:

| | Estimate | Std. Error | t value | Pr(> t) |
|-----------------------|----------|------------|---------|------------|
| factor(data[, 1])fat1 | 274.50 | 3.34 | 82.20 | <2e-16 *** |
| factor(data[, 1])fat2 | 285.00 | 3.34 | 85.34 | <2e-16 *** |
| factor(data[, 1])fat3 | 274.83 | 3.34 | 82.30 | <2e-16 *** |
| factor(data[, 1])fat4 | 262.00 | 3.34 | 78.45 | <2e-16 *** |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.18 on 20 degrees of freedom

Multiple R-squared: 0.9993, Adjusted R-squared: 0.9991

F-statistic: 6742 on 4 and 20 DF, p-value: < 2.2e-16

- A contrast L is defined as a linear combination of the level means where the coefficient add up to zero. That is

$$L = \sum_{i=1}^k c_i \mu_i \quad \text{where} \quad \sum_{i=1}^k c_i = 0$$

- Examples:

- ① $L = \mu_2 - \mu_1$
- ② $L = \mu_3 - (\mu_1 + \mu_2)/2$
- ③ $L = (\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4)/2$

- We estimate $L = \sum_{i=1}^k c_i \mu_i$ by

$$\hat{L} = \sum_{i=1}^k c_i \bar{Y}_{i\bullet}$$

- We have

$$E(\hat{L}) = \sum_{i=1}^k c_i E(\bar{Y}_{i\bullet}) = \sum_{i=1}^k c_i \mu_i = L \quad (\hat{L} \text{ is an unbiased estimator of } L)$$

and

$$\text{Var}(\hat{L}) = \sum_{i=1}^k c_i^2 \text{Var}(\bar{Y}_{i\bullet}) = \sigma^2 \sum_{i=1}^k \frac{c_i^2}{n_i}$$

This implies that

$$SE(\hat{L}) = \sqrt{MSE} \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i}}$$

- A $100(1 - \alpha)\%$ confidence interval for L is

$$\hat{L} \pm t_{n-k}(\alpha/2)SE(\hat{L})$$

- To test $H_0 : L = 0$ against $H_a : L \neq 0$, the test statistic is

$$t = \frac{\hat{L} - 0}{SE(\hat{L})}$$

and we reject H_0 is

$$|t| > t_{n-k}(\alpha/2)$$

Same technique works for linear combinations. Later we will look at multiple contrasts.

- Suppose we have k population with medians $\eta_1, \eta_2, \dots, \eta_k$.
- Test

$H_0 : \eta_1 = \eta_2 = \dots = \eta_k$ against $H_a : \text{at least two of these medians are not equal}$

- We apply the Kruskal-Wallis test. And to do so we pool the responses from all groups and rank them; then we apply one way ANOVA to the ranks, not to the original observations.
- If $R_{i\bullet}$ = sum of the ranks corresponding to the data from i th sample, the Kruskal-Wallis test statistic is

$$KW = \frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_{i\bullet}^2}{n_i} - 3(n+1)$$

and we reject H_0 if $KW > \chi_{k-1}^2(\alpha)$ or if $p\text{-value} < \alpha$.

```
> kruskal.test(data[,2]~data[,1])
```

Kruskal-Wallis rank sum test

data: data[, 2] by data[, 1]

Kruskal-Wallis chi-squared = 13.249, df = 3, p-value = 0.004128