Lecture 2: Single-Factor Studies

STA 106: Analysis of Variance

Suggested reading: ALSM Chapter 16 & 17

Xiner Zhou
Department of Statistics, University of California, Davis

Single-Factor Studies

- Single-Factor ANOVA Model
- Analysis of Variance
- F Test for Equality of Factor Level Means
- Analysis of Factor Level Means
- Planning of Sample Size

Substantive Research Questions of Interest:

Whether the factor levels or treatments differ in terms of response?
If the factor levels differ in terms of response, in what way do they differ or how do they differ?

These research questions lead to statistical questions usually performed in two steps, correspondingly.

Statistical Questions of Interest:

F Test for Equality of Factor Level Means

Whether the factor level means $\mu_i's$ are all equal or not?

To test whether: $\mu_1 = \mu_2 = \ldots = \mu_r$

If factor level means $\mu_i's$ are equal

No relation between the factor and the response variable is present.

No further analysis is needed

If factor level means $\mu_i's$ differ

A relation between the factor and the response variable is present.

Thorough analysis of the nature of the factor level means, how do they differ?

Analysis of Factor Level Means

Quantities about Factor Level Means

Quantities about Factor level Means that might be of interest

- A single factor level mean
- A difference between two factor level means
- A contrast among factor level means
- A linear combination of factor level means

μ_1 :

Estimate the average number of days required for below average fitness patients?
Is the average number of days required for below average fitness patients > 40 days?

$\mu_1 - \mu_3$:

Estimate the difference in average number of days required between below average fitness and above fitness patients?
Is the difference > 30 days?

$\mu_1 - \frac{\mu_2 + u_3}{2}$:

Suppose we consider average and above average fitness patients as "ideal patients", while below average fitness patients as "not ideal".
Estimate the difference in average number of days required between the "not ideal" and "ideal" patients?
Is the difference > 10 days?

$0.3\mu_1 + 0.4\mu_2 + 0.3\mu_3$

Suppose we know the population composed of 30%, 40%, 30% of below average, average and above average patients.
Estimate the average number of days required for a randomly selected patient from the population?
Is it < 30 days?

Inference (Estimation & Testing) of a Single Quantity

A single factor level mean μ_i :

1. Point estimate: $\hat{\mu}_i = \overline{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$

CLT : $\overline{Y}_{i\cdot} \sim N(\mu_i, \underbrace{\frac{\sigma^2}{n_i}}_{\text{unknown}})$

est. variance of $\overline{Y}_{i\cdot}$: $s^2(\cancel{\overline{Y}_{i\cdot}}\ \hat{\mu}_i) = \frac{MSE}{n_i}$

$$\frac{\overline{Y}_{i\cdot} - \mu_i}{s(\overline{Y}_{i\cdot})} = \frac{(\overline{Y}_{i\cdot} - \mu_i)/\sigma/\sqrt{n_i}}{\sqrt{MSE/n_i}\,/\,\sigma/\sqrt{n_i}}$$

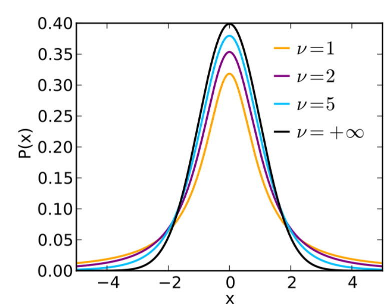
$$= \frac{(\overline{Y}_{i\cdot} - \mu_i)/\sigma/\sqrt{n_i}}{\sqrt{\frac{MSE}{\sigma^2}}}$$

- $\overline{Y}_{i\cdot} \perp\!\!\!\perp MSE$
- $(\overline{Y}_{i\cdot} - \mu_i)/\sigma/\sqrt{n_i} \sim N(0,1)$
- $(n_T - r)\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}$

general t-dist. with df=r is : $\frac{z}{\sqrt{\frac{\chi^2_r}{r}}}$

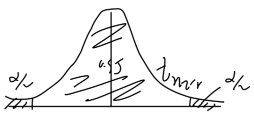
$\sim t_{df=r}$.

$$\Rightarrow \frac{\overline{Y}_{i\cdot} - \mu_i}{s(\overline{Y}_{i\cdot})} \sim t_{n_T - r}$$



Inference (Estimation & Testing) of a Single Quantity

2. $(1-\alpha)100\,\%$ confidence interval:



95% C.I.: $P\left(-t(0.975;\, n_T-r) \le \dfrac{\hat{\mu}_i - \mu_i}{s(\hat{\mu}_i)} \le t(0.975;\, n_T-r)\right) = 0.95$

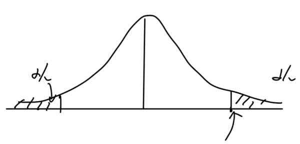
$P\left(\hat{\mu}_i - t(0.975;\, n_T-r)\, s(\hat{\mu}_i) \le \mu_i \le \hat{\mu}_i + t(0.975;\, n_T-r)\, s(\hat{\mu}_i)\right) = 0.95$

$$\hat{\mu}_i \pm t\left(1-\frac{\alpha}{2}; n_T - r\right) s\left(\hat{\mu}_i\right)$$

3. Hypothesis testing (t-test)

$$H_0 : \mu_i = c \;\; H_a : \mu_i \neq c$$

$t^* = \dfrac{\hat{\mu}_i - c}{s(\hat{\mu}_i)} \sim t_{n_T - r}$ if H_0 is true



critical value is $t(1-\frac{\alpha}{2};\, n_T-r)$

If $|t^*| \le t(1-\alpha/2; n_T - r)$, conclude H_0

If $|t^*| > t(1-\alpha/2; n_T - r)$, conclude H_a

Inference (Estimation & Testing) of a Single Quantity

Difference between two factor level means $D = \mu_i - \mu_j$: (a pairwise comparison)

1. Point estimate: $\hat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}$

CLT : $\bar{Y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n_i})$

$\bar{Y}_{i'\cdot} \sim N(\mu_{i'}, \frac{\sigma^2}{n_{i'}})$

$\bar{Y}_{i\cdot} \perp\!\!\!\perp \bar{Y}_{i'\cdot} \Rightarrow \hat{D} \sim N(\overbrace{\mu_i - \mu_{i'}}^{=D}, \sigma^2(\frac{1}{n_i} + \frac{1}{n_{i'}}))$

$s^2(\hat{D}) = MSE(\frac{1}{n_i} + \frac{1}{n_{i'}})$

$$\frac{\hat{D} - D}{s(\hat{D})} = \frac{(\hat{D} - D)/\sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}}{\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_{i'}})}/\sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}}$$

$$= \frac{(\hat{D} - D)/\sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}}{\sqrt{MSE/\sigma^2}}$$

- $\hat{D} \perp\!\!\!\perp MSE$
- $(\hat{D} - D)/\sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}} \sim N(0,1)$
- $(n_T - r)\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}$

$$\Rightarrow \frac{\hat{D} - D}{s(\hat{D})} \overset{d}{=} \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n_T - r}}{n_T - r}}} \sim t_{df = n_T - r}$$

Inference (Estimation & Testing) of a Single Quantity

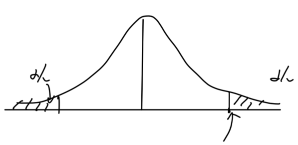
2. $(1-\alpha)100\,\%$ confidence interval:

$$\hat{D} \pm t\left(1-\frac{\alpha}{2}; n_T - r\right) s\left(\hat{D}\right)$$

3. Hypothesis testing (t-test)

$$H_0 : D = 0 \quad H_a : D \neq 0$$

$$t^* = \frac{\hat{D}}{s(\hat{D})} \sim t_{n_T - r} \text{ if } H_0 \text{ is true}$$



critical value is $t(1-\frac{\alpha}{2}; n_T - r)$

If $|t^*| \leq t(1-\alpha/2; n_T - r)$, conclude H_0

If $|t^*| > t(1-\alpha/2; n_T - r)$, conclude H_a

Inference (Estimation & Testing) of a Single Quantity

A contrast of several factor level means $L = \sum_{i=1}^{r} c_i \mu_i$ where $\sum_{i=1}^{r} c_i = 0$

1. Point estimate: $\hat{L} = \sum_{i=1}^{r} c_i \bar{Y}_{i\cdot}$

CLT : $\bar{Y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n_i})$

$\bar{Y}_{i\cdot}$'s indep

$\sigma^2(\hat{L}) = \sigma^2(c_1 \bar{Y}_{1\cdot}) + \cdots + \sigma^2(c_r \bar{Y}_{r\cdot})$

$= \sum_{i=1}^{r} c_i^2 \frac{\sigma^2}{n_i} = \sigma^2 \sum \frac{c_i^2}{n_i}$

$\Rightarrow \hat{L} \sim N(L, \sigma^2 \sum_{i=1}^{r} \frac{c_i^2}{n_i})$

$s^2(\hat{L}) = MSE(\sum \frac{c_i^2}{n_i})$

$\frac{\hat{L} - L}{s(\hat{L})} = \frac{(\hat{L} - L)/\sigma\sqrt{\sum \frac{c_i^2}{n_i}}}{\sqrt{MSE(\sum \frac{c_i^2}{n_i})}/\sigma\sqrt{\sum \frac{c_i^2}{n_i}}}$

$= \frac{(\hat{L} - L)/\sigma\sqrt{\sum \frac{c_i^2}{n_i}}}{\sqrt{MSE/\sigma^2}}$

- $\hat{L} \perp\!\!\!\perp MSE$
- $(\hat{L} - L)/\sigma\sqrt{\sum \frac{c_i^2}{n_i}} \sim N(0,1)$
- $(n_T - r)\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}$

$\Rightarrow \frac{\hat{L} - L}{s(\hat{L})} \overset{d}{=} \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n_T - r}}{n_T - r}}} \sim t_{df = n_T - r}$

Inference (Estimation & Testing) of a Single Quantity

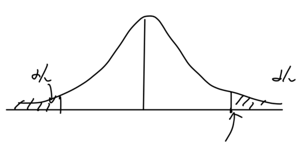
2. $(1-\alpha)100\ \%$ confidence interval:

$$\hat{L} \pm t\left(1-\frac{\alpha}{2}; n_T - r\right) s\left(\hat{L}\right)$$

3. Hypothesis testing (t-test)

$$H_0 : L = 0 \ H_a : L \neq 0$$

$$t^* = \frac{\hat{L}}{s(\hat{L})} \sim t_{n_T - r} \text{if } H_0 \text{ is true}$$



critical value is $t(1-\frac{\alpha}{2}; n_T - r)$

If $|t^*| \leq t(1-\alpha/2; n_T - r)$, conclude H_0

If $|t^*| > t(1-\alpha/2; n_T - r)$, conclude H_a

Inference (Estimation & Testing) of a Single Quantity

A linear combination of several level means $L = \sum_{i=1}^{r} c_i \mu_i$

- Example of linear combination?

  Overall mean $\bar{\mu}_{\cdot} = \sum_{i=1}^{r} \frac{n_i}{n_T} \mu_i$

- All previous quantities are special cases of "linear combination"

- Same inference as "contrast"

Example

The researcher is interested in estimating the mean number of days required for patients with average physical fitness, report the 99 percent confidence interval and give an interpretation.

$$\overline{Y_{2.}} \pm t\left(0.995; n_T - r\right) s\left(\overline{Y_{2.}}\right) = \overline{Y_{2.}} \pm t\left(0.995; n_T - r\right) \sqrt{\frac{MSE}{n_2}}$$

$$\bar{Y}_2. = 32, t\left(0.995; n_T - r\right) = 2.83, MSE = 19.8, n_2 = 10$$

$\longrightarrow$ [28.01,35.99]

Meaning of a confidence interval:

if repeated the same experiment many times, and each time we construct a confidence interval as above, then we would expect that 99% of times, the confidence intervals contracted this way will include the true average number of days required for patients with average fitness.

Interpret in the context:

the mean number of days required for patients with average fitness is estimated to be somewhere between 28.01 and 35.99 days, with 99% confidence.

Simultaneous Inference Procedures for Multiple Quantities about Factor Level Means

we aim to conclude something like:

“ The below average fitness group has the longer time required for rehabilitation than the average fitness group,
The average fitness group has the longer time required for rehabilitation than the above average fitness group”

That is equivalent to test the following comparisons:

$$\mu_1 > \mu_2$$
$$\mu_2 > \mu_3$$
$$\mu_1 > \mu_3$$

It is tempting to just conduct 3 separate tests as above section, then piece them together:

$$\left(\begin{array}{l} H_0^1 : \mu_1 = \mu_2 \text{ vs } H_a^1 \mu_1 > \mu_2 \\ H_0^2 : \mu_2 = \mu_3 \text{ vs } H_a^2 \mu_2 > \mu_3 \\ H_0^3 : \mu_1 = \mu_3 \text{ vs } H_a^3 \mu_1 > \mu_3 \end{array} \right.$$

Q: What might go wrong?

Simultaneous Inference Procedures for Multiple Quantities about Factor Level Means

What might go wrong?

What is considered “mistake” is different
when we make a composite statement that includes results of multiple tests or multiple C.I.s:

In a composite statement, if one sub-hypothesis test has a wrong conclusion, then it makes the whole statement wrong. That is, we have much more ways to make mistakes

The confidence level of C.I.s and significance level of tests apply only to each quantity considered individually, it will fail to compensate for multiple comparisons when it’s desired to have a statement which include multiple tests or multiple C.I.s.

E.g. Suppose we consider the efficacy of a drug in terms of reduction of any one of the disease symptoms. If we just consider more symptoms, it’s almost guaranteed that the drug will appear to be an improvement in terms of at least one symptom.

Data snooping

Simultaneous Inference Procedures for Multiple Quantities about Factor Level Means

“Family-wise” confidence coefficient $1 - \alpha$:

The proportion of correct families, when repeated sets of samples are selected and simultaneous tests or C.I.s are calculated each time.

> A family including simultaneous inference for multiple quantities is considered correct, if everyone single quantity inference is correct.

Thus, a family-wise confidence coefficient indicates that all conclusions in this family will be correct in $(1 - \alpha)100$ percent of repetitions.

Multiple Comparison Procedure: Bonferroni

Suppose we're interested in making inferenece about multiple quantities,
that are linear combinations of factor level means,
i.e., a family containing g linear combinations of factor level means

$$\mathscr{L} = \{L_1 = \sum_{i=1}^{r} c_{1i}\mu_i, \ldots, L_g = \sum_{i=1}^{r} c_{gi}\mu_i\}$$

family-wise error rate $= P($Making Type I error for any of the hypothesis in the family$)$

$= P(H_0^i$ falsely rejected for some $i)$

$= p(\cup_{i=1}^{g} \{H_0^i$ falsely rejected $\})$

$\leq \sum_{i=1}^{g} P(H_0^i$ falsely rejected $)$

If we control individual test at significance level α_0
$= g\alpha_0$

Bonferroni's idea:
One very easy and conservative way to control family-wise error rate at α is to control individual test's significance level at $\alpha_0 = \dfrac{\alpha}{g}$

Multiple Comparison Procedure: Bonferroni

$(1-\alpha)100\,\%$ confidence interval for individual quantity in this family:

$$\hat{L}_i \pm Bs\left(\hat{L}_i\right) \text{ for } i = 1 \ldots g$$

$$B = t\left(1 - \frac{\alpha}{2g}; n_T - r\right)$$

Guarantee:
family-wise confidence coefficient is at least $(1-\alpha)100\,\%$

Meaning:
in at least $(1-\alpha)100\,\%$ of repetition of experiments, all the intervals in the family cover the true corresponding $L_i's$ $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Hypothesis testing (t-test) for individual quantity in this family:

$$H_0^i : L_i = 0 \;\; H_a^i : L_i \neq 0$$

$$t^* = \frac{\hat{L}_i}{s(\hat{L}_i)} \sim t_{n_T - r} \text{if } H_0 \text{ is true}$$

If $|t^*| \leq B$, conclude H_0

If $|t^*| > B$, conclude H_a

Guarantee:
family-wise Type I error is at most α

Meaning:
in at most $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Multiple Comparison Procedure: Sheffe

Suppose we're interested in making inference about <u>all possible contracts of factor level means</u>

i.e., a family containing all possible contracts of factor level means

$$\mathscr{L} = \{L = \sum_{i=1}^{r} c_i\mu_i \text{ where } \sum_{i=1}^{r} c_i = 0\}$$

Infinitely many claims or quantities

$\forall L \in \mathscr{L}$:

$$\hat{L} - L = \sum_{i=1}^{r} c_i(\bar{Y}_{i\cdot} - \mu_i) = \sum c_i[(\bar{Y}_{i\cdot} - \mu_i) + (\bar{Y}_{\cdot\cdot} - \mu)]$$

$$= \sum_{i=1}^{r} \frac{c_i}{\sqrt{n_i}} \sqrt{n_i}[(\bar{Y}_{i\cdot} - \mu_i) + (\bar{Y}_{\cdot\cdot} - \mu)]$$

$$\le \sqrt{\sum \frac{c_i^2}{n_i}} \sqrt{\underbrace{\sum n_i[(\bar{Y}_{i\cdot} - \mu_i) + (\bar{Y}_{\cdot\cdot} - \mu)]^2}_{\stackrel{H_0}{=} \sum n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}}$$

$$= \sqrt{SSTR}.$$

$$s(\hat{L}) = s(\sum c_i \bar{Y}_{i\cdot}) = \sqrt{\sum c_i^2 s^2(\bar{Y}_{i\cdot})}$$

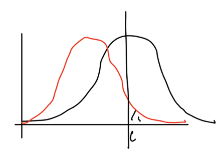
$$= \sqrt{\sum c_i^2 \frac{MSE}{n_i}}$$

$$= \sqrt{MSE}\sqrt{\sum \frac{c_i^2}{n_i}}$$

$$\Rightarrow \frac{\hat{L} - L}{s(\hat{L})} \le \sqrt{\frac{SSTR = (r-1)MSTR}{MSE}} = \sqrt{r-1}\sqrt{\frac{MSTR}{MSE}}$$

Under H_0, $\frac{r-1}{\sigma^2} MSTR \sim \chi^2_{r-1}$

$\frac{n_T - r}{\sigma^2} MSE \sim \chi^2_{n_T - r}$

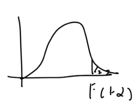


$$\stackrel{d}{=} \sqrt{\frac{\sigma^2 \chi^2_{r-1} / (r-1)}{\sigma^2 \chi^2_{n_T-r} / (n_T - r)}} \sqrt{r-1}$$

$$\stackrel{d}{=} \sqrt{(r-1) F(r-1, n_T-1)}$$

$$\Rightarrow P\left(\left|\frac{\hat{L} - L}{s(\hat{L})}\right| \ge c\right) \text{ for all } L$$

$$\le P\left(\underbrace{\sqrt{r-1}\sqrt{\frac{MSTR}{MSE}} \ge c}_{\frac{MSTR}{MSE} \ge \frac{c^2}{r-1}}\right)$$



$$\frac{c^2}{r-1} = F(1-\alpha)$$

$$\Rightarrow c = \sqrt{(r-1)F(1-\alpha)}$$

$\Rightarrow$ critical value $S = \sqrt{(r-1) F(1-\alpha; r-1, n_T - r)}$

Multiple Comparison Procedure: Sheffe

$(1-\alpha)100\,\%$ confidence interval for individual quantity in this family:

$$\hat{L}_i \pm Ss\left(\hat{L}_i\right)$$

$$S = \sqrt{(r-1)F(1-\alpha; r-1, n_T - r)}$$

Guarantee:
family-wise confidence coefficient is at least $(1-\alpha)100\,\%$

Meaning:
in at least $(1-\alpha)100\,\%$ of repetition of experiments, all the intervals in the family cover the true corresponding $L_i's$ $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Hypothesis testing (t-test) for individual quantity in this family:

$$H_0^i : L_i = 0 \;\; H_a^i : L_i \neq 0$$

$$t^* = \frac{\hat{L}_i}{s(\hat{L}_i)}$$

If $|t^*| \leq S$, conclude H_0

If $|t^*| > S$, conclude H_a

Guarantee:
family-wise Type I error is at most α

Meaning:
in at most $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Multiple Comparison Procedure: Tukey

Suppose we're interested in making inference about all pairwise comparisons of factor level means
i.e., a family containing <u>all pairwise comparisons of factor level means</u>

$$\mathscr{L} = \{D_{ii'} = \mu_i - \mu_{i'} \text{ for } i \neq i'\}$$

$\frac{r(r-1)}{2}$ Pairwise comparisons

$\bar{Y}_{i\cdot} - \mu_i \overset{iid}{\sim} N(0, \frac{\sigma^2}{n})$

MSE is unbiased estimator of σ^2. $\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}$

$$\text{Studentized range} = \frac{\text{range of data}}{\text{estimated s.d.}} = \frac{W}{S}$$

$$= \frac{\max\{\bar{Y}_{i\cdot} - \mu_i\} - \min\{\bar{Y}_{i\cdot} - \mu_i\}}{\sqrt{\frac{MSE}{n}}}$$

$$\sim \underbrace{q(r, n_T - r)}_{\text{Studentized range distribution}}$$

$$|\hat{D}_{ii'} - D_{ii'}| = |(\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}) - (\mu_i - \mu_{i'})|$$

$$= |(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{i'\cdot} - \mu_{i'})|$$

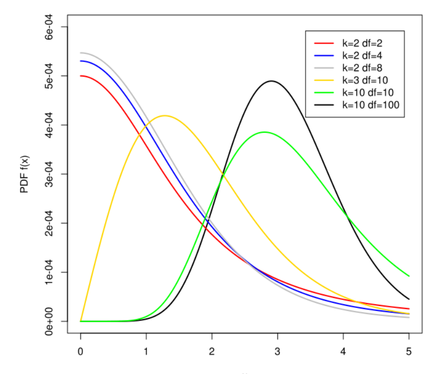
$$\leq \max\{\quad\} - \min\{\quad\}$$

$$s(\hat{D}_{ii'}) = s(\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot})$$

$$= \sqrt{s^2(\bar{Y}_{i\cdot}) + s^2(\bar{Y}_{i'\cdot})}$$

$$= \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}$$

$$\Rightarrow \frac{|\hat{D}_{ii'} - D_{ii'}|}{s(\hat{D}_{ii'})} \leq \frac{1}{\sqrt{2}} \text{ Studentized range for all } i \neq i'$$



Multiple Comparison Procedure: Tukey

$(1-\alpha)100\,\%$ confidence interval for individual quantity in this family:

$$\hat{D}_{ii'} \pm Ts\left(\hat{D}_{ii'}\right)$$

$$T = \frac{1}{\sqrt{2}} q(1-\alpha; r, n_T - r)$$

Guarantee:
family-wise confidence coefficient is at least $(1-\alpha)100\,\%$

Meaning:
in at least $(1-\alpha)100\,\%$ of repetition of experiments, all the intervals in the family cover the true corresponding $L_i's$ $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Hypothesis testing (t-test) for individual quantity in this family:

$$H_0^i : D_{ii'} = 0 \ H_a^i : D_{ii'} = \neq 0$$

$$q^* = \frac{\hat{D}_{ii'}}{s(\hat{D}_{ii'})}$$

If $|q^*| \leq T$, conclude H_0

If $|q^*| > T$, conclude H_a

Guarantee:
family-wise Type I error is at most α

Meaning:
in at most $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Compare: Bonferroni vs Scheffe vs Tukey

Compare 3 Simultaneous Inference Procedure

- **Bonferroni**
  - most general: quantities of interest can be mixture of any type: factor level mean, difference between factor level mean, contrast, or linear combination
  - can't handle data snooping, since need to specify the number of "data snooping" beforehand
- **Sheffe**
  - all possible contrasts
  - allow for data snooping as long as they are all contrasts
  - since applications never involve all contrasts, the actual family confidence coefficient is larger than stated, and the actual significance is smaller than stated
  - if only contrasts are of interest, Sheffe should be preferred, especially when the number of contrasts are large. However, when the number of contrasts is not that large, Bonferron might be better at times. (narrower C.I.s and smaller Type I error)
- **Tukey**
  - all possible pairwise differences
  - allow for data snooping as long as they are all differences
  - if only pairwise differences are of interest, Tukey should be preferred, especially when the number of differences are large. However, when the number of differences is not that large, Bonferroni might be better at times. (narrower C.I.s and smaller Type I error)

Compare: Bonferroni vs Scheffe vs Tukey

All three procedures are of the form "estimator $\pm$ multiplier $\times$SE."

The only difference among the three procedures is the multiplier.

In any given problem, one may compute the Bonferroni multiple as well as the Scheffé multiple and, when appropriate, the Tukey multiple, and select the one that is smallest.

Smaller multiple means: more efficient

- narrower C.I.s —> more precise estimate
- Smaller actual Type I error, larger power to detect true difference in hypothesis testing

Example

Suppose before seeing the data (a priori), the researcher wants to estimate the confidence intervals for

$$D_1 = \mu_2 - \mu_3, D_2 = \mu_1 - \mu_2$$

With a 95 percent family confidence coefficient.

Since the quantities of interest are pairwise comparisons (and contrasts of course), all three methods apply.

$$\hat{D}_1 = \overline{Y_{2.}} - \overline{Y_{3.}} = 8 \qquad \hat{D}_2 = \bar{Y}_{1.} - \bar{Y}_{2.} = 6$$

$$s\left(\hat{D}_1\right) = \sqrt{\text{MSE}\left(\frac{1}{n_2} + \frac{1}{n_3}\right)} = 2.298378 \qquad s\left(\hat{D}_2\right) = \sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 2.111195$$

Bonferroni:

$B = t(1 - \frac{\alpha}{2\times 2}; n_T - r) = 2.41384$

$$D_1 : \hat{D}_1 \pm Bs(\hat{D}_1) = 8 \pm 2.41384 \times 2.298378 = [2.452083, 13.54792]$$

$$D_2 : \hat{D}_2 \pm Bs(\hat{D}_2) = 6 \pm 2.41384 \times 2.111195 = [0.9039131, 11.09609]$$

Sheffe: $\sqrt{(r-1)F(1-\alpha; r-1, n_T - r)} = 2.633173$

$$D_1 : \hat{D}_1 \pm Ss(\hat{D}_1) = 8 \pm 2.63 \times 2.298378 = [1.955266, 4.04473]$$

$$D_2 : \hat{D}_2 \pm Ss(\hat{D}_2) = 6 \pm 2.63 \times 2.111195 = [0.4475572, 11.55244]$$

Tukey: $T = \frac{1}{\sqrt{2}} q(1-\alpha; r, n_T - r) = 2.52$

$$D_1 : \hat{D}_1 \pm Ts(\hat{D}_1) = 8 \pm 2.52 \times 2.298378 = [2.2, 13.79]$$

$$D_2 : \hat{D}_2 \pm Ts(\hat{D}_2) = 6 \pm 2.52 \times 2.111195 = [0.6797886, 11.32021]$$

Since the Bonferroni Multiple is the smallest, Bonferroni is the most efficient, it leads to tightest confidence intervals at the same level of family confidence coefficient.

Example

If the researcher also wants to estimate $D_3 = \mu_1 - \mu_3$, still with a 95 percent family confidence coefficient.
Would the multiples change?
What procedure(s) is appropriate for “data snooping” if the researcher wants to look at many such comparisons not a priori determined?

The Bonferroni multiple will change, but Sheffe and Tukey won't.

Tukey and Sheffe are applicable for data snooping as they are developed to include all pairwise comparisons and all contrasts, correspondingly.

But Bonferroni is not open for data snooping.

Example

The researcher is interested in comparing all pairs of factor level means, test for all pairs of factor level means whether or not they differ.

Use the procedure that is most efficient with $\alpha = 0.05$.

Set up groups by whether or not their factor levels means differ.

There are 3 pairwise comparisons.

$$B = t\left(1 - \frac{\alpha}{2 \times 3}; n_T - r\right) = 2.60135$$

$$S = \sqrt{(r-1)F\left(1-\alpha; r-1, n_T - r\right)} = 2.633173$$

$$T = \frac{1}{\sqrt{2}} q\left(1-\alpha; r, n_T - r\right) = 2.52$$

Therefore Tukey's procedure is most efficient.

$$\left(\begin{array}{l} H_0^1 : \mu_1 = \mu_2 \text{ vs } H_a^1 : \mu_1 \neq \mu_2 \\ H_0^2 : \mu_2 = \mu_3 \text{ vs } H_a^2 : \mu_2 \neq \mu_3 \\ H_0^3 : \mu_1 = \mu_3 \text{ vs } H_a^3 : \mu_1 \neq \mu_3 \end{array}\right.$$

Compare with critical value T, we reject all null hypotheses.

We conclude that, the average number of days required for different physical fitness group are all different from each other, with familywise confidence coefficient of 0.95.

We can set up groups within which factor level means do not differ:

Group 1: Below Average
Group 2: Average
Group 3: Above Average

Example

If before seeing the data, the researcher wants to estimate the difference in difference of mean days required between adjacent factor levels with a 99 percent confidence interval.

That is, to estimate the contrast $L = (\mu_1 - \mu_2) - (\mu_2 - \mu_3)$

Interpret the interval estimate.

This is a contrast with $c_1 = 1, c_2 = -2, c_3 = 1$

$$\hat{L} = (\bar{Y}_1 - \bar{Y}_2.) - (\bar{Y}_2. - \bar{Y}_3.) = -2$$

$$s(\hat{L}) = \sqrt{\text{MSE}\left(\sum \frac{c_i^2}{n_i}\right)}$$

A 99 percent confidence interval is

$$\hat{L} \pm t\left(1 - \frac{\alpha}{2}; n_T - r\right) s(\hat{L}) = [-12.48, 8.48]$$

Example

Estimate the following comparisons using all appropriate procedures with a 95 percent family confidence coefficient:

$$D_1 = \mu_1 - \mu_2, D_2 = \mu_1 - \mu_3, D_3 = \mu_2 - \mu_3, L_1 = D_1 - D_3,$$

Which procedure is more efficient?
Interpret results and describe findings using the one that is most efficient.

Both Bonferroni and Sheffe are appropriate.

$$B = t\left(1 - \frac{\alpha}{2 \times 4}; n_T - r\right) = 2.731632$$

$$S = \sqrt{(r-1)F\left(1 - \alpha; r-1, n_T - r\right)} = 2.633173$$

The Sheffe multiple is smaller, therefore Sheffe is most efficient.

$$D_1 : \hat{D}_1 \pm Ss\left(\hat{D}_1\right) = 6 \pm 2.63 \times 2.1112 = [0.45, 11.55]$$

$$D_2 : \hat{D}_2 \pm S_s\left(\hat{D}_2\right) = 14 \pm 2.63 \times 2.4037 = [7.68, 20.32]$$

$$D_3 : \hat{D}_3 \pm S_s\left(\hat{D}_3\right) = 8 \pm 2.63 \times 2.2984 = [1.96, 14.04]$$

$$L_1 : \hat{L}_1 \pm Ss\left(\hat{L}_1\right) = -2 \pm 2.63 \times 3.7 = [-11.73, 7.73]$$

Since the intervals for D_1, D_2, D_3 do not include zero, while the one for L does, we can conclude the statement that:

> the average days required for different physical fitness are all different with each other, and the difference between "below average" and "average" is the same as the difference between "average" and "above average".