- Inference for contrast of factor level means.

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

Ex: $L = \mu_1 - \mu_2$

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

$$L = \mu_1 - \frac{\mu_2 + \mu_3 + \mu_4}{3}$$

Estimator: $\hat{L} = \sum_{i=1}^{r} c_i \bar{y}_{i\cdot}$, $SE(\hat{L}) = \sqrt{\left(\sum_{i=1}^{r} \frac{c_i^2}{n_i}\right) MSE}$

- Multiple comparisons and error rates

Ex: 100 tests with sig-level .05, all H_0 true

Then X = # of times H_0 rejected $\sim$ Binomial (100, .05)

$$EX = 100 \times .05 = 5$$

- Running multiple tests on the same data set at the same stage of an analysis increases the chance of obtaining at least one invalid ~~test~~ result.

Significance level
- due to real difference
- due to chance

- performing more than one stat inference procedure on the same data set without adjusting the type I error rate accordingly is a common error in practice.

"Data snooping".

test what "seems" to be significant.

"hunt around through the data for a big contrast and then pretend that you've only done one comparison".

- Now assume you perform 2 tests using the same data. both H_0 are true, there is no ~~general~~ reason ~~that~~ the sample giving a type I error for one test will also give a type I error for the other test. So we need to consider the joint Type I error.

Def experimentwise error rate / familywise error rate.

Keep the familywise error rate at $\alpha = .05$

$P\{$ in the experiment, $\exists$ at least one incorrect inference $\}$

Given a family of null hypotheses $H_{01}, H_{02}, \cdots H_{0K}$
a familywise type I error occurs if all $H_{01}, \cdots H_{0K}$
are true, but at least one of them is rejected.

- Tukey's test

focus on the pairs of means

Tukey's procedure (balanced $n_1 = n_2 = \cdots n_r = n$

$$q = \frac{\bar{Y}_{max} - \bar{Y}_{min}}{\sqrt{\frac{MSE}{n}}} \sim \text{Tukey's studentized range distr. with parameters } (r, n_T - r)$$

Testing: $H_0: \mu_i = \mu_{i'}$, for all $i, i' = 1, \cdots r$

vs H_A: H_0 is not true

gives us the simultaneous testing result for

$H_0: \mu_i = \mu_{i'}$ for all i, i'.

$$1 - \alpha = 1 - \text{FWER (family wise error rate)}$$

$$= P(\text{none of the pairwise comparison is rejected})$$

$$= P\left(\frac{|\bar{y}_{i\cdot} - \bar{y}_{i'\cdot}|}{\sqrt{MSE(\frac{1}{n} + \frac{1}{n})}} < \text{critical value for all } i, i'\right)$$

$$\leq \quad = P\left(\frac{\bar{Y}_{max} - \bar{Y}_{min}}{\sqrt{MSE\left(\frac{1}{n}+\frac{1}{n}\right)}} < c.v. \right)$$

$$\frac{\bar{Y}_{max} - \bar{Y}_{min}}{\sqrt{\frac{MSE}{n}}\sqrt{2}} = \frac{q}{\sqrt{2}}$$

$$= P\left(\frac{q}{\sqrt{2}} < \frac{q_\alpha(r, N_T - r)}{\sqrt{2}} \right)$$

$$= 1 - \alpha$$

- If not balanced, approximately.

~~Fall~~ For all $1 \le i, i' \le r$, the $100(1-\alpha)\%$ Tukey simultaneous C.I. for $\mu_i - \mu_{i'}$ is

$$\bar{y}_{i\cdot} - \bar{y}_{i'\cdot} \pm \frac{q_\alpha(r, n_T - r)}{\sqrt{2}} SE(\bar{y}_{i\cdot} - \bar{y}_{i'\cdot})$$

$$SE(\bar{y}_{i\cdot} - \bar{y}_{i'\cdot}) = \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}$$

For $H_0: \mu_i - \mu_{i'} = 0$ for all i, i' vs H_A

reject H_0 if $\dfrac{|\bar{y}_{i\cdot} - \bar{y}_{i'\cdot}|}{SE(\bar{y}_{i\cdot} - \bar{y}_{i'\cdot})} > \dfrac{q_\alpha(r, n_T - r)}{\sqrt{2}}$

EX: treatment

A	B	C
5	3	1
2	3	0
5	0	1
4	2	2
2	2	1
$\bar{y}_{1\cdot} = 3.6$	$\bar{y}_{2\cdot} = 2$	$\bar{y}_{3\cdot} = 1$

$$T = q \cdot \sqrt{\frac{MSE}{n}}$$

$$= 3.77 \sqrt{\frac{1.43}{5}} = 1.99$$

$\uparrow$

$q(.95, 3, 12)$

group 1 & 2 $\quad |\bar{y}_{1\cdot} - \bar{y}_{2\cdot}| < T$

2   3 $\quad |\bar{y}_{2\cdot} - \bar{y}_{3\cdot}| < T$

1   3 $\quad |\bar{y}_{1\cdot} - \bar{y}_{3\cdot}| = 2.6 > T$

Scheffé multiple comparison Procedure

- Suppose there are r factor ~~levels~~ in total

$$L = \sum_{i=1}^{r} c_i \mu_i \qquad \left(\sum_i c_i = 0 \right)$$

$$\hat{L} = \sum_{i=1}^{r} c_i \bar{y}_{i\cdot}, \qquad SE(\hat{L}) = \sqrt{MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}}$$

- $100(1-\alpha)\%$ CI

$$\hat{L} \pm \sqrt{(r-1) F_{\alpha, r-1, n_T - r}} \; SE(\hat{L})$$

- For testing $H_0: L = 0$ vs $H_A: L \neq 0$

test statistic $T = \dfrac{|\hat{L}|}{SE(\hat{L})} > \sqrt{(r-1) F_{\alpha, r-1, n_T - r}}$

- Most conservative (least powerful)

$H_0: L = 0$
include infinite statements

Proof: Because $\sum_{i=1}^{r} c_i = 0$

$$\hat{L} = \sum_i c_i \bar{y}_{i\cdot} = \sum_i c_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})$$

By the Cauchy-schwartz inequ.

$$\left|\sum_i a_i b_i\right| \le \sqrt{\sum_i a_i^2 \sum_i b_i^2}$$

we have

$$|\hat{L}| = \left|\sum_i \frac{c_i}{\sqrt{n_i}} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})\sqrt{n_i}\right|$$

$$\le \sqrt{\sum_i \frac{c_i^2}{n_i} \sum_i n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}$$

$$= \sqrt{\sum_i \frac{c_i^2}{n_i} SSTR}$$

$$T = \frac{|\hat{L}|}{SE(\hat{L})} = \frac{|\hat{L}|}{\sqrt{MSE \sum_i \frac{c_i^2}{n_i}}} \le \frac{\sqrt{\sum_i \frac{c_i^2}{n_i} SSTR}}{\sqrt{\sum_i \frac{c_i^2}{n_i} MSE}} = \sqrt{\frac{SSTR}{MSE}}$$

Recall $F = \frac{MSTR}{MSE}$ is the ANOVA F-statistic

$$T \le \sqrt{\frac{SSTR}{MSE}} = \sqrt{\frac{(r-1)\,MSTR}{MSE}} = \sqrt{(r-1)F}$$

We ~~thus~~ get a uniform upper bound for the test statistic for any contrast L

$$T(L) \le \sqrt{(r-1)F}$$

- F has a F-distr. with df. $(r-1, n_T-r)$

$$P(F > F_{\alpha, r-1, n_T-r}) = \alpha$$

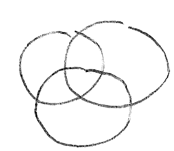
$$\begin{aligned}
\text{FWER} &= P(\text{any contrast } L \text{ is rejected}) \\
&= P\left(T(L) > \sqrt{(r-1) F_{\alpha, r-1, n_T-r}} \text{ for some contrast}\right) \\
&\le P\left(\sqrt{(r-1)F} > \sqrt{(r-1) F_{\alpha, r-1, n_T-r}}\right) \\
&= P(F > F_{\alpha, r-1, n_T-r}) \\
&= \alpha
\end{aligned}$$

- Bonferroni's method

$$\hat{L} \pm B \, SE(\hat{L})$$

$$B \downarrow \; t_{(1-\frac{\alpha}{2g}, n_T-r)} \qquad g: \# \text{ of contrast}$$

$$P\left(\bigcap_{i=1}^{g} A_i\right) = 1 - P\left(\bigcup_{i=1}^{g} A_i^c\right)$$



$$\ge 1 - \sum_{i=1}^{g} P(A_i^c) = 1 - \sum_{i=1}^{g} (1 - P(A_i))$$

- Holm Method (Refined Bonferroni Method, more powerful than Bonferroni)

Order P-values $P_{(1)} \le P_{(2)} \le \cdots P_{(K)}$

Then. if $P_{(1)} > \frac{\alpha}{K} \Rightarrow$ accept $H_0^{(1)}, \cdots H_0^{(K)}$

o.w. reject $H_0^{(1)}$, go to $P_{(2)}$

If $P_{(2)} > \frac{\alpha}{K-1} \Rightarrow$ accept $H_0^{(2)}, \cdots H_0^{(K)}$

o.w. reject $H_0^{(2)}$, go to $P_{(3)}$

If $P_{(3)} > \frac{\alpha}{K-2} \Rightarrow$ accept $H_0^{(3)}, \cdots H_0^{(K)}$

etc.

Proof: why Holm's method control ~~ex~~ FWER.

Denote $J_T :=$ index set of true hypotheses

Define $j :=$ first index of a true hypotheses.

$\Rightarrow H_0^{(1)}, \cdots H_0^{(j-1)}$ are all false

$\Rightarrow (j-1) \le K - |J_T|$

$|J_T|$: # of index in J_T

$j \le K+1-|J_T|$

$K+1-j \ge |J_T|$

Now for $\forall J_T$

$P(\text{at least 1 type I error})$

$= P(\text{Reject } H_0^{(j)})$

$= P\left(P_{(j)} \le \frac{\alpha}{K+1-j}\right) \le P\left(P_{(j)} \le \frac{\alpha}{|J_T|}\right)$

$= P\left(\min_{j \in J_T} \underbrace{P_j}_{\sim U(0,1)} \le \frac{\alpha}{|J_T|}\right) \le \sum_{j \in J_T} \frac{\alpha}{|J_T|} = \alpha$

$P\left(\cancel{\text{some}}\ P_i \le \frac{\alpha}{|J_T|} \text{ for some } i\right)$

- Fisher's least significant Difference (LSD)

  The LSD is the minimun amount by which two means must differ in order to be considered statistically different.

  - a usual t-test
  - No adjustment is made for multiple comparisons.

- Summary of multiple comparison adjustments

powerful ↑ ↓ Conservative

Method	family of contrast	C.V.
Fisher LSD	a single pairwise comparison	$t_{\frac{\alpha}{2}, N_T - r}$
Tukey	all pairwise comparison	$q_\alpha(r, N_T - r)/\sqrt{2}$
Bonferroni	varies	$t_{\frac{\alpha}{2k}, N_T - r}$ k: # of test
Scheffé	all contrasts	$\sqrt{(r-1) F_{\alpha, r-1, n_T - r}}$

Ex. Recall

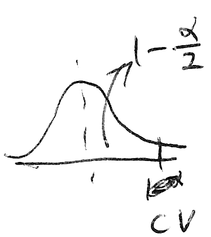
treatment	A	B	C
n_i	5	5	5
$\bar{y}_{i\cdot}$	3.6	2	1

$MSE = 1.43$

$H_0: \mu_i = \mu_{i'}$ for all i, i'

the C.V. at $\alpha = .05$ are

```
> alpha <- .05
  r <- 3
  k <- (3 choose 2) = 3
  NT <- 15
```



$> qt(1 - \frac{alpha}{2}, df = N_T - r)$ # Fisher's LSD

2.178813

$> qt(1 - \frac{alpha/2}{k}, df = N_T - r)$ # Bonferroni

2.7794

$> qtukey(1 - alpha, r, N_T - r)/\sqrt{2}$ # Tukey

2.6679

$> sqrt((r-1)qf(1-alpha, r-1, n_T - r)$ # scheffé

2.7876

C.I width

$$C.V \times SE(\bar{y}_{i\cdot} - \bar{y}_{i'\cdot})$$

$$C.V. \cdot \sqrt{MSE\left(\frac{1}{5}+\frac{1}{5}\right)} = C.V. \sqrt{1.43 \times \frac{2}{5}} = C.V. \cdot .7563$$

power.

scheffé < Bonferroni < Tukey

Scheffe : conservative, works better for "many comparison".

Tukey : pairwise (not work for contrast like $\mu_1 - \frac{\mu_2+\mu_3}{2}$)

Bonferroni. conservative, works for smaller # of comparisons

In R

pairwise. t. test (y, x, p. adject = "none", "holm", "bonferroni")

TukeyHSD. A B C

(AB) (AC), (BC)