Lecture 17 Wednesday February 22

- Quiz 3 is due tonight.
- Project 1 is available, is due Wed March 8

PLAN TODAY

- Finish example from end of Lec16, Individual CI for contrast L.
- Define, illustrate contrast.
- Simultaneous inference - Introduction
- Tukey's studentized range method for all pairwise comp's.

EX. Kenton Food Company, 95% individual CI for $L = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4}{2}$ → (cartoons minus no cartoons)

First get $\hat{L} = \frac{\bar{Y}_{1\cdot} + \bar{Y}_{3\cdot}}{2} - \frac{\bar{Y}_{2\cdot} + \bar{Y}_{4\cdot}}{2}$

$= \frac{14.6 + 19.5}{2} - \frac{13.4 + 27.2}{2}$ from summary stats p.86

$= 17.05 - 20.3 = -3.25$

Last time we found $Var(\hat{L}) = \sigma^2 \frac{17}{80}$, and $\hat{Var}(\hat{L}) = MSE\left(\frac{17}{80}\right)$

so $\hat{Var}(\hat{L}) = 10.547\left(\frac{17}{80}\right) = 2.24124$

and $s(\hat{L}) = \sqrt{2.24124} = 1.4971$

The critical constant for 95% confidence is $t_{15,.975} = 2.131$, and the allowance is $2.131 \times 1.4971 = 3.19$

The 95% CI for L is $(-6.44, -0.06)$

One conclusion is that we are 95% confident that "no-cartoon" design sells better than "cartoon" design) at least slightly.

Assume we are given a one-way ANOVA setting with cell means $\mu_i, i = 1, \ldots, r$.

We define a *linear combination* L to be $L = \sum_{i=1}^{r} c_i \mu_i$, for some set of constants c_i, with no restrictions on the c_i's. $c_i,\ i = 1, \ldots, r$

Examples The three examples for the Kenton Food Company that we just discussed are all linear combinations of the cell means. Set of cell means: $\{\mu_1, \mu_2, \mu_3, \mu_4\}$

Ex. 1 $\mu_4 = 0 \cdot \mu_1 + 0 \cdot \mu_2 + 0 \cdot \mu_3 + 1 \cdot \mu_4$ Ex. 2 $\mu_1 - \mu_2 = \underbrace{1}_{c_1}\mu_1 + \underbrace{(-1)}_{c_2}\mu_2 + \underbrace{0}_{c_3}\cdot\mu_3 + \underbrace{0}_{c_4}\cdot\mu_4$

Ex. 3 below

A *contrast* is any linear combination of the cell means where the constants sum to zero. EX, Kenton Check whether $\sum_{i=1}^{4} c_i = 0$

Examples

- A single cell mean μ_4 is **not a contrast** $\sum_{i=1}^{4} c_i = 0 + 0 + 0 + 1 = 1 \neq 0$
- Any pairwise difference is a contrast. eg $\mu_1 - \mu_2$ $\sum_{i=1}^{4} c_i = 1 + (-1) + 0 + 0 = 0$
- Another contrast:

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

$$= \frac{1}{2}\mu_1 - \frac{1}{2}\mu_2 + \frac{1}{2}\mu_3 - \frac{1}{2}\mu_4$$

$$\sum_{i=1}^{4} c_i = \frac{1}{2} - \frac{1}{2} + \frac{1}{2} - \frac{1}{2} = 0$$

How about $L_2 = \mu_4 - \dfrac{\mu_1 + \mu_2 + \mu_3}{3}$? c_i's: $\frac{-1}{3}, \frac{-1}{3}, \frac{-1}{3}, 1$ $\sum^{4} c_i = 0$ so L_2 is a contrast

How about $L_3 = \dfrac{\mu_1 + \mu_3}{2}$

c_i's $\frac{1}{2},\ 0,\ \frac{1}{2},\ 0$

$\sum c_i = 1 \neq 0$

so L_3 is not a contrast

Need for Simultaneous Inference in One-Way ANOVA

With r groups, we are interested in up to $\binom{r}{2}$ comparisons of means, and may be interested in other *contrasts* as well.

pairwise comparisons

How many pairwise comparisons are there?

Answer. With r groups, there are $\binom{r}{2} = \frac{r!}{2!\,(r-2)!} = \frac{r(r-1)}{2}$ pairwise comparisons.

r	$\binom{r}{2}$
3	3
4	6

r	$\binom{r}{2}$
5	10
8	28

Methods are needed to control the family-wise confidence level or significance level. In experimental studies, this makes conclusions more interpretable. In observational studies, these methods make *data snooping* possible.

EX. Kenton $r=4$

Chart of all pairwise comparisons (outline)

Parameter D_i	Estimate $\hat{D}_i$	St. Error $s(\hat{D}_i)$	95% CI
$D_1 = \mu_1 - \mu_2$			
$D_2 = \mu_1 - \mu_3$			
$D_3 = \mu_1 - \mu_4$			
$D_4 = \mu_2 - \mu_3$			
$D_5 = \mu_2 - \mu_4$			
$D_6 = \ldots$			

We often use notation "D" for pairwise differences

With six individual CIs we can conclude, for each CI,

"We are 95% confident that CI 1 contains D_1."
" " " " " " CI 2 " D_2."
⋮
" " " " " " CI 6 " D_6."

That is - we have six ~~a~~ individual conclusions.

MUCH BETTER to have one statement which enables us to read the chart as a whole.
We'd like to say: "We are 95% confident that each of the six CIs contains its respective parameter."

90-1

We will consider three multiple comparison methods:

1 Tukey's studentized range, for all possible pairwise comparisons. Useful in practice. Has to be tweaked to handle unequal group sizes.

Requires a balanced design — all n_i must be equal.
So, this method doesn't apply to the Kenton example.

2 Scheffé method, for all possible contrasts. Amazing that such a method exists, but the intervals tend to be very wide.

3 Bonferroni method. Useful, flexible, easy. Requires that you specify contrasts to be estimated in advance of getting the data.

No data snooping

Tukey Studentized Range Intervals

Consider the usual one-way ANOVA model with r treatments, and suppose we want to do $\binom{r}{2}$ tests to compare pairs μ_i, μ_j for all $i \neq j$, and also to get CI's for all possible differences $\mu_i - \mu_j$.

Suppose the sample sizes are all equal, and let n denote the common value.

The Tukey method is ideal for comparing all pairs of means, when sample sizes are equal.

Tukey's

Studentized range distribution Probability Background

Suppose:

- $Y_1, Y_2, \ldots, Y_r$ are independent and normally distributed with mean μ and standard deviation σ.
- The statistic s^2 is an unbiased estimate of σ^2, is independent of the Y_i's, and the quantity νs^2 has a χ^2 distribution with ν d.f..

Let $R = \max(Y_i) - \min(Y_i)$ be the range of the Y's.

The studentized range is

$$q = \frac{R}{s}$$

The distribution of q depends on the parameters r and ν, is available in tables and in software.

Example For $r = 7$, $\nu = 21$, the .95 quantile of the studentized range distribution is:

nmean

```
> qtukey(.95,~~x~~=7,df=21)
[1] 4.597302
```

For the stud. range distr. w/ r = 7 and df = 21, 95% of the distr. is between 0 and 4.597

OPTIONAL

In one-way ANOVA to apply the studentized range distribution first consider a little intuition:

① $\bar{Y}_{1\cdot} - \mu_1,\ \bar{Y}_{2\cdot} - \mu_2 \ldots \bar{Y}_{r\cdot} - \mu_r$
Here we have r iid $N\left(0, \frac{\sigma^2}{n}\right)$ r.v.'s

② Note that $\frac{MSE}{n}$ is an unbiased estimate of $\frac{\sigma^2}{n}$

And, note that the range of the $\bar{Y}_{i\cdot} - \mu_i$'s is the largest pairwise difference of

$$\left\{ (\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{j\cdot} - \mu_j) \right\}_{\substack{i=1,\ldots,r \\ j=1,\ldots,r \\ i \neq j}} = \left\{ (\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}) - (\mu_i - \mu_j) \right\}$$

Pairwise comparisons of means

r groups with n observations per group, MSE estimates σ^2.

We want to give a confidence interval formula which used the standard error of the difference of means. Recall:

$$\text{SE}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}) = \sqrt{\text{MSE}}\sqrt{\frac{2}{n}}$$

The Tukey studentized range method for all pairwise comparisons of means forms $\binom{r}{2}$ intervals as follows. For estimation of $\mu_i - \mu_j$, take $\hat{D} = \overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}$.

THE WAY TO CONSTRUCT TUKEY CIs

$$\hat{D} \pm \frac{1}{\sqrt{2}} q(1-\alpha, r, n_T - r)\text{SE}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})$$