Stat 445/545: Analysis of Variance and Experimental Design

Chapter 17: Analysis of factor level means

Instructor: Yan Lu

THE UNIVERSITY OF NEW MEXICO

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

- $i = 1, 2, \cdots, r; j = 1, 2, \cdots n_i$
- $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$

Parameter	μ_i	σ^2
Estimator	$\hat{\mu}_i = \bar{Y}_{i.}$	$\hat{\sigma}^2 = MSE = \frac{\sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2}{n - r}$
Expected value	$E(\bar{Y}_{i.}) = \mu_i$	$E(MSE) = \sigma^2$
Variance	$V(\bar{Y}_{i.}) = \sigma^2 / n_i$	
Estimated variance	$s^2(\bar{Y}_{i.}) = MSE / n_i$	

Inference for Single Factor Level Mean

- $\bar{Y}_{i.} \sim N(\mu_i, \sigma^2/n_i)$
- $\dfrac{\bar{Y}_{i.} - \mu_i}{s(\bar{Y}_{i.})} \sim t(n-r)$
- $(1-\alpha)100\%$ confidence interval for μ_i is

$$\bar{Y}_{i.} \pm t\left(1 - \frac{\alpha}{2}, n - r\right) s(\bar{Y}_{i.})$$

$$= \bar{Y}_{i.} \pm t\left(1 - \frac{\alpha}{2}, n - r\right) \sqrt{\frac{MSE}{n_i}}$$

—— $t\left(1 - \frac{\alpha}{2}, n - r\right)$ is the upper tail $100(1-\alpha/2)$ percentile of the t-distribution with $n-r$ degrees of freedom

—— $1-\alpha$ is the confidence level

Testing:

$$H_0 : \mu_i = c \text{ versus } \mu_i \neq c$$

- check if c is in the confidence interval
  ——If c is not in the CI, reject H_0 at level of significance α
  ——Otherwise, don't reject H_0.

Equivalently, we can compute the test statistic

$$t^* = \frac{\bar{Y}_{i.} - c}{s(\bar{Y}_{i.})}$$

Test statistic t^* follows a t distribution with $n - r$ degrees of freedom when H_0 is true.

- Reject H_0 when $|t^*| > t\left(1 - \frac{\alpha}{2}; n - r\right)$,
- Otherwise, we conclude H_0.

Example: In the Kenton Food Company example, the sales manager wished to estimate mean sales for package design 1 with a 95 percent condifence interval. Using the results from Table 17.1, we have

$$\bar{Y}_{1.} = 14.6, n_1 = 5, MSE = 10.55$$

we require $t(0.975; 15) = 2.131$. Finally,

$$s^2(\bar{Y}_{i.}) = \frac{MSE}{n_i} = \frac{10.55}{5} = 2.110$$

so that $s(\bar{Y}_{i.}) = 1.453$. Hence, CI is $14.6 \pm 2.131(1.453)$, the 95% confidence interval is

$$11.5 \leq \mu_1 \leq 17.7$$

Thus, we estimate with confidence coefficient 0.95 that the mean sales per store for package design 1 are between 11.5 and 17.7 cases.

Differences between two means

$$D = \mu_i - \mu_j, \hat{D} = \bar{Y}_{i.} - \bar{Y}_{j.}$$

$$E(\hat{D}) = \mu_i - \mu_j$$

Since $\bar{Y}_{i.}$ and $\bar{Y}_{j.}$ are independent,

$$V[\hat{D}] = V[\bar{Y}_{i.}] + V[\bar{Y}_{j.}] = \sigma^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)$$

$$\hat{V}[\hat{D}] = MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)$$

$$\frac{\hat{D} - D}{s(\hat{D})} \sim t(n-r)$$

$(1-\alpha)100\%$ CI for D is

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

Testing

$$H_0 : \mu_i = \mu_j \text{ or } \mu_i - \mu_j = 0$$

$$H_\alpha : \mu_i \neq \mu_j \text{ or } \mu_i - \mu_j \neq 0$$

Test statistic

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

Conclusion of H_0 is reached if $|t^*| \leq t\left(1 - \frac{\alpha}{2}; n - r\right)$, Otherwise, H_α is concluded.

Example: For the Kenton Food Company example, package designs 3 and 4 used 5-color printing. We wish to estimate the difference in mean sales for 5-color designs 3 and 4 using a 95 percent condifence interval. That is, we wish to estimate $D = \mu_3 - \mu_4$. From Table 17.1, we have

$$\bar{Y}_{3.} = 19.5, n_3 = 4, MSE = 10.55$$

$$\bar{Y}_{4.} = 27.2, n_4 = 5$$

Hence

$$\hat{D} = \hat{Y}_{3.} - \hat{Y}_{4.} = 19.5 - 27.2 = -7.7$$

The estimated variance of $\hat{D}$ is:

$$s^2(\hat{D}) = MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right) = 10.55\left(\frac{1}{4} + \frac{1}{5}\right) = 4.748$$

so that $s(\hat{D}) = 2.179$. Hence, CI is $-7.7 \pm 2.131(2.179)$, the 95% confidence interval is

$$-12.1 \leq \mu_3 - \mu_4 \leq -3.1$$

Thus, we estimate with confidence coefficient 0.95 that the mean sales for package design 3 fall short of those for package design 4 by somewhere between 3.1 and 12.3 cases per store.

Note that the only difference between package desgin 3 and 4 is the presence of cartoons: both designs used 5-color printing. The sales manager may therefore wish to test whether the addition of cartoons affects sales for 5-color designs.

$$H_0 : \mu_3 - \mu_4 = 0$$

$$H_\alpha : \mu_3 - \mu_4 \neq 0$$

Since the hypothesized difference zero in H_0 is not contained within the 95% CI $-12.1 \leq \mu_3 - \mu_4 \leq -3.1$, we conclude H_α, that the presence of cartoons has an effect.
We could also obtain

$$t^* = \frac{\hat{D}}{s(\hat{D})} = \frac{-7.7}{2.179} = -3.53$$

since $|t^*| = 3.53 > t(0.975, 15) = 2.131$, we conclude H_α. The two sided P-value for this test is 0.003.

Contrast of factor level means

Contrast: a comparison of two or more factor level means of the form

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

Example: (a) $\mu_1 - \mu_2, c_1 = 1, c_2 = -1, c_1 + c_2 = 0$

(b) $\mu_1 + \mu_2 - (\mu_3 + \mu_4)$

(c) $\frac{1}{3}\mu_1 + \frac{2}{3}\mu_2 - \mu_3$

Estimator:

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

- $E(\hat{L}) = L$
- $V(\hat{L}) = \sum_{i=1}^{r} c_i^2 V(\bar{Y}_{i.}) = \sum_{i=1}^{r} c_i^2 \frac{\sigma^2}{n_i} = \sigma^2 \sum_{i=1}^{r} \frac{c_i^2}{n_i}$
- $s^2(\hat{L}) = \hat{V}(\hat{L}) = MSE \sum_{i=1}^{r} c_i^2 / n_i$

Under model assumptions,

$$\frac{\hat{L} - L}{s(\hat{L})} \sim t(n - r)$$

$(1 - \alpha)100\%$ CI for $L = \sum_{i=1}^{r} c_i \mu_i$ is

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

Testing

$$H_0 : L = 0$$

Test statistic

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

Reject H_0, if

$H_\alpha : L < 0,\ t^* < -t\,(1-\alpha; n-r)$

$H_\alpha : L > 0,\ t^* > t\,(1-\alpha; n-r)$

$H_\alpha : L \neq 0,\ |t^*| > t\left(1-\frac{\alpha}{2}; n-r\right)$

Review probability of intersection of two events:
Suppose we have two statements: s_1 and s_2

- Statement 1 (Event A) is correct with probability $1 - \alpha$.
- Statement 2 (Event B) is correct with probability $1 - \alpha$.
- What is the probability that both statements are simultaneously correct?
  (1) If the statements are independent, then the probability that both are correct is $(1 - \alpha)(1 - \alpha)$.
  (2) If they are not independent, the probability is hard to determine, but it should be less than $1 - \alpha$.

$$P(A \cap B) = P(A) - P(A \cap B^C) < 1 - \alpha$$

Comments:

(1) the level of the confidence $(1-\alpha)$ applies only to one particular CI and not to an entire collection of CIs that might be of interest

——Example: $r = 3$, want CI for

$\mu_1 - \mu_2, 95\% CI$

$\mu_1 - \mu_3, 95\% CI$

$\mu_2 - \mu_3, 95\% CI$

the confidence coefficient for all 3 statements together, is not 0.95, but less than 0.95.

(2) the level of confidence $(1-\alpha)$ applies to the CI for a function of factor level means, only if the function was determined without reference to the data. If the data suggested the function to be considered, the level of confidence is not $(1-\alpha)$.

—-want to explore the data and test hypothesis suggested by the data, this is called data snooping. if the data suggests a hypothesis to test, the level of significance of the test will be bigger than the one specified.

——Example: $r=6$, if always compares the min and max factor level means at $\alpha=0.05$, then aobut 40% of the times the test will conclude that the means are different even in fact there are no difference among any factor level means $(\mu_1=\mu_2=\cdots=\mu_r)$

Tukey Multiple Comparisons

Need simutaneous inference procedure methods for constructing confidence intervals for
(1) a number of pre-specified functions of the level means for which we want confidence intervals or conduct tests of hypothesis
(2) avoid data snooping
The family of interest is the set of all pairwise comparisons of factor level means, for $i = 1, 2, \cdots, r; j = 1, 2, \cdots, n_i$,

$$H_0 : \mu_i - \mu_j = 0$$

$$H_\alpha : \mu_i - \mu_j \neq 0$$

Studentized Range Distribution

Idea: suppose $Y_1, \cdots, Y_r \sim N(\mu, \sigma^2)$, let w be the range for this set of observations, such that

$$w = max\{Y_i\} - min\{Y_i\}.$$

- $-w \leq Y_i - Y_j \leq w$.
- If we have distribution for w, we have simultaneous CI's for all $\mu_i - \mu_j$
- Suppose that we have an estimate s^2 of the variance σ^2 which is based on v degrees of freedom and is independent of Y_i. The ratio w/s is called the studentized range denoted by

$$\frac{w}{s} = q(r, v)$$

- Distribution of q has been tabulated in Table B.9.

$$q(0.95; 5, 10) = 4.65$$

i.e.

$$P\left(\frac{w}{s} = q(5, 10) \leq 4.65\right) = 0.95$$

	Tukey Multiple Comparison for $i = 1, \cdots, r; j = 1, \cdots, n_i$ total of r choose 2 pairs	Single Pairwise Comparison for i and j one pair
H_0	$D = \mu_i - \mu_j = 0$	$D = \mu_i - \mu_j = 0$
Estimator $\hat{D}$	$\bar{Y}_{i.} - \bar{Y}_{j.}$	$\bar{Y}_{i.} - \bar{Y}_{j.}$
$s^2(\hat{D})$	$MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$	$MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$
Distribution	studentized range T	t
CI	$\hat{D} \pm \frac{1}{\sqrt{2}} q(1-\alpha; r, n-r) s(\hat{D})$	$\hat{D} \pm t(1-\frac{\alpha}{2}; n-r) s(\hat{D})$
Reject H_0	$\lvert\frac{\hat{D}}{s(\hat{D})}\rvert > T$	$\lvert\frac{\hat{D}}{s(\hat{D})}\rvert > t$
	$T = \frac{1}{\sqrt{2}} q(1-\alpha; r, n-r)$	

Comments:

(1) Can be used for data snooping

(2) if not all the pairwise comparisons are of interest, level of confidence for the comparisons of interest is greater than or equal to $(1-\alpha)$

Example 1: equal sample size case, In a study of the effectiveness of different rust inhibitors,

- four brands A, B, C, D were tested.
- Altogether, 40 experimental units were randomly assigned to the four brands, with 10 units assigned to each brand.
- A portion of the results after exposing the experimental units to severe weather conditions is given in coded form in Table 17.2 a.
  —-The higher the coded value, the more effective is the rust inhabitor.

This study is a compeletly randomized design, where the levels of the single factor correspond to the four rust inhibitor brands.

Example 1: continued

$$r = 4, n = 40, n - r = 36, MSE = 6.140,$$

$q(0.95, 4, 36) = 3.814$ from table B.9 , so that

$$T = \frac{1}{\sqrt{2}} * 3.814 = 2.70$$

$$s^2(\hat{D}) = MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right) = 6.140\left(\frac{1}{10} + \frac{1}{10}\right) = 1.23$$

$$s(\hat{D}) = 1.11, \ T * s(\hat{D}) = 2.70 * 1.11 = 3.0$$

For $\mu_2 - \mu_1$

$$\hat{D} = \bar{Y}_{2.} - \bar{Y}_{1.} = 89.44 - 43.14 = 46.3$$

CI:

$$46.3 \pm 3.0, \ \text{i.e. } 43.3 \leq \mu_2 - \mu_1 \leq 49.3$$

Similarly, we can construct CIs for other differences betweent he factor level means.

Tukey multiple comparison CI

$$46.3 \pm 2.70 * 1.11, \text{ i.e. } 43.3 \leq \mu_2 - \mu_1 \leq 49.3$$

Single pairwise comparison CI
$t(0.975, 36) = 2.028$

$$46.3 \pm 2.028 * 1.11, \text{ i.e. } 44.04892 \leq \mu_2 - \mu_1 \leq 48.55108$$

The multiple comparison CI is wider than the single comparison CI for the same confidence level.

Table : Simultaneous Confidence Intervals FOR Pairwise Differences using the Tukey Procedure–Rust Inhibitor Example

CI	H_0	H_α
$43.3 \leq \mu_2 - \mu_1 \leq 49.3$	$\mu_2 = \mu_1$	$\mu_2 \neq \mu_1$
$21.8 \leq \mu_3 - \mu_1 \leq 27.8$	$\mu_3 = \mu_1$	$\mu_3 \neq \mu_1$
$-0.3 \leq \mu_1 - \mu_4 \leq 5.7$	$\mu_1 = \mu_4$	$\mu_1 \neq \mu_4$
$18.5 \leq \mu_2 - \mu_3 \leq 24.5$	$\mu_2 = \mu_3$	$\mu_2 \neq \mu_3$
$46.0 \leq \mu_2 - \mu_4 \leq 52.0$	$\mu_2 = \mu_4$	$\mu_2 \neq \mu_4$
$24.5 \leq \mu_3 - \mu_4 \leq 30.5$	$\mu_3 = \mu_4$	$\mu_3 \neq \mu_4$

Line plot

- List sample means from smallest to largest
  — $\bar{Y}_{A.} = 43.14$, $\bar{Y}_{B.} = 89.44$, $\bar{Y}_{C.} = 67.95$, $\bar{Y}_{D.} = 40.47$
- Draw a line below the means that are determined to be equal based on Tukey's method

```
D  A        C     B
-----
Performance Score
```

The line between D and A indicates that there is no clear evidence whether D or A is the better rust inhibitor. The absence of a line signifies that a difference in performance has been found and the location of the points indicates the direction of the difference.

- With a 95% family confidence, B is the best inhibitor (better by somewhere between 18.5 and 24.5 units than the second best), C is the second best, and A and D follow substantially behind with little or no difference between them.

Example 2, unequal size

$$s(\hat{D}) = MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)$$

need to recalculated for each pairwise comparison. When the Tukey procedure is used with unequal sample sizes, it is sometimes called the Tukey-Kramer procedure.

Comments on incosistency

Why F test rejects $H_0 : \mu_1 = \cdots = \mu_r$, but Tukey's method indicate that all means are equal

- The F test gives a critical region that is elliptical in shape, while the Tukey method gives critical regions that are rectangular in shape.

Scheffé's method
Gives simultaneous CIs for all possible contrasts (even infinitely many, even contrasts suggested by data) of the form

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

- $\hat{L} = \sum_{i=1}^{r} c_i \bar{Y}_{i.}$
- $s^2(\hat{L}) = MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}$
- CI: $\hat{L} \pm \mathcal{S} * s(\hat{L})$
  ——$\mathcal{S}^2 = (r-1)F(1-\alpha; r-1, n-r), \mathcal{S} = \sqrt{(r-1)F_{1-\alpha,r-1,n-r}}$
  ——$\mathcal{S}^2$ is called scheffé's multiplier for any number of contrasts.

Bonferroni Comparisons

Suppose we have two statements: s_1 and s_2

- Statement 1 is correct with probability $1 - \alpha$.
- Statement 2 is correct with probability $1 - \alpha$.
- What is the probability that both statements are simultaneously correct?
  (1) If the statements are independent, then the probability that both are correct is $(1 - \alpha)(1 - \alpha)$.
  (2) But they are not independent. The actual probability is difficult to compute.
- p(s_1 is true and s_2 is true)
  =p(both s_i's are simultaneously true)
  $\geq 1 - 2\alpha$
  —-this gives a lower bound on the probability that both statements are simultaneously true.

- Bonferroni Inequality
  Let $s_1, s_2 \cdots s_g$ be statements with
  $$p(s_i \text{ is true}) = 1 - \alpha_i$$
  then
  p(s_1 is true, s_2 is true $\cdots$ and s_g is true)
  =p(all s_i's are simultaneously true)
  $\geq$ 1-$\sum_{i=1}^{g} \alpha_i$
- If α_is are equal, p(s_1 is true, s_2 is true $\cdots$ and s_g is true) $\geq$ 1-$g\alpha$ or
  $$\text{Family Error Rate } < g\alpha.$$
  Example: Suppose $1 - \alpha_i = .90$, $g = 10$
  $p(\text{All} \quad 10 \quad s_i's \quad \text{true}) \geq 1 - \sum_{i=1}^{10} .10 = 0$
  The Bonferroni inequality works, but might not work very well.

Genreal Case

To get a joint confidence coefficient of at least $(1-\alpha)$ for g parameters, we construct each interval estimate with statement confidence coefficient $1-\alpha/g$

- The confidence coefficient is at least

$$1 - g * \frac{\alpha}{g} = 1 - \alpha.$$

- The Bonferroni method controls the **family error rate** FER by reducing the individual comparison error rate.
- We have at least $100(1-\alpha)\%$ confidence that all pairwise t-test statements hold simultaneously!

When the user has a specified collection of contrast or linear combinations of interest specified before hand (no data snooping)

$$L_j = \sum_{i=1}^{r} d_{ji}\mu_i$$

The $(1-\alpha)100\%$ confidence intervals for the collection of g linear combinations $(L_1, L_2, \cdots, L_g)$ are

$$\hat{L}_j \pm Bs(\hat{L}_j), \text{ where } \hat{L}_j = \sum_{i=1}^{r} d_{ji}\bar{Y}_{i.}$$

where $\hat{L}_j = \sum_{i=1}^{r} d_{ji}\bar{Y}_{i.}$
$s^2(\hat{L}_j) = MSE \sum_{i=1}^{r} d_{ji}^2/n_i$
$B = t\left(1 - \frac{\alpha}{2g}, n-r\right)$
Each confidence interval has confidence coefficient of $1-\alpha/g$, and the confidence coefficient for the family of g statements is at least $1-\alpha$.

Summaries and Comments:

- Tukey: used for all pairwise comparisons, when n_i's are constant; extended to unequal samploe size called Tukey-Kramer method.
- Scheffé: for all possible contrasts
- Bonferroni: g tests, perform each at α/g level

In general

- If only paiwise comparisons are to be made, the Tukey procedure gives narrower CI and is preferred; If not all pairwise comparisons are of interest, the Bonferroni procedure may be the better one at times
- In the case of general contrasts, the scheffé procedure tends to give narrower confidence limits and is therefore the preferred method.
- All the three procedures are of the form of estimator $\pm$ multiplier $*$ se, may compute all three multipliers, choose the smallest one.
- Tukey's and Scheffé's methods are ok with data snooping, but Bonferroni is not suitable with data snooping.