Stat 206: Linear Models

Lecture 18

Dec. 2, 2019

Analysis of Factor Level Means

Upon the rejection of $H_0 : \mu_1 = \cdots = \mu_I$:

- Investigate the nature of the differences among the factor level means.
- Comparison between two factor level means: $D = \mu_i - \mu_j$.
- *Contrast* of factor level means: $L = \sum_{i=1}^{I} c_i \mu_i$, where $\sum_{i=1}^{I} c_i = 0$.

Comparison Between Two Means

$D = \mu_i - \mu_j$ for some $i \neq j$.

- $\widehat{D} =$ is an estimator of D.
- $\mathrm{Var}(\widehat{D}) =$, $s(\widehat{D}) =$
- $\frac{\widehat{D}-D}{s(\widehat{D})} \sim$: $(1-\alpha)$ - confidence interval of D:

$$\widehat{D} \pm s(\widehat{D}) t(1 - \frac{\alpha}{2}; n_T - I).$$

- Test $H_0 : D = 0$ vs. $H_a : D \neq 0$. At the significance level α, check whether

$$0 \in \widehat{D} \pm s(\widehat{D}) t(1 - \frac{\alpha}{2}; n_T - I).$$

If , reject H_0 at level α and conclude the two means are different.

Comparison Between Two Means

$D = \mu_i - \mu_j$ for some $i \neq j$.

- $\widehat{D} = \overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}$ is an unbiased estimator of D.
- $\mathrm{Var}(\widehat{D}) = \mathrm{Var}(\overline{Y}_{i\cdot}) + \mathrm{Var}(\overline{Y}_{j\cdot}) = \sigma^2\{\frac{1}{n_i} + \frac{1}{n_j}\}$
- $s(\widehat{D}) = \sqrt{MSE(1/n_i + 1/n_j)}$.
- $\frac{\widehat{D}-D}{s(\widehat{D})} \sim t_{(n_T - I)}$: $(1-\alpha)$ - confidence interval of D:

$$\widehat{D} \pm s(\widehat{D})t(1 - \frac{\alpha}{2}; n_T - I).$$

- Test $H_0 : D = 0$ vs. $H_a : D \neq 0$. Check whether

$$0 \in \widehat{D} \pm s(\widehat{D})t(1 - \frac{\alpha}{2}; n_T - I).$$

If not, reject H_0 at level α and conclude the two means are different.

Rust Inhibitors

In a study of the effectiveness of different rust inhibitors, four brands (1,2,3,4) were tested. Altogether, 40 experimental units were randomly assigned to the four brands, with 10 units assigned to each brand. The resistance to rust was evaluated in a coded form after exposing the experimental units to severe conditions. This is a *balanced complete randomized design (CRD)*.

$\overline{Y}_{1\cdot} = 43.14, \ \overline{Y}_{2\cdot} = 89.44, \ \overline{Y}_{3\cdot} = 67.95, \ \overline{Y}_{4\cdot} = 40.47.$

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (df)	MS
Between treatments	SSTR=15953.47	$I-1=3$	MSTR=5317.82
Within treatments	SSE=221.03	$n_T - I = 36$	MSE=6.140
Total	SSTO=16174.50	$n_T - 1 = 39$	

95% C.I and testing for $D = \mu_1 - \mu_2$.

- $\widehat{D} = 43.14 - 89.44 = -46.3$.
- $s(\widehat{D}) = \sqrt{MSE(\frac{1}{n_1} + \frac{1}{n_2})} = \sqrt{6.14 \times \frac{2}{10}} = 1.11$.
- $t(1 - \frac{\alpha}{2}; n_T - I) = t(0.975; 36) = 2.03$.
- 95% C.I: $-46.3 \pm 1.11 \times 2.03 = [-48.6, \ -44]$.
- Since $0 \notin [-48.6, \ -44]$, reject $H_0 : \mu_1 = \mu_2$ at the 0.05 significance level.

Multiple Comparison

A family of statistical inferences are considered **simultaneously**:

- Errors are more likely to occur.
  - Suppose one tests 100 null hypotheses which are indeed all true. If the type I error rate of each test is 5% and if these tests are independent, then the probability of making at least one false rejection is $1 - 0.95^{100} = 99.4\%$.
- Simultaneously control the probability of committing such errors.
  - Multiple hypothesis testing: Control the family-wise type-I error rate (FWER).
  - Simultaneous confidence region: Maintain a family-wise confidence level.

Family-wise Confidence Intervals for Pairwise Comparisons

- For I factor levels, there are $I(I-1)/2$ distinct pairwise comparisons of the form $D_{ij} = \mu_i - \mu_j$ $(1 \le i < j \le I)$.
- Denote the $(1-\alpha)$-C.I. for D_{ij} by $C_{ij}(\alpha)$:

$$C_{ij}(\alpha) = \widehat{D}_{ij} \pm s(\widehat{D}_{ij}) \times t(1 - \frac{\alpha}{2}; n_T - I).$$

- $t(1 - \frac{\alpha}{2}; n_T - I)$ is the multiplier that gives the desired confidence coefficient $1-\alpha$:

$$P(D_{ij} \in C_{ij}(\alpha)) = 1 - \alpha,$$

i.e., the probability that D_{ij} falls out of C_{ij} is at most α.

- *Family-wise confidence coefficient* of this family of confidence intervals is defined as:

  i.e., the probability that these C.Is cover their respective parameter.

-

$$P(D_{ij} \in C_{ij}(\alpha), \text{ for all } 1 \leq i < j \leq I)$$
$$P(D_{ij} \in C_{ij}(\alpha)) = 1 - \alpha.$$

- How to construct C.Is such that the family-wise confidence coefficient is at least $1 - \alpha$?
- We should replace $t(1 - \frac{\alpha}{2}; n_T - I)$ by a multiplier (resulting in C.Is).

- *Family-wise confidence coefficient* of this family of confidence intervals is defined as:

$$P(D_{ij} \in C_{ij}(\alpha),\ \text{for } \textbf{all } 1 \le i < j \le I),$$

i.e., the probability that these C.Is **simultaneously** cover their respective parameter.

- Note

$$\begin{aligned} & P(D_{ij} \in C_{ij}(\alpha),\ \text{for all } 1 \le i < j \le I) \\ \le\ & P(D_{ij} \in C_{ij}(\alpha)) = 1 - \alpha. \end{aligned}$$

- How to construct C.Is such that the family-wise confidence coefficient is at least $1 - \alpha$?
- We should replace $t(1 - \frac{\alpha}{2}; n_T - I)$ by a larger multiplier (resulting in wider C.Is).

Tukey's Procedure

Tukey's procedure for families of pairwise comparisons:

$$C_{ij}^{T}(\alpha) := \widehat{D}_{ij} \pm s(\widehat{D}_{ij}) \times T$$

with the multiplier

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

where $q(I, n_T - I)$ is the *studentized range distribution* with parameters I and $n_T - I$.

- T is larger than the corresponding t-multiplier.
- The family-wise confidence coefficient is at least $1-\alpha$:

$$P(D_{ij} \in C_{ij}^{T}(\alpha), \text{ for all } 1 \leq i < j \leq I) \geq 1 - \alpha.$$

  - "=" holds for balanced designs.

Rust Inhibitors

Tukey's multiple comparison confidence intervals for all pairwise comparisons with a family-wise confidence coefficient 95%.

- $I = 4$, there are 6 pairwise comparisons:
  $\mu_1 - \mu_2,\ \mu_1 - \mu_3,\ \mu_1 - \mu_4,\ \mu_2 - \mu_3,\ \mu_2 - \mu_4,\ \mu_3 - \mu_4.$
- $T =$
- Note $T = 2.7$ is greater than the corresponding t-multiplier $t(0.975; 36) = 2.03$, so the Tukey's intervals are
  .
- Tukey's C.I for $\mu_1 - \mu_2$:

```
> qtukey(0.95,4,36)
[1] 3.808798
```

Rust Inhibitors

Tukey's multiple comparison confidence intervals for all pairwise comparisons with a family-wise confidence coefficient 95%.

- $I = 4$, there are 6 pairwise comparisons:
  $\mu_1 - \mu_2,\ \mu_1 - \mu_3,\ \mu_1 - \mu_4,\ \mu_2 - \mu_3,\ \mu_2 - \mu_4,\ \mu_3 - \mu_4.$
- $T = \frac{1}{\sqrt{2}} q(1-\alpha; I, n_T - I) = \frac{1}{\sqrt{2}} q(0.95; 4, 36) = \frac{1}{\sqrt{2}} 3.81 = 2.7.$
- Note $T = 2.7$ is greater than the corresponding t-multiplier $t(0.975; 36) = 2.03$, so Tukey's intervals are wider.
- Tukey's C.I for $\mu_1 - \mu_2$:

$$-46.3 \pm 1.11 \times 2.7 = [-49.3, -43.3].$$

```
> qtukey(0.95,4,36)
[1] 3.808798
```

- All six Tukey's confidence intervals:

$$-49.3 \le \mu_1 - \mu_2 \le -43.3, \; -27.8 \le \mu_1 - \mu_3 \le -21.8,$$
$$-0.3 \le \mu_1 - \mu_4 \le 5.7, \; 18.5 \le \mu_2 - \mu_3 \le 24.5,$$
$$46.0 \le \mu_2 - \mu_4 \le 52.0, \; 24.5 \le \mu_3 - \mu_4 \le 30.5.$$

- Zero is contained in one of the C.Is, but is not in the other five C.Is.
- Therefore, at the family-wise significance level 0.05, we should $\mu_1 = \mu_4$, but should the other five null hypotheses.
- Such a decision rule will for simultaneously testing of

- All six Tukey's confidence intervals:

$$-49.3 \le \mu_1 - \mu_2 \le -43.3, \; -27.8 \le \mu_1 - \mu_3 \le -21.8,$$
$$-0.3 \le \mu_1 - \mu_4 \le 5.7, \; 18.5 \le \mu_2 - \mu_3 \le 24.5,$$
$$46.0 \le \mu_2 - \mu_4 \le 52.0, \; 24.5 \le \mu_3 - \mu_4 \le 30.5.$$

- Zero is contained in one of the C.Is, but is not in the other five C.Is.
- Therefore, at the family-wise significance level 0.05, we should not reject $\mu_1 = \mu_4$, but should reject the other five null hypotheses.
- Such a decision rule will control FWER at level 0.05 for simultaneously testing of

$$H_{ij,0} : D_{ij} = 0, \quad 1 \le i < j \le I.$$

Studentized Range Distribution

Optional Reading.

- $X_1, \cdots, X_r \sim_{i.i.d.} N(\mu, \sigma^2)$.
- Let $W = \max\{X_i\} - \min\{X_i\}$ be the *range statistic*.
- Let S^2 be an estimator of σ^2, which has distribution $\sigma^2\chi^2_{(\nu)}/\nu$ and is independent with X_i's.
- Then the distribution of W/S is called a studentized range distribution with the number of groups being r and the degrees of freedom being ν, denoted by

$$\frac{W}{S} \sim q(r, \nu).$$

Tukey's Procedure: Derivation

Optional Reading.

Consider balanced design: $n_1 = \cdots = n_I = n$.

- $\overline{Y}_{1\cdot} - \mu_1, \cdots, \overline{Y}_{I\cdot} - \mu_I$ are i.i.d. $N(0, \frac{\sigma^2}{n})$.
- $MSE \sim \sigma^2 \chi^2_{(n_T - I)}/(n_T - I)$ is an estimator of σ^2 and is independent with $\overline{Y}_{i\cdot} - \mu_i$. *Why?*
- By definition of the studentized range distribution:

$$\frac{\max_i\{\overline{Y}_{i\cdot} - \mu_i\} - \min_i\{\overline{Y}_{i\cdot} - \mu_i\}}{\sqrt{MSE/n}} \sim q(I, n_T - I).$$

- Note

$$\begin{aligned}
& \max_i\{\overline{Y}_{i\cdot} - \mu_i\} - \min_i\{\overline{Y}_{i\cdot} - \mu_i\} \\
= \quad & \max_{i,j} |(\overline{Y}_{i\cdot} - \mu_i) - (\overline{Y}_{j\cdot} - \mu_j)| = \max_{i,j} |\widehat{D}_{ij} - D_{ij}|
\end{aligned}$$

- $s(\widehat{D}_{ij}) = \sqrt{MSE(\frac{1}{n} + \frac{1}{n})} = \sqrt{2}\sqrt{\frac{MSE}{n}}$.
- Family-wise confidence coefficient for Tukey's C.Is:

$$
\begin{aligned}
& P(D_{ij} \in C_{ij}^{T}(\alpha), \text{ for all } 1 \le i < j \le I) \\
= \; & P\left(\frac{|\hat{D}_{ij} - D_{ij}|}{s(\hat{D}_{ij})} \le T, \text{ for all } 1 \le i < j \le I\right) \\
= \; & P\left(\frac{\max_{i,j} |\hat{D}_{ij} - D_{ij}|}{\sqrt{2}\sqrt{\frac{MSE}{n}}} \le T\right) \\
= \; & P\left(\frac{\max_i\{\overline{Y}_{i.} - \mu_i\} - \min_i\{\overline{Y}_{i.} - \mu_i\}}{\sqrt{\frac{MSE}{n}}} \le \sqrt{2}T\right)
\end{aligned}
$$

which is $1 - \alpha$ if $T = \frac{1}{\sqrt{2}} q(1 - \alpha; I, n_T - I)$.

Why Not Just Look at the Largest Difference?

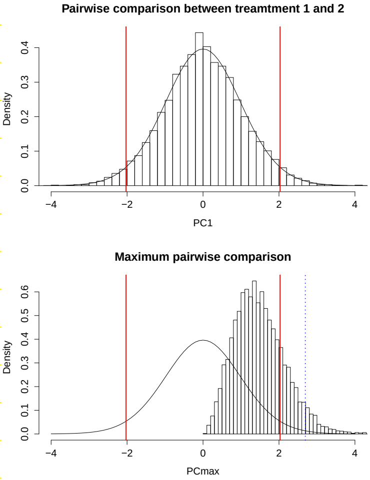
"Data snooping refers to statistical inference that the researcher decides to perform after looking at the data "

- Consider a single factor study with 4 levels. We used a CRD with 10 experimental units per treatment. Our goal is to see whether there is treatment effect.
- After collecting the data, we decided to construct a 95% confidence interval using the difference between the largest treatment sample mean and the smallest treatment sample mean: $D_{\text{max}} = \max\{\bar{Y}_{i\cdot}\} - \min\{\bar{Y}_{i\cdot}\}$:

$$D_{\text{max}} \pm t(0.975; 36) \sqrt{MSE \times \frac{2}{10}}$$

- The rationale is that if this interval does not contain zero, the treatment means are significantly different. *Is this correct?*

Top panel: $t_{(36)}$ distribution. Bottom Panel: Distribution of $\frac{D_{\max}}{\sqrt{MSE \times \frac{2}{10}}} \sim \frac{1}{\sqrt{2}} q(4, 36)$. Notice, how the bottom distribution shifts towards right compared to the $t_{(36)}$ distribution.



Bonferroni's Procedure

Suppose we want to construct g **prespecified** C.Is simultaneously.

- Bonferroni procedure: Construct each C.I at level
  . Then the familywise confidence coefficient is
  .
- Construct C.Is for g pairwise comparisons.
  - The Bonferroni's C.Is are of the form:

$$C^B(\alpha) = \widehat{D} \pm s(\widehat{D}) \times B.$$

    where $B =$
  - Then

$$P(D_{ij} \in C^B_{ij}(\alpha), \text{for } \textbf{all } g \text{ comparisons}) \quad 1 - \alpha.$$

Bonferroni's Procedure

Suppose we want to construct g **prespecified** C.Is simultaneously.

- Bonferroni procedure: Construct each C.I at level $1 - \alpha/g$. Then the familywise confidence coefficient is at least $1 - \alpha$.
- Construct C.Is for g pairwise comparisons.
  - The Bonferroni's C.Is are of the form:

$$C^B(\alpha) = \widehat{D} \pm s(\widehat{D}) \times B.$$

    where $B = t(1 - \frac{\alpha}{2g}; n_T - I)$.
  - Then

$$P(D_{ij} \in C_{ij}^B(\alpha), \text{for } \textbf{all } g \text{ comparisons}) \geq 1 - \alpha.$$

Bonferroni Inequality

Optional Reading.
If $A_1, \cdots, A_g$ are g events with $P(A_k) \geq 1 - \alpha/g \ \ (k = 1, \cdots, g)$, then

$$P(\bigcap_{k=1}^{g} A_k) \geq 1 - \alpha.$$

Proof.

$$\begin{aligned} P(\bigcap_{k=1}^{g} A_k) &= 1 - P(\bigcup_{k=1}^{g} A_k^c) \geq 1 - \sum_{k=1}^{g} P(A_k^c) \\ &\geq 1 - \sum_{k=1}^{g} \alpha/g = 1 - \alpha. \end{aligned}$$

Rust Inhibitors

Construct simultaneous C.Is for all 6 pairwise comparisons with $1 - \alpha = 0.95$.

- Bonferroni's multiplier: $I = 4, n_T = 40, g =$
- 95% Bonferroni's C.I. for $\mu_1 - \mu_2$:

$$-46.3 \pm 1.11 \times 2.79 = [-49.4, -43.2].$$

- Recall 95% Tukey's C.I. is $[-49.3, -43.3]$: Tukey's interval is . This is because Tukey's multiplier is $T = 2.7$, which is smaller than $B = 2.79$.
- If the family consists of **all pairwise comparisons**, then $T \quad B$ and thus Tukey's procedure is .

Rust Inhibitors

Construct simultaneous C.Is for all 6 pairwise comparisons with $1 - \alpha = 0.95$.

- Bonferroni's multiplier: $I = 4, n_T = 40, g = 6$

$$\begin{aligned} B &= t(1 - \frac{\alpha}{2g}; n_T - I) = t(1 - \frac{0.05}{12}; 36) \\ &= t(0.9958; 36) = 2.79. \end{aligned}$$

- 95% Bonferroni's C.I. for $\mu_1 - \mu_2$:

$$-46.3 \pm 1.11 \times 2.79 = [-49.4, -43.2].$$

- Recall 95% Tukey's C.I. is $[-49.3, -43.3]$: Tukey's interval is slightly narrower. This is because Tukey's multiplier is $T = 2.7$, which is smaller than $B = 2.79$.
- If the family consists of **all pairwise comparisons**, then $T < B$ and thus Tukey's procedure is better.

Contrasts

$$L = \sum_{i=1}^{I} c_i \mu_i, \quad with$$

- Examples. Pairwise comparisons: $\mu_i - \mu_j$ for $i \neq j$; $\frac{\mu_1+\mu_2}{2} - \mu_3$.
- Unbiased estimator:

$$\widehat{L} = \qquad\qquad\qquad\qquad , \quad \mathrm{Var}(\widehat{L}) =$$

- Standard error:

$$s(\widehat{L}) = \sqrt{\mathrm{MSE} \sum_{i=1}^{I} \frac{c_i^2}{n_i}}.$$

- $(1-\alpha)$ – confidence interval of L:

$$\widehat{L} \pm s(\widehat{L}) t(1 - \frac{\alpha}{2}; n_T - I).$$

Contrasts

$$L = \sum_{i=1}^{I} c_i \mu_i, \quad \textit{with} \quad \sum_{i=1}^{I} c_i = 0.$$

- Examples. Pairwise comparisons: $\mu_i - \mu_j$ for $i \neq j$; $\frac{\mu_1 + \mu_2}{2} - \mu_3$.
- Unbiased estimator:

$$\widehat{L} = \sum_{i=1}^{I} c_i \overline{Y}_{i\cdot}, \quad \mathrm{Var}(\widehat{L}) = \sum_{i=1}^{I} \sigma^2 c_i^2 / n_i.$$

- Standard error:

$$s(\widehat{L}) = \sqrt{\mathrm{MSE} \sum_{i=1}^{I} \frac{c_i^2}{n_i}}.$$

- $(1 - \alpha)$ – confidence interval of L:

$$\widehat{L} \pm s(\widehat{L}) t(1 - \frac{\alpha}{2}; n_T - I).$$

Package Design

Package Design (i)			Store (j)					
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	$Y_{i\cdot}$	$\overline{Y}_{i\cdot}$	n_i
1	11	17	16	14	15	73	14.6	5
2	12	10	15	19	11	67	13.4	5
3	23	20	18	17	miss	78	19.5	4
4	27	33	22	26	28	136	27.2	5
All Designs				$Y_{\cdot\cdot}$	$= 354$	$\overline{Y}_{\cdot\cdot}$	$= 18.63$	19

Package Design

Designs 1 and 2 are 3-color designs, while designs 3 and 4 are 5-color designs. We want to compare 3-color designs to 5-color designs in terms of their effects on sales.

- Consider the contrast:

- $c_1 =$, $c_2 =$, $c_3 =$, $c_4 =$: They add up to .
- An unbiased estimator of L:

$$\begin{aligned}
\widehat{L} &= \frac{\overline{Y}_{1\cdot} + \overline{Y}_{2\cdot}}{2} - \frac{\overline{Y}_{3\cdot} + \overline{Y}_{4\cdot}}{2} \\
&= \frac{14.6 + 13.4}{2} - \frac{19.5 + 27.2}{2} = -9.35.
\end{aligned}$$

Package Design

Designs 1 and 2 are 3-color designs, while designs 3 and 4 are 5-color designs. We want to compare 3-color designs to 5-color designs in terms of their effects on sales.

- Consider the contrast: $L = \frac{\mu_1+\mu_2}{2} - \frac{\mu_3+\mu_4}{2}$.
- $c_1 = c_2 = 0.5,\ c_3 = c_4 = -0.5$: They add up to zero.
- An unbiased estimator of L:

$$
\begin{aligned}
\widehat{L} &= \frac{\overline{Y}_{1\cdot} + \overline{Y}_{2\cdot}}{2} - \frac{\overline{Y}_{3\cdot} + \overline{Y}_{4\cdot}}{2} \\
&= \frac{14.6 + 13.4}{2} - \frac{19.5 + 27.2}{2} = -9.35.
\end{aligned}
$$

- Standard error:

$$
\begin{aligned}
s(\widehat{L}) &= \sqrt{MSE \sum_{i=1}^{I} \frac{c_i^2}{n_i}} \\
&= \sqrt{10.55 \times \left(\frac{(0.5)^2}{5} + \frac{(0.5)^2}{5} + \frac{(-0.5)^2}{4} + \frac{(-0.5)^2}{5} \right)} \\
&= \sqrt{10.55 \times 0.2125} = 1.5.
\end{aligned}
$$

- A 90%–C.I for L :

$$
\begin{aligned}
& \widehat{L} \pm s(\widehat{L}) \times t(0.95; 15) = -9.35 \pm 1.5 \times 1.753 \\
= \; & [-11.98, \; -6.72].
\end{aligned}
$$

- We are 90% confident that 5-color designs work better than 3-color designs. We can reject $H_0 : L = 0$ at significance level 0.1.

Scheffe's Procedure

There are infinitely many contrasts. How to control family-wise confidence coefficient or FWER if **all contrasts** or a large number of them are simultaneously considered?

- The family of all contrasts:

$$\mathcal{L} = \left\{ L = \sum_{i=1}^{I} c_i \mu_i : \quad \sum_{i=1}^{I} c_i = 0 \right\}.$$

Notes: All contrasts equal to zero if and only if $\mu_1 = \cdots = \mu_I$.

- *Scheffe's procedure*: Define the C.I. for a contrast L as

$$C_L^S(\alpha) := \hat{L} \pm s(\hat{L}) \times S,$$

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

- The family-wise confidence coefficient of $\{C_L^S(\alpha) : L \in \mathcal{L}\}$:

$$P(L \in C_L^S(\alpha), \text{ for all } L \in \mathcal{L}) = 1 - \alpha.$$

- Simultaneous testing: Reject $H_{0L} : L = 0$, if and only if

- Such a decision rule has a

What happens if the F test of equality of means is not rejected at level α?

- Simultaneous testing: Reject $H_{0L} : L = 0$, if and only if zero is not contained in the corresponding C.I. $C_L^S(\alpha)$.
- Such a decision rule has a family-wise type-I error rate at most α.

What happens if the F test of equality of means is not rejected at level α? Then all Scheffe's C.I contain zero.

Package Design

Suppose we want to maintain a family-wise confidence coefficient at 90% for all possible contrasts simultaneously.

- $S =$

- Scheffe's C.I. of $L = \frac{\mu_1+\mu_2}{2} - \frac{\mu_3+\mu_4}{2}$:

$$-9.35 \pm 1.50 \times 2.73 = [-13.4,\ -5.3].$$

- 0 is not contained in this interval, so we $H_0 : L = 0$ at familywise significance level 0.1.
- Scheffe's multiplier $S = 2.73$ is (much) larger than the multiplier $t(0.95; 15) = 1.753$ when we are only interested in . Consequently, Scheffe's C.I. is

Package Design

Suppose we want to maintain a family-wise confidence coefficient at 90% for all possible contrasts simultaneously.

- $S^2 = (I-1)F(1-\alpha; I-1, n_T - I) = 3 \times F(0.9; 3, 15) = 7.47$, $S = \sqrt{7.47} = 2.73$.
- Scheffe's C.I. of $L = \frac{\mu_1+\mu_2}{2} - \frac{\mu_3+\mu_4}{2}$:

$$-9.35 \pm 1.50 \times 2.73 = [-13.4,\ -5.3].$$

- 0 is not contained in this interval, so we can reject $H_0 : L = 0$ at familywise significance level 0.1.
- Scheffe's multiplier $S = 2.73$ is (much) larger than the multiplier $t(0.95; 15) = 1.753$ when we are only interested in a single contrast. Consequently, Scheffe's C.I. is (much) wider.

Compare Three Multiple Comparison Procedures

All three procedures have confidence intervals of the form:

- Tukey's procedure: Applicable to families of
  .
- Scheffe's procedure: Applicable to families of finite or infinite number of .
- Bonferroni's procedure: Applicable to families of finite number of .
- In practice, one could compute all **applicable multipliers** and use the multiplier to construct the C.Is.

Compare Three Multiple Comparison Procedures

All three procedures have confidence intervals of the form:

$$\text{Estimator} \pm \text{SE} \times \text{Multiplier}.$$

- Tukey's procedure: Applicable to families of pairwise comparisons.
- Scheffe's procedure: Applicable to families of finite or infinite number of contrasts. So Scheffe's procedure is more generally applicable than Tukey's procedure.
- Bonferroni's procedure: Applicable to families of finite number of **pre-specified** inferences.
- In practice, one could compute all **applicable multipliers** and use the smallest multiplier to construct the C.Is.

Model Diagnostics

Single factor ANOVA model:

$$Y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, \cdots, I, \; j = 1, \cdots, n_i.$$

Model assumptions:

- Normality: ϵ_{ij}'s are normal random variables (with mean zero).
- Equal Variance: ϵ_{ij}'s have the same variance.
- Independence: ϵ_{ij}'s are independent random variables.

- Effects of violation of model assumptions.
  - F-test and related procedures are pretty robust to the normality and equal variance assumptions.
  - Pairwise comparisons could be substantially affected by unequal variances.
  - Non-independence can have serious side effects and is hard to correct. So it is important to apply randomization whenever necessary.
- Diagnostic tools:
  - Residual plots: Check equal variance, normality, independence, outliers, etc.
- Remedial measures:
  - Transformations: Variance stabilizing transformations; BoxCox procedure.
  - Non-parametric tests: Rank F test.

Residuals

- Fitted values and residuals:

  $\widehat{Y}_{ij} =$, $e_{ij} =$, $i = 1, \cdots, I$, $j = 1, \cdots, n_i$.

- Studentized residuals:

  $$r_{ij} := \frac{e_{ij}}{s(e_{ij})},$$

  where $s(e_{ij}) =$.

- Studentized residuals adjust for difference in in different treatment groups and are across treatment groups even when the design is unbalanced.

Residuals

- Fitted values and residuals:

$$\widehat{Y}_{ij} = \overline{Y}_{i\cdot}, \quad e_{ij} = Y_{ij} - \overline{Y}_{i\cdot}, \quad i = 1, \cdots, I, \; j = 1, \cdots, n_i.$$

- Studentized residuals:

$$r_{ij} := \frac{e_{ij}}{s(e_{ij})},$$

  where $s(e_{ij}) = \sqrt{MSE \times (n_i - 1)/n_i}$. *Why?*

- Studentized residuals adjust for difference in sample size in different treatment groups and are comparable across treatment groups even when the design is unbalanced.

Check Equal Variance

By plots.

- Constancy of the error variance is supported by the residuals having (around zero) across different treatment groups.

Check Equal Variance

By residual vs. fitted value plots.

- Constancy of the error variance is supported by the residuals having similar extent of dispersion (around zero) across different treatment groups.

Check Normality

By

- Normality is supported by the Q-Q plots being

  .

Check Normality

By Normal Q-Q plots of the residuals.

- Normality is supported by the Q-Q plots being (nearly) linear.

Other things that can be examined by residual plots:

- Independence: if measurements are obtained in a time/space sequence, a residual sequence plot can be used to check whether the error terms are serially correlated.
- Outliers are identified by residuals with big magnitude.
- Existence of other important (but un-accounted for) explanatory variables can be identified to see whether residual plots show certain patterns.

Remedial Measures in ANOVA

How to make up for unequal variance and/or nonnormality?

- to stabilize the variance, which often also makes the distribution closer to Normal.
  - Variance stabilizing transformation.
  - Box-cox procedure.
- If the departures are too extreme such that transformations do not work, then use .
  - Rank F test for equality of means.

Remedial Measures in ANOVA

How to make up for unequal variance and/or nonnormality?

- Transformation of the response variable to stabilize the variance, which often also makes the distribution closer to Normal.
  - Variance stabilizing transformation.
  - Box-cox procedure.
- If the departures are too extreme such that transformations do not work, then use *nonparametric methods*.
  - Rank F test for equality of means.

Variance Stabilizing Transformations

When factor level variance is a function of the factor level mean, i.e., $\sigma_i^2 = \phi(\mu_i)$ for $i = 1, \cdots, I$, we can find a transformation $f(\cdot)$:

- Factor level variances of the transformed data $Y_{ij}^* = f(Y_{ij})$ are approximately equal.
- Often the Normality assumption also holds better for the transformed data.

Optional Reading.

Suppose $\mathrm{E}(Y) = \mu,\ \mathrm{Var}(Y) = \sigma^2 = \phi(\mu)$.

- Find a transformation $Y^* = f(Y)$ such that variance of Y^* is a constant (i.e., not depend on μ).
- By first order Taylor expansion:

$$Y^* \approx f(\mu) + f'(\mu)(Y - \mu).$$

- Therefore $\mathrm{E}(Y^*) \approx f(\mu)$ and $\mathrm{Var}(Y^*) \approx (f'(\mu))^2\phi(\mu)$
- Choose f such that $(f'(\mu))^2\phi(\mu) = 1$.
- Thus

$$f(\mu) = \int \frac{1}{\sqrt{\phi(\mu)}} d\mu.$$

Commonly Used Transformations

- If $\sigma_i^2 \propto \mu_i$, then use
- If $\sigma_i \propto \mu_i$, then use
- If $\sigma_i \propto \mu_i^2$, then use
- How to decide on which one to use?
  - Calculate
  - The approximate of one of the three statistics across treatment groups suggests the corresponding transformation.

Commonly Used Transformations

- If $\sigma_i^2 \propto \mu_i$, then use the square root transformation: $Y^* = \sqrt{Y}$.
- If $\sigma_i \propto \mu_i$, then use the log transformation: $Y^* = \log(Y)$.
- If $\sigma_i \propto \mu_i^2$, then use the inverse transformation: $Y^* = 1/Y$.
- How to decide on which one to use?
  - Calculate $\frac{s_i^2}{\overline{Y}_{i\cdot}}$, $\frac{s_i}{\overline{Y}_{i\cdot}}$, $\frac{s_i}{(\overline{Y}_{i\cdot})^2}$ for $i = 1, \ldots, I$.
  - The approximate constancy of one of the three statistics across treatment groups suggests the corresponding transformation.

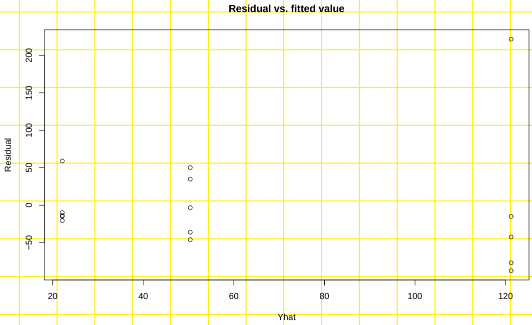
Computers

A company operates computers at three different locations. The computers are identical as to make and model, but are subject to different degrees of voltage fluctuation. The length of time between computer failures are recorded for three locations, each for five failure intervals. This is an observational study since no randomization of treatments to experimental units occurred.

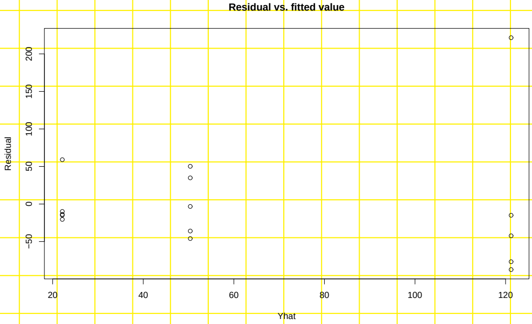
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	$\overline{Y}_{i\cdot}$
1	4.41	100.65	14.45	47.13	85.21	50.37
2	8.24	81.16	7.35	12.29	1.61	22.13
3	106.19	33.83	78.88	342.81	44.33	121.2

i	s_i^2	$\frac{s_i^2}{\overline{Y}_{i\cdot}}$	$\frac{s_i}{\overline{Y}_{i\cdot}}$	$\frac{s_i}{(\overline{Y}_{i\cdot})^2}$		
1	1788.7	35.5	0.84	0.017		
2	1103.454	49.9	1.5	0.068		
3	16167.4	133.4	1.05	0.009		
	$\overline{Y}_{\cdot\cdot}$	=	64.6;	MSE	=	6353.2

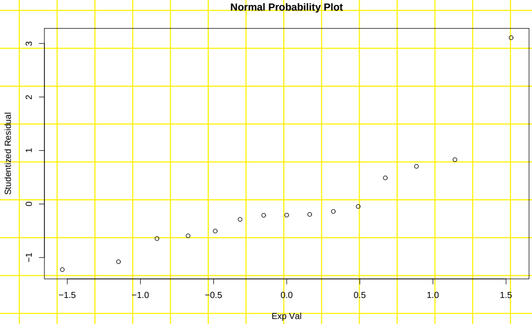
Residual vs. fitted value plot indicates



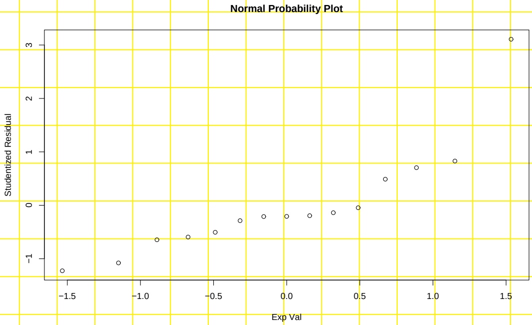
Residual vs. fitted value plot indicates unequal variances.



Normal Q-Q plot indicates



Normal Q-Q plot indicates Nonnormality.



Since is nearly constant across i,
we should use the .

i	Y^*_{i1}	Y^*_{i2}	Y^*_{i3}	Y^*_{i4}	Y^*_{i5}	$\overline{Y^*}_{i\cdot}$	$(s^*_i)^2$
1	1.484	4.612	2.671	3.853	4.445	3.413	1.742
2	2.109	4.396	1.995	2.509	0.476	2.297	1.974
3	4.665	3.521	4.368	5.837	3.792	4.437	0.818
	$\overline{Y^*}_{\cdot\cdot}$	=	3.38	;	MSE^*	=	1.511

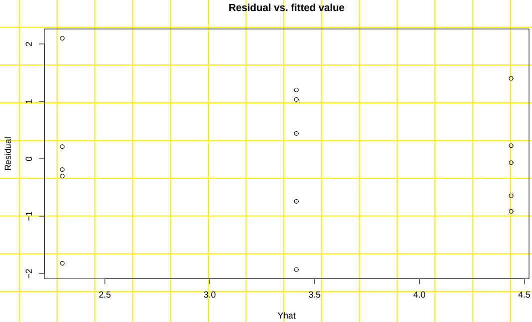
Since $\frac{s_i}{\overline{Y}_{i\cdot}}$ is nearly constant across i, we should use the log-transformation: $Y^* = \log(Y)$.

Table: log-transformed data

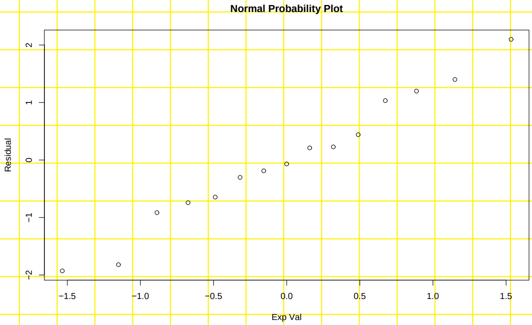
i	Y^*_{i1}	Y^*_{i2}	Y^*_{i3}	Y^*_{i4}	Y^*_{i5}	$\overline{Y^*}_{i\cdot}$	$(s_i^*)^2$
1	1.484	4.612	2.671	3.853	4.445	3.413	1.742
2	2.109	4.396	1.995	2.509	0.476	2.297	1.974
3	4.665	3.521	4.368	5.837	3.792	4.437	0.818
	$\overline{Y^*}_{\cdot\cdot}$	=	3.38	;	MSE^*	=	1.511

What would the Box-Cox procedure suggest?

Residual vs. fitted value plot of the log-transformed data: Nearly equal variance.



Normal Q-Q plot of the log-transformed data: Nearly linear.



ANOVA on the log-transformed data Y_{ij}^*.

- $MSTR^* = 5.726$, $MSE^* = 1.511$.
- F test for equal means: $F_{log}^* = \frac{5.726}{1.511} = 3.789$.
- $I = 3, n_T = 15$, $pvalue = P(F_{2,12} > 3.789) = 0.053$.
- Can not reject $H_0 : \mu_1^* = \mu_2^* = \mu_3^*$ at 0.05 significance level, but can reject H_0 at 0.1 significance level.
- This is often referred to as "nearly significant".

Rank Transformation and Rank F Test

Test $H_0 : \mu_1 = \cdots = \mu_I$ **without the normality assumption**.

- Assumptions:
  - $Y_{ij} = \mu_i + \epsilon_{ij}$.
  - ϵ_{ij} have the **same continuous distribution** which is centered at zero.
  - Consequently, the distributions of Y_{ij} are the same up to a location translation (by μ_i).
- **Rank transformation**: Get the rank R_{ij} for each observation Y_{ij}.
  - For example, for the data set $3, 4, 2, 5, 6, 4$, the ranks are $2, 3.5, 1, 5, 6, 3.5$.

Rank F-test.

- Apply ANOVA on the ranks R_{ij}.
- Derive the F ratio: $F_R^* = MSTR(R)/MSE(R)$, where,

$$MSTR(R) = \frac{\sum_{i=1}^{I} n_i(\overline{R}_{i\cdot} - \overline{R}_{\cdot\cdot})^2}{I-1},$$

$$MSE(R) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{n_i}(R_{ij} - \overline{R}_{i\cdot})^2}{n_T - I}.$$

- The null distribution of F^* is approximately F_{I-1,n_T-I} provided that n_i's are not too small.
  - If $F_R^* > F(1-\alpha; I-1, n_T - I)$, then reject $H_0 : \mu_1 = \cdots = \mu_I$ at significance level α.

Computer

Table: Ranks of the data

i	R_{i1}	R_{i2}	R_{i3}	R_{i4}	R_{i5}	$\overline{R}_{i\cdot}$	$s_i^2(R)$
1	2	13	6	9	12	8.4	20.3
2	4	11	3	5	1	4.8	14.2
3	14	7	10	15	8	10.8	12.7
	$\overline{R}_{\cdot\cdot}$	=	8	;	$MSE(R)$	=	15.7

- $MSTR(R) = 45.6,\ MSE(R) = 15.7.$
- $F_R^* = \frac{45.6}{15.7} = 2.90.$
- $pvalue = P(F_{2,12} > 2.90) = 0.094.$
- Reject H_0 at significance level 0.1, but can not reject H_0 at level 0.05.

- Under the log-transformation, the p-value is 0.053, which is than the p-value under the rank transformation (0.094).
- In practice, for small data sets, simple transformations such as logarithm transformation are the rank transformation since the ANOVA tests tend to have under these simple transformations.

- Under the log-transformation, the p-value is 0.053, which is more significant than the p-value under the rank transformation (0.094).
- In practice, for small data sets, simple transformations such as logarithm transformation are often preferred over the rank transformation since the ANOVA tests tend to have greater power under these simple transformations.