Stat 206: Linear Models

Lecture 16

Nov. 25, 2015

Difference Between Two Means

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

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

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

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

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)*.
Summary statistics and ANOVA table: $n_1 = n_2 = n_3 = n_4 = 10$ and $\overline{Y}_{1\cdot} = 43.14$, $\overline{Y}_{2\cdot} = 89.44$, $\overline{Y}_{3\cdot} = 67.95$, $\overline{Y}_{4\cdot} = 40.47$.

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.

Type I and Type II errors

In hypothesis testing, there are two types of errors.

- Type I error: Reject the null hypothesis when it is

  .

  - Type I error rate:

    $$P(\text{reject } H_0 | H_0 \text{ true})$$

  - When testing H_0 at a pre-specified significance level α, the type I error rate is controlled to be

    α.

- Type II error: Accept the null hypothesis when it is

  .

  - Type II error rate:

    $$P(\text{accept } H_0 | H_0 \text{ wrong}).$$

  - Type II error rate is usually controlled.
  - *Power*: The probability of rejecting H_0 when it is wrong:

    $$\text{Power} = P(\text{reject } H_0 | H_0 \text{ wrong})$$

    $$= \text{1-type II error rate.}$$

The power of a test is influenced by which factors?

Multiple Comparison

It refers to the situation where a family of statistical inferences are considered simultaneously, such as constructing a family of confidence intervals, or testing multiple hypotheses.

- Errors are to occur when multiple inferences are conducted.
  - If one conducts 100 hypotheses testing, each at the 0.05 significance level. Even when all 100 null hypotheses are true, on average, one would reject of them purely by chance.
  - If these tests are independent, then the probability of at least one wrong rejection is .
- We want to simultaneously control the probability of committing such errors.
  - Multiple hypothesis testing: Control the family-wise type-I error rate.
  - Simultaneous confidence region: Maintain a family-wise confidence level.

Family-wise Type-I Error Rate

- Let Θ denote the parameter space.
  - One-way ANOVA:

$$\Theta = \{(\mu_1, \cdots, \mu_I, \sigma^2) : \mu_1, \cdots, \mu_I \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+\}.$$

- Testing m null hypotheses $H_{i,0}$, $i = 1, \cdots, m$.
  - For example: $H_{1,0} : \mu_1 - \mu_2 = 0$, $H_{2,0} : \mu_3 - \mu_4 = 0$.
  - For $\theta \in \Theta$, let $I_0(\theta) = \{1 \le i \le m : H_{0,i} \text{ is true}\}$.
- Let $\phi = \{\phi_1, \cdots, \phi_m\}$ be the testing procedure, where ϕ_i is a test for $H_{0,i}$. Then

$$\begin{aligned} FWER_\theta(\phi) &= P_\theta(\text{at least one null hypothesis is falsely rejected}). \\ &= P_\theta(\cup_{i \in I_0(\theta)}\{H_{0,i} \text{ is rejected}\}). \end{aligned}$$

- For a given $\alpha \in (0, 1)$, the procedure ϕ is said to have FWER controlled at level α (in the **strong sense**) if

$$FWER(\phi) := \sup_{\theta \in \Theta} FWER_\theta(\phi) \le \alpha.$$

Family-wise Confidence Coefficient

- For I factor levels, there are $I(I-1)/2$ pairwise comparisons of the form $D_{ij} = \mu_i - \mu_j$ $(1 \leq i < j \leq I)$.
- 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:

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

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

- Note

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

- 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 multiplier (resulting in 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: $\mu_1 - \mu_2,\ \mu_1 - \mu_3,\ \mu_1 - \mu_4,\ \mu_2 - \mu_3,\ \mu_2 - \mu_4,\ \mu_3 - \mu_4.$
- $T = \frac{1}{\sqrt{2}} q(1 - \alpha; I, n_T - I) = \frac{1}{\sqrt{2}} q(0.95; 4, 36) = \frac{1}{\sqrt{2}} 3.81 = 2.7.$
- Note $T = 2.7$ is greater than the corresponding t-multiplier $t(0.975; 36) = 2.03.$
- 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 confidence intervals:

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

- 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 control FWER at level 0.05 for simultaneously testing:

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

Studentized Range Distribution: Definition

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

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

Tukey's Procedure: Derivation

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

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

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

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

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

Bonferroni's Procedure

Suppose we want to construct g C.Is simultaneously.

- Bonferroni procedure: Construct each C.I at level $1 - \alpha/g$. Then the familywise confidence coefficient is at least $1 - \alpha$.
- Example: 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

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

Bonferroni Inequality

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 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

$L = \sum_{i=1}^{I} c_i\mu_i$: c_i's are pre-specified constants satisfying $\sum_{i=1}^{I} c_i = 0$.

- 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

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:

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

- 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.

Scheffe's Procedure

There are infinitely many contrasts. How to control family-wise confidence coefficient or FWER if **all contrasts** are considered?

- Consider the family of all possible contrasts:

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

- *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 $S^2 = (I - 1)F(1 - \alpha; I - 1, n_T - I)$.

- 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 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.

Package Design

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

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

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

- 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 L. Consequently, Scheffe's C.I. 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

  .
- 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 smallest multiplier to construct the C.Is.

- If the family of interest consists of all pairwise comparisons, then Tukey's procedure is the best.
- If the family consists of some (but not all) of the pairwise comparisons, Bonferroni's procedure may or may not be better than Tukey's depending on the number of pairwise comparisons of interests.
- If the family consists of finite number of contrasts no larger than the number of factor level means, then Bonferroni's procedure is better than Scheffe's.
- Otherwise, Bonferroni's procedure may or may not be better than Scheffe's.

Rust Inhibitors

If all 6 pairwise comparisons are of interest, then at $\alpha = 0.05$:

- Tukey's multiplier:

$$T = \frac{1}{\sqrt{2}} q(1-\alpha; I, n_T - I) = \frac{1}{\sqrt{2}} q(0.95; 4, 36) = 2.7.$$

- Scheffe's multiplier:

$$\begin{aligned} S &= \sqrt{(I-1)F(1-\alpha; I-1, n_T - I)} \\ &= \sqrt{3 * F(0.95; 3, 36)} = 2.97. \end{aligned}$$

- Bonferroni's multiplier: note $g = 6$

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

- $T < B < S$: Tukey's procedure is the best.

If only four pairwise comparisons are of interest, then at $\alpha = 0.05$:

- T and S remain the same.
- Bonferroni's multiplier B decreases. Now $g = 4$:

$$B = t(1 - \frac{\alpha}{2g}; n_T - I) = t(1 - \frac{0.05}{8}; 36) = 2.63.$$

- $B < T < S$: Bonferroni's procedure is the best.
- Note for Bonferroni's procedure to be applicable, these pairwise comparisons need to be **pre-specified** before looking at the data. *Why?*

Alternative Formulations of One-way ANOVA

There are several alternative formulations of the one-way ANOVA model. In the `lm` function in `R`, the treatments are represented by $I-1$ indicator variables.

- Specify a reference treatment, e.g., treatment 1, and define

$$\mu = \mu_1.$$

- Define treatment contrasts:

$$\alpha_i = \mu_i - \mu_1 = \mu_i - \mu, \quad i = 1, \cdots, I.$$

  Note $\alpha_1 = 0$.

- The model equation can be re-written as :

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, \cdots, I, j = 1, \cdots, n_i, \qquad \textit{with}, \quad \alpha_1 = 0.$$

- Then $\mu_1 = \cdots = \mu_I \Longleftrightarrow \alpha_2 = \cdots = \alpha_I = 0$.
- `lm` summary output: $\mu (= \mu_1)$ corresponds to intercept, α_i corresponds to the coefficient associated with treatment i (for $i = 2, \cdots, I$); F test for regression relation $\Longleftrightarrow$ F test for equality of means.

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: Used to 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:

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

- Studentized residuals:

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

  where $s(e_{ij}) =$. *Why?*

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

Check Equal Variance

By residual vs. fitted value plots.

- When the design is approximately balanced: Plot residuals e_{ij}'s against the fitted values $\overline{Y}_{i.}$'s.
- When the design is very unbalanced: Plot the studentized residuals r_{ij}'s against the fitted values $\overline{Y}_{i.}$'s.
- Constancy of the error variance is supported by the residuals having similar extent of dispersion (around zero) across different treatment groups.

Check Normality

By Normal Q-Q plots of the residuals.

- When sample size is small: Use the combined studentized residuals across all treatment groups.
- When sample size is large: Draw separate plot for each treatment group.
- When there is evidence for unequal variances and combined residuals are used, then use "treatment-specific" studentized residuals:

$$\tilde{r}_{ij} := \frac{e_{ij}}{\sqrt{s_i^2 \frac{n_i - 1}{n_i}}}, \quad s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2.$$

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

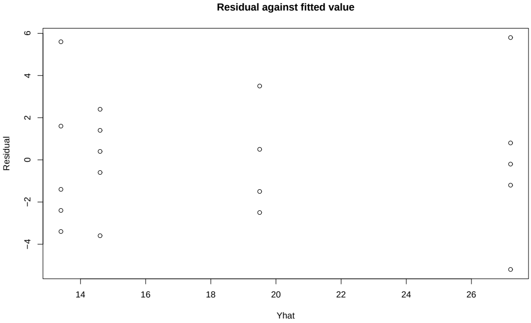
Other things that can be examined by residual plots:

- 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.

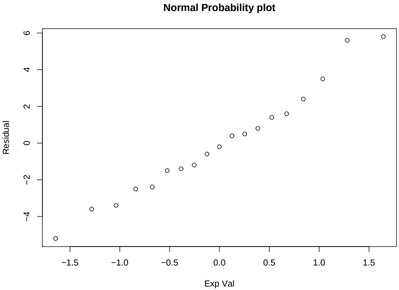
Package Design

Package Design (i)			Store	(j)			
i	e_{i1}	e_{i2}	e_{i3}	e_{i4}	e_{i5}	$\overline{Y}_{i.}$	n_i
1	-3.6	2.4	1.4	-0.6	0.4	14.6	5
2	-1.4	-3.4	1.6	5.6	-2.4	13.4	5
3	3.5	0.5	-1.5	-2.5	miss	19.5	4
4	-0.2	5.8	-5.2	-1.2	0.8	27.2	5

Residual versus fitted value plot: The 4 treatments show about the same extent of dispersion of the residuals around zero, thus the error variances are about equal across these treatments.



Normal Q-Q plot using combined residuals across all 4 treatments. The plot shows a strong linear pattern, thus normality assumption holds well.



Remedial Measures

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

- Transformation of the response variable to make its distribution closer to normal and to stabilize the variance.
  - Variance stabilizing transformation.
  - Box-cox procedure.
- If the departures are too extreme such that transformations do not work, then use *nonparametric tests*.
  - 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 $f(\cdot)$ such that:

- Factor level variances of the transformed data $Y_{ij}^* = f(Y_{ij})$ will be approximately equal.
- Often the distribution of the transformed data will also be closer to Normal.

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

- We want to find a transformation $Y^* = f(Y)$ such that variance of Y^* is a constant, say 1.
- 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

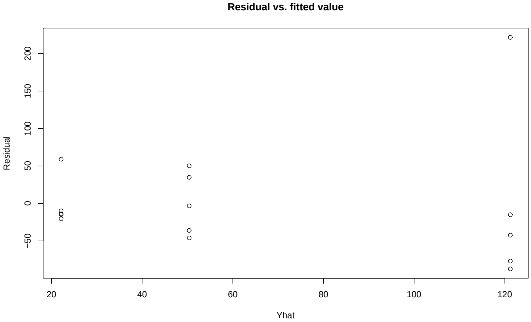
- If $\sigma_i^2 \propto \mu_i$, then use
- If $\sigma_i \propto \mu_i$, then use
- If $\sigma_i \propto \mu_i^2$, then use
- How to decide on which one to use?
  - Calculate $\frac{s_i^2}{\overline{Y}_{i\cdot}}$, $\frac{s_i}{\overline{Y}_{i\cdot}}$, $\frac{s_i}{(\overline{Y}_{i\cdot})^2}$ for $i = 1,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

Computers

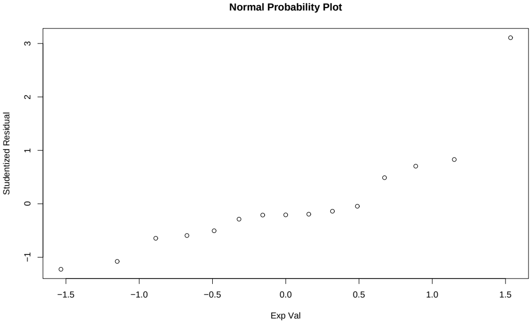
A company operates computers at three different locations. The computers are identical as to make and model, but are subject to different degrees of voltage fluctuation. The length of time between computer failures are recorded for three locations, each for five failure intervals. This is an observational 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 unequal variances.



Normal Q-Q plot indicates non-normality.

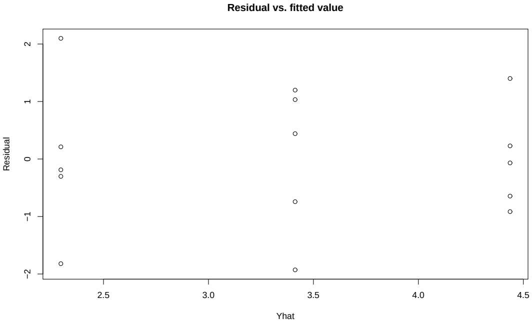


Since is nearly constant across i,
we should use the .

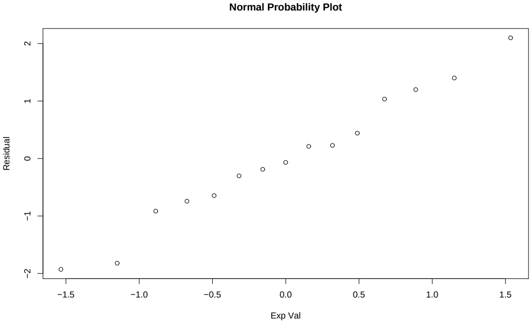
Table: log-transformed data

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

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 F Test

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

- Nonparametric procedures assume:
  - Model equation $Y_{ij} = \mu_i + \epsilon_{ij}$.
  - 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 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.