Introduction to Applied Statistics
STAT 5005
Lecture 6: Inferences about More Than Two Population Central Values (Chapter 8) / Multiple Comparisons (Chapter 9)

Jingyu Sun

Fall 2021

ANOVA for a Completely Randomized Design

Checking ANOVA Assumptions

Multiple Comparisons (Chapter 9)

ANOVA for a Completely Randomized Design

- In Lecture 5, we presented a method for testing the equality of two population means
- We wish to extend this method to test the equality of more than two population means

Example

- We wish to compare the mean hourly wage for farm laborers from three different classifications (union-documented, nonunion-documented, nonunion-undocumented)
- Independent random samples of farm laborers would be selected from each of the three classifications
- How do we determine the size of difference in the sample means necessary for us to state with some degree of certainty that the population means are different?
- The statistical procedure called **Analysis of Variance** (ANOVA) will provide us with the answer to this question

- **Table 8.1:** Data set 1

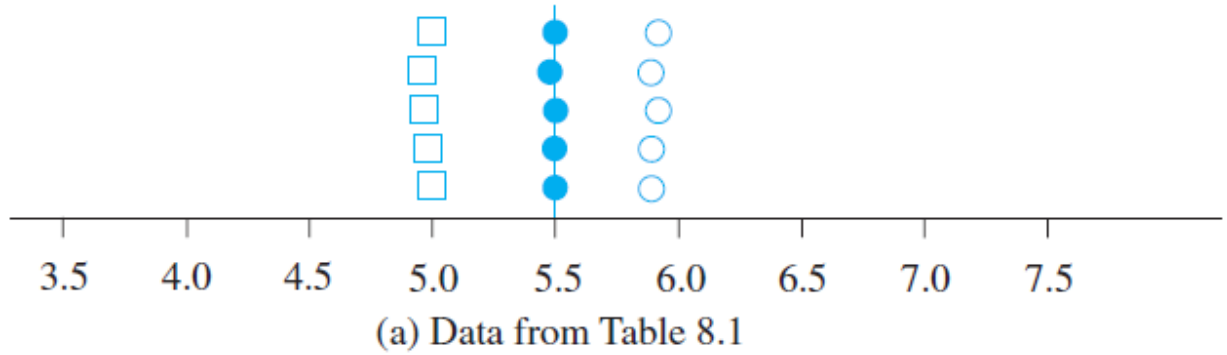
Sample 1	Sample 2	Sample 3
5.90	5.51	5.01
5.92	5.50	5.00
5.91	5.50	4.99
5.89	5.49	4.98
5.88	5.50	5.02

- **Table 8.2:** Data set 2

Sample 1	Sample 2	Sample 3
5.90	6.31	4.52
4.42	3.54	6.93
7.51	4.73	4.48
7.89	7.20	5.55
3.78	5.72	3.52

- Sample means for both data sets: $\bar{y}_{1\bullet} = 5.90, \bar{y}_{2\bullet} = 5.50, \bar{y}_{3\bullet} = 5.00$
- Draw the corresponding dot diagrams for both data sets
- Which data set presents sufficient evidence to indicate differences among the three population means?

Dot Diagrams



(a) Data from Table 8.1

(b) Data from Table 8.2

Completely Randomized Design

- The experimental setting in which a random sample of observations is taken from each of t different populations is called a **completely randomized design**
- Example: College students from five regions of the United States (northeast, southeast, midwest, southwest, and west-were) interviewed to determine their attitudes toward industrial pollution
- Each student selected was asked a set of questions related to the impact on economic development of proposed federal restrictions on air and water pollution. A total score reflecting each student's responses was then produced.
- 250 students are randomly selected in each of the five regions
- We want to know if there are significant differences among the mean scores for the five regions

	Population				
	I	II	III	IV	V
Sample mean	$\overline{y}_1$	$\overline{y}_2$	$\overline{y}_3$	$\overline{y}_4$	$\overline{y}_5$
Sample variance	s_1^2	s_2^2	s_3^2	s_4^2	s_5^2
Sample size	250	250	250	250	250

- The analysis of variance (ANOVA) procedures are developed under the following conditions:

1. Each of the five populations has a normal distribution
2. The variances of the five populations are equal: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma^2$
3. The five sets of measurements are independent random samples from their respective populations

- What would be an estimate for the variance between samples?
- What would be a sensible estimate for the common variance σ_2^2?

Notation Needed for the ANOVA of a Completely Randomized Design

Suppose there are t different populations

- y_{ij}: The jth sample observation selected from population i
- n_i: The number of sample observations selected from population i
- n_T: The total sample size $n_T = \sum_{i=1}^{t} n_i$
- $\bar{y}_{i\bullet}$: The average of the n_i sample observations drawn from population i, $\bar{y}_{i\bullet} = \sum_{j=1}^{n_i} y_{ij}/n_i$
- $\bar{y}_{\bullet\bullet}$: The average of all sample observations, also called **overall mean**, $\bar{y}_{\bullet\bullet} = \sum_{i=1}^{t} \sum_{j=1}^{n_i} y_{ij}/n_T$
- Example: Give the values of $t, n_i, n_T, \bar{y}_{i\bullet}$ for the pollution perception survey

Sums of Squares

- Let s_T^2 be the sample variance of the n_T. The **total sum of squares** (TSS) of the measurements about the overall mean is

$$TSS = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\bullet\bullet})^2 = (n_T - 1)s_T^2$$

- A measure of the within-sample variability is given by the **within-sample sum of squares** (SSW), also called error sum of squares

$$SSW = \sum_{i=1}^{t} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\bullet})^2 = \sum_{i=1}^{t} (n_i - 1)s_i^2$$

- A measure of the between-sample variability is given by the **between-sample sum of squares** (SSB), also called treatment sum of squares

$$SSB = \sum_{i=1}^{t} n_i (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$$

Partition and Mean Squares

- It is possible to partition the total sum of squares as follows:

$$TSS = SSW + SSB$$

- An estimate of the common variance is the **mean square within samples**, also called mean square error is

$$s_W^2 = \frac{SSW}{n_T - t}$$

- An estimate of the variance between samples is the **mean square between samples**, also called treatment mean square is

$$s_B^2 = \frac{SSB}{t - 1}$$

Hypothesis Testing for ANOVA

- An analysis of variance for a completely randomized design with t populations has the following null and alternative hypotheses:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_t$$

$$H_a : \text{at least one of the } \mu_i\text{'s differs from the rest}$$

- The test statistic is $F = s_B^2/s_W^2$
- For a fixed level of significance α, reject H_0 if $F > F_\alpha$ for an F distribution with $df_1 = t - 1$ and $df_2 = n_T - t$ degrees of freedom

ANOVA Table

We summarize the results of a study in an ANOVA table

Source	Deg. of Freedom	Sum of Sq.	Mean Sq.	F
Between samples	$t-1$	SSB	$SSB/(t-1)$	s_B^2/s_W^2
Within samples	n_T-t	SSW	$SSW/(n_T-t)$	
Total	n_T-1	TSS		

Example

An educator wishes to conduct a study to find out whether the difficulty levels of different majors are the same. For such a study, a random sample of major grade point averages (GPA) of 11 graduating seniors at a large university is selected for each of the four majors mathematics, English, education, and biology.

- Mathematics: sample mean 2.90, sample variance 0.188
- English: sample mean 3.34, sample variance 0.148
- Education: sample mean 3.36, sample variance 0.229
- Biology: sample mean 3.02, sample variance 0.157

Test, at the 5% level of significance, whether the data contain sufficient evidence to conclude that there are differences among the average major GPAs of these four majors

Example

The National Transportation Safety Board (NTSB) wants to examine the safety of compact cars, midsize cars, and full-size cars. It collects a sample of three for each of the car types. The data provided below give the pressure applied to the driver's head during a crash test for each type of car.

Compact	Midsize	Full-size
643	469	484
655	427	456
702	525	402
573		512

Test, at the 1% level of significance, whether the data contain sufficient evidence to conclude that there are differences among the mean pressures for the three car classifications

Performing ANOVA using R

```r
# Download the file "cars.csv" and load data into R
df.car <- read.csv("cars.csv")
df.car$carsize <- as.factor(df.car$carsize) # convert to factor

fit <- aov(pressure ~ carsize , data=df.car)
summary(fit)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## carsize       2  78482   39241   15.72 0.00169 **
## Residuals     8  19970    2496
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Model for Observations in a Completely Randomized Design

We will impose the following conditions:

1. The samples are independent random samples. Results from one sample in no way affect the measurements observed in another sample
2. Each sample is selected from a normal population
3. The mean and variance for population i are, respectively, μ_i and σ_i^2 $(i = 1, \ldots, t)$. The t variances are equal $\sigma_1^2 = \ldots = \sigma_t^2 = \sigma^2$

A model (equation) that encompasses these three assumptions is

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

where

- μ is called the **overall mean**
- τ_i is the **effect** due to population i
- ϵ_{ij}'s are the **error terms**

Note that here *error* does not mean mistake!

- The null hypothesis for a one-way analysis of variance is that $H_0 : \mu_1 = \ldots = \mu_t$
- Using our model, this would be equivalent to the null hypothesis is $H_0 : \tau_1 = \ldots = \tau_t = 0$
- The corresponding alternative hypothesis is H_a : at least one of the τ_i's differs from 0

Checking ANOVA Assumptions

Checking ANOVA Assumptions

- In ANOVA, our inferences are based on the assumptions that the data follow the model

$$y_{ij} = \mu + \tau_i + \epsilon_{ij} \, ,$$

  where

  - the τ_i's are fixed but unknown
  - the ϵ_{ij}'s are (1) independent and (2) normally distributed with (3) constant variance

- Accuracy of inference depends on assumptions being true
- Assuming that (1) is true, we need to check Assumptions (2) and (3) on the ϵ_{ij}'s
- Our assessments of assumptions about the errors are based on residuals $r_{ij} = y_{ij} - \bar{y}_{i\bullet}$

Assessing Nonnormality

- The normal quantile-quantile plot, or **normal Q-Q plot**, is a graphical procedure for assessing normality
- The theoretical p-th percentile of any distribution is the value such that $p\%$ of the measurements fall below the value
- The sample p-th percentile of any data set is the value such that $p\%$ of the measurements fall below the value
- If the data follow a normal distribution, then a plot of the theoretical percentiles of the normal distribution versus the observed sample percentiles should be approximately **linear**
- Since we are concerned about the normality of the error terms, we create a normal probability plot of the residuals

Example

- 60 participants are selected to taste 3 cheeses
- Using a CRD, the participants are divided into 3 groups at random
- They are asked to give a rating on a scale of 1 to 10
- To view the dataset, download the file `cheese.csv` in your working directory and type

```
df.cheese <- read.csv("cheese.csv")
df.cheese$type <- as.factor(df.cheese$type)
```

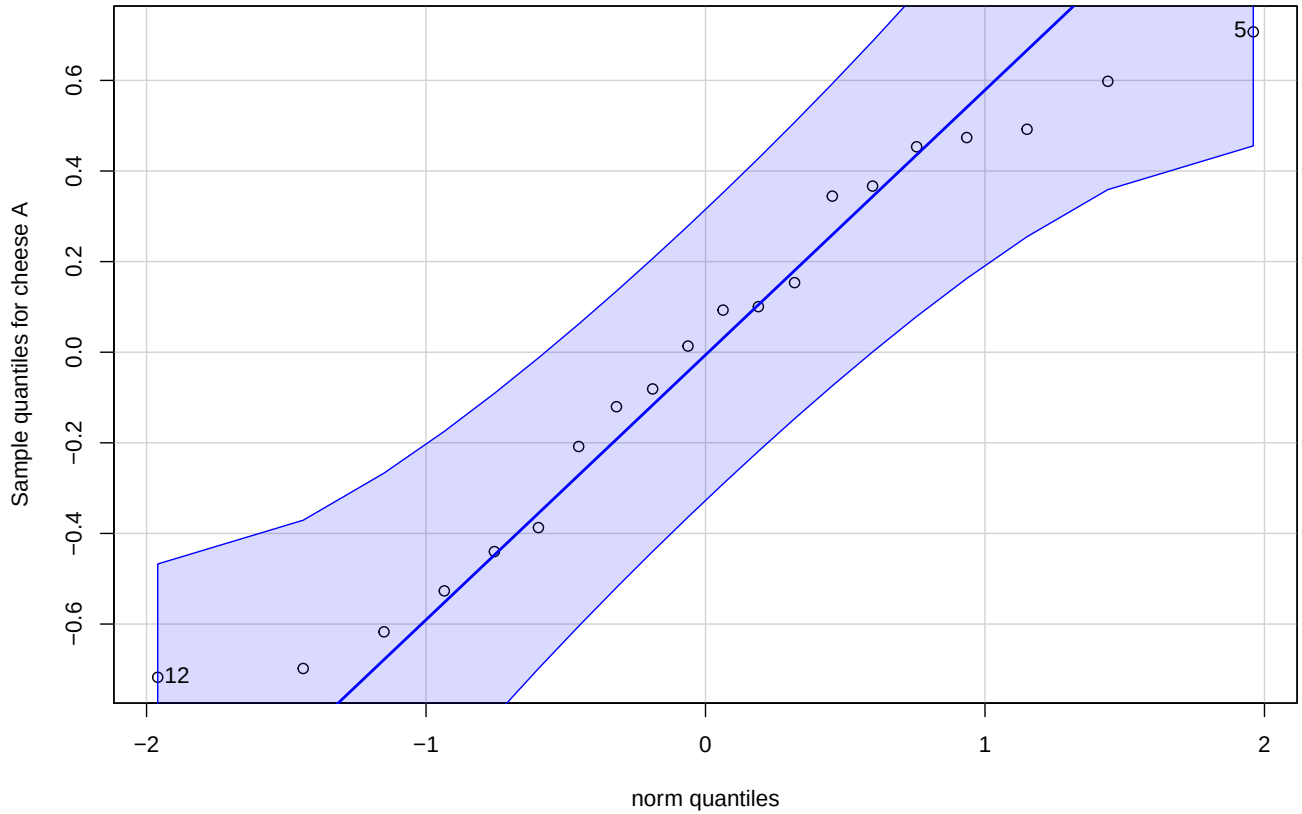
- Run ANOVA:

```
fit <- aov(rating ~ type, data = df.cheese)
```

- To obtain the residuals $r_{ij} = y_{ij} - \bar{y}_{i\bullet}$, type

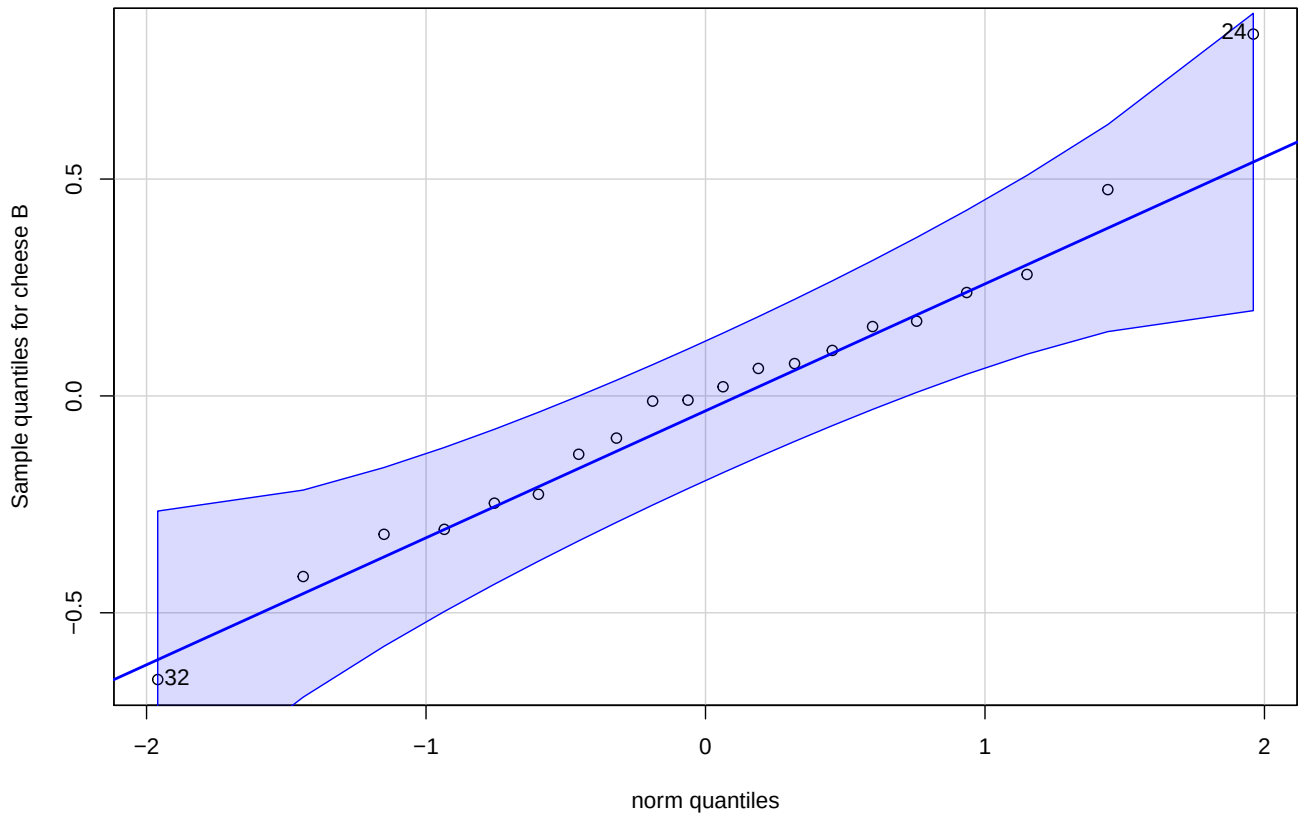
```
rsd <- fit$residuals
```

```r
library(car)
qqPlot(rsd[df.cheese$type == "A"],
       ylab="Sample quantiles for cheese A")
```



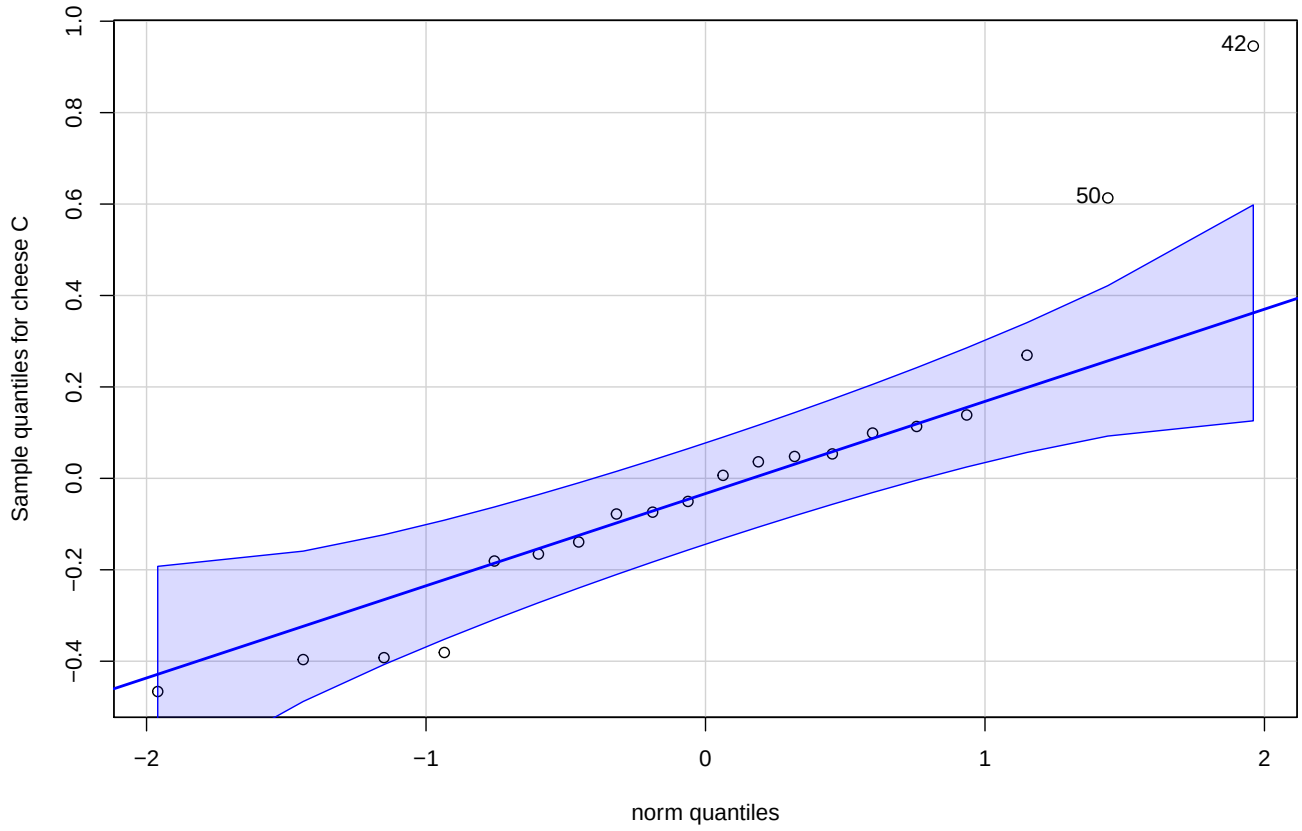

```
[1] 12  5
```

```
qqPlot(rsd[df.cheese$type == "B"],
       ylab="Sample quantiles for cheese B")
```



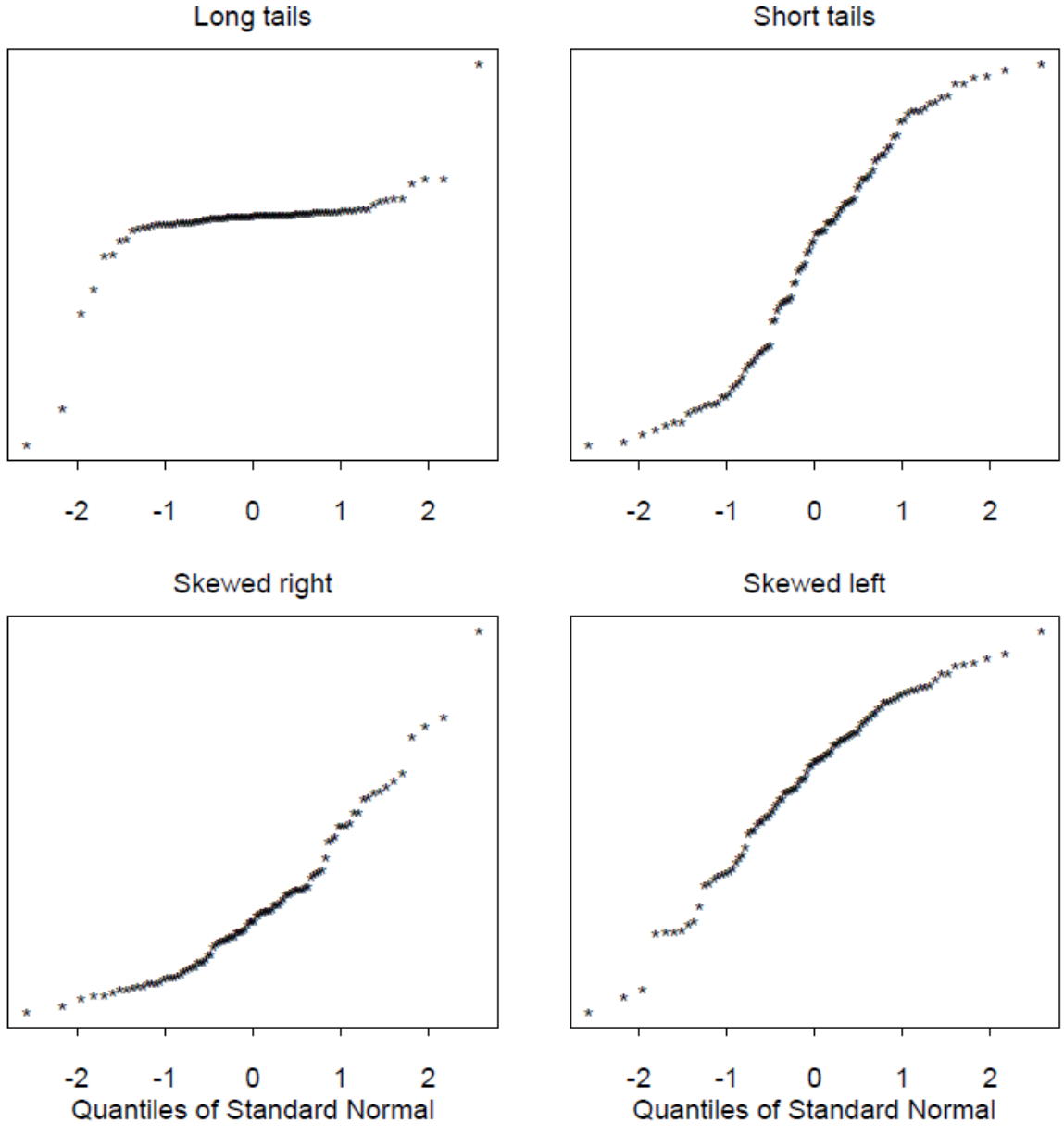

```
24 32
 4 12
```

```
qqPlot(rsd[df.cheese$type == "C"],
       ylab="Sample quantiles for cheese C")
```




```
42 50
 2 10
```

Examples of Nonnormal Residuals

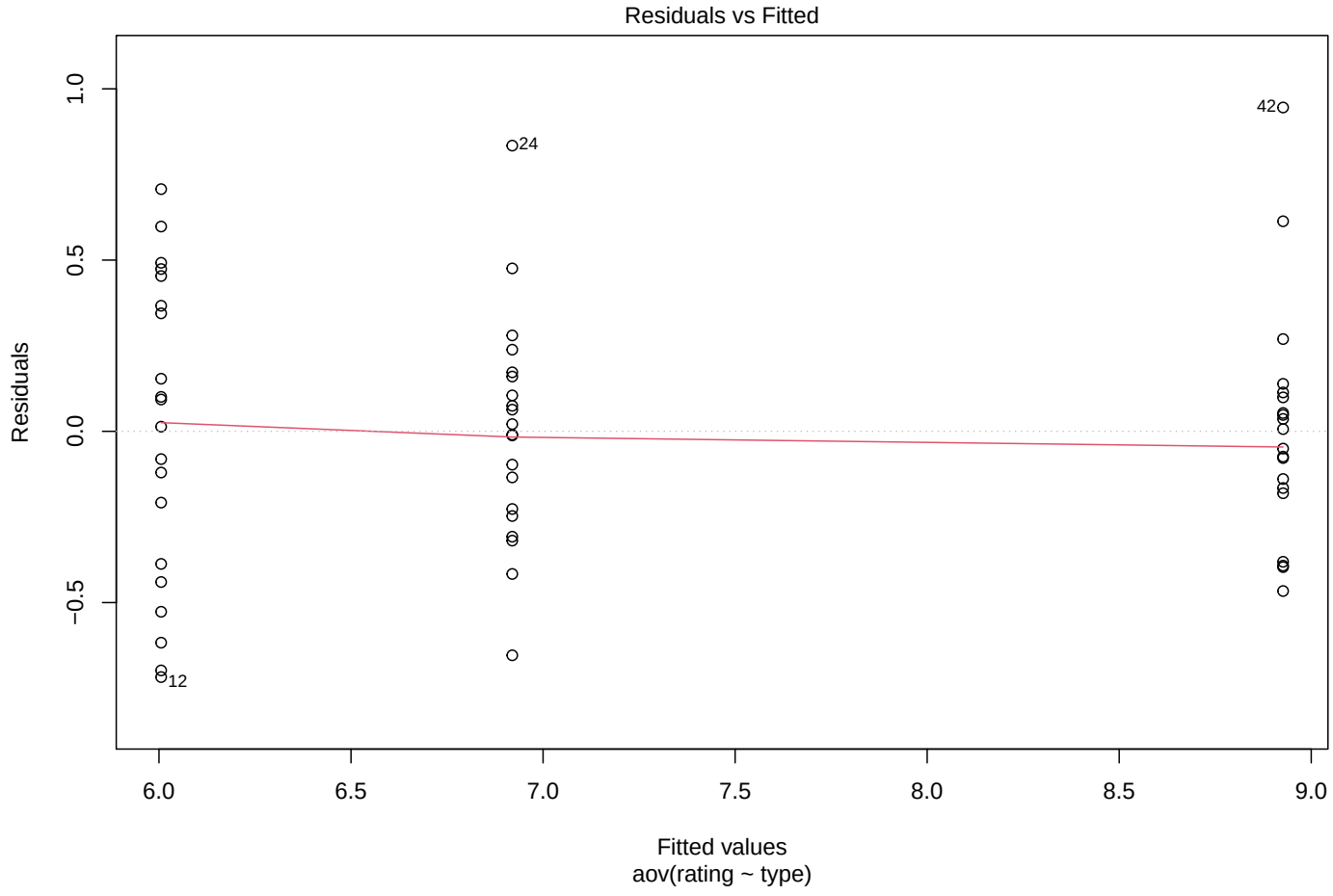


Assessing Nonconstant Variance

- We will look for nonconstant variance that occurs when the responses within a treatment group all have the same variance σ_i^2, but the variances differ between groups
- We assess nonconstant variance by making a plot of the residuals r_{ij} against the fitted values $\bar{y}_{i\bullet}$
- If the variance is constant, the vertical spread in the stripes will be about the same

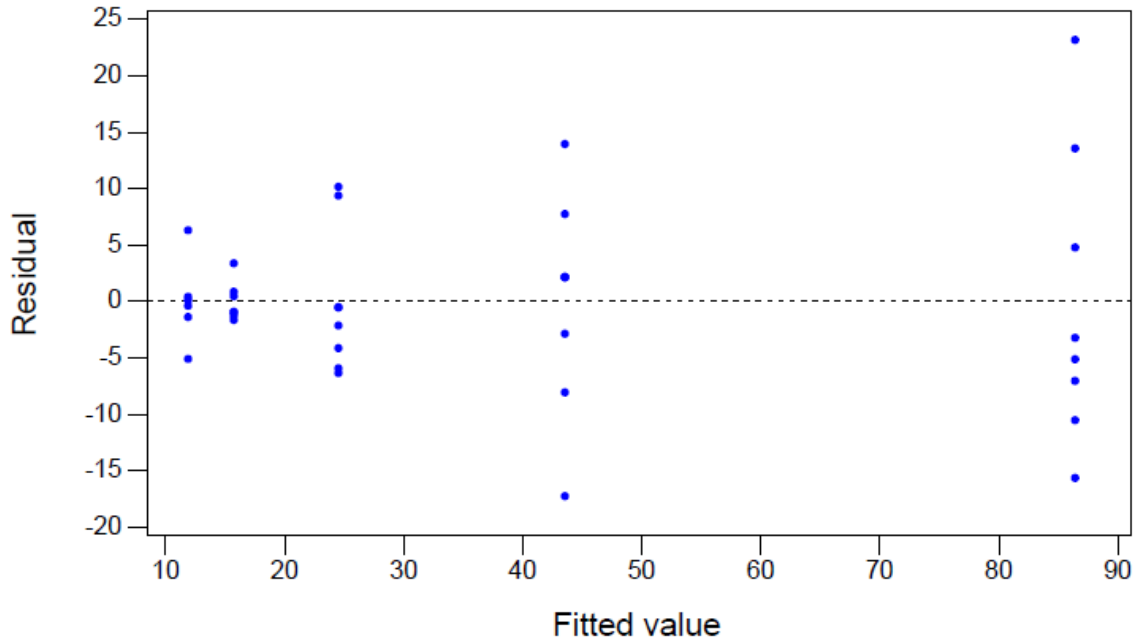
Cheese Example

```
plot(fit, 1)
```



Examples of Nonconstant Variance

- Nonconstant variance is revealed as a pattern in the spread of the residuals
- The most common deviations from constant variance are those where the residual variation depends on the mean



Fixing Problems

When our assessments indicate that our data do not meet our assumptions, we must

- either modify the data so that they do meet the assumptions,
- or modify our methods so that the assumptions are less important

Accommodating Nonnormality and Nonconstant Variance

- Nonnormality, particularly asymmetry, and nonconstant error variances can sometimes be lessened by transforming the response to a different scale (e.g. a square root, logarithm, or other transformation to a certain power)

Multiple Comparisons (Chapter 9)

Contrasts

- Assume we have 6 treatments
- An ANOVA procedure indicates whether to reject the null hypothesis that all the 6 treatment groups have the same mean response
- But rejection of the null hypothesis does not tell us which treatments are different or in what ways they differ
- For example, does μ_1 differ from μ_2 or μ_3? Does μ_3 differ from the average of μ_2, μ_4, and μ_5?
- **Contrasts** have been developed to answer questions such as these

- Consider a completely randomized design where we wish to make comparisons among the t population means $\mu_1, \ldots, \mu_t$
- Formally, a contrast is a **linear combination** of treatment means

$$C = \sum_{i=1}^{t} a_i \mu_i$$

where the coefficients a_i add to 0

- Example 1: Write a contrast to compare the mean for population 1 and the mean for population 2
- Example 2: Write a contrast to compare the mean for population 1 to the average of the means for populations 2 and 3
- Rescaling contrasts provides the same information
- We often write the contrasts with all the a_i's as integer values

- Note (prove) that $C = \sum_{i=1}^{t} a_i \tau_i$ where the τ_i is the effect due to population i
- An estimate of C, designated by $\hat{C}$, is formed by replacing the μ_i's in C with their corresponding sample means $\bar{y}_{i\bullet}$

$$\hat{C} = \sum_{i=1}^{t} a_i \bar{y}_{i\bullet}$$

Several kinds of contrasts

- The most common contrasts are pairwise comparisons, where we contrast the mean response in one treatment with the mean response in a second treatment
- How many pairwise contrasts are there for t treatments?

Several kinds of contrasts

- The most common contrasts are pairwise comparisons, where we contrast the mean response in one treatment with the mean response in a second treatment
- How many pairwise contrasts are there for t treatments?
- A second classic example of contrasts occurs in an experiment with a control and two or more new treatments
- Suppose that treatment 1 is a control, and treatments 2 and 3 are new treatments. On average, do the new treatments have the same response as the control? Propose a relevant contrast.

Orthogonal Contrasts

- Two contrasts C and C' with respective lists of coefficients $(a_1, \ldots, a_t)$ and $(a'_1, \ldots, a'_t)$ are said to be orthogonal if

$$\sum_{i=1}^{t} a_i a'_i / n_i = 0$$

- Consider a completely randomized design for comparing 4 populations means, with sample sizes $n_1 = 5$, $n_2 = 4$, $n_3 = 6$, and $n_4 = 5$. Are the following contrasts orthogonal?
  $C = \mu_1 - \mu_3$ and $C' = \mu_1 + \mu_2 + \mu_3 - 3\mu_4$
- Reconsider the question if the sample sizes were all equal

- If two contrasts are orthogonal, then one contrast conveys no information about the other contrast
- The random error of one contrast is not correlated with the random error of an orthogonal contrast
- If there are t treatments, you can find a set of $t-1$ contrasts that are pairwise orthogonal
- The sum of squares for a contrast C is

$$SS_C = \frac{\hat{C}^2}{\sum_{i=1}^{t} \frac{a_i^2}{n_i}}$$

- It indicates the amount of variation in the treatment means that can be explained by that particular contrast

- Orthogonal contrasts partition the between-groups sum of squares: If $\hat{C}_1, \ldots, \hat{C}_{t-1}$ are a full set of pairwise orthogonal contrasts, then

$$SSB = \sum_{i=1}^{t-1} SS_{C_i}$$

- Thus, we can take the $t-1$ degrees of freedom associated with the treatment sum of squares that describe any differences among the treatment means and break them into $t-1$ independent explanations of how the treatment means may differ

Example

- Researchers conducted a study on the effectiveness of various agents to control weeds in crops. The study consisted of a control (no agent), two biological agents (Bio1 and Bio2), and two chemical agents (Chm1 and Chm2)
- Thirty 1-acre plots of land were planted with hay. Six plots were randomly assigned to receive one of the five treatments
- The hay was harvested and the total yield in tons per acre was recorded

Agent	1	2	3	4	5
Type	None	Bio1	Bio2	Chm1	Chm2
$\bar{y}_{i\bullet}$	1.175	1.293	1.328	1.415	1.500
s_i	.1204	.1269	.1196	.1249	.1265

- Q1: Compute the ANOVA table. Is there any significant difference in the mean efficacy of the five agents? Use the p-value approach at 1% level of significance.

Source	df	SS	MS	F	p-value
Treatment	4	.3648	.0912	5.96	.0016
Error	25	.3825	.0153		
Total	29	.7472			

- Q2: Determine four orthogonal contrasts and verify that the four contrasts are pairwise orthogonal.

```r
sample_mean <- c(1.175, 1.293, 1.328, 1.415, 1.500)
sample_sd <- c(.1204, .1269, .1196, .1249, .1265)

ss_trt <- 6*sum((sample_mean-mean(sample_mean))**2)

w_ctrl_agents <- c(4,-1,-1,-1,-1)
w_bio_chem <-  c(0,1,1,-1,-1)
w_bio1_bio2 <- c(0,1,-1,0,0)
w_chem1_chem2 <- c(0,0,0,1,-1)
C_ctrl_agents <- sample_mean%*%w_ctrl_agents
C_bio_chem <- sample_mean%*%w_bio_chem
C_bio1_bio2 <- sample_mean%*%w_bio1_bio2
C_chem1_chem2 <- sample_mean%*%w_chem1_chem2
#ss_ctrl_agents <- (6*C_ctrl_agents**2)
#/sum(w_ctrl_agents**2)

ss_ctrl_agents <- 0.2097
ss_bio_chem <- .1297
ss_bio1_bio2 <- .0037
ss_chem1_chem2 <- .0217
```

Confidence Intervals for Contrasts

- A $100(1-\alpha)\%$ confidence interval for a contrast $C = \sum_{i=1}^{t} a_i \mu_i$ is

$$\hat{C} \pm t_{\alpha/2} \sqrt{s_W^2} \sqrt{\sum_{i=1}^{t} \frac{a_i^2}{n_i}}$$

where $t_{\alpha/2}$ is the value of of a t-distribution with $n_T - t$ degrees of freedom with right-tail area $\alpha/2$

Hypothesis Testing for Contrasts

- Suppose that we want to test of the null hypothesis: $H_0 : C = \sum_{i=1}^{g} a_i \mu_i = \delta$ (often, $\delta = 0$). We can do a t-test by computing the test statistic

$$t = \frac{\hat{C} - \delta}{\sqrt{s_W^2}\sqrt{\sum_{i=1}^{t} \frac{a_i^2}{n_i}}}$$

- Under H_0, this statistic follows a t-distribution with $n_T - t$ degrees of freedom

Example

- Is there a significant difference between the control treatment and the four active agents for weed control with respect to their effect on average hay production?
- Test each of the four contrasts for significance. Use a level of significance $\alpha = 0.05$

Which Error Rate is Controlled?

- An experimenter wishes to compare t population (treatment) means using m contrasts.
- Each of the m contrasts can be tested using the t-test we introduced in the previous section
- Suppose each of the contrasts is tested with the same value of α, which we will denote as α_I, called the **individual comparisons** Type I error rate
- The probability of falsely rejecting at least one of the m null hypotheses is called the **experimentwise or familywise** Type I error rate (abbrev. FWER) and denoted by α_E
- If the m tests are independent, what is the value of α_E?

- In practice, the tests will not be independent, but the following upper bound exists for the experimentwise error rate

$$\alpha_E \leq 1 - (1 - \alpha_I)^m$$

m, Number of Contrasts	α_I Probability of a Type I Error on an Individual Test		
	.10	.05	.01
1	.100	.050	.010
2	.190	.097	.020
3	.271	.143	.030
4	.344	.185	.039
5	.410	.226	.049
.	.	.	.
.	.	.	.
.	.	.	.
10	.651	.401	.096

- What value of α_I must we use on 8 tests to achieve an experimentwise error rate of $\alpha_E = 0.05$?

Different Approaches To Control Type I Errors

- We will present several techniques to control the FWER
- There is a trade-off between Type I error and Type II error (type II error describes the error that occurs when one fails to reject a null hypothesis that is actually false.)
- As we go to more and more stringent Type I error rates, we become more confident in the rejections that we do make, but it also becomes more difficult to make rejections

The Bonferroni Technique

- When each of the m tests have the same individual error rates α_I, the **Bonferroni inequality** yields, for the experimentwise error rate α_E that

$$\alpha_E \leq m\alpha_I$$

- If we wanted to guarantee that the chance of a Type I error was at most α, we could select

$$\alpha_I = \frac{\alpha}{m}$$

- This procedure may be very conservative with respect to the experimentwise error rate, and hence an inflated probability of Type II error may result

Pairwise Comparisons

- A pairwise comparison is a contrast that examines the difference between two treatment means $\bar{y}_{i\bullet} - \bar{y}_{j\bullet}$
- Fisher's Least Significant Difference and Tukey's procedures have been developed for pairwise comparisons among t population means

Fisher's Least Significant Difference Procedure

1. Perform an analysis of variance to test $H_0 : \mu_1 = \mu_2 = \ldots = \mu_t$ against the alternative hypothesis that at least one of the means differs from the rest
2. If there is insufficient evidence to reject H_0, proceed no further
3. If H_0 is rejected, define the least significant difference (LSD) to be the observed difference between two sample means necessary to declare the corresponding population means different. For a specified level of significance α, the least significant difference for comparing μ_i to μ_j is

$$LSD_{ij} = t_{\alpha/2}\sqrt{s_W^2\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

   where $t_{\alpha/2}$ is the value of a t-distribution with $n_T - t$ degrees of freedom $\alpha/2$ with right-tail area $\alpha/2$
4. Then compare all pairs of sample means. If $|\bar{y}_{i\bullet} - \bar{y}_{j\bullet}| \geq LSD_{ij}$ declare the corresponding population means μ_i and μ_j different

- The Fisher's LSD test is basically a set of individual two-sample t-tests
- If you choose to use the Fisher's LSD test, you'll need to account for multiple comparisons when you interpret the data, since the computations themselves do not correct for multiple comparisons
- The only difference between a set of two-sample t-tests and the Fisher's LSD test, is that t tests compute the pooled sample standard deviation from only the two groups being compared, while the Fisher's LSD test computes the pooled sample standard deviation from all the groups

- In the weed agent example, test the following comparisons using the Fisher's LSD procedure at level 0.05:

1. Control vs Biological 1
2. Chemical 1 vs Chemical 2

The Studentized Range

- The **Studentized range** for a pair of treatment means is

$$\frac{\text{largest } \bar{y}_{i\bullet} - \text{smallest } \bar{y}_{i\bullet}}{\sqrt{s_W^2/n}}$$

assuming that all n_i's are the same and equal to some given n

- If all the treatments have the same mean (H_0), then the Studentized range statistic follows a **Studentized range distribution**
- Large Studentized values of the Studentized range are less likely under H_0

Tukey's Procedure

- Two population means μ_i and μ_j are declared different if

$$|\bar{y}_{i\bullet} - \bar{y}_{j\bullet}| \geq \frac{q_\alpha(t, n_T - t)}{\sqrt{2}} \sqrt{s_W^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

  where $q_\alpha(t, n_T - t)$ is the upper-tail value of area α of the Studentized range distribution with t and $n_T - t$ degrees of freedom

- Example: R command for the 5% quantile of the Studentized distribution with 5 and 20 degrees of freedom

`qtukey(0.95, 5, 20)` $= 4.2318567$

- This procedure controls the FWER at level α
- Tukey's procedure is more conservative than Fisher's LSD

- In the weed agent example, test the following comparisons using the Tukey's procedure at level 0.05:

1. Control vs Biological 1
2. Chemical 1 vs Chemical 2

The Scheffé Method

- It is tempting to analyze only those comparisons that appear to be interesting *after* seeing the sample data. This practice has sometimes been called **data dredging** or **data snooping**
- The Scheffé method (1954) is a multiple comparisons technique for **all** possible contrasts
- Should be used if you have not planned contrasts in advance
- Appropriate for assessing contrasts suggested by the data (data snooping)
- Most conservative (least powerful) of all tests

- Suppose that we are testing the null hypothesis $H_0 : C = \sum_{i=1}^{t} a_i \mu_i = 0$ against a two-sided alternative
- Test statistic: $|\hat{C}| = |\sum_{i=1}^{t} a_i \bar{y}_{i\bullet}|$
- Compute

$$S = \sqrt{s_W^2 \sum_{i=1}^{t} \frac{a_i^2}{n_i}} \sqrt{(t-1)F_\alpha}$$

  where F_α is the upper-tail value of area α of an F-distribution with $t-1$ and $n_T - t$ degrees
- For a specified level α, reject H_0 if

$$|\hat{C}| > S$$

- This procedure controls the FWER at level α

- In the weed agent example, use Scheffé's method to determine which if any of the four contrasts are significantly different from zero
- We have $\hat{C}_{\text{control vs rest}} = -0.836$, $\hat{C}_{\text{bio vs chem}} = -0.294$, $\hat{C}_{\text{bio1 vs bio2}} = -0.035$, $\hat{C}_{\text{chem1 vs chem2}} = -0.085$

Procedure Usage Summary

- Use Bonferroni technique when only interested in a small number of planned contrasts (or pairwise comparisons)
- Use Tukey's procedure when only interested in all (or most) pairwise comparisons of means
- Use Scheffé's method when doing anything that could be considered data snooping - i.e. for any unplanned contrasts
- Always keep in mind trade-off between Type I error and power