PLOTS FOR THE DESIGN AND ANALYSIS OF EXPERIMENTS

by

Jenna Christina Haenggi

Bachelor of Science in Mathematics, Southern Illinois University, 2007

A Research Paper
Submitted in Partial Fulfillment of the Requirements for the
Master of Science Degree

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
July 1, 2009

ACKNOWLEDGMENTS

My sincere thanks goes to Dr. Olive for his invaluable assistance leading to the writing of this paper. Many of the examples and definitions have been quoted directly from his book. I would also like to thank the authors of the books and research papers that were used in the writing of this paper, and also a thanks to the members of my graduate committee, Dr. Hughes and Dr. Pericak-Spector. Finally, I would like to extend a special thanks to my friends and family for their patience, love, and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PLOTS AND TESTS FOR EXPERIMENTAL DESIGNS

1.1 INTRODUCTION

This chapter follows Olive (2009, Ch.5) closely.

Definition 1.1. Models in which the response variable Y is quantitative, but all of the predictor variables are qualitative are called **analysis of variance** (ANOVA) models or **experimental design** models. Such models are the focus of **design of experiments** (DOE). Each combination of the levels of the predictors gives a different distribution for Y. A predictor variable W is often called a factor and a factor level a_i is one of the categories W can take.

Definition 1.2. A **lurking variable** is not one of the variables in the study, but may affect the relationships among the variables in the study. A **unit** is the experimental material assigned **treatments**, which are the conditions the investigator wants to study.

Definition 1.3. In an **experiment**, the investigators use **randomization** to assign treatments to units. To assign p treatments to $n = n_1 + \cdots + n_p$ experimental units, draw a random permutation of $\{1, ..., n\}$. Assign the first n_1 units treatment 1, the next n_2 units treatment $2, ...,$ and the final n_p units treatment p.

Randomization allows one to do valid inference such as F tests of hypotheses and confidence intervals. Randomization also washes out the effects of lurking variables and makes the p treatment groups similar except for the treatment. The effects of lurking variables are present in observational studies defined in Definition 1.4.

Definition 1.4. In an **observational study**, investigators simply observe the response, and the treatment groups need to be p random samples from p populations (the levels) for valid inference.

Example 1.1. Consider using randomization to assign the following nine people (units) to three treatment groups.

Carroll, Collin, Crawford, Halverson, Lawes,

Stach, Wayman, Wenslow, Xumong

Balanced designs have the group sizes the same: $n_i = n/p$. Label the units alphabetically so Carroll gets 1, ..., Xumong gets 9. The *R/Splus* function `sample` can be used to draw a random permutation. Then the first 3 numbers in the permutation correspond to group 1, the next 3 to group 2 and the final 3 to group 3. Using the output shown below, gives the following 3 groups.

group 1: Stach, Wayman, Xumong

group 2: Lawes, Carroll, Halverson

group 3: Collin, Wenslow, Crawford

```
> sample(9)
[1] 6 7 9 5 1 4 2 8 3
```

Often there is a table or computer file of units and related measurements, and it is desired to add the unit's group to the end of the table. The *regpack* function `rand` reports a random permutation and the quantity groups[i] = treatment group for the ith person on the list. Since persons 6, 7 and 9 are in group 1, groups[7] = 1. Since Carroll is person 1 and is in group 2, groups[1] = 2, et cetera.

```
> rand(9,3)
$perm
[1] 6 7 9 5 1 4 2 8 3
$groups
[1] 2 3 3 2 2 1 1 3 1
```

Definition 1.5. Replication means the response variables $Y_{i,1}, ..., Y_{i,n_i}$ are approximately independently and identically distributed (iid) random variables.

Example 1.2. a) If ten students work two types of paper mazes three times each, then there are 60 measurements that are not replicates. Each student should work the six mazes in random order since speed increases with practice. For the ith student, let Z_{i1} be the average time to complete the three mazes of type 1, let Z_{i2} be the average time for mazes of type 2 and let $D_i = Z_{i1} - Z_{i2}$. Then $D_1, ..., D_{10}$ are replicates.

b) Cobb (1998, p. 126) states that a student wanted to know if the shapes of sponge cells depends on the color (green or white). He measured hundreds of cells from one white sponge and hundreds of cells from one green sponge. There were only two units so $n_1 = 1$ and $n_2 = 1$. The student should have used a sample of n_1 green sponges and a sample of n_2 white sponges to get more replicates.

c) Replication depends on the goals of the study. Box, Hunter and Hunter (2005, p. 215-219) describes an experiment where the investigator times how long it takes him to bike up a hill. Since the investigator is only interested in his performance, each run up a hill is a replicate (the time for the ith run is a sample from all possible runs up the hill by the investigator). If the interest had been on the effect of eight treatment levels on student bicyclists, then replication would need $n = n_1 + \cdots + n_8$ student volunteers where n_i ride their bike up the hill under the conditions of treatment i.

1.2 FIXED EFFECTS ONE WAY ANOVA

Definition 1.6. Let $f_Z(z)$ be the probability density function (pdf) of Z. Then the family of pdf's $f_Y(y) = f_Z(y - \mu)$ indexed by the **location parameter** μ, $-\infty < \mu < \infty$, is the **location family** for the random variable $Y = \mu + Z$ with **standard pdf** $f_Z(z)$.

Definition 1.7. A **one way fixed effects ANOVA model** has a single qualitative predictor variable W with p categories a_1, ..., a_p. There are p different distributions for Y, one for each category a_i. The distribution of

$$Y|(W = a_i) \sim f_Z(y - \mu_i)$$

where the location family has second moments. Hence all p distributions come from the same location family with different location parameter μ_i and the same variance σ^2.

Definition 1.8. The **one way fixed effects normal ANOVA model** is the special case where

$$Y|(W = a_i) \sim N(\mu_i, \sigma^2).$$

Example 1.3. The pooled 2 sample t–test is a special case of a one way ANOVA model with $p = 2$. For example, one population could be ACT scores for men and the second population ACT scores for women. Then W = *gender* and Y = *score*.

Notation. It is convenient to relabel the response variable $Y_1, ..., Y_n$ as the vector $\boldsymbol{Y} = (Y_{11}, ..., Y_{1,n_1}, Y_{21}, ..., Y_{2,n_2}, ..., Y_{p1}, ..., Y_{p,n_p})^T$ where the Y_{ij} are independent and $Y_{i1}, ..., Y_{i,n_i}$ are iid. Here $j = 1, ..., n_i$ where n_i is the number of cases from the ith level where $i = 1, ..., p$. Thus $n_1 + \cdots + n_p = n$. Similarly use double subscripts on the errors. Then there will be many equivalent parameterizations of the one way fixed effects ANOVA model.

Definition 1.9. The **cell means model** is the parameterization of the one way fixed effects ANOVA model such that

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

where Y_{ij} is the value of the response variable for the jth trial of the ith factor level. The μ_i are the unknown means and $E(Y_{ij}) = \mu_i$. The e_{ij} are iid from the location family with pdf $f_Z(z)$ and unknown variance $\sigma^2 = \text{VAR}(Y_{ij}) = \text{VAR}(e_{ij})$. For the normal cell means model, the e_{ij} are iid $N(0, \sigma^2)$ for $i = 1, ..., p$ and $j = 1, ..., n_i$.

The cell means model is an ordinary least squares (OLS) model (without intercept) of the form $\boldsymbol{Y} = \boldsymbol{X}_c\boldsymbol{\beta}_c + \boldsymbol{e} =$

$$\begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1,n_1} \\ Y_{21} \\ \vdots \\ Y_{2,n_2} \\ \vdots \\ Y_{p,1} \\ \vdots \\ Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1,n_1} \\ e_{21} \\ \vdots \\ e_{2,n_2} \\ \vdots \\ e_{p,1} \\ \vdots \\ e_{p,n_p} \end{bmatrix}. \tag{1.1}$$

Notation. Let $Y_{i0} = \sum_{j=1}^{n_i} Y_{ij}$ and let

$$\hat{\mu}_i = \overline{Y}_{i0} = Y_{i0}/n_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}. \tag{1.2}$$

Hence the "dot notation" means sum over the subscript corresponding to the 0, e.g. j. Similarly, $Y_{00} = \sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{ij}$ is the sum of all of the Y_{ij}.

Notice that the indicator variables used in the OLS cell means model (1.1) are $x_{hi} = 1$ if the hth case has $W = a_i$, and $x_{hi} = 0$, otherwise, for $i = 1, ..., p$ and $h = 1, ..., n$. So Y_{ij} has $x_{hi} = 1$ where $j = 1, ..., n_i$. The model can use p indicator variables for the factor instead of $p - 1$ indicator variables because the model does not contain an intercept. Also notice that

$$E(\boldsymbol{Y}) = \boldsymbol{X}_c\boldsymbol{\beta}_c = (\mu_1, ..., \mu_1, \mu_2, ..., \mu_2, ..., \mu_p, ..., \mu_p)^T,$$

$(\boldsymbol{X}_c^T \boldsymbol{X}_c) = \text{diag}(\text{n}_1, ..., \text{n}_\text{p})$ and $\boldsymbol{X}_c^T \boldsymbol{Y} = (Y_{10}, ..., Y_{10}, Y_{20}, ..., Y_{20}, ..., Y_{p0}, ..., Y_{p0})^T$. Hence $(\boldsymbol{X}_c^T \boldsymbol{X}_c)^{-1} = \text{diag}(1/\text{n}_1, ..., 1/\text{n}_\text{p})$ and

$$\hat{\boldsymbol{\beta}}_c = (\boldsymbol{X}_c^T \boldsymbol{X}_c)^{-1} \boldsymbol{X}_c^T \boldsymbol{Y} = (\overline{Y}_{10}, ..., \overline{Y}_{p0})^T = (\hat{\mu}_1, ..., \hat{\mu}_p)^T.$$

Thus $\hat{\boldsymbol{Y}} = \boldsymbol{X}_c \hat{\boldsymbol{\beta}}_c = (\overline{Y}_{10}, ..., \overline{Y}_{10}, ...\overline{Y}_{p0}, ..., \overline{Y}_{p0})^T$. Hence the ijth fitted value is

$$\hat{Y}_{ij} = \overline{Y}_{i0} = \hat{\mu}_i \tag{1.3}$$

and the ijth residual is

$$r_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \hat{\mu}_i. \tag{1.4}$$

Definition 1.10. Consider the one way fixed effects ANOVA model. The **response plot** is a plot of $\hat{Y}_{ij} \equiv \hat{\mu}_i$ versus Y_{ij} and the **residual plot** is a plot of $\hat{\mu}_i$ versus r_{ij}.

Since the cell means model is an OLS model, there is an associated response plot and residual plot. However, many of the interpretations of the OLS quantities for ANOVA models differ from the interpretations for multiple linear regression (MLR) models. First, for MLR models with continuous predictors $\boldsymbol{x}$, the conditional distribution $Y|\boldsymbol{x}$ makes sense even if $\boldsymbol{x}$ is not one of the observed $\boldsymbol{x}_i$ if $\boldsymbol{x}$ is not far from the $\boldsymbol{x}_i$. This fact makes MLR very powerful. For the one way fixed effects ANOVA model, the p distributions $Y|\boldsymbol{x}_i$ make sense are where $\boldsymbol{x}_i$ is a column of $\boldsymbol{X}_c$.

Also, the OLS MLR ANOVA F test for the cell means model tests $H_0 : \boldsymbol{\beta} = 0 \equiv H_0 : \mu_1 = \cdots = \mu_p = 0$, while the one way fixed effects ANOVA F test given after Definition 1.12 tests $H_0 : \mu_1 = \cdots = \mu_p$.

The points in the response plot scatter about the identity line with unit slope and zero intercept and the points in the residual plot scatter about the $r = 0$ line, but the scatter need not be in an evenly populated band. A *dot plot* of $Z_1, ..., Z_m$ consists of an axis and m points each corresponding to the value of Z_i. The response plot consists of p dot plots, one for each value of $\hat{\mu}_i$. The dot plot corresponding to $\hat{\mu}_i$ is the dot plot of $Y_{i1}, ..., Y_{i,n_i}$. The p dot plots should have roughly the same amount of spread, and each $\hat{\mu}_i$ corresponds to level a_i. If a new level a_f corresponding to $\boldsymbol{x}_f$ was of interest, hopefully the points in the response plot corresponding to a_f would form a dot plot at $\hat{\mu}_f$ similar in spread to the other dot plots, but it will not be possible to predict the value of $\hat{\mu}_f$.

Similarly, the residual plot consists of p dot plots, and the plot corresponding to $\hat{\mu}_i$ is the dot plot of $r_{i1}, ..., r_{i,n_i}$. Assume that each $n_i \geq 10$. Under the assumption that the Y_{ij} are from the same location scale family with different parameters μ_i, each of the p dot plots should have roughly the same shape and spread. This assumption is easier to judge with the residual plot. If the response plot looks like the residual plot, then a horizontal line fits the p dot plots about as well as the identity line, and there is not much difference in the μ_i. If the identity line is clearly superior to any horizontal line, then at least some of the means differ.

The assumption of the Y_{ij} coming from the same location scale family with different location parameters μ_i and the same constant variance σ^2 is a big assumption and often does not hold. Another way to check this assumption is to make a box plot of the Y_{ij} for each i. The box in the box plot corresponds to the lower, middle, and upper quartiles of the Y_{ij}. The middle quartile is just the sample median of the data m_{ij}: at least half of the $Y_{ij} \geq m_{ij}$ and at least half of the $Y_{ij} \leq m_{ij}$. The p boxes should be roughly the same length and the median should occur in roughly the same position (e.g. in the center of each box). The "whiskers" in each plot should also be roughly similar. Histograms for each of the p samples could also be made. All of the histograms should look similar in shape.

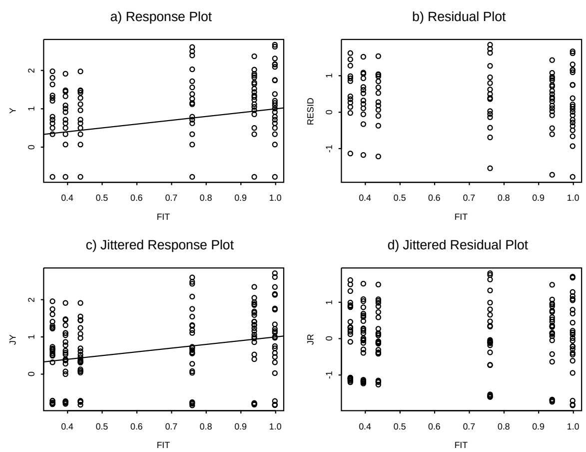

Figure 1.1. Plots for Crab Data

Example 1.4. Kuehl (1994, p. 128) gives data for counts of hermit crabs on 25 different transects in each of six different coastline habitats. Let Z be the count. Then the response variable $Y = \log_{10}(Z + 1/6)$. Although the counts Z varied greatly, each habitat had several counts of 0 and often there were several counts of 1, 2 or 3. Hence Y is not a continuous variable. The cell means model was fit with $n_i = 25$ for $i = 1, ..., 6$. Each of the six habitats was a level. Figure 1.1a and b shows the response plot and residual plot. There are 6 dot plots in each plot. Because several of the smallest values in each plot are identical, it does not always look like the identity line is passing through the six sample means $\overline{Y}_{i0}$ for $i = 1, ..., 6$. In particular, examine the dot plot for the smallest mean (look at the 25 dots furthest to the left that fall on the vertical line FIT ≈ 0.36). Random noise (jitter) has been added to the response and residuals in Figure 1.1c and d. Now it is easier to compare the six dot plots. They seem to have roughly the same spread.

The plots contain a great deal of information. The response plot can be used to explain the model, check that the sample from each population (treatment) has roughly the same shape and spread, and to see which populations have similar means. Since the response plot closely resembles the residual plot in Figure 1.1, there may not be much difference in the six populations. Linearity seems reasonable since the samples scatter about the identity line. The residual plot makes the comparison of "similar shape" and "spread" easier.

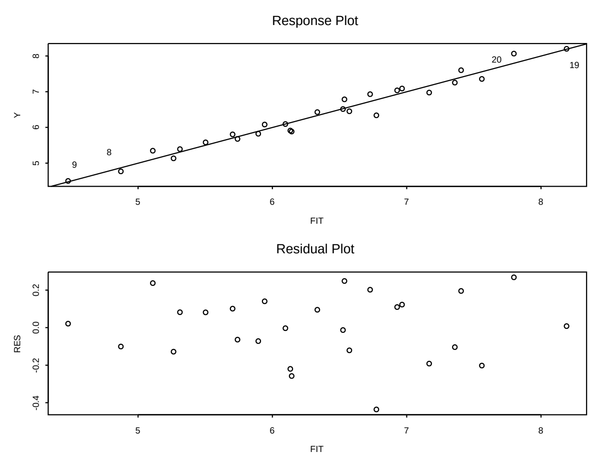

Figure 1.2. Plots for Textile Data

Example 1.5. Box and Cox (1964) analyze data from a 3^3 experiment on the behavior of worsted yarn under cycles of repeated loadings. The response $Y =$

$log(Z)$ where Z is the number of cycles to failure and the three predictors are the length, amplitude and load. To make Figure 1.2, a constant was used in the design matrix, but no interactions. For this data set, there is one value of the response for each of the 27 treatment level combinations.

Figure 1.2 shows that linearity with constant variance is reasonable, and that the signal to noise ratio is high. To use the response plot to visualize the conditional distribution of $Y|\boldsymbol{x}^T\boldsymbol{\beta}$, use the fact that the fitted values $\hat{Y} = \boldsymbol{x}^T\hat{\boldsymbol{\beta}}$. Notice that cases 19 and 20 had the largest time until failure. These cases correspond to wool specimens with long length, short amplitude of loading cycle, and low load. Cases 8 and 9 had the shortest times with low length, high amplitude, and high load.

Definition 1.11. a) The **total sum of squares**

$$SSTO = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{00})^2.$$

b) The **treatment sum of squares**

$$SSTR = \sum_{i=1}^{p} n_i(\overline{Y}_{i0} - \overline{Y}_{00})^2.$$

c) The residual sum of squares or **error sum of squares**

$$SSE = \sum_{i=1}^{p}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{io})^2.$$

Definition 1.12. Associated with each sum of squares (SS) in Definition 1.11 is a **degrees of freedom** (df) and a **mean square** $= SS/df$. For SSTO, df $= n-1$ and $MSTO = SSTO/(n-1)$. For SSTR, df $= p-1$ and $MSTR = SSR/(p-1)$. For SSE, $df = n - p$ and $MSE = SSE/(n-p)$.

Let $S_i^2 = \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i0})^2/(n_i - 1)$ be the sample variance of the ith group. Then the residual mean square (MSE) is a weighted sum of the S_i^2:

$$\hat{\sigma}^2 = MSE = \frac{1}{n-p} \sum_{i=1}^{p} \sum_{j=1}^{n_i} r_{ij}^2 = \frac{1}{n-p} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i0})^2 =$$

$$\frac{1}{n-p} \sum_{i=1}^{p} (n_i - 1) S_i^2 = S_{pool}^2$$

where S_{pool}^2 is known as the pooled variance estimator.

The one way ANOVA table is the same as that for regression, except that SSTR replaces the regression sum of squares. The MSE is again an estimator of σ^2. The ANOVA F test tests whether all p means μ_i are equal. Shown below is an ANOVA table given in symbols. Sometimes "Treatment" is replaced by "Between treatments," "Between Groups," "Model," "Factor" or "Groups." Sometimes "Error" is replaced by "Residual," or "Within Groups." Sometimes "p-value" is replaced by "P", "$Pr(> F)$" or "PR > F."

Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Treatment	$p-1$	SSTR	MSTR	$F_0 = MSTR/MSE$	for H_0:
Error	$n-p$	SSE	MSE		$\mu_1 = \cdots = \mu_p$

Next the 4 step fixed effects one way ANOVA F test of hypotheses is given:

i) State the hypotheses H_0: $\mu_1 = \mu_2 = \cdots = \mu_p$ and H_a: not H_0.

ii) Find the test statistic $F_0 = MSTR/MSE$ or obtain it from output.

iii) Find the p–value from output or use the F–table: p–value =

$$P(F_{p-1,n-p} > F_0).$$

iv) State whether you reject H_0 or fail to reject H_0. If the p–value $< \delta$, reject H_0 and conclude that the mean response depends on the factor. Otherwise fail to reject H_0 and conclude that the mean response does not depend on the factor. Give a nontechnical sentence.

Rule of Thumb 1.1. Moore (2000, p. 512). If

$$\max(S_1, ..., S_p) \leq 2 \min(S_1, ..., S_p),$$

then the one way ANOVA F test results will be approximately correct if the response and residual plots suggest that the remaining one way ANOVA model assumptions are reasonable.

Remark 1.1. If the units are a representative sample of some population of interest, then randomization of units into groups makes the assumption that $Y_{i1}, ..., Y_{i,n_i}$ are iid hold to a useful approximation. Random sampling from populations also induces the iid assumption. Linearity can be checked with the response plot, and similar shape and spread of the location families can be checked with both the response and residual plots. Also check that outliers are not present. If the p dot plots in the response plot are approximately symmetric, then the sample sizes n_i can be smaller than if the dot plots are skewed.

Remark 1.2. When the assumption that the p groups come from the same location family with finite variance σ^2 is violated, the one way ANOVA F test may not make much sense because unequal means may not imply the superiority of one category over another. Suppose Y is the time in minutes until relief from a headache and that $Y_{1j} \sim N(60, 1)$ while $Y_{2j} \sim N(65, \sigma^2)$. If $\sigma^2 = 1$, then the type 1 medicine gives headache relief 5 minutes faster, on average, and is superior, all other things being equal. But if $\sigma^2 = 100$, then many patients taking medicine 2 experience much faster pain relief than those taking medicine 1, and many experience much longer time until pain relief. In this situation, predictor variables that would identify which medicine is faster for a given patient would be very useful.

```
fat1  fat2  fat3  fat4 One way Anova for Fat1 Fat2 Fat3 Fat4
64    78    75    55   Source      DF    SS     MS      F    P
72    91    93    66   treatment  3    1636.5 545.5  5.41 0.0069
68    97    78    49   error      20   2018.0 100.9
77    82    71    64
56    85    63    70
95    77    76    68
```

Example 1.6. The output above represents grams of fat (minus 100 grams) absorbed by doughnuts using 4 types of fat. See Snedecor and Cochran (1967, p. 259). Let μ_i denote the mean amount of fat i absorbed by doughnuts, $i = 1, 2, 3$ and 4. a) Find $\hat{\mu}_1$. b) Perform a 4 step ANOVA F test.

Solution: a) $\hat{\boldsymbol{\beta}}_{1c} = \hat{\mu}_1 = \overline{Y}_{10} = Y_{10}/n_1 = \sum_{j=1}^{n_1} Y_{1j}/n_1 =$

$(64 + 72 + 68 + 77 + 56 + 95)/6 = 432/6 = 72.$

b) i) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ H_a: not H_0

ii) F = 5.41

iii) pvalue = 0.0069

iv) Reject H_0, the mean amount of fat absorbed by doughnuts depends on the type of fat.

Definition 1.13. A **contrast** $C = \sum_{i=1}^p k_i \mu_i$ where $\sum_{i=1}^p k_i = 0$. The estimated contrast is $\hat{C} = \sum_{i=1}^p k_i \overline{Y}_{i0}$.

If the null hypothesis of the fixed effects one way ANOVA test is not true, then not all of the means μ_i are equal. Researchers will often have hypotheses, before examining the data, that they desire to test. Often such a hypothesis can be put in the form of a contrast. For example, the contrast $C = \mu_i = \mu_j$ is used to compare the means of the ith and jth groups while the contrast $\mu_1 = -(\mu_2 + \cdots + \mu_p)/(p-1)$ is used to compare the last $p-1$ groups with the 1st group. This contrast is useful when the 1st group corresponds to a new treatment while the remaining groups correspond to new treatments.

Assume that the normal cell means model is a useful approximation to the data. Then the $\overline{Y}_{i0} \sim N(\mu_i, \sigma^2/n_i)$ are independent, and

$$\hat{C} = \sum_{i=1}^p k_i \overline{Y}_{i0} \sim N\left(C, \sigma^2 \sum_{i=1}^p \frac{k_i^2}{n_i}\right).$$

Hence the standard error

$$SE(\hat{C}) = \sqrt{MSE \sum_{i=1}^p \frac{k_i^2}{n_i}}.$$

The degrees of freedom is equal to the MSE degrees of freedom $= n - p$.

1.3 RANDOM EFFECTS ONE WAY ANOVA

For the **random effects one way ANOVA**, the levels of the factor are a random sample of levels from some population of levels Λ. The cell means model for the random effects one way ANOVA is $Y_{ij} = \mu_i + e_{ij}$ for $i = 1, ..., r$ and $j = 1, ..., n$. The μ_i are independent $N(\mu_o, \sigma^2_\mu)$ random variables, and the e_{ij} are iid $N(0, \sigma^2)$ random variables. The e_{ij} and the μ_i are independent. Note that the population of levels $\Lambda \sim N(\mu_o, \sigma^2_\mu)$ and if $\sigma^2_\mu = 0$, then $\mu_i \equiv \mu_o$ for all $i \in \Lambda$.

The 4 step random effects one way ANOVA test is

i) H_0: $\sigma^2_\mu = 0$ H_a: $\sigma^2_\mu > 0$.

ii) $F_0 = MSTR/MSE \sim F_{r-1,nr-1}$ if H_0 is true.

iii) p-value $= P(F_{r-1,nr-1} > F_0)$ is usually obtained from output.

iv) If p–value $< \delta$ reject H_0, conclude that $\sigma^2_\mu > 0$ and that the mean response depends on the factor. Otherwise, fail to reject H_0, conclude that $\sigma^2_\mu = 0$ and that the mean response does not depend on the factor.

1.4 RESPONSE TRANSFORMATIONS FOR EXPERIMENTAL DESIGNS

A model for an experimental design is $Y_i = E(Y_i) + e_i$ for $i = 1, ..., n$ where the error $e_i = Y_i - E(Y_i)$ and $E(Y_i) \equiv E(Y_i|\boldsymbol{x}_i)$ is the expected value of the response Y_i for a given vector of predictors $\boldsymbol{x}_i$. Many models can be fit with least squares (OLS) and have the form

$$Y_i = x_{i,1}\beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$$

for $i = 1, \ldots, n$. Often $x_{i,1} \equiv 1$ for all i. In matrix notation, these n equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ design matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. If the fitted values are $\hat{Y}_i = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$, then $Y_i = \hat{Y}_i + r_i$ where the residuals $r_i = Y_i - \hat{Y}_i$.

The applicability of an experimental design model can be expanded by allowing response transformations. An important class of *response transformation models* adds an additional unknown transformation parameter λ_o, such that

$$Y_i = t_{\lambda_o}(Z_i) \equiv Z_i^{(\lambda_o)} = E(Y_i) + e_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i.$$

If λ_o was known, then $Y_i = t_{\lambda_o}(Z_i)$ would follow the experimental design model.

Definition 1.14. Assume that *all* of the values of the "response" Z_i are *positive*. A **power transformation** has the form $Y = t_\lambda(Z) = Z^\lambda$ for $\lambda \neq 0$ and $Y = t_0(Z) = \log(Z)$ for $\lambda = 0$ where $\lambda \in \Lambda_L = \{-1, -1/2, 0, 1/2, 1\}$.

A graphical method for response transformations computes the fitted values $\hat{W}_i$ from the experimental design model using $W_i = t_\lambda(Z_i)$ as the "response" for each of the five values of $\lambda \in \Lambda_L$. The plotted points follow the identity line in a (roughly) evenly populated band if the experimental design model is reasonable for $(\hat{W}, W)$. If more than one value of $\lambda \in \Lambda_L$ gives a linear plot, consult subject matter experts and use the simplest or most reasonable transformation.

Definition 1.15. A **transformation plot** is a plot of $(\hat{W}, W)$ with the identity line added as a visual aid.

After selecting the transformation, the usual checks should be made. A variant of the method would plot the residual plot or both the response and the residual plot for each of the five values of λ. Residual plots are also useful, but they no not distinguish between nonlinear monotone relationships and nonmonotone relationships. See Fox (1991, p. 55).

In the following two examples, the plots show $t_\lambda(Z)$ on the vertical axis. The label "TZHAT" of the horizontal axis are the fitted values that result from using $t_\lambda(Z)$ as the "response" in the software.

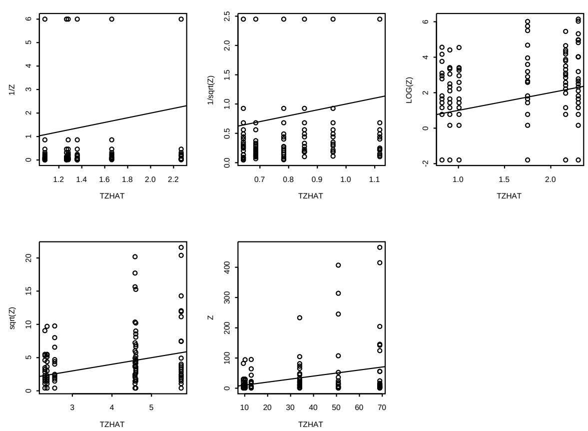

Figure 1.3. Transformation Plots for Crab Data

Example 1.4, continued. Following Kuehl (1994, p. 128), let C be the count of crabs and let the "response" $Z = C + 1/6$. Figure 1.3 shows the five *transformation plots*. The transformation $\log(Z)$ results in dot plots that have roughly the same shape and spread. The transformations $1/Z$ and $1/\sqrt{Z}$ do not handle the 0 counts well, while the transformations $\sqrt{Z}$ and Z have variance that increases with the mean.

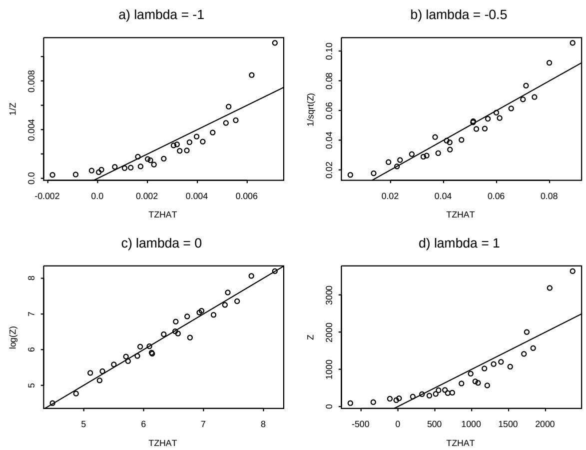

Figure 1.4. Transformation Plots for Textile Data

Example 1.5, continued. For the textile data, Z = number of cycles until failure. Figure 1.4 shows four of the five transformation plots. The plotted points for the log transformation follow the identity line with roughly constant variance.

1.5 COMMENTS ON THE DESIGN AND ANALYSIS OF EXPERIMENTS

All of the parameterizations of the one way fixed effects ANOVA model yield the same predicted values, residuals and ANOVA F test, but the interpretations of the parameters differ.

Three excellent tests on the design and analysis of experiments are Box, Hunter and Hunter (2005), Cobb (1998) and Kuehl (1994). The response plot of fitted values versus the response is useful for visualizing many of these models.

Wilcox (2005) gives an excellent discussion of the problems that outliers and skewness can cause for the one and two sample t–intervals, the t–test, tests for comparing 2 groups and the ANOVA F test. Wilcox (2005) replaces ordinary population means by truncated population means and uses trimmed means to create analogs of one way ANOVA and multiple comparisons.

Box and Cox (1964) give a numerical method for selecting the response transformation for the modified power transformations. Although the method gives a point estimator $\hat{\lambda}_0$, often an interval of "reasonable values" is generated (either graphically or using a profile likelihood to make a confidence interval), and $\hat{\lambda} \in \Lambda_L$ is used if it is also in the interval.

The *modified power transformation family*

$$Y_i = t_\lambda(Z_i) \equiv Z_i^{(\lambda)} = \frac{Z_i^\lambda - 1}{\lambda}$$

for $\lambda \neq 0$ and $t_0(Z_i) = \log(Z_i)$ for $\lambda = 0$ where $\lambda \in \Lambda_L$.

There are several reasons to use a coarse grid Λ_L of powers. First, several of the powers correspond to simple transformations such as the log, square root, and reciprocal. These powers are easier to interpret than $\lambda = .28$, for example. Secondly, if the estimator $\hat{\lambda}_n$ can only take values in Λ_L, then sometimes $\hat{\lambda}_n$ will converge in probability to $\lambda^* \in \Lambda_L$. Thirdly, Tukey (1957) showed that neighboring modified power transformations are often very similar, so restricting the possible powers to a coarse grid is reasonable.

The graphical method for response transformations is due to Olive (2004, 2008), and alternative methods are given by Cook and Olive (2002) and Box, Hunter and Hunter (2005, p. 321).

1.6 ALTERNATIVES TO ONE WAY ANOVA

An alternative to one way ANOVA is to use feasible weighted least squares (FWLS) on the cell means model with $\sigma^2 \boldsymbol{V} = \text{diag}(\sigma_1^2, ..., \sigma_p^2)$ where σ_i^2 is the variance of the ith group for $i = 1, ..., p$. Then $\hat{\boldsymbol{V}} = \text{diag}(S_1^2, ..., S_p^2)$ where $S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i0})^2$ is the sample variance of the Y_{ij}. Hence the estimated weights for FWLS are $\hat{w}_{ij} \equiv \hat{w}_i = 1/S_i^2$. Then the FWLS cell means model has $Y = \boldsymbol{X}_c \boldsymbol{\beta}_c + \boldsymbol{e}$ as in (1.1) except $\text{Cov}(\boldsymbol{e}) = \text{diag}(\sigma_1^2, ..., \sigma_p^2)$. Hence $\boldsymbol{Z} = \boldsymbol{U}_c \boldsymbol{\beta}_c + \boldsymbol{\epsilon}$ or

$$\begin{bmatrix} \sqrt{\hat{w}_1} Y_{1,1} \\ \vdots \\ \sqrt{\hat{w}_1} Y_{1,n_1} \\ \sqrt{\hat{w}_2} Y_{21} \\ \vdots \\ \sqrt{\hat{w}_2} Y_{2,n_2} \\ \vdots \\ \sqrt{\hat{w}_p} Y_{p,1} \\ \vdots \\ \sqrt{\hat{w}_p} Y_{p,n_p} \end{bmatrix} = \begin{bmatrix} \sqrt{\hat{w}_1} & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \sqrt{\hat{w}_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\hat{w}_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \sqrt{\hat{w}_2} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{\hat{w}_p} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{\hat{w}_p} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \mu_p \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{1,n_1} \\ \epsilon_{21} \\ \vdots \\ \epsilon_{2,n_2} \\ \vdots \\ \epsilon_{p,1} \\ \vdots \\ \epsilon_{p,n_p} \end{bmatrix}. \tag{1.5}$$

Then $\boldsymbol{U}_c^T \boldsymbol{U}_c = \text{diag}(n_1 \hat{w}_1, ..., n_p \hat{w}_p)$, $(\boldsymbol{U}_c^T \boldsymbol{U}_c)^{-1} = \text{diag}(S_1^2/n_1, ..., S_p^2/n_p)$ $= (\boldsymbol{X} \hat{\boldsymbol{V}}^{-1} \boldsymbol{X}^T)^{-1}$, and $\boldsymbol{U}_c^T \boldsymbol{Z} = (\hat{w}_1 Y_{10}, ..., \hat{w}_p Y_{p0})^T$. Thus

$$\hat{\boldsymbol{\beta}}_{FWLS} = (\overline{Y}_{10}, ..., \overline{Y}_{p0})^T = \hat{\boldsymbol{\beta}}_c.$$

That is, the FWLS estimator equals the one way ANOVA estimator of $\boldsymbol{\beta}$ based on OLS applied to the cell means model. The ANOVA F test generalizes the pooled t test in that the two tests are equivalent for $p = 2$. The FWLS procedure is also known as the Welch one way ANOVA and generalizes the Welch t test. The Welch t test is thought to be much better than the pooled t test. See Brown and Forsythe (1974), Kirk (1982, p. 100, 101, 121, 122) and Welch (1947, 1951).

Four tests for $H_0 : \mu_1 = \cdots = \mu_p$ can be used if Rule of Thumb 1.1: $\max(S_1, ..., S_p) \leq 2\min(S_1, ..., S_p)$ fails. Let $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$, and let $Y_{(1)} \leq Y_{(2)} \cdots \leq Y_{(n)}$ be the order statistics. Then the rank transformation of the response is $\boldsymbol{Z} = rank(\boldsymbol{Y})$ where $Z_i = j$ if $Y_i = Y_{(j)}$ is the jth largest order statistic. For example, if $\boldsymbol{Y} = (7.7, 4.9, 33.3, 6.6)^T$, then $\boldsymbol{Z} = (3, 1, 4, 2)^T$. The first test performs the one way ANOVA F test with $\boldsymbol{Z}$ replacing $\boldsymbol{Y}$. See Montgomery (1984, p. 117-118). The two of the next three tests are described in Brown and Forsythe (1974). Let $\lceil x \rceil$ be the smallest integer $\geq x$, e.g. $\lceil 7.7 \rceil = 8$. Then the Welch (1951) ANOVA F test uses test statistic

$$F_W = \frac{\sum_{i=1}^p w_i(\overline{Y}_{i0} - \tilde{Y}_{00})^2/(p-1)}{1 + \frac{2(p-2)}{p^2-1}\sum_{i=1}^p (1 - \frac{w_i}{u})^2/(n_i - 1)}$$

where $w_i = n_i/S_i^2, u = \sum_{i=1}^p w_i$ and $\tilde{Y}_{00} = \sum_{i=1}^p w_i \overline{Y}_{i0}/u$. Then the test statistic is compared to an F_{p-1,d_W} distribution where $d_W = \lceil f \rceil$ and

$$1/f = \frac{3}{p^2-1}\sum_{i=1}^p (1 - \frac{w_i}{u})^2/(n_i - 1).$$

For the modified Welch (1947) test, the test statistic is compared to an $F_{p-1,d_{MW}}$

distribution where $d_{MW} = \lceil f \rceil$ and

$$f = \frac{\sum_{i=1}^p (S_i^2/n_i)^2}{\sum_{i=1}^p \frac{1}{n_i-1}(S_i^2/n_i)^2} = \frac{\sum_{i=1}^p (1/w_i)^2}{\sum_{i=1}^p \frac{1}{n_i-1}(1/w_i)^2}.$$

Some software uses f instead of d_W or d_{MW}, and variants on the denominator degrees of freedom d_W or d_{MW} are common.

The modified ANOVA F test uses test statistic

$$F_M = \frac{\sum_{i=1}^p n_i(\overline{Y}_{i0} - \overline{Y}_{00})^2}{\sum_{i=1}^p (1 - \frac{n_i}{n})S_i^2}$$

The test statistic is compared to an F_{p-1,d_M} distribution where $d_M = \lceil f \rceil$ and

$$1/f = \sum_{i=1}^p c_i^2/(n_i - 1)$$

where

$$c_i = (1 - \frac{n_i}{n})S_i^2/[\sum_{i=1}^p (1 - \frac{n_i}{n})S_i^2].$$

The *regpack* function `anovasim` can be used to compare the five tests.

CHAPTER 2

SIMULATIONS

The simulations were performed in *R/Splus* using the `anovasim` function. Each simulation used 40,000 runs and had various sample sizes (n_i), population means (μ_i), and standard deviations (σ_i). The significance level was set at 0.05 for the ANOVA F test (F), the modified ANOVA F test (F_M), the Welch ANOVA F test (F_W), the modified Welch ANOVA F test (F_{MW}), and the ANOVA F rank test (F_R). This simulation is similar to that of Rodriguez (2007), which used 5000 runs and did not include F_{MW}.

For Table 2.1, we want levels near 0.05. Overall, the Welch test had the best performance, giving simulated ANOVA F statistics closest to 0.05, ranging from 0.04875 to 0.0671. The modified Welch test is almost as good with statistics ranging from 0.0466 to 0.069425. The modified ANOVA F test is not quite as good, with statistics ranging from 0.037975 to 0.062525. The rank and ANOVA F test are not as good. The rank test did not perform well when n_i and σ_i were not equal. The ANOVA F test had levels near 0.05 when $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$, but has values ranging from 0.030175 to 0.1413 when σ_i were not equal.

In Tables 2.3, 2.4, 2.5, and 2.6, we want levels near 0.05 if $\mu_1 = \mu_2 = \mu_3 = \mu_4$. Otherwise we want good power with levels near 1. For Table 2.3, all five tests gave similar statistics. The ANOVA F test gives the most favorable statistics, and the modified Welch test also gives favorable statistics, but is not as good. For Table

2.4 and 2.5, the modified Welch test performs the best, and the Welch test is next best with very similar values. The rank and modified ANOVA F tests seem to be the weaker tests in these tables. For Table 2.6, the Welch test had the best levels except when $\mu_4 = 1$ and $\mu_1 = \mu_2 = \mu_3 = 0$. In this case, we want the statistic to be near 1, but the Welch test gives values near 0.05. For this case, the ANOVA F test performed consistently better than the other four tests.

Table 2.1. $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, F = proportion of times the ANOVA F test rejected H_0 with level .05, 40,000 runs

n_1, n_2, n_3, n_4	σ_1, σ_2, σ_3, σ_4	F	F_M	F_W	F_{MW}	F_R
4, 4, 4, 4	1, 1, 1, 1	0.0502	0.0380	0.0536	0.0664	0.0582
	1, 2, 2, 3	0.0678	0.0496	0.0621	0.0679	0.0719
	3, 2, 2, 1	0.0678	0.0497	0.0635	0.0694	0.0724
4, 8, 10, 12	1, 1, 1, 1	0.0505	0.0487	0.0596	0.0673	0.0513
	1, 2, 2, 3	0.0302	0.0585	0.0502	0.0632	0.0353
	3, 2, 2, 1	0.1413	0.0654	0.0671	0.0620	0.1003
8, 16, 20, 24	1, 1, 1, 1	0.0511	0.0506	0.0529	0.0562	0.0519
	1, 2, 2, 3	0.0302	0.0614	0.0508	0.0584	0.0353
	3, 2, 2, 1	0.1353	0.0608	0.0542	0.0466	0.0915
11, 11, 11, 11	1, 1, 1, 1	0.0534	0.0515	0.0534	0.0626	0.0535
	1, 2, 2, 3	0.0634	0.0573	0.0512	0.0570	0.0596
	3, 2, 2, 1	0.0670	0.0616	0.0535	0.0594	0.0640
11, 16, 16, 21	1, 1, 1, 1	0.0501	0.0493	0.0513	0.0573	0.0517
	1, 2, 2, 3	0.0378	0.0607	0.0493	0.0557	0.0423
	3, 2, 2, 1	0.1058	0.0605	0.0525	0.0522	0.0817
22, 32, 32, 42	1, 1, 1, 1	0.0485	0.0486	0.0490	0.0526	0.0482
	1, 2, 2, 3	0.0360	0.0606	0.0503	0.0539	0.0410
	3, 2, 2, 1	0.1046	0.0617	0.0512	0.0511	0.0825

Table 2.2. (Table 2.1 continued)

n_1, n_2, n_3, n_4	σ_1, σ_2, σ_3, σ_4	F	F_M	F_W	F_{MW}	F_R
5, 5, 5, 5	1, 1, 1, 1	0.0494	0.0417	0.0520	0.0669	0.0559
	1, 2, 2, 3	0.0654	0.0514	0.0602	0.0676	0.0670
	3, 2, 2, 1	0.0657	0.0516	0.0577	0.0646	0.0657
10, 10, 10, 10	1, 1, 1, 1	0.0505	0.0482	0.0496	0.0596	0.0514
	1, 2, 2, 3	0.0644	0.0580	0.0540	0.0608	0.0628
	3, 2, 2, 1	0.0640	0.0582	0.0533	0.0590	0.0618
20, 20, 20, 20	1, 1, 1, 1	0.0505	0.0500	0.0500	0.0554	0.0519
	1, 2, 2, 3	0.0633	0.0603	0.0509	0.0544	0.0604
	3, 2, 2, 1	0.0639	0.0604	0.0507	0.0543	0.0600
50, 50, 50, 50	1, 1, 1, 1	0.0516	0.0515	0.0519	0.0543	0.0522
	1, 2, 2, 3	0.0628	0.0616	0.0503	0.0518	0.0610
	3, 2, 2, 1	0.0629	0.0617	0.0498	0.0512	0.0590
100, 100, 100, 100	1, 1, 1, 1	0.0498	0.0498	0.0497	0.0508	0.0490
	1, 2, 2, 3	0.0633	0.0628	0.0509	0.0516	0.0621
	3, 2, 2, 1	0.0630	0.0625	0.0488	0.0497	0.0598

Table 2.3. $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, F = proportion of times the ANOVA F test rejected H_0 with level .05, 40,000 runs

n_1, n_2, n_3, n_4	$\mu_1, \mu_2, \mu_3, \mu_4$	F	F_M	F_W	F_{MW}	F_R
4, 4, 4, 4	0, 0, 0, 0	0.0491	0.0386	0.0542	0.0676	0.0577
	1, 0, 0, 0	0.2100	0.1757	0.1841	0.2232	0.2168
	1, 0, 0, 0.7	0.2093	0.1736	0.1853	0.2234	0.2232
	5, 0, 0, 0.5	1	1	0.9982	0.9987	0.9999
5, 5, 5, 5	0, 0, 0, 0	0.0500	0.0428	0.0535	0.0676	0.0566
	1, 0, 0, 0	0.2757	0.2482	0.2373	0.2848	0.2718
	1, 0, 0, 0.7	0.2734	0.2468	0.2390	0.2872	0.2792
	5, 0, 0, 0.5	1	1	1	1	1
10, 10, 10, 10	0, 0, 0, 0	0.0489	0.0470	0.0519	0.0617	0.0522
	1, 0, 0, 0	0.5692	0.5618	0.5280	0.5644	0.5511
	1, 0, 0, 0.7	0.5782	0.5706	0.5447	0.0582	0.5676
	5, 0, 0, 0.5	1	1	1	1	1
20, 20, 20, 20	0, 0, 0, 0	0.0521	0.0517	0.0509	0.0562	0.0525
	1, 0, 0, 0	0.9017	0.9009	0.8841	0.8926	0.8836
	1, 0, 0, 0.7	0.9086	0.9079	0.8982	0.9068	0.8961
	5, 0, 0, 0.5	1	1	1	1	1
100, 100, 100, 100	0, 0, 0, 0	0.0515	0.0515	0.0513	0.0525	0.0501
	1, 0, 0, 0	1	1	1	1	1
	1, 0, 0, 0.7	1	1	1	1	1
	5, 0, 0, 0.5	1	1	1	1	1

Table 2.4. $\sigma_1 = 3$, $\sigma_2 = 2$, $\sigma_3 = 2$, $\sigma_4 = 1$, F = proportion of times the ANOVA F test rejected H_0 with level .05, 40,000 runs

n_1, n_2, n_3, n_4	μ_1, μ_2, μ_3, μ_4	F	F_M	F_W	F_{MW}	F_R
4, 4, 4, 4	0, 0, 0, 0	0.0671	0.0500	0.0613	0.0676	0.0715
	1.5, 0, 0, 0	0.1516	0.1162	0.1046	0.1174	0.1430
	0, 0, 0, 1	0.0921	0.0670	0.1199	0.1314	0.1069
	1.3, 0, 0, 1.3	0.1374	0.1036	0.1572	0.1700	0.1535
5, 5, 5, 5	0, 0, 0, 0	0.0642	0.0503	0.0582	0.0653	0.0641
	1.5, 0, 0, 0	0.1761	0.1454	0.1155	0.1307	0.1550
	0, 0, 0, 1	0.1003	0.0797	0.1398	0.1552	0.1154
	1.3, 0, 0, 1.3	0.1617	0.1309	0.1904	0.2109	0.1804
10, 10, 10, 10	0, 0, 0, 0	0.0626	0.0566	0.0489	0.0551	0.0592
	1.5, 0, 0, 0	0.3085	0.2889	0.1932	0.2082	0.2492
	0, 0, 0, 1	0.1490	0.1347	0.2821	0.3014	0.1988
	1.3, 0, 0, 1.3	0.2987	0.2771	0.4208	0.4427	0.3514
20, 20, 20, 20	0, 0, 0, 0	0.0628	0.0599	0.0491	0.0524	0.0592
	1.5, 0, 0, 0	0.5444	0.5347	0.3767	0.3893	0.4475
	0, 0, 0, 1	0.2739	0.2626	0.5799	0.5924	0.4014
	1.3, 0, 0, 1.3	0.5861	0.5741	0.7845	0.7940	0.6827
100, 100, 100, 100	0, 0, 0, 0	0.0609	0.0602	0.0502	0.0509	0.0592
	1.5, 0, 0, 0	0.9964	0.9964	0.9875	0.9877	0.9884
	0, 0, 0, 1	0.9890	0.9889	0.9999	0.9999	0.9974
	1.3, 0, 0, 1.3	1	1	1	1	1

Table 2.5. $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 2$, $\sigma_4 = 3$, F = proportion of times the ANOVA F test rejected H_0 with level .05, 40,000 runs

n_1, n_2, n_3, n_4	$\mu_1, \mu_2, \mu_3, \mu_4$	F	F_M	F_W	F_{MW}	F_R
4, 4, 4, 4	0, 0, 0, 0	0.0691	0.0505	0.0621	0.0687	0.0730
	1.3, 0, 0, 0	0.1151	0.0854	0.1604	0.1754	0.1344
	0, 0, 0, 1	0.1045	0.0784	0.0818	0.0896	0.1050
	1, 0, 0, 1	0.1085	0.0811	0.1162	0.1255	0.1191
5, 5, 5, 5	0, 0, 0, 0	0.0686	0.0540	0.0595	0.0665	0.0672
	1.3, 0, 0, 0	0.1290	0.1023	0.2050	0.2277	0.1589
	0, 0, 0, 1	0.1161	0.0947	0.0835	0.0951	0.1054
	1, 0, 0, 1	0.1200	0.0965	0.1344	0.1493	0.1302
10, 10, 10, 10	0, 0, 0, 0	0.0646	0.0585	0.0521	0.0580	0.0635
	1.3, 0, 0, 0	0.2245	0.2044	0.4601	0.4837	0.3186
	0, 0, 0, 1	0.1721	0.1594	0.1096	0.1201	0.1412
	1, 0, 0, 1	0.1926	0.1773	0.2609	0.2787	0.2216
20, 20, 20, 20	0, 0, 0, 0	0.0641	0.0602	0.0497	0.0530	0.0606
	1.3, 0, 0, 0	0.48883	0.4732	0.8301	0.8385	0.6575
	0, 0, 0, 1	0.2857	0.2777	0.1816	0.1915	0.2299
	1, 0, 0, 1	0.3596	0.3489	0.5343	0.5468	0.4334
100, 100, 100, 100	0, 0, 0, 0	0.0632	0.0628	0.0524	0.0532	0.0617
	1.3, 0, 0, 0	0.9999	0.9999	1	1	1
	0, 0, 0, 1	0.8781	0.8770	0.7675	0.7694	0.7946
	1, 0, 0, 1	0.9913	0.9911	0.9994	0.9994	0.9969

Table 2.6. $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\sigma_4 = 9$, F = proportion of times the ANOVA F test rejected H_0 with level .05, 40,000 runs

n_1, n_2, n_3, n_4	μ_1, μ_2, μ_3, μ_4	F	F_M	F_W	F_{MW}	F_R
4, 4, 4, 4	0, 0, 0, 0	0.1496	0.0793	0.0632	0.0385	0.1113
	1.3, 0, 0, 0	0.1552	0.0834	0.2467	0.1462	0.1543
	0, 0, 0, 1	0.1539	0.0830	0.0664	0.0399	0.1170
	1, 0, 0, 1	0.1553	0.0836	0.1648	0.0939	0.1392
5, 5, 5, 5	0, 0, 0, 0	0.1410	0.0792	0.0584	0.0327	0.0802
	1.3, 0, 0, 0	0.1513	0.0863	0.3286	0.2025	0.2106
	0, 0, 0, 1	0.1461	0.0841	0.0620	0.0344	0.0856
	1, 0, 0, 1	0.1492	0.0838	0.2118	0.1245	0.1558
10, 10, 10, 10	0, 0, 0, 0	0.1196	0.0824	0.0504	0.0317	0.0899
	1.3, 0, 0, 0	0.1395	0.0974	0.7165	0.6304	0.4921
	0, 0, 0, 1	0.1405	0.1005	0.0591	0.0372	0.1032
	1, 0, 0, 1	0.1363	0.0961	0.4808	0.3861	0.3115
20, 20, 20, 20	0, 0, 0, 0	0.1133	0.0944	0.0503	0.0389	0.0835
	1.3, 0, 0, 0	0.1526	0.1261	0.9766	0.9687	0.8960
	0, 0, 0, 1	0.1521	0.1284	0.0643	0.0509	0.1083
	1, 0, 0, 1	0.1489	0.1248	0.8462	0.8153	0.6531
100, 100, 100, 100	0, 0, 0, 0	0.1030	0.0994	0.0499	0.0470	0.0802
	1.3, 0, 0, 0	0.4588	0.4388	1	1	1
	0, 0, 0, 1	0.3038	0.2974	0.1303	0.1253	0.2029
	1, 0, 0, 1	0.3342	0.3239	1	1	1

REFERENCES

[1] Box, G.E.P., and Cox, D.R. (1964), An Analysis of Transformations, *Journal of the Royal Statistical Society*, B, 26, 211-246.

[2] Box, G.E.P, Hunter, J.S., and Hunter, W.G. (2005), *Statistics for Experimenters*, 2nd ed., Wiley, NY.

[3] Brown, M.B. and Forsythe, A.B. (1974), "The Small Sample Behavior of Some Statistics Which Test the Equality of Several Means," *Technometrics*, 16, 129-132.

[4] Cobb, G.W. (1998), *Introduction to Design and Analysis of Experiments*, Key College Publishing, Emeryville, CA.

[5] Cook, R.D., and Olive, D.J. (2001), A Note on Visualizing Response Transformations in Regression, *Technometrics*, 43, 443-449.

[6] Fox, J. (1991), *Regression Diagnostics*, Sage Publications, Newbury Park, CA.

[7] Kirk, R.E. (1982), *Experimental Design: Procedures for the Behavioral Sciences*, 2nd ed., Brooks/Cole Publishing Company, Belmont, CA.

[8] Kuehl, R.O. (1994), *Statistical Principles of Research Design and Analysis*, Duxbury Press, Belmont, CA.

[9] Montgomery, D.C. (1984), *Design and Analysis of Experiments*, 2nd ed., Wiley, NY.

[10] Moore, D.S. (2000), *The Basic Practice of Statistics*, 2nd ed., W.H. Freeman, NY.

[11] Olive, D.J. (2009), *Multiple Linear and 1D Regression Models*, Unpublished Manuscript available from (www.math.siu.edu/olive/regbk.htm).

[12] Olive, D.J. (2008), *Response Plots for Experimental Design*, Unpublished Manuscript available from (www.math.siu.edu/olive/ppresplted.pdf).

[13] Olive, D.J. (2004), Visualizing 1D Regression, in *Theory and Applications of Recent Robust Methods*, eds. Hubert, M., Pison, G., Struyf, A., and Van Aelst S., Series: Statistics for Industry and Technology, Birkhauser, Basel.

[14] Rodriguez, E.A. (2007), *Regression and Anova Under Heterogeneity*, Unpublished Manuscript available from (www.math.siu.edu/olive/selmer.pdf).

[15] Snedecor, G.W., and Cochran, W.G. (1967), *Statistical Methods*, 6th ed., Iowa State College Press, Ames, Iowa.

[16] Tukey, J.W. (1957), Comparative Anatomy of Transformations, *Annals of Mathematical Statistics*, 28, 602-632.

[17] Welch, B.L. (1947), The Generalization of Students Problem When Several Different Population Variances are Involved, *Biometrika*, 34, 28-35.

[18] Welch, B.L. (1951), On the Comparison of Several Mean Values: an Alternative Approach, *Biometrika*, 38, 330-336.

[19] Wilcox, R.R. (2005), *Introduction to Robust Estimation and Testing*, 2nd ed., Elsevier Academic Press, San Diego, CA.