Day By Day Notes for MATH 386/586

Spring 2012

Day 1

Activity: Introduction / Math 301 review.

This course is about analyzing averages from more than one sample of data, and about how to design experiments to give meaningful results. Today I want to review some of the key theoretical results we will need to accomplish our goals. Often in class we will use spreadsheets and MINITAB files to analyze data and demonstrate the techniques. I will save these spreadsheets and analyses on TitanFiles in a Math386 folder so that you all have access to our work.

I will begin the review with Random Variables. The easiest examples of random variables are counting heads on coins and totaling the sums on dice. Basically, a random variable is an assignment of a value to each of the elements of a sample space. We are usually interested in how random variables are distributed, e.g. what the possible and probable number of heads in n coin tosses are. A complete enumeration of the values of a random variable and their associated probabilities constitutes a probability distribution function (pdf) or a probability mass function (pmf). In the case of a continuous random variable, the pdf will likely be given in equation form (e.g. the normal density curve). In the case of a discrete random variable, we may have an equation (e.g. the binomial curve) or we may just have a table of values describing the distribution.

Given a pdf, either from a continuous or a discrete random variable, we usually need to know the typical value, and how much the probabilities are spread out. Two keys calculations from a distribution are the Expected Value and Variance. The Expected Value (EV) is really an average or a mean, while the Variance (Var) is very similar to the mean of the squared deviations from average. High variance indicates a distribution that has quite disparate values while a low variance indicates a distribution that has homogenous values. For the common distributions (uniform, binomial, normal, gamma, among others) we calculate the EV and Var theoretically, using either sums (discrete pdf's) or integrals (continuous pdf's). The chief distributions we will need in this course are the normal, the χ^2, the t, and the F (see below).

To understand distributions and expectation, simulation can help us. If we generate a series of observations, and calculate the mean and variance of our resulting list, we should find results that agree with our EV and Var calculations, respectively. On the TI-84, try this: seq(rand, X, 1, 100) -> L1 . List L1 should now have 100 entries, each a random number between 0 and 1. The theoretical mean and variance of the Uniform(0, 1) distribution are 1/2 and 1/12, so the mean of L1 should be close to .5 and the standard deviation of L1 should be close to √(1/12). (After we discuss the Central Limit Theorem below, we should be able to predict just how close these estimates should be to the theoretical values!)

The next step in our progression toward ANOVA and experimental design is the concept of a sampling distribution. Revisiting our simulation in the previous paragraph, the mean we calculated can be thought of as one repetition of an experiment. This experiment has a

sample size of 100 and represents taking 100 observations from a continuous uniform distribution. By repeating this simulation again and again, we can create a new simulation of sample averages. The way this set of values is distributed is what we call the sampling distribution of the sample average. While most of our interest in statistical analyses will be about sample averages, we could have considered **any** statistic, such as the sample variance. In this case, we would calculate the sample variance of our 100 observations, and repeat that process again and again. The distribution of the resulting sample variances would represent the sampling distribution of the sample variance. In general, finding the theoretical distribution of a statistic can be a daunting task. In fact, in some cases the pdf's involved are so complicated that the formulas are intractable. However, the next result will save us in many cases, by giving us an approximate solution to the difficult theoretical result.

Regardless of what distribution data comes in, if we take a large enough sample, the distribution of the sample average should be close to a normal curve. This famous result is the Central Limit Theorem (CLT). If a distribution is close to the normal curve, then this approximation is very good even for small samples. If the distribution is skewed or very non-normal, we need larger samples before the approximation is adequate. Unfortunately, there is no magic number that will satisfy all distributions as a standard for "large". We may need to resort to some sort of simulation (e.g. bootstrapping) before we can be satisfied that our n is large enough to invoke the CLT. To use the CLT, we need to know the standard deviation of the statistic we are working with, which in most cases is the sample average, where $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$. This formula says that sample averages are better estimates of the population average than any individual measurement.

To demonstrate the CLT, try this simulation: seq(mean(randInt(1, 25, 50)), X, 1, 100) -> L1. This time, we are generating 100 repetitions of samples of size 50 from a discrete uniform distribution (picking a number between 1 and 25). The theoretical mean and variance for this discrete uniform distribution are 12.5 and 48, respectively. Therefore, the standard deviation of the mean for a sample of 50 items from this distribution should be $\sqrt{48}/\sqrt{50} = 0.98$. This L1 then should have a mean of 12.5 and a standard deviation of around 0.98. (Caution: this command will take a few minutes to complete. After all, we are generating 5,000 observations!) Make sure you look at a graph of this distribution in addition to summary statistics. You should see data that is approximately normal, despite the original distribution being uniform!

From the normal distribution, we can derive other useful distributions. The standard normal distribution squared produces the χ^2 distribution with 1 degree of freedom. If two independent χ^2 random variables are added together, the result is also a χ^2 distribution, and the degrees of freedom are the sum of the two degrees of freedom. The χ^2 distribution is most useful in describing the distribution of the sample variance, when sampling from the normal distribution.

When we do not know σ, we estimate it using the sample standard deviation, and the resulting z-type statistic, $t = \dfrac{(\overline{X} - \mu)}{s/\sqrt{n}}$, has what is known as the t distribution with $n - 1$

degrees of freedom. The formal definition of a t distribution is $T = \dfrac{Z}{\sqrt{W/df}}$, where Z is a standard normal and W is an independent χ^2 distribution with df degrees of freedom. The last major distribution is the F distribution. The formal definition of an F distribution is $F = \dfrac{W}{df1} \bigg/ \dfrac{Y}{df2}$, where W and Y are two independent χ^2 distributions with $df1$ and $df2$ degrees of freedom. Our major use of the F distribution is the ratio of two sums of squares, such as the numerators of variance.

We have two main applications of these theoretical results: confidence intervals, and hypothesis tests. In hypothesis intervals, we "guess" the value of some population fact, like an average, using a range of values based on a sample statistic. Examples of questions we can answer with confidence intervals are "What percentage of m&m's are blue?" and "What is the mean credit load for UWO students?" The usual method of forming a confidence interval is to use a reasonable sample statistic as a point estimate of the population parameter, and then add and subtract a "margin of error" from this value, producing an interval of "reasonable guesses" for the parameter. For example, for estimating μ, the population mean, we use $\bar{x} \pm t\sqrt{s/n}$. The value of t we use is based on our confidence coefficient, which represents the chance of our method producing an interval that contains the true unknown population parameter.

Our other inference procedure is hypothesis tests. In these procedures, we have some claim about a parameter that we want to test by collecting some data and calculating an appropriate test statistic. We use a proof by contradiction, where we assume the opposite of our claim is true, and then use the collected data to either support that claim or reject it. Examples of questions we can answer with hypothesis testing are "Is this die fair?", "Is the dairy cheating us by under filling our milk jugs?", and "Do these two populations have the same mean?" The usual method is to use our test statistic and calculate the chance of exceeding this value. This "P-value" represents the chance of making a Type I error, which is rejecting a true null hypothesis. If the P-value is small enough, we conclude the null hypothesis is in error, and therefore reject it in favor of the alternate, which is usually that the means in several populations are different.

Goals: Recall basic statistics and results from your introductory statistics course.

Skills:

- **Recall the prerequisites.** Remember all those things you learned in your introductory statistics course. This includes Expected Values, Variances, sampling distributions, normal curves, t distributions, confidence intervals, and hypothesis tests. Of course we will examine individual examples of these techniques as we encounter them in the days to come.

Reading: Chapter 15. (The reading assignments will always refer to the next day's reading, so you can prepare ahead of the lecture.)

Day 2

Activity: Overview of Experimental Design. Examples from Hicks.

Completely Randomized Design: In a completely randomized design, we collect samples of data from several populations to test whether the population means are equal or not. Samples do not have to be of equal size, but we do make assumptions about the variances being equal. To assign treatments to subjects, we essentially put all the names in a hat and draw them out one by one. Hicks 3.6 p 62.

Randomized Complete Block Design: In a block design, the experiment is essentially repeated within each block of subjects. Blocks are chosen so that subjects within a block are as alike as possible, and subjects between blocks are possibly quite different. The reasoning is that experimental error will be small within blocks and large between blocks, so that any within-block differences in means should be attributable to treatment differences, and not subject differences. One drawback of this type of design is that block size must be a multiple of the number of treatments, and all blocks must be of equal size. Hicks 4.1 p 81

Factorial Designs: In factorial designs, two or more factors are under study at once. Each level of each factor appears in a treatment combination with each level of each other factor. Advantages of factorial designs are efficiency: many factors can be measured at the same time; possible interactions between factors can also be studied. Disadvantages include the large number of subjects needed to fill out the design. Hicks 5.1 p 105

Nested Designs: In some designs, each of the levels of one factor is unique to one particular level of another factor. On the surface, these appear to be factorial designs, but are correctly analyzed as nested designs. Hicks 11.1 p 245

Repeated Measures Designs: Sometimes repeated measurements are taken on subjects, such as a before-treatment and after-treatment situation. These repeated measures situations may involve several factors, or several levels of randomization. KNNL 27.11 p 1167

Incomplete Block Designs: One of the drawbacks mentioned earlier in RCB designs was the need for the block sizes to be exactly equal to the number of treatments, or a multiple thereof. Sometimes structural constraints make full blocks impossible. In these cases, analyses may still be possible on smaller, incomplete blocks. Hicks 16.9 p 365

Fractional Factorial Designs: Just as incomplete block designs may be impossible to perform due to sample size constraints, full factorial designs may be impractically large. In these cases, we may be able to design a small fraction of the full factorial experiment, and still make valuable conclusions. Hicks 15.12 p 315

Goals: Differentiate between various experimental designs.

Skills:

- **Recognize various designs.** Know the structure of these seven designs. While we will work out the details of these various designs as the course proceeds, you should know from the beginning about the various differences in the way data is gathered for these studies.

- **Assign treatments.** By knowing the structure of the designs we will study, you should also know how to randomize treatments to subjects. In the case of a completely randomized

design, it is basically a "names in a hat" situation. For a more complicated design, like a nested design, we have to understand which names go in an appropriately sized hat.

Reading: Sections 16.1 to 16.4.

Day 3

Activity: Single Factor ANOVA. Extend from two-sample Math 301 type problem to the three-or-more-sample ANOVA problem. Lecture on t-tests and t-distributions.

In Math 301 you might see the two-sample t-tests and intervals. If not, I will fill in the details for you here. In this setting, we have two independent samples taken from separate populations, and we want to test the hypothesis that the two means are equal. If the data is distributed normally, then from Math 301 we know the individual sample means also have normal distributions with mean μ and variance $\dfrac{\sigma^2}{n}$. The typical procedure to test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$ is to work with the difference in the means as a test statistic.

The Expected Value of $\overline{x}_d = \overline{x}_1 - \overline{x}_2$ is $\mu_1 - \mu_2$ and the variance is $\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_1^2}{n_2}$. Because we do not know the population variances and we estimate them with sample variances, the resulting statistic is a t statistic and not a z statistic. In the one-sample case, the standardized statistic has a t distribution. However, in the two-sample case, this is not the case in general. If we are willing to assume that the two variances are equal, then the resulting statistic (called the **pooled** t statistic) does in fact have a t distribution. Pooling is not common nowadays as the computer intensive Welch's test from 1947 http://en.wikipedia.org/wiki/Welch%27s_t_test gives an acceptable approximation and has been programmed into most computer packages and calculators.

The approach we will take in ANOVA actually mirrors the pooled test procedures, so we will be interested in verifying the assumption of equal variances. The pooled t statistic has the following form: $t_{pooled} = \dfrac{\overline{x}_1 - \overline{x}_2}{s_{pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, where $s_{pooled}^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$. In class, we will focus on this pooled variance formula, as it holds the key to how to extend our work to more than two samples.

The model we will use in ANOVA is $Y_{ij} = \mu_i + \varepsilon_{ij}$. The ε terms are assumed to have a normal distribution with mean zero and variance σ^2, and the μ's are population means. Notice we are simply saying that each sample is coming from a population and the data should be normally distributed around the mean. To fit this model, we will look at some calculus results (pages 687 and 689 in the text). Essentially, all we will do is use the sample averages to estimate the population averages.

Goals: Observe that ANOVA is an extension of the two-sample pooled t-test. Define the cell means model.

- **Recognize how the *t*-test extends to two samples.** Understand that the two-sample *t*-statistic has a *t* distribution **only** if we assume equal variances. Otherwise it is only an approximation, with likely fractional degrees of freedom.

- **Know the single factor ANOVA model.** Understand the terms and assumptions in the model: $Y_{ij} = \mu_i + \varepsilon_{ij}$.

Reading: Section 16.5.

Day 4

Activity: Extend pooled variance ideas to the ANOVA equivalent: SSE. Introduce dot notation. Partition the sums of squares.

By pooling the variances for more than two samples, we can extend our results so far to the ANOVA situation. The important observation is that the numerator of this pooled variance is a sum of squared errors. We want to show that the variance of the observations without worrying about which sample they come from can be "broken down" into components that reflect the pooled variance and that reflect the effect of the treatments. This partitioning of the variance, or rather the sums of squares, is the basis for the ANOVA table included in all our analyses.

Before we can efficiently work with the various averages we will encounter, it is helpful to introduce a new notation. If we replace a subscript with a •, we can indicate that a sum has been taken over the range of that subscript's values. Thus, $Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}$, $Y_{\bullet j} = \sum_{i=1}^{r} Y_{ij}$, and.

$Y_{\bullet\bullet} = \sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij} = \sum_{i=1}^{r} Y_{i\bullet}$. If we add an overbar, we indicate an average, dividing by the

appropriate sample size. So, $\overline{Y}_{i\bullet} = \dfrac{\sum_{j=1}^{n_i} Y_{ij}}{n_i}$ and $\overline{Y}_{\bullet\bullet} = \dfrac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}}{\sum_{i=1}^{r} n_i}$. At first this notation may seem

cumbersome, but with practice you will see its utility and effectiveness.

Using our new dot notation, we now want to show that the numerator of the pooled variance is the SSE given on page 691. We will also look at the breakdown of SSTO on page 692. Along with being able to break down the sums of squares, we are also able to divide up the degrees of freedom associated with each sum of squares. Because we know that deviations from average always add up to zero, we effectively "lose" one piece of information; i.e. if we know all but one of the errors or residuals, we know the last one. This is the reasoning and justification for calling this fact a "degree of freedom".

We summarize our breakdown of the sums of squares and the degrees of freedom in a chart known as an ANOVA table. In this table, we have columns for the sums of squares, the degrees of freedom, the Mean Squares (MS), and test statistics and p-values, as appropriate. The MS's are simply the ratio of the SS to their df, and in the language of our work on Day

1 represent random variables with χ^2 distributions. The ratio of two appropriate MS's is then an F distribution. Deciding which null hypothesis each of these F statistics tests is one of our challenges in the upcoming material.

Goals: Model estimation. Dot notation algebra. ANOVA table facts.

Skills:

- **Use dot notation.** To simplify our algebra (but sometimes requiring closer attention to detail) we use dots (•) in place of subscripts to indicate a summation. We also add overbars to indicate averages.

- **Decompose sums of squares.** Understand algebraically why the sums of squares decompose (add up with no cross products). This understanding hinges on the fact that deviations from average add to zero, when summed over appropriate subscripts. We also observe that the degrees of freedom similarly decompose.

Reading: Sections 16.6 and 16.7.

Day 5

Activity: F-Test. Simulation. Alternative model.

One of the theoretical features we will use concerning ANOVA tables is the Expected Mean Squares. The derivation of these expectations is beyond the scope of our course (you need Math 401 knowledge) but will we examine some algorithms designed to allow us to make use of the results. One basic fact that we **can** prove at this point is the unbiasedness of MSE as an estimator for σ^2. I will show you a proof of this today, which elaborates on a comment on page 696. The authors, on pages 697 and 698, detail the derivation for E(MSTR). At this point, we are more interested in the form of the result than the details of the derivation.

Focusing on E(MSTR), we see that when the means of the populations are all equal, the E(MSTR) = E(MSE). This forms the justification of our F-test. The null hypothesis then is $H_0 : \mu_1 = \mu_2 = ... = \mu_r$ and the alternate hypothesis is that not all means are equal. Note that this alternate hypothesis is difficult to specify exactly; there are **many** combinations of population means that satisfy the alternate hypothesis, but only one way to satisfy the null hypothesis. This multiplicity of alternate hypothesis values makes power calculations quite complicated. We will explore power on Day 6.

At this point, it is not a bad idea to conduct a simulation to see just how ANOVA works. I will generate data using either MINITAB or Excel, and we will see the various theoretical facts at work. In particular, we can see how MSE, MSTR, and F are distributed.

So far we have been considering the cell means model, as our parameters have been the means of the individual populations. An alternative but equivalent model is the factor effects model. (This model will more closely parallel the regression model that you saw in Math 385, for those who took that course.) The factor effects model is needed for our more complicated future models, so it's important for us to understand it early on. The key fact is that the models are equivalent; they lead to the same fitted values, and the same

hypothesis tests. The only difference in the two approaches is how we phrase the parameters.

Derive F-test for equal means.

Skills:

- **Perform ANOVA tests.** Know the details of the *F*-test from the ANOVA table. in particular, you should understand the null hypothesis being tested, and that the alternate hypothesis is quite complicated.

- **Understand the Factor Effects model.** The cell means model uses a population mean for each individual population. The factor effects model uses a deviation from average for each population. Fitted values and hypothesis tests are equivalent between the two models, so which one we use will be a matter of convenience in interpretations.

Reading: Sections 16.10 and 16.11.

Day 6

Activity: Details of Calculation. Power.

Today we will use MINITAB to calculate an ANOVA, including the *F*-test decision. In addition, I want to use our simulation from Day 5 to explore power and Type II errors. The theoretical work is complicated. Most practitioners use prepared tables instead of calculations from scratch. We will take a look at both approaches.

The complication comes when we examine the distribution of the test statistic when the alternate hypothesis is true instead of the null hypothesis. In this situation, a new distribution results, called the non-central *F* distribution. Because of the complexity of the alternate hypothesis, this distribution is quite complicated, and using it can be a burden.

Goals: Calculate an ANOVA. Understand Type II error calculations.

Skills:

- **Know ANOVA formulas.** While we may often use a computer to perform ANOVA, we should realize the formulas are simple enough to understand and use without powerful computing. For example, just knowing means and standard deviations of the various sub-samples is sufficient to produce the entire ANOVA table.

- **Power calculations.** In ANOVA, the distribution under the alternate hypothesis is a complicated *non-central F* distribution. We must use tabulated values to calculate tail probabilities. Power calculations, and sample size determination are typical uses of this distribution.

Reading: Sections 17.1 to 17.3

Day 7

Activity: Contrasts and Linear Combinations. **Homework 1 due today.**

Once we have rejected a null hypothesis of equal means, we usually want to know which means differ from the others. We accomplish this by examining various differences of means, or linear combinations of means. If the weights in a linear combination sum to zero, we call that linear combination a contrast. In all cases, our rules for means and variances help us develop the test needed.

Specifically, if $L = \sum_i c_i \mu_i$ then we estimate L using appropriate averages: $\hat{L} = \sum_i c_i \overline{Y}_{i\bullet}$. The mean and variance of this estimate are L and $\sum_i c_i^2 \dfrac{\sigma^2}{n_i} = \sigma^2 \sum_i \dfrac{c_i^2}{n_i}$, respectively. If we use MSE as our estimate for σ^2, we have the basis for t intervals and tests.

The typical contrasts one examines are often pairwise comparisons between the sample means. With r treatments, there are $\binom{r}{2}$ pairwise comparisons between sample means. There may be other comparisons that make sense, such as when one of the treatments is control group, and the researcher wants to compare each treatment level with the control.

Goals: Introduce contrasts and linear combinations.

Skills:

- **Expected value and variance formulas.** Recall the Expected Value and Variance formulas. To calculate confidence intervals and tests for various combinations of the factor means, we need to rely on our expected value and variance formulas. The key result is that means and variances for sums are additive for independent samples.

- **Recognize contrasts and know their means and variances.** Typical contrasts involve comparing the average of several treatment means to a control mean, or grouping several treatment means together.

Reading: Sections 17.4 and 17.5.

<center>Day 8</center>

Activity: Multiple comparisons I.

When we make multiple tests or confidence intervals, we need to consider the chance of **all** such statements being simultaneously true. The individual success rates will not be accurate if many tests are made. For example, if 100 hypothesis tests are conducted, and if the null hypothesis is true in every case, we would expect 5 of the confidence intervals to be incorrect, just by chance. And yet we would not want to tell a client that we're pretty sure five of the 100 statements we are making are not correct, but we don't know which five. Instead, we need a procedure that ensures that there is a 95% chance that **all** of our 100 statements are simultaneously true. This issue is called multiple comparisons.

The first procedure we will look at is the Tukey procedure. This procedure is only appropriate when the contrasts we are interested in are the pairwise comparison of sample means. The procedure is based on the distribution of the difference between the largest and smallest values taken from a normal distribution. Table B9 is based on this Studentized

Range distribution. Essentially we construct tests and confidence intervals using formulas from Day 7 for our standard errors, but with a different multiplier in place of t.

Goals: Introduce the need for multiple comparisons. Tukey's Procedure.

Skills:

- **Understand the issue of multiple comparisons.** When we make multiple tests or confidence intervals, we do not have faith that all of our statements are true with the same confidence we have that a single statement is true. To adjust for this issue, we will use techniques that account for multiple statements.

- **Know Tukey's Procedure.** Tukey's procedure may only be used when we are comparing means in a pairwise manner. There are $\binom{r}{2}$ possible pairwise comparisons.

Reading: Sections 17.6 and 17.7.

<div align="center">Day 9</div>

Activity: Multiple comparisons II.

One significant drawback to Tukey's procedure is that the only contrasts or linear combinations we can discuss are the pairwise comparisons between treatment means. Today's procedures both allow us to test other contrasts and combinations. Bonferroni's is by far the simplest multiple comparison procedure to use, but Scheffé's is the most flexible.

Scheffé's procedure is used when we are making a large number of tests or statements. The idea is that the overall F-test will indicate if there is **any** contrast that is significant. If the F-test is not significant, then no contrast will be found to be significant. Therefore, this procedure allows us to data snoop, that is to look through all possible contrasts and find the interesting or significant ones. However, it is often the case that the contrasts of interest do not include the ones that happen to be causing the F-test to be significant. Therefore the significance levels or confidence levels tend to be conservative, and we can relax the levels when using Scheffé's procedure.

Bonferroni's procedure is based on the Bonferroni inequality: $P(\overline{A}_1 \cap \overline{A}_2) \geq 1 - P(\overline{A}_1) - P(\overline{A}_2)$. Essentially, we will divvy up our alpha level into however many statements or tests we are making. Of course, the larger the confidence level, the wider the confidence interval, so the effect of using Bonferroni's procedure is that the intervals get wider as more statements are made. Therefore, we should only use Bonferroni's procedure if we have a small number of statements of interest. If we have a large number of statements, either Tukey's procedure (for the case of pairwise comparisons) or Scheffé's procedure should yield narrower intervals. It is completely appropriate to calculate all three multipliers (if applicable) and use the smallest one.

We can understand these procedures a little better if you expand on our simulation from Day 5. I will add several confidence intervals and hypothesis tests to each sample, and we will keep track of the success rates involved.

Goals: Scheffé's Procedure, Bonferroni's Procedure

Skills:

- **Know Scheffé's Procedure.** Scheffé's procedure involves a "gateway" F-test; then any contrasts or means comparisons are valid. Scheffé's procedure is most efficient when we are making a large number of statements or tests, or when we do not know ahead of time what contrasts might be of interest (data snooping is allowed).

- **Know Bonferroni Procedure.** The Bonferroni procedure is best suited for a small number of pre-selected statements or tests. This procedure divides up the significance level among all the statements or tests desired. Then by the Bonferroni inequality the family confidence coefficient or significance level is controlled.

Reading: Sections 18.1 to 18.2.

<div align="center">

Day 10

</div>

Activity: Diagnostics. Hicks p 46.

We have assumptions about the error terms in the model that need to be checked before we can certify that our model is an adequate description of our data. In particular, we assume the error terms follow a normal distribution, with constant variance, independent of other error terms.

To test whether the data follow the normal distribution, we can use Table B.6. We calculate the correlation of the residuals to their expected value under normality (the NSCO command in MINITAB) and refer to the entry in Table B.6. If the correlation is greater than the tabled value, we conclude the assumption of normality is acceptable.

To test for independence of the error terms, we can plot the residuals against time, or some other logical sequence of measurements, such as a geographic sequence. If a trend or pattern is apparent in these plots, we have an indication that the errors are not independent, at least for that variable or sequence.

To test for constancy of variance (homoskedasticity) we can use either the Hartley test or the Brown-Forsythe test. The Hartley test is only appropriate for equal size samples, while the Brown-Forsythe test is always appropriate.

The Hartley test compares the largest sample variance to the smallest sample variance in a ratio. If this ratio is large, we have evidence that the variances are not equal. Table B.10 contains the percentiles of the Hartley test statistic.

The Brown-Forsythe test examines the absolute size of the residuals around the sample medians. If the error variances are constant, these absolute deviations should all be equal. The appropriate test is then an ANOVA on the absolute deviations.

Goals: Residual plots. Hartley test. Brown-Forsythe test.

- **Know how residual plots are used to assess model adequacy.** The residuals should have the same distribution as the error terms, so making plots of them and analyzing them will give us some clues as to whether the model we've chosen is adequate. Specifically, the use of the normal probability plot will let us know about the normality assumption, and side-by-side box plots will help assess the equal variance hypothesis.

- **Know how to use the Hartley test for equal variances.** One simple formal test for equal variances is the Hartley test. There is a special table (B.10) of the percentiles from this distribution in our text. The test statistic is the ratio of the largest sample variance to the smallest sample variance. Clearly, large values are evidence of differing variances.

- **Know how to use the Brown-Forsythe test for equal variances.** Samples with equal variances will produce residuals that are essentially all of the same size. The Brown-Forsythe test statistic measures whether the absolute deviations of the deviations from the medians are equal.

Reading: Sections 18.5 to 18.7.

Day 11

Activity: Remedial measures.

Goals: Transformations. Nonparametric test.

Often we can transform our response variable, Y, to yield a model that more closely matches our assumptions on the error terms. Based on Taylor's series expansions for the variance formulas, we can use the following transformations for common situations where the variances are related to the means for the various populations. Variance proportional to the mean: Use square root. Standard deviation proportional to the mean: Use logarithm. Standard deviation proportional to the mean squared: Use reciprocal. Response is a proportion: Use arcsin. Pages 789 and 790 in our text have the details of the transformations.

Another approach to transformations is to use the Box-Cox procedure. After an appropriate transformation (equation 3.36) we use different powers of the response variable to find the transformation that yields the smallest SSE when fitting an ANOVA. I have a spreadsheet on TitanFiles that you can use as a template for this procedure.

If we are unable to correct the model's deficiencies with transformations, another option available to us is to use a non-parametric procedure. These procedures involve using techniques that are not dependent on the distribution of the error terms being normal. Often, non-parametric techniques will use ranks in place of raw data. Then the ranks are analyzed instead of the original data values. This is the basic idea in the Kruskal-Wallis test, which is equivalent to the rank F-test statistic.

Skills:

- **Know which transformations may help with unequal variances.** When the population variances or standard deviations are proportional to the population means, some common

transformations are square root, logarithm, and the reciprocal. The more involved Box-Cox transformation may help too, but could also be more trouble than it is worth, considering the resulting power may not be very interpretable, i.e. $Y_{ij}^{0.4}$ as opposed to square root.

• **Know how the Kruskal-Wallis test works.** If we replace the data with the ranks (from one to n, the total sample size) before doing ANOVA, we essentially have the Kruskal-Wallis test. This test is most useful when the normality assumption doesn't hold.

Reading: Chapters 15 to 18.

<div align="center">Day 12</div>

Activity: Review. **Homework 2 due today.**

Reading: Sections 19.1 to 19.3.

<div align="center">Day 13</div>

Activity: Exam 1.

Know everything about one-way ANOVA.

<div align="center">Day 14</div>

Activity: Two factor ANOVA with replicates and equal sample sizes.

Today we begin ANOVA with more than one factor present. Some of the calculations we make will be identical to one-way ANOVA. However, there is a new profoundly different concept: interaction. The difference between the true cell means and the cell means using an additive model is our definition of the interaction terms. We will examine situations today with and without this additive structure.

One common tool used to understand two-factor ANOVA is the Treatment Means plot. This plot will visually give us information on the existence of interaction. We must be careful though to remember the idea of variability when analyzing these plots. The text ignores this issue, so I want to make sure we keep it in mind. The main idea is that if interaction is not present, the lines in a Treatment Means plot will be parallel. In practice, however, we will not know the true means, and we will be plotting cell means instead. These are of course only estimates, and so carry with them a notion of imprecision. By attaching error bars to our estimates, we can better assess parallelness.

We can analyze our two-factor ANOVA using either a cell means model or a factor effects model. The cell means model will not, however, account for the additive structure, if present. We may often prefer the factor effects model then. The estimates we use for the model components will be quite similar to the one-way ANOVA estimates. The notable exception is the interaction term.

Goals: Model and fit.

- **Recognize interaction.** With two factors, we have a new term in the model: interaction. Our examples in class show that we can add an extra term to account for this new idea. Plots are especially useful for characterizing the nature of the interactions.

- **Understand the two models; differences and similarities.** The cell means model and factor effects models have different terms (parameters) but yield the same cell estimates and thus the same ANOVA table. The **interpretation** is what differs.

- **State model and fits.** Appropriate averages will be our estimators for the additive terms in the model. Pooled variance will be our estimate of variance, once again.

Reading: Sections 19.4 to 19.7.

Day 15

Activity: Two-factor ANOVA with replicates.

Goals: ANOVA and test.

Body

Skills:

- **Be able to decompose the SSTR into its components.** Balance is important to the orthogonal decomposition of the sums of squares.

- **Know F-tests for 2-way ANOVA.** With the addition of an interaction term, we have a new test we can perform, and which should be done before we examine main effects.

Reading: Sections 19.8 to 19.10.

Day 16

Activity: Two factor multiple comparisons.

Goals: Understand the issues of multiple comparisons in two-factor studies.

Body

Skills:

- **Tukey's method.** Because Tukey's method deals with the distribution of means in multiple samples, there is no way for us to deal with the two factors at once. What is usually done then is to perform **two** Tukey's tests and combine the results using a Bonferroni approach. For example, if we desire a family $\alpha = 0.05$ level, we can use an $\alpha = 0.025$ level for each of the two factors, and we can then compare all possible mean differences separately for each factor.

- **Scheffé's method.** To make statements about **both** factors, we can modify the Scheffé's method by using an appropriate numerator degrees of freedom; then any statements we would like to make are valid. On the other hand, the Bonferroni approach for Tukey's an

also be used here. As was the case for the one-way ANOVA, it is acceptable to perform "data-snooping" with this technique.

- **Bonferroni's method.** Everything we learned from one-factor ANOVA is still true for two-factor ANOVA. Specifically, if we make g statements, we can be jointly confident of the truth of our statements by dividing up the alpha level into g pieces, usually equal, but not necessarily equal. This technique allows us to make statements about either factor, or about interaction terms, as long as the statements, tests, contrasts, etc are specified beforehand.

Reading: Chapter 20

Day 17

Activity: Two factor ANOVA with no replicates.

Goals: Tukey test for additivity.

Body

Skills:
- **Interaction is used as error when there are no within-cell replicates.** When the sample size in each cell is only 1, we have zero degrees of freedom left to estimate the population variance σ^2. Our only recourse is to then assume that no interaction exists and use that variation as an estimate for σ^2.

- **Tukey's one degree of freedom test for additivity can be used to detect interaction.** If we model the interaction as a multiplicative effect, then one degree of freedom can be taken from the error term and used as a test for this sort of interaction. While this will not handle all sorts of interaction, it is at least **something** that we can do other than to just assume interaction is not there.

Reading: Sections 21.1 to 21.4.

Day 18

Activity: Randomized Complete Blocks (RCB's). **Homework 3 due today.**

Goals: Randomization.

Body

Skills:
- **Understand the definition of RCB's.** In an RCB we assign each treatment to n subjects, and repeat this arrangement b times. In structure then, the data appear as in Chapter 20 as a two-way ANOVA without replication. Each block is a separate repetition of the whole experiment. Blocks are chosen to create more homogenous groups of subjects.

- **Know how to interpret the ANOVA table.** In the ANOVA table, we generally are not interested interesting for block effects. In fact, blocks are generally chosen specifically

because we know or suspect that they are contributing to the sums of squares. It may be of interest though to know if interaction exists, and we can still perform the Tukey's one degree of freedom test for additivity.

- **Know the mechanics of actually assigning treatments to subjects.** The simplest way to assign treatment to experimental unit is to generate a random number (like rand from the TI-83) for each item in the block. The smallest of these n random numbers gets treatment 1, the next smallest treatment 2, etc.

Day 19

Activity: RCB's continued using MINITAB. Consider empty cells.

Goals: Use MINITAB to perform an analysis. Understand the Friedman test, a non-parametric alternative.

Body

Skills:

- **The MINITAB command is still a two-way ANOVA.** When there are no replicates, we can still use two-way ANOVA to perform the calculations.

- **Know a way to deal with missing data.** When a cell is empty, we can use calculus to estimate a value so that the sums of squares are not unduly influenced by not having data in that cell.

- **Know the Friedman test details.** This non-parametric test is similar to the Kruskal-Wallis procedure, but we rank each block separately. Thus we have absolutely no ability to detect block differences with this procedure.

Day 20

Activity: Two factor ANOVA with unequal sample sizes.

Goals: Regression approach.

Body

Skills:

- **Understand the lack of orthogonality.** When the sample sizes are unequal, the sums of squares for treatments do not decompose into orthogonal components, as in balanced two-way ANOVA, or with one-way ANOVA. Therefore to do any tests we must use a reduced model regression approach. Basically we will compare SSE's from different models.

- **Know how to use and interpret MINITAB output.** The procedure GLM will produce the output we need. The SSE from the full model is subtracted from the reduced model SSE to

get an adjusted SS. This value is what is reported in MINITAB. Our tests proceed as usual.

- **Use Reduced Model Tests to deal with missing cells.** When a cell is missing data, so that the sample size is zero for that cell, we can use the MINITAB command GLM to conduct our test. No new difficulties result.

Day 21

Activity: Two factor ANOVA with unequal sample sizes multiple comparisons. **Homework 4 due today.**

Goals: Multiple comparison and empty cells.

Body

Skills:

- **Apply the multiple comparison techniques to unequal sample sizes.** The fitting of the model uses our standard averaging techniques. The multiple comparison techniques we have already encountered then are applicable with little modification. It **is** important, however, to pay attention to sample sizes and Expected Value and Variance formulas.

Day 22

Activity: Multi factor models.

Goals: Model and fits.

Body

Skills:

- **Realize that no new problems exist when adding additional factors.** When we have three or more factors, we proceed in the same way as before. Sums of squares are calculated based on orthogonality in the case of balanced designs, and using GLM in the case of unbalanced designs.

- **Understand that the new higher order interactions are very difficult to interpret.** With three factors, we have three main effects, three two-way interactions, and one three-way interaction. With four factors, we have 4 main effects, 6 2-way interactions, 4 3-way interactions, and a 4-way interaction. Due to the complicated nature of these interactions, most analysts will opt to assume they do not exist; this is analogous to using a model with fewer terms to avoid "overfitting" the model.

<div align="center">Day 23</div>

Body

Skills:

- **Skill.** Description

<div align="center">Day 24</div>

Body

Skills:

- **Identify the Random Model.** Instead of adding terms that are constants (parameters) the random model consists of terms that have random distributions. The analysis proceeds as usual, but the EMS terms differ.

- **Understand the MINITAB displays.** The analysis of random models is very similar to fixed models, but there are a few differences we need to recognize. Specifically, the way we estimate variances changes.

<div align="center">Day 25</div>

Body

Skills:

- **Know the details of the interactions and fixed terms in a mixed model.** When we have both fixed and random terms in a model, some thought must be put into the assumptions of the model. For example, which terms have a variance term instead of a sum of parameters squared in the EMS terms? The tests we perform to test the various null hypotheses will vary according to which terms are random; the EMS column is our guide to these tests.

<div align="center">**Day 26**</div>

Activity: Review. **Homework 5 due today.**

Reading: Appendix D.

<div align="center">**Day 27**</div>

Activity: Exam 2.

Know everything.

<div align="center">**Day 28**</div>

Activity: EMS rules.

Goals: Look at examples from Appendix D concerning model creation.

Body

Skills:
- **Become familiar with the Appendix D rules.** A set of rules is given in Appendix D that allow us to formulate models, calculate sums of squares, and derive EMS terms.

Reading: Chapter 26

<div align="center">**Day 29**</div>

Activity: Nested designs.

Goals: Understand the nuances of the nested model.

Body

Skills:
- **Understand how nesting changes our models and tests.** When a factor is nested within another factor, our model changes drastically; we have more parameters, and the ANOVA and EMS's change as well. We have more parameters because each level of the nested variable is replicated for **each** level of the factor it is nested within.

- **Know how to use the EMS column to select the proper test.** By specifying which terms are random and which are fixed, we can select a proper test for each potential null hypothesis. We are looking for terms in the EMS equations that are exactly the same except for the term involving the null hypothesis parameter(s).

Reading: Chapter 27

<div align="center">**Day 30**</div>

Activity: Repeated measures.

Goals: Distinguish and identify a repeated measures layout.

            Body

Skills:

  •   **Understand how a repeated measures model is an example of an RCB design.** Each subject is exposed to each treatment; in this sense then each subject is a block. Usually subjects are thought to be a random sample from the population of all subjects, so this block variable is typically considered a random block effect.

Reading: Hicks Chapter 13

<p align="center" style="color:red">Day 31</p>

Activity: Split plots.

Goals: Recognize when a split plot design is appropriate.

            Body

Skills:

  •   **Identify factors that cannot be randomized within a block.** Sometimes the nature of an experiment makes it impossible or infeasible to use more than one level of a factor within a block. This situation leads to the idea of whole plots and split plots. Our chief examples will be the irrigation example and the baking temperatures example.

  •   **Recognize how this split plot design is an example of a repeated measures design.** By looking carefully at the components of the model, we should be able to fit this split plot methodology into our repeated measures methodology. Tests and analyses will follow normally.

Reading: Sections 28.1 and 28.2.

<p align="center" style="color:red">Day 32</p>

Activity: BIBD's. **Homework 6 due today.**

Goals: Construct BIBD's. Notice that some combinations cannot be formed, and that some combinations are simply duplicates of smaller designs.

            BIBD's:

1) $r_b * n_b = r * n$. Note that this value is the total number of experimental units in a BIBD.

2) $n_p = n (r_b - 1)/(n - 1)$, where r_b is the number of treatments per block, n_b is the number of blocks, r is the number of treatments, n is the number of replications of each treatment, and n_p is the number of times each **pair** of treatments appears together.

- **Recognize a Balanced Incomplete Block Design.** BIBD's are used when the block size in an experiment is smaller than the number of treatments. We call them **balanced** because each treatment appears the same number of times throughout the experiment. We call them **incomplete** because not **all** treatments appear in each block. They are somewhat restrictive because not all combinations of the indices exist.

- **Know the relationships among the indices of a BIBD.** We have two equations that define our BIBD's. 1) $r_b * n_b = r * n$, and 2) $n_p = n (r_b - 1)/(n - 1)$.

Reading: Hicks Chapter 16.

Day 33

Activity: BIBD's.

Goals: Analyze a BIBD using MINITAB.

Skills:

- **Know the appropriate MINITAB commands to analyze a BIBD.** Because the BIBD is a linear model in the parameters, we can use GLM from MINITAB to analyze it. There are no additional complications.

Reading: Sections 28.3 to 28.5.

Day 34

Activity: Latin Squares.

Goals: Understand Latin Square designs.

Body

Skills:

- **Recognize a Latin Square Design.** In a Latin Square design, there are **two** blocking variables, in addition to the treatment variable. Each variable, blocking or treatment, has the same number of levels. (These designs thus may be more appropriately called Latin Cubes.) For three treatments, one design exists; for larger numbers of treatments, many designs exist. I have not found a simple way to enumerate them.

- **Know the appropriate MINITAB commands to analyze a Latin Square.** As in BIBD's, Latin Squares are simply linear models, with missing (empty) cells. The GLM command in MINITAB will produce estimates and tests for us.

Reading: Sections 29.1 to 29.3.

Day 35

Activity: 2^f Factorials.

Become familiar with factorial design notation.

Body

- **Know the notation when all factors have just two levels.** When all factors have just two levels, the notation for designating which treatment combinations are present is abbreviated. Each letter will now indicate whether a factor is at the high level or not. A missing letter designates the factor at low level. Thus the code "ab" in a 3 factor study means Factors A and B are at high level, and Factor C is at low level. (1) is the code to indicate all factors at low level.

Hicks Chapter 15.

Day 36

3^f Factorials. **Homework 7 due today.**

Understand the effects of running just a fraction of the treatment combinations.

Body

- **Realize the benefits and drawbacks of not running all treatments.** When we have physical restraints that prevent us from performing all treatment combinations in an experiment, we may wish to run just a fraction of the combinations. This "fractional factorial" will typically suffer from some confounding. Thus, we may only run half of our treatment combinations and will therefore be unable to estimate uniquely half of our parameters. If these parameters are chosen wisely, we may still be able to estimate all factor effects; typically some interaction terms will have to be assumed to be absent.

- **Recognize what effects have been confounded.** In a fractional factorial, the columns of the design matrix that are identical (or off by a minus sign) indicate which effects are confounded with the effect (1), the intercept term. Once this confounded effect (or effects) has been identified, we can complete the entire confounding scheme and see if any main effects have been confounded with each other, an undesirable consequence. If the confounded effects are chosen wisely, we will be able to estimate all main effects uniquely.

Hicks Chapter 15.

Day 37

Fractional Factorials.

Understand how experiments can be blocked with confounded effects.

Body

- **Identify which effects in a fractional factorial are confounded with blocks.** If we let each block in a blocked experiment be a fractional factorial, then those effects will be confounded with block effects. By appropriate choices of interaction terms to be confounded, we can still recover enough useful information to make the design feasible.

Reading: Hicks Chapter 15.

Day 38

Activity: Fractional Factorials.

Goals:

Body

Skills:

- **Skill.** Description.

- **Skill.** Description.

Reading: Hicks Chapter 14.

Day 39

Activity: Confounding.

Body.

Goals:

Skills:

- **Skill.** Description.

- **Skill.** Description.

Reading: Hicks Chapter 14.

Day 40

Activity: Confounding.

Body.

Goals:

Skills:

- **Skill.** Description.

- **Skill.** Description.

Reading: Chapters 26 to 29, Hicks Chapters 13 to 16.

<p style="text-align:center;">Day 41</p>

Activity: Review. **Homework 8 due today.**

<p style="text-align:center;">Day 42</p>

Activity: Exam 3.

Know everything.

<p style="text-align:center;"><a>Return to Chris' Homepage</p>

<a>Return to UW Oshkosh Homepage

Managed by: <a>chris edwards
Last updated January 10, 2012