**Handout#5: Evaluation of Model Conditions**

Model Conditions: model . We impose the condition that . This condition implied three things

(C1) The t treatment pops have a normal population, (C2) The t treatment pops. have the same sd (C3) Experiments are conducted so that the obs. are independent

* If the conds. are violated, then he inferences we make would yield invalid p-values, tests with incorrect powers and CIs that are either too wide or too narrow.

Evaluation of Normality Condition: **Case1:** if each of the are large, we can run the SW test for each of the t random samples from the t treatment pops. **Rarely occurs**.

* **Case 2 (one or more of the is small)**: its not possible to obtain meaning full normal probability plots or SW tests. Can't apply tests of normality to all n data values because we don’t know if there's a difference in the treatment means… if there is, even if each of the treatment pops is Normal, the pooled data would be multimodal.
* Thus, when the are small, we examine the sample residuals from the fitted model

Properties of the Residuals: (1) (2) (3) are correlated obs. from the same treatment sample.

Evaluation of Normality Using the Sample Residuals: (1) SW-test; (2) Construct QQ plots; (3) Construct box plots; (4) When samp. sizes are unequal, the results are somewhat affected by the unequal variances, but only minimally; (5) The slight correlation in the residuals has a slight affect on the sensitivity of the SW-test; (6) To overcome the problems we can standardize the residuals where ; ; (7) In order to detect if an observation we declare

Evaluation of Equal Variance Condition: . Violations of equal variance conditions effect our inferences greater than moderate deviations from normality. Although the AOV F-test is robust to moderate deviations from constant variances **provided** we have equal sample sizes.

Brown-Forsythe-Levene Test of Homogeneity of Variance: **NOTE:** **requires .** For a specified value of α, reject if .

Transformations of the data to obtain constant variance: (1) Power transformations: If reg. is valid use trans . -> log.

(2) Box-Cox Transformation: used to transform data to normality. **NOTE:** in both cases, the have approx. equal vars. and we can then conduct the AOV F-test and multiple comparison procedures. **IMPORTANT:** the tests and confidence intervals are being constructed using the transformed data, . Thus we are making inferences about and not about . Ex. we're testing when in fact we want to test . **If** the t population variances are equal; **then** the two tests are equivalent. E.g. testing if the treatment means of the transformed data are equal is equivalent to testing if the treatment means of the o.g. are equal.

A Distribution-Free AOV Procedure: **Kruskal-Wallis Test:** A generalization of the Wilcoxon-Rank Sum procedure. *It retains all the conditions required of the AOV F-test except the Normality condition.* The conditions are equivalent to the following model: where is the overall median.

* B/c the KW-test is appropriate for any cdf G, whereas the AOV F-test is just for normally distributed responses, the KW-test **tends to be less powerful** than the AOV F-test when the data is normally distributed.
* The KW-procedure: (4) The distribution of KW, under H0 doesn’t depend on the distribution of (5) We reject at level where are given in table A.12. (6) A large sample approximation: Reject at level **(we almost always use this)**.

Multiple Comparison Procedures using Ranks:

* Hollander-Wolfe: declare if where is the critical value from the KW-test.
* Miller-Large sample approximation for equal sample size (Has Low Power): declare where can be found in table VII
* Dunn – Large Sample approx. when unequal sample size (Has Low Power): declare where

Generalized Linear Models: in generalized linear models, the response is assumed to possess a probability distribution of the exponential form

* Log-Likelihood Functions: Log-likelihood functions for the distributions that are available in GENMOD are parameterized in terms of the means and the **dispersion parameter** . The log-likelihood functions are of the form:
* Over-dispersion: Over-dispersion is a phenomenon that sometimes occurs in data that are modelled with binomial or Poisson distributions. If the estimate of dispersion after fitting, as measured by the deviance or Pearson's chi-squared, divided by degrees of freedom is not near 1, then the data may be over-dispersed if the dispersion estimate is >1 or under-dispersed if the dispersion estimate is <1. A simple way to model this situation is to allow the variance functions of these distributions to have a multiplicative over-dispersion factor . Binomial: Poisson: **NOTE:** Param. estimates aren't affected by the value of
* Goodness of Fit: The deviance of Poisson regression model is defined to be where is the maximum of the log likelihood for the model and is the saturated model (where there is a parameter for every observation). The **deviance** statistics provides us with a test of model fit.

Tests For Correlation in the Residuals: When we have **positive** correlation in the data, the inferences in the AOV F-test and multiple comparisons procedures have a dramatic increase in the size of the test, that it, the Type I error rate may be much larger than the nominal value. This also causes an increase in the power, but the increase in the size of the test (type I error rate) negates the positive aspect of an increase in power. Cis for treatment means and model parameters with have a coverage probability less than the stated value.

* **Dubin Watson Test**: requires that the residuals have a normal distribution. Test STAT:
* One sided test of -> if (k = # of treatments)
* **Runs Test for Correlation**: (1) Center the observations: (2) Count the number of runs (R), where a run is defined as a sequence of observations of all positive values or all negative values. (3) Count the number of positive 's (n1) and the number of negative 's (n2). (4) When n1 and n2 ≤ 20, we can use the following decision rule (a) positive correlation if R ≤ RL (b) negative correlation if R ≥ RU (c) indeterminant if RL ≤ R ≤ RU
* (5): Large sample size approx: where . Say data correlated if
* NOTE: correlation is affected by outliers

Impact of Correlated Data on Inference Procedures: if correlation > 0 -> this results in Cis that are two narrow and hence our coverage probability is less than the stated value. Also, the F-test statistic is too large in comparison to the ratio if the correct value for the estimated variance was used in the denominator. Thus the type 1 error rate is inflated. However, the power of the test is also inflated but this gain in power is paid for by an inflated Type I error rate.

**Handout#6: Experiments Investigating Variance Components**

Random Effects: A factor has **random effects** if the levels of the factor to be used in the experiment or study are randomly selected from a population of potential levels or the levels have varying experimental conditions.

Model for Random Effects: A random effects experiment has the following elements (1) Randomly select t treatments (factor levels) from a population of all treatments of interest. (2) Randomly assign ni experimental units (EU's) to the *i*th treatment or randomly select ri EU's from the ith treatment population. (3) Observe or record

* with ; thus are fixed pop. Params.
* To test if there are treatment differences **in the population of treatments (not just ones observed in the study)** we test

Random Effects Model: ; **(1)** independently of ; **(2)** but conditional on the selected treatments we have ; **(3)** ; **(4)** ; **(5)** For j≠h,-> the observations from a given treatment are positively correlated; **(6)** for j ≠ h and i≠k, -> observations from different treatments are independent; **(7)** The intraclass correlation coefficient this is the correlation between two observations receiving the same treatment.

How do the Random Effects and Fixed Effects models Differ: **(1)** In Fixed effects there are only t treatments of interest, in the random effects there is a population of treatments from which we select t for use in the experiment. We are interested in making inferences concerning the population of treatments not just the t treatments used in the experiment. **(2)** In fixed effects, the treatment effects are fixed population parameters. Thus; equivalent to . In Random Effects the treatment effects are random variables with reflecting the difference in the treatment means across all treatments in the population of treatments. Thus is equivalent to . **(3)** In the Fixed effects model, the observations are independent RVs. In the Random Effects model the observations from the same treatment are correlated.

Estimation of Variance Components: In the random effects model: there are **two components to**

* Point Estimation – Using MOM (Type3 in SAS): Partition . will use the same formulas that were used in the FIXED EFFECT Model:

using MOM:

* **One of the problems with MOM estimator** of is that it could be negative, in that case, we use MLE or REML to estimate . *The moment matching estimators from AOV should not be used when the data is unbalanced or when one or more of the moment matching estimators from AOV are negative.*

Confidence Intervals for Variance Components: (1) 100(1-α)% CI for : (2) Conservative 100(1- α)% CI for

* Where ; ; . **NOTE:** interval (2) requires equal sample size.

**Test of:** ; where

* Decision Rule: Reject

Satterthwaite df (Using Method = REML in SAS):

Attributing Variation to Treatments: Proportion Due to Treatment Diffs =

CI's and Hypothesis Tests when **Unequal Sample Sizes**: (under we have the same test statistic and rejection region). However, with unequal sample sizes, under F no longer has a central-F distribution (it is unknown, and the power function is intractable)

**Models for Subsampling Experiments:**

Case 1: Fixed Treatment Effect: where **(1)** **(2)** **(3)** **(4)** **(5)** set

**(6)** **(7)** **(8)** **(9)** **(10)** **(11)** The multiple responses on the same EU are correlated:

* Goal: Test for differences in treatments- where  and

Case 2: Random Treatment Effect: where **(1)** **(2)** **(3)** **(4)** **(5)** ) **(6)** **(7)** **(8)** **(9)** **(10)**

**(11)** )

* Goal: Obtain estimates of the variance components and evaluate the relative sizes of

Sum of Squares for Both Fixed and Random Effects Models: = # treatments, = # EU's assigned to treatment i,

# of subsamples taken on the jth EU assigned to TRT i,

* Sum of Squares Subsample within Rep: Sub(Rep,TRT) : Pooled Variation w/in Subsamples for same EU – Variation w/in EU or Due to Measurement Error

*S*

* Sum of Squares Rep w/in Treatment: R(T) – Pooled Variation Between Replications within Same Treatment

*S*

* Sum of Squares Treatment: Variation Between the t TRTs:

Expected Mean Squares: To determine the proper test statistic it is necessary to derive the Expected Mean Squares:

* where:
* *C1*

Case1: Equal # of subsample: When (equal number of subsamples for all reps,

; F has a **Central**

Under

Case2: Unequal # of Subsamples: when are not all equal, then doesn't have an F distribution. An approximate test statistic can be constructed:

(Satterthwiate Approximation)

Estimation of in Fixed Effects Model: EU's observed on treatment i, subsamples on the jth EU receiving treatment i, The LSE of is given as follows:

* This results from the fact that the subsamples are not valid replications of the TRT means. In fact, the AOV F-test for testing differences in the treatment means is equivalent to running an AOV on the means over the subsample means:
* . When

Finding Power From Table Given r,t,: (1) (2) Use the set of graphs with and and

Determine Minimum Number of Reps, r, to Achieve a Specified Power for Fixed number of Treatments, t: We want to determine the minimum number of replications, r, such that a level test of will have power of at least to detect that ; using tables

Determine Minimum Number of TRTs, t, to Achieve Specified Power for Fixed Number of Reps, r: Same Calculations, just look at different graphs for values of v1.

**Handout # 7: CRD with Factorial Treatment Structure**: (i.e. We now have multiple Factors)

CRD with 2 Factors: that is, we have independent data from t=ab populations having normal distributions with equal variances **(3)** is the mean response of the treatment consisting of the ith level of Factor 1 and the jth level of Factor 2. Thus, we have a\*b populations, each having a normal distribution with equal variances but possibly different means

* Effects Model: with constraints
* **Simple Effects** of a factor are contrasts between the levels of one Factor at **FIXED** levels of the second Factor. Fix F1 at ith level, construct contrasts across levels of F2
* **Main Effects** of the factors are contrasts between levels of one factor **averaged over the levels of the second Factor**.
  + For Example: Suppose F1 has 2 levels and F2 has 3 levels: Main Effects F1: ; Main Effects of F2: (+2 more contrasts)
* **Interaction Effects** between two factors measure differences between the Simple Effects of one Factor at different levels of the second Factor. For Example, suppose F1 has 2 levels and F2 has 3 levels. Compare the two levels of F1 at the three different levels of F2 (or vice versa):
* **Definition of Interaction:** Two Factors F1 and F2, INTERACT if the differences in the average response at all pairs of levels F1 are NOT the same at all levels of F2. There is an interaction between F1 and F2 if

LS Estimation of Model parameters: **(1)** **(2)**  **(3)**  **(4)**

Development of Test Statistics: In experiments involving a factorial TRT structure, there are several types of hypotheses that the researchers want to test

* **Overall TRT Differences:** **(a)** Cell Means Model: **(b)** Effects Model:
* **Interactions btwn Factors:** **(a)** Cell Means Model:

**(b)** Effects Model:

* **Main Effect of Factor F1:** (\*only if interaction isn't significant) **(a)** Cell Means Model:

**(b)** Effects Model: (it’s the same basic thing for Main effect of F2 but with

Table

Description automatically generatedCase 1: Equal Number of Replications: When the experiment has an equal number of replications per treatment , the test statistics are obtained by a decomposition of into its several components: Then is decomposed into

* The types of inferences that are conducted post AOV depend on the conclusions of the test for an interact.
* Case1: If the interaction is not significant, then we can look at the main effects (i.e. we can make inferences

about the differences in the mean TRT responses across the levels of F1 averaged over the levels of F2)

* Case2: If the interaction is significant, the inferences concerning the differences in the mean TRT responses

for F1 must be conducted separately for each level of F2. Because when there is an interaction, the diffs

in the TRT mean responses across the levels of F1 may differ depending on the level of F2

Two Factor Experiment with Unequal Number of Reps Per Treatment: Let be the number of EU's assigned

to treatment (i,j) and suppose the are not all equal. Then the estimation of TRT means and tests of hypos

becomes more complicated.

* Standard Errors of LSE Estimates: ;
* Confidence Intervals: for the above LSE means a 100(1-α)% CI is given by

CRD w/ an axbxc factorial Treatment Structure and R reps: Suppose we have t TRTs which are constructed by combining the levels of three factors. t=a\*b\*c. To each of the t treatments r EU's are randomly assigned, yielding n=r\*a\*b\*c observations. Further suppose that the levels of the three factors are FIXED.

* Diagram

  Description automatically generated3-Way Interaction: the

3- way interaction is evaluating whether the 2-way interactions between F1 and F2 are consistent

across all levels of F3

* Three Sets of 2-Way interactions: F1\*F2-
* Three Sets of Main Effects: F1-
* The Effects Model:
* Constraints: Highest subscripted value is set equal to 0

Table

Description automatically generated

**Sample Size Calculations:** or if we assume no interaction effects

**Selected Problems and Answers:**

Midterm 2 Practice Exam:

**(1)** *Based on the two plots and assuming the standard errors for estimating the treatment means is very small, there appears to be a three-way interaction. The Water by Day interaction for Chamber C1 has a different pattern than the Water by Day interaction for Chamber C2***.**

**(1)** In a study to determine if there is a difference in the mean strength of cotton fibers produced by the thousands of fiber manufacturers located in North America, the **researcher randomly selects 10 manufacturers** to be included in the study. From each manufacturer, 20 samples of cotton fiber are randomly selected from their warehouse and a tensile strength measurement is determined for each sample**. The researcher used Tukey’s HSD procedure** to determine which pairs of manufacturers had significantly different mean responses. What is a major criticism of the researcher’s methodology?

**C** - *Tukey’s HSD procedure should only be used when the treatment levels have fixed effects. (The factor Manufacturer has Random levels and hence HSD would not be appropriate)*

**(3)** After conducting a CRD with the factors A and B having 3 fixed levels each, the researcher conducts a residual analysis and determines there are a large number of outliers. Because a transformation of the data often leads to a situation where the conclusions are hard to interpret, the research would like you to suggest an alternative approach. Which one of the following methods would be a valid method to determine the existence of an interaction between the factors A and B?

**B** - *Apply the multiple comparison procedure associated with the Kruskal-Wallis rank based procedure to the pairs of levels of factor A at each level of factor B. (Compare the levels of factor A separately for each level of Factor B using the ranks, AOV only robust to moderate deviations from normality)*

**(5)** An entomologist designs an experiment to evaluate the effectiveness of five Dose levels of a pestivide to control fire ants. She **randomly** selects 100 1-acre plots of land and randomly assigns 20 plots to each dose level. Next, she randomly selects 15 fire ant hills in each plot and records the weight, W, of the fire ants killed after two weeks of treatment. The scientist runs the following code in SAS to analyze her data:

**PROC GLM;** CLASS DOSE PLOT; MODEL W=DOSE PLOT(DOSE); RANDOM PLOT(DOSE)/TEST; LSMEANS DOSE/PDIFF ADJUST = TUKEY

She then uses the output from LSMEANS to group the five Doses according to the mean weight of fire ants killed. The conclusions reached using the SAS output will be incorrect b/c: **C.** **the calculation of ̂SE (ˆμi) is incorrect, SAS only considers and not in the calculation. (PROC GLM ignores random effects, use Proc mixed)**

**(6)** A three factor completely crossed experiment is run with Factor A-fixed levels, Factor B-fixed levels, Factor C-fixed levels. The following effects were significant: interaction between factor A and factor B, and the interaction between factor A and factor C. The following effects were not significant: main effect of A, main effect of B, main effect of C and the 3-way interaction between A, B, and C.

**B** - *comparison of the differences in the levels of Factor B can be conducted using Tukey’s HSD on the means for the levels of Factor B computed separately at each level of factors A and C.- (Factor B interacts with Factor A and the 3- way interaction is not significant. But whether or not Factor B interacts with Factor C is not given. Therefore, to be conservative, assume that the B by C interaction is significant and select Answer B or select answer E implying that not enough information is given.)*

**(7)** A Completely Randomized Design with t = 4 treatments, r = 5 reps/treatment, and m = 7 subsamples/rep was run. The researcher wanted to perform a multiple comparison procedure to pairwise group the 4 treatment means with an overall error rate of αF = .05. The form of HSD for this type of experiment is given by:

**B**: HSD = where ; With subsamples, the correct error term is with each of the treatment means averaged over (5)(7)=35 data values

**Assignment05:**

**(4)** What are the necessary conditions on the distributions of the data in a CRD in order for the Kruskal-Wallis statistic to be valid in testing for treatment differences? How are these conditions different from the conditions required for applying the test statistic? *The t treatment population distributions have same location-scale family with equal scale but different location parameters. Data iid within treatments and independent between treatments. Conditions are identical with standard AOV model except non-Normal distributions are allowed.*

**Assignment07:**

**(3a)** In an experiment having the levels of factor F1-qualitative and the levels of factor F2-quantitative, the F1 ∗F2 interaction was found to be significant. The experimenter wants to compare the mean responses across the levels of factor F1, averaged over the levels of factor F2.

**R-***Comparison of marginal means is not appropriate.*

**(3b)** In an experiment having factors F1 - quantitative and F2 - qualitative, the F1 ∗ F2 interaction was found to be significant. The experimenter wants to determine if there is an increasing relationship in the mean responses across the levels of factor F1 at each level of factor F2

**B**- *Trend analysis in the levels of F1 separately at each level of F2*

**(4a)** F1 ∗ F2 ∗ F3 was not significant and F1 ∗ F3, F2 ∗ F3, F2, F3 were significant, but F1 ∗ F2 was not significant. The experimenter wants to compare the levels of factor F1, a quantitative factor.

***C or K*** *: Because F1 ∗ F3 was significant but F1 ∗ F2 and F1 ∗ F2 ∗ F3 were not significant, we should examine trends in levels of F1 separately over the levels of F3, but averaged over the levels of F2. That is, trends in separately for each value of k = 1, . . . , c or you could do a Tukey comparison of the means of F1 separately over the levels of F3, but averaged over the levels of F2.*

**(4b)** F1 ∗ F2 ∗ F3 was significant, F1 ∗ F3, F1 ∗ F2 were not significant, but F2 ∗ F3, F1, F2,and F3 were significant. (levels of F2, F3 are qualitative, levels of F1 is quantitative). The experimenter wants to evaluate trends in the levels of factor F1.

**D-***Trend analysis in the levels of F1 at each combination of (F2, F3)*

**(4c)** F1 ∗ F2 ∗ F3 was not significant and F1 ∗ F3, F2 ∗ F3, F1 ∗ F2 were significant, but F1, F2, F3 were not significant. The experimenter wants to compare the levels of factor F1 to a control treatment.

**G&H-***Dunnett’s comparison technique applied to the levels of factor F1 separately at each level of F2 & Dunnett’s comparison technique applied to the levels of factor F1 separately at each level of F3*

**(4D)**- F1 ∗ F2 ∗ F3, F1 ∗ F2, F1 ∗ F3, and F2 ∗ F3 were significant, but F1, F2, F3 were not significant. The levels of F2, F3 are qualitative, levels of F1 are quantitative. The experimenter wants to determine the level of factor F1 which yields the smallest treatment mean.

**P-***Hsu’s comparison technique applied toto the levels of factor F1 separately at each combination of (F2, F3)*

(5d)- In a CRD with a qualitative factor A at 2 levels and a quantitative factor B at 3 levels, the researcher wants to know if the quadratic trend in the mean responses across the levels of factor B differ for the two levels of factor (A). Which of the following contrasts would address this question:

B-