Chapter 4

Analysis of Variance

(Balanced Case)

Review: Categorical Data

In Chapter 3:

- Combinations of categorical variables defined cells
- Had counts for each cell
- Looked at association between categorical variables

Review: Two-Sample T-Test

- Classification variable with two possible values
- A population measure of interest
- Observed values from each population
- Assumed underlying normality (in each population)

Limitations of T-Test

The two-sample t-test limitations:

- Single classification variable with two possible values
- Assumption of normality
- Use rank-based methods when not normal
- Case of more than one categorical variable?
- Case of more than two categorical values?

ANOVA Model

- Continuous response
- Assumed iid normal errors
- One or more nominal categorical predictor variables
- No continuous predictor variables

Some Definitions

- **One-way analysis of variance**: ANOVA based on a single categorical predictor variable
- **Two-way analysis of variance**: based on 2 independent categorical predictor variables
- **N-way analysis of variance**: based on *n* independent categorical variables

More Definitions…

- **Main effect**: effect of single categorical predictor
- **Interaction**: the combined effect of combination of categorical predictors
- **First-order interaction**: an interaction between two categorical predictors
- **N-th order interaction**: interaction of a categorical predictor with *n* other categorical predictors
- **Balanced data:** data with an equal number of observations in each cell

proc tabulate

- Can be used to get cross-classified measures
- Can use it to tabulate basic stats by cell
- Some similar structure to **proc freq**

proc anova

- Procedure for analysis of variance for balanced data
- Also fine for unbalanced one-way ANOVAs
- Some additional experimental designs are also fine
- Use **proc glm** for more general cases (next chapter…)

model Statement

- Specifies response and effects
- **model *response = terms***
- Main effect: ***variable***
- Interaction: use ***** between variables (e.g. ***A*B***)
- Bar (|) syntax: include interaction and all lower interactions and main effects (e.g. ***A|B|C***)

Example: Hypertension Data

- Response: blood pressure (**bp**)
- **drug**: **X**, **Y** or **Z**
- **biofeed**: **P**resence or **A**bsence
- **Diet**: **Y**es or **N**o

Tabulation of **bp** by **drug**

- Obtain descriptive stats like mean, standard deviation, and count
- Indications of differences of means
- Indications of assumed common variance
- Identify balanced or unbalanced data

Cross-classified Tabulations

- Can do the same for combinations of **drug**, **biofeed**, and **diet**

Example: One-way ANOVA

- Perform analysis of variance of **bp** as a function of **drug**
- ANOVA table interpretation
- R-Square value for predictive power of the model
- Significance of **drug** as a predictive variable
- Consistency with tabulation?
- Conclusion about impact of **drug** on blood pressure?

Exercise: Two-way ANOVA

- Fit ANOVA model of **bp** with main effects **drug** and **diet**
- Interpret significance of model, terms, etc. like in the one-way case
- Draw conclusions about impact of **drug** and **diet** on blood pressure

Exercise: One-way Using **cell**

- Analysis using **cell** as the only predictor
- What does this tell us about the individual impacts of **drug**, **diet**, and **biofeed**back on blood pressure?
- Conclusions based on this analysis?

Exercise: Three-way ANOVA

- ANOVA model with **drug**, **biofeed**, and **diet** main effects (no interactions)
- Interpret results

Exercise: Add Interactions

- Add all interactions to the previous model
- Compare to the one-way ANOVA based on **cell**
- Compare to the three-way main effects ANOVA
- Insignificant interactions at .05 level?
- Include an **ods output** statement to save the means of the three-way interaction term to a variable **outmeans**

Testing Equal Variance in One-way

- Assume common error variance in model
- In one-way case can test this with Levene's test
- Related to F test we used for t-tests
- Use **hovtest** option to **means** statement
- Could use Welch adjustment (with **welch** option) if equal variance is not reasonable

Multiple Comparisons

- Compare two group means like before
- That gives t tests and t confidence intervals
- Making many comparisons at once
- Need to account for increased probability of making wrong decision
- Want probability of making some mistake to be small
- Use **means** statement in SAS (main effects only)

Bonferroni’s Method

- Dividing significance level by number of comparisons being made
- Very conservative comparison
- Results in an overall error rate less than or equal to significance level

Tukey’s Method

- Assumes all pairwise comparisons to be made
- Most appropriate if comparing all differences of means

Scheffe’s Method

- Considers all possible contrasts

Other Methods

- Other methods built around other assumptions exist
- See help for **The ANOVA Procedure> MEANS Statement** for a few

Example: One-way example

- Test for equal variance in drug effect model
- Interpret significance of differences of mean for the drug groups

Example: Choose Best Model

- Back to three-way model with interactions…
- Based on this model what terms should we keep?

Exercise: Means Analysis

For the model just chosen:

- Perform pairwise means analysis on main effects
- Which test should we use?
- What do we conclude about significantly different groups?

The Log Example from the Text

- Transformation to hopefully reduce higher order interaction effect
- Has benefits and drawbacks…