Final Exam Part 1: Jade O’Connor

1. Pet Food Company: One-way ANOVA

Introduction: For this report, I am interested in investigating whether a pet food should expand their cat food product line based on data collected from the existing chicken and shrimp food and newly developed salmon and beef food. For this data, ten cats were randomly assigned to a flavor of cat food and were each given 3 ounces of their assigned food. The amount of food consumed in a 10 minute interval (beginning when filled food dish was presented) was measured and data is listed below:

**Type** **Quantity**

Shrimp 2.26, 2.69, 2.25, 2.45, 2.34, 2.37, 2.22, 2.56, 2.36, 2.59

Chicken 2.29, 2.23, 2.41, 2.68, 2.25, 2.17, 2.37, 2.26, 2.45, 2.57

Salmon 1.79, 2.33, 1.96, 2.05, 2.26, 2.24, 1.96, 1.58, 2.18, 1.93

Beef 2.09, 1.87, 1.67, 1.64, 2.16, 1.75, 1.38, 1.92, 1.32, 1.94

This is a good sampling technique because random assignment protects against bias of the cats preferences of food and 10 cats for each flavor allows for equal comparison among the flavors of cat food.

Analysis: I will run a one-way ANOVA because we have one independent variable and are comparing it’s multiple levels (one factor with more than two levels/groups).

Our hypotheses are as follows:

μ₁: Mean ounces of Shrimp flavor eaten in 10 min

μ₂: Mean ounces of Chicken flavor eaten in 10 min

μ₃: Mean ounces of Salmon flavor eaten in 10 min

μ₄: Mean ounces of Beef Flavor eaten in 10 min

H₀: μ₁=μ₂=μ₃=μ₄

Hₐ: At least one mean is different

\*Not finished\*

1. Chi square Test of Independence

Introduction: For this report, I am investigating the type of entree ordered and the type of dessert ordered by patrons during the Friday to Sunday weekend time period at a continental-style restaurant. We are investigating a relationship between entree and dessert ordered at this particular restaurant on the weekend time period. Data was collected from 607 customers on the type of entree ordered and type of dessert ordered. Data is given below:

 **Dessert**

**Meal** Cake Fruit Ice Cream None Total

 Beef 51 11 13 124 199

 Fish 33 11 12 140 196

 Pasta 12 10 10 42 74

 Poultry 22 10 8 98 138

Total 118 42 43 404 607

 I will assume this is a good sampling technique, but I was not given detailed information regarding the sampling technique. From the information given, the restaurant owner sampled all customers on one given weekend and recorded their meal/entree choice and dessert choice. This sampling method is good, because customers are sampled on all days of the weekend and the large sample size reduces variability in the data. A better sampling method may have been to sample multiple weekends but it would’ve been more time consuming. Conditions are met with one SRS, each individual categorized according to both variables, and all expected counts are at least 5 (smallest is 5.12>5).

 Expected Counts:

 **Meal Dessert**

 Cake Fruit Ice Cream None

 Beef 38.68 13.77 14.1 132.45

 Fish 38.1 13.56 13.88 130.45

 Pasta 14.39 5.12 5.24 49.25

 Poultry 26.83 9.55 9.78 91.85



Above is a bar graph of the customer dessert choices including the meals they ordered located within the bar. The graph shows that most customers did not order dessert, but if they did it was mainly cake. Fruit and ice cream appear to have equally low amounts of orders.

Analysis: I will run a chi-square test of independence because we are working with nominal/categorical data with two categorical variables (meal and dessert).

H₀: there is no relationship between meal order and dessert order

Hₐ: there is a relationship between meal order and dessert order



The results of our Chi-Square Test of independence shows p<.05 which means we reject the null hypothesis and have sufficient evidence to conclude that there is a relationship between meal order and dessert order, $X^{2}(9, N=607)$= 19.28, p<.05.

Conclusion: The Chi-Square test of independence shows a relationship between type of dessert and type of entree ordered in this particular restaurant over the Friday to Sunday weekend time period.

1. Two-way ANOVA 2x3

Introduction: I am investigating the best method of applying the reflective stripe that is used to guide Automated Guided Vehicles (AGVs, or robotic vehicles) along their path. I am comparing the application methods of paint and coated adhesive tape which will be compared to the three different floor types, linoleum, Concrete A and Concrete B (slicker). The strip application is tested on each floor type’s “test track” and replicated 5 times randomly. Time (in minutes) is recorded for the time the AGV travels before deviating from its path. Data in minutes is shown below:



This is a good sampling technique because the times were replicated to produce less bias in our dataset as well as increasing the data sample and reducing variability. Our data is from independent SRS’s. Our data does not show the same standard deviations (.311/.11>2) but we will use the sample because our standard deviation calculation is still less than 3.

The boxplots are shown below:



The boxplots show an outlier for floor type A. It also appears that floor type B has a larger mean time than both Linoleum and concrete A. The paint application appears to have larger variability than the adhesive application.

Analysis: I will run a two-way ANOVA to compare our two qualitative variables on the time before the AGV deviates from its path (in minutes).



Our two-way ANOVA shows a significant main effect for floor type (p<.001) and a significant interaction of floor type and application type (p<.01). Bonferroni Post hoc test was run for Floor Type because it has more than 2 levels and it has a significant main effect (p= 4.544e-11)



The bonferroni post hoc test shows that there was a statistically significant difference in time between Concrete A and Concrete B and Linoleum and Concrete B. Both comparisons show Concrete B as having a significantly higher time (in minutes) than both other floor types.

Conclusion: After running the two-way ANOVA to determine the best method of applying the reflective strip and the best floor type for the AGVs. Our data analysis showed a main effect of floor type (p<.001) which led to our post hoc test that determined Concrete B had the highest time (in minutes) until the AGVs derailed from their path and that Concrete A had the lowest time (in minutes) until the AGVs derailed. Our analysis also showed an interaction between floor type and application type, F=8.019, p<.05.This means that one variable differs among levels of the other variable. Application type did not have a significant main effect (p>.05) which means there was no statistically significant difference between the two application types.

Looking at this analysis, I would recommend using Paint for a slightly faster time (in minutes) of AGVs before derailing, but the difference between paint and adhesive was not large enough to be a significant finding. I also recommend not using Concrete B, because it had the highest time (in minutes) which was a significant finding.

1. Multiple regression

Introduction: For this report, I am investigating the value of office buildings for a commercial developer and will build a model to predict the value of office buildings given multiple variables, floor space (square feet), number of offices, number of entrances, and age of the building (months). Data was collected on 30 buildings in the area. I will also use the model to predict the value of an office building with 2500 square feet, 3 entrances, 2 offices, and 25 months old. Data is below:





Plot of the data shows no linear connections of the predictor variables.



Shows no collinearity.



Y= 17.734(X1) + 5706.29(X2) + 32642.98(X3) - 449.16(X4) + 133934.51

P<.001 indicates at least one predictor variable is not 0. We see that “Entrances” is not significant so we can remove this from our model.



Y= 17.831(X1) + 32586.363(X2) - 407.577(X3) + 147580.387

 Floor Offices Age Intercept

To interpret the values in my model: For Floor space, as the value of the office building increases by 1, the floor space increases by 17.831 square feet. For offices, as the value of the building increases by 1, the number of offices increases by 32,586.363. For age, as the value increases by 1, age decreases by 407.577 months.

An office space with 2500 square feet, 3 entrances, 2 offices, and 25 months old:

Y= 17.831(2500) + 32586.363(2) - 407.577(25) + 147580.387

Y= 44577.5 + 65172.726 - 10189.425 + 147580.387

Y= 247,141.188

The value of an office building with 2500 square feet, 3 entrances, 2 offices, and 25 months old according to my model is $247,141.188. This model is useful and the variables included show to be statistically significant predictors of value except entrance which was not included in my final model.