Statistics and Experimental Design

Time Limit: 3 hours

Good Mythical Morning!! The final exam is worth 100 points. Be sure to report the results of all statistical tests in the format outlined in class. Ambiguous marks will be considered incorrect. Let’s talk about that.

Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1) I want to determine whether homesickness during the first year of college is related to the distance between one’s college campus and one’s home. What type of study – observational or experimental – do you think would be best suited to answering my question? What type of analysis would you conduct? Would it be better to collect data from UMass or Amherst College? Why? (4 pts)

* Observational study because you can’t randomly assign kids to colleges at different distances from their homes.
* Regression or correlation
* Amherst College would be better because the variability in distance from home is going to be greater.

2) When you get right down to it, every episode of Good Mythical Morning is good, but I have a kind of a preference for the food mash-up episodes. My three favorites are: ‘Will it taco?’, ‘Will it ice cream sandwich?’, and ‘Will it popsicle?’. I decided to see whether I could crown a grand champion episode by asking all of the members of my family to watch and rate each episode on a 10-point scale (higher ratings = higher preference). The data appear below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Taco | Ice Cream Sandwich | Popsicle |  |
| Matt | 9 | 10 | 4 | 23 |
| Tammy | 5 | 10 | 2 | 17 |
| Jake | 6 | 10 | 4 | 20 |
| Abby | 8 | 8 | 8 | 24 |
| Zoe (dog) | 7 | 7 | 7 | 21 |
| Aragorn (gecko) | 7 | 9 | 8 | 24 |
|  |  |  |  |  |
| **Σ(X)** | 42 | 54 | 33 |  |
| M | 7 | 9 | 6.5 |  |
| **Σ(X2)** | 304 | 494 | 213 |  |

Conduct an ANOVA to determine whether these data indicate there is any difference in preference for the three episodes amongst my family. Your answer should include a statement of the null and alternative hypotheses, all relevant SS, the observed value of your test statistic, and a decision regarding the null. Fcrit = 4.10; Tukey’s Honestly Significant Difference = 2.5. Be sure to INTERPRET the results of all statistical tests. (20 pts)



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | SS | df | MS | F |
| Model | 37.00 |  2 | 18.5 | 5.00 |
| Within | 49.50 | 15 |  |  |
| Bt SS | 12.50 |  5 |  |  |
| Error | 37.00 | 10 | 3.70 |  |
| Total | 86.50 | 17 |  |  |

The ANOVA indicates there is a significant difference between episodes: F (2, 10) = 5.00. MSE = 3.7. The Tukey test indicates that people rated ICS higher than Popsicle, but that there was no significant difference between Taco and either of the other episodes.

3) According to the ‘Swedish Stereotypes’ episode of GMM, Swedes celebrate Vaffeldagen, which translates roughly into Waffle Day. My family celebrates Vaffeldagen by having waffles for breakfast. Mmm-mmm. What a holiday!!!!! We’re trying to figure out if there are cross-cultural differences in what constitutes the best topping for Vaffel (waffles). So, we ran an experiment. We gathered together 12 Swedes and 12 Americans. Half of the people in each group consumed a waffle topped with Vermont maple syrup. The other half of the people in each group consumed a waffle topped with Swedish light syrup, which (according to the Google) is brewed from beets. Each person rated their waffle/topping combination on a 10-point scale (higher ratings = greater preference). The relevant data appear below; to repeat, there are 6 observations per treatment. Use the data in the table to conduct a 2-way ANOVA to determine the effects of the two independent variables (nationality; topping) and the interaction on Vaffel (waffle) preference. Be sure to report AND interpret the results of all appropriate tests. Fcrit for the omnibus test = 3.10; Fcrit for the remaining tests = 4.35. (20 pts)

|  |  |  |
| --- | --- | --- |
|  | American | Swedish |
| Vermont Maple Syrup | Σ(x) = 48Σ(x2) = 386 | Σ(x) = 36Σ(x2) = 228 |
| Swedish Light Syrup | Σ(x) = 18Σ(x2) = 64 | Σ(x) = 54Σ(x2) = 492 |





|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source | SS | df | MS | F |
| Model | 126 | 3 | 42 | 28 |
| Error | 30 | 20 | 1.5 |  |
| Total | 156 | 23 |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nationality | 24 | 1 | 24 | 16 |
| Syrup | 6 | 1 | 6 | 4 |
| N x S | 96 | 1 | 96 | 64 |

The ANOVA revealed a significant omnibus effect: F (3, 20) = 28, p <.05. This indicates that at least one of our treatment means differed from the others. The main effect of nationality was significant: F (1, 20) = 16, p <.05. Swedish people liked waffles more than Americans. The main effect of Syrup was not significant: F (1, 20) = 4, p <.05. There was not enough evidence to conclude that one type of syrup was preferred over another. The interaction between nationality and syrup was significant: F (1, 20) = 64, p <.05. Whereas American preferred waffles with Vermont Maple Syrup, Swedes preferred waffles with their type of syrup. Kind of makes sense.4) Fill in the blanks in the ANOVA table below (1/2 pt each), and answer the questions that follow. Note: 40 subjects participated in each treatment.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | F | p-value |
| ModelErrorTotal | 252.002808.003060.00 | 5234239 | 50.4012.00 | 4.20 | .001 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | SS | df | MS | F | p-value |
| ABAxB | 90.0084.0078.00 | 212 | 45.0084.0039.00 | 3.757.003.25 | .025.009.041 |

a) How many observations were in the experiment? (2 pts)

240.

b) Was the interaction between the two factors significant? Explain. (1 pts)

Yes. P-value was less than .05.

BONUS I: What proportion of the variance in the dependent measure can the three independent factors (A, B and the interaction) explain? (2 pts)

Proportion of Variance = SSModel / SSTotal = 252 / 3060 = .082