SOC 282 Take-Home FINAL EXAM

**For the exam, you will be asked to use SPSS to run and interpret ANOVA and linear regressions using data** **from a 2016 study conducted on a Cosmetic Brand’s Facebook page. The relevant abstract from the article that was published about the study can be found below:**

Moro, S., Rita, P., & Vala, B. (2016). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. Journal of Business Research, 69(9), 3341-3351.

Abstract

This study presents a research approach using data mining for predicting the performance metrics of posts published in brands' Facebook pages. Twelve posts' performance metrics extracted from a cosmetic company's page including 790 publications were modeled, with the two best results achieving a mean absolute percentage error of around 27%. One of them, the “Lifetime Post Consumers” model, was assessed using sensitivity analysis to understand how each of the seven input features influenced it (category, page total likes, type, month, hour, weekday, paid). The type of content was considered the most relevant feature for the model, with a relevance of 36%. A status post captures around twice the attention of the remaining three types (link, photo, video)….

**Directions for preparing the dataset:**

1. *Click “Content” on the D2L course site.*
2. *Go to “Table of Contents”*
3. *Click on the “Final Exam Materials” folder/module.*
4. *Download the dataset “Facebook Metrics”*
5. *Open SPSS*
6. *Click “File”*
7. *Hover over “open”*
8. *Click “Data”*
9. *Choose your download folder*
10. *Open the dataset you just downloaded*
11. *Be certain to label the* ***variable values*** *for “****category****” before you run your analysis. The categories are described as: action (special offers and contests), product (direct advertisement, explicit brand content), and inspiration (non-explicit brand related content):*

*1=action*

*2=product*

*3=inspiration*

***\*\*\*\*\*\*\*There are multiple extra credit problems on this exam. You may choose to do only select problems or do all of them, but the extra credit maxes out at 10 points (i.e., you cannot get more than 10 extra credit points even if you get all the extra credit problems correct).\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\****

***Part A****. Conduct a one-way ANOVA analyzing if* ***category*** *makes a difference in regard to* ***lifetime***

***Post consumptions (variable #12 on the variable view page)****. Please make certain you are conducting your analysis using the correct variables since many of them have similar names! It might be useful to widen the column that gives the variable name on SPSS.*

***Please provide the following: descriptive statistics for your dependent variable (you need to run this separately from your ANOVA), null hypothesis and alternative hypothesis, the f-statistic (including degrees of freedom), and the p-value. All problems worth 1 point unless otherwise noted.***

1. **=**

**=**

**3. :**

**4. :**

**5. =**

**6. =**

**7.**

**8. Explain your results, including which hypothesis is favored (2pts):**

**Extra credit (2 pts): What is the critical value associated with the df? (you can look this up in the table on D2L)**

Critical value ≈

**9. Please attach your output for your descriptive statistics and ANOVA below (3 pts):**

**POST HOC ANALYSIS**

**You hypothesize that the mean for Liftetime Post Consumptions will be higher for inspiration posts compared to product posts. You also hypothesize that mean for Liftetime Post Consumptions will be higher for action posts compared to inspiration posts. Write the corresponding null and alternative hypotheses below:**

**10. :**

**11. :**

**12.** :

**13. :**

**Provide the following information Below:**

**Run a post hoc analysis using the Bonferonni test**

**14. Do “inspiration” posts have a higher mean for Lifetime Post Consumptions** **compared to “product” posts?**

**15. If yes, how much higher? (if not, write N/A)**

**16. Do you reject or fail to reject the null?**

**17. Why?**

**18. Do “action” posts have a higher mean for Lifetime Post Consumptions** **compared to “inspiration” posts?**

**19. If yes, how much higher? (If not, write N/A)?**

**20. Do you reject or fail to reject the null?**

**21. Why?**

**22. Attach your output below (6 pts):**

**Extra Credit (5 pts): create a bar chart with confidence intervals (also called “error bars”) for the means of each category for Lifetime Post Consumptions. Explain in a couple of sentences what information the reader can get just by looking at the bar chart. Attach your output below:**

***Part B****. Conduct a linear regression with* ***Totalinteractions*** *as the dependent variable and* ***comments*** *as the independent variable****.***

***Please provide the following: null hypothesis and alternative hypothesis, the t-statistic and p-value for the independent variable, value. Lastly, write the regression equation with the values from your output and predict Total interactions from the value of comments provided below.***

1. **:**
2. **:**
3. **=**
4. **<**
5. **=**
6. **Regression equation (3 pts):**
7. **Total interactions this model predicts when there are 100 comments (3 pts):**
8. **Attach all pieces of your output (4 pts):**

***Extra Credit (5 pts): Create a scatterplot with the total fit line and attach it below. What information does the scatterplot provide that we might not have known just from running the regression?***

*Conduct a multiple linear regression with* ***Totalinteractions*** *as the dependent variable and* ***comments and shares*** *as the independent variables****. Answer the following questions:***

1. **How much of the variance do comments and shares account for in Total interactions? (2 pts)**
2. **Which variable is a stronger predictor of total interactions? (2 pts)**
3. **How do you know? (2 pts)**
4. **Create a table using the information from the output and paste below (5 pts):**
5. **Attach all pieces of your output (4 pts):**

**Extra Credit (5 pts): What might be a problem with our independent variables (are any rules being violated)? Please attach output to support your answer.**