OPIM 5510 Assignment 3

**Fall 2022**

This is a team-based assignment and has in total 10 points. Include your answer in a **single** Word document (named **assignment3-team-x.docx**) and submit it on HuskyCT before its posted deadline. Members of the same team share the same link to submit assignment on HuskyCT. Each team only needs to submit one solution. HuskyCT allows for three attempts for this assignment before the deadline. Only the **last attempt** will be graded. Failure to submit your assignment by the posted deadline will result in **a score of 0** on this assignment – no exception will be made.

Please properly format your report to make it easy to find the answer to each question. You will be graded based on correctness, clarity, easiness to read, thoughtfulness in crafting the reports (i.e., non-trivial and novel), and the significance of the business insights in the report. Please note that although the assignment will be turned in as a team, each student is required to understand all the materials. This helps you prepare for the exam.

1. ***Multiple Choice Questions (one choice per question)***

Write down your answer for Questions 1 and 2 in the following table:

|  |  |
| --- | --- |
| **Question #** | **Answer** |
| 1 | A |
| 2 | B |
| 3 | A |
| 4 | B |

* 1. Suppose you would like to know whether reducing the number of words in your ad can increase the number of clicks on your ad to help you make decision on whether you should change the number of words in your ad. What method should you use for this analysis?

Causal Analysis

1. Predictive Analysis
   1. Which of the following is not a statistic that can help you assess the goodness of fit of your model?
      1. Adjusted R2

Estimated coefficients

* + 1. Root Mean Square Error
    2. R2
  1. Which of the following is an example of Type I error? = accept Ha but Ho is true
     1. Including more links on the landing page increased the bounce rate, but your model suggests that including more links on the landing page did not have an impact on the bounce rate
     2. Including more links on the landing page increased the bounce rate, and your model suggests that including more links on the landing page had an impact on the bounce rate
     3. Removing coupon box did not increase your sales, but your model suggests that removing coupon box increased your sales.

not increase your sales.

Removing coupon box did not increase your sales, and your model suggests that removing coupon box did

Answer = accept Ha from the model, when Ho is true

Ho = Including more links on the landing page increased the bounce rate

Ha = Including more links on the landing page does not impact the bounce rate

* 1. Suppose you would like to include income groups in your regression. Your data has 3 income groups: 0-10000, 10000-20000, and 20000+. How many income group dummy variables do you need to create and include in your regression?

A. 1

B. 2

C. 3

D. 4

1. ***Short Answer Questions***
2. Suppose you observe in historical data that the firms running Twitter promotions have higher sales than the firms not running Twitter promotions. You run a regression that has “sales” as the dependent variable and “Twitter promotion dummy” as the independent variable. Twitter promotion dummy equals 1 if the firm runs Twitter promotions and 0 otherwise. The estimated coefficient of Twitter promotion dummy is positive, and its p-value is 0.02. Does this mean that for those firms not running Twitter promotions, starting to run Twitter promotions will increase sales? Why?

Answer: Sales = Bo + B1 x Twitter, so Twitter (0) = off = Bo, and Twitter (1) = Bo+B1 = and we are looking for a causal relationship = so we will focus on B1. Since B1 is positive, and the p Value for B1 is less than 0.05, then we can say that a change in B1 would have a positive impact on Sales = Yes

1. Suppose you randomly assign three promotions to 21 users and observe the data in PromotionData.xlsx (which is posted on HuskyCT). Specifically, for each user, you observe how much purchase (in dollars) the user made, the user’s tenure (i.e., how long this user has been your customer), and the promotion that was sent to the user. Based on the data, answer the questions below and use R for data analysis.
   1. Run a simple regression in R: Purchase = 𝛽0 + 𝛽1 × Tenure + 𝜀. Based on the result of this regression, does Tenure have a statistically significant impact on Purchase? Why?

# Question 1:Run a simple regression in R: Purchase = 𝛽0 + 𝛽1 × Tenure + 𝜀. Based on the result of this regression, does

#Tenure have a statistically significant impact on Purchase? Why? = regress Purchase $ on Tenure = Set

a3.hw1<-lm(Purchase ~ Tenure, data=PromotionData)

# use summary to see regression output

summary(a3.hw1)

#Answer = Pvalue of Tenure is below 0.05 = Tenure is sigificant and 𝛽1 is positive +2.25 so for every year of tenure, purchase goes up by 2.25

#lm(formula = Purchase ~ Tenure, data = PromotionData)

#Residuals:

# Min 1Q Median 3Q Max

#-292.71 -87.89 -22.03 137.95 243.85

#Coefficients:

# Estimate Std. Error t value Pr(>|t|)

#(Intercept) 286.6905 196.5800 1.458 0.16107

#Tenure 2.2563 0.5864 3.848 0.00108 \*\*

---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

#Residual standard error: 155.3 on 19 degrees of freedom

#Multiple R-squared: 0.438, Adjusted R-squared: 0.4084

#F-statistic: 14.81 on 1 and 19 DF, p-value: 0.001084

Graphical user interface, text, application, email

Description automatically generated

* 1. Run a simple regression in R that has Purchase as the department variable and Tenure, Promotion dummies, and the interactions between Tenure and Promotion dummies as independent variables. When you create the Promotion dummies, Promotion 1 should be the reference category. Based on the result of this regression and any additional hypothesis tests you’d like to do, answer the following questions:
     1. Does Tenure have a significant impact on Purchase for users receiving Promotion 1? Why?

a3.hw2<-lm(Purchase ~ Tenure + as.factor(Promotion), data=PromotionData)

summary(a3.hw2)

# Purchase = 𝛽0 + 𝛽1 × Tenure + P1 x Promo2 + P2 x Promo3 + C1 x Tenure x Promo2 + C2 x Tenure x Promo3

# purchase = 1) 𝛽0 + 𝛽1 × Tenure = Promo1, 2) 𝛽0 + P1 + (𝛽1 + C1) x Tenure = Promo2, 3) 𝛽0 + P2 + (𝛽1 + C2) x Tenure = Promo3

# Tenure effect = 1) 𝛽1 = Promo1, 2) 𝛽1 + C1 = Promo2, 3) 𝛽1 + C2 = Promo3

# C1 = measures the difference in Tenure effect between Promo 1 and 2

# C2 = measures the difference in Tenure effect between Promo 1 and 3

#Call:

# lm(formula = Purchase ~ Tenure + as.factor(Promotion), data = PromotionData)

#Residuals:

# Min 1Q Median 3Q Max

#-219.76 -39.37 13.08 71.82 203.45

#Coefficients:

# Estimate Std. Error t value Pr(>|t|)

#(Intercept) 182.6239 155.9298 1.171 0.25768

#Tenure 3.0398 0.5028 6.046 1.31e-05 \*\*\*

# as.factor(Promotion)2 -228.8483 69.4973 -3.293 0.00430 \*\*

# as.factor(Promotion)3 -235.1909 69.7604 -3.371 0.00363 \*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

#

#Residual standard error: 121.2 on 17 degrees of freedom

#Multiple R-squared: 0.6936, Adjusted R-squared: 0.6395

#F-statistic: 12.83 on 3 and 17 DF, p-value: 0.0001258

#Question 2.a: Does Tenure have a significant impact on Purchase for users receiving Promotion 1? Why?

#Answer 2.a: Yes, P value for 𝛽1 = 1.31e-05 \*\*\*, with 𝛽1 = 3.0398

#Tenure effect = 1) 𝛽1 = Promo1, 2) 𝛽1 + C1 = Promo2, 3) 𝛽1 + C2 = Promo3

Graphical user interface, text, application, email

Description automatically generated

* + 1. Does Tenure have a significant impact on Purchase for users receiving Promotion 3? Why?

#Question 2.b: Does Tenure have a significant impact on Purchase for users receiving Promotion 3? Why?

#Anser 2.b: Yes = Tenure effect on purchase with Promo3 = 𝛽1 + C2 = as.factor(Promotion)3 = P Value = 0.00363 \*\*

#Tenure effect = 1) 𝛽1 = Promo1, 2) 𝛽1 + C1 = Promo2, 3) 𝛽1 + C2 = Promo3

# lm(formula = Purchase ~ Tenure + as.factor(Promotion), data = PromotionData)

#Residuals:

# Min 1Q Median 3Q Max

#-219.76 -39.37 13.08 71.82 203.45

#Coefficients:

# Estimate Std. Error t value Pr(>|t|)

#(Intercept) 182.6239 155.9298 1.171 0.25768

#Tenure 3.0398 0.5028 6.046 1.31e-05 \*\*\*

# as.factor(Promotion)2 -228.8483 69.4973 -3.293 0.00430 \*\*

# as.factor(Promotion)3 -235.1909 69.7604 -3.371 0.00363 \*\*

Text, application

Description automatically generated

* + 1. Is the effect of Tenure on Purchase the same for users receiving Promotion 2 and users receiving Promotion 3? Why?

#Awnser 2.c: Run a Linear Hypothesis Test = linearHypothesis(a3.hw2,"as.factor(Promotion)2 = as.factor(Promotion)3")

linearHypothesis(a3.hw2,"as.factor(Promotion)2 = as.factor(Promotion)3")

#Linear hypothesis test

#Hypothesis:

# as.factor(Promotion)2 - as.factor(Promotion)3 = 0

#Model 1: restricted model

#Model 2: Purchase ~ Tenure + as.factor(Promotion)

#Res.Df RSS Df Sum of Sq F Pr(>F)

#1 18 249911

#2 17 249771 1 140.78 0.0096 0.9232

#Awnser 2.c.cont: Yes, the effect of Tenure on Purchase with Promo2 = Effect of Tenure of Purchase Promo3 = pvalue of LHT = 0.9232 which is Greater than 0.05, accept the null = they are the same.

Graphical user interface, application, Word

Description automatically generated

* + 1. Is the effect of Promotion 3 on Purchase the same for users with different tenure? Why?

(Hint: in this question, we look at the effect of Promotion 3 on Purchase and wonder if this effect varies for users having different tenure – e.g., is the effect of Promotion 3 on Purchase different for users with a long tenure vs. users with a short tenure?)

#Question 2.d: Is the effect of Promotion 3 on Purchase the same for users with different tenure? Why?

#Answer 2.d: 3) 𝛽0 + P2 + (𝛽1 + C2) x Tenure = Promo3 effect on Purchase with different Tenure =

a3.hw2.d<-lm(Purchase ~ Tenure + as.factor(Promotion) + Tenure:factor(Promotion), data=PromotionData)

summary(a3.hw2.d)

#Call:

# lm(formula = Purchase ~ Tenure + as.factor(Promotion) + Tenure:factor(Promotion),

# data = PromotionData)

#Residuals:

# Min 1Q Median 3Q Max

#-183.76 -65.41 -16.99 78.73 164.83

#Coefficients:

# Estimate Std. Error t value Pr(>|t|)

#(Intercept) 878.9527 425.7077 2.065 0.0567 .

#Tenure 0.6908 1.4290 0.483 0.6358

#as.factor(Promotion)2 -696.4607 522.3047 -1.333 0.2023

#as.factor(Promotion)3 -1186.4231 476.0732 -2.492 0.0249 \*

# Tenure:factor(Promotion)2 1.6888 1.6703 1.011 0.3280

#Tenure:factor(Promotion)3 3.0818 1.5500 1.988 0.0653 .

#---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

#Residual standard error: 112.4 on 15 degrees of freedom

#Multiple R-squared: 0.7676, Adjusted R-squared: 0.6901

#F-statistic: 9.906 on 5 and 15 DF, p-value: 0.0002437

#Answer 2.d cont: test the significance of Tenure effect in Promo3 in a3.hw2 = significance is true

linearHypothesis(a3.hw2.d,"Tenure + Tenure:factor(Promotion)3 = 0") = Reject the Null accept the Alternative.

#Linear hypothesis test

#Hypothesis:

# Tenure + Tenure:factor(Promotion)3 = 0

#Model 1: restricted model

#Model 2: Purchase ~ Tenure + as.factor(Promotion) + Tenure:factor(Promotion)

#Res.Df RSS Df Sum of Sq F Pr(>F)

#1 16 688209

#2 15 189460 1 498749 39.487 1.465e-05 \*\*\*

# ---

# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Graphical user interface, text

Description automatically generated

**Note**: For each question, first provide the answer to the question and then copy the screenshot from R that includes both the R commands and the entire regression output - see a sample screenshot below in Figure 1. Highlight the numbers that you use to answer the question in the screenshot.

***Figure 1. A sample screenshot that includes both the R commands and the entire regression output with the numbers used to answer the questions highlighted.***



* ml1<-lm(sales ~ quality, data=BookSales)
* summary(ml1)

Call:

lm(formula = sales ~ quality, data = BookSales)

Residuals:

Min 1Q Median 3Q Max

-87.098 -27.098 7.136 33.253 72.902

Coefficients:

Estimate Std. Error t value Pr(>|t|) (Intercept) 74.630

Number highlighted

Entire Regression

Output

quality

---

14.694 5.079 6.44e-06 \*\*\*

23.117 4.224 5.47

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 42.8 on 47 degrees of freedom Multiple R-squared: 0.3892, Adjusted R-squared: 0.3762

F-statistic: 29.95 on 1 and 47 DF, p-value: 1.674e-06

R Commands

***By submitting the assignment, you are confirming that this represents your own work, and that no part of the work submitted has been copied from others. Consequences for cheating are very serious. It can lead to not only failing the course but also dismissal from the program.***

3 1.67e-06 \*\*\*

***Full R code:***

***# first use "Import Dataset" to import data from PromotionData.xlsx***

***# the default name of the dataset is Promotion Data***

***# view the imported data***

***View(PromotionData)***

***# Assignment 3 use lm command to run some simple regression analyses = Specifically, for each user, you observe how much purchase (in dollars) the user made, the***

***#user’s tenure (i.e., how long this user has been your customer), and the promotion that was sent to the user. Based on***

***#the data, answer the questions below and use R for data analysis.***

***# Question 1:Run a simple regression in R: Purchase = 𝛽0 + 𝛽1 × Tenure + 𝜀. Based on the result of this regression, does***

***#Tenure have a statistically significant impact on Purchase? Why? = regress Purchase $ on Tenure = Set***

***a3.hw1<-lm(Purchase ~ Tenure, data=PromotionData)***

***# use summary to see regression output***

***summary(a3.hw1)***

***#Answer = Pvalue of Tenure is below 0.05 = Tenure is sigificant and 𝛽1 is positive +2.25 so for every year of tenure, purchase goes up by 2.25***

***#lm(formula = Purchase ~ Tenure, data = PromotionData)***

***#Residuals:***

***# Min 1Q Median 3Q Max***

***#-292.71 -87.89 -22.03 137.95 243.85***

***#Coefficients:***

***# Estimate Std. Error t value Pr(>|t|)***

***#(Intercept) 286.6905 196.5800 1.458 0.16107***

***#Tenure 2.2563 0.5864 3.848 0.00108 \*\****

***---***

***# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1***

***#Residual standard error: 155.3 on 19 degrees of freedom***

***#Multiple R-squared: 0.438, Adjusted R-squared: 0.4084***

***#F-statistic: 14.81 on 1 and 19 DF, p-value: 0.001084***

***# Question 2: Run a simple regression in R that has Purchase as the department variable and Tenure, Promotion dummies, and***

***#the interactions between Tenure and Promotion dummies as independent variables. When you create the***

***#Promotion dummies, Promotion 1 should be the reference category. Based on the result of this regression and any***

***#additional hypothesis tests you’d like to do, answer the following questions:***

***a3.hw2<-lm(Purchase ~ Tenure + as.factor(Promotion), data=PromotionData)***

***summary(a3.hw2)***

***# Purchase = 𝛽0 + 𝛽1 × Tenure + P1 x Promo2 + P2 x Promo3 + C1 x Tenure x Promo2 + C2 x Tenure x Promo3***

***# purchase = 1) 𝛽0 + 𝛽1 × Tenure = Promo1, 2) 𝛽0 + P1 + (𝛽1 + C1) x Tenure = Promo2, 3) 𝛽0 + P2 + (𝛽1 + C2) x Tenure = Promo3***

***# Tenure effect = 1) 𝛽1 = Promo1, 2) 𝛽1 + C1 = Promo2, 3) 𝛽1 + C2 = Promo3***

***# C1 = measures the difference in Tenure effect between Promo 1 and 2***

***# C2 = measures the difference in Tenure effect between Promo 1 and 3***

***#Call:***

***# lm(formula = Purchase ~ Tenure + as.factor(Promotion), data = PromotionData)***

***#Residuals:***

***# Min 1Q Median 3Q Max***

***#-219.76 -39.37 13.08 71.82 203.45***

***#Coefficients:***

***# Estimate Std. Error t value Pr(>|t|)***

***#(Intercept) 182.6239 155.9298 1.171 0.25768***

***#Tenure 3.0398 0.5028 6.046 1.31e-05 \*\*\****

***# as.factor(Promotion)2 -228.8483 69.4973 -3.293 0.00430 \*\****

***# as.factor(Promotion)3 -235.1909 69.7604 -3.371 0.00363 \*\****

***# ---***

***# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1***

***#***

***#Residual standard error: 121.2 on 17 degrees of freedom***

***#Multiple R-squared: 0.6936, Adjusted R-squared: 0.6395***

***#F-statistic: 12.83 on 3 and 17 DF, p-value: 0.0001258***

***#Question 2.a: Does Tenure have a significant impact on Purchase for users receiving Promotion 1? Why?***

***#Answer 2.a: Yes, P value for 𝛽1 = 1.31e-05 \*\*\*, with 𝛽1 = 3.0398***

***#Tenure effect = 1) 𝛽1 = Promo1, 2) 𝛽1 + C1 = Promo2, 3) 𝛽1 + C2 = Promo3***

***#Question 2.b: Does Tenure have a significant impact on Purchase for users receiving Promotion 3? Why?***

***#Anser 2.b: Yes = Tenure effect on purchase with Promo3 = 𝛽1 + C2 = as.factor(Promotion)3 = P Value = 0.00363 \*\****

***#Tenure effect = 1) 𝛽1 = Promo1, 2) 𝛽1 + C1 = Promo2, 3) 𝛽1 + C2 = Promo3***

***#Question 2.c: Is the effect of Tenure on Purchase the same for users receiving Promotion 2 and users receiving***

***#Promotion 3? Why?***

***#Awnser 2.c: Run a Linear Hypothesis Test = linearHypothesis(a3.hw2,"as.factor(Promotion)2 = as.factor(Promotion)3")***

***linearHypothesis(a3.hw2,"as.factor(Promotion)2 = as.factor(Promotion)3")***

***#Awnser 2.c.cont: Yes, the effect of Tenure on Purchase with Promo2 = Effect of Tenure of Purchase Promo3 = pvalue of LHT = 0.9232 which is Greater than 0.05, accept the null = they are the same.***

***#Linear hypothesis test***

***#Hypothesis:***

***# as.factor(Promotion)2 - as.factor(Promotion)3 = 0***

***#Model 1: restricted model***

***#Model 2: Purchase ~ Tenure + as.factor(Promotion)***

***#Res.Df RSS Df Sum of Sq F Pr(>F)***

***#1 18 249911***

***#2 17 249771 1 140.78 0.0096 0.9232***

***#Question 2.d: Is the effect of Promotion 3 on Purchase the same for users with different tenure? Why?***

***#Answer 2.d: 3) 𝛽0 + P2 + (𝛽1 + C2) x Tenure = Promo3 effect on Purchase with different Tenure =***

***a3.hw2.d<-lm(Purchase ~ Tenure + as.factor(Promotion) + Tenure:factor(Promotion), data=PromotionData)***

***summary(a3.hw2.d)***

***#Answer 2.d cont: test the significance of Tenure effect in Promo3 in a3.hw2 = significance is true***

***linearHypothesis(a3.hw2.d,"Tenure + Tenure:factor(Promotion)3 = 0")***

***#Call:***

***# lm(formula = Purchase ~ Tenure + as.factor(Promotion) + Tenure:factor(Promotion),***

***# data = PromotionData)***

***#Residuals:***

***# Min 1Q Median 3Q Max***

***#-183.76 -65.41 -16.99 78.73 164.83***

***#Coefficients:***

***# Estimate Std. Error t value Pr(>|t|)***

***#(Intercept) 878.9527 425.7077 2.065 0.0567 .***

***#Tenure 0.6908 1.4290 0.483 0.6358***

***#as.factor(Promotion)2 -696.4607 522.3047 -1.333 0.2023***

***#as.factor(Promotion)3 -1186.4231 476.0732 -2.492 0.0249 \****

***# Tenure:factor(Promotion)2 1.6888 1.6703 1.011 0.3280***

***#Tenure:factor(Promotion)3 3.0818 1.5500 1.988 0.0653 .***

***#---***

***# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1***

***#Residual standard error: 112.4 on 15 degrees of freedom***

***#Multiple R-squared: 0.7676, Adjusted R-squared: 0.6901***

***#F-statistic: 9.906 on 5 and 15 DF, p-value: 0.0002437***

***#Linear hypothesis test***

***#Hypothesis:***

***# Tenure + Tenure:factor(Promotion)3 = 0***

***#Model 1: restricted model***

***#Model 2: Purchase ~ Tenure + as.factor(Promotion) + Tenure:factor(Promotion)***

***#Res.Df RSS Df Sum of Sq F Pr(>F)***

***#1 16 688209***

***#2 15 189460 1 498749 39.487 1.465e-05 \*\*\****

***# ---***

***# Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1***