**ISOM 670 Prof. Stuk Campus ID Number** \_\_\_\_\_\_\_\_\_2234611\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Individual Regression Assignment, Fall 2019**

HONOR CODE STATEMENT

The Goizueta Business School Honor Code is the standard of professional behavior on this exam. When you have completed your exam, please read the following pledge and add your signature if you have complied with the Honor Code:

*I pledge that I have neither given nor received any unauthorized assistance on this exam, and that any violations of the Honor Code by others that I have observed or otherwise become aware of will be reported by me to the Honor Council.*

**Type Name (electronic signature)** \_\_\_\_\_\_Carl Jia Hua Xi\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

INSTRUCTIONS — READ CAREFULLY

**Exam Mechanics:**

Type your name legibly in the space provided below the Honor Code. This serves as an electronically signed pledge, after you have finished the exam.

Use only your own notes and exam prep materials. Sharing materials with other students during the exam period is not permitted. You may use anything posted in our course conference, whether you downloaded it before the exam or not.

Computers are permitted throughout and are necessary for some parts. You are not required to use computers if there is another way to get to an answer.

We have not provided space in the exam booklet itself for you to show your work. Please adjust the spacing accordingly when you create the printed version that you will be turning in.

**PLEASE Transfer your answers to the front Answer Sheet when indicated.** Failure to do this may cost you some points. Of course, your work pages will contain any long answers & any required explanations that might accompany the short answers. **Please transfer all calculations to the exam document.** They will also show your assumptions and how you got your answers. (Note: we recommend that you support all answers by showing your work, as time permits.)

9

**Post your completed exam to Canvas by 9 am 19 August**

**Suggestions for taking the exam:**

These questions are “fresh baked” for this year’s class, so there is the very real possibility that parts of them are half-baked. Contact the professor (via First Class) if something doesn’t seem right. If we do make changes and/or clarifications, we will post them in our course conference right away. (If there are designated on-campus work rooms, we will come by and put that information on the chalk boards, too.) PLEASE — IT IS YOUR RESPONSIBILITY TO CHECK Canvas REGULARLY!

Your best opportunity for clarification of the questions is during the exam, not afterwards. The exam questions are not intended to be ambiguous. If there are words or phrases that you do not fully understand, please ask us about them; this is not a test about American English vocabulary. Ask any questions you like; we just may not be able to answer some questions that are too close to exam content.

Read each section carefully, and spend some time thinking before you try to answer the questions. The questions range greatly in difficulty (some of the tougher are marked “Challenge”). Read through the entire exam before you start working, to gauge the difficulty of the sections and budget your time.

When making assumptions about the problems, try to use the simplest set of assumptions that is consistent with all the information in the problem. Of course, more elaborate complications arise in real life, but here you’ll benefit from keeping things simple.

Partial credit may be important for some questions, so (as time permits) make sure your work pages clearly show your line of thinking and the specific steps of any analysis you performed. (State your assumptions! Draw your pictures!)

Good Luck.

**Deadline: exams are due to Canvas by 9 am 19 August.**

***ANSWER SHEET***

**PART A** *(30 pts)*1. - 8. *Put answers on the exam page.*

*TOTAL = 30 points*

**Part A (30 points)**

Use the data Seawatch C. Generate a simple regression model using r for the dependent variable GROSS: the dollars collected from a town, and the independent variable is Pop80: population from the 1980 census

Provide a copy of the Regression output (i.e. summary (model)

1. Put Regression output here

> Model1 <- lm(SeaWatch\_C\_data$GROSS~SeaWatch\_C\_data$POP80)

> summary(Model1)

Call:

lm(formula = GROSS ~ POP80, na.rm = TRUE)

Residuals:

Min 1Q Median 3Q Max

-13193.0 -1678.9 -657.2 949.7 29473.1

Coefficients:

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

(Intercept) 1.857e+03 2.438e+02 7.616 2.07e-13 \*\*\*

POP80 8.283e-02 8.208e-03 10.091 < 2e-16 \*\*\*

---

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

Residual standard error: 3645 on 383 degrees of freedom

(11 observations deleted due to missingness)

Multiple R-squared: 0.21, Adjusted R-squared: 0.208

F-statistic: 101.8 on 1 and 383 DF, p-value: < 2.2e-16

1. How much of the variability has been explained?

* The “Multiple R-squared” output from the regression summary tells us the percentage of the variations that our model explains. It does so by statically measuring how close the data points are to the fitted regression line. In our case, 0.21 or 21% of the variability has been explained.

1. Is there a relationship between Gross and Pop80? Support your answer?

* To see if there is a relationship between Gross and Pop80, we can look at the coefficient for SeaWatch\_C\_data$POP80 in the summary table. With an estimate of 8.283e-02 or 0.08283, we know that there is a relationship between the two variables. We know this because an estimate of 0 would indicate a lack of relationship, but our estimate is non-zero. We then look at the standard error, t value and p value. Running qt(0.975, 383) gave us the true “2” of 1.966177, which our t value of 10.091 easily surpassed, indicating that our variable is significant. Our p value of 2.2e-16, being much lower than 0.05 further confirms that POP80 is a statically significant variable for explaining gross.

1. Interpret the constant coefficient? Be specific! ( ie. Explain in words)

* If the constant coefficient refers to the estimated intercept, we can say that the model indicates that towns with a population of 0 will on average produce a gross of 1857. If analysis of the t and p values are also expected for this question, please see the latter half of my answer for question 3.

1. How much will Gross increase for an increase of Pop80 by 1? 1000? ( not for a value of 1 and 1000)

* Looking back at the coefficient, we can say that on average ‘Gross’ will increase by 8.283e-02 or 0.08283 for an increase of Pop80 by 1, or 82.83 for an increase of Pop80 by 1000.

1. State a confidence interval for the first answer in question 5 (i.e. the slope).

* We can calculate the tcritical using qt(0.975, 383), which gives us 1.966177. Multiplying our tcritical with the standard error of the slope will give us the 95% confidence interval for our slope.

> slopeup <- (8.283\*10^-2)+(8.208\*10^-3)\*tcritical

> slopedown <- (8.283\*10^-2)-(8.208\*10^-3)\*tcritical

> slopeup

[1] 0.09896838

> slopedown

[1] 0.06669162

* In summary, the slope on average is 0.08283, and 95% of the time it falls within 0.06669 and 0.09897

1. How much variability will you have if you predict Gross with this model?

The residual standard error gives us the variability, so we should expect about a variability of 3645 if we predict Gross with this model.

1. What would be your best *point* estimate for Gross in a town with a Population of 20,000?

> predict.lm(SeaWatchmodel,newdata=data.frame(POP80=20000))

3513.269

* Running the predict.lm function in R with POP80 set to 20000 gave us a result of 3513.269

1. What would be an *interval* estimate for question 8? (Use 95% limits.)

> predict.lm(SeaWatchmodel,newdata=data.frame(POP80=20000), interval="prediction", level=.95)

|  |  |  |
| --- | --- | --- |
| Fit | Lwr | Upr |
| 3513.269 | -3663.145 | 10689.68 |

* Adding a 95% prediction interval to our predict.lm fuction from question 8 gives us an interval with a lower bound of -3663.145 and an upper bound of 10689.68

## END OF EXAM