ISOM 670 Business Statistics, Individual Regression Problem

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 (ie summary (model)

1.) Put Regression output here

```
> model <- lm(GROSS ~ POP80)
> summary(model)

Call:
lm(formula = GROSS ~ POP80)

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
```

2.) How much of the variability has been explained?

According to the R-squared, 21% of the variability in GROSS has been explained.

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

```
> cor(GROSS,POP80,use="complete.obs")
[1] 0.458274
```

YES. GROSS and POP80 have a moderate positive relationship, with a correlation level of 0.458. Also, from the linear regression model above, POP80's P-value is very small, indicating POP80's effect on GROSS is statistically significant.

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

In this regression model, when the total population from the 1980 census (POP80) is 0, the dollars collected from a town (GROSS) is expected to be about 1857.

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

Holding other variables constant, for each 1 increase in POP80, GROSS is expected to increase by 0.0828. Holding other variables constant, for every 1000 increase in POP80, GROSS is expected to increase by 82.83.

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

```
> qt(0.975,383)
[1] 1.966177
```

With a Degrees of Freedom of 383, the critical value for a two-sided 95% confidence interval is 1.966.
The upper limit is Estimate + 1.966 * Std. Error = 0.0828 + 1.966 * 0.00821 = 0.0989
The lower limit is Estimate – 1.966 * Std. Error = 0.0828 – 1.966 * 0.00821 = 0.0667
We have 95% confidence that the true mean of POP80's coefficient is within [0.0667, 0.0989].

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

According to the R-squared, 21% of the variability in GROSS has been explained in this model. Thus, if I predict GROSS with this model, I will have 1-21% = 79% of the variability.

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

In this model, GROSS = 1857 + 0.0828*POP80 = 1857 + 0.0828*20000 = 3513.

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

Since the 95% confidence interval for POP80's coefficient is [0.0667, 0.0989],
the upper limit is GROSS = 1857 + 0.0989*POP80 = 1857 + 0.0989*20000 = 3835;
the lower limit is GROSS = 1857 + 0.0667*POP80 = 1857 + 0.0667*20000 = 3191.
Thus, when POP80 is 20000, GROSS is estimated to be within [3191, 3835] (95% limits).