Karla Cruz

1/30/21

Week 4 Assignment

1. (**1pt**) For each correlation coefficient below, calculate what proportion of variance is shared by the two correlated variables:

1. *r* = 0.25 *r2* = .0625
2. *r* = 0.33 *r2*= .1089
3. *r* = 0.80 *r2*= .64
4. *r* = 0.14 *r2*= .0196

2. (**1pt**) For each coefficient of determination below, calculate the value of the correlation coefficient:

1. *r2* = 0.54 *r* = 0.734
2. *r2* = 0.13 *r* = 0.360
3. *r2* = 0.39 *r* = 0.624
4. *r2* = 0.07 *r* = 0.265

3. (**1pt**) Suppose a researcher regressed surgical patients’ length of stay (dependent variable) in the hospital on a scale of functional ability measured 24 hours after surgery. Given the following, solve for the value of the intercept constant and write out the full regression equation:

 Mean length of stay = 6.5 days; mean score on scale = 33; slope = -0.10

Simple Linear regression equation: Y= a+bX

6.5=a+ (-.10\*33)

6.5=a-3.3

+3.3 +3.3

**a=9.8**

4. (**1pt**) Using the regression equation calculated in Exercise 3, compute the predicted value of Y (length of hospital stay) for patients with the following functional ability scores:

1. X = 42

Y= a+bX

Y= 9.8+(-.10\*42)

Y=9.8+ (-.42)

**Y=5.6**

1. X = 68

Y=a+bX

Y=9.8+(-.10\*68)

Y=9.8+(-6.8)

**Y= 3**

5. (**4 pts**) Reexamine the Fellows Research Groupdata we used in week 2’s assignment (the excel file is named “**Fellows Mental Health Data**” and is also in this week’s assignment area). Using this data set again, perform a correlation analysis using the General Practitioner Assessment of Cognition score (“**GPCOG**”) as a predictor of the Mental Health Survey Score (“**MentalHlth**”) of all subjects.

1. Plot the data with a regression line and perform a regression with the appropriate statistical test in StatCrunch. Copy and paste your graph and regression output into your Word document for full credit.



**Simple linear regression results:**

Dependent Variable: MentalHlth
Independent Variable: GPCOG
MentalHlth = 43.036035 + 0.29198815 GPCOG
Sample size: 150
R (correlation coefficient) = 0.10496464
R-sq = 0.011017575
Estimate of error standard deviation: 10.879158

**Parameter estimates:**

| **Parameter** | **Estimate** | **Std. Err.** | **Alternative** | **DF** | **T-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 43.036035 | 2.3863377 | ≠ 0 | 148 | 18.034344 | <0.0001 |
| Slope | 0.29198815 | 0.22739748 | ≠ 0 | 148 | 1.284043 | 0.2011 |

**Analysis of variance table for regression model:**

| **Source** | **DF** | **SS** | **MS** | **F-stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| Model | 1 | 195.14152 | 195.14152 | 1.6487665 | 0.2011 |
| Error | 148 | 17516.698 | 118.35607 |  |  |
| Total | 149 | 17711.84 |  |  |  |

1. What is the correlation coefficient *r* and what is the coefficient of determination here? What does the coefficient of determination mean in this case?

Coefficient *r =* 0.10496464

Coefficient of determination=0.011017575

Indicating a positive relationship between GPCOG scores and mental health survey scores.

The coefficient of determination means there is 1% of the variation in GPCOG scores as a predictor of

Mental Health Survey Score.

1. Is there a statistically significant correlation between cognition and mental health in this sample?

The value of .2011 is greater than .05 and the Coefficient r-value0.10496464 is less than the table value of .174 making the correlation not statistically significant.

1. What hypotheses did you test?

H0: There is no relationship between GPCOG scores as a predictor and Mental Health Survey scores.

H1: There is a relationship between GPCOG scores as a predictor and Mental Health Survey scores.

6. (**5 pts**) According to the University of Chicago, as men age, their cholesterol level goes up. A new drug (XAB) is being tested to determine if it can lower cholesterol in aging males and at what dose. The data for the first test subject is below:

 Dose (mg): 2 3 5 6 8 10

Cholesterol level (mg/dL): 310 124 190 110 52 20

1. Plot the data with a regression line and perform a regression with the appropriate statistical test in StatCrunch. Copy and paste your graph and regression output into your Word document for full credit.



**Simple linear regression results:**
Dependent Variable: Cholesterol level (mg/dL):
Independent Variable: Dose (mg):
Cholesterol level (mg/dL): = 303 - 29.764706 Dose (mg):
Sample size: 6
R (correlation coefficient) = -0.85838665
R-sq = 0.73682764
Estimate of error standard deviation: 59.884939

**Parameter estimates:**

| **Parameter** | **Estimate** | **Std. Err.** | **Alternative** | **DF** | **T-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 303 | 56.017231 | ≠ 0 | 4 | 5.40905 | 0.0057 |
| Slope | -29.764706 | 8.8942387 | ≠ 0 | 4 | -3.3465153 | 0.0287 |

**Analysis of variance table for regression model:**

| **Source** | **DF** | **SS** | **MS** | **F-stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| Model | 1 | 40162.51 | 40162.51 | 11.199165 | 0.0287 |
| Error | 4 | 14344.824 | 3586.2059 |  |  |
| Total | 5 | 54507.333 |  |  |  |

**Predicted values:**

| **X value** | **Pred. Y** | **s.e.(Pred. y)** | **95% C.I. for mean** | **95% P.I. for new** |
| --- | --- | --- | --- | --- |
| 7 | 94.647059 | 27.172348 | (19.204528, 170.08959) | (-87.935445, 277.22956) |

1. What is the correlation coefficient *r* and what is the coefficient of determination here? What does the coefficient of determination mean in this case?

Correlation coefficient *r* =-0.85838665

Coefficient of determination =0.73682764 this correlation also indicates a negative association.

The coefficient of determination means 74% of the variation we see in the reduction of cholesterol can be explained by taking the new drug XAB.

1. Is there a statistically significant correlation between the dose and cholesterol in this sample?

The pvalue of .0287 is less than .05, and the Correlation coefficient r-value of -0.85838665 is greater than the table value of .811 making the correlation statistically significant.

1. What hypotheses did you test?

H0: There is no relationship between taking drug XAB and reducing cholesterol levels.

H1: There is a relationship between taking drug XAB and reducing cholesterol levels.

1. Using StatCrunch, what is the predicted cholesterol level for a person taking a dose of 7 mg? Copy and paste your regression output from StatCrunch into your Word document for full credit.

The predicted cholesterol level for a person taking a dose of 7 mg is **94.647059.**

**Predicted values:**

| **X value** | **Pred. Y** | **s.e.(Pred. y)** | **95% C.I. for mean** | **95% P.I. for new** |
| --- | --- | --- | --- | --- |
| 7 | 94.647059 | 27.172348 | (19.204528, 170.08959) | (-87.935445, 277.22956) |

7. (**5 pts**) A study by the University of Montreal examined a new test to deepen the evaluation of mental states through the creation and validation of a practical measure, the Mental States Task (MST). The researchers examined many factors including age and the valid and reliable PANAS test. The data is in an excel file named “**Montreal\_Study\_on\_Mental\_States**” and is also in this week’s assignment area. Using this data set, perform a correlation analysis using the subject’s age (“**AGE**”) as a predictor of their PANAS score (“**PANAS**”) of all the subjects.

1. Plot the data with a regression line and perform a regression with the appropriate statistical test in StatCrunch. Copy and paste your graph and regression output into your Word document for full credit.



**Simple linear regression results:**
Dependent Variable: PANAS
Independent Variable: age
PANAS = 0.93798014 + 0.055259298 age
Sample size: 125
R (correlation coefficient) = 0.21675735
R-sq = 0.046983747
Estimate of error standard deviation: 0.93691191

**Parameter estimates:**

| **Parameter** | **Estimate** | **Std. Err.** | **Alternative** | **DF** | **T-Stat** | **P-value** |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 0.93798014 | 0.52767125 | ≠ 0 | 123 | 1.7775843 | 0.0779 |
| Slope | 0.055259298 | 0.022440325 | ≠ 0 | 123 | 2.4624999 | 0.0152 |

**Analysis of variance table for regression model:**

| **Source** | **DF** | **SS** | **MS** | **F-stat** | **P-value** |
| --- | --- | --- | --- | --- | --- |
| Model | 1 | 5.3229204 | 5.3229204 | 6.0639059 | 0.0152 |
| Error | 123 | 107.96988 | 0.87780393 |  |  |
| Total | 124 | 113.2928 |  |  |  |

**Predicted values:**

| **X value** | **Pred. Y** | **s.e.(Pred. y)** | **95% C.I. for mean** | **95% P.I. for new** |
| --- | --- | --- | --- | --- |
| 23 | 2.208944 | 0.083940014 | (2.0427899, 2.3750981) | (0.34695614, 4.0709318) |

1. What is the correlation coefficient *r* and what is the coefficient of determination here? What does the coefficient of determination mean in this case?

Correlation coefficient r =0.21675735

 Coefficient of determination =0.046983747 this correlation indicates a positive association.

The coefficient of determination means 5% of the variation we see in the PANAS score can be explained by age.

1. Is there a statistically significant correlation between age and the PANAS score in this sample?

The p-value 0.0152 is less than .05 and the Correlation coefficient r value of 0.21675735 is greater than the table value of .195 making this statistically significant.

1. What hypotheses did you test?

H0: There is no relationship between age and PANAS Score.

H1: There is a relationship between age and PANAS Score.

1. Using StatCrunch, what is the predicted PANAS score for a subject that is 23 years old? Copy and paste your regression output from StatCrunch into your Word document for full credit.

The predicted PANAS score for a subject that is 23 years old is **2.208944.**

**Predicted values:**

| **X value** | **Pred. Y** | **s.e.(Pred. y)** | **95% C.I. for mean** | **95% P.I. for new** |
| --- | --- | --- | --- | --- |
| 23 | 2.208944 | 0.083940014 | (2.0427899, 2.3750981) | (0.34695614, 4.0709318) |

8. (**1pt**) Use the multiple regression equation below for predicting graduate GPA for the two subjects below:

 Y′ = -1.636 + 0.793(undergrad GPA) + 0.004(GREverbal) – 0.0009(GREquant) +0.009(Motivation)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Subject** | **undergrad GPA** | **GREverbal** | **GREquant** | **Motivation** |
| 1 | 2.9 | 560 | 540 | 55 |
| 2 | 3.2 | 550 | 590 | 65 |

**Subject 1:**

Y=-1.636+0.793(2.9) + 0.004(560) – 0.0009(540) +0.009(55)

Y= -1.636+2.2997+2.24-.486+.495

**Y=2.9127**

**Subject 2:**

Y=-1.636+0.793(3.2) + 0.004(550) – 0.0009(590) +0.009(65)

Y=-1.636+2.5376+2.2-.531+.585

**Y=3.1556**

9. (**1pt**) Using the following information for *R2*, *k*, and *N*, calculate the value of the *F* statistic for testing the overall multiple regression equation and determine whether *F* is statistically significant at the 0.05 level:

1. *R2* = 0.13, *k* = 5, *N* = 120

F= (.13/5)/((1-.13)/(120-5-1))

F=.026/((.87)/(114))

F=.026/.0076

**F=3.421**

The F value of 3.421 is greater than the table value of 2.45 making the results statistically significant.

1. *R2* = 0.18, *k* = 3, *N* = 64

F= (.18/3)/((1-.18)/(64-3-1))

F=.06/((.82)/(60))

F=.06/.01367

**F=4.389**

The F value of 4.389 is greater than the table value of 2.76 making the results statistically significant.