**Statistics for Psychology - PSYCH-UH 1004Q)**

**Homework #6**

28 points

(The homework assignments will never require you to use R unless the problem explicitly says “use this R code”. For other problems, can use R if you find it useful, they should be completed easily by hand.)

1. One of the most common tools of the cognitive psychologist is the lexical decision task, in which a string of letters is flashed on a screen and an experimental participant must decide as quickly as possible whether those letters form a word. This exercise asks whether the reaction time required by subjects to recognize a word is a function of the frequency of that word’s use. The following data report the frequency and the mean reaction time of 12 English words. Calculate Pearson’s correlation coefficient for the data and test for statistical significance. (The frequency is log-transformed. You don’t need to worry about this for the problem. But, if you are curious, we log-transform the frequency counts of words because the frequency of words, and indeed a lot of things in the world, has a logarithmic distribution.)

|  |  |  |
| --- | --- | --- |
| **Word** | **Frequency** | **Reaction Time in milliseconds** |
| koala | 6.68 | 933 |
| afire | 5.09 | 819 |
| cobalt | 6.06 | 745 |
| slimy | 7.11 | 726 |
| vegan | 7.42 | 716 |
| shrimp | 7.6 | 705 |
| misery | 7.97 | 705 |
| curve | 8.98 | 676 |
| lesson | 8.99 | 673 |
| scroll | 8.66 | 603 |
| booth | 8.4 | 595 |
| turkey | 9.53 | 584 |

You can use the following R commands to load this data into R (just copy and paste into R):

library(tidyverse)

data <- tibble(word = c("koala", "afire", "cobalt", "slimy", "vegan", "shrimp", "misery", "curve", "lesson", "scroll", “booth”, "turkey"), frequency = c(6.68, 5.09, 6.06, 7.11, 7.42, 7.6, 7.99, 8.98, 8.99, 8.66, 8.4, 9.53), reactionTime = c(933, 819, 745, 726, 716, 705, 705, 676, 673, 603, 595, 584))

1. Use R to create a scatterplot of frequency and reaction time. (2 points)
2. Use R to calculate a correlation coefficient (r) between frequency and reaction time. Report the r here: (2 points)
3. What is the direction of this correlation? (1 point)
4. Do you find it to be a weak, moderate, or strong correlation? (1 point)
5. Use the equation for a *t*-test for *r* shown below to test whether this sample of words comes from a population that is different from a population with a correlation of 0 (ρ0). Plug in the values, and solve. Report the *t* here. (4 points)

$$t=\frac{r- ρ\_{0}}{\sqrt{\frac{1-r^{2}}{n-2}}}$$

1. Use the pt() function in R to determine the one-tailed *p*-value for this *t*. Report the *p*-value here. (2 points)

2. A researcher has found that caffeine intake increases heart rate. They find that for each increase of 25 milligrams (mg) of caffeine, heart rate increases by 2 beats per minute (bpm). This is based on the fact that average heart rate is 70bpm without any caffeine.

This result can be modeled with linear regression. Write the linear equation for this result in the form $ŷ=b\_{1}x+b\_{0}$ using the values the researcher found. (2 points)

3. A statistics professor has devised a diagnostic quiz, which, when given during the first class, can accurately predict a student’s performance on the final exam. So far data are available for 10 students, as follows:

|  |  |  |
| --- | --- | --- |
| **student** | **quiz** | **exam**  |
| 1 | 5 | 82 |
| 2 | 8 | 80 |
| 3 | 3 | 75 |
| 4 | 1 | 60 |
| 5 | 10 | 92 |
| 6 | 6 | 85 |
| 7 | 7 | 86 |
| 8 | 4 | 70 |
| 9 | 2 | 42 |
| 10 | 6 | 78 |

You can use the following R commands to load this data into R (just copy and paste into R):

library(tidyverse)

data <- tibble(student = c(1:10), quiz = c(5, 8, 3, 1, 10, 6, 7, 4, 2, 6), exam = c(82, 80, 75, 60, 92, 85, 86, 70, 42, 78))

First, let’s find the regression equation for predicting the final exam score from the quiz.

1. State the general linear regression equation. (1 point)
2. Use R to calculate the correlation coefficient (*r*) between quiz and exam*.* Report *r* here.(2 points)
3. Is it a positive or negative correlation? (1 point)
4. Is the strength of the relationship weak, moderate, or strong? (1 point)

1. Use R to create a linear model that uses the quiz score to predict the exam score. Report the full equation for the line with the correct slope and intercept values entered: (4 points)
2. Given the equation you found using R, what final exam score would be predicted for a student who scored a 9 on the quiz? Show your work by plugging values into the equation and solving. (2 points)
3. Use R to calculate a *t*-test for the slope of the linear regression. Report the following: (2 points)

*t*-statistic for the slope:

*p*-value for the slope *t*:

1. Calculate *r2* for this linear model. You can use R or do it by hand. Report the *r2* here. (1 point)