Exam 1.

Student name: Kalynn Wesby

Date: March 7th 2024

1) What is Descriptive and Inferential Statistics**? 5 points**

Descriptive statistics summarizes the characteristics of a data set while inferential statistics allows you to test a hypothesis or assess whether your data is generalizable to the broader population.

2) What is standard deviation**? 5 points**

The standard deviation is the average amount of variability in your data set

3) What is median value? **5 points**

The median value is the mean of the values at positions n/2 and (n/2) + 1.

4What is the mode and types of modes? **5 points**

Mode is the most common number to appear in the data set.

Types: unimodal - 1 mode

Bimodal – 2 modes

Trimodal – 3 modes

Multimodal – 4 modes

5) What’s the difference between the range and interquartile range? **5 points**

Range gives you the spread of the whole data set while the interquartile range gives you the spread of the middle half of a data set.

6) What is a normal distribution? **5 points**

A normal distribution is a data set that is symmetrically distributed with no skew.

7) What is the empirical rule**? 5 points**

This rule tells you where most of the values lie in a normal distribution and a quick way to get an overview of your data and check for any outliers or extreme values that don’t follow this pattern.

**8) What does correlation coefficient tell you? 5 points**

Correlation coefficient is a single number that describes the strength and direction of the relationship between your variables.

9) **What are the assumptions of the Pearson correlation coefficient**? **5 points**

Both variables are on an interval or ratio level of measurement

Data from both variables follow normal distributions.

Your data have no outliers.

Your data is from a random or representative samples.

You expect a linear relationship between the two variables

10) What are the main assumptions of T-test? **5 points**

The main assumptions of a T-test are the scale of measurement, random sampling, normality of data distribution, adequacy of sample size, and equality of variance in standard deviation.

**11) Calculate One-way ANOVA test with R: 50 points**

**Using** following data find a F stat value from One-way ANOVA test between columns “new\_cases\_smoothed”, and “total\_deaths”.

https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv

Email R program Exam1.R and output in answer document to get full points.

#Author: Wesby; Date: 03/07/2024; Purpose : Midterm

library(dplyr)

PATH <-"https://raw.githubusercontent.com/oowid/covid-19-data/master/public/data/owid-covid-data.csv"

df &lt;- read.csv(PATH)

anova\_one\_way &lt;- aov(new\_cases\_smoothed~total\_deaths, data = df)

summary(anova\_one\_way)

Df Sum Sq Mean Sq F value Pr(>F)

poison 1 0.9316 0.9316 20.67 3.96e-05 \*\*\*

Residuals 46 2.0735 0.0451

---

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

F value : - 16