University of Potsdam

Statistics Exercises 2019-06-16 **Exercise ID HW 2**

Name: ____________________

Student ID: ____________________

Declaration: This submission is my work alone; I did not consult anyone about it, and I did not use any other unfair means for obtaining the answer(s). [Your signature below implies that you have made this declaration.]

Signature: ____________________

Grades:

1. (a) ☐☐☐☐☐☐ . ☐☐☐

   (b) ☐☐☐☐☐☐ . ☐☐☐

   (c) ☐☐☐☐☐☐ . ☐☐☐

   (d) ☐☐☐☐☐☐ . ☐☐☐

   (e) ☐☐☐☐☐☐ . ☐☐☐

   (f) ☐☐☐☐☐☐ . ☐☐☐

2. (a) | | | | | | . | | |

(b) | | | | | | . | | |

(c) | | | | | | . | | |

(d) | | | | | | . | | |

(e) | | | | | | . | | |

(f) | | | | | | . | | |

(g) | | | | | | . | | |

(h) | | | | | | . | | |

(i) | | | | | | . | | |

(j) | | | | | | . | | |

(k) | | | | | | . | | |

1. All answers must be to three decimal places. I have set the tolerance limit to this problem such that the system should accept the correct answer to a tolerance of ± 0.1.

   You are given a sample in the text file called `sample2.txt`. Read in the file into R:

   ```
   > fl<-"http://www.ling.uni-potsdam.de/~vasishth/data/sample2.txt"
   > data<-read.table(fl,
   +                  header=TRUE)
   > colnames(data)

   [1] "x"
   ```

   Carry out a t-test on this sample of scores using R; the null hypothesis (H_0) and alternative hypothesis (H_1) are:

   $H_0 : \mu = 92$ and $H_1 : \mu \neq 92$.

   We are going to do a t-test by hand. **All numerical answers must be to three decimal places**.

   (a) the sample mean
   (b) the sample standard deviation
   (c) the sample size
   (d) the standard error
   (e) the **absolute** critical t-value
   (f) the **absolute** observed t-value

2. [**All numerical answers must be to three decimal places**. I have set the tolerance limit to this problem such that the system should accept the correct answer to a tolerance of ± 0.1.]

   This is your first real data-set! This is data from a famous paper in sentence processing, by Grodner and Gibson, 2005 (Experiment 1).

   The full paper can be downloaded from moodle (week 6).

   The data contains differences in reading time (milliseconds) in subject versus object relatives in English. Each row represents data from one participant; each participant sees subject as well as object relatives (a so-called within subjects design).

   The research question is that subject relatives are easier to process than object relatives. We are going to test this hypothesis.

   First, load the data:

```
> fl<-"http://www.ling.uni-potsdam.de/~vasishth/data/rcdata.txt"
> data<-read.table(fl,header=TRUE)
> head(data)
```

```
  objgap subjgap
1 335.00  310.50
2 419.38  398.88
3 247.25  346.75
4 616.38  416.38
5 395.75  362.88
6 330.25  320.50
```

Next, take the difference between the two columns (the order matters!). **Do not round the differences yet**.

```
> ## difference between object and subject relatives
> d<-data[,1]-data[,2]
> ## vector of differences:
> str(d)
```

```
 num [1:42] 24.5 20.5 -99.5 200 32.9 ...
```

Carry out a t-test on this sample of scores d using R; the null hypothesis (H_0) and alternative hypothesis (H_1) are:

$H_0 : \mu = 0$ and $H_1 : \mu \neq 0$.

The above null hypothesis states that there is no difference between subject and object relatives' processing times. The alternative hypothesis states that the difference is not zero, i.e., it could be positive or negative.

We are going to do a t-test using the t.test function, and also by hand.

(a) Run the function t.test(d), and then write the observed t-value from the model, rounded to three decimal places.

(b) By looking at the output of the function t.test(d), write the p-value from the model. Again, round to three decimal places.

(c) From the output of the function t.test(d), write the lower bound of the 95% confidence interval. Round to three decimal places.

(d) From the output of the function t.test(d), write the upper bound of the 95% confidence interval. Round to three decimal places.

(e) From the output of the function t.test(d), write the sample mean. Round to three decimal places.

(f) Now, using the mean() function, compute the sample mean of the vector d, and write it here. Round your answer to three decimal places.

(g) From the data d, compute and write down the standard error. Round your answer to three decimal places.

(h) Using the standard error and the sample mean, write the absolute observed t-value. Round your answer to three decimal places.

(i) In the t-distribution with 41 degrees of freedom, what is the quantile q such that the area to the right of it is 0.10? I.e., for a t-distribution with 41 degrees of freedom, what is the quantile value q such that $Prob(T > q) = 0.10$? Round your answer to three decimal places.

(j) Computing the **95% confidence interval** is purely conventional. One could compute, for example, an 80% confidence interval. Compute the lower bound of the 80% confidence interval. Round your answer to three decimal places.

(k) Following up on the preceding question, compute the upper bound of the 80% confidence interval. Round your answer to three decimal places.