# University of Potsdam

**Statistics Exercises 2019-06-16** **Exercise ID HW 3**

**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) | | | | | | | . | | |

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

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

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

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

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

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

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

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

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

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

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

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

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

(l) | | | | | | | . | | |

## Exercises: HW 3



1. [**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 $\pm 0.1$.]

   This is a real data-set. This is data from experiment 1 of:

   https://doi.org/10.1371/journal.pone.0100986

   The data contains a repeated measures design that investigates reading time (milliseconds) in subject versus object relatives in Chinese. The research question is that object relatives are easier to process than subject relatives. We are going to test this hypothesis by setting up the null hypothesis that there is no difference in reading time between object and subject relatives.

   **Be careful to subtract the subject relative reading time from the object relative reading time. Otherwise you will get the wrong sign on the t-value.**

   First, load the data:

```
> fl<-"http://www.ling.uni-potsdam.de/~vasishth/data/gibsonwu2012datarepeat.txt
> data<-read.table(fl,
+                  header=TRUE)
```

   (a) What is the observed t-value for the by-subjects paired t-test?

   (b) What is the p-value for the by-subjects paired t-test?

   (c) What is the lower bound of the 95% confidence interval of the difference between the two conditions in the by-subjects paired t-test?

   (d) What is the upper bound of the 95% confidence interval of the difference between the two conditions in the by-subjects paired t-test?

   (e) What is the observed t-value for the by-items paired t-test?

   (f) What is the p-value for the by-items paired t-test?

   (g) What is the lower bound of the 95% confidence interval of the difference between the two conditions in the by-items paired t-test?

   (h) What is the upper bound of the 95% confidence interval of the difference between the two conditions in the by-items paired t-test?

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 $\pm 0.1$.]

   This is a real data-set. This is data from experiment 7 of:

   https://doi.org/10.1016/j.jml.2018.07.004.

   There are four conditions (2x2 repeated measures factorial design):

- One factor is syntactic difficulty: Conditions a,b: easy sentences; conditions c,d: hard sentences
- The other factor is syntactic predictability: Conditions a,c: unpredictable; conditions b,d: predictable.

The research questions are:

- is there a main effect of difficulty (**we want to compare reading times for easy minus hard; the order matters otherwise the sign of the t-value will be wrong**.)
- is there a main effect of predictability (**we want to compare reading times for predictable minus unpredictable; the order matters otherwise the sign of the t-value will be wrong**.)
- is there an interaction between difficulty and predictability? **In other words, we want the difference between easy-minus-hard and predictable-minus-unpredictable**.

The dependent measure is total reading time from an eyetracking study at a particular word of the sentences (the verb).

First, load the data, isolate the relevant columns, and create new colums for the two factors (difficulty and predictability).

```
> fl<-"http://www.ling.uni-potsdam.de/~vasishth/data/data_LK13rep100subj.txt"
> data<-read.table(fl,
+                  header=TRUE)
> head(data)

  subject trial itemid condition list answer RESPONSE_ACCURACY roi FFD FFP SFD
1       1     1      1         p   40      0                 -2   1 164   1   0
2       1     1      1         p   40      0                 -2   2 155   1   0
3       1     1      1         p   40      0                 -2   3 208   1   0
4       1     1      1         p   40      0                 -2   4 176   1 176
5       1     1      1         p   40      0                 -2   5 240   1   0
6       1     1      1         p   40      0                 -2   6   0   0   0
  FPRT RBRT TFT RPD CRPD RRT RRTP RRTR RBRC TRC LPRT
1  297  297 297 297  297   0    0    0    0   0  297
2  290  290 606 290  587 316    0  316    0   0  316
3  208  208 449 208  795 241    0  241    0   0  241
4  176  176 176 176  971   0    0    0    0   0  176
5  240  240 442 240 1211 202    0  202    0   0  202
6    0    0   0   0 1211   0    0    0    0   0    0
```

```
> data<-subset(data,
+              condition%in%c("a","b","c","d"))
> data<-data[,c(1,3,4,14)]
> data$condition<-factor(data$condition)
> str(data)

'data.frame':        83440 obs. of  4 variables:
 $ subject  : int  1 1 1 1 1 1 1 1 1 1 ...
 $ itemid   : int  24 24 24 24 24 24 24 24 24 24 ...
 $ condition: Factor w/ 4 levels "a","b","c","d": 1 1 1 1 1 1 1 1 1 1 ...
 $ TFT      : int  271 428 0 279 507 535 228 694 364 406 ...

> data$difficulty<-factor(ifelse(data$condition%in%c("a","b"),
+                                "easy","hard"))
> data$predictability<-factor(ifelse(data$condition%in%c("a","c"),
+                                    "unpredictable","predictable"))
```

(a) What is the observed t-value for the by-subjects paired t-test for the effect of difficulty?

(b) What is the p-value for the by-subjects paired t-test for the effect of difficulty?

(c) What is the observed t-value for the by-items paired t-test for the effect of difficulty?

(d) What is the p-value for the by-items paired t-test for the effect of difficulty?

(e) What is the observed t-value for the by-subjects paired t-test for the effect of predictability?

(f) What is the p-value for the by-subjects paired t-test for the effect of predictability?

(g) What is the observed t-value for the by-items paired t-test for the effect of predictability?

(h) What is the p-value for the by-items paired t-test for the effect of predictability?

(i) What is the observed t-value for the by-subjects paired t-test for the distance x predictability interaction?

(j) What is the p-value for the by-subjects paired t-test for the distance x predictability interaction?

(k) What is the observed t-value for the by-items paired t-test for the distance x predictability interaction?

(l) What is the p-value for the by-items paired t-test for the distance x predictability interaction?