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) □□□□□□□□ □□□□□

2. (a) □□□□□□□□ □□□□□
   (b) □□□□□□□□ □□□□□
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 ±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.

First, load the data:

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

Then, aggregate the data by subject:

```r
> rctypebysubj<-aggregate(rt~subj+condition,mean,data=data)
```

Then fit a linear mixed model investigating the effect of condition on reading time (rt) using the command shown in the linear modeling lecture.

(a) What is the observed t-value for the by-subjects effect of condition in the linear mixed model?

(b) What is the lower bound of the 95% confidence interval of the difference between the two conditions in the by-subjects linear mixed model? Use the approximation that the 95% confidence interval of the effect is \( \bar{x} \pm 2 \times SE \), where \( \bar{x} \) is the estimated difference in means between the two conditions, and SE is the standard error of this difference.

(c) What is the upper bound of the 95% confidence interval of the difference between the two conditions in the by-subjects linear mixed model? Use the approximation that the 95% confidence interval of the effect is \( \bar{x} \pm 2 \times SE \), where \( \bar{x} \) is the estimated difference in means between the two conditions, and SE is the standard error of this difference.

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

There are four conditions (2x2 repeated measures factorial design):
Exercises: Class activity 3

• 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?
• is there a main effect of predictability?

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

First, load the data, isolate the relevant columns, and create new columns for the two factors (difficulty and predictability) with sum contrast coding. Then aggregate data by subjects, with difficulty and predictability as predictors.

```R
> 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 FPRT RBRT TFT RPD CRPD RRT RTP RBRC TRC LPRT
1 1 1 1 1 p 40 0 -2 1 164 1 0 297 0 0 0 0 0 297
2 1 1 1 1 p 40 0 -2 2 155 1 0 290 0 0 0 0 0 290
3 1 1 1 1 p 40 0 -2 3 208 1 0 208 0 0 0 0 0 208
4 1 1 1 1 p 40 0 -2 4 176 1 176 176 0 0 0 0 0 176
5 1 1 1 1 p 40 0 -2 5 240 1 0 240 0 0 0 0 0 240
6 1 1 1 1 p 40 0 -2 6 0 0 0 0 0 0 297 0 0 0 0 0 297

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

'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$difficulty<-ifelse(data$condition%in%c("a","b"),1,-1)
> data$predictability<-ifelse(data$condition%in%c("a","c"),1,-1)
> bysubj<-aggregate(TFT~subject+predictability+difficulty,mean,data=data)

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