University of Potsdam

Quiz Solutions: Frequentist foundations

Date March 11, 2020

Approximate time needed: 120 minutes

Name: _____

Instructions on how to answer questions:

- **Always write down your code/work** so that I can give partial credit for answers that basically are in the right direction. If you just write the answer and it turns out to be wrong, I don't know why it was wrong and you would lose all the points.
- When I ask you to write text, please keep your answers brief and to the point.

The maximum number of points is 100.

In all questions, assume that Type I error (otherwise known as α) is 0.05 unless stated otherwise.

Question 1

[10+10+10+10=40 points]

In the space for answers below, please write down only the key commands for doing the t-test and the linear mixed model. But after the exam is over, please also email me your code for Question 1 (Email: vasishth.shravan@gmail.com) with the subject line "Quiz Frequentist Foundations" and the file name YOURLASTNAME.R.

Download the data-set provided here:

https://tinyurl.com/y3g7lg8u

Then run the following commands to load the lme4 library and to set up your data for analysis:

```
library(lme4)

## Loading required package:  Matrix

## load data:
dat<-read.csv("data/E1_data.csv",header=TRUE)
## convert RT to milliseconds:
dat$RT<-dat$RT*1000
## choose critical region:
word_n<-4
## subset critical data:
## use this data in all questions below.
crit<-subset(dat,Position==word_n)
```

The data consist of a repeated measures experiment comparing two conditions which are labeled Type 1 and Type 2. The column Sub refers to subject id, and the column ID refers to item id. RT refers to reading time in seconds (we have converted it above to milliseconds); NA is missing data. You can ignore the other columns. This is a standard Latin square design.

(i) Using the data frame called crit above, with RT as a dependent variable, carry out the appropriate **by-subjects t-test** to evaluate the null hypothesis that there is no difference between the conditions labeled Type 1 and 2. Write down **all** the R commands needed to do the appropriate t-test,

and the resulting t-value and p-value. State whether we can reject the null hypothesis given the results of the t-test; explain why.

Answer:

```
crit$SO<-ifelse(crit$Type==1,1,-1)
## a positive slope will mean SRs are harder

## correct t-test
crit_subj<-aggregate(RT~Sub+Type,mean,na.rm=TRUE,
                     data=crit)
t.test(RT~Type,crit_subj,paired=TRUE)


##
##  Paired t-test
##
## data:  RT by Type
## t = 1.79, df = 45, p-value = 0.08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.6412 80.1111
## sample estimates:
## mean of the differences
##                  37.735
```

No we can't reject the null hypothesis, as the critical t-value is larger than the observed t-value.

(ii) Now, **using the object called crit above as the data frame**, fit a linear mixed model (called M0) with the column called Type coded as **sum contrasts**. Assume random intercepts for subjects and random intercepts for items (random intercepts are sometimes called varying intercepts—these terms are synonymous). Write down the linear mixed models command, and write down the fixed-effects estimates (Intercept and slope) along with their standard errors. State whether we can reject the null hypothesis given the results of the t-value shown in the linear mixed model output; explain why.

Answer:

```
M0<-lmer(RT~SO+(1|Sub)+(1|ID),
           crit)
summary(M0)

## Linear mixed model fit by REML ['lmerMod']
## Formula: RT ~ SO + (1 | Sub) + (1 | ID)
##    Data: crit
##
## REML criterion at convergence: 38936
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -4.148 -0.423 -0.136  0.238  7.498
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  ID       (Intercept)   1864    43.2
##  Sub      (Intercept) 100336   316.8
##  Residual             130431   361.2
## Number of obs: 2651, groups:  ID, 64; Sub, 46
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   686.01      47.54   14.43
## SO             18.94       7.03    2.69
##
## Correlation of Fixed Effects:
##    (Intr)
## SO 0.002
```

Answer: Yes, we can reject the null because the t-value is larger than 2, which is the approximate critical t-value here.

(iii) Why do the results of the t-test and the linear mixed model M0 differ?

Answer:

Because we aggregate over items in the by-subjects t-test, but we do not aggregate by items in the linear mixed model. In the latter case, there is more source of variance, due to items.

(iv) The researcher wants to achieve 80% statistical power in a future study.

Based on the available data above, she determines that the **standard error** (note: not the standard deviation!) of the difference in means between the conditions Type 1 and Type 2 is 21. She assumes that the true difference in means is 30 ms. What is the number of participants (to the nearest whole number) needed to achieve approximately 80% power? Use the power.t.test function to compute your answer. Write down the exact power.t.test function specification you used, as well as the number of participants needed, based on the output of the power.t.test function.

Answer:

```
d<-subset(crit_subj,Type==1)$RT-subset(crit_subj,Type==2)$RT
stddev<-sd(d)
n<-length(d)
SE<-round(stddev/sqrt(n))
```

```
nsubj<-length(unique(crit$Sub))
power.t.test(delta = 30,
             sd=SE*sqrt(nsubj),
             power = .80,
             sig.level = 0.05,
             type = "one.sample",
             alternative = "two.sided")$n

## [1] 178.84
```

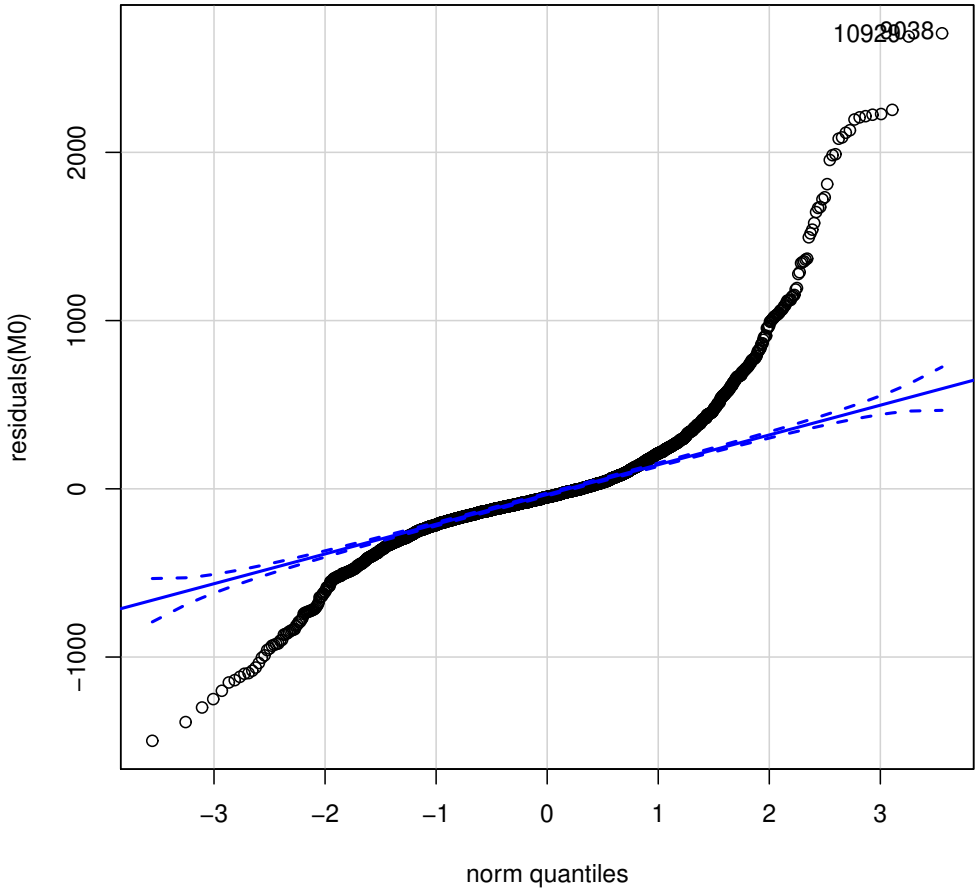
Question 2

[10 points]

The plot below shows the distribution of the residuals from model M0 plotted against the standard normal distribution with mean 0 and standard deviation 1. Explain what the plot tells us about one of the model assumptions of the linear mixed model M0 that you fit earlier.

(You can ignore the numbers below the plot.)

```
library(car)
qqPlot(residuals(M0))
```



9038 10929 726 2544

Answer:

The normality assumption of the residuals is not satisfied.

Question 3

[10+10=20 points]

(i) Using only your estimates (computed in Question 1) of the intercept and the slope in model M0's fixed effects output, write down the mean of the condition labeled Type 1 in the data, and the mean of the condition labeled Type 2.

Answer: Intercept + slope = mean for condition coded +1, Intercept - Slope= mean for ondition coded -1.

```
Intercept<-summary(M0)$coefficients[1,1]
Slope<-summary(M0)$coefficients[2,1]

## condition coded +1, Ttpe 1:
Intercept + Slope

## [1] 704.95

## condition coded -1, Type 2:
Intercept - Slope

## [1] 667.08
```

(ii) Suppose that the model M0's output for the fixed effects analysis were as follows. SO is a +/-1 sum-coded contrast specification for the conditions labeled Type.

```
results<-summary(M0)$coefficients
results[2,3]<-2
results[2,2]<-NA
results

##             Estimate Std. Error t value
## (Intercept)  686.011     47.538  14.431
## SO            18.936         NA   2.000
```

Answer:

What is the value of the standard error of the slope (SO), which is labeled NA above?

Answer:

t= 18.936/SE

Solve for SE.

```
18/2
```

```
## [1] 9
```

Question 4

[5+5=10 points]

(i) A paired t-test is done with data from 10 participants. The t-value from the test is 2.1. What is the p-value associated with a **two-sided null hypothesis test**? You are given the following output to help you:

```
pt(2.1,df=10)
```

```
## [1] 0.96896
```

```
pt(2.1,df=9)
```

```
## [1] 0.96744
```

No calculation is needed; just show how you used the above information to get an expression that represents the p-value. But if you do calculate the answer, please give the answer to three decimal places. Be sure to show your work, based on the above information.

Answer:

```
round(2*pt(-2.1,df=9),3)
```

```
## [1] 0.065
```

(ii) If the p-value from a **two-sided null hypothesis test** had been 0.09, what would be the associated **absolute** t-value (i.e., ignoring the sign on the t-value)? Use the following information to work out the answer. The number of participants is 10, as above.

```r
qt(0.09,df=10)
```

```
## [1] -1.4416
```

```r
qt(0.09,df=9)
```

```
## [1] -1.4537
```

```r
qt(0.09/2,df=10)
```

```
## [1] -1.8768
```

```r
qt(0.09/2,df=9)
```

```
## [1] -1.8992
```

```r
qt(2*0.09,df=10)
```

```
## [1] -0.95936
```

```r
qt(2*0.09,df=9)
```

```
## [1] -0.96449
```

No calculation is needed. Be sure to show your reasoning so I know how you got the answer.

Answer:

```r
abs(qt(0.09/2,df=9))
```

```
## [1] 1.8992
```

Question 5

[10+10=20 points]

(i) A researcher fits a linear mixed model to compare the reading times between two conditions (a) and (b), just like in the above study in Question 1. Her hypothesis is that the mean for condition (a) is larger than the mean for (b). She observes that condition a has sample mean 500 ms, and condition (b) has sample mean 450 ms. She also establishes from the linear mixed model that the t-value is 1.94. The approximate p-value associated with this t-value is 0.052. Answer the following: (A) Do we have evidence against the null hypothesis and (B) do we have evidence for the particular research hypothesis that the researcher has?

Answer: A: Not strictly speaking, but we do have some weak evidence against the null.

B: No. The significance test only tells us the evidence against the null, not the evidence for a particular alternative.

(ii) The researcher runs the same analysis as above on a new data-set that has the same design as above, and now gets a p-value of 0.001. Now she has stronger evidence than in the above case where the p-value was 0.052. What does she have stronger evidence for? Explain your answer in one sentence.

Answer: She has stronger evidence against the null. She doesn't have stronger evidence for the particular alternative of interest.

Question 6 (Extra credit)

[10 points]

Given that Type I error is 0.01; what is the highest value possible for Type II error?

Answer: The lowest possible power that we can have is 0.01. That means the highest possible Type II error is 1-0.01=0.99.