University of Potsdam
Introduction to Statistical Data Analysis, Final Exam, SU15

Instructions: Answer all questions. Be brief in your responses, and answer the question asked. Answers can be in English or German.

This is a closed-book exam; you cannot use any lecture notes or any other notes. The internet or any other electronic means cannot be used either for answering questions.

Do not write your name on this sheet!

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: ______________________________

Question 1

[20 points]

Circle the letter corresponding to the correct answer.

1. Standard error is
   - a the standard deviation of the sample scores
   - b the standard deviation of the distribution of sample means
   - c the sample variance
   - d 2 times the standard deviation of sample scores
2. A p-value is
   - a the probability of the null hypothesis being true
   - b the probability of the null hypothesis being false
   - c the probability of the alternative hypothesis being true
   - d the probability of getting the sample mean that you got (or a value more extreme) assuming the null hypothesis is true
   - e the probability of getting the sample mean that you got (or a value less extreme) assuming the null hypothesis is true
3. If Type I error probability, alpha, is 0.05 in a t-test, then
   - a we have a 5% probability of rejecting the null hypothesis when it is actually true
   - b we have a 95% probability of rejecting the null hypothesis when it is actually true
   - c we necessarily have low power
   - d we necessarily have high power
4. Type II error probability is
   - a the probability of "accepting" the null when it's true
   - b the probability of "accepting" the null when it's false
   - c the probability of rejecting the null when it's true
   - d the probability of rejecting the null when it's false
5. When power increases
   - a Type II error probability decreases
   - b Type II error probability increases
   - c Type II error probability remains unchanged

Question 2

[10 points]

We have data from men and women which represent F1 (first formant) frequencies of vowels. You can assume that for each gender, all the vowel F1 values within a language come from different subjects (note that no subject information is provided). Here is the entire dataset:

```
dat

##    female male vowel  language
## 1     391  339     i  W.Apache
## 2     561  512     e  W.Apache
## 3     826  670     a  W.Apache
## 4     453  427     o  W.Apache
## 5     358  291     i CAEnglish
## 6     454  406     e CAEnglish
## 7     991  706     a CAEnglish
## 8     561  439     o CAEnglish
## 9     398  324     u CAEnglish
## 10    334  307     i   Ndumbea
## 11    444  361     e   Ndumbea
## 12    796  678     a   Ndumbea
## 13    542  474     o   Ndumbea
## 14    333  311     u   Ndumbea
## 15    343  293     i      Sele
## 16    520  363     e      Sele
## 17    989  809     a      Sele
## 18    507  367     o      Sele
## 19    357  300     u      Sele
```

You are asked to do a t-test. For each of the t-tests below, state whether it is an appropriate test for testing the null hypothesis that there is no difference between the F1 frequencies by gender. For each test below, briefly explain why it is appropriate, or why it is not appropriate. **Keep your answers short!**

1. `t.test(dat$female,dat$male)`

   Answer:

2. `t.test(dat$female-dat$male)`

   Answer:

3. ```r
   t.test(dat$female,dat$male,paired=TRUE)
   ```

   Answer:

4. ```r
   t.test(dat$male,dat$female,paired=TRUE)
   ```

   Answer:

5. ```r
   t.test(dat$male)
   t.test(dat$female)
   ```

   Answer:
   ```

Question 3

[20 points]

An eyetracking (reading) study has been done with 30 subjects. 157 sentences were read by each subject, and total reading times on each word are recorded. We are interested in determining whether there is a statistically significant effect of word complexity (computed using a method we are not concerned with) and word frequency on total reading time.

Each of these predictors is centered; that is, the mean word complexity is subtracted from each value of word complexity, and mean word frequency is subtracted from each value of word frequency.

We fit a linear mixed model with varying intercepts:

```
> summary(m_TFT)
Linear mixed model fit by REML ['lmerMod']
Formula: TFT ~ c_word_complex + c_word_freq + (1 | subj)
   Data: etm_TFT

Random effects:
 Groups   Name        Variance Std.Dev.
 subj     (Intercept)  4182     64.67
 Residual             47222    217.31
Number of obs: 47996, groups:  subj, 30

Fixed effects:
               Estimate Std. Error t value
(Intercept)    325.4946    11.8524   27.46
c_word_complex  16.0319     0.9338   17.17
c_word_freq    -36.1198     1.1740  -30.77
```

Explain the interpretation of each of the estimated coefficients, and explain whether there is a statistically significant effect of word length and word frequency on total reading time (at Type I error 0.05).

Answer:

Question 4

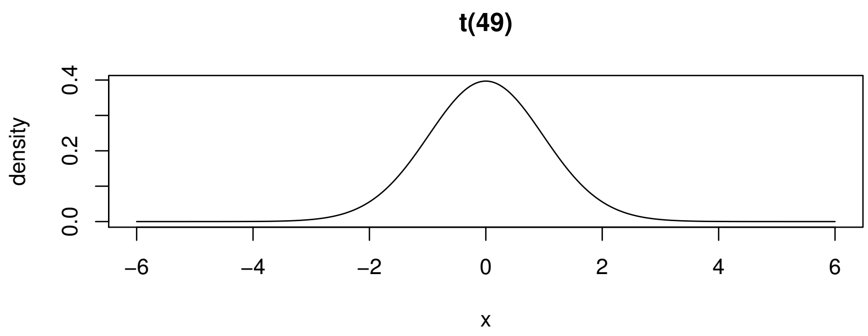
[10 points]

Given data with sample size $n = 50$, we want to do a two-sided hypothesis test, assuming that the null hypothesis is that $\mu = 0$. We use the t-distribution with n-1 degrees of freedom for this. We can compute the critical value for the rejection regions for different Type I errors (α values) as follows:

R command	critical t-value
`qt(0.025,df=49)`	-2
`qt(0.05,df=49)`	-1.7
`qt(0.10,df=49)`	-1.3

Table 1: Different t-values for various α.

The null hypothesis distribution is shown below.



(a) In the figure above, draw two vertical lines to mark the critical rejection regions for $\alpha = 0.10$; these lines mark the points on the x-axis beyond which one would reject the null. An approximate positioning of the lines is acceptable. Shade the rejection regions as well.

(b) Suppose that the true distribution were centered around 1. Draw this true distribution either on top of the null distribution shown above, or in the space below the null distribution shown above.

(c) Shade the Type II error region in the appropriate figure.

(d) If the true distribution had mean -2 instead of 1, would Type II error increase or decrease? No figure needs to be drawn.

Answer:

(e) The sample mean is 5.72. Would you reject the null hypothesis at $\alpha = 0.10$? Would you reject the null if we had decided on $\alpha = 0.05$?

Answer:

Question 5

[10 points]

In this question, you have to use the fact that $t = \frac{\bar{x}-\mu}{SE}$, where $\bar{x}$ is the sample mean, μ is the null hypothesis mean, and $SE = \hat{\sigma}/\sqrt{n}$, where $\hat{\sigma}$ is the estimated standard deviation of the sample, and n is the sample size.

Assume here that the critical t-value for rejecting the null hypothesis at $\alpha = 0.05$ is approximately 2.

Suppose we run an experiment with sample size $n = 100$. The sample mean is 46.76 and the sample standard deviation is 103.15. Let the null hypothesis be that the true mean is 50.

(a) What t-value does the data give us? You do not need to do the full calculation, just write the expression for t.

Answer:

(b) Would the t-value increase or decrease if the null hypothesis had been that the true mean is -50 instead of 50? (Note: No calculation needs to be done here.)

Answer:

(c) If the null hypothesis had been that the true mean is 0, what t-value would the data give us? You do not need to do the full calculation, just write the expression for t.

Answer:

(d) Define the p-value.

Answer:

(e) A t-test on the data, assuming that the null hypothesis is 50, gives us the p-value 0.75. Would you reject the null hypothesis based on this p-value, at Type I error 0.05? Explain why or why not.

Answer:

Question 6

[10 points]

1. A researcher is planning to do an experiment, and wants to work out what her sample size should be to obtain Type II error of 0.05 and Type I error probability 0.05. She plans to do a one-sample t-test (two-sided). The researcher estimates that the true effect will have value 50. In order to estimate the standard deviation, she looks at some relevant previous research and finds out that a **standard error** is reported as 10, for an experiment with sample size $n = 100$.

   Complete the following command that she would have to type to find her sample size given the above constraints. Fill in the three values marked ??.

   ```
   power.t.test(d=??,
                type="one.sample",
                alternative="two.sided",
                sd=??,
                power=??)
   ```

   Answer:

   (a) d=
   (b) sd=
   (c) power=

2. Then she runs the experiment with the appropriate sample size provided by the above command.

   If her p-value is 0.13, would it be reasonable to conclude that her evidence favors the null hypothesis?

   Answer:

3. If her Type II error had been 0.80 instead of 0.05, and if she had obtained a p-value of 0.13, would it be reasonable to conclude that her evidence favors the null hypothesis?

   Answer:

Question 7

[20 points]

A self-paced reading experiment has been done to study the effect of relative clause type (subject vs object relatives) and sentence complexity (simple vs complex) on reading time. We have repeated measures data from 60 subjects. Each subject saw each condition 6 times; this is a fully balanced standard Latin Square design. There are no missing values.

There are therefore four conditions, labeled a,b,c,d.

1. a,b: subject relative
2. c,d: object relatives
3. a,c: simple
4. b,d: complex

The data look like this:

```
##     subj item cond   rt
## 103    1   14    b  438
## 121    1   16    d  531
## 154    1   15    c  422
## 171    1   18    b 1000
## 219    1    4    d  344
## 325    1   17    a  406
```

Because we are interested in the effect of RC type, we define a column vector in the data frame `dat` called `RCType` with the following contrast coding: subject relatives: -1, object relatives: +1.

Next, we fit a linear mixed model with varying intercepts for subjects:

```
library(lme4)
m<-lmer(rt~RCType+(1|subj),dat)
summary(m)

## Linear mixed model fit by REML ['lmerMod']
## Formula: rt ~ RCType + (1 | subj)
##    Data: dat
##
## REML criterion at convergence: 22840.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9274 -0.4894 -0.2265  0.1715 10.2617
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subj     (Intercept)  85656   292.7
##  Residual             426495   653.1
## Number of obs: 1440, groups:  subj, 60
```

```
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   935.61      41.52  22.535
## RCType        -66.14      17.21  -3.843
##
## Correlation of Fixed Effects:
##        (Intr)
## RCType 0.000
```

The questions are about the above output.

(a) Explain what the intercept and slope terms in the fixed effects mean.

Answer:

(b) If the research question is, are subject relatives easier to process than object relatives, what is the conclusion from this study? Explain with reference to the statistical hypothesis test.

Answer:

(c) What is the estimated variance of the residual error ε?

Answer:

(d) What does the variance in the random effects output (shown below) represent:

```
Random effects:
 Groups   Name        Variance Std.Dev.
 subj     (Intercept)  85656   292.7
```

Answer:

(e) What are the estimated means of the subject relative and object relatives?

Answer:

(e) If we had had treatment contrast coding for the factor relative clause type, with 0 representing **subject** relatives and 1 representing object relatives, what would have the intercept and slope estimates been in the linear mixed model above?

Answer:

(f) If the coding had been 0 for **object** relatives and 1 for subject relatives, what would the intercept and slope estimates have been in the above model?

Answer:

(g) Next, we look at the effect of complexity. The research question is: are complex sentences more difficult to process than simple sentences?

We create a new column vector called `comp`, in which simple sentences are coded 1 and complex sentences are coded -1. So this is also a sum contrast as done above for RC Type.

You are given that the mean for the complex condition is 1018.37 and the mean for the simple condition is 852.85. A linear mixed model with varying intercepts is fit as below:

```
m2<-lmer(rt~comp+(1|subj),dat)
```

(i) Using the means provided above for the two conditions, simple and complex, give the estimated fixed effects intercept and slope coefficients for this linear mixed model. In other words, write down the numbers that should appear in the column labeled Estimate in the output below:

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)    ????      41.52  22.535
comp           ????      17.16  -4.824
```

Answer:

(ii) Is the difference between complex and simple sentences statistically significant at Type I error 0.05?

Answer:

END OF PAPER