Data Analytics & Communication Assignment 5

Micky Labreche s4021290 & Yann Amado s5091128

23-12-2021

1 T-test

(a) We perform a unpaired T-test on the data. Despite the fact that the low- and high coherence trials are both performed by a single subject, we need to have an equal amount of samples to do a paired test. This is not the case, so we do an unpaired test.

```
subj1 <- dat %>% filter(subjNo == 1)

onlyDots <- subj1 %>% filter(isDots == 1)

# selecting only correct trials
onlyDots <- onlyDots[onlyDots$ER != 1,]

# We remove the outliers
cleaned <- onlyDots %>%
  filter(!(abs(RT - median(RT)) > 2*sd(RT)))



cohFacHigh <- onlyDots %>% filter(cohFac == 1)
cohFacLow <- onlyDots %>% filter(cohFac == 0)

sd(cohFacHigh$RT)
```

```
## [1] 0.2958375
```

```
sd(cohFacLow$RT)
```

```
## [1] 0.5464982
```

```
t.test(cohFacLow$RT, cohFacHigh$RT, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  cohFacLow$RT and cohFacHigh$RT
## t = 11.63, df = 645.85, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  0.2741558 0.3855458
## sample estimates:
## mean of x mean of y
## 1.2268374 0.8969866
```

(b) After performing the T-test, we find a significant difference in the response times between a high coherence of dots and a low coherence of dots, $t(869) = 12.3$, $p < .05$. The participant responded slower in the low coherence condition (M=1.22, SD=0.5) than in the high coherence condition (M=0.90, SD=0.3).

(c) The T-test we need is an paired T-test, because the two groups contain the same participants.

(d) Now we use the median to get the average RT per participant. This is more robust to outliers.

```
onlyDotsAll <- dat[dat$isDots==1,]
average <- onlyDotsAll %>%
  group_by(cohFac, subjNo) %>%
  summarise_all(median)

cohFacHighAll <- average %>% filter(cohFac == 1)
cohFacLowAll <- average %>% filter(cohFac == 0)

t.test(cohFacLowAll$RT, cohFacHighAll$RT, paired=TRUE)
```

```
##
##  Paired t-test
##
## data:  cohFacLowAll$RT and cohFacHighAll$RT
## t = 5.9424, df = 22, p-value = 5.575e-06
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  0.1446318 0.2997043
## sample estimates:
## mean difference
##        0.222168
```

After performing the T-test, we find a significant difference in the response times between a high coherence of dots and a low coherence of dots, $t(22) = 5.94$, $p < .05$. The mean of the differences is 0.22.

(e) We calculate the effect size:

```
library(lsr)
cohensD(cohFacLowAll$RT, cohFacHighAll$RT)
```

```
## [1] 1.201406
```

A value of 1.2 corresponds to a quite large effect.

2 Non-parametric alternatives

(a) As an alterative we perform a Kruskall Wallis test for non-parametric data. This ranks the data instead of using the real values. This is because the data is not normally distributed. You can see this if you look at all the outliers for instance.

```
kruskal.test(RT ~ cohFac, data = average)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  RT by cohFac
## Kruskal-Wallis chi-squared = 13.703, df = 1, p-value = 0.0002141
```

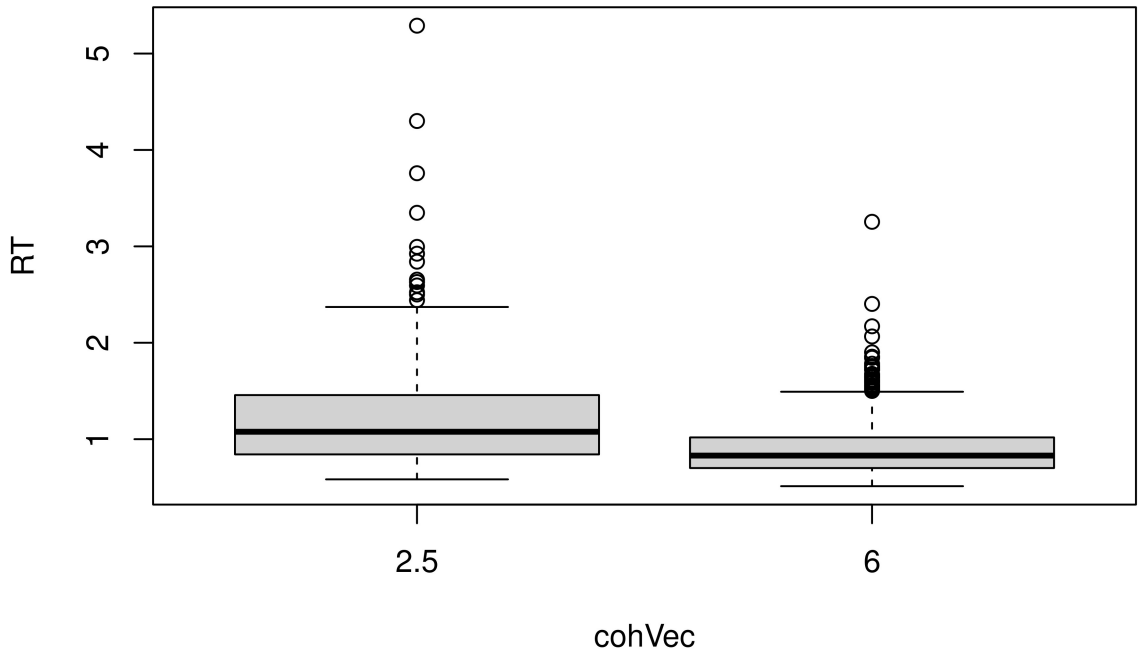
(b) We find that H(1)=13.7, p $<$ .05, which concludes a significant difference in the response times between a high coherence of dots and a low coherence of dots.

3 One-way ANOVA

```
onlyDots$cohVec <- as.factor(onlyDots$cohVec)
aov.onlyDots <- aov(onlyDots$RT ~ onlyDots$cohVec, data = onlyDots)
summary(aov.onlyDots)
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## onlyDots$cohVec    1  28.23  28.228   159.1 <2e-16 ***
## Residuals       1059 187.91   0.177
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
boxplot(RT ~ cohVec, data = onlyDots)
```



(a) A one-way ANOVA indicated a significant effect of task condition (dots versus control) on RT ($F(1, 1059) = 159.1$, $p < 0.001$). This indicates that there is a significant difference in the response times between a high coherence of dots and a low coherence of dots. The boxplot also indicates that the central tendency and the overall distribution is different.

b) Both tests presented very similar results in relation to the p-value, with their p-values being $2.2 * 10^{-16}$ for the t-test and $2 * 10^{-16}$ for one-way ANOVA. This is expected since ANOVA is a generalization of the t-test used for multiple conditions.

c) No, it's not necessary to correct for multiple comparisons since only two different groups are being compared (the low and high coherence trials) in the same subject, meaning that only one comparison is being made at all times in the one-way ANOVA test.

```
etaSquared(aov.onlyDots, type = 2, anova = FALSE)
```

```
##                     eta.sq eta.sq.part
## onlyDots$cohVec 0.1306018   0.1306018
```

d) The effect size computed using the lsr package is $\eta^2 = 13.06\%$. According to the benchmark on the article, this corresponds to a medium to large effect size (medium = 0.06, large = 0.14 in the article).

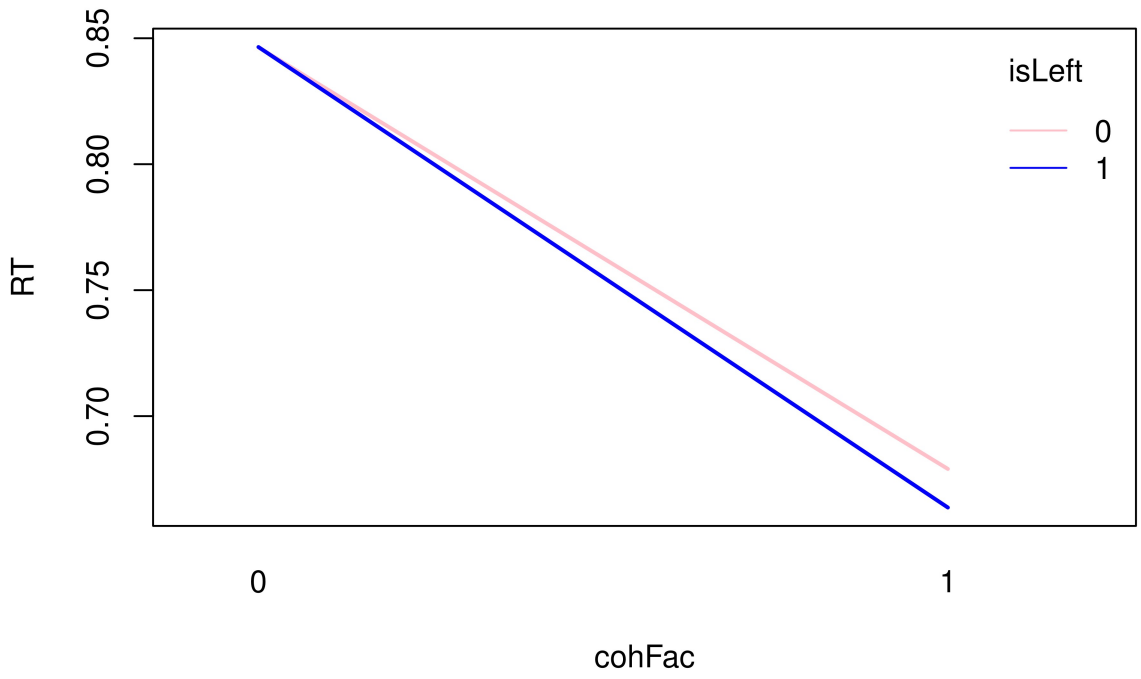
4 Two-way ANOVA

```r
# getting the data of all participants and only correct responses
onlyDots <- dat %>% filter(isDots == 1)
onlyDots <- onlyDots[onlyDots$ER != 1,]

aov.onlyDots <- aov(onlyDots$RT ~ onlyDots$isLeft * onlyDots$cohFac, data = onlyDots)
summary(aov.onlyDots)
```

```
##                                    Df Sum Sq Mean Sq  F value Pr(>F)
## onlyDots$isLeft                     1      0     0.2    0.514  0.473
## onlyDots$cohFac                     1    402   401.6 1369.217 <2e-16 ***
## onlyDots$isLeft:onlyDots$cohFac     1      0     0.2    0.774  0.379
## Residuals                       23155   6792     0.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
interaction.plot(x.factor = onlyDots$cohFac,#x-axis variable
                 trace.factor = onlyDots$isLeft, #variable for lines
                 response = onlyDots$RT, #y-axis variable
                 fun = median, #metric to plot
                 ylab = "RT",
                 xlab = "cohFac",
                 col = c("pink", "blue"),
                 lty = 1, #line type
                 lwd = 2, #line width
                 trace.label = "isLeft")
```



A two-way ANOVA was performed to analyze the effect of whether the dots are moving to the left or right and a factor that represents the proportion of dots that move into a single direction, which can be interpreted as a measure on difficulty, on response time.

A two-way ANOVA revealed that there was not a statistically significant interaction between the effects of the direction of the dots and the level of difficulty ($F(1, 23155) = 0.774$, $p = 0.379$). This result can be observed on the plot since the lines do not seem to have any meaningful difference between each other, with only minimal distance and difference in slope between them, which can be attributed to noise.

Simple main effects analysis showed that the direction did not have a statistically significant effect on Response time ($p = 0.473$). While for level off difficulty, however, it did have a statistically significant effect on response time ($p < 0.001$).

5 Repeated-measures ANOVA

a) Since we're analysing the data between different subjects, or in other words, different groups, we're performing multiple comparisons in the same variable. In this case, the two-way ANOVA is not correcting for this multiple comparisons, being necessary an external method to do so. Repeated measures ANOVA, on the other hand, takes into account this characteristic.

b)

c)

6 What test for what data?

a) Difference of yield between two fields in a farm
b) We want to know whether or not three different exam prep programs lead to different mean scores on a college entrance exam
c) We want to to know whether walruses are heavier in early or late mating season and if that depends on the sex of the walrus
d) See if participants' weight loss differs between a weight loss therapy program only, the program plus a walking regimen, or the program plus a biking regimen. # Division of labor

Micky: 1, 2, 3a

Yann: 3, 4, 5a, 6