Data Analytics & Communication Assignment 5

Micky Labreche s4021290 & Pavel Katsitadze s3240932

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)

# 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] 0.2982633
```

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

```
## [1] 0.5497926
```

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

```
##
##  Welch Two Sample t-test
##
## data:  cohFacLow$RT and cohFacHigh$RT
## t = 12.302, df = 869.29, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.2656177 0.3664616
## sample estimates:
## mean of x mean of y
## 1.2210231 0.9049834
```

(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 difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1446318 0.2997043
## sample estimates:
## mean of the differences
##                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)
```

```
## Warning: package 'lsr' was built under R version 4.1.2
```

```
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
```

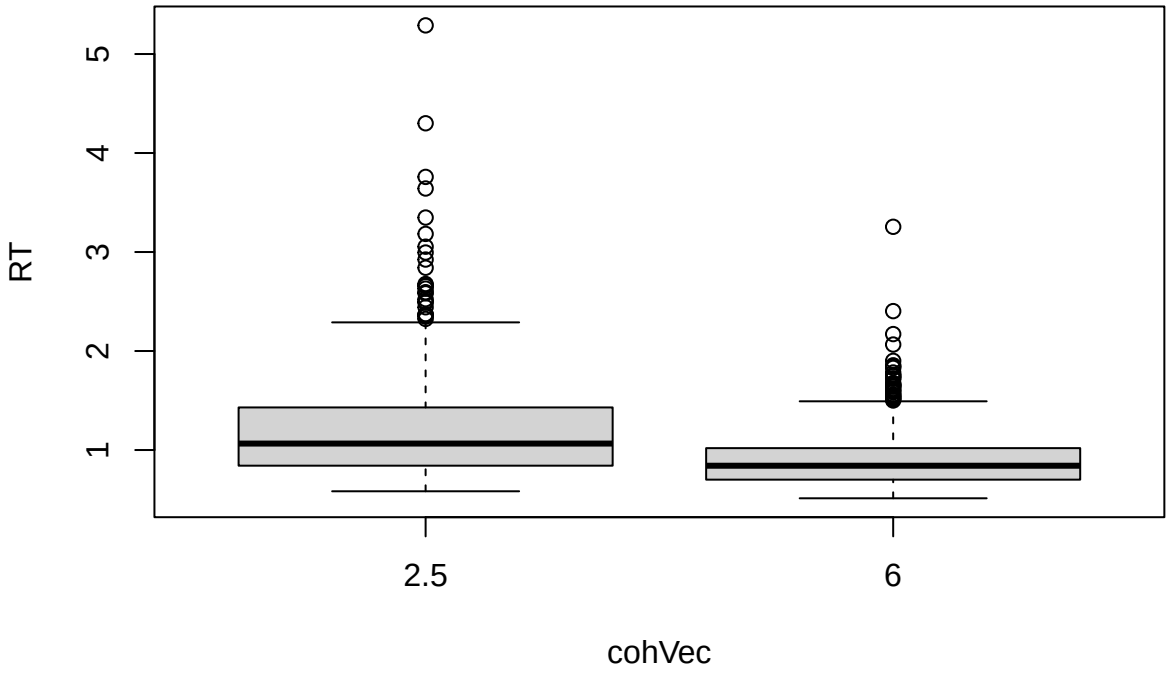
(b) We find that $H(1)=13.7$, $p < .05$, which concludes a significant difference in the response times between a high coherence of dots and a low coherence of dots.

3 One-way ANOVA

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

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## cohVec         1  30.47   30.48   160.5 <2e-16 ***
## Residuals   1222 232.00    0.19
## ---
## 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 ($F(1, 1222) = 160.5$, $p < 0.001$). 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.

4 Division of labor

This is a short one, because my teammate was sick and I didn't have the time to take over his exercises.

Micky: 1, 2 & 3a

Pavel: - (sick)