COSC6323 - Exercise 4

Sachin Shubham

2/27/2021

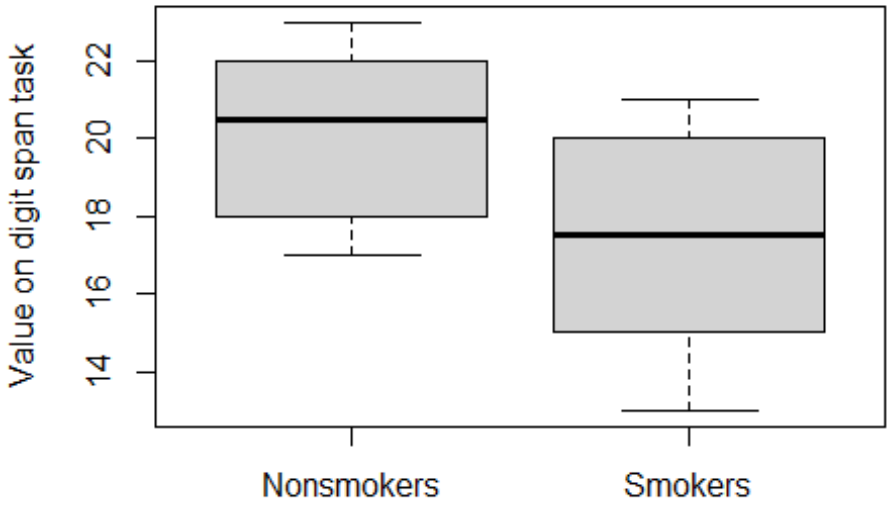
Task 1: As part of his senior research project in the Fall semester of 2001, Scott Keats looked for a possible relationship between marijuana smoking and a deficit in performance on a task measuring short term memory – the digit span task from the Wechsler Adult Intelligence Scale. Two groups of ten subjects were tested. One group, the ”nonsmokers,” claimed not to smoke marijuana. A second group, the ”smokers,” claimed to smoke marijuana regularly. Is there a significant difference between the smokers and non-smokers? Content of the data file: nonSmokers:(18,22,21,17,20,17,23,20,22,21) smokers:(16,20,14,21,20,18,13,15,17,21)

```
library(dplyr)

nonSmokers<-c(18,22,21,17,20,17,23,20,22,21)
smokers<-c(16,20,14,21,20,18,13,15,17,21)

my_data <- data.frame(
  group = rep(c("nonSmokers", "smokers"), each = 10),
  quantity = c(nonSmokers,  smokers)
)

# Summary statistics
group_by(my_data, group) %>%
  summarise(
    count = n(),
    mean = mean(quantity, na.rm = TRUE),
    sd = sd(quantity, na.rm = TRUE)
  )
```

```
## # A tibble: 2 x 4
##   group      count  mean    sd
## * <chr>      <int> <dbl> <dbl>
## 1 nonSmokers    10  20.1  2.13
## 2 smokers       10  17.5  2.95
```

Our study finds that mean of Non-smokers and Smokers are 20.1 and 17.5 respectively.Standard Deviation of Smokers and Non-smokers are 2.95 and 1.13 respectively.

```
boxplot(nonSmokers, smokers, ylab="Value on digit span
task",names=c("Nonsmokers","Smokers"))
```




```
d <- with(my_data,
              quantity[group == "nonSmokers"] -
                quantity[group == "smokers"])
shapiro.test(d)

##
##  Shapiro-Wilk normality test
##
## data:  d
## W = 0.97512, p-value = 0.9339
```

RESULTS: From the output, the p-value 0.9339 is greater than the significance level 0.05 implying that the distribution of the differences (d) are not significantly different from normal distribution. In other words, we can assume the normality.

```
t.test(nonSmokers, smokers)

##
##  Welch Two Sample t-test
##
## data:  nonSmokers and smokers
## t = 2.2573, df = 16.376, p-value = 0.03798
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1628205 5.0371795
## sample estimates:
```

```
## mean of x mean of y
##      20.1      17.5
```

Our study finds that mean sample data of non-smoking group is 20.1 compared to the mean 17.5 in smoking group (t-statistic 2.2573, p=0.03798, α=0.05, 95% CI [0.1628205 5.0371795]). Since α>p-value, a small p(<=0.05), reject the null hypothesis. Therefore, this is strong evidence that the null hypothesis is invalid

```
t.test(nonSmokers, smokers, paired = TRUE)

##
##  Paired t-test
##
## data:  nonSmokers and smokers
## t = 1.9723, df = 9, p-value = 0.08004
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3820841  5.5820841
## sample estimates:
## mean of the differences
##                     2.6
```

Our study finds that the mean of the differences between non-smoking group and smoking group is 2.6 (t-statistic 1.9723, df=9, p=0.08004, 95% CI [-0.3820841 5.5820841]). A big p(<=0.05), fail to reject the null hypothesis. Therefore, this is strong evidence that the null hypothesis is valid The p-value is larger than 0.05, we cannot conclude that is there a significant difference between the smokers and non-smokers exists