Exercise4

Mohammad Imtiaz Nur

2/26/2021

ID: 1878074

```
# Intelligence score data
nonsmokers <- c(18,22,21,17,20,17,23,20,22,21)
smokers <- c(16,20,14,21,20,18,13,15,17,21)


# Creating a data frame
my_data <- data.frame(
    group = rep(c("nonsmokers", "smokers"), each = 10),
    score = c(nonsmokers,  smokers)
)
```

1. Summary Statistics

```
# Summary statistics (using dplyr)
library(dplyr)
```

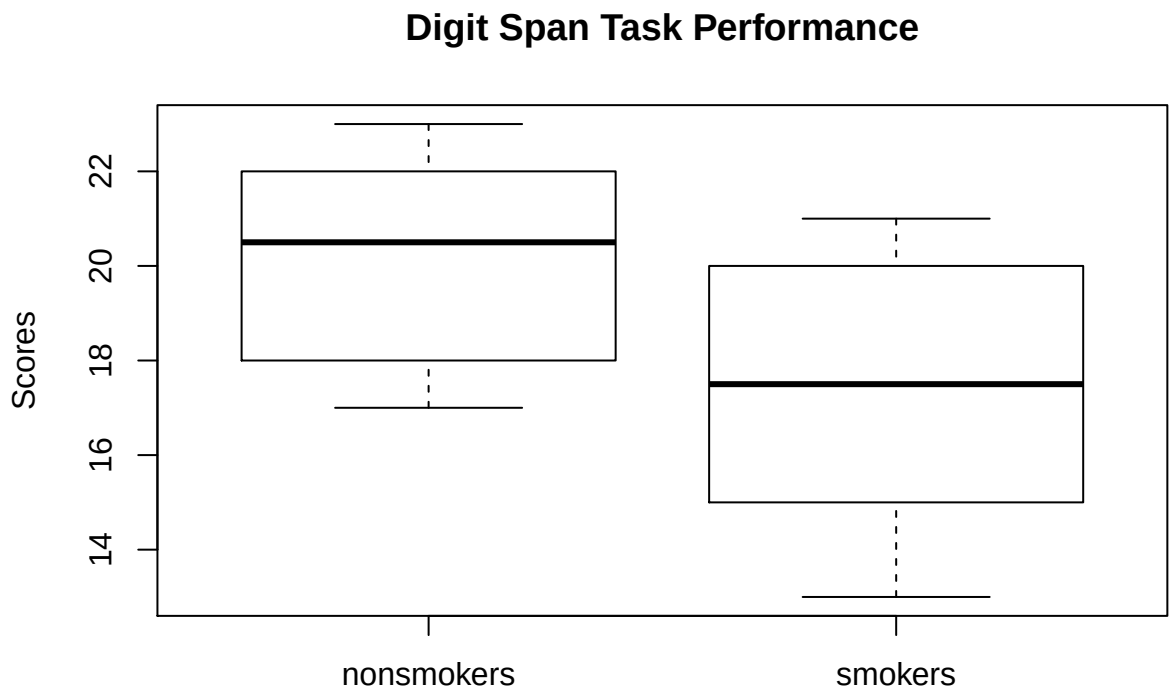
```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
group_by(my_data, group) %>%
  summarise(count = n(),
            mean = mean(score, na.rm = TRUE),
            sd = sd(score, na.rm = TRUE)
            )
```

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

2. Boxplots

```
boxplot(nonsmokers, smokers, ylab="Scores", names=c("nonsmokers","smokers"),
        main="Digit Span Task Performance")
```



3. Shapiro-Wilk test

```
# Use Shapiro-Wilk
d <- with(my_data,
          score[group == "nonsmokers"] -
              score[group == "smokers"])

# Shapiro-Wilk normality test for the differences
shapiro.test(d)
```

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

Here, Null hypothesis: the data are normally distributed Alternative hypothesis: data not normally distributed

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.

4. t-test

```
res <- t.test(nonsmokers, smokers, paired=TRUE)
res
```

```
##
##  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
```

The p-value of the test is 0.08004, which is greater than the significance level alpha = 0.05. We can then accept null hypothesis and conclude that the average score of the nonsmokers isn't significantly different from the average score of the smokers with a p-value = 0.08004.

5. Paired t-test

```
res <- t.test(score ~ group, data = my_data, paired = TRUE)
res
```

```
##
##  Paired t-test
##
## data:  score by group
## 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
```

The p-value of the test is 0.08004, which is greater than the significance level alpha = 0.05. We can then accept null hypothesis and conclude that the average score of the nonsmokers isn't significantly different from the average score of the smokers with a p-value = 0.08004.