Data Assignment 5

Yash Patel

15 April 2022

Question 1

The command "rnorm" returns random numbers from a normal distribution.

Create a vector of 10 numbers using the mean of 100 and a standard deviation of 1. (See the "cookbook" for details about "rnorm()". Find the mean and standard deviation of the vector. (Cut and paste your work.)

```
q1a <- rnorm(n = 10, mean = 100, sd = 1)
mean(q1a)
```

```
## [1] 99.75666
```

```
sd(q1a)
```

```
## [1] 0.9314004
```

Repeat the process from above, but change the standard deviation from 1 to 10. Find the mean and standard deviation of the new vector. Describe the change. (5 pts.) (What happened?) What can you say about the scores generated using the larger standard deviation? (5 pts.) Cut and paste the output here.

```
q1b <- rnorm(n = 10, mean = 100, sd = 10)
mean(q1b)
```

```
## [1] 100.4277
```

```
sd(q1b)
```

```
## [1] 11.04854
```

By increasing the standard deviation, the range and variance of the data set increased, meaning that the data points are farther out from the expected mean. This affects the rnorm's mean as well, since rnorm, after being given a target mean, varies the data based on the standard deviation given, so the final calculated mean is most likely not going to be be the given mean, but rather a value close to it. By increasing the standard deviation, this difference from the given mean is greater.

Repeat the process from above (find the mean and standard deviation), but now

change the number of output elements from 10 to 100. Describe the new data. (How is it different from (b) (5 pts.) What does that say about sample size? (5 pts.)

```
q1c <- rnorm(n = 100, mean = 100, sd = 10)
mean(q1c)
```

```
## [1] 101.0234
```

```
sd(q1c)
```

```
## [1] 9.604456
```

The difference between the given mean and the resulting mean is lesser than the difference from the previous data set. By increasing the sample size, the effect of variation from individual data points is lessened, so increases in standard deviation have less of an effect on the resulting mean.

Question 2

A researcher wants to know if people on vacation, engage in an "inner dialogue" less than when working. The researcher selects a starts by obtaining a sample of 10 individuals who are about to go on a week's vacation and agree to note (on an app) each time they "hear" themselves mentally talking. Each person in the sample is asked to keep a log for the week. The daily average instances (based on the week) appears below.

Create a vector with the following observations (3 pts.)

```r
q2 <- c(50, 40, 46, 49, 40, 58, 45, 47, 46, 43)
```

Complete a one-sample t-test where the population mean is 50. (See the "cookbook" for details about running a t-test. Cut and paste the output.)

```r
t.test(q2, mu = 50, alternative = "less")
```

```
##
##  One Sample t-test
##
## data:  q2
## t = -2.1583, df = 9, p-value = 0.02962
## alternative hypothesis: true mean is less than 50
## 95 percent confidence interval:
##      -Inf 49.45763
## sample estimates:
## mean of x
##      46.4
```

What is the t-value? (3 pts.)

$t = -2.1583$

What is the p-value? (3 pts.)

$p = 0.02962$

What is your interpretation of the Null Hypothesis Significance Test? (3 pts.)

We reject the null hypothesis since $p < 0.05$.

This is a one-tailed t-test. In which direction (as compared to the mean)? (3pts.)

We are using a left-tailed t-test, meaning that we are testing if the sample mean is less than the population mean.

Could this experiment be converted to a two-tailed t-test? If so, state the hypothesis. (5pts.)

Yes it can. The hypothesis would be that there is a general difference (any difference, not just a "less than" relationship) between people on vacation and people working.

Question 3

From the experiment in (2), the researcher also obtains data from a second sample (of the same size) from individuals during a regular week of work. The daily average instances (based on a week of data) of inner dialogue appear below.

Create a vector with the following observations (3 pts.)

```
q3 <- c(53, 40, 51, 50, 43, 62, 49, 47, 51, 39)
```

Complete a two-sample independent t-test of your first vector against the second. (Cut and paste the output.)

```
t.test(q2, q3, var.equal = TRUE, alternative = "less")
```

```
##
##  Two Sample t-test
##
## data:  q2 and q3
## t = -0.77378, df = 18, p-value = 0.2246
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 2.606172
## sample estimates:
## mean of x mean of y
##      46.4      48.5
```

What is the t-value? (3 pts.)

t = -0.77378

What is the p-value? (3pts.)

p = 0.2246

What is your interpretation of the Null Hypothesis Significance Test? (3 pts.)

We fail to reject the null hypothesis since $p > 0.05$.

Could this experiment be converted to a two-tailed t-test? If so, state the hypothesis. (5pts.)

Yes it can. The hypothesis would be that there is a general difference (any difference, not just a "less than" relationship) between people on vacation and people working.

Question 4

Redo the t-test above, but instead of a two-sample, independent t-test, compute the t-test as a paired samples t-test. (Assume that the same people are measured during vacation and then again at a later time during a workweek. (Cut and paste the output.)

```
t.test(q2, q3, paired = TRUE, alternative = "less")
```

```
##
##  Paired t-test
##
## data:  q2 and q3
## t = -2.3333, df = 9, p-value = 0.02225
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.4501984
## sample estimates:
## mean of the differences
##                    -2.1
```

What is the t-value? (3 pts.)

t = -2.3333

What is the p-value? (3pts.)

p = 0.02225

What is your interpretation of the Null Hypothesis Significance Test? (3 pts.)

We reject the null hypothesis since $p < 0.05$.

Compared to the two-sample, independent t-test, what changed? (3 pts.) What explains the change? (7 pts.)

The t-value, p-value, and df changed. This change can be explained by the fact that instead of the data coming from two seperate groups, the data is from the same group at two different time points. This changes the comparison from a "between group" comparison to a "within group" comparison and reduces variability.