Lab 5 Key

Anwesha Guha & Janette Avelar

2/11/2022

First, I imported all relevant libraries are imported. Your process may look different.

library(rio)  
library(here)

## here() starts at /Users/aguha/Documents/r\_projects/educ614/lab5/lab5

library(lsr)  
library(psych)

### Part 1. Understanding the data structure for an independent samples t-test

The time variable contains all the times of death, for both control and experiment groups. The first 11 numbers belong to the control group, while the last 8 numbers belong to the experimental group. To specify which time of death is in which group, we need to create a second variable that labels each time as either “control” or “experiment.”

time <- c(70, 77, 83, 87, 92, 93, 100, 102, 102, 103, 96,  
 34, 36, 48, 48, 65, 91, 98, 102)  
category <- c(rep("control", 11),  
 rep("experiment", 8))

We then join both “time” and “category” to create a data frame with two columns.

death\_df <- data.frame(time, category)

In R, we have to do an additional step of turning the category variable into a factor, so we can use it for analysis.

death\_df$category <- as.factor(death\_df$category)

**QUESTION 1: Now you’re ready to run an independent samples t-test. Copy/paste the R output below.**

Assuming equal variances:

independentSamplesTTest(formula = time ~ category,   
 data = death\_df,   
 var.equal = TRUE,  
 one.sided = "control")

##   
## Student's independent samples t-test   
##   
## Outcome variable: time   
## Grouping variable: category   
##   
## Descriptive statistics:   
## control experiment  
## mean 91.364 65.250  
## std dev. 11.012 28.065  
##   
## Hypotheses:   
## null: population means are equal, or smaller for group 'control'   
## alternative: population mean is larger for group 'control'   
##   
## Test results:   
## t-statistic: 2.825   
## degrees of freedom: 17   
## p-value: 0.006   
##   
## Other information:   
## one-sided 95% confidence interval: [10.035, Inf]   
## estimated effect size (Cohen's d): 1.313

Not assuming equal variances:

independentSamplesTTest(formula = time ~ category,   
 data = death\_df,   
 var.equal = FALSE,  
 one.sided = "control")

##   
## Welch's independent samples t-test   
##   
## Outcome variable: time   
## Grouping variable: category   
##   
## Descriptive statistics:   
## control experiment  
## mean 91.364 65.250  
## std dev. 11.012 28.065  
##   
## Hypotheses:   
## null: population means are equal, or smaller for group 'control'   
## alternative: population mean is larger for group 'control'   
##   
## Test results:   
## t-statistic: 2.496   
## degrees of freedom: 8.58   
## p-value: 0.018   
##   
## Other information:   
## one-sided 95% confidence interval: [6.826, Inf]   
## estimated effect size (Cohen's d): 1.225

### Part 2. Understanding the data structure for a dependent samples t-test

We’ll again create a dataset in R by using the following code.

silent <- c(6, 7, 7, 5, 4, 4, 6, 9, 5, 7)  
noisy <- c(9, 9, 6, 7, 6, 7, 9, 11, 7, 9)  
mistakes\_df <- data.frame(silent, noisy)

**QUESTION 2: Using the oneSampleTTest function, run a dependent samples t-test and paste a screenshot of your output below.**

This time, we compute the difference scores in a separate column.

mistakes\_df$difference <- mistakes\_df$silent - mistakes\_df$noisy

Conduct one-sample t-test of differnces.

oneSampleTTest(x = mistakes\_df$difference,  
 mu = 0,  
 one.sided = "less")

##   
## One sample t-test   
##   
## Data variable: mistakes\_df$difference   
##   
## Descriptive statistics:   
## difference  
## mean -2.000  
## std dev. 1.155  
##   
## Hypotheses:   
## null: population mean greater than or equal to 0   
## alternative: population mean less than 0   
##   
## Test results:   
## t-statistic: -5.477   
## degrees of freedom: 9   
## p-value: <.001   
##   
## Other information:   
## one-sided 95% confidence interval: [-Inf, -1.331]   
## estimated effect size (Cohen's d): 1.732

Alternatively, if you calculated mistakes\_df$difference as noisy minus silent (mistakes\_df$noisy - mistakes\_df$silent), you would do one.sided = "greater" in your oneSampleTTest.

### Part 3. Conducting an independent samples t-test and writing up results

**a) State the null and alternative hypotheses in symbols.**

*where* *average reading score for students who are first-time kindergartners, and*  *average reading score for students who are not first-time kindergarteners*

**b) Run an independent samples t-test (two-sided) to compare the average reading score for those kids who are first-time kindergartners and those who are not. Copy/Paste the software output below.**

Import data, whichever way is most comfortable to you.

ecls <- import(here("data", "ecls-k-sub.csv"))

Next, remember to filter data, removing values that don’t make sense, like -9. Also, remember to convert the variable to a factor for easier interpretability.

ecls$X1FIRKDG[ecls$X1FIRKDG==-9] <- NA  
ecls$X1RSCALK5[ecls$X1RSCALK5==-9] <- NA  
ecls$X1FIRKDG <- as.factor(ecls$X1FIRKDG)

Check out data descriptives.

describeBy(ecls$X1RSCALK5, group = ecls$X1FIRKDG)

##   
## Descriptive statistics by group   
## group: 1  
## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 1460 54.1 11.54 52.82 52.74 8.67 33.7 116.78 83.09 1.87 5.88 0.3  
## ------------------------------------------------------------   
## group: 2  
## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 75 54.53 10.93 53.76 53.8 11.25 34.87 90.4 55.54 0.69 0.55 1.26

Then, conduct t-test. We are doing a two sided test since we are checking mean differences.

independentSamplesTTest(formula = X1RSCALK5 ~ X1FIRKDG,   
 data = ecls,   
 var.equal = FALSE,  
 one.sided = FALSE)

## Warning in independentSamplesTTest(formula = X1RSCALK5 ~ X1FIRKDG, data =  
## ecls, : 284 case(s) removed due to missingness

##   
## Welch's independent samples t-test   
##   
## Outcome variable: X1RSCALK5   
## Grouping variable: X1FIRKDG   
##   
## Descriptive statistics:   
## 1 2  
## mean 54.103 54.533  
## std dev. 11.542 10.935  
##   
## Hypotheses:   
## null: population means equal for both groups  
## alternative: different population means in each group  
##   
## Test results:   
## t-statistic: -0.331   
## degrees of freedom: 82.699   
## p-value: 0.741   
##   
## Other information:   
## two-sided 95% confidence interval: [-3.012, 2.152]   
## estimated effect size (Cohen's d): 0.038

**c) Provide a summary write-up for your results from this analysis.**

The average reading score is 54.1 with a standard deviation of 11.54 for 1460 who are first-time kindergarteners and 54.53 with a standard deviation of 10.93 for the 75 who are not. An independent samples t-test was conducted to test whether or not the average reading score for first-time and not-first-time kindergarteners significantly differed. The findings indicate that the difference in average scores between two groups was not significantly significant (95% confidence interval [-3.01, 2.15]), t(83) = -0.33, p = 0.75, Cohen’s d = 0.04).

### Part 4. Conducting a dependent samples t-test and writing up results

**a) State the null and alternative hypotheses in symbols.**

*where* *average focus score for students in the fall, and* *average focus score for students in the spring*

**b) Run a dependent samples t-test to compare the average rating for attentional focus in the fall of kindergarten and the average rating for attentional focus in the spring of kindergarten. Copy/Paste the software output below.**

First, remember to clean your data.

ecls$X1ATTNFS[ecls$X1ATTNFS==-9] <- NA  
ecls$X1ATTNFS[ecls$X1ATTNFS==-1] <- NA  
ecls$X2ATTNFS[ecls$X2ATTNFS==-9] <- NA  
ecls$X2ATTNFS[ecls$X2ATTNFS==-1] <- NA

Check out descriptives.

describe(ecls$X1ATTNFS)

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 1444 4.7 1.33 4.83 4.78 1.48 1 7 6 -0.47 -0.43 0.03

describe(ecls$X2ATTNFS)

## vars n mean sd median trimmed mad min max range skew kurtosis se  
## X1 1 1598 4.93 1.32 5.17 5.02 1.43 1 7 6 -0.55 -0.37 0.03

Create a third variable calculating the difference in scores.

ecls$growth <- ecls$X2ATTNFS - ecls$X1ATTNFS

We are testing whether attention grew, so we want to test whether growth is greater than zero.

oneSampleTTest(x = ecls$growth,  
 mu = 0,  
 one.sided = "greater")

## Warning in oneSampleTTest(x = ecls$growth, mu = 0, one.sided = "greater"): 461  
## case(s) removed due to missingness

##   
## One sample t-test   
##   
## Data variable: ecls$growth   
##   
## Descriptive statistics:   
## growth  
## mean 0.223  
## std dev. 1.056  
##   
## Hypotheses:   
## null: population mean less than or equal to 0   
## alternative: population mean greater than 0   
##   
## Test results:   
## t-statistic: 7.79   
## degrees of freedom: 1357   
## p-value: <.001   
##   
## Other information:   
## one-sided 95% confidence interval: [0.176, Inf]   
## estimated effect size (Cohen's d): 0.211

**c) Provide a summary write-up for your results from this analysis.**

The average rating for attentional focus in the fall of kindergarten was 4.70 with a standard deviation of 1.33, and the average for the spring was 4.93 with a standard deviation of 1.32. A dependent samples t-test was conducted to test whether or not the average focus score increase over the school year was significant. The findings indicated that the increase of 0.223 was statistically significant, t(1357) = -7.79, p < 0.001, Cohen’s d = 0.211.