PS 5 Part II

Olivia Bogiages

2025-12-11

Conduct a hypothesis test for a difference in means. You decide what the hypotheses are, whether you use a t-test or a z-test, and what the level of significance is. Explain your decisions, and interpret your results both substantively and statistically.

I chose to use a t-test to test my hypotheses because it is more conservative. I set the level of significance at 95%, and I am testing these hypotheses: H1:Small class size will result in higher math scores H0: Small class size will have no affect on math scores

The t-statistic is above 2 and the p-value is below .05, thus these findings are statistically significant. There is a statistically significant affect of small class size on math scores.

```r
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
```

```r
#read in the data
star<-read.csv("https://raw.githubusercontent.com/MLBurnham/pols_602/refs/heads/main/data/STAR.csv")
star$treatment<-ifelse(star$class=="small", 1, 0)

#creating new datasets to hold treatment and control groups
star_treatment <-star[star$treatment==1,]
star_control<-star[star$treatment==0,]
#running difference in means test using welch two sample t-test
two_tailed <- t.test(star_treatment$math,
                     star_control$math,
                     mu = 0,
                     alternative = "two.sided",
                     conf.level = 0.95, na.rm=TRUE)
print(two_tailed)
```

```
##
##  Welch Two Sample t-test
##
## data:  star_treatment$math and star_control$math
## t = 2.7404, df = 1219.3, p-value = 0.006227
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##   1.701545 10.278265
## sample estimates:
## mean of x mean of y
##  634.8274  628.8374
```

Using the same data, fit a linear model. Interpret the coefficient, standard error, tvalue, and p-value.

The coefficient is 5.99. This means that for students who were in class sizes, there was a 5.99 point increase in their math test score. The standard error is the margin within which the estimate of the coefficient differs from the true value. In this case, the value of 5.99 for the coefficient is the estimated value and the true value lies within the 2.178 up or down from that value. The p-value signifies that these values are statistically significant. If 100 observations were to be made, 95 of them would be consistent with my model The t-value is the number of stnadard deviations my coefficient is away from a mean of zero. This means that this coefficient is 2.7 standard deviations from a mean of zero, making it significant at a 95% level.

```r
model<-lm(math~treatment, data=star)
summary(model)
```

```
##
## Call:
## lm(formula = math ~ treatment, data = star)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -119.827  -27.585   -0.827   26.163  145.163
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  628.837      1.476  426.09  < 2e-16 ***
## treatment      5.990      2.178    2.75  0.00604 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 38.74 on 1272 degrees of freedom
## Multiple R-squared:  0.005911,   Adjusted R-squared:  0.00513
## F-statistic: 7.564 on 1 and 1272 DF,  p-value: 0.006039
```