Hw7

Jaxon Lara

jtl3236

(Github_R_Code)

1A

The number of male students: 106
The number of female students: 111
The sample proportion of males who folded their left arm on top: 0.47
The sample proportion of females who folded their left arm on top: 0.42

1B

The observed difference in proportions between the two groups (males - females): 0.048

1C

The 95% confidence interval for the difference in proportions (males - females). (R's built in function)
Lower bound: -0.087, Upper bound: 0.179

standard error for the difference in proportions:
sqrt(((samplePropGroup1(1-samplePropGroup1)) / sampleSize1) + ((samplePropGroup2(1-samplePropGroup2)) / sampleSize2)
or
sqrt(((propMalesLeft(1-propMalesLeft)) / numMales) + ((propFemalesLeft(1-propFemalesLeft)) / numFemales))

standard error for difference in proportions: 0.067

sample mean for the difference in proportions: 0.048

z, standard deviations: 1.96
because i wanted to create a 95% confidence interval I would have to go out 1.96 standard deviations to either side to contain 95% of sample difference in proportions.

upperbound = (sample mean) + (1.96) * (standard error)
lowerbound = (sample mean) - (1.96) * (standard error)
upper bound: 0.18
lower bound: -0.085

1D

If we were to take many many samples of students who were asked to fold their arms across their chest and calculate their corresponding confidence intervals. Then we would expect that 95% of those confidence intervals would contain the true difference in proportions.

1E

The standard error for the difference in proportions represents how varied a sample mean for the difference in proportions could be from the true population difference in proportions.

1F

The term sampling distribution refers to the distribution of bootstrapped samples created from resampling the proportion of Males/Females who folded their left arm on top. Specifically, isolating the cases whose sex was male (fixed number of cases corresponding to sex) and resampling whether or not they folded their left arm on top with replacement (varies from sample to sample). Then doing the same for cases whose sex was female and subtracting the female proportion from the male proportion to get that sample's difference in proportions. Then repeat.

1G

The central limit theorem, because the sampling distribution is reasonably assumed to be made up of many many samples. The central limit theorem says that the distribution of a sample mean or in this case proportions will become approximately normal with a reasonably large sample size.

1H

I would say that their assessment is correct because the 95% confidence interval includes 0 and so there is not enough evidence to suggest there is a difference in arm folding based on sex.

1I

Yes the confidence intervals will be different across the samples but the true difference in proportions of folding left arms on top for males and females should be contained within 95% of those confidence intervals.

2A

The proportion of those receiving a GOTV call who voted in 1998: 0.65
The sample proportion of those not reciving a GOTV call who voted in 1998: 0.44

Large-sample 95% confidence interval for the difference in these two proportions:
(prop of those who received a call) - (prop of those who didn't receive a call)
Lower bound: 0.144, Upper bound: 0.263

2B

The correlation between voted1996 and voted1998 using a large-sample confidence interval
is estimated to be the following.
Lower bound: 0.395, Upper bound: 0.428
The correlation between voted1996 and GOTV_call call using a large-sample confidence interval
is estimated to be the following.
Lower bound: 0.036, Upper bound: 0.071
voted1996 is highly correlated to both the treatment(GOTV_call) and the outcome(voted1998)
because 0 is included in neither interval which is evidence that voted1996 is a confounder.

The correlation between AGE and voted1998 using a large-sample confidence interval
is estimated to be the following.
Lower bound: 0.261, Upper bound: 0.296
The correlation between AGE and GOTV_call using a large-sample confidence interval
is estimated to be the following.
Lower bound: 0.05, Upper bound: 0.091
AGE is highly correlated to both the treatment(GOTV_call) and the outcome(voted1998)
because 0 is included in neither interval which is evidence that AGE is a confounder.

The correlation between MAJORPTY and voted1998 using a large-sample confidence interval
is estimated to be the following.
Lower bound: 0.097, Upper bound: 0.134
The correlation between MAJORPTY and GOTV_call using a large-sample confidence interval
is estimated to be the following.
Lower bound: 0.002, Upper bound: 0.036
MAJORPTY is highly correlated to both the treatment(GOTV_call) and the outcome(voted1998) because
0 is included in neither interval which is evidence that MAJORPTY is a confounder.

2C

The correlation between voted1996 and GOTV_call call using a large-sample confidence interval
is estimated to be the following.
Lower bound: -0.053, Upper bound: 0.049
voted1996 is no longer highly correlated to the treatment(GOTV_call) because 0 is included
in the confidence interval which is evidence that voted1996 is no longer a confounder.

The correlation between AGE and GOTV_call using a large-sample confidence interval
is estimated to be the following.
Lower bound: -0.049, Upper bound: 0.052
AGE is no longer highly correlated to the treatment(GOTV_call) because 0 is included
in the confidence interval which is evidence that AGE is no longer a confounder.

The correlation between MAJORPTY and GOTV_call using a large-sample confidence interval
is estimated to be the following.
Lower bound: -0.058, Upper bound: 0.045
MAJORPTY is no longer highly correlated to the treatment(GOTV_call) because 0 is included
in the confidence interval which is evidence that MAJORPTY is no longer a confounder.

Repeating the analysis from part A with the matched data:
The proportion of those receiving a GOTV call who voted in 1998: 0.65
The sample proportion of those not reciving a GOTV call who voted in 1998: 0.57

Large-sample 95% confidence interval for the difference in these two proportions:
(prop of those who received a call) - (prop of those who didn't receive a call)
Lower bound: 0.015, Upper bound: 0.142

To conclude: There is enough evidence to suggest the GOTV call increases the likelihood of voting in the 1998 election because the interval is positive and 0 is not included in the confidence interval.