Stat Inference Lab Answers

Aaron Gullickson

8/9/2019

For students born in January, February, March, or April:

Conduct a hypothesis test on the politics data, where the null hypothesis is that the same proportion of Democrats and Republicans in the US population have no religious affiliation.

```r
props <- prop.table(table(politics$party,
                          politics$relig),1)[c("Democrat","Republican"),
                                             "Non-religious"]
props
```

```
##   Democrat Republican
## 0.18361775 0.04834811
```

```r
pdiff <- diff(props)
pdiff
```

```
## Republican
## -0.1352696
```

```r
n <- table(politics$party)[c("Democrat","Republican")]
n
```

```
##
##   Democrat Republican
##       1465       1241
```

```r
se <- sqrt(sum(props*(1-props)/n))
se
```

```
## [1] 0.01180668
```

```r
tstat <- pdiff/se
tstat
```

```
## Republican
##  -11.45704
```

```r
2*pt(-abs(tstat), min(n)-1)
```

```
##   Republican
## 5.858835e-29
```

Assuming the same proportion of Democrats and Republican have no religious affiliation in the population, there is less than a 0.00000001% chance of observing a sample proportion difference of 0.1353 or greater in a sample of this size. Therefore, I **reject** the null hypothesis and conclude that a higher proportion of Democracts are non-religious than Republicans.

Conduct a hypothesis test on the sexual frequency data, where the null hypothesis is that individuals with no religious affiliation in the US population have the same sexual frequency as mainline Protestants.

```
means <- tapply(sex$sexf, sex$relig, mean)[c("None","Mainline Protestant")]
means
```

```
##                 None Mainline Protestant
##             59.84862            44.15595
```

```
mdiff <- diff(means)
mdiff
```

```
## Mainline Protestant
##           -15.69267
```

```
sds <- tapply(sex$sexf, sex$relig, sd)[c("None","Mainline Protestant")]
sds
```

```
##                 None Mainline Protestant
##             57.94731            49.58294
```

```
n <- table(sex$relig)[c("None","Mainline Protestant")]
n
```

```
##
##                 None Mainline Protestant
##                  327                 529
```

```
se <- sqrt(sum(sds^2/n))
se
```

```
## [1] 3.862145
```

```
tstat <- mdiff/se
tstat
```

```
## Mainline Protestant
##             -4.0632
```

```
2*pt(-abs(tstat), min(n)-1)
```

```
## Mainline Protestant
##        6.070374e-05
```

Assuming that Mainline Protestants and those with no religious affiliation have the same sexual frequency in the population, there is less than a 0.007% chance of observing a sample mean difference of 15.693 in a sample of this size. Therefore, I **reject** the null hypothesis and conclude that those with no religious affilation have more sex than Mainline Protestants, on average.

Conduct a hypothesis test on the sexual frequency data, where the null hypothesis is that there is no association between years of education and sexual frequency.

```
r <- cor(sex$educ, sex$sexf)
r
```

```
## [1] 0.001409977
```

```
n <- nrow(sex)
n
```

```
## [1] 2103
```

```
se <- sqrt((1-r^2)/(n-2))
se
```

```
## [1] 0.02181657
```

```
tstat <- r/se
tstat
```

```
## [1] 0.0646287
```

```
2*pt(-abs(tstat), n-2)
```

```
## [1] 0.9484758
```

Assuming there is no relationship between education and sexual frequency in the population, there is a 95% chance of observing a sample correlation coefficient of 0.001 in absolute magnitude just by random chance in a sample this size. I therefore **fail to reject** the null hypothesis that there is no relationship between education and sexual frequency.

For students born in May, June, July, or August:

Conduct a hypothesis test on the earnings data, where the null hypothesis is that STEM workers and managers have the same hourly wage, on average, in the US population.

```
means <- tapply(earnings$wages, earnings$occup, mean)[c("STEM","Manager")]
means
```

```
##     STEM  Manager
## 37.23067 34.67611
```

```
mdiff <- diff(means)
mdiff
```

```
##   Manager
## -2.554559
```

```
sds <- tapply(earnings$wages, earnings$occup, sd)[c("STEM","Manager")]
sds
```

```
##     STEM  Manager
## 18.33703 18.81986
```

```
n <- table(earnings$occup)[c("STEM","Manager")]
n
```

```
##
##    STEM Manager
##    9858   11642
```

```
se <- sqrt(sum(sds^2/n))
se
```

```
## [1] 0.254032
```

```
tstat <- mdiff/se
tstat
```

```
##   Manager
## -10.05605
```

```
2*pt(-abs(tstat), min(n)-1)
```

```
##      Manager
## 1.123442e-23
```

Assuming that STEM workers and managers have the same average hourly wage in the population , there is less than a 0.00000001% chance of observing a sample mean difference of 2.555 in a sample of this size. Therefore, I **reject** the null hypothesis and conclude that those STEM workers make more than managers, on average.

Conduct a hypothesis test on the popularity data, where the null hypothesis is that the number of sports played in school is not associated with the number of friend nominations received in school among US adolescents.

```
r <- cor(popularity$nsports, popularity$indegree)
r
```

```
## [1] 0.1749571
```

```
n <- nrow(popularity)
n
```

```
## [1] 4397
```

```
se <- sqrt((1-r^2)/(n-2))
se
```

```
## [1] 0.01485148
```

```
tstat <- r/se
tstat
```

```
## [1] 11.78044
```

```
2*pt(-abs(tstat), n-2)
```

```
## [1] 1.462129e-31
```

Assuming there is no relationship between number of sports played and the number of friend nominations recieved in the population of adolescents, there is less than a 0.00000001% chance of observing a sample correlation coefficient of 0.175 in absolute magnitude just by random chance in a sample this size. I therefore **reject** the null hypothesis and conclude that the number of sports played is positively associated with the number of friend nominations recieved.

Conduct a hypothesis test on the popularity data, where the null hypothesis is that the same proportion of men and women drink frequently.

```
props <- prop.table(table(popularity$sex,
                          popularity$alcoholuse),1)[,"Drinker"]
props
```

```
##    Female      Male
## 0.1439098 0.1866029
```

```
pdiff <- diff(props)
pdiff
```

```
##       Male
## 0.04269303
```

```
n <- table(popularity$sex)
n
```

```
##
##  Female    Male
##    2307    2090
```

```
se <- sqrt(sum(props*(1-props)/n))
se
```

```
## [1] 0.01122612
```

```
tstat <- pdiff/se
tstat
```

```
##    Male
## 3.80301
```

```
2*pt(-abs(tstat), min(n)-1)
```

```
##         Male
## 0.0001470506
```

Assuming the same proportion of boys and girls drink frequently in the population, there is less than a 0.015% chance of observing a sample proportion difference of 0.0427 or greater in a sample of this size. Therefore, I **reject** the null hypothesis and conclude that a higher proportion of boys drink than girls.

For students born in September, October, November, or December:

Conduct a hypothesis test on the popularity data where the null hypothesis is that there is no association between parental income and the number of friend nominations received in school among US adolescents.

```
r <- cor(popularity$parentinc, popularity$indegree)
r
```

```
## [1] 0.1247392
```

```
n <- nrow(popularity)
n
```

```
## [1] 4397
```

```
se <- sqrt((1-r^2)/(n-2))
se
```

```
## [1] 0.01496633
```

```
tstat <- r/se
tstat
```

```
## [1] 8.334658
```

```
2*pt(-abs(tstat), n-2)
```

```
## [1] 1.027781e-16
```

Assuming there is no relationship between parental income and the number of friend nominations recieved in the population of adolescents, there is less than a 0.00000001% chance of observing a sample correlation coefficient of 0.125 in absolute magnitude just by random chance in a sample this size. I therefore **reject** the null hypothesis and conclude that parental income is positively associated with the number of friend nominations recieved.

Conduct a hypothesis test on the sexual frequency data where the null hypothesis is that mainline Protestants and fundamentalist Protestants have the same sexual frequency, on average, in the US population.

```
means <- tapply(sex$sexf, sex$relig, mean)[c("Fund Protestant","Mainline Protestant")]
means
```

```
##     Fund Protestant Mainline Protestant
##            49.59892            44.15595
```

```
mdiff <- diff(means)
mdiff
```

```
## Mainline Protestant
##           -5.442966
```

```
sds <- tapply(sex$sexf, sex$relig, sd)[c("Fund Protestant","Mainline Protestant")]
sds
```

```
##     Fund Protestant Mainline Protestant
##            52.84669            49.58294
```

```
n <- table(sex$relig)[c("Fund Protestant","Mainline Protestant")]
n
```

```
##
##     Fund Protestant Mainline Protestant
##                 556                 529
```

```
se <- sqrt(sum(sds^2/n))
se
```

```
## [1] 3.10972
```

```
tstat <- mdiff/se
tstat
```

```
## Mainline Protestant
##           -1.750307
```

```
2*pt(-abs(tstat), min(n)-1)
```

```
## Mainline Protestant
##          0.08064611
```

Assuming that Mainline Protestants and Fundamentalist Protestants have the same sexual frequency in the population, there is about an 8% chance of observing a sample mean difference of 5.443 in a sample of this size. Therefore, I **fail to reject** the null hypothesis that Fundamentalist and Mainline Protestants have the same average sexual frequency in the population.

Conduct a null hypothesis test on the politics data where the null hypothesis is that the same proportion of men and women voted for Trump in the US population.

```
props <- prop.table(table(politics$gender,
                          politics$president),1)[c("Male","Female"),"Trump"]
props
```

```
##      Male    Female
## 0.3160544 0.2754464
```

```
pdiff <- diff(props)
pdiff
```

```
##      Female
## -0.04060792
```

```
n <- table(politics$gender)[c("Male","Female")]
n
```

```
##
##   Male Female
##   1987   2240
```

```
se <- sqrt(sum(props*(1-props)/n))
se
```

```
## [1] 0.01406718
```

```
tstat <- pdiff/se
tstat
```

```
##    Female
## -2.886715
```

```
2*pt(-abs(tstat), min(n)-1)
```

```
##     Female
## 0.00393494
```

Assuming the same proportion of men and women voted for Trump in the population, there is less than a 0.4% chance of observing a sample proportion difference of `abs(r round(pdiff, 4))` or greater in a sample of this size. Therefore, I **reject** the null hypothesis and conclude that a higher proportion of men voted for Trump than women.