Public Policy 529
Fall 2023 Problem Set #6

Francisco Brady

2023-10-25

Due Wednesday, October 25th, end of day

1. Suppose your friend claims that 65% of people in Ann Arbor sort their household waste into trash and recyclables. If this is true, there is a probability of .65 that a randomly selected person sorts their household waste. To test your friend's claim, you randomly select 6 people and find that only 2 of them say that they sort their household waste.

(a) With what probability does exactly 2 people say they sort household waste? One person? Zero people? Use the binomial formula to calculate the probabilities:

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \text{ for x = 1,..., n}$$

$$P(X = x) = \frac{6!}{2!(6-2)!} .65^2 (1-.65)^{6-2}, \text{ for x = 2}$$

```r
# 2 wins
# 6 trials
# prob .65
prob_x <- ((6 * 5 * 4 * 3 * 2 * 1) /
             ((2 * 1) * (4 * 3 * 2 * 1))) *
  (.65 ^ 2) * (1 - .65) ^ (4)
prob_x
```

```
## [1] 0.09510211
```

$$P(X = x) = \frac{6!}{1!(6-1)!} .65^1 (1-.65)^{6-1}, \text{ for x = 1}$$

```r
# 1 win
# 6 trials
# prob .65
prob_x <- ((6 * 5 * 4 * 3 * 2 * 1) /
             ((1) * (5 * 4 * 3 * 2 * 1))) *
  (.65 ^ 1) * (1 - .65) ^ (5)
prob_x
```

```
## [1] 0.02048353
```

$$P(X = x) = \frac{6!}{0!(6-0)!}.65^0(1-.65)^{6-0}, \text{ for x = 0}$$

```
# 0 wins
# 6 trials
# prob .65
prob_x <- ((6 * 5 * 4 * 3 * 2 * 1) /
             ((1 # 0! = 1
              ) * (6 * 5 * 4 * 3 * 2 * 1))) *
  (.65 ^ 0) * (1 - .65) ^ (6)
prob_x
```

```
## [1] 0.001838266
```

(b) If the true proportion of people that sort their trash is .65, what is the probability that your "survey" with a random sample of 6 would produce 2 or fewer people who sort household waste? (you can think of this as a one-sided p-value)

$$P(X \leq 2) = P(0) + P(1) + P(2) = 0.001838266 + 0.02048353 + 0.09510211 = 0.1174239$$

2. In a sample of 31 seniors at a local high school, the mean combined SAT score was 990 with a standard deviation of 75. Perform a significance test in which the null hypothesis states the population mean is 960. Use α=.05. Be sure to include all the steps: consider assumptions, state hypotheses, calculate test statistic and its critical value, find the p-value, and state a conclusion. Show your work.

a) assumptions: we assume that since the sample has more than 30 records, the sampling distribution is normally distributed. Since we are testing means, we can use a t-table.
b) hypothesis: $H_0 : \mu = 960$, $H_a = \mu \neq 960$
c) calculate test statistic and critical values: The critical value: $t_{.025}$ with 30 degrees of freedom (n -1) = 2.042.
   The test statistic:
$$t = \frac{\bar{y} - \mu}{\hat{\sigma}_{\bar{y}}}, \text{ where } \hat{\sigma}_{\bar{y}} = \frac{s}{\sqrt{n}}$$
$$\hat{\sigma}_{\bar{y}} = \frac{s}{\sqrt{n}} = \frac{75}{\sqrt{31}} = 13.4704$$
$$t = \frac{990 - 960}{13.4704} = 2.227106$$
d) find the p-value:
   Using the t-table, our p-value is between $t_{.025}$ (2.042) and $t_{.010}$ (2.457). Since we are doing a two-tailed test, we need to double the p-value.
$$2 \cdot (t_{.010} < p < t_{.025})$$
$$= t_{.02} < p < t_{.05}$$
e) conclude: Since the absolute value of the observed test statistic exceeds that of the critical value (at α), we can reject H_0. Similarly, since the p-value is less than .05, we can reject the null hypothesis that the population mean SAT score is 960.

3. Suppose you are running a study of a program intended to prepare students for state math tests. The mean test score in your sample of students is 3 points higher than the value specified by the null hypothesis (in which the program has no effect). The standard deviation of their test scores (s) is 40. How large does your sample size need to be for you to reject the null hypothesis at α=.05? Hint: start with the formula for the z statistic, write it such that the result must be greater than 1.96, then use algebra to solve for n.

$$Z = \frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Dividing by $\frac{\sigma}{\sqrt{n}}$:

$$Z\frac{\sigma}{\sqrt{n}} = \bar{y} - \mu$$

Squaring each term to get rid of the $\sqrt{n}$:

$$Z^2\frac{\sigma^2}{n} = (\bar{y} - \mu)^2$$

Combine and multiply through by n to get it out of the numerator:

$$n\left(\frac{Z^2\sigma^2}{n}\right) = (\bar{y} - \mu)^2 n \rightarrow Z^2\sigma^2 = (\bar{y} - \mu)^2 n$$

Divide by $(\bar{y} - \mu)^2$ to isolate n

$$\frac{Z^2\sigma^2}{(\bar{y} - \mu)^2} = \frac{(\bar{y} - \mu)^2}{(\bar{y} - \mu)^2}n \rightarrow \frac{Z^2\sigma^2}{(\bar{y} - \mu)^2} = n$$

Rearrange:

$$n = \left(\frac{Z\sigma}{\bar{y} - \mu}\right)^2$$

Now we can apply what is given – $Z = 1.96, s = 40,$ and $(\bar{y} - \mu) \leq 3$

$$n = \left(\frac{Z\sigma}{\bar{y} - \mu}\right)^2 \rightarrow \left(\frac{1.96 \cdot 40}{3}\right)^2 \rightarrow 682.951$$

Since this results in the minimum sample we need, we should round up to 683.

```
# Note: this is code for question 4:
anes2020 <- read_dta('anes2020subset.dta')
```

4. For this question, use the dataset `anes2020subset`. The variable `VotedPres` is a dichotomous measure for whether the respondent says they voted in the presidential election (0=no; 1=yes).

```
freq(table(anes2020$VotedPres), cum = T, valid = F)
```

(a) Use your software to obtain the information you need to find $\hat{\pi}$, the proportion of people who say they voted.

```
##      n    %  %cum
## 0 1116 15.8  15.8
## 1 5956 84.2 100.0
```

84.2% of respondents said they voted.

```r
# Note: this is code for question 4b:
prop.test(5956, 7072, p = .86, correct = FALSE)
```

(b) Now use your software to test the null hypothesis that $\pi = .86$. In Stata, the command `prtest VotedPres = .86`. In R, the command is `prop.test(#cat, n, p = .86, correct = FALSE)`. In this command, you need to insert the number of people who say they voted for #cat, and the overall sample size for n. Interpret the output from your software to provide the results of the significance test.

```
##
##  1-sample proportions test without continuity correction
##
## data:  5956 out of 7072, null probability 0.86
## X-squared = 18.622, df = 1, p-value = 1.594e-05
## alternative hypothesis: true p is not equal to 0.86
## 95 percent confidence interval:
##  0.8335125 0.8505051
## sample estimates:
##         p
## 0.8421946
```

The p-value from the test is: 0.00001594, which is less than the α (.05) for a 95% confidence level, so we can reject the null hypothesis that the true proportion is .86

5. For this question, use the dataset `anes2020subset`. In opinion surveys, a feeling thermometer is a 0-100 scale that measures how warmly the respondent feels toward a person, group, or institution. The idea is to connect feelings to a temperature scale, which is interval-level. The variable `CongressTherm` is the feeling thermometer for the U.S. Congress.

```r
# note: this is code for question 5a
# length(na.omit(anes2020$CongressTherm))
mean(anes2020$CongressTherm, na.rm = T)
```

(a) Use your software to get the mean and standard deviation of this variable.

```
## [1] 44.35548
```

```r
sd(anes2020$CongressTherm, na.rm = T)
```

```
## [1] 21.72543
```

(b) Using the formulas to perform the calculations, test the null hypothesis that $\mu = 45.0$ at $\alpha = .05$. Be sure to perform all the steps.

a) assumptions: We assume this is a random sample. This is a mean so we would normally use a t distribution, however when n is much larger than 30 (n = 7359), the t distribution will mirror the z distribution, so we can use a z test.

b) hypothesis: $H_0 : \mu = 45.0, H_a : \mu \neq 45.0$
c) calculate test statistic and critical values: The critical value: $Z_{\alpha=.05} = |1.96|$
The test statistic:

$$Z = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}}, \text{ where } \hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{21.72543}{\sqrt{7359}} = 0.2532556$$

$$Z = \frac{44.35548 - 45}{0.2532556} = -2.544939$$

d) calculate the p-value: To calculate the p-value, we locate the observed z-statistic (~2.54) in the z distribution table: .0055. Since we are doing a two-sided test, we need to double this proportion to get the p-value: 0.011
e) conclude: Since the absolute value of the test statistic (~2.54) xceeds the critical value (1.96), we can reject the null hypothesis that 45 is the true mean of the sample. Similarly, since the p-value (0.011) is less than alpha of 0.05, we can reject the null hypothesis.

(c) Now use your software to test the null hypothesis that $\mu = 45.0$. In your output, locate the t-statistic and p-value (in Stata, use the p-value for the two-sided test). Interpret this output to state the results of the significance test. In Stata, the command is `ttest CongressTherm=45.0`. In R, the command is `t.test(anes2020$CongressTherm, mu = 45.0)`.

```
t.test(anes2020$CongressTherm, mu = 45.0)
```

```
##
##  One Sample t-test
##
## data:  anes2020$CongressTherm
## t = -2.5449, df = 7358, p-value = 0.01095
## alternative hypothesis: true mean is not equal to 45
## 95 percent confidence interval:
##  43.85903 44.85194
## sample estimates:
## mean of x
##  44.35548
```

The command outputs a p-value of 0.01095, which is less than the α of .05, (and very close to what we calculated above: 0.011) so we can reject the null hypothesis that the true mean is 45.

```
t.test(anes2020$CongressTherm, mu = 44.0)
```

(d) Now, using your software, perform a significance test in which H_0 is $\mu = 44.0$. Interpret the results.

```
##
##  One Sample t-test
##
## data:  anes2020$CongressTherm
## t = 1.4037, df = 7358, p-value = 0.1605
## alternative hypothesis: true mean is not equal to 44
```

```
## 95 percent confidence interval:
##  43.85903 44.85194
## sample estimates:
## mean of x
##  44.35548
```

In this case, the p-value output from the function is 0.1605, which is larger than α of 0.05, so we **cannot** reject the null hypothesis that the true mean is 45.