Sections 8.2, 9.1, and 10.2 Population Proportion

Sections 8.2 – Distribution of the Sample Proportion

We might be interested in

- the proportion of US males who have health insurance,
- the proportion of imported cars in the US,
- the proportion of Americans who have Type II diabetes,

P: Population Proportion.

$\widehat{P} = \frac{X}{n}$: Sample Proportion. $\widehat{P}$ is a point estimate of P, the population proportion.

Example 1: A sample of 200 married couples selected from throughout the United States showed that for 84 of the couples, both the husband and wife held full-time jobs. Use the sample results to estimate the proportion of all married couples in the United States for which both the husband and wife hold full-time jobs.

Solution: $\widehat{P} = \frac{x}{n} = \frac{84}{200} = 0.42$

The Sampling Distribution of the proportion-CLT

The $\mu_{\widehat{P}} = E\left(\widehat{P}\right) = P$ and the $\sigma_{\widehat{P}} = \sqrt{\frac{p(1-p)}{n}}$. For a simple random sample of size n from a population with a proportion, P, the sampling distribution of the sample proportion, $\widehat{P}$, is approximately normally distributed with mean, P, and standard deviation of $\sigma_{\widehat{P}} = \sqrt{\frac{p(1-p)}{n}}$, **if n < 0.05N and np(1 – p) ≥ 10**, i.e $\widehat{P} \sim N\left(P, \sqrt{\frac{P(1-P)}{n}}\right)$.

Example 2: A county in West Virginia has a 9% unemployment rate. To monitor the unemployment rate in that county, a monthly survey of 800 individuals is conducted by a state agency.

(a) Describe the shape of the sampling distribution of $\widehat{P}$ when a sample size of 800 is used?

(b) In the sample of 800, what is the probability at least 64 people will be unemployed?

(a) The shape is approximately Normal since n < 0.05N and np(1 – p) ≥ 10 i.e $\widehat{P} \sim N\left(P, \sqrt{\frac{P(1-P)}{n}}\right) \Rightarrow \widehat{P} \sim N\left(0.09, \sqrt{\frac{0.09(1-0.09)}{800}}\right) \Rightarrow \widehat{P} \sim N(0.09, 0.0101)$

(b) $P(X \geq 64) = P\left(\frac{X}{800} \geq \frac{64}{800}\right) = P\left(\widehat{P} \geq 0.08\right) = Normalcdf\left(0.08, E99, 0.09, 0.0101\right) = 0.838937$

Sections 9.1 Confidence interval for a population proportion

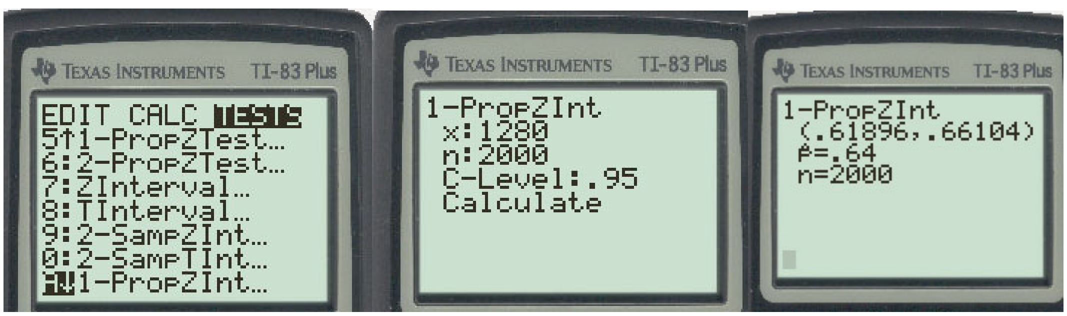
Confidence interval for a proportion: $\widehat{P} \pm Z_{\frac{\alpha}{2}}\sqrt{\dfrac{\widehat{P}\left(1-\widehat{P}\right)}{n}}$

Example 1. In a Roper Organizational poll of 2,000 adults, 1,280 have money in regular saving accounts. Find a 95% confidence interval for the true proportion of adults who have money in regular saving accounts.

Solution. $\widehat{P} = \dfrac{x}{n} = \dfrac{1280}{2000} = 0.64$, (0.619, 0.661) **From TI-83/84**

We are 95% confident that the percentage of adults having money in the saving account is **61.9% to 66.1%.**

Using TI-83/84: stat, TESTS, 1-PropZInterval (is A).
Enter X: 1280, n: 2000, C-Level: 95 or .95, put your cursor on “calculate” and hit enter. You should get the range (0.619, 0.661).



Determination of the Sample Size: **Let** $\mathbf{E} = Z_{\frac{\alpha}{2}}\sqrt{\dfrac{p(1-p)}{n}}$

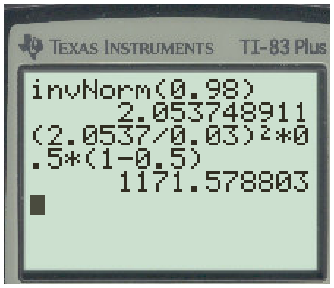
Solving for n gives us, n = $\left(\dfrac{Z_{\frac{\alpha}{2}}}{E}\right)^2 P(1-P)$**.** To use this formula we need P**,** E, and $Z_{\frac{\alpha}{2}}$**.** If a preliminary estimate on P is given use it, if not use P **= 0.5 .** E is always given in the word problem. $Z_{\frac{\alpha}{2}}$ needs to be determined using the conficence level given in the word problem. The TI83/84 command is Invnormal(percent, μ , σ). We practice the Invnormal in Chapter 7.

Example 2. We want to estimate with a maximum error of 3%, the true proportion of VSU students who would like Friday classes during the summer semester and we want 96% confidence in our results. How many VSU students must we survey?

$\alpha = 1 - 0.96 = 0.04 \Rightarrow \frac{\alpha}{2} = \frac{0.04}{2} = 0.02$. We need to find the value of $Z_{0.02}$ that has an area to the right of 0.02 and an area to the left of 1-0.02 = 0.98 . The calculator can only find Z scores using the area to the left. Hence; Invnormal(0.98).

Using TI-83/84: 2nd and vars, invnorm(is #3). Invnormal(0.98)=2.053748911

To apply the sample size formula, take 4 decimals without rounding; i.e. 2.0537. Since P is NOT given, then P=0.5. E=0.03. Now we can use the formula.



Solution.n= $\left(\frac{Z_{\frac{\alpha}{2}}}{E}\right)^2 P(1-P) = \left(\frac{2.0537}{0.03}\right)^2$ (0.5)(1-0.5)

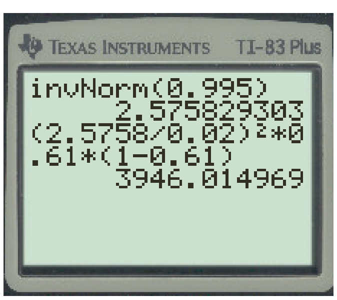
=1171.57 $\approx$ 1172. So, we need to survey 1172 voters. Please note: For the final answer for **n** we always round up regardless of the decimal. It does not follow the rounding rules.

Example 3. The current percentage of voters favoring the Republican nominee is 61%. A pollster is hired to determine the percentage of voters favoring the Republican nominee. If we require 99% confidence that the estimated value is within 2% of the true value, how large should the sample be?

$\alpha = 1 - 0.99 = 0.01 \Rightarrow \frac{\alpha}{2} = \frac{0.01}{2} = 0.005$. We need to find the value of $Z_{\frac{\alpha}{2}=0.005}$ that has an area to the right of 0.005 and an area to the left of 1-0.005 = 0.995 . The calculator can only find Z scores using the area to the left. Hence; Invnormal(0.995).

Using TI-83/84: 2nd and vars, invnorm(is #3). Invnormal(0.995)=1.575829303

To apply the sample size formula, take 4 decimals without rounding; i.e. 1.5758. P is given P=0.61 and E=0.02. Now we can use the formula.



Solution. n = $\left(\frac{Z_{\frac{\alpha}{2}}}{E}\right)^2 P(1-P) = \left(\frac{2.5758}{0.02}\right)^2$ (0.61)(1-0.61)=

3946.01 $\approx$ 3,947 . So, we need to survey 3,947 voters. Please note: For the final answer for **n** we always round up regardless of the decimal. It does not follow the rounding rules.

Homework: Sections 8.2 and 9.1 Online - MyStatLab

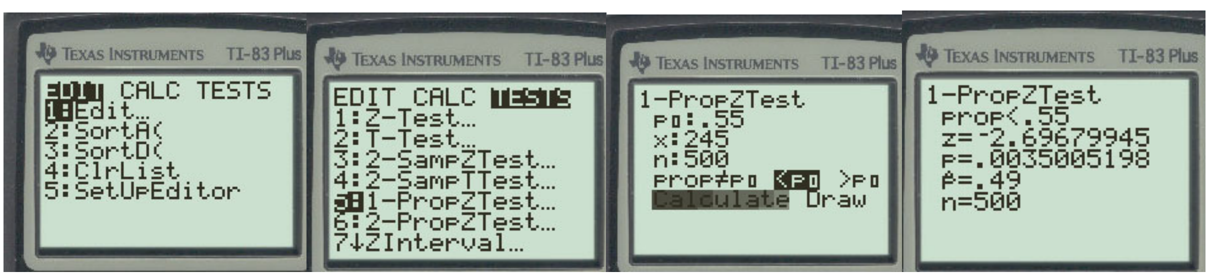
Section 10.2 Hypothesis Testing for a Population Proportion

This is a hypothesis testing for Proportion(P) NOT the Mean(μ). Therefor, we are changing from μ to P ; i.e. μ_0 to P_0.

$H_0 : P = P_0 \quad vs \quad H_1 : \begin{cases} 1. & P \neq P_0 \\ 2. & P < P_0 \\ 3. & P > P_0 \end{cases}$. Everything else we know about Hypothesis Testing remains the same. Such as : Reject H_0 if the P-Value is less than α and Do NOT Reject H_0 if the P-Value is greater or equal to α.

Example 6. Mr. Dixon, a Republican, claims that he has the support of 55% of all voters in the 23rd U.S. Congressional District. Can the Central Committee conclude that less than 55% of all voters support Mr. Dixon, if, out of a random sample of 500 registered voters, only 245 expressed their preference for Mr. Dixon? Use α = 0.01 as the level of significance.

Using TI-83/84: stat, TESTS, 1-PropZTest (is #5), then choose "stats".
Enter P_o: 0.55, X: 245, n: 500, Prop: $< P_o$ (put your cursor on the appropriate Test and hit "enter"), put your cursor on "calculate" and hit enter. Look for the P-Value: p=.0035005198. (see calculator pictures below)



Soln. Step1: State the null and alternative hypotheses.

$$\mathbf{H}_0 : \text{P} = 0.55 \quad \text{vs} \quad \mathbf{H}_1 : \text{P} < 0.55$$

Step2: Write the given information using proper symbols.

n=500, $\widehat{P} = \frac{245}{500} = 0.49$, $\widehat{P} \sim N\left(P_0 , \sqrt{\frac{P_0(1-P_0)}{n}}\right) \Rightarrow \widehat{P} \sim N\left(0.55 , \sqrt{\frac{0.55(1-0.55)}{500}} = 0.0222\right)$, α=.01

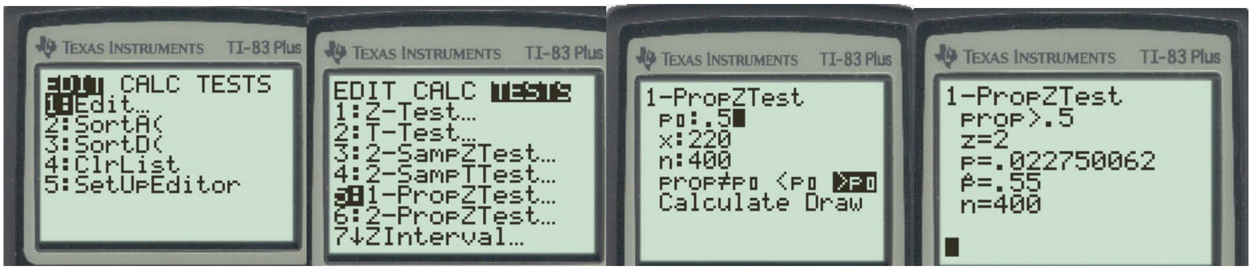
Test statistic. $Z = \frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$ = -2.69679945 (From TI-83/84)

Step3: From TI83/84. P-value= 0.0035. Since the **P-value** is less than **α=0.01**, Reject $\mathbf{H}_0$.

Step4: Conclusion. Since the P-value=.0035 is less than α=.01, we reject the null hypothesis and claim that the Central Committee has sufficient evidence to conclude that less than 55% of all voters support Mr. Dixon.

Example 7. The sponsor of a weekly television show believes that the studio audience consists of an equal number of men and women. Out of 400 persons attending the show on a given night, 220 are men. Using a level of significance of 0.03, can we conclude that more than 50% of the people attending the show are men?

Using TI-83/84: stat, TESTS, 1-PropZTest (is #5), then choose "stats".
Enter P_o: 0.5, X: 220, n: 400, Prop: $> P_o$ (put your cursor on the appropriate Test and hit "enter"), put your cursor on "calculate" and hit enter. Look for the P-Value: p=.0035005198. (see calculator pictures below)



Soln. Step1: State the null and alternative hypotheses.

$$\mathbf{H}_0: p = 0.5 \quad \text{vs} \quad \mathbf{H}_a: p > 0.5$$

Step2: Write the given information using proper symbols.

n=400, $\widehat{P} = \frac{220}{400} = 0.55$, $\widehat{P} \sim N\left(P_0\,,\ \sqrt{\frac{P_0(1-P_0)}{n}}\right) \Rightarrow \widehat{P} \sim N\left(0.5\,,\ \sqrt{\frac{0.5(1-0.5)}{400}} = 0.025\right)$, α= 0.03

Test statistic. $Z = \frac{\hat{P}-P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = 2$ (From TI-83/84)

Step3: From TI83/84. P-value= 0.0228. Since the **P-value** is less than **α=0.03**, Reject $\mathbf{H}_0$.

Step4: Conclusion. Since the P-value=.0228 is less than α=.03, we reject the null hypothesis and claim we have sufficient evidence to conclude that more than 50% of the people attending the show are men.

Helpful Hints for the Homework

Since is a clinical trial we can reasonably assume the data to be random.

6. In a clinical trial, 17 out of 829 patients taking a prescription drug daily complained of flulike symptoms. Suppose that it is known that 1.7% of patients taking competing drugs complain of flulike symptoms. Is there sufficient evidence to conclude that more than 1.7% of this drug's users experience flulike symptoms as a side effect at the $\alpha = 0.01$ level of significance?

829(0.017)(1−0.017) = 13.8534

Because $np_0(1-p_0)$ = 13.89 (1) > 10, the sample size is (2) less than 5% of the population size, and the sample (3) can be reasonably assumed to be random the requirements for testing the hypothesis (4) are satisfied.
(Round to one decimal place as needed.)

Note: in (1) the answer is always " > " and (2) the answer is always "less than". For (3) Since this is a clinical trial the answer is "can be reasonably assume to be random"

7. According to a certain government agency for a large country, the proportion of fatal traffic accidents in the country in which the driver had a positive blood alcohol concentration (BAC) is 0.38. Suppose a random sample of 109 traffic fatalities in a certain region results in 51 that involved a positive BAC. Does the sample evidence suggest that the region has a higher proportion of traffic fatalities involving a positive BAC than the country at the $\alpha = 0.01$ level of significance?

109(0.38)(1−0.38) = 25.6804 ≈ 25.7

Because $np_0(1-p_0)$ = 25.7 (1) > 10, the sample size is (2) less than 5% of the population size, and the sample (3) is given to be random. the requirements for testing the hypothesis

Note: in (1) the answer is always " > " and (2) the answer is always "less than". For (3) Since this is given to be a random sample the answer is "is given to be random"

Similarly, #8 and #9 in the homework are like #6.

Also, #10 in the homework is like #7.

Homework-Section 10.2 Online - MyStatLab