

AP Statistics – 9.2b	(2020 KEY)	Name:
Goal: 2-Sided TOH and CI for Population Proportion (p)		Date:

I. 2-Sided Test of Significance for Proportions -- Example #1 "Benford's law and fraud"

When the accounting firm AJL and Associates audits a company's financial records for fraud, they often use a test based on Benford's law. Benford's law states that the distribution of first digits in many real-life sources of data is not uniform. In fact, when there is no fraud, about 30.1% of the numbers in financial records begin with the digit 1. However, if the proportion of first digits that are 1 is significantly different from 0.301 in a random sample of records, AJL and Associates does a much more thorough investigation of the company. Suppose that a random sample of 300 expenses from a company's financial records results in only 68 expenses that begin with the digit 1. Should AJL and Associates do a more thorough investigation of this company?

- Parameter of Interest $P = \text{TRUE PROPORTION OF EXPENSES THAT BEGIN WITH 1}$
- Level of Significance $\alpha = .05$
- Choice of Test $1 \text{ SAMPLE Z TEST FOR PROPORTIONS}$
- Null Hypothesis $H_0: p = 0.301$
- Alternative Hypothesis $H_A: p \neq 0.301$ (IS IT DIFFERENT FROM 0.301)
- Conditions of Test

Random sample stated ✓

INDEPENDENT: 10% CONDITION $n = 300 \leq \frac{1}{10}$ (all records)

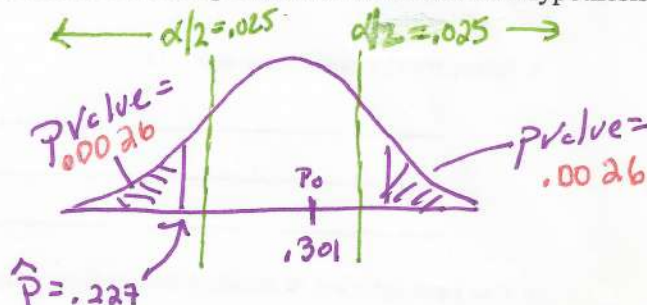
NORMAL: $np_0 \rightarrow 300(.301) = 90.3 \geq 10$ ✓
 $n(1-p_0) \rightarrow 300(.699) = 209.7 \geq 10$ ✓

- Sampling Distribution (Sketch of the sampling distribution of the sample statistic under the null hypothesis, indicating the mean)

$$n = 300$$

$$\hat{p} = \frac{68}{300} = 0.227$$

$$p_0 = 0.301$$



- Test Statistic (clearly show calculation)

Formula: $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ FILLIN $Z = \frac{.227 - .301}{\sqrt{\frac{(.301)(.699)}{300}}} = -2.79$

- P-value (Use correct probability notation.) 2 TAIL TEST (MULT * 2)

$$P\text{-value} = P(Z \leq -2.79) * 2 = 2(.0026) = .0052$$

CALC:
1 PROP Z TEST

$$Z = -2.81$$

$$P\text{-value} = 0.0050$$

- Conclusions (in context)

SINCE THE PVALUE (.0052) IS LESS THAN $\alpha = .05$,
 WE REJECT H_0 . THERE IS CONVINCING EVIDENCE
 THAT THE PROPORTION OF EXPENSES THAT HAVE
 THE FIRST DIGIT OF 1 IS NOT .301.

THEREFORE AJL SHOULD DO MORE INVESTIGATION
 OF THIS COMPANY'S RECORDS FOR FRAUD.

This is a review from CH8

II. Confidence Interval for Proportions -- Example #1 "Benford's law and fraud(continued)"

Find and interpret an appropriate confidence interval for the true proportion of expenses that begin with the digit 1 for the company in the previous alternate example. Use your interval from (a) to decide whether this company should be investigated for fraud.

- Parameter of Interest p = the true proportion of expenses that begin with the digit 1

- Confidence Level: $TOH \alpha = 0.05 \rightarrow CL = 1 - \alpha = 95\%$

- Choice of Test: 1 SAMPLE Z INTERVAL FOR PROPORTIONS

- Conditions of Test

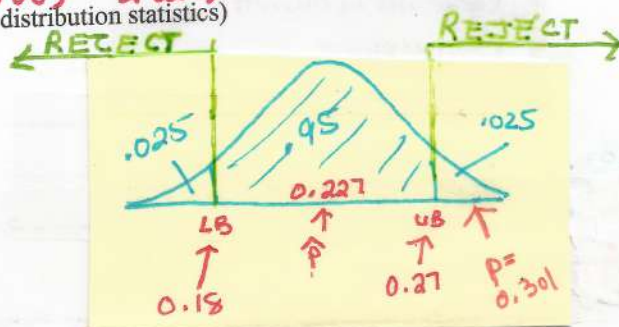
Random AND INDEPENDENT - SAME AS TOH

Normal:

$$n\hat{p} \rightarrow (.227)(300) = 68 \geq 10$$

$$n(1-\hat{p}) \rightarrow (.773)(300) = 232 \geq 10$$

- Sampling Distribution (Sketch graph and provide sampling distribution statistics)



$$n = 300$$

$$\hat{p} = \frac{68}{300} = 0.227$$

$$CL = .95 \rightarrow Z^* = \pm 1.96$$

- Confidence Interval (clearly show calculation)

$$.227 \pm 1.96 \sqrt{\frac{(.227)(.773)}{300}}$$

$$.227 \pm 1.96 (.0242)$$

$$.227 \pm 0.047$$

$$[0.180, 0.274]$$

$$\hat{p} \pm Z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- Interpret the confidence interval

CALC:

1-PROPZINT

[0.179, 0.274]

WE ARE 95% CONFIDENT THE TRUE PROPORTION OF EXPENSES AT THIS COMPANY BEGINS WITH THE DIGIT 1 IS BETWEEN 0.180 AND 0.274.

Use your interval to decide whether this company should be investigated for fraud.

SINCE 0.301 DOES NOT FALL IN OUR 95% CONFIDENCE INTERVAL [0.180, 0.274], WE HAVE SUFFICIENT EVIDENCE TO REJECT H_0 AT $\alpha = .05$.

* NOTE: 0.301 IS NOT A PLAUSIBLE POPULATION PARAMETER

III. Compare the differences between doing a 2-tail TOH versus a CI $CL = 1 - \alpha$

① NOTICE IN THIS EXAMPLE A 2-TAIL TOH AND CI GAVE US THE SAME DECISION

② A TWO TAIL TOH AND CI GIVE THE SAME DECISION

③ A CI GIVES US MORE INFORMATION SINCE IT PROVIDES ALL THE PLAUSIBLE VALUES

FOR THE POPULATION PARAMETER.

Test of Significance Template

Parameter of Interest	$P = \text{true proportion of restaurant workers who say WORK STRESS HAS A NEGATIVE IMPACT}$		
Choice of Test	1 Sample Z test for proportions		
Level of Significance	$\alpha = 0.05$		
Null Hypothesis	English: $H_0: P = 0.75 \leftarrow \text{NATIONAL SURVEY RESULTS}$ Symbols:		
Alternative Hypothesis	English: $H_a: P \neq 0.75 \leftarrow \text{DOES THE CHAIN DIFFER FROM NAT'L}$ Symbols:		
Conditions of Test	Random sample of 100 employees IND. - 10% Condition $n = 100 \leq \frac{1}{10} (\text{all emps in chain})$ NORMAL: $.75(100) = 75 \geq 10 \checkmark$ $.25(100) = 25 \geq 10 \checkmark$ $P_0 = 0.75$		
Sampling Distribution	Sketch of the sampling distribution of the sample statistic under the null hypothesis, indicating the mean: $n = 100$ $P_0 = 0.75$ $\hat{P} = 68/100 = 0.68$ 		
Test Statistic	Formula: $Z = \frac{\hat{P} - P}{\sqrt{\frac{P(1-P)}{n}}}$	Plug-ins & Value: $Z = \frac{.68 - .75}{\sqrt{\frac{(.75)(.25)}{100}}} = -1.62$	CALC 5:1PROPTTEST $Z = -1.62$
P-value	Use correct probability notation. $P\text{Value} = 2 * P(Z \leq -1.62) = .053 * 2 = .106$		$P\text{value} = .1060$
Meaning of the P-value	ASSUMING THE TRUE PROPORTION OF EMPLOYEE WHO SAY WORK STRESS HAS A NEGATIVE IMPACT ON THEM IS 0.75, THERE IS A 0.106 PROBABILITY OF GETTING		
Conclusions	☐ Reject null hypothesis ☒ Fail to reject null hypothesis English: SINCE THE PVALUE (0.106) IS GREATER THAN $\alpha = 0.05$, WE FAIL TO REJECT H_0 . WE DO NOT HAVE CONVINCING EVIDENCE TO SAY EMPLOYEES AT THIS RESTAURANT DIFFERS FROM THE NATIONAL AVERAGE FOR NEGATIVE IMPACT OF WORK STRESS A SAMPLE PROPORTION OF 0.68 OR FURTHER AWAY, PURELY BY CHANCE		

IV. CYU – page 558 – 2-sided Test (complete on template)

V. CYU – page 561 – Using computer output for CI to compare to 2-tail TOH on pg 558

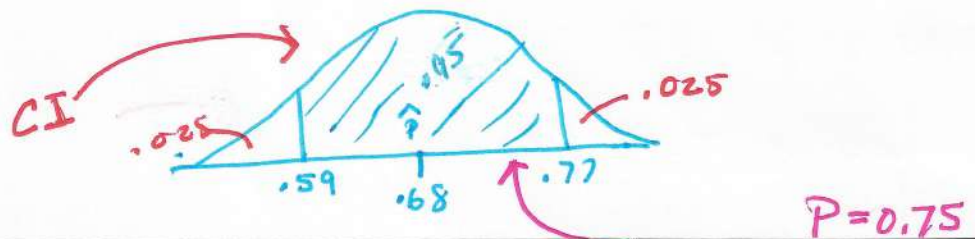
THIS COMPUTER OUTPUT IS FOR THE RESTAURANT STRESS.
COMPLETE ABOVE IN SECTION IV.

Interpret CI:

95% CI [0.589, 0.771]

WE ARE 95% CONFIDENT THE INTERVAL
0.589 TO 0.771 CAPTURE THE TRUE PROPORTION
OF RESTAURANT WORKERS WHO SAY STRESS
HAS A NEGATIVE IMPACT ON THEIR WORK

CI decision:



SINCE THE CONFIDENCE INTERVAL [0.59, 0.77]
INCLUDES OUR POPULATION PARAMETER (0.75)
OUR DECISION WOULD BE THE SAME
"FAIL TO REJECT H_0 "

The CI gives more information. It gives all
the plausible values for a population parameter.

VI. Page 563 – #50 (2-tail TOH), #52 (versus CI) - complete on template

#50 TON

COMPARE 2-TAIL TESTS VS CI'S (pg 563-4)

#52 CI

Parameter of Interest	p = true proportion of 1st Year College students who believe its important to be financially well off	
Choice of Test	1 sample Z test for Proportions	1 sample Z-interval for proportions
Level of Significance	$\alpha = 0.05$	$CL = .95 (1 - \alpha)$
Null Hypothesis	$H_0: p = .73$	Same
Alternative Hypothesis	$H_a: p \neq .73$	Same
Conditions of Test	Random Sample stated INDEPENDENT: $200 \geq \frac{1}{10}$ (1st yr students) Normal: $200(.73) = 146 > 10 \checkmark$ $p_0 = .73$ $200(.27) = 54 > 10 \checkmark$	Random - same as TON INDEPENDENT - same as TON NORMAL: $200(.66) = 132 > 10 \checkmark$ $\hat{p} = .66$ $200(.34) = 68 > 10 \checkmark$
Sampling Distribution	Sketch of the sampling distribution of the sample statistic under the null hypothesis, indicating the mean: $H_0:$ $n = 200$ $\hat{p} = 132/200 = .66$ 	CI for $\hat{p}$
CALCULATIONS	$Z = \frac{.66 - .73}{\sqrt{\frac{(.73)(.27)}{200}}} = -2.23$	$.66 \pm 1.96 \sqrt{\frac{(.66)(.34)}{200}}$ $.66 \pm 0.066$ $[.594, .726]$
P-value	$p\text{-value} = 2 * P(Z \leq -2.23) = 0.026$	
CONCLUSIONS	Since the pvalue (.026) is less than $\alpha = 0.05$, we reject H_0 . We have convincing evidence that students at this college thinking being well off differs from the national average.	We are 95% confident the true proportion for this population parameter is between .594 and .726. Since the population parameter (.73) falls outside the confidence interval, we reject H_0 .