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 #~~3~~ 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= TRUE PROPORTION OF EXPENSES THAT BEGIN WITH 1
- **Level of Significance** $\alpha = .05$
- **Choice of Test** 1 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: $np_0 \longrightarrow 300(.301) = 90.3 \geq 10$ ✓

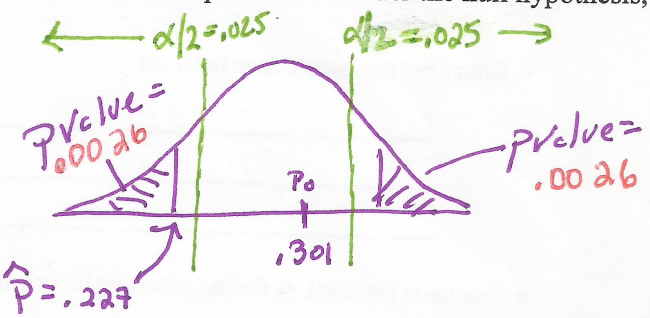
$n(1-p_0) \longrightarrow 300(.699) = 209.7 \geq 10$ ✓

- **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: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ FILL IN $z = \frac{.227 - .301}{\sqrt{\frac{(.301)(.699)}{300}}} = -2.79$

CALC: 1 PROPZTEST

$z = -2.81$

Pvalue = 0.0050

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

Pvalue $= P(z \leq -2.79) * 2 = 2(.0026) = .0052$

- **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 BE DO MORE INVESTIGATION OF THIS COMPANY'S RECORDS FOR FRAUD.

This is a review from CH8

II. Confidence Interval for Proportions -- Example #~~4~~ 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 $\alpha = 0.05$ $\longrightarrow$ $CL = 1 - \alpha = 95\%$
- **Choice of Test:** 1 SAMPLE ZINTERVAL FOR PROPORTIONS
- **Conditions of Test**

Random AND INDEPENDENT – SAME AS TOH

NORMAL:

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

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

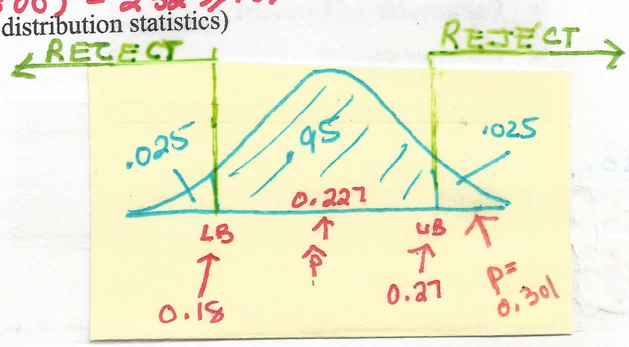
Samp dist for $\hat{p}$

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

$n = 300$

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

$CL = .95 \longrightarrow z^* = \pm 1.96$



- **Confidence Interval** (clearly show calculation)

FORMULA

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

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

$.227 \pm 1.96\ \underbrace{(.0242)}_{SE(\hat{p})}$

$.227 \pm \underbrace{0.047}_{ME}$

$[0.180, 0.274]$

CALC: 1-PROPZINT $[0.179, 0.274]$

- **Interpret the confidence interval**

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.**

$H_0: p = 0.301$

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

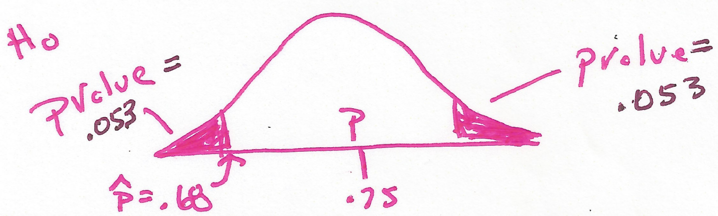
III. Compare the differences between doing a 2-tail TOH versus a CI

$\boxed{CL = 1 - \alpha}$

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.

CYU pg 558

Test of Significance Template

Parameter of Interest	P = 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: $H_0: P = 0.75$ ← NATIONAL SURVEY Results Symbols:
Alternative Hypothesis	English: $H_a: P \neq 0.75$ ← DOES THE CHAIN DIFFER FROM NAT'L Symbols:
Conditions of Test	Random sample of 100 employees IND. – 10% CONDITION $n = 100 \leq \frac{1}{10}$(all emps in chain) NORMAL: $.75(100) = 75 \geq 10$ ✓ $P_0 = 0.75$ $.25(100) = 25 \geq 10$ ✓
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: 1 PROPZTEST $Z = -1.62$
P-value	Use correct probability notation. Pvalue $= 2 * P(Z \leq -1.62) = .053 * 2 = .106$ Pvalue = .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 A SAMPLE PROPORTION OF 0.68 OR FURTHER AWAY, PURELY BY CHANCE
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

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

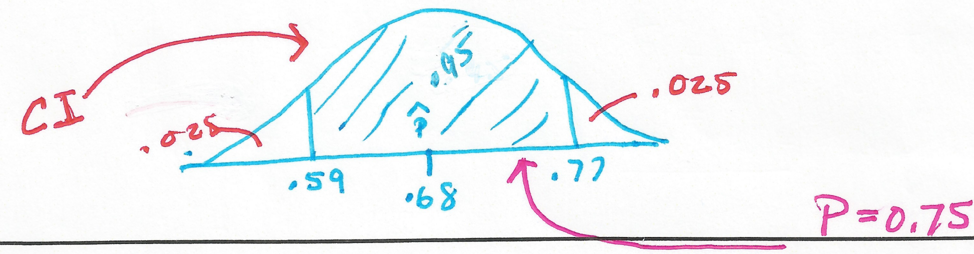
THIS COMPUTER OUTPUT IS FOR THE RESTAURANT STRESS.
COMPLETE ABOVE IN SECTION IV.

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:

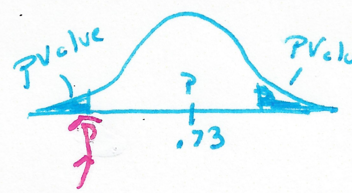
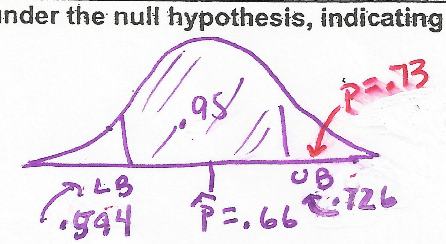


SINCE THE CONFIDENCE INTERVAL [.59, .77] INCLUDES OUR POPULATION PARAMETER (0.75) OUR DECISION WOULD BE THE SAME "FAIL TO REJECT H_0"

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

Compare 2-Tail Tests vs CI's (pg 563-4)

	#50 TOH	#52 CI
Parameter of Interest	p = true proportion of 1st year college students who believe its importent to be finencilly 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: $200 \le \frac{1}{10}$ (1st yr students) Normal: $p_0 = .73$ $200(.73) = 146 \ge 10$ ✓ $200(.27) = 54 \ge 10$ ✓	Random – same as TOH INDEPENDENT – same as TOH NORMAL: $\hat{p} = .66$ $200(.66) = 132 \ge 10$ ✓ $200(.34) = 68 \ge 10$ ✓
Sampling Distribution	Sketch of the sampling distribution of the sample statistic under the null hypothesis, indicating the mean: Ho:  n=200 $\hat{p} = 132/200 = .66$	CI for $\hat{p}$ 
CALCULATIONS	$Z = \dfrac{.66 - .73}{\sqrt{\frac{(.73)(.27)}{200}}} = -2.23$	$.66 \pm 1.96 \sqrt{\frac{(.66)(.34)}{200}}$ $.66 \pm 0.066$
P-value	pvalue $= 2 * P(Z \le -2.23) = 0.026$	$[.594, .726]$
CONCLUSIONS	Since the pvalue (.026) is less then $\alpha = 0.05$, we reject Ho. 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 Ho.