AP STATS NOTES [12.2]

12.2 Tests about a Population Proportion

When the three important conditions are met—SRS, Normality, and independence—the sampling distribution of $\hat{p}$ is approximately Normal with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. For confidence intervals, we substitute $\hat{p}$ for p in the last expression to obtain the standard error. When performing a significance test, however, the null hypothesis specifies a value for p, which we will call p_0. We assume that this value is correct when performing our calculations. If we standardize $\hat{p}$ by subtracting its mean p_0 and dividing by its standard deviation, we obtain the following z statistic:

$$z = \frac{\text{estimate} - \text{hypothesized parameter value}}{\text{standard deviation of estimate}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Here is a summary of the details for a ***one-proportion z test.***

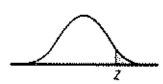
The One-Proportion z Test

Choose an SRS of size n from a large population with unknown proportion p of successes. To test the hypothesis $H_0: p = p_0$, compute the **z statistic**

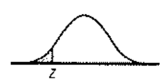
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

In terms of a random variable Z having the standard Normal distribution, the approximate P-value for a test of H_0 against

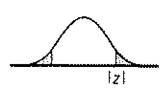
$H_a: p > p_0$ is $P(Z \geq z)$



$H_a: p < p_0$ is $P(Z \leq z)$



$H_a: p \neq p_0$ is $2P(Z \geq |z|)$



Normality condition: Use this test when the expected number of successes np_0 and failures $n(1-p_0)$ are both at least 10.

NAME PER

Example 12.7 *Work stress*
Performing a one-proportion z test

According to the National Institute for Occupational Safety and Health, job stress poses a major threat to the health of workers. A national survey of restaurant employees found that 75% said that work stress had a negative impact on their personal lives.[6] A random sample of 100 employees from a large restaurant chain finds that 68 answer "Yes" when asked, "Does work stress have a negative impact on your personal life?" Is this good reason to think that the proportion of all employees in this chain who would say "Yes" differs from the national proportion $p_0 = 0.75$?

Example 12.9 Estimating work stress

Confidence intervals give more information

The restaurant worker survey in Example 12.7 found that 68 of a random sample of 100 employees agreed that work stress had a negative impact on their personal lives. We checked the three important conditions for performing the significance test earlier. Before we construct a confidence interval for the population proportion p, we should verify that both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10. Since the number of successes and failures in the sample are 68 and 32, respectively, we can proceed with the calculation.

Minitab

```
Session
Test and Confidence Interval for One Proportion
Test of p = 0.75 vs p not = 0.75
Sample    X     N    Sample p        95.0 % CI            Z-Value   P-Value
1        68   100    0.680000   (0.588572, 0.771428)      -1.62     0.106
```

CrunchIt!

Session

Hypothesis test results:

```
p = proportion of successes for population
Parameter: p
H0 : Parameter = 0.75
HA : Parameter not = 0.75
```

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	68	100	0.68	0.04330127	-1.6165807	0.106

CAUTION

Confidence intervals provide additional information that significance tests do not—namely, a range of plausible values for the true population parameter p. For inference about a population proportion, confidence intervals and two-tailed tests do not have the perfect correspondence that we saw with inference about a population mean. But the relationship is similar.

Note that the standard error used for the confidence interval is estimated from the data, whereas the denominator for the test statistic z is based on the value assumed in the null hypothesis.

12.26 Side effects An experiment on the side effects of pain relievers assigned arthritis patients to one of several over-the-counter pain medications. Of the 440 patients who took one brand of pain reliever, 23 suffered some "adverse symptom."

(a) Does this experiment provide strong evidence that fewer than 10% of patients who take this medication have adverse symptoms? Perform an appropriate test.

AP STATS NOTES [12.2]

12.2 Tests about a Population Proportion → p

SAMPLE PROPORTION → $\hat{p} = \frac{\text{\# OF SUCCESSES}}{\text{SAMPLE SIZE}}$

When the three important conditions are met—SRS, Normality, and independence—the sampling distribution of $\hat{p}$ is approximately Normal with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. For confidence intervals, we substitute $\hat{p}$ for p in the last expression to obtain the standard error. When performing a significance test, however, the null hypothesis specifies a value for p, which we will call p_0. We assume that this value is correct when performing our calculations. If we standardize $\hat{p}$ by subtracting its mean p_0 and dividing by its standard deviation, we obtain the following z statistic:

$$z = \frac{\text{estimate} - \text{hypothesized parameter value}}{\text{standard deviation of estimate}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Here is a summary of the details for a one-proportion z test.

- USE PROPORTIONS TO SUMMARIZE CATEGORICAL VARIABLES

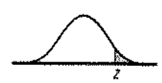
The One-Proportion z Test

Choose an SRS of size n from a large population with unknown proportion p of successes. To test the hypothesis H_0: $p = p_0$, compute the **z statistic**

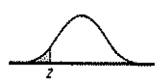
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

In terms of a random variable Z having the standard Normal distribution, the approximate P-value for a test of H_0 against

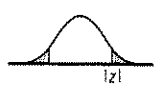
H_a: $p > p_0$ is $P(Z \geq z)$



H_a: $p < p_0$ is $P(Z \leq z)$



H_a: $p \neq p_0$ is $2P(Z \geq |z|)$



Normality condition: Use this test when the expected number of successes np_0 and failures $n(1-p_0)$ are both at least 10.

$np_0 \geq 10 \quad n(1-p_0) \geq 10$

NAME MCCREARY PER X

Example 12.7 *Work stress*
Performing a one-proportion z test

According to the National Institute for Occupational Safety and Health, job stress poses a major threat to the health of workers. A national survey of restaurant employees found that 75% said that work stress had a negative impact on their personal lives.[6] A random sample of 100 employees from a large restaurant chain finds that 68 answer "Yes" when asked, "Does work stress have a negative impact on your personal life?" Is this good reason to think that the proportion of all employees in this chain who would say "Yes" differs from the national proportion $p_0 = 0.75$?

$\hat{p} = 68/100 = .68$ $n = 100$

I HYPOTH

$H_0: p = .75$
$H_a: p \neq .75$

p REPRESENTS THE TRUE PROPORTION OF THE EMPLOYEES FROM THIS CHAIN THAT SAY WORK STRESS HAS A NEGATIVE IMPACT ON THEIR PERSONAL LIVES

II CONDITIONS ① SRS → RANDOM SAMPLE OF 100

② NORMALITY → $np_0 = (100)(.75) = 75 \geq 10$

③ INDEPENDENCE $n(1-p_0) = (100)(.25) = 25 \geq 10$

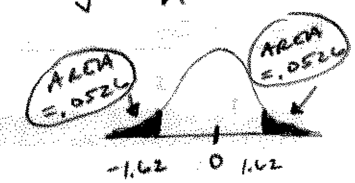
→ $N \geq 10 \cdot n$ # OF EMPLOYEES $\geq 10 \cdot 100$

WE MUST ASSUME THIS "LARGE" CHAIN HAS AT LEAST 1000 EMPLOYEES

III CALCULATIONS $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

$$z = \frac{.68 - .75}{\sqrt{\frac{(.75)(.25)}{100}}} = \frac{-.07}{.0433}$$

$z = -1.62$



P-VALUE = 2 * P(Z ≤ -1.62)
= 2 * (.0526)

P-VALUE = .1052

IV INTERPRETATION

WE FAIL TO REJECT H_0 B/C .1052 NOT ≤ .05

WE DO NOT HAVE ENOUGH EVIDENCE TO CONCLUDE THE TRUE PROPORTION OF THIS CHAIN'S EMPLOYEES WHO SUFFER FROM WORK STRESS IS DIFFERENT FROM NATIONAL PROPORTION OF .75

Example 12.9 *Estimating work stress*

Confidence intervals give more information $\hat{p} = .68$

The restaurant worker survey in Example 12.7 found that 68 of a random sample of 100 employees agreed that work stress had a negative impact on their personal lives. We checked the three important conditions for performing the significance test earlier. Before we construct a confidence interval for the population proportion p, we should verify that both $n\hat{p}$ and $n(1-\hat{p})$ are at least 10. Since the number of successes and failures in the sample are 68 and 32, respectively, we can proceed with the calculation.

$z^* = 1.96$

NORMALITY

$n\hat{p} = (100)(.68) = \boxed{68 \geq 10}$ ✓

$n(1-\hat{p}) = (100)(.32) = \boxed{32 \geq 10}$ ✓

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

$$= .68 \pm 1.96 \sqrt{\frac{(.68)(.32)}{100}}$$

$$= .68 \pm (1.96)(.0466) = .68 \pm .091$$ ✓

$\boxed{CI = (.589, .771)}$ ✓

WE ARE 95% CONFIDENT THAT BETWEEN 59% & 77% OF THIS CHAIN'S EMPLOYEES FEEL THAT WORK STRESS DAMAGES THEIR PERSONAL LIVES

Minitab

Session

```
Test and Confidence Interval for One Proportion
Test of p = 0.75 vs p not = 0.75
Sample    X     N   Sample p        95.0 % CI          Z-Value   P-Value
1        68   100   0.680000   (0.588572, 0.771428)     -1.62     0.106
```

CrunchIt! $CI = .68 \pm 1.96 \sqrt{\frac{(.68)(.32)}{100}} = .68 \pm 1.96(.0466)$

Session

Hypothesis test results:

```
p = proportion of successes for population
Parameter: p
H0 : Parameter = 0.75
HA : Parameter not = 0.75
```

$z = \frac{.68 - .75}{\sqrt{\frac{(.75)(.25)}{100}}} = \frac{-.07}{.0433} = \boxed{-1.62}$

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	68	100	0.68	0.04330127	-1.6165807	0.106

12.26 Side effects An experiment on the side effects of pain relievers assigned arthritis patients to one of several over-the-counter pain medications. Of the 440 patients who took one brand of pain reliever, 23 suffered some "adverse symptom." $\hat{p} = 23/440$

(a) Does this experiment provide strong evidence that fewer than 10% of patients who take this medication have adverse symptoms? Perform an appropriate test.

$\boxed{H_0: p = .10 \quad H_a: p < .10}$

(p) REPRESENTS THE TRUE PROPORTION OF PATIENTS WHO WOULD HAVE AN ADVERSE SYMPTOM FROM THIS BRAND OF PAIN RELIEVER

$\hat{p} = .052$

$n = 440$

① SRS → ? UNKNOWN

→ MAY NOT BE ABLE TO GENERALIZE RESULTS

② NORMALITY $np_0 = (440)(.10) = \boxed{44 \geq 10}$ ✓

$n(1-p_0) = (440)(.90) = \boxed{396 \geq 10}$ ✓

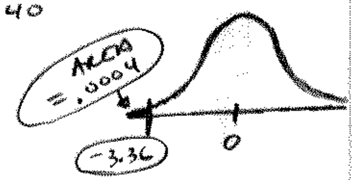
③ INDEPENDENCE

$N \geq 10 \cdot n$

$\boxed{\#\text{ OF ALL PATIENTS TAKING PAIN RELIEVER} \geq 4{,}400}$ ✓

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.052 - .10}{\sqrt{\frac{(.10)(.90)}{440}}} = \frac{-.048}{.0143}$$

$\boxed{z = -3.36}$ ✓



$\boxed{p\text{-VALUE} = .0004}$ ✓

$\boxed{\text{WE REJECT } H_0 \text{ B/C } .0004 \leq .05}$ ✓ ✓

WE HAVE VERY STRONG EVIDENCE TO CONCLUDE THAT FEWER THAN 10% OF PATIENTS WHO TAKE THIS MEDICATION HAVE ADVERSE SYMPTOMS

CAUTION

Confidence intervals provide additional information that significance tests do not—namely, a range of plausible values for the true population parameter p. For inference about a population proportion, confidence intervals and two-tailed tests do not have the perfect correspondence that we saw with inference about a population mean. But the relationship is similar.

Note that the standard error used for the confidence interval is estimated from the data, whereas the denominator for the test statistic z is based on the value assumed in the null hypothesis.

$$CI = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$