STAT 100 Lecture 18 : Inferences about a Population Proportion

Nate Strawn

November 10

Last Time...

1. Testing hypothesis about a population mean
2. Test criterion and rejection region
3. Two types of error and their probabilities
4. P–value
5. Five steps of hypothesis testing

Today's Agenda

1. Point estimation of p.
2. Confidence interval for p.
3. Determining the sample size.
4. Large sample test about p.

Inferences about a Population Proportion

The sample proportion $\hat{p} = \frac{X}{n}$. is an estimator of p.
The sample count X has the binomial distribution with mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$.
If $np \geq 15$ and $n(1-p) \geq 15$ the binomial variable X is well approximated by a normal with mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$. That is

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is approximately standard normal. Observe that

$$Z = \frac{(X - np)/n}{\sqrt{np(1-p)}/n} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}.$$

Point estimation of p.

Since $\hat{p} = X/n$, the properties of expectation give

$$E(\hat{p}) = p, \quad sd(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}.$$

Summary

Point Estimation of a Population Proportion.

Parameter: *Population proportion* p.

Data: X = *Number having the characteristic in a random sample of size* n

Estimator: $\hat{p} = \frac{X}{n}$.

$$S.E.(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} \quad Estimated\ S.E.(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

For large n, *an approximate* $100(1-\alpha)\%$ *error margin is* $z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$.

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose if a particular car were available in either blue or red body color. Of the 90 respondents 53 said BLUE. *Give a point estimate of the proportion of people preferring the blue color and attach a 95.4% error margin.*

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose if a particular car were available in either blue or red body color. Of the 90 respondents 53 said BLUE. *Give a point estimate of the proportion of people preferring the blue color and attach a 95.4% error margin.*

Solution

$\hat{p} = \frac{53}{90} = 0.59.$

Estimated $S.E.(\hat{p}) = \sqrt{\frac{0.59 \times 0.41}{90}} = 0.052.$

95.4% error margin is$= 2 \times 0.052 = 0.104.$

Confidence interval for p.

A common formula for the confidence interval is Estimator$\pm$(z value)(estimated standard error).

Definition

Large Sample Confidence Interval for p.

For large n a $100(1-\alpha)\%$ *confidence interval for p is given by*

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose if a particular car were available in either blue or red body color. Of the 90 respondents 53 said BLUE. *Compute 95.4% confidence interval for the population proportion.*

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose if a particular car were available in either blue or red body color. Of the 90 respondents 53 said BLUE. *Compute 95.4% confidence interval for the population proportion.*

Solution

$\hat{p} = \frac{53}{90} = 0.59$. *Since* $n\hat{p} = 53 \geq 15$ *and* $n(1 - \hat{p}) = 37 \geq 15$ *a normal approximation to the distribution of* $\hat{p}$ *is justified. Since* $1?\alpha = .954$, *we have* $\alpha/2 = .023$ *and* $z_{.023} = 2$. *Calculating* 0.59 ± 0.104 *we find the 95.4% confidence interval*

$$(0.486, 0.694)$$

Determining sample size

The required sample size is obtained by equating $z_{\alpha/2}\sqrt{p(1-p)/n} = d$, where d is the specified error margin. Thus

$$n = p(1-p)\left[\frac{z_{\alpha/2}}{d}\right]^2.$$

Formula

If the value of p is **known** *to be roughly in the neighborhood of a value p^*, then n can be determined from*

$$n = p^*(1-p^*)\left[\frac{z_{\alpha/2}}{d}\right]^2.$$

Without prior knowledge *of p, we can replace $p(1-p)$ by its maximum possible value $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$ and hence n determined from the relation*

$$n = \frac{1}{4}\left[\frac{z_{\alpha/2}}{d}\right]^2.$$

Example

A public health survey is to be designed to estimate the proportion p of a population having defective vision. How many persons should be examined if the public health commissioner wishes to be 98% certain that the error of estimation is below .05 when:
(a) There is no knowledge about the value of p?
(b) p is known to be about .3?

Example

A public health survey is to be designed to estimate the proportion p of a population having defective vision. How many persons should be examined if the public health commissioner wishes to be 98% certain that the error of estimation is below .05 when:
(a) There is no knowledge about the value of p?
(b) p is known to be about .3?

Solution

The tolerable error is $d = .05$. *Also* $1 - \alpha = .98$, *so* $\alpha/2 = .01$.
From the normal table, we know that $z_{.01} = 2.33$.
(a) Since p is unknown, the conservative bound on n yields
$n = \frac{1}{4}\left[\frac{2.33}{0.05}\right]^2 = 543$.
A sample of size 543 would suffice.

Example

A public health survey is to be designed to estimate the proportion p of a population having defective vision. How many persons should be examined if the public health commissioner wishes to be 98% certain that the error of estimation is below .05 when:
(a) There is no knowledge about the value of p?
(b) p is known to be about .3?

Solution

The tolerable error is $d = .05$*. Also* $1 - \alpha = .98$*, so* $\alpha/2 = .01$*.*
From the normal table, we know that $z_{.01} = 2.33$*.*
(a) Since p is unknown, the conservative bound on n yields
$n = \frac{1}{4}\left[\frac{2.33}{0.05}\right]^2 = 543.$
A sample of size 543 would suffice.
(b) If $p^* = .3$*, the required sample size is*
$n = 0.3 \times 0.7 \left[\frac{2.33}{0.05}\right]^2 = 456.$

Large sample test about p.

$H_0 : p = p_0$ versus $H_1 : p \neq p_0$.
With a large number of trials n, the sample proportion $\hat{p} = X/n$ is approximately normally distributed. Under the null hypothesis, p has the specified value p_0 and the distribution of $\hat{p}$ is approximately $N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$. Thus

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

has the $N(0,1)$ distribution.
Since the alternative hypothesis is two-sided, the rejection region of a level α test is given by $R : |Z| \geq z_{\alpha/2}$.
For one-sided alternatives, we use a one-tailed regions
$R : Z \geq z_\alpha$ to test $H_1 : p > p_0$ or
$R : Z \leq -z_\alpha$ to test $H_1 : p < p_0$.

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose Of the 90 respondents 53 said BLUE. *The manufacturer wants to learn if a majority of buyers would pick blue with the level of significance* $\alpha = 0.05$.

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose Of the 90 respondents 53 said BLUE. *The manufacturer wants to learn if a majority of buyers would pick blue with the level of significance* $\alpha = 0.05$.

Solution

1. $H_0 : p = 0.5$, $H_1 : p > 0.5$.

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose Of the 90 respondents 53 said BLUE. *The manufacturer wants to learn if a majority of buyers would pick blue with the level of significance* $\alpha = 0.05$.

Solution

1. $H_0 : p = 0.5,\ H_1 : p > 0.5.$
2. $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose Of the 90 respondents 53 said BLUE. *The manufacturer wants to learn if a majority of buyers would pick blue with the level of significance* $\alpha = 0.05$.

Solution

1. $H_0 : p = 0.5,\ H_1 : p > 0.5.$
2. $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
3. $z = \frac{0.59 - 0.5}{\sqrt{0.5(1-0.5)/90}} = 1.71.$

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose Of the 90 respondents 53 said BLUE. *The manufacturer wants to learn if a majority of buyers would pick blue with the level of significance* $\alpha = 0.05$.

Solution

1. $H_0 : p = 0.5,\ H_1 : p > 0.5.$
2. $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
3. $z = \frac{0.59 - 0.5}{\sqrt{0.5(1-0.5)/90}} = 1.71.$
4. *From the normal table*
   $p - value = P(Z > 1.71) = 1 - 0.9564 = 0.0436.$

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose Of the 90 respondents 53 said BLUE. *The manufacturer wants to learn if a majority of buyers would pick blue with the level of significance* $\alpha = 0.05$.

Solution

1. $H_0 : p = 0.5,\ H_1 : p > 0.5.$
2. $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
3. $z = \frac{0.59 - 0.5}{\sqrt{0.5(1-0.5)/90}} = 1.71.$
4. *From the normal table* $p - value = P(Z > 1.71) = 1 - 0.9564 = 0.0436.$
5. *Since* $0.0436 < 0.05$ *we* **reject** *the null hypothesis and conclude that our result is significant.*

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose Of the 90 respondents 53 said BLUE. *Is the claim that the buyers are not equally likely to choose red and blue cars justified with the level of significance* $\alpha = 0.05$*?*

Choosing car color

Example

In a marketing survey for an automobile manufacturer, 90 randomly selected adults are asked which car color they would choose Of the 90 respondents 53 said BLUE. *Is the claim that the buyers are not equally likely to choose red and blue cars justified with the level of significance* $\alpha = 0.05$*?*

Solution

1. $H_0 : p = 0.5,\ H_1 : p \neq 0.5.$
2. $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$
3. $z = \frac{0.59 - 0.5}{\sqrt{0.5(1-0.5)/90}} = 1.71.$
4. *From the normal table* $p - value = P(|Z| > 1.71) = P(Z > 1.71) + P(Z < -1.71) = 2 \times (1 - 0.9564) = 0.0872.$
5. *Since* $0.0872 > 0.05$ *we do not have enough evidence to reject the the null hypothesis.*

For Next Time

- Read Section 8.5 from Johnson and Bhattacharyya
- Online homework 8.5: 8.57, 8.61, 8.67, 8.71, 8.73.