Topic 9. Inference for Population Proportions

1

Learning outcomes

- Calculate the standard error of a proportion
- Calculate and interpret 95% confidence interval for the population proportion
- Use R to calculate 95% confidence intervals for the population proportion
- Estimate the sample size needed for a specific margin of error

2

General approach for interval estimation (confidence intervals)

- Add a margin of error to a numerical summary
- Numerical summary plus and minus the margin of error is the CI
- for a population proportion π a CI is given by

z-value for required confidence level

Point estimate $\rightarrow$ $p \pm \underbrace{z * se}_{\text{Margin of error}}$ $\leftarrow$ Standard error

3

Standard error for a proportion

- Recall the Binomial distribution is approximated by Normal for large *n*
- $X \sim Bin(n, \pi) \approx N(n\pi, n\pi(1-\pi))$
- $P = \frac{X}{n} \approx N(\pi, \frac{n\pi(1-\pi)}{n^2})$
- ie the s.e. of a proportion for large *n* is $\frac{\pi(1-\pi)}{n}$
- We don't know the true proportion so estimate the s.e. by plugging in the observed proportion

  p: estimated standard error (ese) = $\frac{p(1-p)}{n}$

4

Approximate 100(1-α)% Confidence Interval for a Binomial Proportion

If p is the proportion of observations resulting in a "success" in a random sample of size n, an approximate $100(1-\alpha)\%$ confidence interval for the population proportion π is

$$p \pm z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

where $z_{1-\alpha/2}$ is the 1-α/2 percentage point of the standard normal distribution.

5

Assumptions

$$\boldsymbol{np} \geq \mathbf{10}$$

$$\boldsymbol{n}(\mathbf{1}-\boldsymbol{p}) \geq \mathbf{10}$$

6

Example: Crankshaft Bearings

In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow.

Construct a 95% confidence interval for the population proportion.

7

Example: Crankshaft Bearings

- A point estimate of π, the proportion of bearings in the population that exceeds the roughness specification is

$$p = \frac{x}{n} = \frac{10}{85} = 0.1176$$

A 95% confidence interval for π is computed

$$p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$$

$$= \frac{10}{85} \pm 1.96\sqrt{\frac{10/85(1-10/85)}{85}} = (0.05, 0.19) \text{ to 2dp}$$

8

<u>Interpretation:</u>

It is likely that the proportion of bearings in the population that exceeds the roughness specification is between 0.05 and 0.19.

Any values outside of this interval are not supported by the data.

$$np \geq 10 \qquad n(1-p) \geq 10$$

np = 10; n(1-p) = 75
sample is just big enough for the normal approximation to be OK

9

```
> prop.test(x=10, n=85)

        1-sample proportions test with continuity correction

data:  10 out of 85, null probability 0.5
X-squared = 48.188, df = 1, p-value = 3.872e-12
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.06091797 0.21013931
sample estimates:
        p
0.1176471
```

10

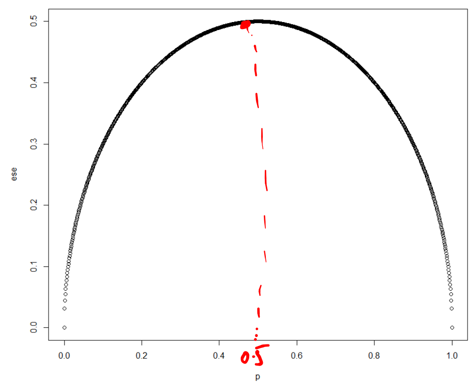
What value of p will result in the largest estimated standard error ?

$$p \pm z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

```
p <- seq(from=0, to=1, by=0.001)
ese <- sqrt(p*(1-p))
plot(p, ese)
```

11

Estimated standard error is largest when p=0.5



12

Estimated standard error is largest when p=0.5

- If a sample of size 1000 is taken what is the likely margin of error in the estimated proportion ?

13

Estimated standard error is largest when p=0.5

- If a sample of size 1000 is taken what is the likely margin of error in the estimated proportion ?

- Assume worst case – this is where *p=0.50*

95 % CI (approximately, using z = 2 instead of 1.96)

$$p \pm 2\sqrt{\frac{0.5(1-0.5)}{1000}} = p \pm 2(0.015) = p \pm 0.03$$

Margin of error is ~ 3% at worst when using a sample of size 1000

14

Sample size to estimate a proportion with specified precision

$$n = \frac{z^2_{1-\frac{\alpha}{2}} * \pi * (1-\pi)}{e\text{^}2}$$

Where:

- $z_{1-\alpha/2}$ is value from standard normal distribution corresponding to desired confidence level ($z_{1-\alpha/2}$ =1.96 or approximately 2 for 95% CI)
- π is expected true proportion
- e is desired precision (half desired CI width)

In the worst case scenario π=1/2 which results in:

$$n = \frac{z^2_{1-\frac{\alpha}{2}}}{4e\text{^}2}$$

15

Calculating (rough) 95% CI for a proportion quickly …

- Assume worst case (i.e. where π=*0.50*)

$$95\% \text{ CI } = p \pm 2\sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{n}}$$

$$= p \pm \frac{1}{2} 2\sqrt{\frac{1}{n}}$$

$$= p \pm \sqrt{\frac{1}{n}}$$

If n =1000 this works out as ± 0.03 (as expected) …..

16

Calculating (rough) 95% CI for a proportion quickly …

- A medical device company want to estimate the proportion of failures that will happen in the population of interest.
- They plan to take a sample of size 16.
- Estimate the (maximum) margin of error they are likely to get using this sample size.

17

Calculating (rough) 95% CI for a proportion quickly …

$$= p \pm \sqrt{\frac{1}{16}}$$
$$= p \pm 0.25$$

i.e. a very wide interval ….. You would hope the device does better than 0.05 successes ☺

18