STA13: Elementary Statistics

Lecture 15
Book Sections 5.4 - 5.5

Dmitriy Izyumin

February 16 2018

Sample Proportion $\hat{p}$

True proportion p

- Parameter
- Probability of success
- Proportion of successes in population

Sample proportion $\hat{p} = \frac{\#\text{ successes}}{n}$

- Pronounced "p hat"
- Number of successes in sample divided by n
- Statistic used to estimate p

Sampling Distribution of $\hat{p}$

How to determine the sampling distribution of $\hat{p}$:

1. In any case $\hat{p}$ has mean p and s.d. $\sqrt{\frac{p(1-p)}{n}}$.
2. Is the sample large ($n\hat{p} > 15$ and $n(1-\hat{p}) > 15$)?
   - YES! $\rightarrow$ $\hat{p}$ is approx. Normal $\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$
   - No... The normal approximation may not be reliable :(

The blue part of the second step is another application of the Central Limit Theorem.

Working with $\hat{p}$

1. Focus on a population with true proportion p.
2. Take a sample of size n: $x_1, x_2, \cdots, x_n$
3. Obtain the sample proportion $\hat{p} = \frac{\#\text{ successes}}{n}$
4. Find the distribution of $\hat{p}$:
   - $\hat{p}$ has mean p and s.d. $\sqrt{\frac{p(1-p)}{n}}$.
   - $n\hat{p} > 15$ and $n(1-\hat{p}) > 15$
     $\to$ $\hat{p}$ is approx. Normal$\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$

Estimating the Population Proportion

Population: Binomial population with proportion p.

Sample: Sample size n, sample proportion $\hat{p}$

- point estimator of p is $\hat{p}$
- If n is large, $(1-\alpha)100\%$ confidence interval for p is
$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
- margin of error is $z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Interpreting a Confidence Interval

We say:

We are ___1___% confident that the true proportion of ___2___ is between ___3___ and ___4___.

The blanks are:

1. Confidence level $100(1-\alpha)\%$
2. Parameter of interest (in terms of problem)
3. Lower bound of interval
4. Upper bound of interval

Example

A political analyst wishes to estimate the public's approval of the President's recent actions. He gathers a random sample of 1000 registered voters, and records their responses. It is found that 614 of the 1000 approve. Construct a 95% confidence interval for the proportion of of ALL registered voters that approve of the President's actions.

Example

Given: $\hat{p} = \frac{614}{1000} = 0.614,\ n = 1000$

Use this to construct the interval:

- $SE(\hat{p}) \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.614(0.386)}{1000}} = 0.015$
- Since $n\hat{p} = 614 > 15$ and $n(1-\hat{p}) = 386 > 15$, we know $\hat{p}$ is approximately normal (CLT)
- 95% confidence $\rightarrow \alpha = 1 - 0.95 = 0.05 \rightarrow \alpha/2 = 0.025$
- $z_{\alpha/2} = z_{0.025} = 1.96$

The confidence interval is: $0.614 \pm 1.96(0.015)$

Equivalently: $(0.585,\ 0.643)$

We are 95% confident that the proportion of all voters that approve of the President's actions is between 0.584 and 0.643.

Margin of Error and Sample Size

- mean, large sample: $z_{\alpha/2}\frac{s}{\sqrt{n}}$
- mean, small sample: $t_{\alpha/2}\frac{s}{\sqrt{n}}$ (using df$=n-1$)
- proportion, large sample: $z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- In any case, ME decreases if n increases.

Choosing Sample Size

Suppose we are estimating a parameter, and want to make a $100(1-\alpha)\%$ confidence interval.

Once α (or the confidence level) is chosen...

- The larger the sample, the smaller the ME.
- We can decide on the desired maximum size of ME beforehand.
- We can compute the sample size n that would guarantee that the ME is below our desired bound.

Example - Mean

Suppose it is known that the annual earnings of recent UCD graduates varies with standard deviation of \$4,600. If I want to estimate the mean annual earnings of recent UCD graduates to within \$1000 with 90% confidence, what is the smallest number of recent graduates that I need to sample?

- $\alpha = 0.10$
- Margin of error should be ≤ 1000
- Need to find minimum sample size to achieve that

Example - Mean

We want $\text{ME} \leq 1000$.

Equivalently, we want $z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 1000$.

Solve for n: $n \geq \frac{(z_{\alpha/2})^2 \sigma^2}{(1000)^2}$.

Use $\sigma = 4600$.

Use $z_{0.05} = 1.645$ since we want a 90% CI ($\alpha = 0.10$).

Plug in those values: $n \geq \frac{(1.645)^2 (4600)^2}{(1000)^2} = 57.259$.

I need a sample of size at least 58. (ALWAYS ROUND UP!)

Example - Proportion

A political analyst wants to estimate the true proportion of ALL registered voters that approve of the President's actions to within 0.01 with 95% confidence. How many registered voters need to be sampled to ensure this?

- $\alpha = 0.05$
- Margin of error should be ≤ 0.01
- Need to find minimum sample size to achieve that

Example - Proportion

We want $\text{ME} \leq 0.01$.

Equivalently, we want $z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.01$.

Solve for n: $n \geq \frac{(z_{\alpha/2})^2(\hat{p}(1-\hat{p}))}{(0.01)^2}$.

Use $\hat{p} = 0.5$ for the most conservative estimate.

Use $z_{0.025} = 1.96$ since we want a 95% CI ($\alpha = 0.05$).

Plug in those values: $n \geq \frac{(1.96)^2(0.25)}{(0.01)^2} = 9604$.

The analyst needs to sample at least 9604 registered voters.

Summary

Suppose ME is the desired size of the margin of error.
(So in the examples we had ME=1000 and ME=0.01)

- For proportion, $n \geq \frac{(z_{\alpha/2})^2(0.25)}{ME^2}$
- For mean, $n \geq \frac{(z_{\alpha/2})^2\sigma^2}{ME^2}$
- Plug in the given values, and simplify.
- When computing n, always round up.