Topic 7 Inference for the Proportion

Tutorial agenda:

1. go through key concepts
2. tutorial questions
3. Q&A

Key concept

Tuto11 Topic7

Sample Proportion

Sample proportion: $\hat{p} = \frac{X}{n}$

- X: no. of obs
- n: sample size
- Categorical variable - two levels
- Random sample

Follow binomial distribution:

- $X \sim B(n, p)$

Approximate the sampling distribution by a normal distribution.

$$\hat{p} \sim N(\mu_p, {\sigma_p}^2)$$

For large n, approximately normal distribution:

$$np \geq 5 \text{ and } n(1-p) \geq 5$$

(p cannot be too small or too large)

Topic 3: binomial distribution

- What is the probability that x out of n obs meet the criterion?

$$P(X=x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{(n-x)}$$

  - X - event of interest/meet the criterion
  - x - no. of event of interest
  - n - no. of obs
  - p - probability of an event of interest
- Summary measures
  - Center - Expected value/mean $\mu = np$
  - Variation
    - Standard deviation (s.d.) $\sigma = \sqrt{np(1-p)}$

Sampling Distribution of Sample Proportion

Topic 4: sampling distribution

Probability: sample ($\hat{p}$) $\xrightarrow[\text{estimate}]{}$ population (p)

Mean of *sample proportions*

$$\mu_{\hat{p}} = p$$

s.d. of *sample proportions*

$$\sigma_{\hat{p}} = \frac{p}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}} = \frac{\sqrt{np(1-p)}}{n}$$

Priority: population > sample:

- If pop p, is known, use p for
  $$\mu_{\hat{p}} = p \text{ and } \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$
- If pop p, is unknown, use sample to estimate: $\hat{p} \rightarrow p$
  $$\mu_{\hat{p}} = \hat{p} \text{ and } \sigma_{\hat{p}} = S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Mean: sample ($\mu_{\bar{X}}$) $\xrightarrow[\text{estimate}]{}$ population (μ)

Mean of *sample means*

$$\mu_{\bar{X}} = \mu$$

s.d. of *sample means* = standard error of the mean

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Priority: population > sample:

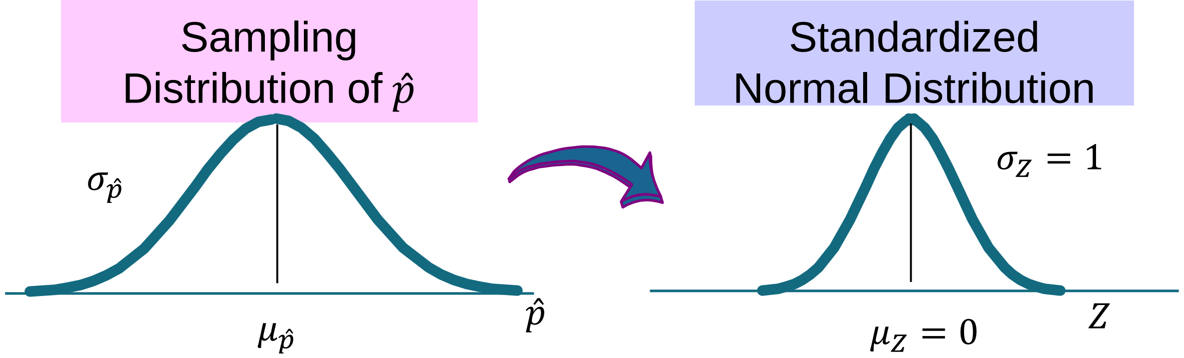
- If pop μ and σ is known, use
  $$\mu_{\bar{X}} = \mu \text{ and } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$
- If pop μ and σ is unknown, use sample to estimate:
  $\bar{X}$-> μ, S -> σ
  $$\mu_{\bar{X}} = \bar{X} \text{ and } \sigma_{\bar{X}} = S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

Sampling Distribution of Sample Proportion

Topic 3 & 4

Standardization (Z-distribution):
For sample proportion:

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



Standardization (Z-distribution):
For population:

$$Z = \frac{X - \mu}{\sigma}$$

For sample:

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Confidence Interval of Sample Proportion

Confidence intervals

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

Standard Error, $\sigma_{\hat{p}}$

Sampling Error, E

Topic 5

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Standard Error, $\sigma_{\bar{X}}$

Sampling Error, E
(Also called margin of error)

Critical Value

Special considerations

- If $\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < 0$, replace lower bound by **0**
- If $\hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} > 1$, replace upper bound by **1**

(because the range of any probability is [0,1])

Width:

- Level of confidence, $(1-\alpha)$
- Sample size, n
- Sample proportion, $\hat{p}$:
  - $\hat{p}$ increases from 0 to 0.5, then $\hat{p}(1-\hat{p})$ ↑, width ↑
  - $\hat{p}$ increases from 0.5 to 1, then $\hat{p}(1-\hat{p})$ ↓, width ↓

Confidence Interval of Sample Proportion

Sampling error (or margin of error):

$$\mathrm{E} = Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

=>

Sample size:

$$n = \frac{(Z_{\alpha/2})^2 p(1-p)}{E^2}$$

(always round up)

Standard deviation:

- p -> use $\hat{p}$ to replace p -> use 0.5 to estimate

(When p= 0.5, $p(1-p)$ becomes the largest)

Topic 5

$$E = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

=>

$$n = (\frac{Z_{\alpha/2}\sigma}{E})^2$$

(always round up)

Standard deviation:

- σ -> use S to replace σ -> use range/4 to estimate

Hypothesis Testing of Sample Proportion

The no. of successes, X, follows Binomial distribution
Normal approximation can be used:

$$n \geq 30$$
$$n\hat{p} \geq 5$$
$$n(1-\hat{p}) \geq 5$$

Test statistic: $Z = \frac{\hat{p}-\mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Reject / Do not reject H_0:

- Reject H_0 if $Z < C.V.$ or $Z > C.V.$ or p-value < α
- There is sufficient evidence that the true mean is not xxx

- Do not reject H_0 as p-value > α
- There is insufficient evidence that the true mean is not xxx

Topic 6

Test statistic: $Z = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

p-value: obtained from test statistic,

- two-tail: P(Z < -| Z |) + P(Z > | Z |)
- lower-tail: P(Z < Z)
- upper-tail: P(Z > Z)

Steps

Step 1: p (if p is not available, use $\hat{p}$ to estimate)

- Mean of *sample proportions:* $\mu_{\hat{p}} = p$
- s.d. of *sample proportions:* $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Step 2: check distribution:

- $n \geq 30$, $np \geq 5$ and $n(1-p) \geq 5$

→ The sampling distribution of $\hat{p}$ follows Normal distribution approximately, i.e.$\hat{p}\sim N(\mu_{\hat{p}}, {\sigma_{\hat{p}}}^2)$

Step 3: standardization i.e. Z distribution

- Probability: $P(\hat{p} < \mathrm{a})$=P($Z < \frac{a-\mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{a-p}{\sqrt{\frac{p(1-p)}{n}}}$)
- Confidence interval: $\hat{p} \ \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- Hypothesis testing: $Z = \frac{\hat{p}-\mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

Step 4: decision/explanation/summary:

- We are xx% confident that the population proportion of xx is between xx and xx
- Reject / Do not reject H_0 at α=xx. There is sufficient / insufficient evidence that the population proportion is not / less than / larger than xxx.

Confidence Interval - Z table

P(z<a)	Z=a
0.01	-2.33
0.025	-1.96
0.05	-1.645
0.1	1.28
0.5	0
0.9	-1.28
0.95	1.645
0.975	1.96
0.99	2.33

Confidence level	P(Z>$Z_{\alpha/2}$)	$Z_{\alpha/2}$
80%	0.1	1.28
90%	0.05	1.645
95%	0.025	1.96
99%	0.005	2.575