5 Foundations for Inference

Thursday, September 17, 2020 11:06 AM

Parameter Estimation
- Sample statistics used as point estimates
- best point estimate is the mean
- how sample statistics varies lets us estimate margin of error for the point estimate

Sample has hat Population has no hat

$\bar{X}, \hat{\sigma}^2, s^2, \hat{\sigma}, s, r_{xy}$ $\mu, \sigma^2, \sigma, \rho_{xy}$

Sample

Pop size = 4 Var X is age $X=\{18, 20, 22, 24\}$

$\mu = \sum X_i / N$

$= \frac{18+20+22+24}{4} = 21$

$\sigma = \sqrt{\sum (X_i-\mu)^2 / N} = 2.236$

Sampling Distribution: any statistic calculated on a sample has a certain sampling distribution

Central Limit Theorem

As sample size increases, the mean will become more normal

Rules:
1) Mean of sampling distribution is equal to the population distribution mean
$\mu = \bar{x}$
2) Variance of the sampling distribution is proportional to the variance of the population and inversely related to the sample size
$\sigma\bar{x} = \sigma x / \sqrt{n}$
3) as $n \uparrow$, distribution becomes more normal

Standard error $= s0/\sqrt{n}$ $\sigma = s0$

↳ as n increases, SE decreases

accuracy vs precision

The standard deviation of student weights for a sample of 100 students is 5.65 kgs. What is the standard error if the sample mean is 65.5 kg.

$SE = s0/\sqrt{n} = 5.65/\sqrt{100} = 5.65/10 = .565$

The distribution of the number of eggs laid by a certain species of hen during their breeding period is 35 eggs with a standard deviation of 18.2. Suppose a group of researchers randomly samples 45 hens of this species, counts the number of eggs laid during their breeding period, and records the sample mean. They repeat this 1,000 times and build a distribution of the sample means.
1. What is this distribution called?
2. Would you expect this distribution to be symmetric, right skewed or left skewed? Explain your answer.
3. Calculate the variability of this distribution and state the appropriate term used to refer to this value.
4. Suppose the researchers' budget is reduced and they are only able to collect random samples of 10 hens. The sample of mean of the number of eggs is recorded, and we repeat this 1,000 times, and build a new distribution of sample means. How will the variability of this new distribution compare to the variability of the original distribution?

1) Sampling distribution
2) Symmetric. As trials increase, the distribution becomes more normal → central limit theorem
3) $SE = s0/\sqrt{n} = 18.2/\sqrt{45} = 2.713$ ← standard error
4) The variability should increase because as n decreases, SE increases

Confidence Intervals

Begin w/ 95% CI

95% CI ≈ 2 SD

Point estimate $\pm z^* \cdot SE = CI$
↑
z critical

- You are interested in measuring the number of hours students spend watching TV on weekdays. You collected 5 different samples and each sample has 40 students.

Sample	Mean (μ)	SD (σ)
1	5	1.2
2	6	1.1
3	4	0.8
4	3	0.5
5	8	1.6

- Calculate the estimate for the average time students watch TV $(3+4+5+6+8)/5 = 5.2$
- You are not very sure about sample 5. Compute the 95% confidence interval for sample 5.

$CI = \bar{x} \pm z\, s/\sqrt{n}$
$= 8 \pm (1.96)(1.6/\sqrt{40})$
$= 8 \pm .496$