1. Confidence Level, Significance Level, Agresti-Coull's Correction

03 September 2024 08:36

Assignment
8, 11, 21

IMPORTANT TERMS

1. Confidence level $(1-\alpha)$
2. Confidence interval: Interval expected to contain the parameter being estimated
3. Significance level (α%): Probability the event could have occurred by chance / is insignificant
4. Right tail test: Alternate hypothesis suggests true value > null hypothesis
5. Left tail test: Alternate hypothesis suggests true value < null hypothesis
6. Hypothesis

CONFIDENCE LEVEL & SIGNIFICANCE LEVEL

A homeowner randomly samples 64 homes similar to her own and finds that the average selling price is \$250,000 with SD = \$15,000. Estimate a 99% confidence interval for the average selling price.

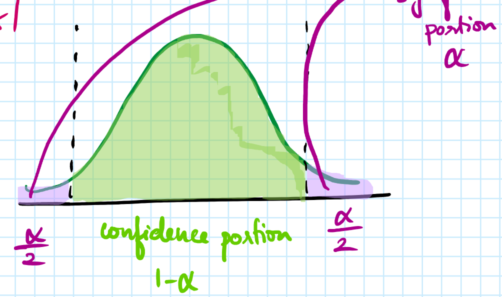
99% confidence %

0.99 confidence level

$1-\alpha = 0.99$

α = significance level = 0.01

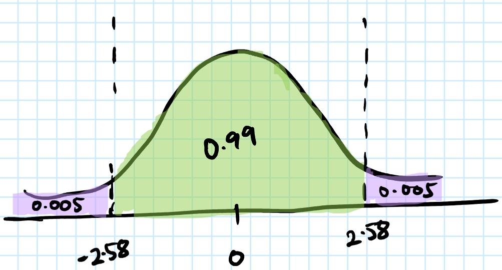
1% chance the occurrence of the event is insignificant / occurred by chance



w.k.t. $\bar{x} = \left(\frac{x_1 + x_2 + \dots + x_n}{n}\right) \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Here, $\bar{x} = \left(\frac{x_1 + x_2 + \dots + x_{64}}{64}\right) \sim N\left(250000, \frac{15000}{8}\right)$

Finding z-score,



$$\Rightarrow P\left(-2.58 \le \frac{\bar{x} - 250000}{15000/8} \le 2.58\right) = 0.99$$

$$P(-4837.5 \le \bar{x} - 250000 \le 4837.5) = 0.99$$

$$P(245162.5 \le \bar{x} \le 254837.5) = 0.99$$

$$P(245162.5 \le \bar{x} \le 254837.5) = 0.99$$

$\therefore$ We are 99% confident that

$$245162.5 \le \bar{x} \le 254837.5$$

CLT vs Confidence Interval

CLT	Confidence Interval
$\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$	$\mu \sim \text{Normal}\left(\bar{x}, \frac{s}{\sqrt{n}}\right)$
μ, σ given	$\bar{x}, s$ given
$P(\bar{x}$ in some interval$) = ?$	Possible range of $\mu = ?$

CI for $(1-\alpha)\%$ Confidence

$$\boxed{\bar{x} + Z'_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\alpha/2}\frac{s}{\sqrt{n}}}$$

> **NOTE: Standard error**
>
> SD in any estimator like $\bar{X}, S_n, \hat{p}$
>
> Std. error in estimating $\mu = \frac{s}{\sqrt{n}}$

A sample of size 40 is drawn from a normally distributed set of marks of students.

$\bar{x} = 35$, $s = 6$

Estimate 95% confidence interval for μ

$\alpha = 0.05$

$Z'_{\alpha/2} = -1.96$ $\quad\quad$ $Z_{\alpha/2} = 1.96$

95% CI :

$$\bar{x} + Z'_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Std. error $= \pm\frac{s}{\sqrt{n}} = 0.9486$

$$35 - 1.96\left(\frac{6}{\sqrt{40}}\right) \leq \mu \leq 35 + 1.96\left(\frac{6}{\sqrt{40}}\right)$$

$$33.14 \leq \mu \leq 36.859$$

250 randomly selected people are surveyed if they own a tablet. Out of these people, 98 own a tablet. Find the 95% confidence interval for proportion of people who own a tablet.

$$\hat{p} = \frac{x}{n} \qquad E(\hat{p}) = \frac{\mu}{n} = \frac{np}{n} = p \qquad V(\hat{p}) = \frac{npq}{n^2} = \frac{pq}{n} \qquad SD = \frac{\sqrt{pq}}{\sqrt{n}}$$

Std error $= \pm\frac{s}{\sqrt{n}} = \pm$

STUDENT'S t DISTRIBUTION

pdf: $$\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

Student proved that if sample size is small (<30), $\bar{x}$ follows t-dist. with $(n-1)$ df

cdf: $$\left[\frac{1}{2} + x\,\Gamma\left(\frac{v+1}{2}\right)\right]\left[\frac{{}_2F_1\left(\frac{1}{2}, \frac{v+1}{2}; \frac{3}{2}; -\frac{x^2}{v}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)}\right]$$

${}_2F_1(\,,\,;\,;\,) \longrightarrow$ hypergeometric function

v: Degrees of freedom

In a relation with n variables, n^{th} value gets fixed after assigning n-1 values randomly. Here we say we have n-1 degrees of freedom.

Confidence interval

$$\bar{x} - \left(t_{n-1,\alpha/2}\right)\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + \left(t_{n-1,\alpha/2}\right)\frac{s}{\sqrt{n}}$$

AGRESTI - COULL'S CORRECTION

Summary

Let X be the number of successes in n independent Bernoulli trials with success probability p, so that $X \sim \text{Bin}(n,\ p)$.

Define $\tilde{n} = n + 4$, and $\tilde{p} = \dfrac{X+2}{\tilde{n}}$. Then a level $100(1-\alpha)\%$ confidence interval for p is

$$\tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \qquad (5.5)$$

If the lower limit is less than 0, replace it with 0. If the upper limit is greater than 1, replace it with 1.

Two datasets are said to be paired when they come from the same observation unit