Module 8: Interval Estimation

- θ is fixed while θ_{hat_n} is a random variable which provides the single best value to estimate θ
- θ_{hat} is unbiased when bias = E($\theta_{hat\_n)\ -}$ θ = 0
- θ_{hat} is consistent when $\theta_{hat\_n\ ->}$ θ
- Mean squared error MSE = E($\theta_{hat\_n\ -}$ θ)^2 = bias(θ_{hat_n}) + V(θ_{hat_n})
- If bias -> and se -> as n -> infinity then θ_{hat_n} is consistent
- Probability is stronger than samples, probability standard error eventually converges to 0 as n approaches infinity but samples converge to a normal distribution which is not necessarily the same as the population distribution.
- **We the estimator variability (se) to provide an interval of parameter values that are "supported" by the sample.**

A 1 - α confidence interval for a parameter θ is an interval C_n = (a; b) where a = a(X_1, ...X_n) and b = b(X_1, ...X_n) are functions of the data such that: P(θ ∈ C_n) >= 1 - α   ; Where θ is the actual population mean. C_n is random and θ is fixed.

The confidence interval (a; b) capture the true mean with confidence 1 - α. We commonly use 95% confidence intervals which corresponds to α = .05. This does **NOT** mean there is 1- α chance/probability the parameter falls in the interval. The correct interpretation: If we repeatedly take samples of size n from a fixed and stable population and build a 95% confidence intervals, 95% of these intervals would contain the true unknown parameter.

CI For Mean of a Normal Distribution

If σ^2 is known: X_{bar} +/- $Z_{\alpha/2}$*α_x

If σ^2 is unknown: X_{bar} +/- $t_{(\alpha/2,n-1)}$*S/sqrt(n); Where S^2 = 1/(n-1) * sum(x_i-x_{bar})^2

Using S in place of SD causes more uncertainty, thus increasing the size of the CI.

We can similarly find the confidence interval of a proportion in a similar manner:

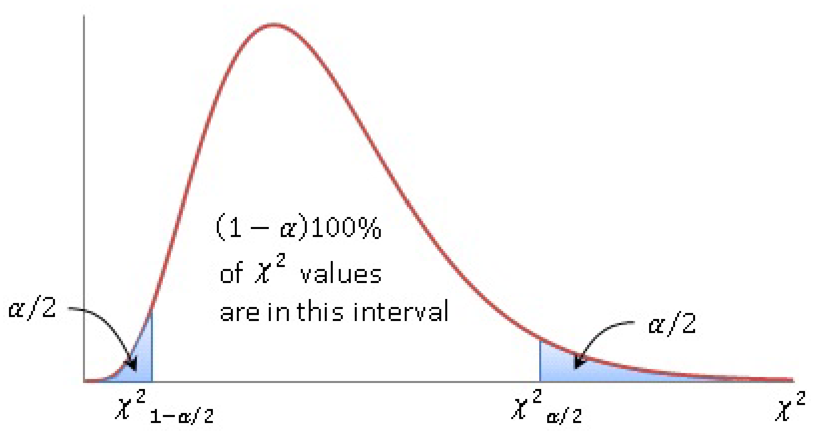
$$C_n = \left(\hat{p}_n - z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}, \hat{p}_n + z_{\alpha/2}\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right)$$

Chi-Square DIstribution

$$Q = \sum_{i=1}^{n} X_i^2$$

The above represents the chi-squared distribution with n degrees of freedom. E(Q) = n and V(Q) = 2n.

The distribution of X^2_{n-1} is not symmetrical, so instead of centering our CI (a,b) on the mean, we look for symmetry so that the bounds P(θ < a) = α/2 and P(θ > b) = α/2.



We derive the variance of a distribution through Fisher's theorem (not shown). The CI comes out to:

$$\mathbb{P}(\sigma^2 < a) = \alpha/2 \Rightarrow \mathbb{P}\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)S^2}{a}\right) = \alpha/2 \qquad (1)$$

$$\mathbb{P}\left(\frac{(n-1)S^2}{\sigma^2} > \chi^2_{(\alpha/2,n-1)}\right) = \alpha/2 \qquad (2)$$

- Thus $a = \frac{(n-1)S^2}{\chi^2_{(\alpha/2,n-1)}} \leq \sigma^2 \leq b = \frac{(n-1)S^2}{\chi^2_{(1-\alpha/2,n-1)}}$

Pearson's product Moment Correlation Coefficient is between -1 and 1 and represents the correlation between 2 variables.

$$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2\sum(y-\overline{y})^2}}$$

Although rarely used, you could find a confidence interval for this value.

$$Z = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right)$$

$$se(Z) = \frac{1}{\sqrt{n-3}}$$

$$\frac{\exp(2F)-1}{\exp(2F)+1} \text{ to } \frac{\exp(2G)-1}{\exp(2G)+1}$$

$$F = Z - \frac{z_{\alpha/2}}{\sqrt{n-3}} \text{ and } G = Z + \frac{z_{\alpha/2}}{\sqrt{n-3}}$$

Revision #7
Created 24 August 2022 14:06:18 by Elkip
Updated 24 August 2022 15:51:56 by Elkip