2.3 Interval Estimation pt2

Steven Cheung
23 February 2022

Theorem:

Let X be a binomial variable defined on n independent trials for which $p = P(success)$. For any a and b

$$lim_{n\to\infty} P(a \leq \frac{X-np}{\sqrt{np(1-p)}} \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{\frac{-z^2}{2}} dz$$

$$T(\frac{X}{n}, p) = \frac{\frac{X}{(n-p)}}{\sqrt{(\frac{X}{n})\frac{(1-\frac{X}{n})}{n}}}, \; T(\frac{X}{n}, p) \tilde{} N(0,1)$$

A *population proportion* is the proportion (percentage) of a population that has a specified attribute.

For X(the number of successes in n independent trials) from Binomial distribution, where n is large and $p = P(success)$ is unknown.

Let $\hat{p} = \frac{X}{n}$ be an estimate for p where X is the number of successes in n trials, then $Var(\frac{X}{n}) = \frac{p(1-p)}{n}$, depends on p

This means that, in general case, we cannot get exact confidence interval

$$[\frac{X}{n} - Z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}}, \frac{X}{n} + Z_{\alpha/2}\sqrt{\frac{(X/n)(1-(X/n))}{n}}]$$

but we can obtain an approximate $100(1-\alpha)\%$ confidence interval for p, if we note that $Var(\frac{X}{n})$ has maximum value that at the point $p = \frac{1}{2}$, then $\frac{Var(\frac{X}{n}) \leq 1}{4n}$ for any $0 < p < 1$.

Symmetric confidence interval can be represented in the form

point estimate ± margin of error

Definition:

The margin of error E for the estimate of μ is

$$E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

The margin of error for the estimate of a population mean indicates the accuracy with which a sample mean estimates the unknown population mean.

Definition:

margin of error for confidence level $100(1-\alpha)\%$ is $E = \frac{Z_{\alpha/2}}{2\sqrt{n}}$

If we can estimate that true value of p is greater than $\frac{1}{2}$ (or less that $\frac{1}{2}$). **margin of error** for confidence level $100(1-\alpha)\%$ is $E = \frac{Z_{\alpha/2}\sqrt{p_g(1-p_g)}}{\sqrt{n}}$

The minimal sample size required for estimation of the population mean μ at level $(1-\alpha)$ with margin of error E is given by

$$n = \frac{Z^2_{\alpha/2}\sigma^2}{E^2}$$

Similarly, a $(1-\alpha)$-level confidence interval for a population proportoin p, if p is unknown that has a margin of error of at most E can be obtained by choosing

$$n = \frac{Z^2_{\alpha/2}}{4E^2}$$

rounded up to the next intege.
Or, if we can make "educated guess" about the value of $p(p_g)$, then the smallest sample required is

$$n = \frac{Z^2_{\alpha/2}}{E^2}p_g(1-p_g)$$

Procedure to calculate large sample confidence intervals for θ

- Find an estimator (such as the MLE) of θ say $\hat{\theta}$;
- Obtain the standard error, $\sigma_{\hat{\theta}}$ of $\hat{\theta}$;
- Find the z-transform: $z = \frac{(\hat{\sigma}-\sigma)}{\sigma_{\hat{\theta}}}$. Then z has an approximately standard normal distribution;
- Using the normal table, find two tail values $-z_{\alpha/2}$ and $z_{\alpha/2}$;
- An approximate $(1-\alpha)100\%$ confidence interval for θ is

$$P(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}) = 1-\alpha$$

- Conclusion: We are $(1-\alpha)100\%$ confident that the true parameter θ lies in the interval

$$(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}})$$

When to Use the One-Mean z-Interval Procedure;

- For small samples - of size less than 15 - the z-interval procedure should be used only when the variable under consideration is normally distributed or very close to being so.
- For samples of moderate size - between 15 and 30 - the z-interval procedure can be used unless the data contain outliers or the variable under consideration is far from being normally distributed.
- For large samples - of size 30 or more - the z-interval procedure can be used essentially without restriction. However, if outliers are present and their removal is not justified, you should compare the confidence intervals obtained with and without the outliers to see what effect the outliers have. If the effect is substantial, use a different procedure or take another sample, if possible.
- If outliers are present but their removal is justified and results in a data set for which the z-interval procedure is appropriate (as previously stated), the procedure can be used.