Confidence Intervals for the Mean and Variance

Data Science for Quality Management: Sampling Distributions, Error and Estimation

with **Wendy Martin**

Learning objective:

Calculate interval estimates for the mean, variance and standard deviation

Calculating Confidence Intervals

- Confidence intervals may be calculated for various statistics
  - Mean
    - Sigma known
    - Sigma unknown
  - Standard Deviation / Variance

Means (Sigma Known)

- If the standard deviation is known, the following formula may be used

$$\mu_{CI} = \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Example

- For example, assume a sample was taken with the following characteristics

$$n = 150,\ \bar{X} = 20,\ s = 5$$

Confidence Level Desired = 95%

Example

- The 95% confidence interval (CI) for the mean is 19.2 to 20.8

$$\mu_{CI} = 20 \pm 1.96 \frac{5}{\sqrt{150}} = 20 \pm 0.80$$

Interval Estimate for the Mean (when σ is known)

In RStudio

```
> z.test.onesample
> z.test.onesample.simple
```

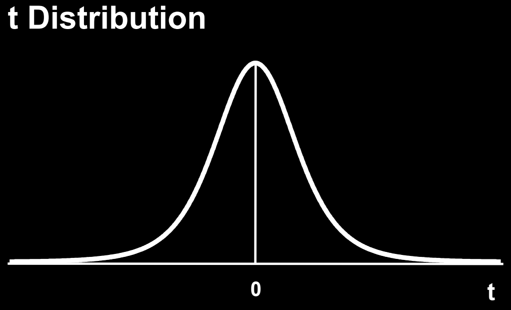
Means (Sigma Unknown)

- When the standard deviation is unknown, there is also uncertainty in the estimate of the standard error

- To account for this, instead of using the *z* distribution, we must use the *t* distribution

The *t* Distribution

- This distribution is commonly employed with small samples and an unknown population σ
- The *t* distribution is lower at the mean and higher at the tails than the normal distribution



The *t* Distribution

- The shape of the *t* distribution is dependent upon υ, or (*df*), the degrees of freedom. Degrees of freedom relate to “the number of values which can be freely chosen”

The *t* Distribution

- The *t* distribution considers the fact there is error associated with the use of *s*, the sample standard deviation to estimate the population (σ)

- This error increases the variability of the resulting statistic, the *t*, relative to a standard normal distribution

Means (Sigma Unknown)

- If the standard deviation is unknown, the following formula may be used

$$\mu_{CI} = \bar{X} \pm t_{\alpha/2,(n-1)df} \frac{s}{\sqrt{n}}$$

Example

- For example, assume a sample was taken with the following characteristics

$$n = 14,\ \bar{X} = 15{,}000,\ \ s = 500$$

Confidence Level desired = 90%

Example

- The df are equal to n - 1 or 13
- The t value corresponding to this is 1.771

$$\mu_{CI} = 15000 \pm 1.771\frac{500}{\sqrt{14}} = 15000 \pm 236.66$$

- The 90% (CI) for the mean is 14763.35 to 15236.65

Interval Estimate for the Mean (when σ is unknown)

In RStudio

```
> t.test.onesample
> t.test.onesample.simple
```

Standard Deviation / Variance

- The following formula may be used to generate a confidence interval for a standard deviation

$$\sqrt{\frac{s^2(n-1)}{\chi^2_{\alpha/2,(n-1)df}}} < \sigma < \sqrt{\frac{s^2(n-1)}{\chi^2_{1-\alpha/2,(n-1)df}}}$$

Standard Deviation / Variance

- This formula assumes that the population sampled from may be approximated by the normal distribution

Example

- For example, a process is studied for variability and a sample is drawn with the following characteristics

$$s = 10 \text{ and } n = 25$$

Confidence Level desired = 95%

Example

- The 95% CI for the standard deviation is 7.81 to 13.91

$$\sqrt{\frac{10^2(24)}{39.364}} < \sigma < \sqrt{\frac{10^2(24)}{12.401}}$$

Interval Estimate for the Variance / Standard Deviation

In RStudio

```
> variance.test.onesample
> variance.test.onesample.simple
```

Sources

- Luftig, J. An Introduction to Statistical Process Control & Capability. Luftig & Associates, Inc. Farmington Hills, MI, 1982