Interval Estimation

Underlying distribution is Normal—Central Limit theorem

Example: Suppose we draw a sample of children and measure Diastolic Blood Pressure (n=10)

$$\bar{x} = 58 \text{ mm Hg}$$

Not certain that μ = 58 exactly—second sample estimate might be $\bar{x}$ = 61 mm Hg

Assume DBP ~ N(μ, σ^2)

Therefore, if μ, σ^2 were known, and because

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

95% of all sample means would fall within

$$\left(\mu - 1.96\frac{\sigma}{\sqrt{n}}, \mu + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

Similarly, if we convert $\bar{x}$ to z where

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}, \text{ then } z \sim N(0, 1)$$

Therefore, 95% of the z values are between (-1.96, 1.96)

t distribution

Calculate $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$

Shape depends on n—not a unique distribution but a family indexed by a parameter called degrees of freedom

If $X_1, X_2, \ldots X_n \sim N(\mu, \sigma^2)$ are independent,

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Notation: $t_{d,u}$

Where d = degrees of freedom
u refers to some percentile, like 95th

Pr(t_d < $t_{d,u}$) □ *u*

For example: $t_{9, .95}$

Properties of *t* distribution

1. **Symmetric about 0**

2. **More spread out than N(0, 1)**

3. **For any percentile, □, □ > 0.5,**

$t_{d, 1-□}$ **is always larger than corresponding percentile for N(0, 1)—**$z_{1-□}$

4. **As *d* increases, *t* approaches N(0, 1)**

5. $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ **has more variability than**

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Using the *t*-table (Page 823)

$t_{10, 0.975} =$

$t_{20, 0.975} =$

$t_{40, 0.975} =$

$t_{120, 0.975} =$

$z_{0.975} =$

$t_{20, 0.025} =$

$t_{20, 0.05} =$

Confidence Intervals

t statistic follows a t distribution with n-1 degrees of freedom

$$\Pr(t_{n-1,\ .025} < t < t_{n-1,\ .975}) = 0.95$$

or

$$\Pr(t_{n-1,\frac{\alpha}{2}} < t < t_{n-1,1-\frac{\alpha}{2}}) = 1-\alpha$$

that is,

$$t_{n-1,\frac{\alpha}{2}} < \frac{\bar{x}-\mu}{s/\sqrt{n}} \quad \text{and} \quad \frac{\bar{x}-\mu}{s/\sqrt{n}} < t_{n-1,1-\frac{\alpha}{2}}$$

therefore,

$$\Pr(\bar{x} - t_{n-1,1-\frac{\alpha}{2}} * s/\sqrt{n} < \mu < \bar{x} + t_{n-1,1-\frac{\alpha}{2}} * s/\sqrt{n}) = 1-\alpha$$

therefore,

$$(\bar{x} - t_{n-1,1-\frac{\alpha}{2}} * s/\sqrt{n} \quad , \quad \bar{x} + t_{n-1,1-\frac{\alpha}{2}} * s/\sqrt{n}) = 100\%(1-\alpha)\ \text{CI}$$

or

$$(\bar{x} \pm t_{n-1,1-\frac{\alpha}{2}} * s/\sqrt{n}) = 100\%(1-\alpha)\ \text{CI}$$

Example: if n = 12, degrees of freedom = 11

$$95\%CI = (\bar{x} - 2.201 * s/\sqrt{12},\ \bar{x} + 2.201 * s/\sqrt{12})$$

Example: Estimate the average level of a certain enzyme in a human population. Take a sample of 10 individuals and determine the level of the enzyme in each.

$$\bar{x} = 22$$
$$s^2 = 45$$

$$
\begin{aligned}
95\%CI &= (22 - 2.262 * \frac{\sqrt{45}}{\sqrt{10}},\ 22 + 2.262 * \frac{\sqrt{45}}{\sqrt{10}}) \\
&= (22 - 2.262 * 2.1213,\ 22 + 2.262 * 2.1213) \\
&= 22 - 4.798,\ 22 + 4.798) \\
&= (17.202,\ 26.798)
\end{aligned}
$$

For □ known, use