Lecture 24, Mar 13, 2023

Other Confidence Intervals

- We may want a one-sided confidence interval $1 - \alpha = P(Z \leq z_\alpha)$
  - Same approach may be used to get $1 - \alpha = P\left(\mu \leq \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}}\right) = P(\mu \leq \bar{X}_U)$
- If the variance is unknown we have to use the t-distribution, thereby assuming the population is normal
  - Let $T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$ with sample variance $S = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
    - T has a t-distribution, $h(t)$
    - Let the CDF be $H(t) = \int_{-\infty}^{t} h(x)\,\mathrm{d}x$
  - Once again define $t_\beta > 0$ for $\beta < 0.5$ such that $H(-t_\beta) = \beta$, that is, the area under the PDF above t_β is β
  - $1 - \alpha = P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = P\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right)$
    - Therefore we find $\bar{X}_L = \bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X}_U = \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$
  - Note if we had known variance we end up getting a smaller confidence interval

Standard Error

- Notice in our confidence interval we used $\frac{\sigma}{\sqrt{n}}$; the size of the interval is proportional to it
- $\frac{\sigma}{\sqrt{n}}$ is referred to as the *standard error*, since in a way it tells us how much error there could be in our estimate of the mean

Prediction Intervals

- If we have samples $X_1, \cdots X_n$, normally distributed, what can we say about the next single sample X_0?
- $\bar{X}$ is a good point estimate, so what is the error $X_0 - \bar{X}$?
  - We know the distribution of this error is normal, with variance $\sigma^2 + \frac{\sigma^2}{n}$
- Let $Z = \frac{\bar{X}_0 - \bar{X}}{\sigma\sqrt{1 + \frac{1}{n}}}$, then Z has distribution $n(z; 0, 1)$
- Therefore $1 - \alpha = P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) \implies P\left(\bar{X} - z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}} \leq X_0 \leq \bar{X} + z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}\right)$
- This gives us our *prediction interval*

Definition

Given samples $X_1, \cdots, X_n$ normally distributed, the *prediction interval* is defined as

$$\left[\bar{x} - z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}, \bar{x} + z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}\right]$$

such that

$$1 - \alpha = P\left(\bar{X} - z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}} \leq X_0 \leq \bar{X} + z_{\alpha/2}\sigma\sqrt{1 + \frac{1}{n}}\right)$$

That is, there is a probability $1 - \alpha$ that the next single sample X_0 falls within this interval

- Prediction intervals can be useful for outlier detection