7.3 CIs based on a Normal Population Distribution:

If we have a small sample size, then we can't apply the CLT to use the results in the last section. Instead, here we assume the population has an (approximately) normal distribution.

Our random sample $X_1, \dots, X_n$ comes from a $N(\mu, \sigma^2)$ distribution with μ and σ^2 both unknown.

Again we start with $\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$. When n is small, this is not approximately normal. Instead it is a t distribution with $n-1$ degrees of freedom (df).

Let $T = \dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$. Then $T \sim t_{n-1}$.

Properties of the t distribution:

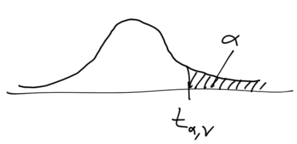
"nu"

Let ν be a positive integer. Let t_ν denote the t distribution with ν degrees of freedom.

1. Each t_ν curve is bell-shaped and centered at 0.
2. Each t_ν curve is more spread out than the standard normal z curve.
3. As ν increases, the spread of t_ν decreases.
4. As $\nu \to \infty$, the sequence of t_ν curves approach the standard normal curve. (z curve is "t with df $= \infty$")

Let $t_{\alpha,\nu}$ be the number for which the area under the t curve with ν degrees of freedom to the right of $t_{\alpha,\nu}$ is α.

$t_{\alpha,\nu}$ is a t critical value.



We can use qt in R (similar to qnorm).

ex: qt(0.95, 12) gives $t_{0.05,12}$.

The t CI:

Prop: Let $\bar{x}$ and s be the sample mean and sample SD for a random sample from a normal population with mean μ. The $100(1-\alpha)\%$ CI for μ is

$$\left(\bar{x} - t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}}, \ \bar{x} + t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}}\right).$$

Compactly: $\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}}$

Example 1: The weight of a certain brand of bread is approximately normally distributed (based on a QQ plot). In a sample of $n=20$ loaves, $\bar{x} = 17$, $s = 0.6$ in ounces. Find a 95% CI for μ, the true mean weight.

$$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}}$$

$$t_{\frac{0.05}{2}, 20-1} = qt(.975, 19) \approx 2.093$$

$$17 \pm 2.093 \cdot \frac{0.6}{\sqrt{20}}$$

$$\boxed{(16.72, 17.28) \text{ is a } 95\% \text{ CI for } \mu}$$

Prediction Interval for a Single Future Value:

Sometimes, instead of estimating the population mean, we want to make a prediction a single value of the variable.

ex: Confidence interval for μ: True mean of some aspect rocket launch

Prediction Interval: We are launching one rocket and want to predict just for that rocket.

Say $X_1, \ldots, X_n$ are a random sample from a normal population. We want to make a prediction for a single new value X_{n+1}.

The point predictor is $\bar{X}$, and the error is $\bar{X} - X_{n+1}$.

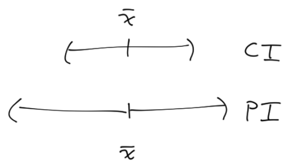
$$E[\bar{X} - X_{n+1}] = E[\bar{X}] - E[X_{n+1}] = \mu - \mu = 0$$

$$Var(\bar{X} - X_{n+1}) = Var(\bar{X}) + (-1)^2 Var(X_{n+1}) = \frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(1 + \frac{1}{n}\right)$$

So $\dfrac{\bar{X} - X_{n+1}}{\sqrt{\sigma^2(1+\frac{1}{n})}} \sim N(0,1)$ and $\dfrac{\bar{X} - X_{n+1}}{\sqrt{s^2(1+\frac{1}{n})}} \sim t_{n-1}$

Prop: A prediction interval (PI) for a single observation from a normal distribution is

$$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \cdot s \cdot \sqrt{1 + \frac{1}{n}}$$



The prediction level is $100(1-\alpha)\%$.

Example 2: We have a random sample of the lifetimes of 15 lightbulbs.

We got $\bar{x} = 210$ days with $s = 14$ days.

a) Find a 95% CI for μ, the true average lifetime of the lightbulb model.

b) Find a 95% PI for the lifetime of a single lightbulb.

a) $\bar{x} \pm t_{\frac{\alpha}{n}, n-1} \cdot \frac{s}{\sqrt{n}}$. $\quad t_{\frac{\alpha}{2}, n-1} = qt(0.975, 14) \approx 2.145$

$$210 \pm 2.145 \cdot \frac{14}{\sqrt{15}}$$

$\boxed{(202.2,\ 217.8) \text{ is a 95\% CI for } \mu}$

b) $\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \cdot s \cdot \sqrt{1 + \frac{1}{n}}$

$$210 \pm 2.145 \cdot 14 \cdot \sqrt{1 + \frac{1}{15}}$$

$\boxed{(179.0,\ 241.0) \text{ is a 95\% PI for the lifetime of 1 light bulb}}$

7.4 CIs for the Variance and SD of a Normal Population :

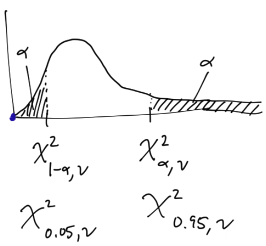
We can also find CIs for σ^2 or σ for a normal distribution.

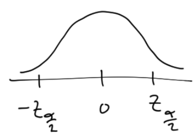
Th^m: Let $X_1, \dots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\Sigma(X_i - \overline{X})^2}{\sigma^2}$$

has a chi-squared (χ^2) distribution with n-1 degrees of freedom.

Let $\chi^2_{\alpha,\nu}$ be the chi-squared critical value, the number such that α of the area under the χ^2 curve with ν df lies to the right of $\chi^2_{\alpha,\nu}$.





The Theorem above tells us

$$P\left(\chi^2_{1-\frac{\alpha}{2},\nu} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha,\nu}\right) = 1-\alpha$$

A $100(1-\alpha)\%$ CI for the variance σ^2 of a normal population is

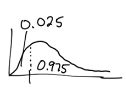
$$\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$$

A $100(1-\alpha)\%$ CI for the SD σ of a normal population is

$$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}}\right)$$

Example 1: A random sample of the breakdown voltage of 20 circuits was found to be approximately normal with $s = 230$.

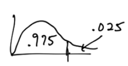
Find a 95% CI for σ^2 and σ.



$$\left(\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}\right)$$

$\chi^2_{1-\frac{\alpha}{2},n-1} = \text{qchisq}(0.025, 19) \approx 8.907$

$\chi^2_{\frac{\alpha}{2},n-1} = \text{qchisq}(0.975, 19) \approx 32.852$



$$\left(\frac{19 \cdot 230^2}{32.852}, \frac{19 \cdot 230^2}{8.907}\right) = \boxed{(30595, 112844) \text{ is a 95\% CI for } \sigma^2}$$

$$\left(\sqrt{\frac{19 \cdot 230^2}{32.852}}, \sqrt{\frac{19 \cdot 230^2}{8.907}}\right) = \boxed{(174.9, 335.9) \text{ is a 95\% CI for } \sigma.}$$