§7.3: Small Sample Normal Population Confidence Intervals

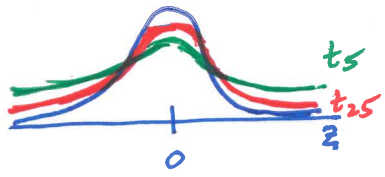
If n is small then we can't use $S^2 \approx \sigma^2$...

Instead we use another distribution!

Thm: $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$ ← "t distribution with $n-1$ degrees of freedom"

Historical Remark: t-distribution has full name "Student's t-distribution". It was developed by William Gosset in 1908. Gosset worked at Guinness Brewery in Dublin Ireland. He developed t-distribution trying to do statistics analyzing quality & effects of different types of barley when making beer. Published his work with pseudonym "Student" to hide identity & keep application to beer making secret.

Remark: If $n > 40$ then $t_{n-1} \approx \text{Normal}(0, 1)$ but for smaller n it has bigger tails



Formula: If $\bar{X}$ is Normal, then

$$\mu = \bar{x} \pm S/\sqrt{n} \cdot qt(\alpha/2, n-1)$$

People often write $t_{\alpha/2, n-1} = -qt(\alpha/2, n-1)$

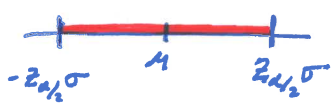
$$\Rightarrow P(T > t_{\alpha/2, n-1}) = \alpha/2$$

Side note on Prediction Intervals (PI):

- If $\bar{X} \sim \text{Normal}(\mu, \sigma)$ then an observation x will be in interval

$$x = \mu \pm z_{\alpha/2}\sigma$$

with probability $(1-\alpha)$



- What if we use $\bar{x}$ instead of μ?

Two sources of variance → from $\bar{x}$. → from x.

$$(\mu - \bar{X}) \sim \text{Normal}(0, \sigma)$$

$$(\bar{X} - X) \sim \; ??$$

$(\bar{X} - X)$ is difference of two Normals, so it is Normal too...

$$E[\bar{X} - X] = E[\bar{X}] - E[X]$$

$$= \mu - \mu = 0$$

$$Var[\bar{X} - X] = Var[\bar{X}] + Var[X]$$

$$= \sigma^2/n + \sigma^2$$

↳ std. dev. is $\sigma \cdot \sqrt{1 + 1/n}$

$$\Rightarrow \frac{\bar{X} - X}{\sigma\sqrt{1+1/n}} \sim \text{Normal}(0, 1)$$

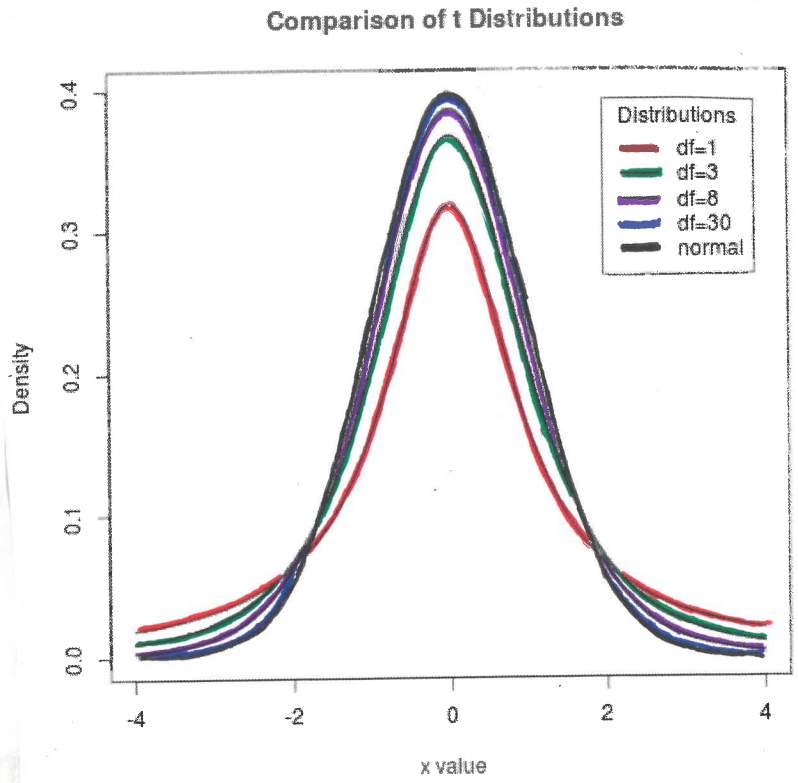
If we use S^2 instead of σ^2 then we should probably also move to t-distribution...

$$\frac{\bar{X} - X}{S\sqrt{1+1/n}} \sim t_{n-1}$$

$$\Downarrow$$

$$x = \bar{x} \pm t_{\alpha/2,\, n-1} \cdot S\sqrt{1 + 1/n} \qquad (1-\alpha)\ CI$$

$-qt(\alpha/2, n-1)$ $\qquad P(T > t_{\alpha/2, n-1}) = \alpha/2$




```
# Display the Student's t distributions with various
# degrees of freedom and compare to the normal distribution

x <- seq(-4, 4, length=100)
hx <- dnorm(x)

degf <- c(1, 3, 8, 30)
colors <- c("red", "blue", "darkgreen", "gold", "black")
labels <- c("df=1", "df=3", "df=8", "df=30", "normal")

plot(x, hx, type="l", lty=2, xlab="x value",
  ylab="Density", main="Comparison of t Distributions")

for (i in 1:4){
  lines(x, dt(x,degf[i]), lwd=2, col=colors[i])
}

legend("topright", inset=.05, title="Distributions",
  labels, lwd=2, lty=c(1, 1, 1, 1, 2), col=colors)
```

Code from statmethods.net
Advanced Graphs / Probability Plots