M358K: November 18th, 2020.

Inference for numerical data

So far: • Normal population distribution w/ an unknown mean μ & a known standard deviation σ

Simple random sample $X_1, X_2, \dots, X_n$ independent and Normal(mean $= \mu$, sd $= \sigma$)

Set $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ ... the sample mean

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

New: Q: What if σ is not known?

→: Idea: Use the sample std deviation: S

w/ $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

You want to use the following statistic:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$ NOT NORMALLY DIST'D.

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(df = n-1)$$

Random Variable

t-distribution (aka the Student dist'n)

Def'n. With $Y \sim \chi^2(df = n)$ and $Z \sim N(0,1)$ independent

we define $T = \dfrac{Z}{\sqrt{Y/n}}$.

T is said to have the t distribution (aka the Student dist'n) w/ n degrees of freedom.

For a normal population distribution:

- $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- $Y = \dfrac{S^2(n-1)}{\sigma^2} \sim \chi^2(df = n-1)$

independent

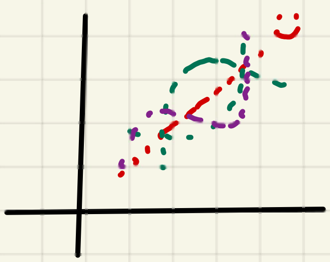
By def'n: $T = \dfrac{Z}{\sqrt{\dfrac{Y}{n-1}}} \sim t\,(df = n-1)$

$$T = \frac{\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{S^2}{\sigma^2}}} = \frac{\dfrac{\bar{X} - \mu}{1\cancel{\sigma}/\sqrt{n}}}{\dfrac{S}{\cancel{\sigma}}} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim \boxed{t\,(df = n-1)}$$

Note : It's essential that the population dist'n be normal for the small sample t-procedures!

Q: How would you check for normality?

- plot the histogram
- box plot
- Q-Q plot (q-q plot) : plot the quantiles of the standard normal against the quantiles of the data in "standard units"



Confidence intervals for μ

(small normal sample w/ unknown σ)

C ... confidence level

point estimate ± margin of error

$$\bar{x} \pm t^*(df = n-1) \cdot \frac{s}{\sqrt{n}}$$

Example. [Ramachandran · Tsokos : available as an ebook in the UT Library]

The scores of a random sample of 16 people had a sample mean of 540 and a sample std deviation of 50.

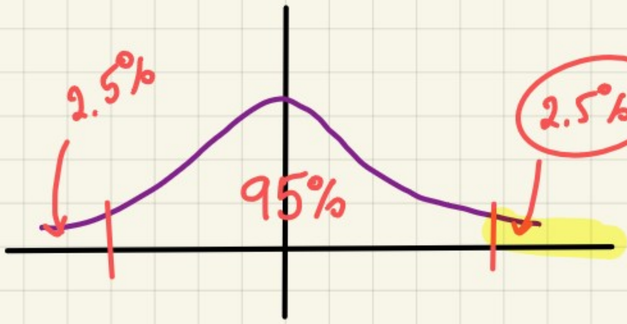
Construct a 95% confidence interval for the population mean μ of the score on the exam assuming that the scores are normal.

→ : $\bar{x} = 540$;

$s = 50$;

$t^* = ?$

$df = 16 - 1 = 15$



$t^* = 2.131$

$$\mu = \bar{x} \pm t^*(df = n-1) \cdot \frac{s}{\sqrt{n}}$$

$$\mu = 540 \pm 2.131 \cdot \frac{50}{\sqrt{16}} = 540 \pm 26.6375$$

C.I. $(513.36, 566.64)$

$$\boxed{513.36 < \mu < 566.64}$$