Stat 145, Fri 22-Oct-2021 -- Fri 22-Oct-2021
Biostatistics
Spring 2021

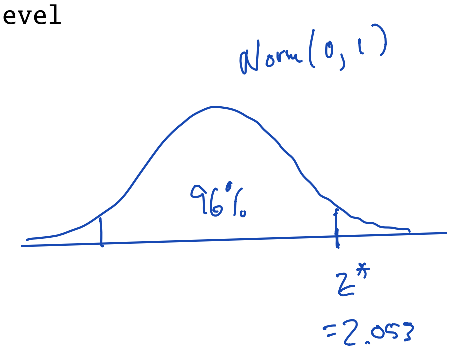
Friday, October 22nd 2021

Due::   PS11 Ellenberg due at 11 pm

Friday, October 22nd 2021

Wk 8, Fr
Topic:: Student t distributions
Read:: Lock5 6.4-6.6

Loose ends from last time
- Last time:
    - z*-critical value for 96% confidence was ...
      main use is for 96% CI construction
    - Did hypothesis test and got a standardized test statistic 2.163
      2.163 is out in the "tail", beyond z*-critical value
      Can know result is statistically significant at 4% level
- continuity correction
    - what it is
    - prop.test() command



Rock-Paper scissors example: $51/119 = \hat{p}$

$H_0: p = 1/3 \qquad H_a: p \neq 1/3$

We had found z^* (critical value) for 96% conf.: 2.054

We standardized $\hat{p}$:

$$z = \frac{(51/119) - 1/3}{\sqrt{\frac{(1/3)(1-1/3)}{119}}} = 2.163$$

corresponds to P-value 0.0305

Ex.] You have a sample of some quant. variable:

BMI

Sample size is $n = 25$ from our population (H.S. students)

Know: $\sigma = 8.2$, $\bar{x} = 23.9$ — point est.

Want: CI for μ. (at 95% level) — use $z^* = 1.96$

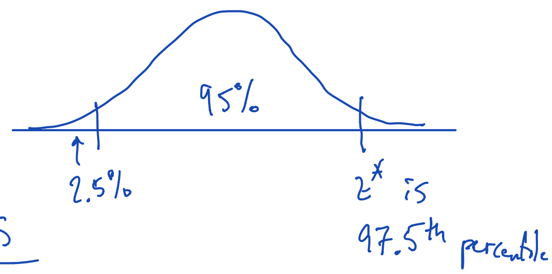
$$23.9 \pm (1.96)\left(\frac{8.2}{\sqrt{25}}\right)$$

lower = 20.69

upper = 27.11

But, unrealistic to act as if I know σ.

Don't know μ — estimate using $\bar{x}$



Is realistic: estimate σ by s

—— Now estimate $SE_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

Using s for σ triggers using t^* (a critical value from a t-distribution) in place of z^*.

before: obtained z^* for 95 conf. qnorm(0.975)

Now: obtain t^* for 95 conf. qt(0.975, df = ____).

$n-1$, one less than the sample size

More realistic version of previous example

Ex.) You have a sample of some quant. variable:

BMI

Sample size is $n = 25$ from our population (H.S. students)

Know: $s = 7.9$ ~~$\sigma = 8.2$~~, $\bar{x} = 23.9$ point est.

Want: CI for μ. (at 95% level) use ~~$z^* = 1.96$~~ $t^* = 2.064$

CI:

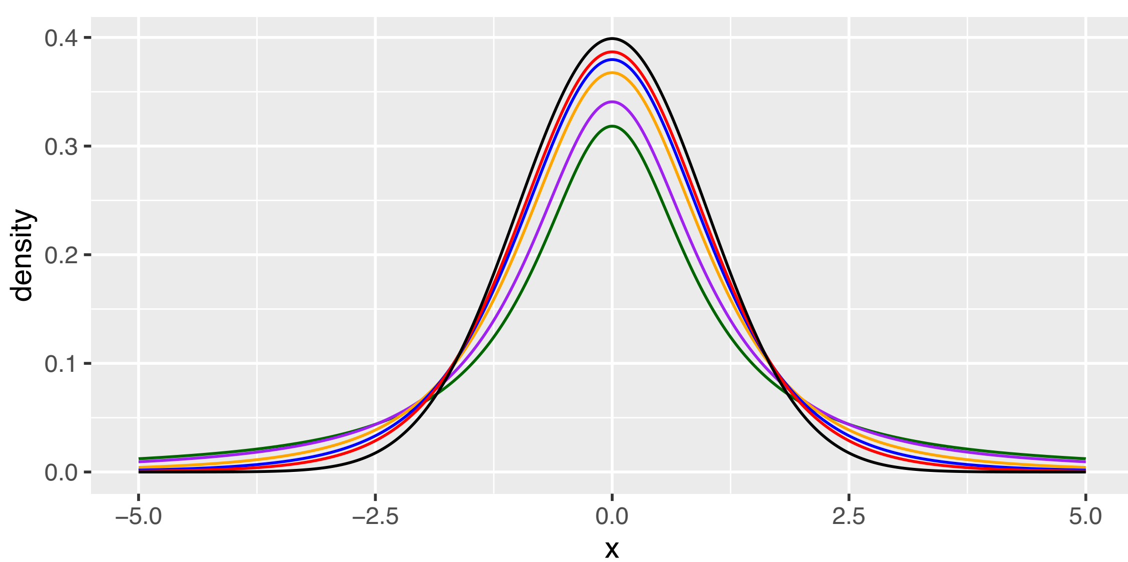
$$23.9 \pm (2.064) \frac{7.9}{\sqrt{25}}$$

Student t-distributions

Plotted below are the

- t-distribution with $df = 1$ (dark green)
- t-distribution with $df = 1.5$ (purple)
- t-distribution with $df = 3$ (orange)
- t-distribution with $df = 5$ (blue)
- t-distribution with $df = 8$ (red)
- standard normal distribution $\text{Norm}(0, 1)$ (black)

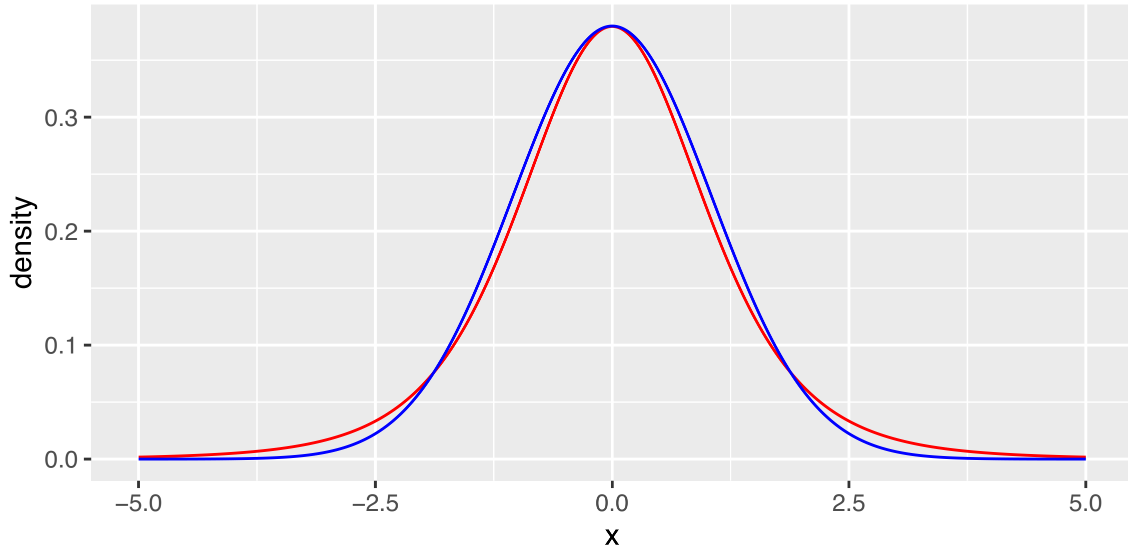
```
gf_dist("t", df=1, xlim=c(-5,5), color="darkgreen") %>%
  gf_dist("t", df=1.5, xlim=c(-5,5), color="purple") %>%
  gf_dist("t", df=3, xlim=c(-5,5), color="orange") %>%
  gf_dist("t", df=5, xlim=c(-5,5), color="blue") %>%
  gf_dist("t", df=8, xlim=c(-5,5), color="red") %>%
  gf_dist("norm", xlim=c(-5,5), color="black")
```



Next is a

- t-distribution with $df = 5$ (red)
- normal distribution $\text{Norm}(0, 1.05)$ (blue)

```
gf_dist("t", df=5 , xlim=c(-5,5), color="red") %>%
  gf_dist("norm", params=list(mean=0,sd=1.05), xlim=c(-5,5), color="blue")
```




```
A new distributional family: t-distributions
 - symmetric, bell(?)-shaped

 - centered on 0
   just a single parameter: degrees of freedom (df)
   Note: These are NOT just normal distributions with mean=0

 - df must be positive, but can be noninteger

 - increasingly like standard normal as df rises
   peak isn't quite so high
   more area in the tails

 - have (on strength of CLT) been
   of mind that sampling/bstrap/null/randomization distributions are normal
   using z* critical values (std. normal dist) corresp to level of confidence

   Q: What if we thought a t-dist with df=15 served as better model?

   1. Say we have x-bar as 15.87, and estimated SE=2.15.
      If we want to use t-dist model with df=11, what is
      a 95% confidence interval for mu?

      Practice obtaining critical t*-values for other levels of confidence

   2. Say we have x-bar as 15.87, and estimated SE=2.35.
```

For
95% Conf., $t^* = qt(0.975, df = \_\_)$
90% Conf., $t = qt(0.9 , df = \_\_)$

```
If we want to use t-dist model with df=11, what is
the strength of evidence against the null hypothesis when
   H_0: mu = 20,  H_a: mu < 20

What if, instead, df=38?
```