Stat 145, Thu 25-Mar-2021 -- Thu 25-Mar-2021
Biostatistics
Spring 2021

Thursday, March 25th 2021

Wk 8, Th
Topic:: Student t distributions
Read:: Lock5 6.4 – 6.6

Comparing P-values with alpha is akin to comparing standardized score with critical z*

modify $H_a: p < 0.25$

Ex.) $H_0: p = 0.25$, $H_a: p \neq 0.25$ $\quad \mu \quad \sigma$

Sample of size $n = 150$, $\hat{p} = \dfrac{28}{150} \sim \text{Norm}(0.25, 0.0354)$

(a) What is the standardized test statistic?

$$\text{est. } SE_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{(0.25)(0.75)}{150}} = 0.0354$$

$$z = \frac{0.187 - 0.25}{0.0354} \doteq -1.78$$

not using continuity correction: $28/150 = 0.187$

using continuity: $\dfrac{28.5}{150}$

(b) What is the z-critical value for significance level $\alpha = 0.05$?

$$z^* = 1.96$$

Variant

positive critical val.	1.96
negative " "	-1.96

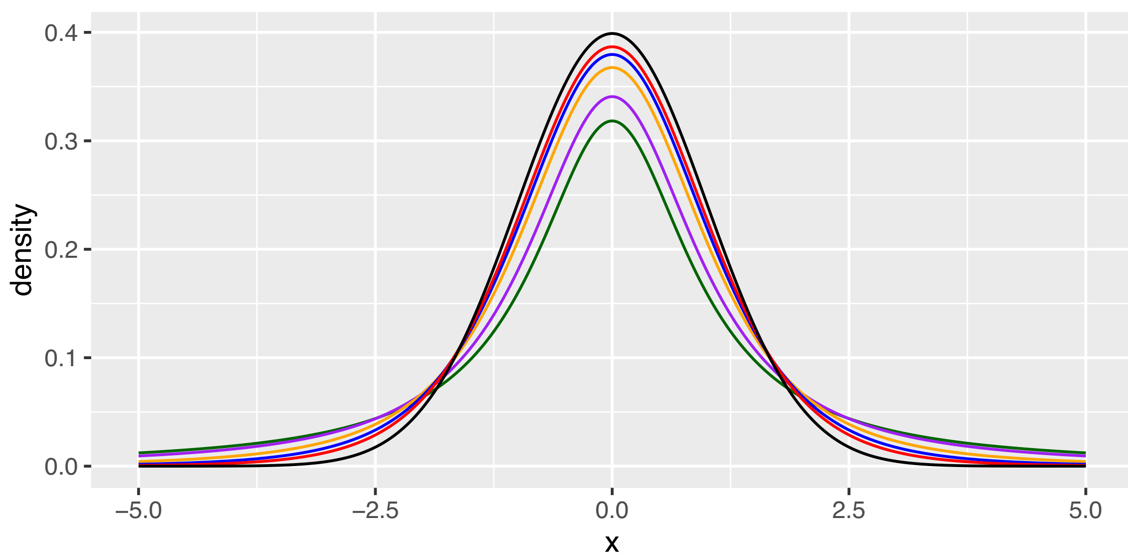
reject if in a left-tail that comprises area 0.05

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 $\mathrm{Norm}(0, 1)$ (black)

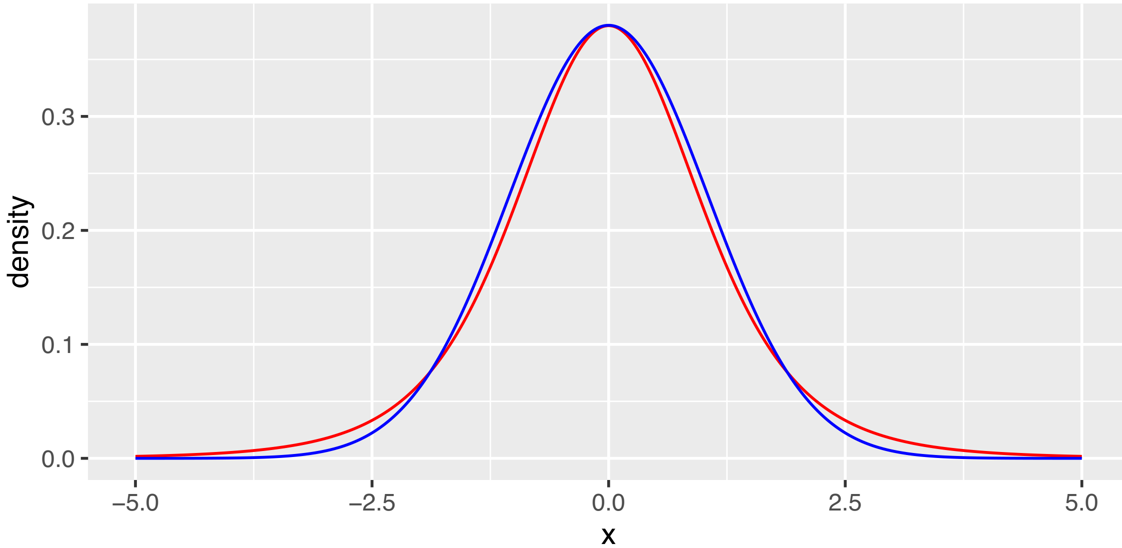
```
gf_dist("t", df=1, xlim=c(-5,5), color="darkgreen") %>%
  gf_dist("t", df=1.5, xlim=c(-5,5), color="purple") %>%
  gf_dist("t", df=3, xlim=c(-5,5), color="orange") %>%
  gf_dist("t", df=5, xlim=c(-5,5), color="blue") %>%
  gf_dist("t", df=8, xlim=c(-5,5), color="red") %>%
  gf_dist("norm", xlim=c(-5,5),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



Next is a

- t-distribution with $df = 5$ (red)
- normal distribution $\mathrm{Norm}(0, 1.05)$ (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?
```

$n \geq 30$

1. Say we have x-bar as 15.87, and estimated SE=2.15.

$\bar{x} = 15.87$ $SE_{\bar{x}} = 2.15$

If we want to use t-dist model with df=11, what is a 95% confidence interval for mu?

$t^* = qt(0.975, df=11)$

$\bar{x} \pm ME$ or $15.87 \pm (2.201)(2.15)$

$\bar{x}$: pt. est

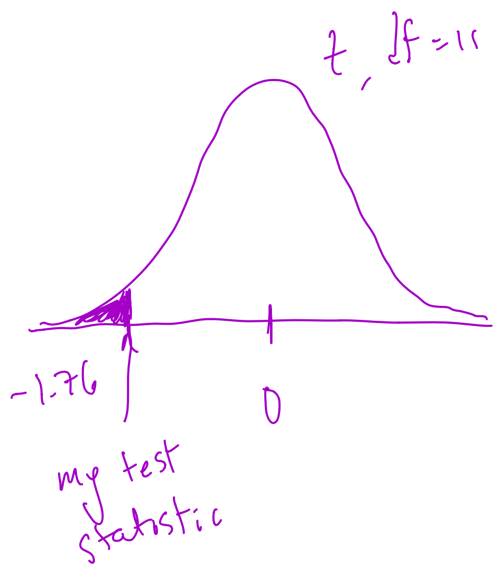
Practice obtaining critical t*-values for other levels of confidence

$t^* = ?$ if 90% CI w/ df=11

$qt(0.89, df=51)$ 78% CI w/ df=51

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

$$t = \frac{15.87 - 20}{2.35} = -1.76$$



What if, instead, df=38?