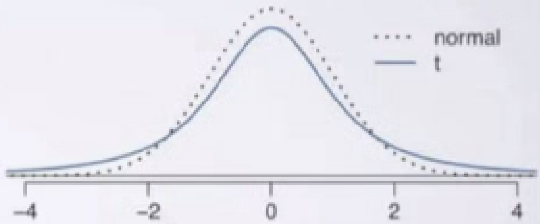
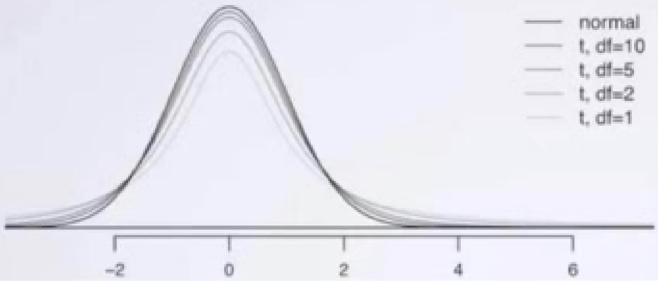
t-distribution

- when σ unknown (almost always), use the t-distribution to address the uncertainty of the standard error estimate
- bell shaped but thicker tails than the normal
  - observations more likely to fall beyond 2 SDs from the mean
  - extra thick tails helpful for mitigating the effect of a less reliable estimate for the standard error of the sampling distribution



- always centered at 0 (like the standard normal)
- has one parameter: degrees of freedom (df) - determines thickness of tails
  - remember, the normal distribution has two parameters: mean and SD



What happens to the shape of the t-distribution as degrees of freedom increases?

approaches the normal dist.

example:

Estimate the average after-lunch snack consumption (in grams) of people who eat lunch **distracted** using a 95% confidence interval.

$\bar{x} = 52.1\ g$

$s = 45.1\ g$

$n = 22$

$t^\star_{21} = 2.08$

$$\bar{x} \pm t^* SE = 52.1 \pm 2.08 \times \frac{45.1}{\sqrt{22}}$$

$$= 52.1 \pm 2.08 \times 9.62$$

$$= 52.1 \pm 20 = (32.1, 72.1)$$

We are 95% confident that distracted eaters consume between 32.1 to 72.1 grams of snacks post-meal.

code:

```
pt (t, df, lower.tail = false)
```

example 2:

$$H_0: \mu = 30 \qquad t = \frac{\bar{x} - \mu}{SE} = 2.3$$

$$H_A: \mu \neq 30 \qquad df = 22 - 1 = 21$$

p-value ≈ 0.0318

Reject H_0

Estimating the difference between independent means

$$\bar{x}_1 - \bar{x}_2 \pm t^*_{df} \cdot SE_{x_1 - x_2}$$

$\min(n_1 - 1, n_2 - 2)$

$$\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Conditions for inference for comparing two independent means:

1. *Independence:*
   - ✓ **within groups:** sampled observations must be independent
     - random sample/assignment
     - if sampling without replacement, n < 10% of population
   - ✓ **between groups:** the two groups must be independent of each other (non-paired)
2. *Sample size/skew:* The more skew in the population distributions, the higher the sample size needed.

Last modified: Jul 23, 2020