t distribution

- review: conditions for inference so far
- large → small n
- introduce the t-distribution

Duke UNIVERSITY

Dr. Mine Çetinkaya-Rundel
Duke University

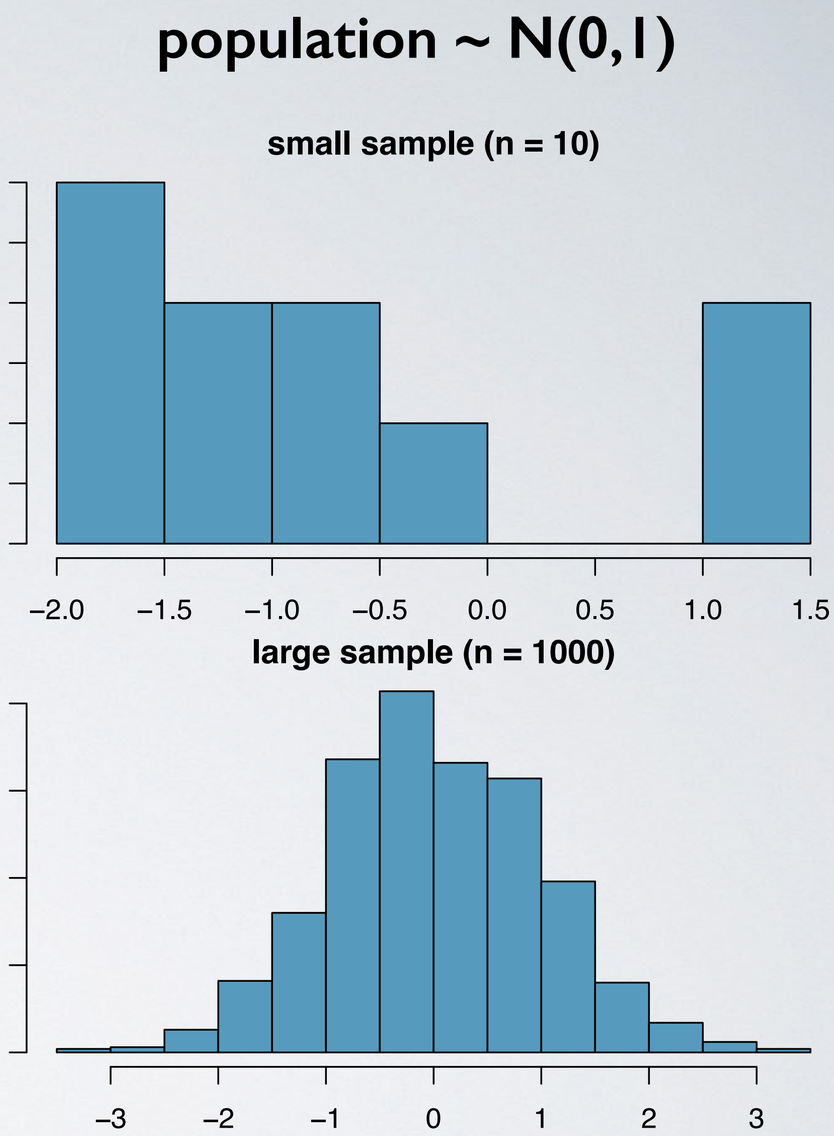
review:
what purpose does a large sample serve?

As long as observations are independent, and the population distribution is not extremely skewed, a large sample would ensure that...

- the sampling distribution of the mean is nearly normal
- the estimate of the standard error is reliable: $\frac{s}{\sqrt{n}}$

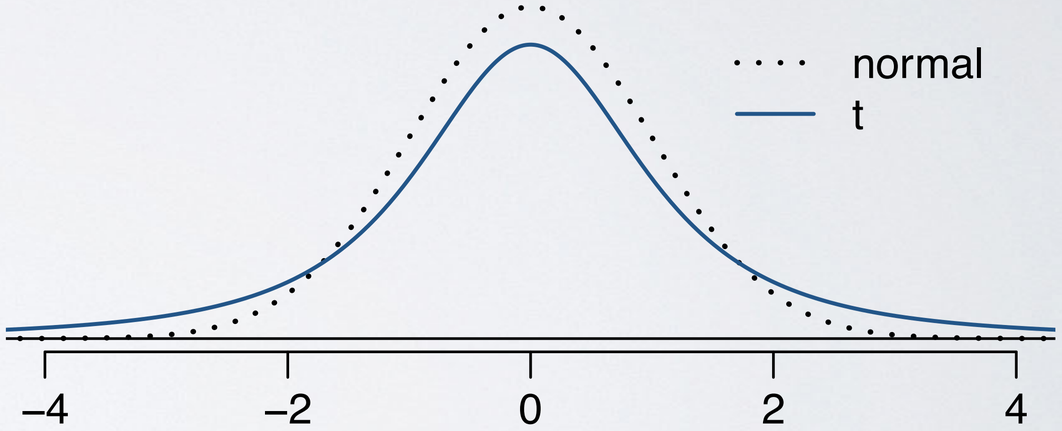
review:
normality of sampling distributions

- CLT: sampling distributions are nearly normal as long as the population distribution is nearly normal, for **any** sample size.
- Helpful special case, but difficult to verify normality in small data sets.
- Careful with the normality condition for small samples: don't just examine the sample, also think about where the data come from.
  - *"Would I expect this distribution to be symmetric, and am I confident that outliers are rare?"*



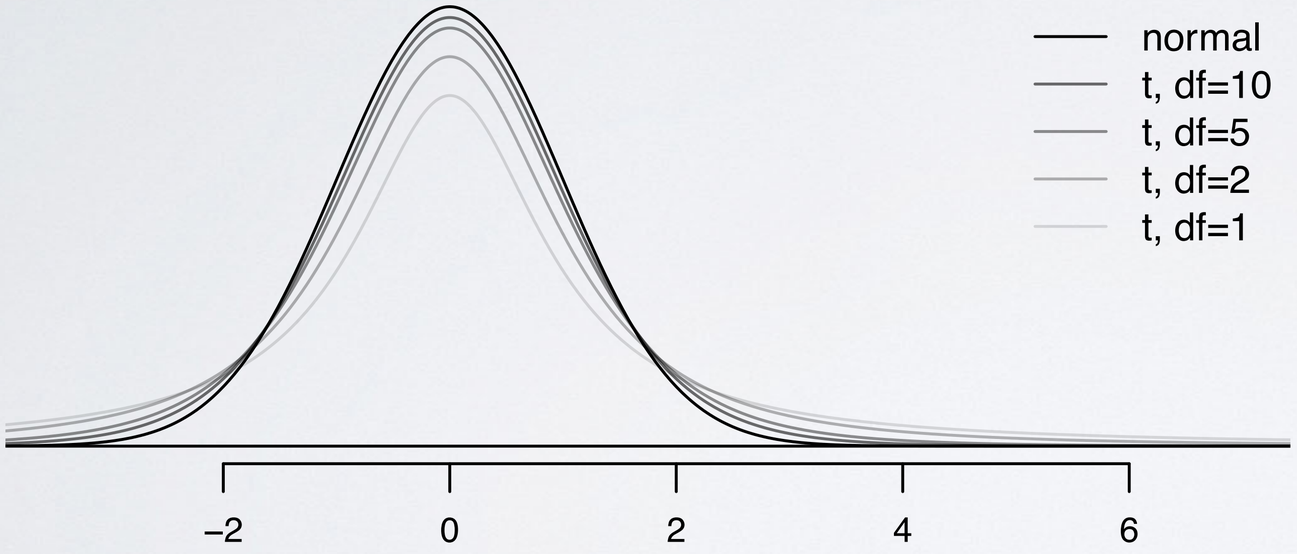
t distribution

- n is small & 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



t 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.

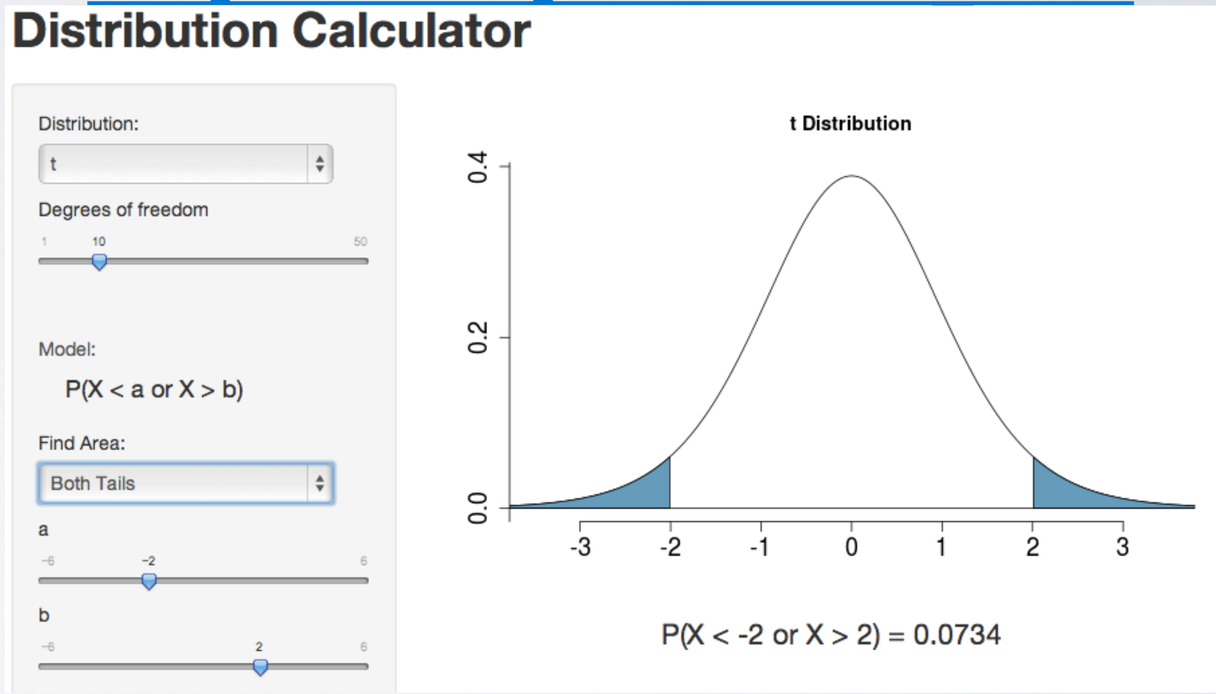
t statistic

- for inference on a mean where
  - σ unknown
  - n < 30
- calculated the same way

$$T = \frac{obs - null}{SE}$$

- p-value (same definition)
  - one or two tail area, based on H_A
  - using R, applet, or table

http://bitly.com/dist_calc



practice

Find the following probabilities.

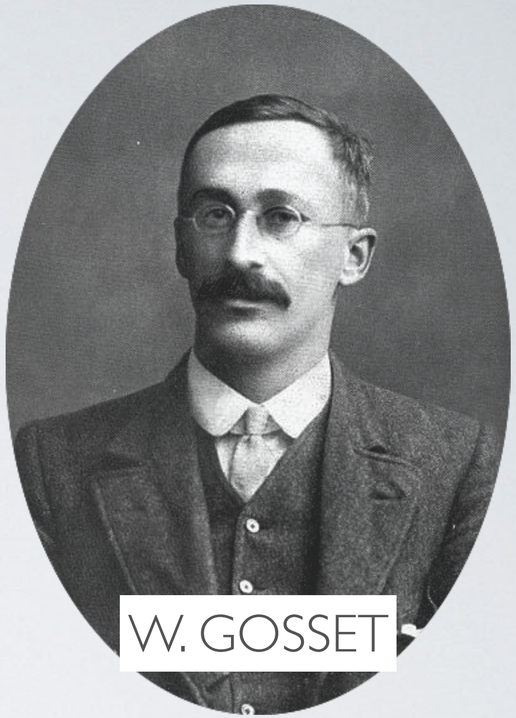
a. $P(|Z| > 2)$ 0.0455 → reject

b. $P(|t_{df = 50}| > 2)$ 0.0509 → fail to reject?

c. $P(|t_{df = 10}| > 2)$ 0.0734 → fail to reject

Say you have a two sided hypothesis test, and your test statistic is 2. Under which of these scenarios would you be able to reject the null hypothesis at the 5% sig. level?

origins of the t distribution



- Student's t
- William Gosset (1876 - 1937)
- "Head Experimental Brewer" at the Guinness brewing company

Gosset: http://commons.wikimedia.org/wiki/File:William_Sealy_Gosset.jpg