Inference for a Normal Population

Introduction to Student's t test

<u>Inference about means:</u>

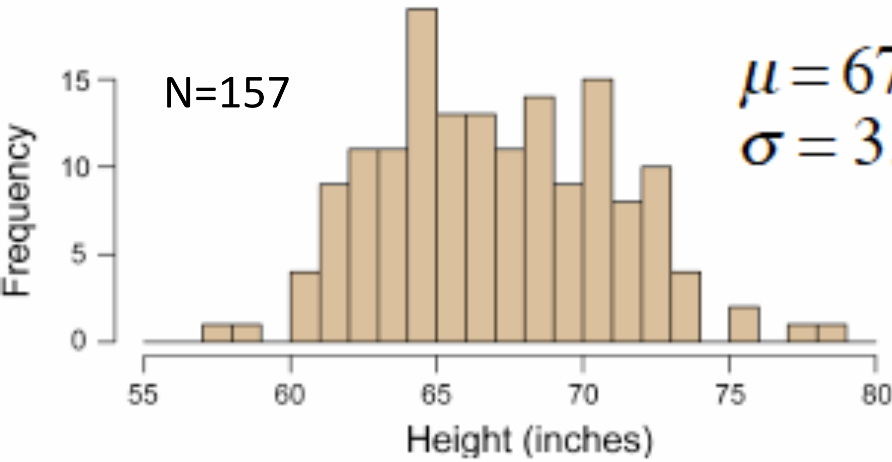
As a reminder….

- To make statistical statements, we need to describe the **<u>sampling distribution</u>** of an estimator.

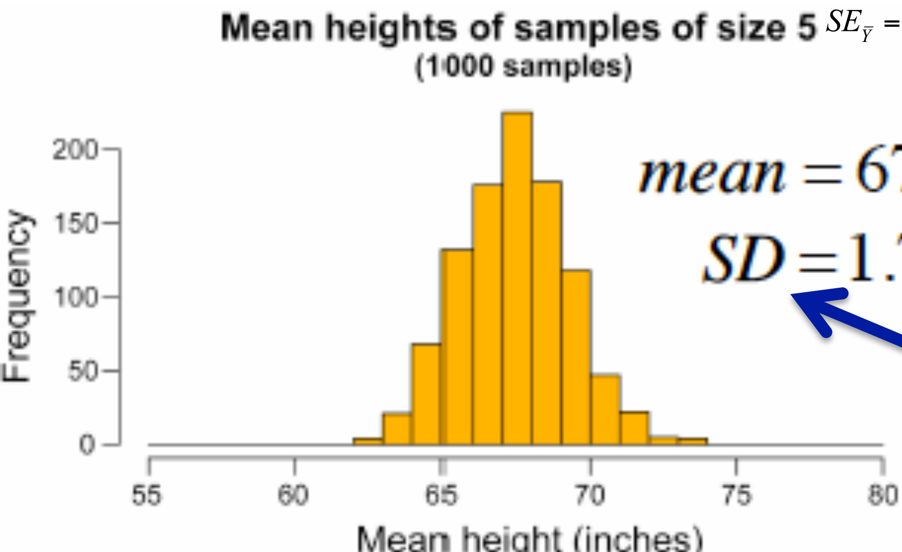
  – The sampling distribution is the probability distribution of all values of an estimate that we might obtain when sampling a population

  – **When the variable, Y, is normally distributed *or <u>n is large</u>*, the sampling distribution for E(Y) is normal***

* thank-you Central Limit Theorem



$$\mu = 67.4$$
$$\sigma = 3.9$$



$$SE_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{3.9}{\sqrt{5}} = 1.7$$

$$mean = 67.4$$
$$SD = 1.7$$

$$SE_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{3.9}{\sqrt{5}} = 1.7$$

<u>The central limit theorem has two problems:</u>

- It depends on a large sample size (n > 30-ish)
- To use it, we need to know σ^2 (i.i.d.)-- but we seldom do.

<u>Inference about means:</u>

Because $\bar{Y}$ is normally distributed, we can convert the distribution to the **<u>standard normal distribution:</u>**

$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}}$$

This gives a probability distribution of the difference between a sample mean and the population mean

A one sample *t*-test is usually used instead of a one sample *z*-test to correct for:

A) X-bar used to estimate μ

B) s used to estimate σ

C) n used to estimate population size

D) none of the above

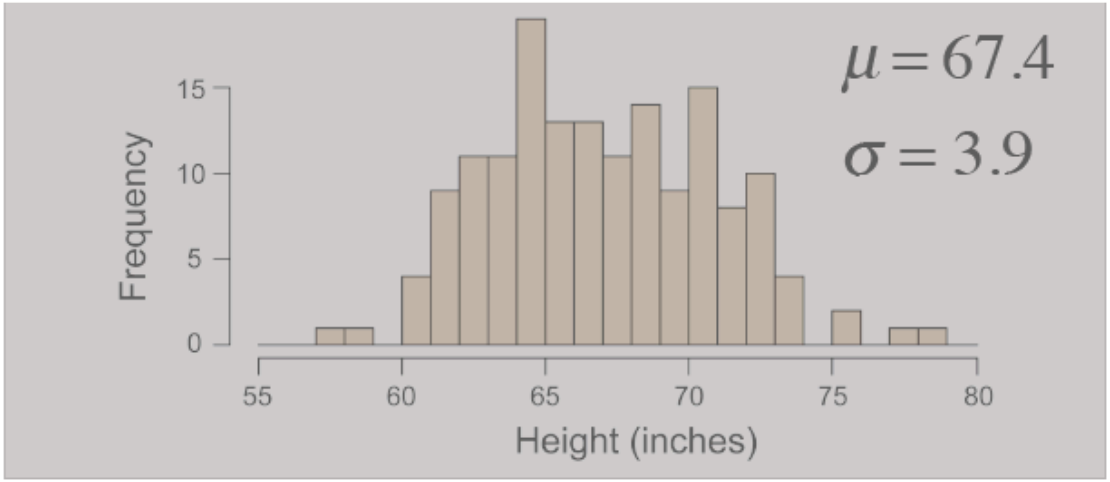
But we don't know σ!

Now what?

- we *do* know ***s***, the standard deviation of our sample, which estimates σ.
  – We can use *s* to get: $SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$

  –This is used as an estimate of $\sigma_{\bar{Y}}$

Inference for a Normal Population

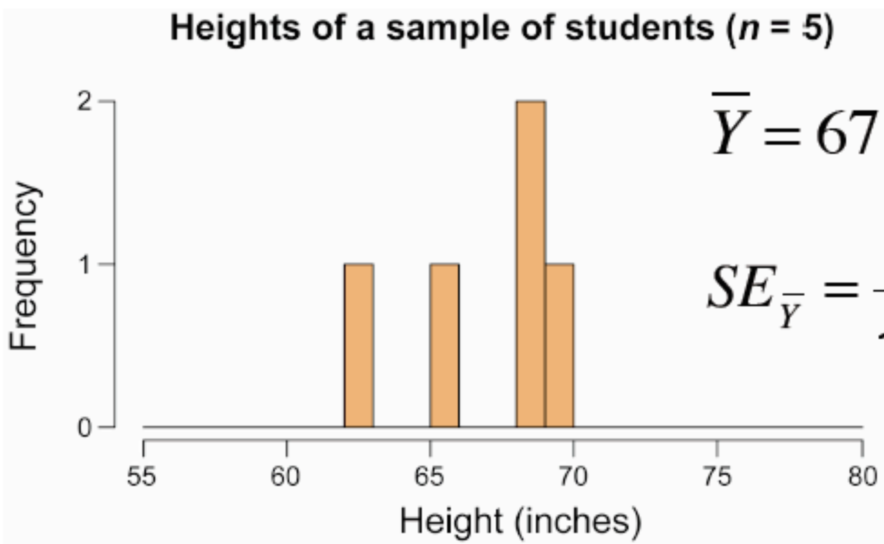


$$\mu = 67.4$$

$$\sigma = 3.9$$

In most cases, we don't know the real population distribution.

We only have a sample.



$$\overline{Y} = 67.1 \qquad s = 3.1$$

$$SE_{\overline{Y}} = \frac{s}{\sqrt{n}} = \frac{3.1}{\sqrt{5}} = 1.4$$

We use this as an estimate of $\sigma_{\overline{Y}}$

The Student's t Distribution:

$$t = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}} = \frac{\bar{Y} - \mu}{s/\sqrt{n}}$$

* Alias of William Gosset of the Guinness Brewing Company

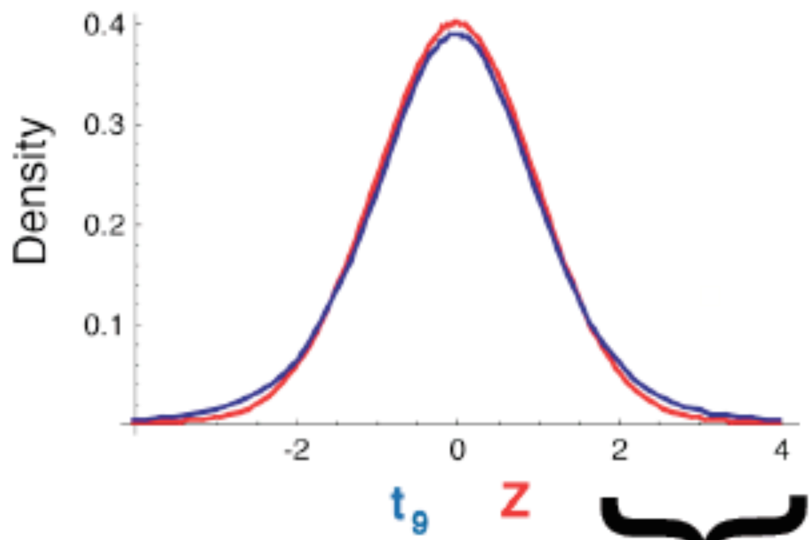
The Student's t Distribution:

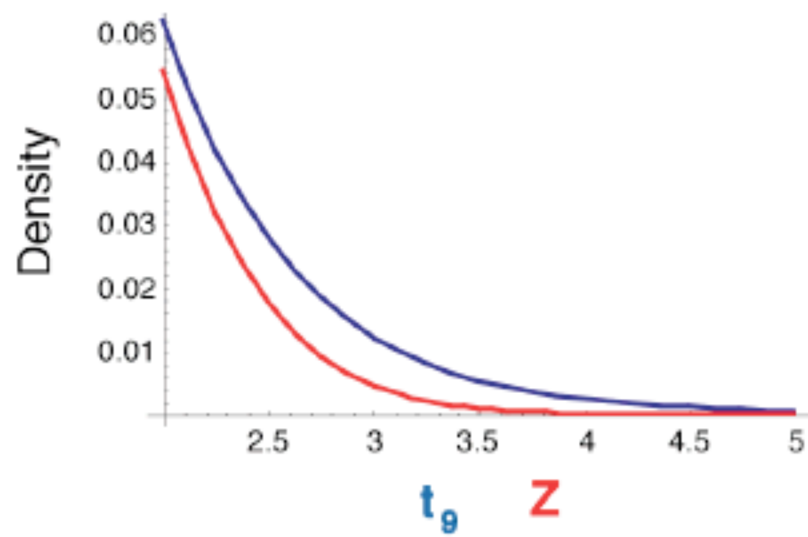
	t	**Z**
Stand. Error	$SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$	$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$
dof	n - 1	n
Sampling Distribution	t-distribution	Normal Distribution

<u>The consequences of using $SE_{\bar{Y}}$ instead of</u> $\sigma_{\bar{Y}}$

- The value of $SE_{\bar{Y}}$ is different for each sample; it doesn't have a constant value like $\sigma_{\bar{Y}}$
  - results in the introduction of some error
    - t-distribution is wider than the Normal distribution
      - not as precise
    - As sample size, *n*, increases the t-distribution narrows and approaches the Normal Distribution
- dof = n - 1 because we have 'used up' one piece of information when we estimate $\sigma_{\bar{Y}}$

Inference for a Normal Population





$$Z = \frac{\bar{Y} - \mu}{\sigma_{\bar{Y}}}$$

$$t = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}}$$

A one sample *t*-test is conducted on H_0: $\mu = 81.6$. The sample has X-bar=84.1, $s = 3.1$, and $n = 25$. The t-test statistic is:

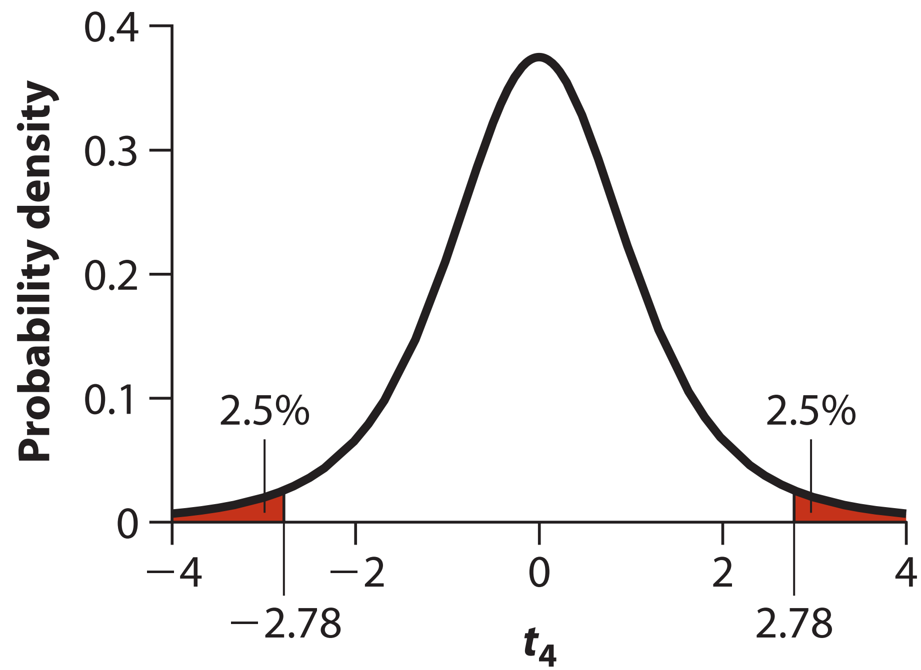
A)0.806

B)1.803

C)4.032

Critical values for the student's t distribution:

- Statistical Table C
- $t_{0.05(2),df}$
  – *the 0.05 is the fraction of the area under the curve shared between the two tails of the distr'n:* **2.5% > $t_{0.05(2),df}$** and **–2.5% <-$t_{0.05(2),df}$**



Use the t-distribution to calculate the confidence interval for the mean of a normal distribution

$$\boxed{-t_{\alpha(2),df} < \frac{\bar{Y} - \mu}{SE_{\bar{Y}}} < t_{\alpha(2),df}}$$

This can be re-arranged (see pg 308):

$$\boxed{\bar{Y} - t_{\alpha(2),df} SE_{\bar{Y}} < \mu < \bar{Y} + t_{\alpha(2),df} SE_{\bar{Y}}}$$

Is a 95% confidence interval always sufficient in a t-test as compared to a z-test?

A) Yes, the area under the tails are equal and therefore have the same likelihood of encompassing the true parameter

B) No, the Z-test has fatter tails and therefore the 99% confidence interval is often more appropriate

C) No, the t-test has fatter tails and therefore the 99% confidence interval is often more appropriate

D) Yes, the size of the tails has no effect on finding the true parameter of the mean

•