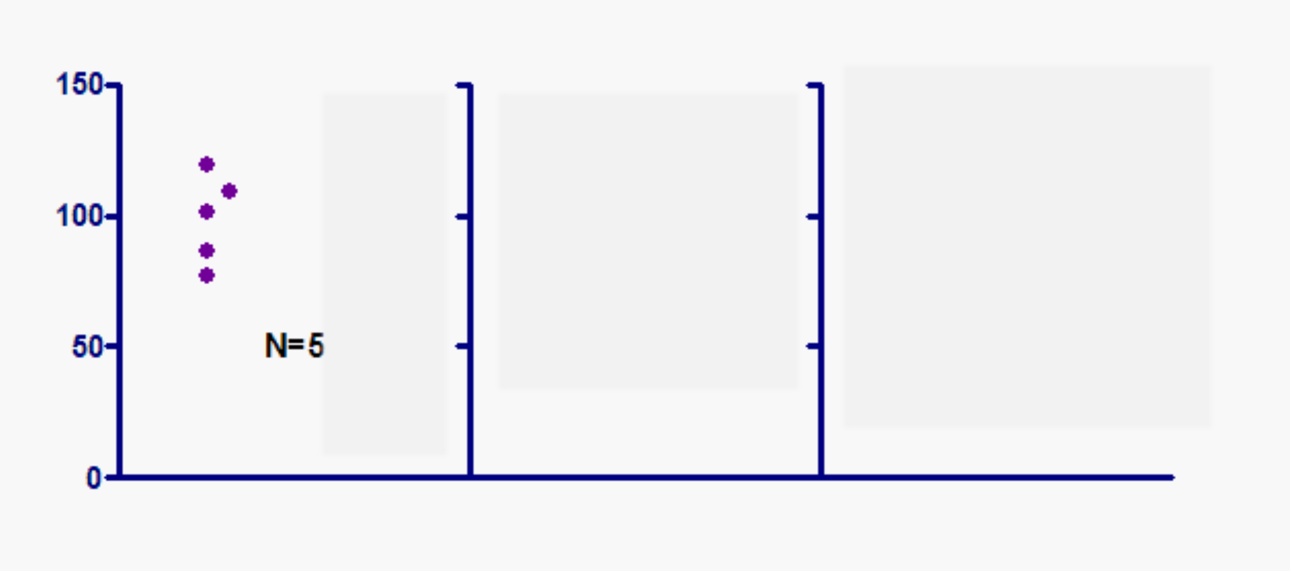
Learning Objectives

- **Describe** *confidence intervals* at the conceptual level
- **Contrast** *confidence intervals* with other measures of variability (e.g., standard deviation)
- **Calculate** *confidence intervals*

Confidence Intervals

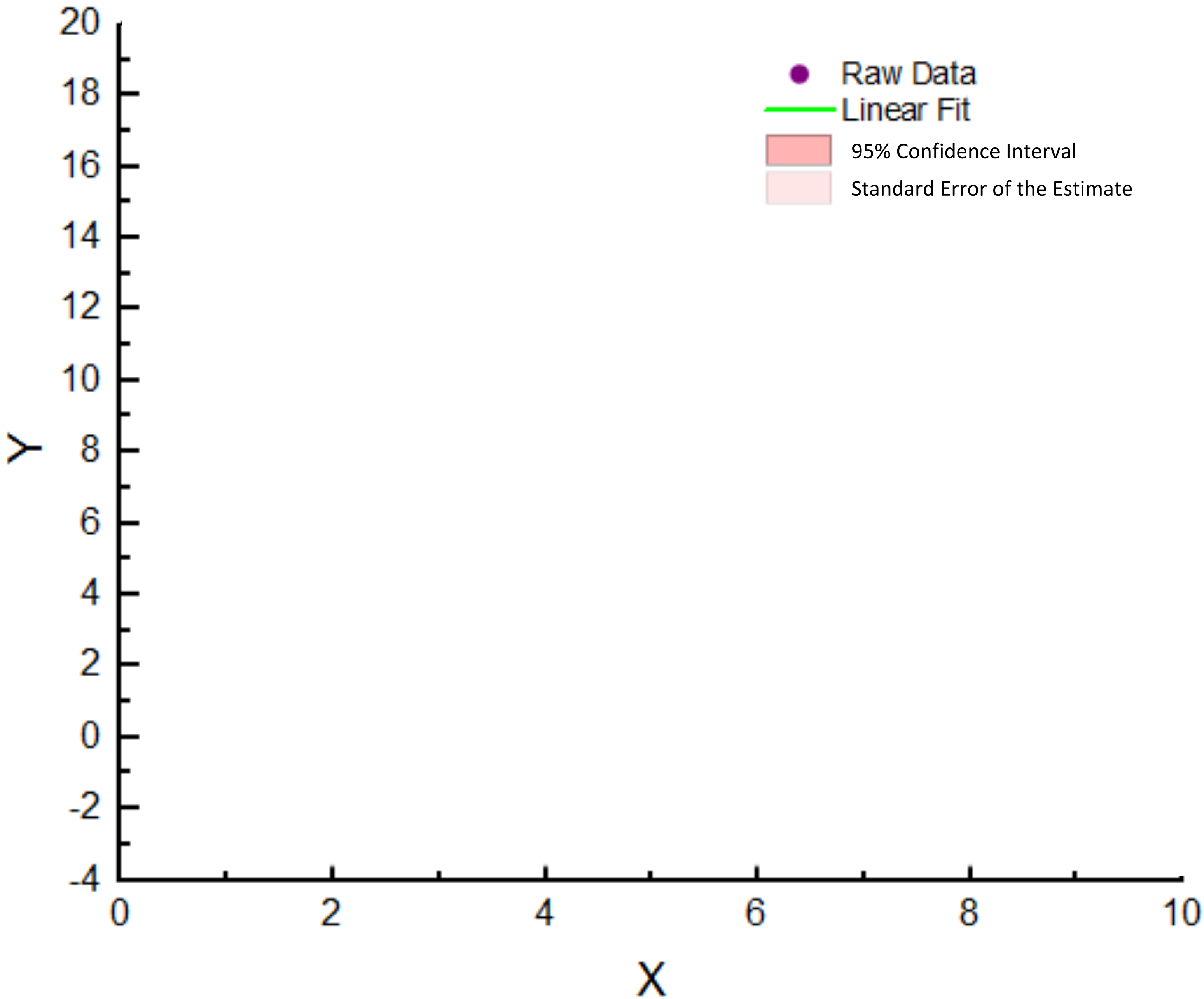
- Characterize our confidence in **parameter** estimates
  - *CI*'s are a range we believe contains the parameter
    - Akin to "margin of error"



s vs. *CI*

- Standard deviation describes the variability of observations around a statistical model
  - Ex. Variability of X's around $\bar{X}$
    - *How much do observations differ from the sample mean or a regression line?*

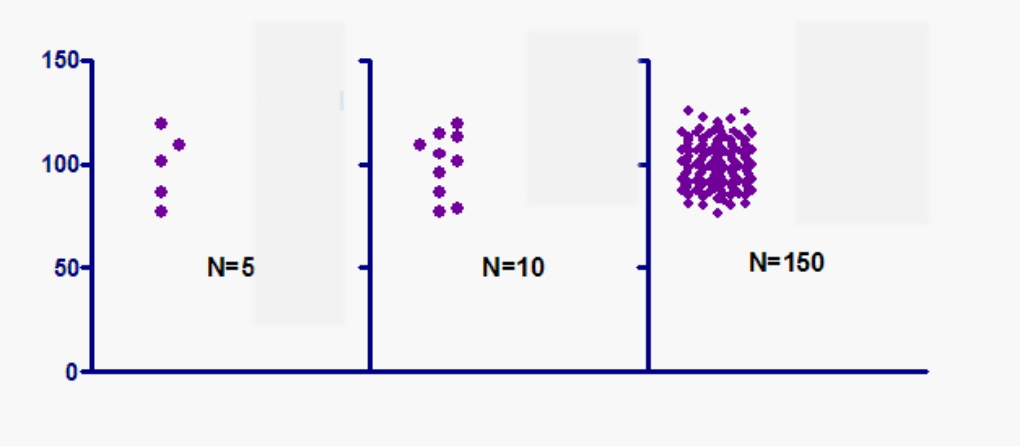
- Confidence intervals describe the variability of the statistical model itself
  - Ex. Variability of $\bar{X}$'s around $\mu_{\bar{X}}$
    - *How different might the sample mean or regression line be if I collected a new sample?*


Raw Data
Linear Fit
95% Confidence Interval
Standard Error of the Estimate
Y
X
20
18
16
14
12
10
8
6
4
2
0
-2
-4
0
2
4
6
8
10

Confidence Intervals

- Like α, we choose our confidence!
  - Larger *CI's more certainly* contain population parameters
    - But are *less practically useful*; & difficult to falsify

- Imagine a 100% Confidence Interval…



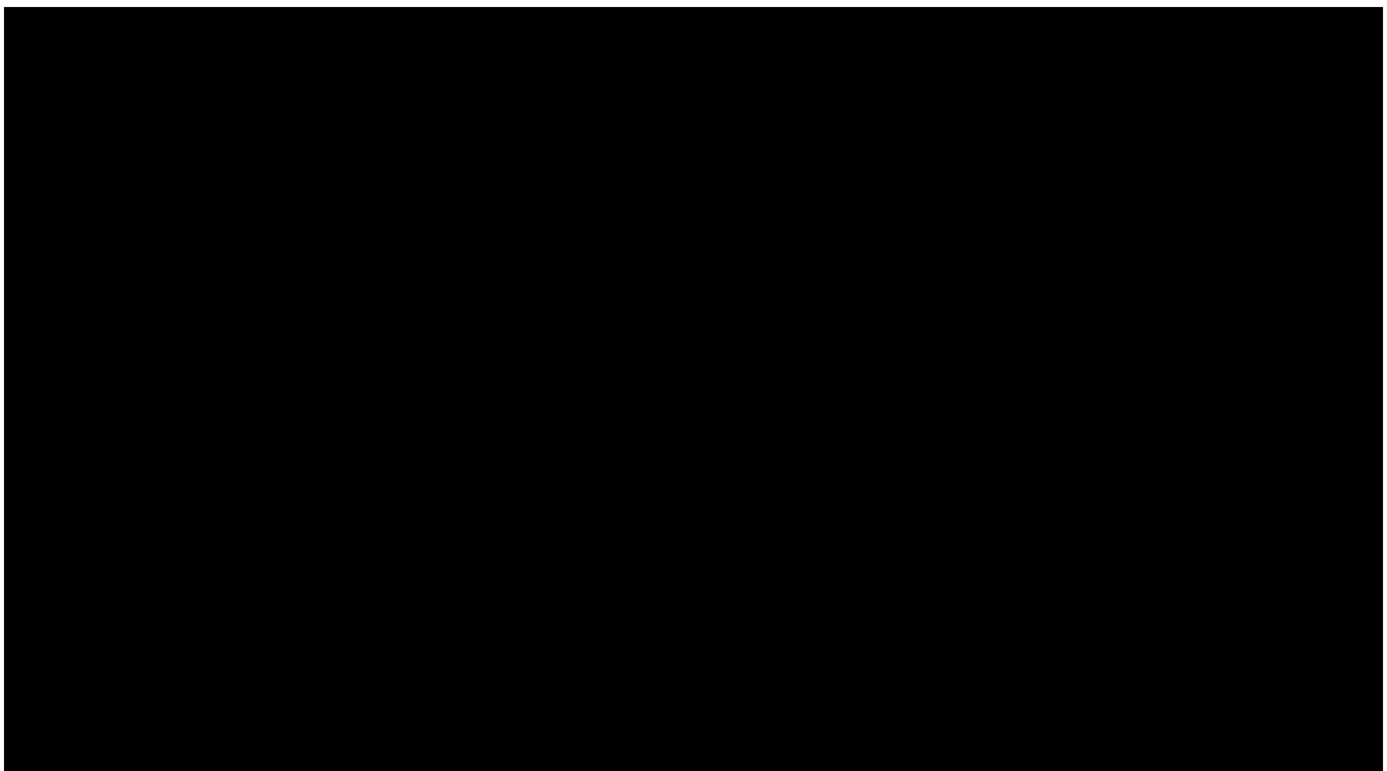
Formula for Confidence Intervals

$$\mu_{\text{lower}} = \boxed{\bar{X}_{\text{obt}}} - (s_{\bar{X}} * t_{.025})$$

- We start by assuming $\bar{X}_{\text{obt}} = \mu$
  - We assume that our observed data IS exactly the population parameter!

<u>NHST</u>: Assuming H_0 ***a priori* expectation is true**, here are plausible *a posteriori* observations (given sampling variability)

<u>CI</u>: Assuming ***a posteriori* observation is true**, here are plausible population parameters (given sampling variability)



Formula for Confidence Intervals

$$\mu_{\text{lower}} = \boxed{\bar{X}_{\text{obt}}} \boxed{-} (\boxed{s_{\bar{X}}} \boxed{*\ t} \boxed{_{.025}})$$

$$\mu_{\text{upper}} = \boxed{\bar{X}_{\text{obt}}} \boxed{+} (\boxed{s_{\bar{X}}} \boxed{*\ t} \boxed{_{.025}})$$

- We start by assuming $\bar{X}_{\text{obt}}$ = μ
- We add or subtract from the best guess $(\bar{X}_{\text{obt}})$ to set upper and lower bounds for plausible population parameters
- Standard error describes the variability of $\bar{X}'s$ given *N*
  - $s_{\bar{X}}$ = $CI_{68\%}$
- Multiply by value of *t* that sets this area into each tail of the distribution of sample means
  - Normally this will be $t_{.050}$ (when $CI_{90\%}$), $t_{.025}$ ($CI_{95\%}$), $t_{.005}$ ($CI_{99\%}$)

Formula for Confidence Intervals

$$\mu_{\text{lower}} = \bar{X}_{\text{obt}} - (s_{\bar{X}} * t_{?})$$

$$\mu_{\text{upper}} = \bar{X}_{\text{obt}} + (s_{\bar{X}} * t_{?})$$

What about when t = 1?

- When t = 1…then $\bar{X}_{\text{obt}}$ is 1 $s_{\bar{X}}$ away from μ
  - What is the area under the curve (*AUC*)?

- When t = 1; then $CI_{68\%}$*

**If N is large*

Formula for Confidence Intervals

$$\mu_{\text{lower}} = \bar{X}_{\text{obt}} - (s_{\bar{X}} * t_{.005})$$

$$\mu_{\text{upper}} = \bar{X}_{\text{obt}} + (s_{\bar{X}} * t_{.005})$$

$t_{.005}$ is appropriate for a **99% *CI***

- 0.5% of the time, actual parameter will be lower than the interval
- 0.5% of the time, actual parameter will be higher than the interval

Formula for Confidence Intervals

$$\mu_{\text{lower}} = \bar{X}_{\text{obt}} - (s_{\bar{X}} * t_{.025})$$
$$\mu_{\text{upper}} = \bar{X}_{\text{obt}} + (s_{\bar{X}} * t_{.025})$$

$t_{.025}$ is appropriate for a **95% *CI***

- 2.5% of the time, actual parameter will be lower than the interval*
- 2.5% of the time, actual parameter will be higher than the interval

*Note this is a simplification

Formula for Confidence Intervals

$$\mu_{\text{lower}} = \bar{X}_{\text{obt}} - (s_{\bar{X}} * t_{.X})$$

$$\mu_{\text{upper}} = \bar{X}_{\text{obt}} + (s_{\bar{X}} * t_{.X})$$

Conceptual steps to creating $CI_{X\%}$

1. Determine total *AUC* that we're willing to be wrong $(1 - X)$
2. Split total *AUC* between the two tails $(\frac{AUC}{2})$
3. Visualize $\frac{AUC}{2}$ in lower/upper tail
4. Determine value of *t* for $\frac{AUC}{2}$, using appropriate *df*

Formula for Confidence Intervals

$$\mu_{\text{lower}} = \bar{X}_{\text{obt}} - (s_{\bar{X}} * t_{.05})$$

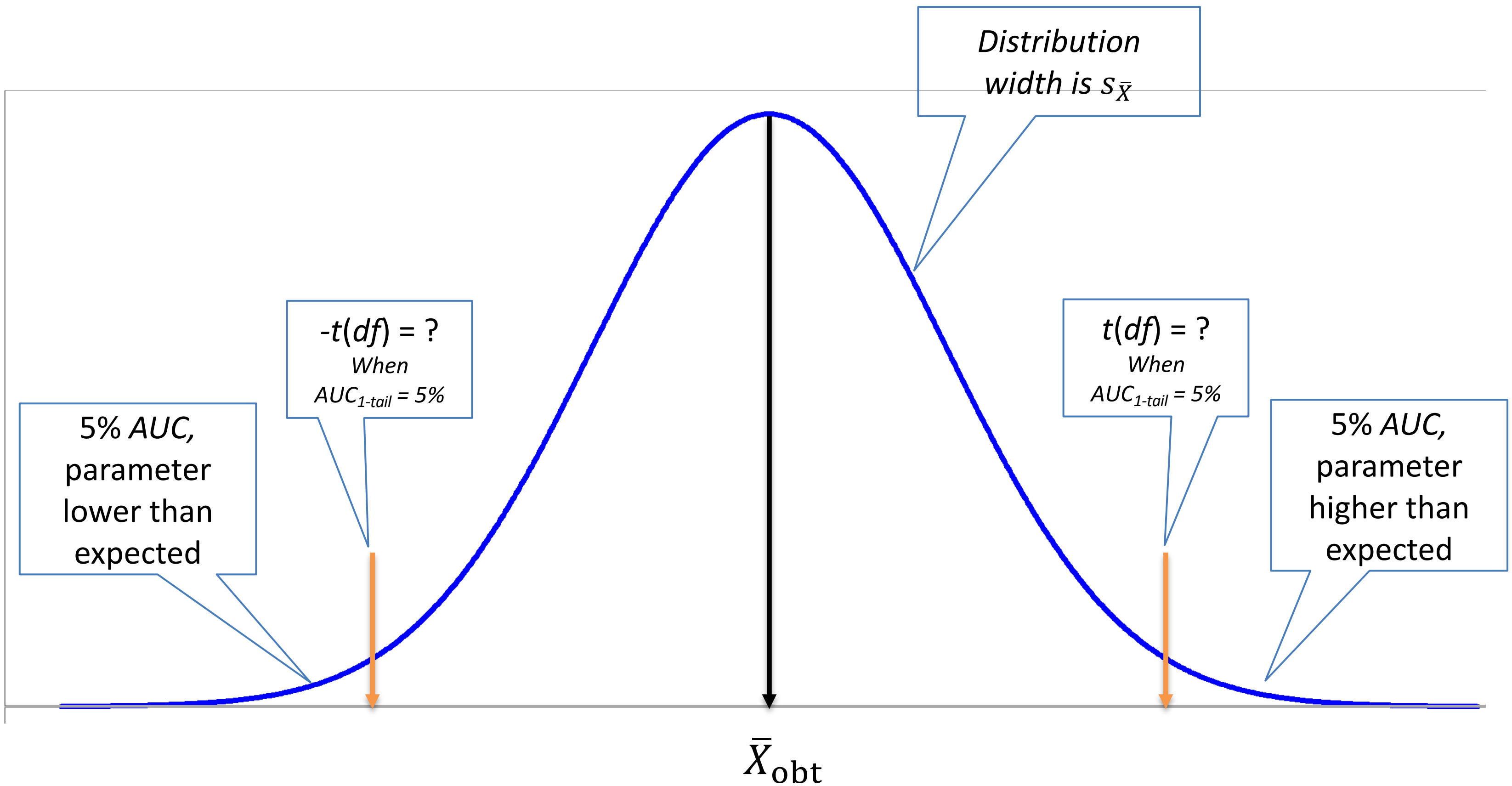
$$\mu_{\text{upper}} = \bar{X}_{\text{obt}} + (s_{\bar{X}} * t_{.05})$$

Practical steps to create ***CI***$_{90\%}$

1. Subtract .90 from 1 — $1 - X = .10$
2. Divide .10 by 2 — $\frac{AUC}{2} = .05$
3. Place result in lower/upper tail
4. Find value of *t* that puts .05 in one tail — $t_{.05}(28) = 1.701$

(let's assume $N = 30$ & independent groups)

Visualize $CI_{90\%}$



Example: Practice Problem 13.4 (pp. 345)

- Manufacturer samples from 200 light bulbs and records how many hours it takes for them to burn out. The sample has a mean life of 215hrs (*s* = 8hrs). Construct the 99% CI using the closest value for *df*

$\bar{X}_{obt} = 215$

$s = 8$

$N = 200$

$s_{\bar{X}} = ???$

$\boldsymbol{t_{???} = ???}$

$$\mu_{lower} = \bar{X}_{obt} - (s_{\bar{X}} * t_?)$$

Determine *t*:

- $1 - .99 = .01$ (total *AUC*)
- $\frac{.01}{2} = .005$ (in each tail)
- $t_{.005}(199) = 2.617^*$

*Approximate using closest *df*
(always defer to lower *df* value)

Example: Practice Problem 13.4 (pp. 345)

- Manufacturer samples from 200 light bulbs and records how many hours it takes for them to burn out. The sample has a mean life of 215hrs (*s* = 8hrs). Construct the 99% CI using the closest value for *df*

$\bar{X}_{\text{obt}} = 215$

$s = 8$

$N = 200$

$s_{\bar{X}} =$

$t_{.005} = 2.617$

$$\mu_{\text{lower}} = \bar{X}_{\text{obt}} - (s_{\bar{X}} * t_{.005})$$

$$s_{\bar{X}} = \frac{s}{\sqrt{N}}$$

Example: Practice Problem 13.4 (pp. 345)

- Manufacturer samples from 200 light bulbs and records how many hours it takes for them to burn out. The sample has a mean life of 215hrs (*s* = 8hrs). Construct the 99% CI using the closest value for *df*

$\bar{X}_{\text{obt}} = 215$

$s = 8$

$N = 200$

$s_{\bar{X}} = .5657$

$t_{.005} = 2.617$

$$\mu_{\text{lower}} = \bar{X}_{\text{obt}} - (s_{\bar{X}} * t_{.005})$$

$$s_{\bar{X}} = \frac{s}{\sqrt{N}} = \frac{8}{\sqrt{200}} = 0.5657$$

$$s_{\bar{X}} * t_{\mathbf{.005}} =$$

Lower = 215 - 1.48 = 213.52

Upper = 215 + 1.48 = 216.48

Example: Practice Problem 13.4 (pp. 345)

- Reporting with confidence intervals:

"Our lightbulbs had a relatively long life, M = 215 hours (s = 8), $CI_{99\%}$ [213.52, 216.48]."

Inference:

"Our lightbulbs last at least 213 hours on average, (with greater than 99% confidence) $CI_{99\%}$ [213.52, 216.48]."