Week 6

t-distributions and Confidence Intervals for a Mean

ST 314
Introduction to Statistics for Engineers

Oregon State University

Estimating a Population Mean

A car manufacturer wants to estimate the average annual mileage of drivers of their top of the line sports car. They would like to be 95% confident in the estimate and assume the population to be normal. A random sample of 10 drivers yields the following data:

11501	8987	12166	9247	10143
8230	3111	13009	7891	10392

According to the central limit theorem, what is the sampling distribution of the sample mean?

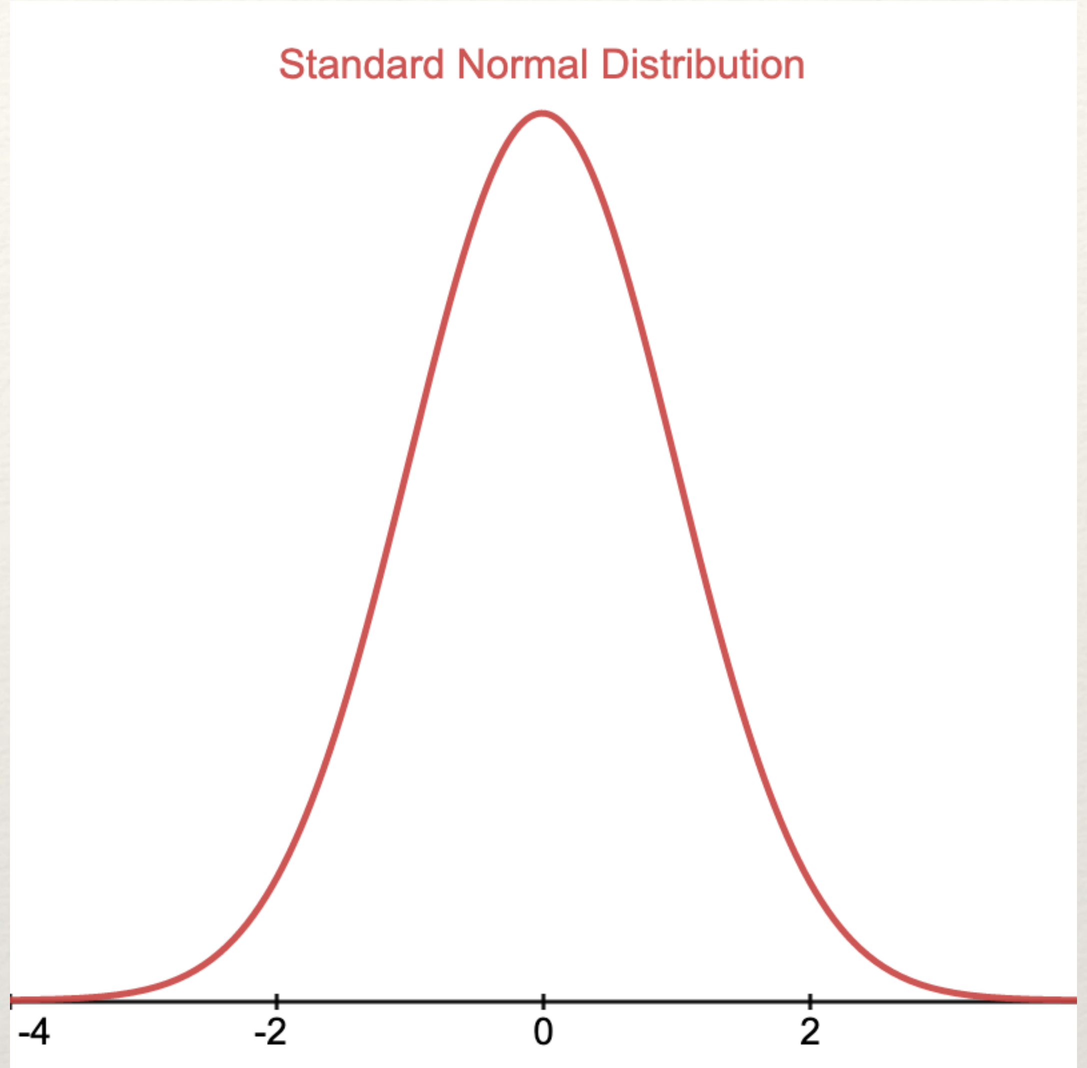
$$\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

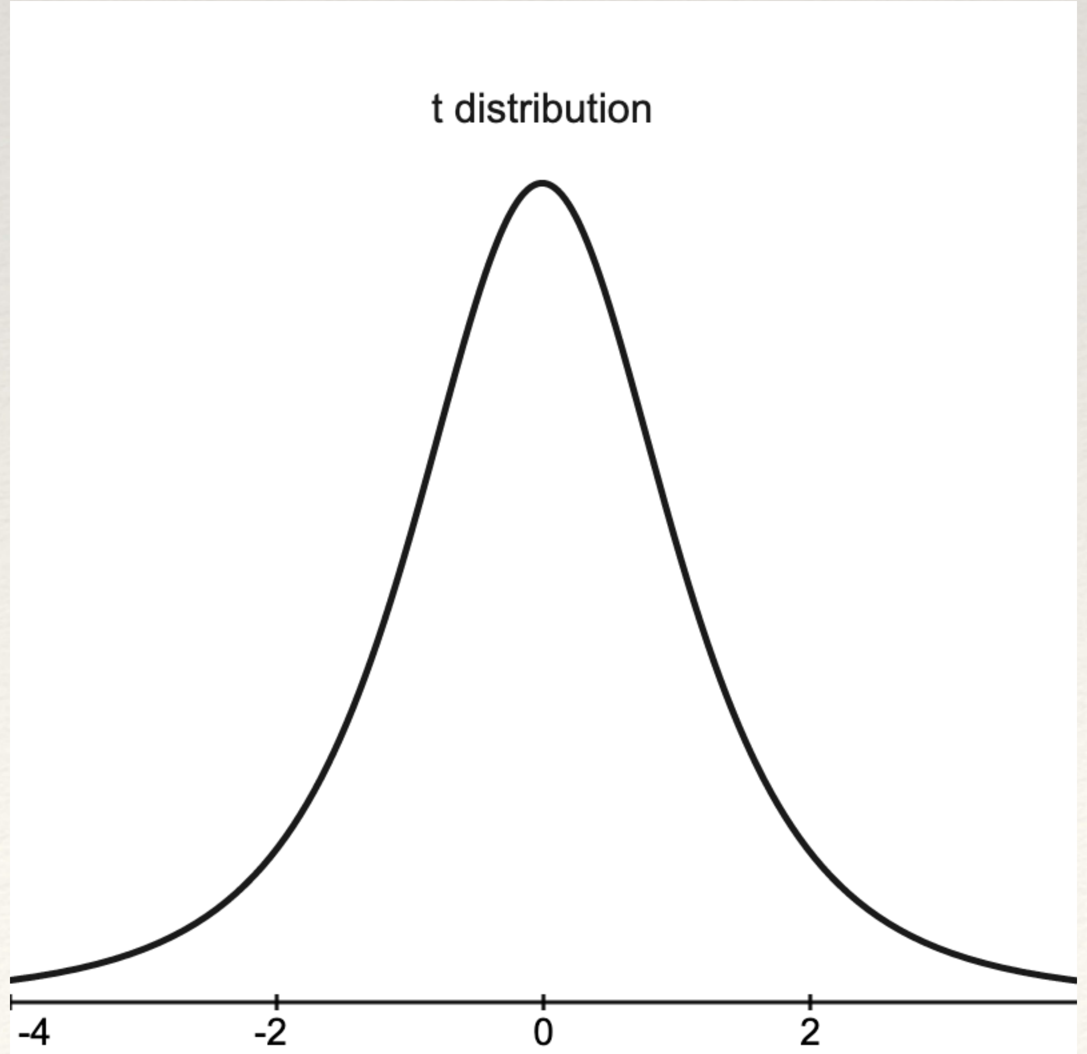
pop. sd

pop. mean

t-distribution

- The sampling distribution is normal when standardized with σ_x. $z = \dfrac{\bar{x} - \mu}{\sigma_x / \sqrt{n}} \sim N(0, 1)$
- If σ_x is unknown, can we use the sample standard deviation s_x instead? yes, but we can't use the normal distribution anymore.
- The distribution is ***not*** normal when we replace σ_x with s_x.
- When we standardize using s_x, our sampling distribution comes from a t-distribution.





t-distribution

- Symmetric and centered at 0.
- Distribution is defined by its degrees of freedom
- As the degrees of freedom increases towards ∞, the t distribution approaches the standard normal distribution $N(0,1)$
- The t distribution has heavier area in the tails when compared to the standard normal distribution.
- Degrees of freedom are based on sample size and the number of unknown parameters.

t-distribution

For sufficiently large n, the distribution of the standardized sample mean follows a t-distribution with $n - 1$ degrees of freedom.

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{\underbrace{n-1}_{\text{degrees of freedom}}}$$

Confidence interval for μ when σ is unknown

- When to use:

want to estimate μ (population mean)
and σ is unknown

- The confidence interval for a population mean is:

point estimate $\pm$ critical value $\times$ standard error of point estimate

$$\bar{X} \pm \underbrace{t^*_{1-\frac{\alpha}{2},\, df}}_{1-\frac{\alpha}{2} \text{ percentile from a t-dist'n with df as the degrees of free.}} \frac{s}{\sqrt{n}}$$

$s \leftarrow$ sample SD

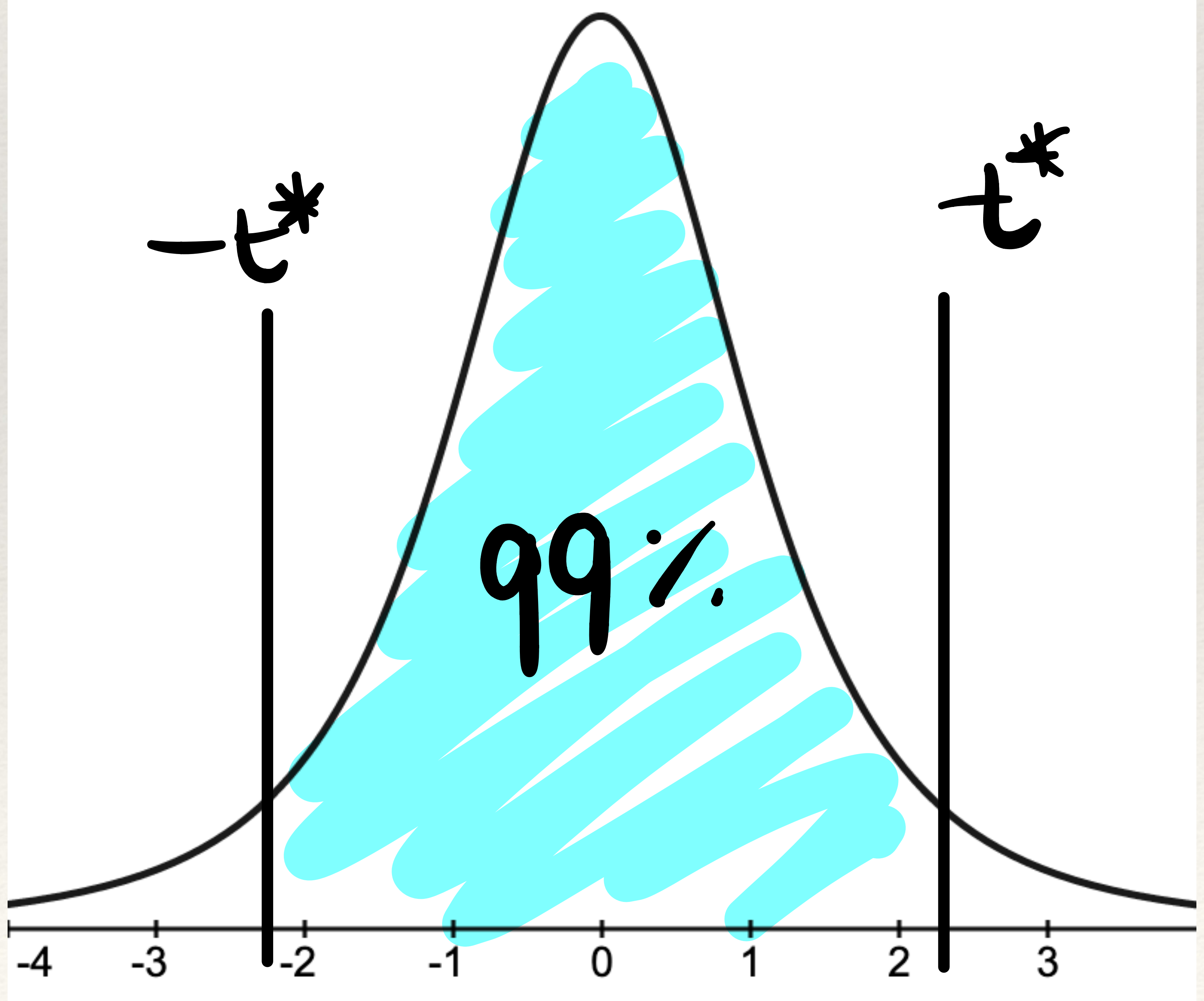
$(1-\alpha)100\%$ CI

Finding t Critical Values

Recall that for a $(1-\alpha)100\,\%$ confidence interval, the critical value is the $\left(1-\frac{\alpha}{2}\right)^{th}$ percentile.

$\alpha = 0.05$

- For a 95% confidence interval, we need to find the 97.5th percentile from the appropriate t distribution.
- For a 90% confidence interval, we need to find the 95th percentile from the appropriate t distribution.
- For a 99% confidence interval, we need to find the 99.5th percentile from the appropriate t distribution.



Finding t Critical Values Using `qt()`

Function	Function Values	What does it do?
`qt(p,df)`	`p` = area under the curve to the left of x_p `df` = degrees of freedom	This is the inverse cumulative distribution function. Finds percentiles for the t distribution. That is, finds x_p for the expression $$P(X \leq x_p) = p$$

$n = 300$

$df = n-1 = 300-1 = 299$

conf. level = 25% → qt(0.625, 299)

conf. level = 95% → qt(0.975, 299)

Confidence Interval Example #1

The FDA's webpage provides some data on mercury content of fish. Based on a sample of 15 croaker white fish (Pacific), a sample mean and standard deviation were computed as 0.287 and 0.069 ppm (parts per million), respectively. From a visualization of the 15 points, we do not find at strong skewness or extreme outliers. We will assume these observations are independent. Calculate the 90% confidence interval for the average mercury content of croaker white fish in the pacific.

$\bar{x} = 0.287$

$s = 0.069$

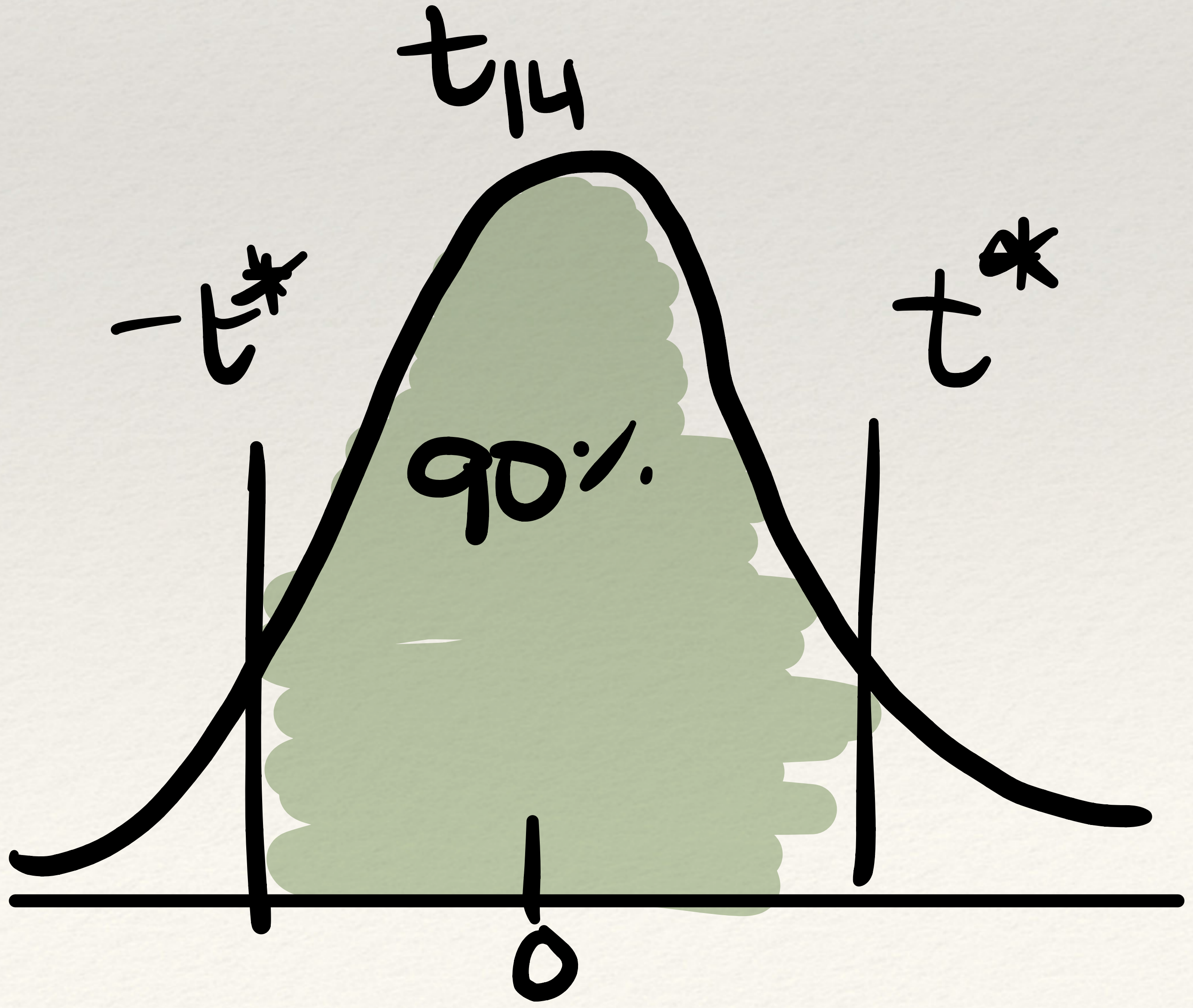
$n = 15$

$df = 14$

$$t^*_{0.95,14} = qt(0.95, 14) = 1.761$$

$$0.287 \pm 1.761\left(\frac{0.069}{\sqrt{15}}\right)$$

$$= (0.257, 0.318)$$



$t^* = 2.008$

Confidence Interval Example #2

Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 μg/l and a SD of 37.74 μg/l. Construct the 95% confidence interval for average lead concentration in the blood of traffic enforcing police officers in this urban environment.

$t^*_{0.975,51} = qt(0.975, 51) = 2.008$

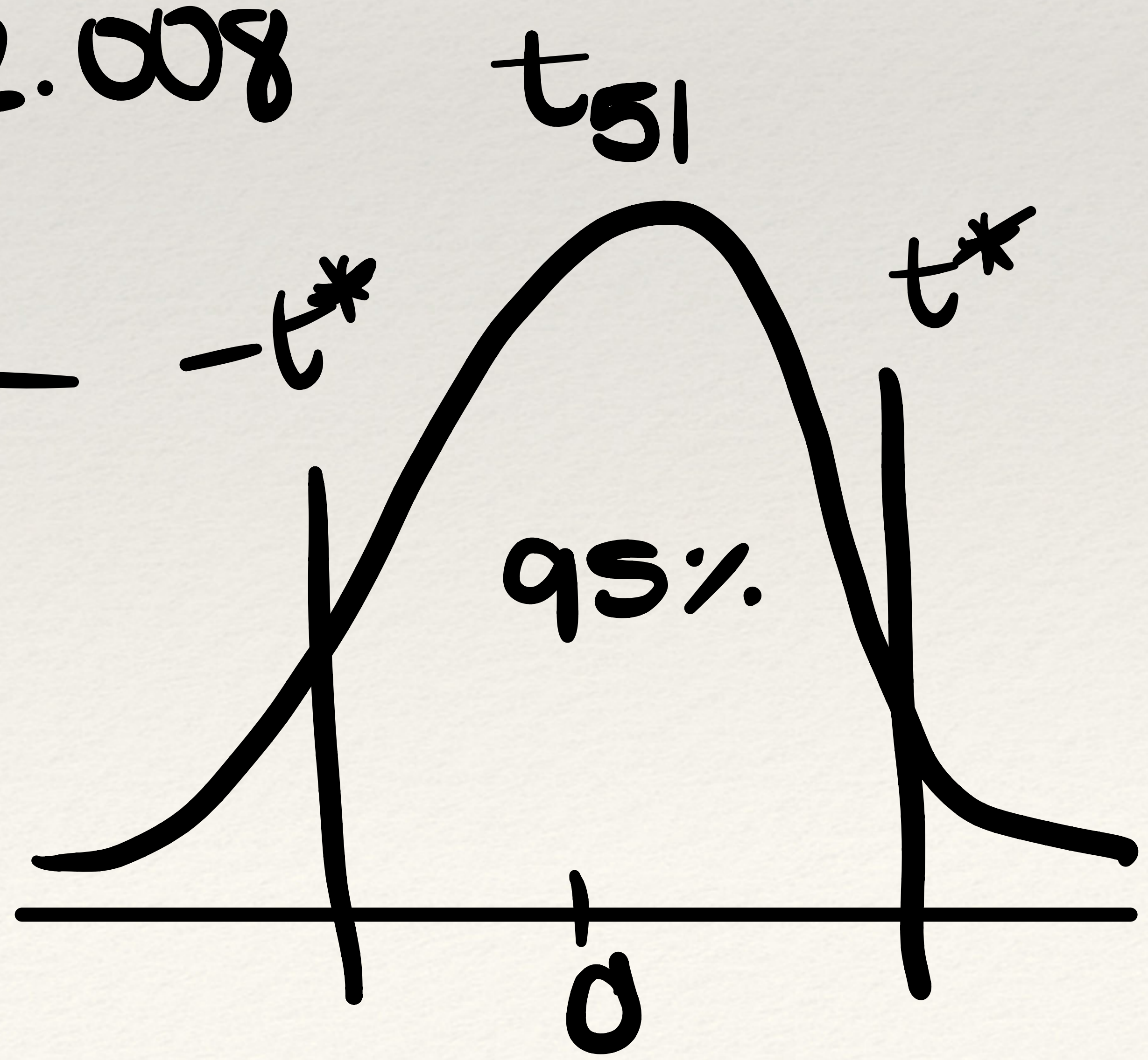
$\bar{x} = 124.32$

$s = 37.74$

$n = 52$

$$124.32 \pm 2.008\left(\frac{37.74}{\sqrt{52}}\right)$$

$$= (113.811, 134.829)$$



Confidence Interval Example #2

Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 μg/l and a SD of 37.74 μg/l. Construct the 95% confidence interval for average lead concentration in the blood of traffic enforcing police officers in this urban environment.

We are 95% confident that the true average lead concentration in the blood of traffic enforcing police officers in this urban area is between 113.811 micrograms per liter and 134.829 micrograms per liter with a point estimate of 124.32 micrograms per liter.