Definitions and Example for Confidence Intervals

Definition: A $(1-\alpha)*100\%$ confidence interval for a parameter Θ is a pair of random variables $\underline{A}$ and $\underline{B}$ such that $P(A \le \Theta \le B) = 1-\alpha$.

- A and B are random variables because they depend on sample data.
- Based on a particular sample, we observe realized values a and b. The interval $[a, b]$ is our <u>observed confidence interval</u>.

Notes:

1) If we want a 95% C.I., then $\alpha = 0.05$:
$(1-\alpha)*100\% = (1-0.05)*100\% = 0.95*100\% = 95\%$
If we want a 99% C.I., then $\alpha = 0.01$.

2) The quantity $(1-\alpha)*100\%$ (ex 95%) is referred to as the <u>confidence level</u> of the interval.

Example: $X_1, \ldots, X_n \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$
(both μ and σ^2 unknown)

From day 2 of class:

$$\boxed{\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean (as a random variable)

and $S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}$ is sample standard deviation (as a random variable)

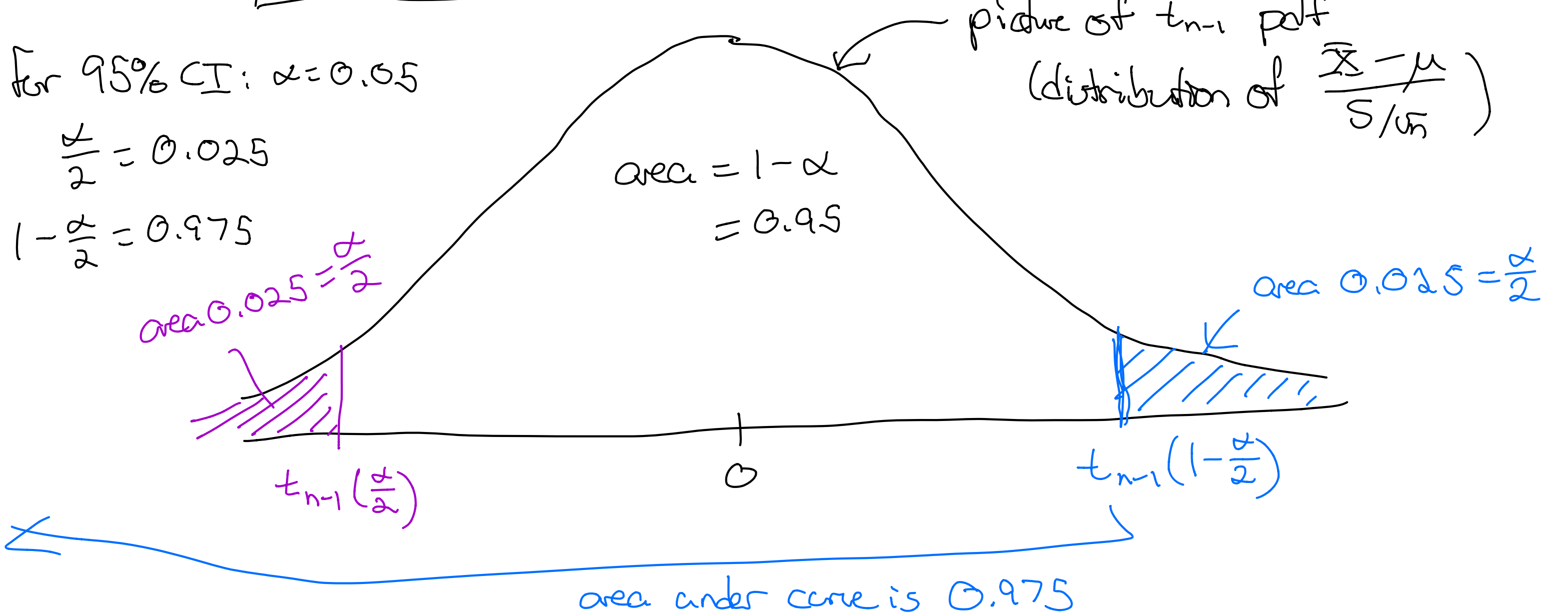
Denote the quantile q of the t_{n-1} distribution by $t_{n-1}(q)$

Then $\boxed{P\left(t_{n-1}\left(\frac{\alpha}{2}\right) \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le t_{n-1}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha}$

For 95% CI: $\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$1 - \frac{\alpha}{2} = 0.975$



Our Goal: A pair of random variables A, B such that

$P(A \le \mu \le B) = 1 - \alpha$

$$1-\alpha = P\left(t_{n-1}\left(\frac{\alpha}{2}\right) \le \frac{\boxed{\bar{X}}-\mu}{\boxed{S}/\sqrt{n}} \le t_{n-1}\left(1-\frac{\alpha}{2}\right)\right)$$

random variables

everything else is a number.

$$= P\left(t_{n-1}\left(\frac{\alpha}{2}\right)\cdot\frac{S}{\sqrt{n}} \le \bar{X}-\mu \le t_{n-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{S}{\sqrt{n}}\right)$$

$$= P\left(-\bar{X}+t_{n-1}\left(\frac{\alpha}{2}\right)\cdot\frac{S}{\sqrt{n}} \le -\mu \le -\bar{X}+t_{n-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{S}{\sqrt{n}}\right)$$

$$= P\left(\bar{X}-t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}} \ge \mu \ge \bar{X}-t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}\right)$$

$$= P\Big(\underbrace{\bar{X}-t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}}_{A} \le \mu \le \underbrace{\bar{X}-t_{n-1}\left(\frac{\alpha}{2}\right)\frac{S}{\sqrt{n}}}_{B}\Big)$$

$-5 \le -3$

$5 \ge 3$

$3 \le 5$

Our $(1-\alpha)*100\%$ confidence interval for μ is

random variables

$$\left[\boxed{\bar{X}}-t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{\boxed{S}}{\sqrt{n}},\ \boxed{\bar{X}}-t_{n-1}\left(\frac{\alpha}{2}\right)\frac{\boxed{S}}{\sqrt{n}}\right]$$

Based on a data set with observed sample mean $\bar{x}$ and sample standard deviation s, our observed confidence interval is

$$\left[\bar{x}-t_{n-1}\left(1-\frac{\alpha}{2}\right)\frac{s}{\sqrt{n}},\ \bar{x}-t_{n-1}\left(\frac{\alpha}{2}\right)\frac{s}{\sqrt{n}}\right]$$

no random variables! all just numbers.

can't make probability statements.

Example: Body Temperatures

- It's generally believed that the average body temperature is 98.6 degrees Farenheit (37 degrees Celsius).
- Let's investigate with measurements of the temperatures of 130 adults.

Confidence Interval Calculation

$$\left[\bar{x} - t_{n-1}\left(1-\tfrac{\alpha}{2}\right)\tfrac{s}{\sqrt{n}},\ \bar{x} + t_{n-1}\left(\tfrac{\alpha}{2}\right)\tfrac{s}{\sqrt{n}}\right)$$

Sample mean and standard deviation:

```
x_bar <- mean(bodytemp$temp)
x_bar
```

```
## [1] 98.24923
```

```
s <- sd(bodytemp$temp)
s
```

```
## [1] 0.7331832
```

Verifying our sample size is $n = 130$

```
n <- nrow(bodytemp)
n
```

```
## [1] 130
```

Finding appropriate quantiles for a 95% confidence interval:

```
t_lower <- qt(0.025, df = n - 1)
t_lower
```

```
## [1] -1.978524
```

```
t_upper <- qt(0.975, df = n - 1)
t_upper
```

```
## [1] 1.978524
```

Calculation of the confidence interval:

t_upper

```
x_bar - t_upper * s / sqrt(n)
```

$\rightarrow 98.249 \boxed{- 1.979} * 0.733 / \sqrt{130}$

```
## [1] 98.122
```

```
x_bar - t_lower * s / sqrt(n)
```

$\rightarrow 98.249 - (-1.979) * 0.733 / \sqrt{130}$

```
## [1] 98.37646
```

$98.249 + 1.979 * 0.733 / \sqrt{130}$

Our observed confidence interval for μ is $[98.122, 98.376]$

Interpretation: We are 95% confident that the population mean body temperature is between 98.122 °F and 98.376 °F.

Non-interpretation (can't say this!!!)
There is probability 0.95 that μ is between 98.122 and 98.376. (This is for a Bayesian credible interval!)

Comparison of Frequentist confidence intervals and Bayesian credible intervals.

In both cases: a range of plausible values for the unknown parameter Θ.

	Bayesian	Frequentist
The random variable is:	Θ (the parameter)	A, B the endpoints of the confidence interval
Why is the random variable random?	Expresses our state of knowledge about Θ.	Each sample we take gives a different confidence interval.
How to interpret:	The probability that Θ is in the interval is 0.95	Before taking sample: · The probability that Θ is in the random interval [A, B] is 0.95 After taking sample: · Θ is either in the interval [a, b] or not, and we don't know for sure whether or not it is. · for 95% of samples, the interval calculated based on that sample would contain Θ.