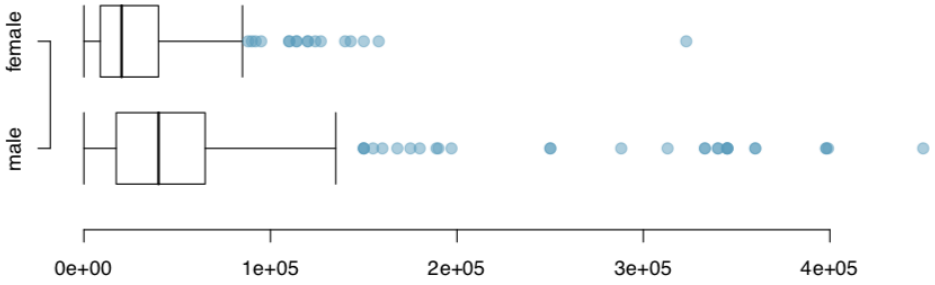
Week 5 Day 4

Stat140-04

Example 1: Gender gap in salaries

Since 2005, the American Community Survey polls approximately 3.5 million households yearly. The following summarizes distribution of salaries of males and females from a random sample of individuals who responded to the 2012 ACS:

	$\bar{x}$	s	n
male	55,890	68,767.88	470
female	29,240	32,025.98	373



(a) What is the population parameter of interest?

(b) Check the conditions for inference with these data

- Randomization Condition: do data come from a random sample or suitably randomized experiment?

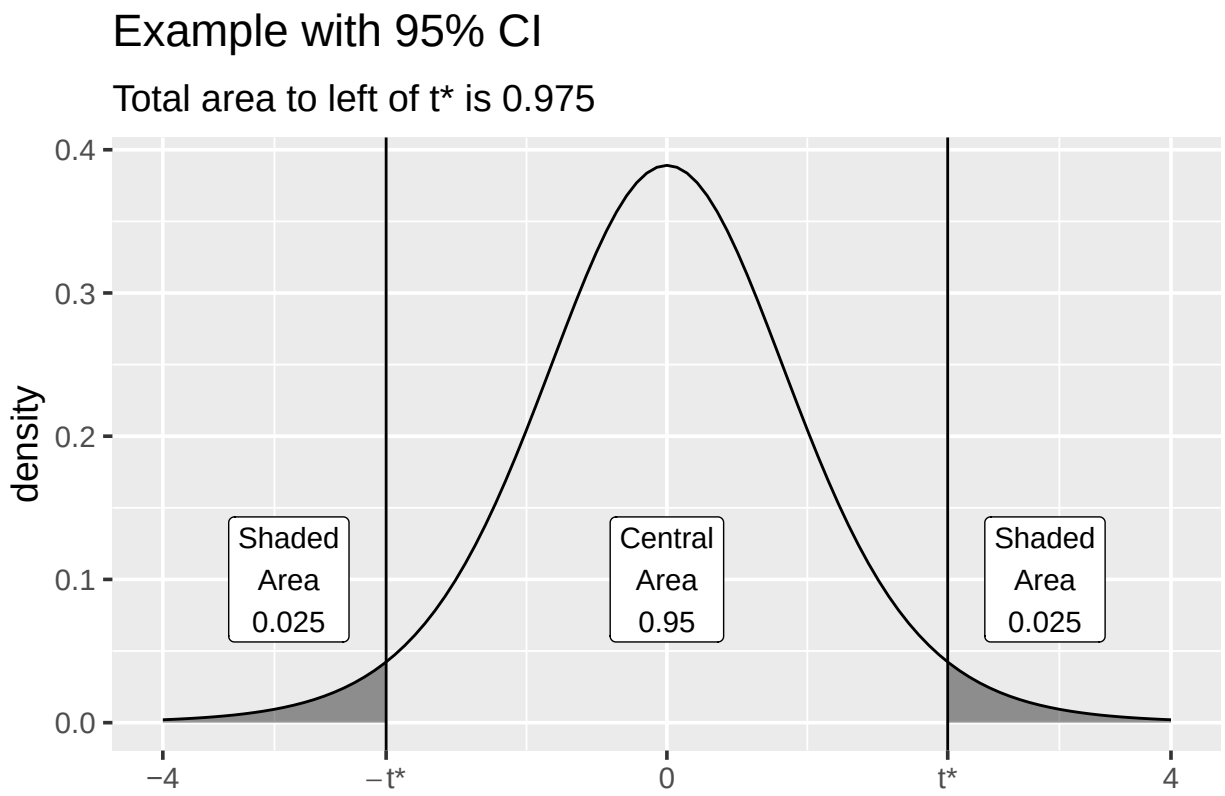
- Nearly normal condition: do data come from a distribution that is unimodal and symmetric.

- Independent Groups Assumption: are the two groups independent?

(c) Construct the 95% onfidence interval and interpret the interval in the context of this question. You may find the note on finding the critical value on the next page useful.

Note on finding the critical value

The **critical value** is t^*: 0.975th quantile of the t_{n-1} distribution.



In R, to look up t^*:

```
qt(0.975, df = 10) # For a 95% CI, sample size is n = 11
```

```
## [1] 2.228139
```

Important things:

- For a 95% CI, the first argument to `qt` is 0.975, not 0.95!
- The second argument to `qt` is $n - 1$.

Example 2: Zinc in water

Trace metals in drinking water affect the flavor and an unusually high concentration can pose a health hazard. Ten pairs of data were taken measuring zinc concentration in bottom water and surface water at 10 randomly sampled locations.

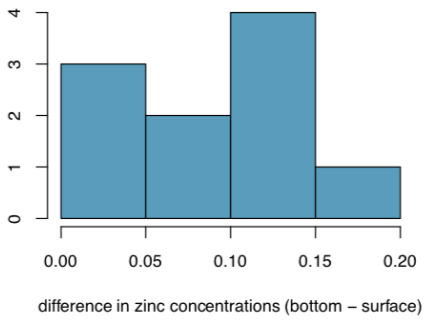
Location	bottom	surface
1	0.43	0.415
2	0.266	0.238
3	0.567	0.39
4	0.531	0.41
5	0.707	0.605
6	0.716	0.609
7	0.651	0.632
8	0.589	0.523
9	0.469	0.411
10	0.723	0.612

(a) What is the population parameter of interest?

(b) How is this example different from Example 2? This is an example of **paired data**:

- We have two measurements on each location (these are **not independent**!)
- We are interested in the **difference** between these measurements
- These **differences are independent** across different locations

Location	bottom	surface	difference
1	0.43	0.415	0.015
2	0.266	0.238	0.028
3	0.567	0.39	0.177
4	0.531	0.41	0.121
5	0.707	0.605	0.102
6	0.716	0.609	0.107
7	0.651	0.632	0.019
8	0.589	0.523	0.066
9	0.469	0.411	0.058
10	0.723	0.612	0.111



(c) Check the conditions for inference with these data

- Randomization Condition: do data come from a random sample or suitably randomized experiment?

- Nearly normal condition: do data come from a distribution that is unimodal and symmetric.

	$\bar{x}$	s	n
bottom	0.5649	0.1468	10
surface	0.4845	0.1312	10
diff	0.0804	0.0523	10

(d) Construct the 95% onfidence interval and interpret the confidence interval in the proper context The formula for making the confidence interval is given below

$$\bar{x}_{\text{diff}} \pm t^* \sqrt{\frac{s^2_{\text{diff}}}{n_{\text{diff}}}}$$

where t^* is the 0.975 quartile of the t-distribution with a degree of freedom

$$df = \min(n_{\text{diff}} - 1).$$

Example 3: Horror films

The slasher horror film has been deplored based on claims that it depicts eroticized violence against predominately female characters as punishment for sexual activities. To test this assertion, a quantitative content analysis was conducted to examine the extent to which gender differences are evident in the association between character survival and engagement in sexual activities. Information pertaining to gender, engagement in sexual activities, and survival was coded for film characters from a simple random sample of 50 English-language, North American slasher films released between 1960 and 2009.[1]

Based on the data provided below, is survival for **male** characters in slasher films associated with sexual activity? If yes, quantify this association using a confidence interval. Make sure to check the necessary conditions.

Gender	Sexual activity	Outcome of physical aggression		*n*
		Survival	Death	
Female				
	Present	13.3% (*n*=11)	86.7% (*n*=72)	83
	Absent	28.1% (*n*=39)	71.9% (*n*=100)	139
Male				
	Present	9.5% (*n*=7)	90.5% (*n*=67)	74
	Absent	14.8% (*n*=28)	85.2% (*n*=161)	189

[1]Welsh, Andrew. "On the perils of living dangerously in the slasher horror film: Gender differences in the association between sexual activity and survival." Sex Roles 62.11-12 (2010): 762-773.