

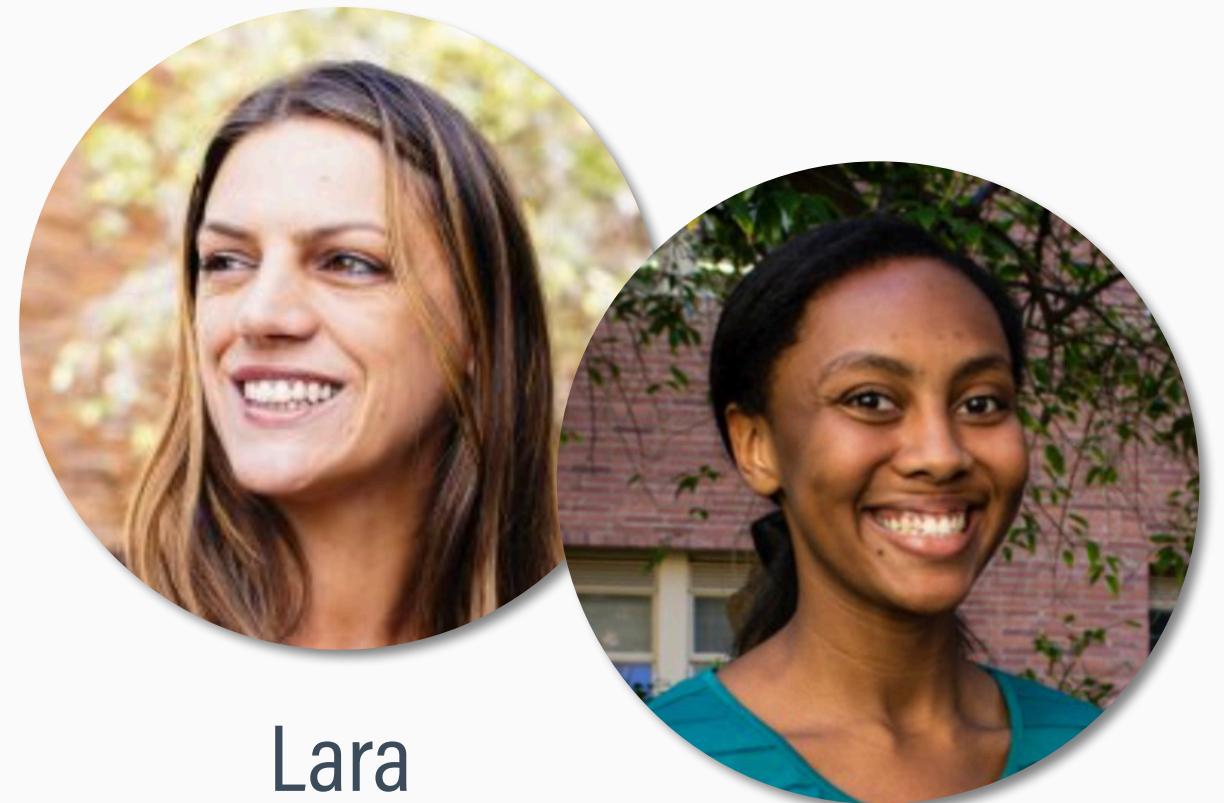
# LAB WEEK 4

# CONFIDENCE INTERVALS

# SMOKING AND DRINKING CESSATION TRIAL

---

Pharmacological treatments that can concomitantly address cigarette smoking and heavy drinking stand to improve health care delivery for these highly prevalent co-occurring conditions. This superiority trial compared the combination of varenicline and naltrexone against varenicline alone for smoking cessation and drinking reduction among heavy-drinking smokers.



Lara  
Ray

ReJoyce  
Green

Ray, L.A., Green, R., Enders, C., et al. (2021). Efficacy of combining varenicline and naltrexone for smoking cessation and drinking reduction: A randomized clinical trial. *American Journal of Psychiatry*, 178, 818–828.

# LOAD PACKAGES AND IMPORT DATA

---

- = data frame name
- = variable name
- = raw data file name

```
# LOAD R PACKAGES ----  
  
# load R packages  
library(ggplot2)  
library(Hmisc)  
library(psych)  
library(summarytools)  
  
# READ DATA ----  
  
# github url for raw data  
filepath <-  
  'https://raw.githubusercontent.com/craigenders/psych250a/main/data/ClinicalTrialData.csv'  
  
# create data frame called ClinicalTrial from github data  
ClinicalTrial <- read.csv(filepath, stringsAsFactors = T)
```

# SUMMARIZING DATA

---

- = data frame name
- = variable name

```
# INSPECT DATA ----
```

```
# summarize entire data frame (summarytools package)
dfSummary(ClinicalTrial)
```

```
# DESCRIPTIVE STATISTICS ----
```

```
# descriptive statistics for entire data frame (psych package)
# both psych and Hisc have describe functions
# psych::describe specifies the psych package's describe function
psych::describe(ClinicalTrial)
```

# R OUTPUT

---

Data Frame Summary

ClinicalTrial

Dimensions: 165 x 2

Duplicates: 125

No	Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid	Missing
1	Condition [factor]	1. Varenicline 2. Varenicline + Naltrexone	82 (49.7%) 83 (50.3%)	IIIIIIIII IIIIIIIIII	165 (100.0%)	0 (0.0%)
2	COWeek8 [integer]	Mean (sd) : 5.5 (6) min < med < max: 0 < 3 < 29 IQR (CV) : 8 (1.1)	25 distinct values	:	165 (100.0%)	0 (0.0%)

# R OUTPUT

---

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Participant	1	165	83.00	47.78	83	83.00	60.79	1	165	164	0.00	-1.22	3.72
Condition*	2	165	1.50	0.50	2	1.50	0.00	1	2	1	-0.01	-2.01	0.04
Gender*	3	165	1.61	0.49	2	1.63	0.00	1	2	1	-0.43	-1.83	0.04
COWeek0	4	165	10.55	6.86	9	9.82	5.93	0	46	46	1.41	3.60	0.53
COWeek4	5	165	5.46	5.10	4	4.69	4.45	0	24	24	1.24	0.96	0.40
COWeek8	6	165	5.53	5.96	3	4.55	2.97	0	29	29	1.75	3.35	0.46
QuitCigsWeek4*	7	165	1.62	0.49	2	1.65	0.00	1	2	1	-0.48	-1.78	0.04
QuitCigsWeek8*	8	165	1.64	0.48	2	1.68	0.00	1	2	1	-0.59	-1.66	0.04
DrinksWeek0	9	165	6.40	4.42	6	5.77	2.97	1	35	34	2.44	10.46	0.34
DrinksWeek4	10	165	3.59	2.98	3	3.30	2.97	0	13	13	0.80	0.30	0.23
DrinksWeek8	11	165	3.23	2.68	3	2.97	2.97	0	13	13	0.89	0.67	0.21
CigsWeek0	12	165	14.22	8.22	12	13.05	5.93	3	51	48	1.59	3.20	0.64
CigsWeek4	13	165	4.18	5.70	2	3.08	2.97	0	41	41	3.01	13.13	0.44
CigsWeek8	14	165	3.16	4.75	2	2.25	2.97	0	35	35	3.44	16.81	0.37

# 95% CONFIDENCE INTERVAL

---

- = data frame name
- = variable name

```
# 95% CONFIDENCE INTERVAL ----  
  
# 95% confidence interval (Hisc package)  
smean.cl.normal(ClinicalTrial$COWeek8)
```

# R OUTPUT

---

Mean	Lower	Upper
5.527273	4.611307	6.443238

# 95% CONFIDENCE INTERVALS BY GROUP

---

- = data frame name
- = variable name
- = grouping variable

```
# 95% CONFIDENCE INTERVAL BY GROUP ----  
  
# 95% confidence interval by group (Hisc package)  
with(ClinicalTrial, tapply(COWeek8, Condition, smean.cl.normal))  
  
# means and 95% confidence interval error bars by group (ggplot2 package)  
ggplot(ClinicalTrial, aes(x = Condition, y = COWeek8)) +  
  stat_summary(fun.data = mean_cl_normal, geom = 'pointrange')
```

# R OUTPUT

---

\$Varenicline

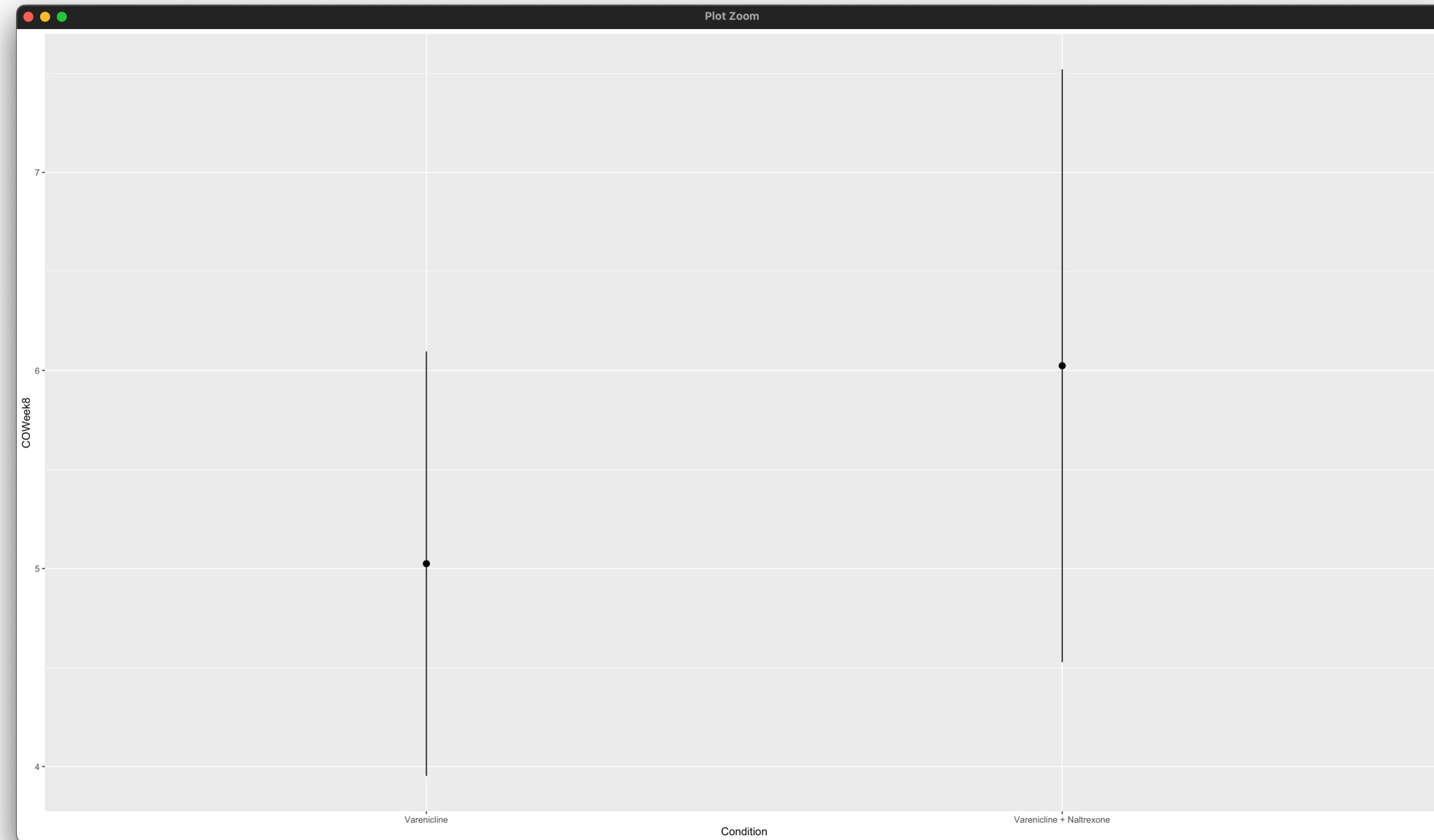
Mean	Lower	Upper
5.024390	3.952419	6.096362

\$`Varenicline + Naltrexone`

Mean	Lower	Upper
6.024096	4.527132	7.521061

# R OUTPUT

---





## SMALL GROUP EXERCISE

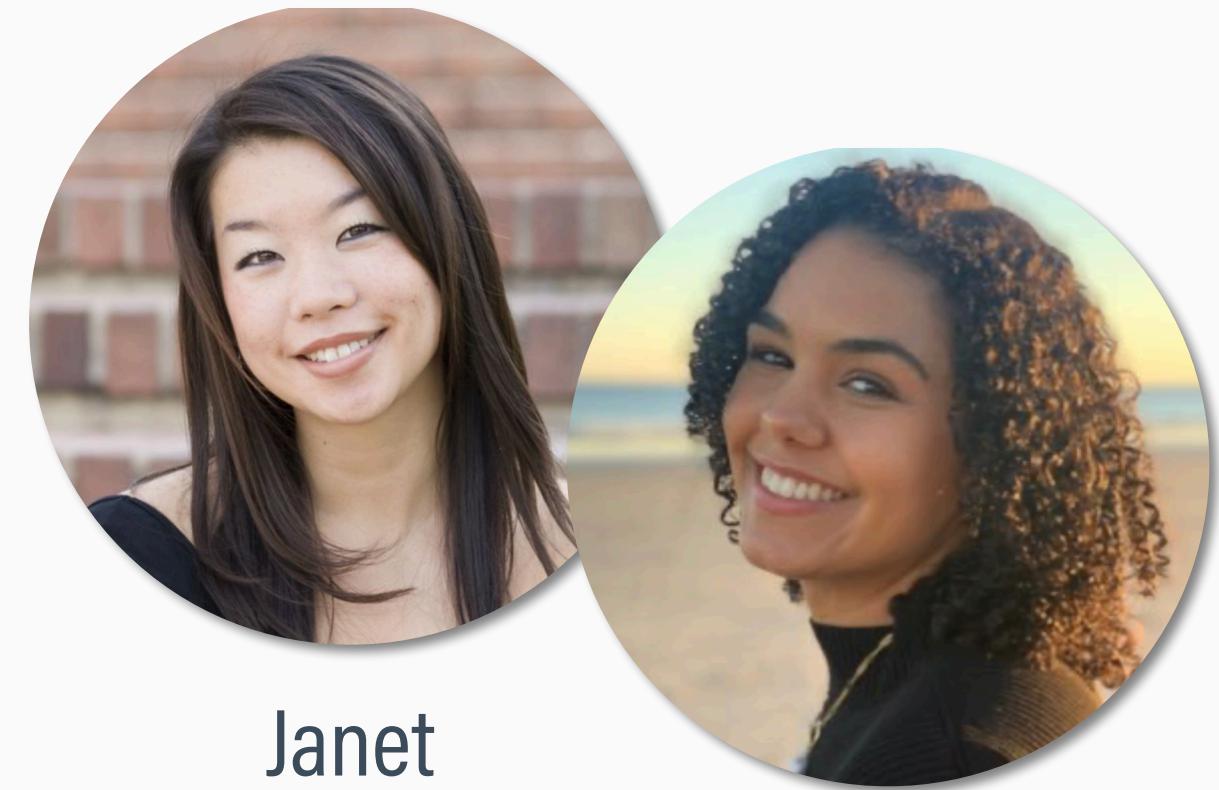
Download two files from Bruin Learn: "Week 4 Lab. Confidence Intervals.R" and "Week 4 Small Group Exercise.R". The Lab script contains the R code we just discussed. The Exercise script contains only the URL for a different data set, BodySatData.csv. In groups of two or three, you will complete a series of R tasks that provide practice for the next assignment. There is no need to write code from scratch; instead, you can copy and paste code chunks from the Lab file into your Exercise script, modifying the data and variable names as needed. The BodySatData.csv file for this exercise contains body satisfaction data from two age cohorts.

# SKIN COLOR SATISFACTION AND BINGE EATING

---

Although it has been demonstrated that (a) body dissatisfaction and internalization of societal appearance standards contribute to disordered eating and (b) that internalization of societal appearance standards leads to decreased skin color satisfaction among Black women, it has not been established whether skin color dissatisfaction contributes to disordered eating among Black women or girls. The objective of the present study is to determine the influence of skin color satisfaction as a potential predictor for binge eating, and its effect through body image in Black girls during the vulnerable developmental period of adolescence.

Parker, J.E., Enders, C.K., Mujahid, M.S., Laraia, B.A., Epel, E.S., Tomiyama, A.J. (2022). Prospective relationships between skin color satisfaction, body satisfaction, and binge eating in Black girls. *Body Image*, 41, 342-353.



Janet  
Tomiyama

Jordan  
Parker



## SMALL GROUP EXERCISE TASK 1

- Use the provided URL to import the BodySatData.csv file into an R data frame (import method #3 from the Week 0 lab script).
- Use the dfSummary function to get numeric and visual summaries of the data frame's variables.



## SMALL GROUP EXERCISE TASK 2

- Use the describe function to get descriptive statistics for the entire sample.
- Consider the standard error for BodySat variable. Interpret the standard error of the mean. Explain what this value tells you about the precision of your estimated mean difference. Relate it to the concept of sampling variability in repeated studies.



## SMALL GROUP EXERCISE TASK 3

- Use the `smean.cl.normal` function in the `Hmisc` package to get the 95% confidence interval for the BodySat mean.
- Interpret the 95% confidence interval. Explain what the lower and upper numeric bounds represent in the context of this study. Then, describe what it means to be “95% confident” from a frequentist perspective.



## SMALL GROUP EXERCISE TASK 4

- Get the 95% confidence intervals for each age cohort (10 versus 18 year olds).
- Use the ggplot function to graph the means and 95% confidence intervals by age.
- Apply the “inference by eye” rule of thumb from Cumming and Finch (2005). What conclusion can you draw about the differences in body satisfaction at ages 10 and 18? Explain your rationale.