Week #1

Definitions:

- Population
- Sample
- Parameter
- Statistic
- Descriptive statistics
- Inferential statistics
- Independent variable
- Dependent variable

Questions:

- Give examples of the above

Summing notation

X_1 = the first element of X
X_i = an arbitrary element of X
X_n = the last element of X
$\Sigma(X_i) = X_1 + X_2 + ... + X_n$
$\Sigma(X_i + 1) = (X_1 + 1) + (X_2 + 1) + ... + (X_n + 1)$
$\Sigma(X_i - Y_i) = (X_1 - Y_1) + (X_2 - Y_2) + ... + (X_n - Y_n)$
$[\Sigma(X_i)]^2 = (X_1 + X_2 + ... + X_n)^2$

Question #1

X = 3 4 5 7 2
Y = 9 10 6 5 1

$\Sigma(X_i) = 21$
$[\Sigma(X_i)]^2 = 441$
$\Sigma(X_i - Y_i) = -10$
$\Sigma(X_iY_i) = 134$
$\Sigma(X_i^2) = 103$
$[\Sigma(X_i - Y_i)]^2 = 100$
$\Sigma(X_i)\Sigma(Y_i) = 651$

Question #2

X = 2 7 5 10 6
Y = 10 2 7 4 8

$\Sigma(X_i) = 30$
$[\Sigma(X_i)]^2 = 900$
$\Sigma(X_i - Y_i) = -1$
$\Sigma(X_iY_i) = 157$
$\Sigma(X_i^2) = 214$
$[\Sigma(X_i - Y_i)]^2 = 1$
$\Sigma(X_i)\Sigma(Y_i) = 930$

Question #3

X = 2 8 4 3 9
Y = 7 2 3 5 4

$\Sigma(X_i) = 26$
$[\Sigma(X_i)]^2 = 676$
$\Sigma(X_i - Y_i) = 5$
$\Sigma(X_iY_i) = 93$
$\Sigma(X_i^2) = 174$
$[\Sigma(X_i - Y_i)]^2 = 25$
$\Sigma(X_i)\Sigma(Y_i) = 546$

Question #4

X = 6 1 3 2 5
Y = 3 7 8 9 1

$\Sigma(X_i) = 17$
$[\Sigma(X_i)]^2 = 289$
$\Sigma(X_i - Y_i) = -11$
$\Sigma(X_iY_i) = 72$
$\Sigma(X_i^2) = 75$
$[\Sigma(X_i - Y_i)]^2 = 121$
$\Sigma(X_i)\Sigma(Y_i) = 476$

Question #5

X = 3 7 8 2 1
Y = 8 5 7 10 1

$\Sigma(X_i) = 21$
$[\Sigma(X_i)]^2 = 441$
$\Sigma(X_i - Y_i) = -10$
$\Sigma(X_iY_i) = 136$
$\Sigma(X_i^2) = 127$
$[\Sigma(X_i - Y_i)]^2 = 100$
$\Sigma(X_i)\Sigma(Y_i) = 651$

Question #6

X = 7 9 3 8 5
Y = 10 9 7 4 1

$\Sigma(X_i) = 32$
$[\Sigma(X_i)]^2 = 1024$
$\Sigma(X_i - Y_i) = 1$
$\Sigma(X_iY_i) = 209$
$\Sigma(X_i^2) = 228$
$[\Sigma(X_i - Y_i)]^2 = 1$
$\Sigma(X_i)\Sigma(Y_i) = 992$

Week #2

Definitions:

- Discrete
- Continuous
- Nominal
- Ordinal
- Interval
- Ratio

Questions:

- Give some examples of the above
- Draw a good bar graph and a bad one, and list the differences

Frequency table construction

For the following data sets, create a frequency table containing the frequency, cumulative frequency, relative frequency, and cumulative relative frequency of each value.

Freq. = number equal to score

C Freq. = number less than or equal to score

R Freq. = frequency divided by sample size

CR Freq. = cumulative frequency divided by n

Question #1

-1, 0, -1, 2, 0, -1, 0, 1, 1, 0, 2, 0

score	freq	cfreq	rfreq	crfreq
-1	3	3	0.25	0.25
0	5	8	0.42	0.67
1	2	10	0.17	0.83
2	2	12	0.17	1.00

Question #2

-1, 0, 2, -1, 0, 0, 1, 0, 2, 0, 0, 1

score	freq	cfreq	rfreq	crfreq
-1	2	2	0.17	0.17
0	6	8	0.50	0.67
1	2	10	0.17	0.83
2	2	12	0.17	1.00

Question #3

-1, 0, 0, -1, 0, 0, 0, 1, -1, 1, -1, -1

score	freq	cfreq	rfreq	crfreq
-1	5	5	0.42	0.42
0	5	10	0.42	0.83
1	2	12	0.17	1.00

Question #4

0, -1, 1, 1, 2, 1, -1, 0, 2, 2, 1, 0

score	freq	cfreq	rfreq	crfreq
-1	2	2	0.17	0.17
0	3	5	0.25	0.42
1	4	9	0.33	0.75
2	3	12	0.25	1.00

Question #5

-1, 1, -1, 0, 2, -1, 0, -1, 0, 0, 1, -1

score	freq	cfreq	rfreq	crfreq
-1	5	5	0.42	0.42
0	4	9	0.33	0.75
1	2	11	0.17	0.92
2	1	12	0.08	1.00

Question #6

0, -1, 1, 2, 0, 0, -1, 1, 0, -1, 2, -1

score	freq	cfreq	rfreq	crfreq
-1	4	4	0.33	0.33
0	4	8	0.33	0.67
1	2	10	0.17	0.83
2	2	12	0.17	1.00

Question #7

2, -1, -1, 0, -1, -1, 1, 0, 0, 2, 0, 3

score	freq	cfreq	rfreq	crfreq
-1	4	4	0.33	0.33
0	4	8	0.33	0.67
1	1	9	0.08	0.75
2	2	11	0.17	0.92
3	1	12	0.08	1.00

Question #8

0, 1, 0, -1, 1, 0, 0, -1, -3, -1, -1, 0

score	freq	cfreq	rfreq	crfreq
-3	1	1	0.08	0.08
-1	4	5	0.33	0.42
0	5	10	0.42	0.83
1	2	12	0.17	1.00

Question #9

-1, -1, 0, 0, 0, -1, 1, 0, 0, 0, 1, 0

score	freq	cfreq	rfreq	crfreq
-1	3	3	0.25	0.25
0	7	10	0.58	0.83
1	2	12	0.17	1.00

Question #10

0, 0, -1, -1, 0, 0, -1, 0, -2, 0, 1, 1

score	freq	cfreq	rfreq	crfreq
-2	1	1	0.08	0.08
-1	3	4	0.25	0.33
0	6	10	0.50	0.83
1	2	12	0.17	1.00

Question #11

-1, 0, -2, -1, 1, -1, 1, 1, 0, -1, -1, 0

score	freq	cfreq	rfreq	crfreq
-2	1	1	0.08	0.08
-1	5	6	0.42	0.50
0	3	9	0.25	0.75
1	3	12	0.25	1.00

Question #12

-1, 2, -1, -1, -2, 0, 0, -1, 0, 0, -1, -1

score	freq	cfreq	rfreq	crfreq
-2	1	1	0.08	0.08
-1	6	7	0.50	0.58
0	4	11	0.33	0.92
2	1	12	0.08	1.00

Question #13

1, 0, 2, 0, 1, 0, 1, 0, 0, 1, -1, 0

score	freq	cfreq	rfreq	crfreq
-1	1	1	0.08	0.08
0	6	7	0.50	0.58
1	4	11	0.33	0.92
2	1	12	0.08	1.00

Question #14

-1, 2, 2, 1, 0, 1, 0, 1, 0, 1, 0, 0

score	freq	cfreq	rfreq	crfreq
-1	1	1	0.08	0.08
0	5	6	0.42	0.50
1	4	10	0.33	0.83
2	2	12	0.17	1.00

Question #15

0, 2, 0, 1, 0, -1, 0, 1, 0, -1, 1, 0

score	freq	cfreq	rfreq	crfreq
-1	2	2	0.17	0.17
0	6	8	0.50	0.67
1	3	11	0.25	0.92
2	1	12	0.08	1.00

Question #16

0, 0, 1, -1, 1, 0, -1, 0, 1, 1, 2, 0

score	freq	cfreq	rfreq	crfreq
-1	2	2	0.17	0.17
0	5	7	0.42	0.58
1	4	11	0.33	0.92
2	1	12	0.08	1.00

Question #17

-1, 0, 0, -1, 1, 0, 1, 2, 0, 0, 1, 1

score	freq	cfreq	rfreq	crfreq
-1	2	2	0.17	0.17
0	5	7	0.42	0.58
1	4	11	0.33	0.92
2	1	12	0.08	1.00

Question #18

1, 2, -2, 0, 0, 0, -1, 2, 0, 2, -2, 1

score	freq	cfreq	rfreq	crfreq
-2	2	2	0.17	0.17
-1	1	3	0.08	0.25
0	4	7	0.33	0.58
1	2	9	0.17	0.75
2	3	12	0.25	1.00

Question #19

-1, 0, 0, -1, 1, 0, 1, -1, 0, 0, 1, 1

score	freq	cfreq	rfreq	crfreq
-1	3	3	0.25	0.25
0	5	8	0.42	0.67
1	4	12	0.33	1.00

Interval construction

For each scenario, find the lower limit, midpoint, and upper limit of the first five intervals.

Question #1

The lowest value in the data set is 38, and the desired interval width is 14.

LL	MP	UL
28	34.5	41
42	48.5	55
56	62.5	69
70	76.5	83
84	90.5	97

Question #2

The lowest value in the data set is 71, and the desired interval width is 10.

LL	MP	UL
70	74.5	79
80	84.5	89
90	94.5	99
100	104.5	109
110	114.5	119

Question #3

The lowest value in the data set is 81, and the desired interval width is 9.

LL	MP	UL
81	85.0	89
90	94.0	98
99	103.0	107
108	112.0	116
117	121.0	125

Question #4

The lowest value in the data set is 2, and the desired interval width is 30.

LL	MP	UL
0	14.5	29
30	44.5	59
60	74.5	89
90	104.5	119
120	134.5	149

Question #5

The lowest value in the data set is 69, and the desired interval width is 11.

LL	MP	UL
66	71.0	76
77	82.0	87
88	93.0	98
99	104.0	109
110	115.0	120

Question #6

The lowest value in the data set is 94, and the desired interval width is 31.

LL	MP	UL
93	108.0	123
124	139.0	154
155	170.0	185
186	201.0	216
217	232.0	247

Question #7

The lowest value in the data set is 40, and the desired interval width is 49.

LL	MP	UL
0	24.0	48
49	73.0	97
98	122.0	146
147	171.0	195
196	220.0	244

Question #8

The lowest value in the data set is 22, and the desired interval width is 24.

LL	MP	UL
0	11.5	23
24	35.5	47
48	59.5	71
72	83.5	95
96	107.5	119

Question #9

The lowest value in the data set is 3, and the desired interval width is 12.

LL	MP	UL
0	5.5	11
12	17.5	23
24	29.5	35
36	41.5	47
48	53.5	59

Question #10

The lowest value in the data set is 31, and the desired interval width is 26.

LL	MP	UL
26	38.5	51
52	64.5	77
78	90.5	103
104	116.5	129
130	142.5	155

Question #11

The lowest value in the data set is 1, and the desired interval width is 15.

LL	MP	UL
0	7.0	14
15	22.0	29
30	37.0	44
45	52.0	59
60	67.0	74

Question #12

The lowest value in the data set is 82, and the desired interval width is 4.

LL	MP	UL
80	81.5	83
84	85.5	87
88	89.5	91
92	93.5	95
96	97.5	99

Question #13

The lowest value in the data set is 25, and the desired interval width is 36.

LL	MP	UL
0	17.5	35
36	53.5	71
72	89.5	107
108	125.5	143
144	161.5	179

Question #14

The lowest value in the data set is 64, and the desired interval width is 13.

LL	MP	UL
52	58.0	64
65	71.0	77
78	84.0	90
91	97.0	103
104	110.0	116

Week #3

Definitions:

- Mean
- Median
- Mode
- Variance
- Standard deviation

Questions:

- Why is the mean more sensitive to extreme scores than the median?
- Why might we prefer the median over the mean?
- What are we squaring in a sum of squares?
- What is the point of the squaring in sum of squares?
- In what sense is a variance a mean?
- When would variance be equal to zero?
- Why do we use standard deviation instead of variance?
- Why do we using $n-1$ instead of n in the denominator of variance?
- Why is the standard deviation so named?
- For the following data set, calculate the mode, median, mean, and standard deviation. Then, add 5 to every number in the data set and calculate again. Then, multiply every number in the data set by 2 and calculate again. What effect does adding/multiplying every score with the same number have, and why? Explain for each statistic.

$$2, 8, 1, 1, 3$$

Median calculation

For each table, calculate the median using this formula:

$$Md = LL + W\left[\frac{0.5(n) - cumF}{fm}\right]$$

LL = (score of the row with the lowest CR freq. ≥ 0.5) - 0.5
W = interval width (= 1 if the data are ungrouped)
n = sample size (= the highest C freq.)
$cumF$ = CR freq. of the row below the one containing LL
fm = frequency of the row containing LL

Question #1

score	freq	cfreq	rfreq	crfreq
14	1	1	0.08	0.08
15	9	10	0.75	0.83
16	2	12	0.17	1.00

Median $= 14.5 + 1\left[\frac{0.5(12)-1}{9}\right] = 15.06$

Question #2

score	freq	cfreq	rfreq	crfreq
9	2	2	0.17	0.17
10	5	7	0.42	0.58
11	3	10	0.25	0.83
12	2	12	0.17	1.00

Median $= 9.5 + 1\left[\frac{0.5(12)-2}{5}\right] = 10.3$

Question #3

score	freq	cfreq	rfreq	crfreq
13	1	1	0.08	0.08
14	2	3	0.17	0.25
15	5	8	0.42	0.67
16	4	12	0.33	1.00

Median $= 14.5 + 1\left[\frac{0.5(12)-3}{5}\right] = 15.1$

Question #4

score	freq	cfreq	rfreq	crfreq
9	2	2	0.17	0.17
10	2	4	0.17	0.33
11	6	10	0.50	0.83
12	2	12	0.17	1.00

Median $= 10.5 + 1\left[\frac{0.5(12)-4}{6}\right] = 10.83$

Question #5

score	freq	cfreq	rfreq	crfreq
14	5	5	0.42	0.42
15	5	10	0.42	0.83
16	2	12	0.17	1.00

Median $= 14.5 + 1\left[\frac{0.5(12)-5}{5}\right] = 14.7$

Question #6

score	freq	cfreq	rfreq	crfreq
12	4	4	0.33	0.33
13	5	9	0.42	0.75
14	3	12	0.25	1.00

Median $= 12.5 + 1\left[\frac{0.5(12)-4}{5}\right] = 12.9$

Question #7

score	freq	cfreq	rfreq	crfreq
13	1	1	0.08	0.08
14	4	5	0.33	0.42
15	4	9	0.33	0.75
16	1	10	0.08	0.83
17	2	12	0.17	1.00

Median $= 14.5 + 1\left[\frac{0.5(12)-5}{4}\right] = 14.75$

Standard deviation calculation

$$SS = \Sigma(X_i - \bar{X})^2$$
$$df = n - 1$$
$$s^2 = \frac{SS}{df}$$
$$s = \sqrt{s}$$

Question #1

4, 0, 4, 9, 3

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
4	0	0
0	-4	16
4	0	0
9	5	25
3	-1	1

$\bar{X} = 4$
$SS = 42$
$df = 5 - 1 = 4$

$s^2 = 42/4 = 10.5$
$s = \sqrt{10.5} = 3.24$

Question #2

10, 3, 1, 2, 9

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
10	5	25
3	-2	4
1	-4	16
2	-3	9
9	4	16

$\bar{X} = 5$
$SS = 70$
$df = 5 - 1 = 4$

$s^2 = 70/4 = 17.5$
$s = \sqrt{17.5} = 4.18$

Question #3

5, 5, 5, 6, 9

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
5	-1	1
5	-1	1
5	-1	1
6	0	0
9	3	9

$\bar{X} = 6$
$SS = 12$
$df = 5 - 1 = 4$

$s^2 = 12/4 = 3$
$s = \sqrt{3} = 1.73$

Question #4

6, 0, 3, 3, 8

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
6	2	4
0	-4	16
3	-1	1
3	-1	1
8	4	16

$\bar{X} = 4$
$SS = 38$
$df = 5 - 1 = 4$

$s^2 = 38/4 = 9.5$
$s = \sqrt{9.5} = 3.08$

Question #5

3, 5, 3, 6, 3

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
3	-1	1
5	1	1
3	-1	1
6	2	4
3	-1	1

$\bar{X} = 4$
$SS = 8$
$df = 5 - 1 = 4$

$s^2 = 8/4 = 2$
$s = \sqrt{2} = 1.41$

Question #6

6, 9, 3, 4, 8

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
6	0	0
9	3	9
3	-3	9
4	-2	4
8	2	4

$\bar{X} = 6$
$SS = 26$
$df = 5 - 1 = 4$

$s^2 = 26/4 = 6.5$
$s = \sqrt{6.5} = 2.55$

Week #4

Definitions:

- z score
- Normal distribution

Questions:

- Why do we transform scores into standard scores?
- Draw these scores on a number line:

  $$4, 6, 4, 3, 8$$

  Then, subtract the mean from each score and draw the result on another number line. Then, divide those by the standard deviation and draw them on yet another number line. What does each step do?
- How does the height of the normal curve correspond to the number line below it?

Z-scores

Calculate the area between the following z scores: -0.63 and 1.12, 0.18 and -0.04, -0.84 and -0.02, 1.6 and 0.94, 0.33 and 0.82, -0.82 and 0.59, 0.49 and 0.92, 0.74 and 0.78, 0.58 and 0.07, -0.31 and -1.99, 1.51 and 0.62, 0.39 and -0.06, -0.62 and -0.16, -2.21 and -1.47

z	Area above z	Area between mean and z
0.02	0.4920	0.0080
0.04	0.4840	0.0160
0.06	0.4761	0.0239
0.07	0.4721	0.0279
0.16	0.4364	0.0636
0.18	0.4286	0.0714
0.31	0.3783	0.1217
0.33	0.3707	0.1293
0.39	0.3483	0.1517
0.49	0.3121	0.1879
0.58	0.2810	0.2190
0.59	0.2776	0.2224
0.62	0.2676	0.2324
0.62	0.2676	0.2324
0.63	0.2643	0.2357
0.74	0.2296	0.2704
0.78	0.2177	0.2823
0.82	0.2061	0.2939
0.82	0.2061	0.2939
0.84	0.2005	0.2995
0.92	0.1788	0.3212
0.94	0.1736	0.3264
1.12	0.1314	0.3686
1.47	0.0708	0.4292
1.51	0.0655	0.4345
1.60	0.0548	0.4452
1.99	0.0233	0.4767
2.21	0.0136	0.4864

z_1	z_2	Area between z_1 and z_2
-0.63	1.12	0.60
0.18	-0.04	0.09
-0.84	-0.02	0.29
1.60	0.94	0.12
0.33	0.82	0.16
-0.82	0.59	0.52
0.49	0.92	0.13
0.74	0.78	0.01
0.58	0.07	0.19
-0.31	-1.99	0.35
1.51	0.62	0.20
0.39	-0.06	0.18
-0.62	-0.16	0.17
-2.21	-1.47	0.06

Week #5

Definitions

- Sampling distribution
- Standard error
- Law of large numbers?

Questions

- What is the difference between standard error and standard deviation?
- Why is standard error so named?
- Why can't we just draw a sample from the population and estimate the parameter by the sample statistic?
- Give an example of a sampling distribution
- What is the difference between a sampling distribution and a sampling distribution of the mean?
- What is the difference between standard error and standard error of the mean?
- Imagine a bowl with five bingo chips in it, numbered 1 through 5. For this "population":
  - List all 25 possible samples of size 2 (sampling with replacement)
  - Calculate the mean for every sample
  - Count the frequency of each value of the mean
  - Draw the frequency distribution as a bar plot
  - Calculate the probability of drawing a sample with:
    - A mean $>=$ 4
    - A mean $<=$ 2.5
    - An error of at least 1.5

Week #6

Definitions:

- Null hypothesis
- Alternate hypothesis
- Alpha (α)
- Beta (β)
- Power
- Type I error
- Type II error
- Null distribution (H_0)
- Alternate distribution (H_1)
- p-value

Questions:

- Draw a null distribution and illustrate the relationship between the critical values and α
- Draw an alternate distribution and illustrate the relationship between the critical values, β, and power
- Draw a distribution and illustrate the relationship between the p-value and z_{obs}
- How do we control Type I error? Type II error?
- How is β affected by change in effect size, α, σ, and sample size?
- How is α affected by change in σ and sample size?
- H_0 and H_1 must be exhaustive (one must be true) and mutually exclusive (they can't both be true). Come up with some invalid hypotheses
- What is wrong with saying "p is the probability that the null hypothesis is true"?
- What is the goal of a confidence interval?
- Why is calculating a 1 - α% confidence interval the same as performing a hypothesis test at α?
- What do we need to know in order to use a z test?
- Give an example of a finding which is statistically significant but practically insignificant. When might this occur?

Z-tests

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$
$z_{\text{crit}} =$ the z score with $\alpha/2$ above it
$$z_{\text{obs}} = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$
$CI_y = \bar{X} \pm (\sigma_{\bar{X}} \times z_y)$

$z_y =$ the z score with $(100 - y)/100$ above it

Critical z values

z	Area between mean and z	Area above z
1.645	0.45	0.05
1.96	0.475	0.025
2.576	0.495	0.005

Question #1

Researchers draw a sample of 8 with a mean of 5.81. The population variance is known to be 21.44. Test $H_0 : \mu = 4$ at an α of 0.05, state your decision, and calculate a 95% confidence interval.

$\sigma = \sqrt{21.44} = 4.63$

$\sigma_{\bar{X}} = 4.63/\sqrt{8} = 1.64$

$z_{\text{obs}} = (5.81 - 4)/1.64 = 1.1$

$z_{\text{crit}} = \pm 1.96$

Fail to reject because $1.96 > 1.1 > -1.96$

$z_{95} = 1.96$

$CI_{95} = 5.81 \pm (1.64 \times 1.96) = [2.6,\ 9.02]$

Question #2

Researchers draw a sample of 8 with a mean of 2.91. The population variance is known to be 2.79. Test $H_0 : \mu = 8$ at an α of 0.01, state your decision, and calculate a 90% confidence interval.

$\sigma = \sqrt{2.79} = 1.67$

$\sigma_{\bar{X}} = 1.67/\sqrt{8} = 0.59$

$z_{\text{obs}} = (2.91 - 8)/0.59 = -8.63$

$z_{\text{crit}} = \pm 2.58$

Reject because $-8.63 =< -2.58$

$z_{90} = 1.64$

$CI_{90} = 2.91 \pm (0.59 \times 1.64) = [1.94,\ 3.88]$

Question #3

Researchers draw a sample of 7 with a mean of 3.06. The population variance is known to be 5.34. Test $H_0 : \mu = 9$ at an α of 0.05, state your decision, and calculate a 99% confidence interval.

$\sigma = \sqrt{5.34} = 2.31$

$\sigma_{\bar{X}} = 2.31/\sqrt{7} = 0.87$

$z_{\text{obs}} = (3.06 - 9)/0.87 = -6.83$

$z_{\text{crit}} = \pm 1.96$

Reject because $-6.83 =< -1.96$

$z_{99} = 2.58$

$CI_{99} = 3.06 \pm (0.87 \times 2.58) = [0.82,\ 5.3]$

Question #4

Researchers draw a sample of 6 with a mean of 6.82. The population variance is known to be 4.45. Test $H_0 : \mu = 1$ at an α of 0.01, state your decision, and calculate a 95% confidence interval.

$\sigma = \sqrt{4.45} = 2.11$

$\sigma_{\bar{X}} = 2.11/\sqrt{6} = 0.86$

$z_{\text{obs}} = (6.82 - 1)/0.86 = 6.77$

$z_{\text{crit}} = \pm 2.58$

Reject because $6.77 >= 2.58$

$z_{95} = 1.96$

$CI_{95} = 6.82 \pm (0.86 \times 1.96) = [5.13,\ 8.51]$

Question #5

Researchers draw a sample of 10 with a mean of 3.04. The population variance is known to be 4.58. Test $H_0 : \mu = 7$ at an α of 0.1, state your decision, and calculate a 99% confidence interval.

$\sigma = \sqrt{4.58} = 2.14$

$\sigma_{\bar{X}} = 2.14/\sqrt{10} = 0.68$

$z_{\text{obs}} = (3.04 - 7)/0.68 = -5.82$

$z_{\text{crit}} = \pm 1.64$

Reject because $-5.82 =< -1.64$

$z_{99} = 2.58$

$CI_{99} = 3.04 \pm (0.68 \times 2.58) = [1.29,\ 4.79]$

Question #6

Researchers draw a sample of 6 with a mean of 7.31. The population variance is known to be 6.35. Test $H_0 : \mu = 10$ at an α of 0.1, state your decision, and calculate a 99% confidence interval.

$\sigma = \sqrt{6.35} = 2.52$
$\sigma_{\bar{X}} = 2.52/\sqrt{6} = 1.03$
$z_{\text{obs}} = (7.31 - 10)/1.03 = -2.61$
$z_{\text{crit}} = \pm 1.64$
Reject because $-2.61 =< -1.64$
$z_{99} = 2.58$
$CI_{99} = 7.31 \pm (1.03 \times 2.58) = [4.65,\ 9.97]$

Question #7

Researchers draw a sample of 5 with a mean of 9.2. The population variance is known to be 1.64. Test $H_0 : \mu = 4$ at an α of 0.05, state your decision, and calculate a 90% confidence interval.

$\sigma = \sqrt{1.64} = 1.28$
$\sigma_{\bar{X}} = 1.28/\sqrt{5} = 0.57$
$z_{\text{obs}} = (9.2 - 4)/0.57 = 9.12$
$z_{\text{crit}} = \pm 1.96$
Reject because $9.12 >= 1.96$
$z_{90} = 1.64$
$CI_{90} = 9.2 \pm (0.57 \times 1.64) = [8.27,\ 10.13]$

Question #8

Researchers draw a sample of 9 with a mean of 4.19. The population variance is known to be 13.03. Test $H_0 : \mu = 3$ at an α of 0.05, state your decision, and calculate a 99% confidence interval.

$\sigma = \sqrt{13.03} = 3.61$
$\sigma_{\bar{X}} = 3.61/\sqrt{9} = 1.2$
$z_{\text{obs}} = (4.19 - 3)/1.2 = 0.99$
$z_{\text{crit}} = \pm 1.96$
Fail to reject because $1.96 > 0.99 > -1.96$
$z_{99} = 2.58$
$CI_{99} = 4.19 \pm (1.2 \times 2.58) = [1.09,\ 7.29]$

Question #9

Researchers draw a sample of 6 with a mean of 3.29. The population variance is known to be 3.46. Test $H_0 : \mu = 1$ at an α of 0.05, state your decision, and calculate a 95% confidence interval.

$\sigma = \sqrt{3.46} = 1.86$
$\sigma_{\bar{X}} = 1.86/\sqrt{6} = 0.76$
$z_{\text{obs}} = (3.29 - 1)/0.76 = 3.01$
$z_{\text{crit}} = \pm 1.96$
Reject because $3.01 >= 1.96$
$z_{95} = 1.96$
$CI_{95} = 3.29 \pm (0.76 \times 1.96) = [1.8,\ 4.78]$

Question #10

Researchers draw a sample of 7 with a mean of 4.13. The population variance is known to be 14.21. Test $H_0 : \mu = 4$ at an α of 0.01, state your decision, and calculate a 95% confidence interval.

$\sigma = \sqrt{14.21} = 3.77$
$\sigma_{\bar{X}} = 3.77/\sqrt{7} = 1.42$
$z_{\text{obs}} = (4.13 - 4)/1.42 = 0.09$
$z_{\text{crit}} = \pm 2.58$
Fail to reject because $2.58 > 0.09 > -2.58$
$z_{95} = 1.96$
$CI_{95} = 4.13 \pm (1.42 \times 1.96) = [1.35,\ 6.91]$

Question #11

Researchers draw a sample of 8 with a mean of 2.77. The population variance is known to be 1.12. Test $H_0 : \mu = 1$ at an α of 0.1, state your decision, and calculate a 99% confidence interval.

$\sigma = \sqrt{1.12} = 1.06$
$\sigma_{\bar{X}} = 1.06/\sqrt{8} = 0.37$
$z_{\text{obs}} = (2.77 - 1)/0.37 = 4.78$
$z_{\text{crit}} = \pm 1.64$
Reject because $4.78 >= 1.64$
$z_{99} = 2.58$
$CI_{99} = 2.77 \pm (0.37 \times 2.58) = [1.82,\ 3.72]$

Week #7

- When would we use a t test instead of a z test?
- What is the expected value of t under the null?
- What is homogeneity of variance?
- Why can't we calculate a t statistic with a sample size of 1?
- When might we employ a Welch correction?
- Give an example of a research question for which a one-sample t-test would be appropriate

One sample t-tests

$$df = n - 1$$
$$s_{\bar{X}} = \frac{s}{\sqrt{n}}$$
$$t_{\text{obs}} = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$
$$CI_y = \bar{X} \pm (s_{\bar{X}} \times t_y)$$
$$t_y = \text{the critical value for } \alpha = (100 - y)/100$$

Critical t values

	α			
df	0.2	0.1	0.05	0.01
4	1.53	2.13	2.78	4.6
5	1.48	2.02	2.57	4.03
6	1.44	1.94	2.45	3.71

Question #1

Researchers draw a sample of 6 with a mean of 4.91 and a standard deviation of 2.19. Test $H_0 : \mu = 3$ at an α of 0.05, state your decision, then calculate a 90% confidence interval.

$s_{\bar{X}} = 2.19/\sqrt{6} = 0.89$

$t_{\text{tobs}} = (4.91 - 3)/0.89 = 2.15$

$t_{\text{crit}} = \pm 2.57$

Fail to reject because $2.57 > 2.15 > -2.57$

$t_{90} = 2.02$

$CI_{90} = 4.91 \pm (0.89 \times 2.02) = [3.11,\ 6.71]$

Question #2

Researchers draw a sample of 7 with a mean of 4.13 and a standard deviation of 2.31. Test $H_0 : \mu = 7$ at an α of 0.01, state your decision, then calculate a 90% confidence interval.

$s_{\bar{X}} = 2.31/\sqrt{7} = 0.87$

$t_{\text{tobs}} = (4.13 - 7)/0.87 = -3.3$

$t_{\text{crit}} = \pm 3.71$

Fail to reject because $3.71 > -3.3 > -3.71$

$t_{90} = 1.94$

$CI_{90} = 4.13 \pm (0.87 \times 1.94) = [2.44,\ 5.82]$

Question #3

Researchers draw a sample of 7 with a mean of 3.11 and a standard deviation of 0.88. Test $H_0 : \mu = 10$ at an α of 0.05, state your decision, then calculate a 99% confidence interval.

$s_{\bar{X}} = 0.88/\sqrt{7} = 0.33$

$t_{\text{tobs}} = (3.11 - 10)/0.33 = -20.88$

$t_{\text{crit}} = \pm 2.45$

Reject because $-20.88 =< -2.45$

$t_{99} = 3.71$

$CI_{99} = 3.11 \pm (0.33 \times 3.71) = [1.89,\ 4.33]$

Question #4

Researchers draw a sample of 5 with a mean of 5.16 and a standard deviation of 2.22. Test $H_0 : \mu = 10$ at an α of 0.05, state your decision, then calculate a 95% confidence interval.

$s_{\bar{X}} = 2.22/\sqrt{5} = 0.99$

$t_{\text{tobs}} = (5.16 - 10)/0.99 = -4.89$

$t_{\text{crit}} = \pm 2.78$

Reject because $-4.89 =< -2.78$

$t_{95} = 2.78$

$CI_{95} = 5.16 \pm (0.99 \times 2.78) = [2.41,\ 7.91]$

Question #5

Researchers draw a sample of 7 with a mean of 5.89 and a standard deviation of 2.1. Test $H_0 : \mu = 3$ at an α of 0.1, state your decision, then calculate a 95% confidence interval.

$s_{\bar{X}} = 2.1/\sqrt{7} = 0.79$

$t_{\text{tobs}} = (5.89 - 3)/0.79 = 3.66$

$t_{\text{crit}} = \pm 1.94$

Reject because $3.66 >= 1.94$

$t_{95} = 2.45$

$CI_{95} = 5.89 \pm (0.79 \times 2.45) = [3.95,\ 7.83]$

Question #6

Researchers draw a sample of 7 with a mean of 6.86 and a standard deviation of 2.01. Test $H_0 : \mu = 8$ at an α of 0.01, state your decision, then calculate a 99% confidence interval.

$s_{\bar{X}} = 2.01/\sqrt{7} = 0.76$
$t_{\text{tobs}} = (6.86 - 8)/0.76 = -1.5$
$t_{\text{crit}} = \pm 3.71$
Fail to reject because $3.71 > -1.5 > -3.71$
$t_{99} = 3.71$
$CI_{99} = 6.86 \pm (0.76 \times 3.71) = [4.04,\ 9.68]$

Question #7

Researchers draw a sample of 6 with a mean of 8.21 and a standard deviation of 1.12. Test $H_0 : \mu = 4$ at an α of 0.05, state your decision, then calculate a 90% confidence interval.

$s_{\bar{X}} = 1.12/\sqrt{6} = 0.46$
$t_{\text{tobs}} = (8.21 - 4)/0.46 = 9.15$
$t_{\text{crit}} = \pm 2.57$
Reject because $9.15 >= 2.57$
$t_{90} = 2.02$
$CI_{90} = 8.21 \pm (0.46 \times 2.02) = [7.28,\ 9.14]$

Question #8

Researchers draw a sample of 5 with a mean of 5.77 and a standard deviation of 2.29. Test $H_0 : \mu = 10$ at an α of 0.1, state your decision, then calculate a 90% confidence interval.

$s_{\bar{X}} = 2.29/\sqrt{5} = 1.02$
$t_{\text{tobs}} = (5.77 - 10)/1.02 = -4.15$
$t_{\text{crit}} = \pm 2.13$
Reject because $-4.15 =< -2.13$
$t_{90} = 2.13$
$CI_{90} = 5.77 \pm (1.02 \times 2.13) = [3.6,\ 7.94]$

Question #9

Researchers draw a sample of 5 with a mean of 3.28 and a standard deviation of 2.37. Test $H_0 : \mu = 10$ at an α of 0.05, state your decision, then calculate a 95% confidence interval.

$s_{\bar{X}} = 2.37/\sqrt{5} = 1.06$
$t_{\text{tobs}} = (3.28 - 10)/1.06 = -6.34$
$t_{\text{crit}} = \pm 2.78$
Reject because $-6.34 =< -2.78$
$t_{95} = 2.78$
$CI_{95} = 3.28 \pm (1.06 \times 2.78) = [0.33,\ 6.23]$

Question #10

Researchers draw a sample of 5 with a mean of 5.53 and a standard deviation of 1.42. Test $H_0 : \mu = 6$ at an α of 0.05, state your decision, then calculate a 95% confidence interval.

$s_{\bar{X}} = 1.42/\sqrt{5} = 0.64$
$t_{\text{tobs}} = (5.53 - 6)/0.64 = -0.73$
$t_{\text{crit}} = \pm 2.78$
Fail to reject because $2.78 > -0.73 > -2.78$
$t_{95} = 2.78$
$CI_{95} = 5.53 \pm (0.64 \times 2.78) = [3.75,\ 7.31]$

Week #8

- Give an example of a research question for which you would use an independent t-test
- What is the standard error of mean differences?
- What is the sampling distribution of mean differences?

Independent t-tests

$$df_i = n_i - 1$$
$$df_{\text{tot}} = df_1 + df_2$$
$$SS_i = s_i^2 \times df_i$$
$$t_{\text{obs}} = \frac{(\bar{X}_1 - \bar{X}_2)}{s_{(\bar{X}_1 - \bar{X}_2)}}$$
$$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$
$$s_p^2 = \frac{SS_1 + SS_2}{df_{\text{tot}}}$$
$$CI_y = (\bar{X}_1 - \bar{X}_2) \pm \ s_{(\bar{X}_1 - \bar{X}_2)} \times t_y$$
$$t_y = \text{is the critical value for } \alpha = (100 - y)/100$$

Critical t values

	α		
df	0.1	0.05	0.01
10	1.81	2.23	3.17
11	1.8	2.2	3.11
12	1.78	2.18	3.05

Question #1

Researchers draw one sample of 6 with a mean of 4.91 and a standard deviation of 2.19, and another sample of 6 with a mean of 2.79 and a standard deviation of 0.79. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.05, state the error, then calculate a 99% confidence interval.

$$s_1^2 = 2.19^2 = 4.79$$
$$s_2^2 = 0.79^2 = 0.63$$
$$df_1 = 6 - 1 = 5$$
$$df_2 = 6 - 1 = 5$$
$$SS_1 = 4.79 \times 5 = 23.95$$
$$SS_2 = 0.63 \times 5 = 3.15$$
$$df_{\text{tot}} = 5 + 5 = 10$$
$$s_p^2 = (23.95 + 3.15)/10 = 2.71$$
$$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(2.71/6) + (2.71/6)} = 0.95$$
$$t_{\text{obs}}(10) = (4.91 - 2.79)/0.95 = 2.23$$
$$t_{\text{crit}} = \pm 2.23$$
Reject because 2.23 >= 2.23
$$t_{99} = 3.17$$
$$CI_{99} = (4.91 - 2.79) \pm (0.95 \times 3.17) = [-0.89,\ 5.13]$$

Question #2

Researchers draw one sample of 7 with a mean of 4.13 and a standard deviation of 2.31, and another sample of 7 with a mean of 6.61 and a standard deviation of 2.06. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.01, state the error, then calculate a 99% confidence interval.

$$s_1^2 = 2.31^2 = 5.35$$
$$s_2^2 = 2.06^2 = 4.24$$
$$df_1 = 7 - 1 = 6$$
$$df_2 = 7 - 1 = 6$$
$$SS_1 = 5.35 \times 6 = 32.1$$
$$SS_2 = 4.24 \times 6 = 25.44$$
$$df_{\text{tot}} = 6 + 6 = 12$$
$$s_p^2 = (32.1 + 25.44)/12 = 4.79$$
$$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(4.79/7) + (4.79/7)} = 1.17$$
$$t_{\text{obs}}(12) = (4.13 - 6.61)/1.17 = -2.12$$
$$t_{\text{crit}} = \pm 3.05$$
Fail to reject because $3.05 > -2.12 > -3.05$
$$t_{99} = 3.05$$
$$CI_{99} = (4.13 - 6.61) \pm (1.17 \times 3.05) = [-6.05,\ 1.09]$$

Question #3

Researchers draw one sample of 7 with a mean of 3.11 and a standard deviation of 0.88, and another sample of 7 with a mean of 8.1 and a standard deviation of 1.24. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.1, state the error, then calculate a 99% confidence interval.

$$s_1^2 = 0.88^2 = 0.78$$
$$s_2^2 = 1.24^2 = 1.53$$
$$df_1 = 7 - 1 = 6$$
$$df_2 = 7 - 1 = 6$$
$$SS_1 = 0.78 \times 6 = 4.68$$
$$SS_2 = 1.53 \times 6 = 9.18$$
$$df_{\text{tot}} = 6 + 6 = 12$$
$$s_p^2 = (4.68 + 9.18)/12 = 1.16$$
$$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(1.16/7) + (1.16/7)} = 0.58$$
$$t_{\text{obs}}(12) = (3.11 - 8.1)/0.58 = -8.6$$
$$t_{\text{crit}} = \pm 1.78$$
Reject because $-8.6 =< -1.78$
$$t_{99} = 3.05$$
$$CI_{99} = (3.11 - 8.1) \pm (0.58 \times 3.05) = [-6.76,\ -3.22]$$

Question #4

Researchers draw one sample of 6 with a mean of 5.91 and a standard deviation of 2.7, and another sample of 6 with a mean of 6.44 and a standard deviation of 1.35. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.1, state the error, then calculate a 99% confidence interval.

$s_1^2 = 2.7^2 = 7.31$
$s_2^2 = 1.35^2 = 1.81$
$df_1 = 6 - 1 = 5$
$df_2 = 6 - 1 = 5$
$SS_1 = 7.31 \times 5 = 36.55$
$SS_2 = 1.81 \times 5 = 9.05$
$df_{\text{tot}} = 5 + 5 = 10$
$s_p^2 = (36.55 + 9.05)/10 = 4.56$
$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(4.56/6) + (4.56/6)} = 1.23$
$t_{\text{obs}}(10) = (5.91 - 6.44)/1.23 = -0.43$
$t_{\text{crit}} = \pm 1.81$
Fail to reject because $1.81 > -0.43 > -1.81$
$t_{99} = 3.17$
$CI_{99} = (5.91 - 6.44) \pm (1.23 \times 3.17) = [-4.43,\ 3.37]$

Question #5

Researchers draw one sample of 7 with a mean of 5.89 and a standard deviation of 2.1, and another sample of 6 with a mean of 3 and a standard deviation of 1.01. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.01, state the error, then calculate a 95% confidence interval.

$s_1^2 = 2.1^2 = 4.4$
$s_2^2 = 1.01^2 = 1.03$
$df_1 = 7 - 1 = 6$
$df_2 = 6 - 1 = 5$
$SS_1 = 4.4 \times 6 = 26.4$
$SS_2 = 1.03 \times 5 = 5.15$
$df_{\text{tot}} = 6 + 5 = 11$
$s_p^2 = (26.4 + 5.15)/11 = 2.87$
$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(2.87/7) + (2.87/6)} = 0.94$
$t_{\text{obs}}(11) = (5.89 - 3)/0.94 = 3.07$
$t_{\text{crit}} = \pm 3.11$
Fail to reject because $3.11 > 3.07 > -3.11$
$t_{95} = 2.2$
$CI_{95} = (5.89 - 3) \pm (0.94 \times 2.2) = [0.82,\ 4.96]$

Question #6

Researchers draw one sample of 7 with a mean of 6.86 and a standard deviation of 2.01, and another sample of 7 with a mean of 6.93 and a standard deviation of 1.7. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.01, state the error, then calculate a 99% confidence interval.

$s_1^2 = 2.01^2 = 4.04$
$s_2^2 = 1.7^2 = 2.89$
$df_1 = 7 - 1 = 6$
$df_2 = 7 - 1 = 6$
$SS_1 = 4.04 \times 6 = 24.24$
$SS_2 = 2.89 \times 6 = 17.34$
$df_{\text{tot}} = 6 + 6 = 12$
$s_p^2 = (24.24 + 17.34)/12 = 3.46$
$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(3.46/7) + (3.46/7)} = 0.99$
$t_{\text{obs}}(12) = (6.86 - 6.93)/0.99 = -0.07$
$t_{\text{crit}} = \pm 3.05$
Fail to reject because $3.05 > -0.07 > -3.05$
$t_{99} = 3.05$
$CI_{99} = (6.86 - 6.93) \pm (0.99 \times 3.05) = [-3.09,\ 2.95]$

Question #7

Researchers draw one sample of 6 with a mean of 8.21 and a standard deviation of 1.12, and another sample of 6 with a mean of 3.57 and a standard deviation of 1.81. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.01, state the error, then calculate a 95% confidence interval.

$s_1^2 = 1.12^2 = 1.25$
$s_2^2 = 1.81^2 = 3.26$
$df_1 = 6 - 1 = 5$
$df_2 = 6 - 1 = 5$
$SS_1 = 1.25 \times 5 = 6.25$
$SS_2 = 3.26 \times 5 = 16.3$
$df_{\text{tot}} = 5 + 5 = 10$
$s_p^2 = (6.25 + 16.3)/10 = 2.25$
$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(2.25/6) + (2.25/6)} = 0.87$
$t_{\text{obs}}(10) = (8.21 - 3.57)/0.87 = 5.33$
$t_{\text{crit}} = \pm 3.17$
Reject because $5.33 >= 3.17$
$t_{95} = 2.23$
$CI_{95} = (8.21 - 3.57) \pm (0.87 \times 2.23) = [2.7,\ 6.58]$

Question #8

Researchers draw one sample of 6 with a mean of 6.37 and a standard deviation of 2.51, and another sample of 6 with a mean of 3.58 and a standard deviation of 1.52. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.05, state the error, then calculate a 99% confidence interval.

$s_1^2 = 2.51^2 = 6.32$

$s_2^2 = 1.52^2 = 2.3$

$df_1 = 6 - 1 = 5$

$df_2 = 6 - 1 = 5$

$SS_1 = 6.32 \times 5 = 31.6$

$SS_2 = 2.3 \times 5 = 11.5$

$df_{\text{tot}} = 5 + 5 = 10$

$s_p^2 = (31.6 + 11.5)/10 = 4.31$

$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(4.31/6) + (4.31/6)} = 1.2$

$t_{\text{obs}}(10) = (6.37 - 3.58)/1.2 = 2.33$

$t_{\text{crit}} = \pm 2.23$

Reject because 2.33 $>=$ 2.23

$t_{99} = 3.17$

$CI_{99} = (6.37 - 3.58) \pm (1.2 \times 3.17) = [-1.01,\ 6.59]$

Week #9

- Give an example of a research design in which you would use a dependent t-test
- How does using a dependent design rather than an independent design affect power? Why?
- Give an example of a research question which could be addressed by ANOVA but not a single t-test
- Define the numerator and denominator of the F statistic, then explain why F is equal to 1 under the null hypothesis
- How much within and between group variability is there in each of the following scenarios?
  - Scenario #1
    - Group 1: [1, 1, 1, 1, 1]
    - Group 2: [1, 1, 1, 1, 1]
    - Group 3: [1, 1, 1, 1, 1]
  - Scenario #2
    - Group 1: [1, 4, 3, 7, 9]
    - Group 2: [2, 3, 4, 8, 8]
    - Group 3: [1, 2, 5, 9, 7]
  - Scenario #3
    - Group 1: [1, 1, 1, 1, 1]
    - Group 2: [15, 15, 15, 15, 15]
    - Group 3: [30, 30, 30, 30, 30]
  - Scenario #4
    - Group 1: [1, 4, 3, 7, 9]
    - Group 2: [15, 26, 17, 29, 21]
    - Group 3: [30, 37, 43, 41, 32]

  Rank them on the likelihoood the null hypothesis (F = 1) will be rejected

Dependent t-tests

$df = n - 1$

$\bar{D} = \Sigma(D_i)/n$

$s_D = \sqrt{\Sigma[(D_i - \bar{D})^2]/df}$

$s_{\bar{D}} = s_D/\sqrt{n}$

$t_{\text{obs}} = \bar{D}/s_{\bar{D}}$

$CI_y = \bar{D} \pm s_{\bar{D}} \times t_{\text{y}}$

t_y is the critical value for $\alpha = (100 - y)/100$

Critical t values

	α		
df	0.1	0.05	0.01
4	2.13	2.78	4.6
5	2.02	2.57	4.03
6	1.94	2.45	3.71

Question #1

Researchers collect the following data:

Pre	Post	D_i	$D_i - \bar{D}$	$(D_i - \bar{D})^2$
8	2	6	2.83	8.01
8	2	6	2.83	8.01
5	2	3	-0.17	0.03
3	4	-1	-4.17	17.39
4	4	0	-3.17	10.05
7	2	5	1.83	3.35

Test $H_0 : \mu_{\bar{D}} = 0$ at an α of 0.05, state the decision/error, then calculate a 99% confidence interval.

$\bar{D} = 3.17$

$\Sigma(D_i - \bar{D})^2 = 46.84$

$df = 5$

$s_D = \sqrt{46.84/5} = 3.06$

$s_{\bar{D}} = 3.06/\sqrt{6} = 1.25$

$t_{\text{obs}}(5) = 3.17/1.25 = 2.54$

$t_{\text{crit}} = 2.57$

Fail to reject because $2.57 > 2.54 > -2.57$

$t_{99} = 4.03$

$CI_{99} = 3.17 \pm (1.25 \times 4.03) = [-1.87,\ 8.21]$

Question #2

Researchers collect the following data:

Pre	Post	D_i	$D_i - \bar{D}$	$(D_i - \bar{D})^2$
5	4	1	-1.8	3.24
8	2	6	3.2	10.24
6	3	3	0.2	0.04
5	3	2	-0.8	0.64
6	4	2	-0.8	0.64

Test $H_0 : \mu_{\bar{D}} = 0$ at an α of 0.01, state the decision/error, then calculate a 99% confidence interval.

$\bar{D} = 2.8$

$\Sigma(D_i - \bar{D})^2 = 14.8$

$df = 4$

$s_D = \sqrt{14.8/4} = 1.92$

$s_{\bar{D}} = 1.92/\sqrt{5} = 0.86$

$t_{\text{obs}}(4) = 2.8/0.86 = 3.26$

$t_{\text{crit}} = 4.6$

Fail to reject because $4.6 > 3.26 > -4.6$

$t_{99} = 4.6$

$CI_{99} = 2.8 \pm (0.86 \times 4.6) = [-1.16,\ 6.76]$

Question #3

Researchers collect the following data:

Pre	Post	D_i	$D_i - \bar{D}$	$(D_i - \bar{D})^2$
8	9	-1	-4	16
10	7	3	0	0
8	4	4	1	1
10	2	8	5	25
9	8	1	-2	4

Test $H_0 : \mu_{\bar{D}} = 0$ at an α of 0.01, state the decision/error, then calculate a 95% confidence interval.

$\bar{D} = 3$

$\Sigma(D_i - \bar{D})^2 = 46$

$df = 4$

$s_D = \sqrt{46/4} = 3.39$

$s_{\bar{D}} = 3.39/\sqrt{5} = 1.52$

$t_{\text{obs}}(4) = 3/1.52 = 1.97$

$t_{\text{crit}} = 4.6$

Fail to reject because $4.6 > 1.97 > -4.6$

$t_{95} = 2.78$

$CI_{95} = 3 \pm (1.52 \times 2.78) = [-1.23,\ 7.23]$

Question #4

Researchers collect the following data:

Pre	Post	D_i	$D_i - \bar{D}$	$(D_i - \bar{D})^2$
4	9	-5	-0.67	0.45
5	8	-3	1.33	1.77
1	10	-9	-4.67	21.81
3	10	-7	-2.67	7.13
7	4	3	7.33	53.73
3	8	-5	-0.67	0.45

Test $H_0 : \mu_{\bar{D}} = 0$ at an α of 0.1, state the decision/error, then calculate a 99% confidence interval.

$\bar{D} = -4.33$

$\Sigma(D_i - \bar{D})^2 = 85.34$

$df = 5$

$s_D = \sqrt{85.34/5} = 4.13$

$s_{\bar{D}} = 4.13/\sqrt{6} = 1.69$

$t_{\text{obs}}(5) = -4.33/1.69 = -2.56$

$t_{\text{crit}} = 2.02$

Reject because $-2.56 =< -2.02$

$t_{99} = 4.03$

$CI_{99} = -4.33 \pm (1.69 \times 4.03) = [-11.14,\ 2.48]$

Question #5

Researchers collect the following data:

Pre	Post	D_i	$D_i - \bar{D}$	$(D_i - \bar{D})^2$
4	5	-1	0.4	0.16
7	8	-1	0.4	0.16
8	8	0	1.4	1.96
2	7	-5	-3.6	12.96
5	5	0	1.4	1.96

Test $H_0 : \mu_{\bar{D}} = 0$ at an α of 0.05, state the decision/error, then calculate a 95% confidence interval.

$\bar{D} = -1.4$

$\Sigma(D_i - \bar{D})^2 = 17.2$

$df = 4$

$s_D = \sqrt{17.2/4} = 2.07$

$s_{\bar{D}} = 2.07/\sqrt{5} = 0.93$

$t_{\text{obs}}(4) = -1.4/0.93 = -1.51$

$t_{\text{crit}} = 2.78$

Fail to reject because $2.78 > -1.51 > -2.78$

$t_{95} = 2.78$

$CI_{95} = -1.4 \pm (0.93 \times 2.78) = [-3.99,\ 1.19]$

Week #10

- Give an example of two variables which would have a correlation of close to 1
- Give an example of two variables which would have a correlation of close to -1
- Give an example of two variable which would have a correlation of close to 0
- When isn't Pearson's r an appropriate measure?
- Why doesn't correlation imply causation? Does causation imply correlation?
- One sample has n = 10 and a variance of 20, the other has n = 15 and variance of 30; will the pooled variance be closer to 15 or 30 (rewrite this)
- Why does $df = n - 2$?
- Rank each set of scores:

$$4, 0, 4, 9, 3$$
$$3, 5, 3, 6, 3$$

Pearson's r

$\bar{X} = \Sigma(X_i)/n$
$df = n - 1$
$SP = \Sigma[(X_i - \bar{X})(Y_i - \bar{Y})]$
$SS_X = \Sigma[(X_i - \bar{X})^2]$
$SS_Y = \Sigma[(Y_i - \bar{Y})^2]$
$r_{XY} = SP/\sqrt{SS_X \times SS_Y}$

Critical r values

	α		
$(n-2)$	0.2	0.1	0.05
2	0.8	0.9	0.95
3	0.69	0.81	0.88
4	0.61	0.73	0.81

Question #1

Calculate r_{XY} and test H_0: $\rho_{XY} = 0$ at $\alpha = 0.2$.

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
4	9	-0	2	0	6	-1
5	8	0	2	0	2	1
7	5	2	-2	6	2	-4
2	4	-2	-2	6	6	6

$\bar{X} = 4.5$
$\bar{Y} = 6.5$
$SS_X = 13$
$SS_Y = 17$

$SP = 2$
$r_{XY} = 2/\sqrt{13 \times 17} = 0.13$
$r_{\text{crit}} = \pm 0.8$
Fail to reject because $0.8 > 0.13 > -0.8$

Question #2

Calculate r_{XY} and test H_0: $\rho_{XY} = 0$ at $\alpha = 0.1$.

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
7	9	2	4	4	16	8
5	2	0	-3	0	9	-0
2	6	-3	1	9	1	-3
6	3	1	-2	1	4	-2

$\bar{X} = 5$
$\bar{Y} = 5$
$SS_X = 14$
$SS_Y = 30$

$SP = 3$
$r_{XY} = 3/\sqrt{14 \times 30} = 0.15$
$r_{\text{crit}} = \pm 0.9$
Fail to reject because $0.9 > 0.15 > -0.9$

Question #3

Calculate r_{XY} and test H$_0$: $\rho_{XY} = 0$ at $\alpha = 0.1$.

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
8	6	2	2	4	6	5
4	1	-2	-2	4	6	5
3	3	-3	-0	9	0	2
9	4	3	0	9	0	2

$\bar{X} = 6$
$\bar{Y} = 3.5$
$SS_X = 26$
$SS_Y = 13$

$SP = 13$
$r_{XY} = 13/\sqrt{26 \times 13} = 0.71$
$r_{\text{crit}} = \pm 0.9$
Fail to reject because $0.9 > 0.71 > -0.9$

Question #4

Calculate r_{XY} and test H$_0$: $\rho_{XY} = 0$ at $\alpha = 0.2$. Error in ifelse(robs > rcrit, ">= ", " =< -") : object 'robs' not found

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
1	7	-3	1	7	1	-3
3	8	-1	2	0	5	-1
2	9	-2	3	3	10	-5
5	1	1	-5	2	23	-7
7	4	3	-2	12	3	-6

Question #5

Calculate r_{XY} and test H$_0$: $\rho_{XY} = 0$ at $\alpha = 0.2$.

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
7	7	2	0	6	0	1
8	5	4	-2	12	3	-6
2	6	-2	-1	6	1	2
1	9	-4	2	12	5	-8

$\bar{X} = 4.5$
$\bar{Y} = 6.75$
$SS_X = 37$
$SS_Y = 8.75$

$SP = -11.5$
$r_{XY} = -11.5/\sqrt{37 \times 8.75} = -0.64$
$r_{\text{crit}} = \pm 0.8$
Fail to reject because $0.8 > -0.64 > -0.8$

Question #6

Calculate r_{XY} and test H_0: $\rho_{XY} = 0$ at $\alpha = 0.1$.

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
9	9	3	3	8	10	9
3	7	-3	1	10	1	-4
8	4	2	-2	3	3	-3
5	1	-1	-5	1	23	6
6	8	-0	2	0	5	-0

$\bar{X} = 6.2$

$\bar{Y} = 5.8$

$SS_X = 22.8$

$SS_Y = 42.8$

$SP = 7.2$

$r_{XY} = 7.2/\sqrt{22.8 \times 42.8} = 0.23$

$r_{\text{crit}} = \pm 0.81$

Fail to reject because $0.81 > 0.23 > -0.81$

Question #7

Calculate r_{XY} and test H_0: $\rho_{XY} = 0$ at $\alpha = 0.2$.

X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
4	9	-1	3	1	9	-3
1	2	-4	-4	16	16	16
8	4	3	-2	9	4	-6
2	8	-3	2	9	4	-6
9	6	4	0	16	0	0
6	7	1	1	1	1	1

$\bar{X} = 5$

$\bar{Y} = 6$

$SS_X = 52$

$SS_Y = 34$

$SP = 2$

$r_{XY} = 2/\sqrt{52 \times 34} = 0.05$

$r_{\text{crit}} = \pm 0.61$

Fail to reject because $0.61 > 0.05 > -0.61$

Week #11

- What does the regression line minimise? Draw a picture
- How many regression lines are possible for a given set of data? How many lines of best fit are possible?
- Draw an example of a nonlinear relationship
- Give an example of restriction of range
- Explain the difference between univariate and regression outliers using a drawing
- How do outliers affect model fit?

Regression

$\bar{Y} = \Sigma(Y_i)/n$
$df_1 = 1$
$df_2 = n - df_1 - 1$
$SP = \Sigma[(X_i - \bar{X})(Y_i - \bar{Y})]$
$SS_X = \Sigma[(X_i - \bar{X})^2]$
$\beta_1 = SP/SS_X$
$\beta_0 = \bar{Y} - \beta_1 \times \bar{X}$

$\hat{Y}_i = \beta_0 + X_i \times \beta_1$
$SS_{\text{tot}} = \Sigma[(Y_i - \bar{Y})^2]$
$SS_{\text{reg}} = \Sigma[(\hat{Y} - \bar{Y})^2]$
$SS_{\text{res}} = SS_{\text{tot}} - SS_{\text{reg}}$
$MS_{\text{reg}} = SS_{\text{reg}}/df_1$
$MS_{\text{res}} = SS_{\text{res}}/df_2$
$F = MS_{\text{reg}}/MS_{\text{res}}$

Critical F values

		df_1		
df_2	α	1	2	3
1	0.05	161.4	199.5	215.71
	0.01	4052	4999	5404
2	0.05	18.51	19	19.16
	0.01	98.94	99	99.17
3	0.05	7.71	6.94	6.59
	0.01	34.12	30.82	29.46
4	0.05	7.71	6.94	6.59
	0.01	21.2	18	16.69

Question #1

Test the model fit at an α of 0.05.

X_i	$(X_i - \bar{X})^2$	Y_i	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$(\hat{Y}_i - \bar{Y})$	$(\hat{Y}_i - \bar{Y})^2$
3	2.25	5	0.06	-0.38	4.6	-0.15	0.02
5	0.25	7	5.06	1.12	4.8	0.05	0
1	12.25	3	3.06	6.12	4.4	-0.35	0.12
9	20.25	4	0.56	-3.38	5.2	0.45	0.2

$SS_X = 35$
$SP = 3.48$
$\beta_1 = 3.48/35 = 0.1$
$\bar{Y} = 4.75$
$\bar{X} = 4.5$
$\beta_0 = 4.75 - (0.1 \times 4.5) = 4.3$
$\hat{Y}_i = 4.3 + (0.1 \times X_i)$
$SS_{\text{tot}} = 8.75$
$SS_{\text{reg}} = 0.35$

$SS_{\text{res}} = 8.75 - 0.35 = 8.4$
$df_1 = 1$
$df_2 = 4 - 1 - 1 = 2$
$MS_{\text{reg}} = 0.35/1 = 0.35$
$MS_{\text{res}} = 8.4/2 = 4.2$
$F = 0.35/4.2 = 0.08$
$F_{\text{crit}} = 18.51$
Fail to reject because $0.08 < 18.51$

Question #2

Test the model fit at an α of 0.01.

X_i	$(X_i - \bar{X})^2$	Y_i	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$(\hat{Y}_i - \bar{Y})$	$(\hat{Y}_i - \bar{Y})^2$
7	4	9	16	8	5.42	0.42	0.18
5	0	2	9	0	5	0	0
2	9	6	1	-3	4.37	-0.63	0.4
6	1	3	4	-2	5.21	0.21	0.04

$SS_X = 14$

$SP = 3$

$\beta_1 = 3/14 = 0.21$

$\bar{Y} = 5$

$\bar{X} = 5$

$\beta_0 = 5 - (0.21 \times 5) = 3.95$

$\hat{Y}_i = 3.95 + (0.21 \times X_i)$

$SS_{\text{tot}} = 30$

$SS_{\text{reg}} = 0.62$

$SS_{\text{res}} = 30 - 0.62 = 29.38$

$df_1 = 1$

$df_2 = 4 - 1 - 1 = 2$

$MS_{\text{reg}} = 0.62/1 = 0.62$

$MS_{\text{res}} = 29.38/2 = 14.69$

$F = 0.62/14.69 = 0.04$

$F_{\text{crit}} = 98.5$

Fail to reject because $0.04 < 98.5$

Question #3

Test the model fit at an α of 0.01.

X_i	$(X_i - \bar{X})^2$	Y_i	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$(\hat{Y}_i - \bar{Y})$	$(\hat{Y}_i - \bar{Y})^2$
8	4	6	6.25	5	4.5	1	1
4	4	1	6.25	5	2.5	-1	1
3	9	3	0.25	1.5	2	-1.5	2.25
9	9	4	0.25	1.5	5	1.5	2.25

$SS_X = 26$

$SP = 13$

$\beta_1 = 13/26 = 0.5$

$\bar{Y} = 3.5$

$\bar{X} = 6$

$\beta_0 = 3.5 - (0.5 \times 6) = 0.5$

$\hat{Y}_i = 0.5 + (0.5 \times X_i)$

$SS_{\text{tot}} = 13$

$SS_{\text{reg}} = 6.5$

$SS_{\text{res}} = 13 - 6.5 = 6.5$

$df_1 = 1$

$df_2 = 4 - 1 - 1 = 2$

$MS_{\text{reg}} = 6.5/1 = 6.5$

$MS_{\text{res}} = 6.5/2 = 3.25$

$F = 6.5/3.25 = 2$

$F_{\text{crit}} = 98.5$

Fail to reject because $2 < 98.5$

Question #4

Test the model fit at an α of 0.01.

X_i	$(X_i - \bar{X})^2$	Y_i	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$(\hat{Y}_i - \bar{Y})$	$(\hat{Y}_i - \bar{Y})^2$
1	11.56	6	1	-3.4	4.69	-0.31	0.1
6	2.56	3	4	-3.2	5.14	0.14	0.02
5	0.36	5	0	0	5.05	0.05	0
2	5.76	4	1	2.4	4.78	-0.22	0.05
8	12.96	7	4	7.2	5.32	0.32	0.1

$SS_X = 33.2$

$SP = 3$

$\beta_1 = 3/33.2 = 0.09$

$\bar{Y} = 5$

$\bar{X} = 4.4$

$\beta_0 = 5 - (0.09 \times 4.4) = 4.6$

$\hat{Y}_i = 4.6 + (0.09 \times X_i)$

$SS_{\text{tot}} = 10$

$SS_{\text{reg}} = 0.27$

$SS_{\text{res}} = 10 - 0.27 = 9.73$

$df_1 = 1$

$df_2 = 5 - 1 - 1 = 3$

$MS_{\text{reg}} = 0.27/1 = 0.27$

$MS_{\text{res}} = 9.73/3 = 3.24$

$F = 0.27/3.24 = 0.08$

$F_{\text{crit}} = 34.12$

Fail to reject because $0.08 < 34.12$

Question #5

Test the model fit at an α of 0.05.

X_i	$(X_i - \bar{X})^2$	Y_i	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$(\hat{Y}_i - \bar{Y})$	$(\hat{Y}_i - \bar{Y})^2$
7	6.25	7	0.06	0.62	5.97	-0.78	0.61
8	12.25	5	3.06	-6.12	5.66	-1.09	1.19
2	6.25	6	0.56	1.88	7.52	0.77	0.59
1	12.25	9	5.06	-7.88	7.83	1.08	1.17

$SS_X = 37$

$SP = -11.5$

$\beta_1 = -11.5/37 = -0.31$

$\bar{Y} = 6.75$

$\bar{X} = 4.5$

$\beta_0 = 6.75 - (-0.31 \times 4.5) = 8.14$

$\hat{Y}_i = 8.14 + (-0.31 \times X_i)$

$SS_{\text{tot}} = 8.75$

$SS_{\text{reg}} = 3.56$

$SS_{\text{res}} = 8.75 - 3.56 = 5.19$

$df_1 = 1$

$df_2 = 4 - 1 - 1 = 2$

$MS_{\text{reg}} = 3.56/1 = 3.56$

$MS_{\text{res}} = 5.19/2 = 2.6$

$F = 3.56/2.6 = 1.37$

$F_{\text{crit}} = 18.51$

Fail to reject because $1.37 < 18.51$

Question #6

Test the model fit at an α of 0.01.

X_i	$(X_i - \bar{X})^2$	Y_i	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$(\hat{Y}_i - \bar{Y})$	$(\hat{Y}_i - \bar{Y})^2$
9	7.84	9	10.24	8.96	6.7	0.9	0.81
3	10.24	7	1.44	-3.84	4.78	-1.02	1.04
8	3.24	4	3.24	-3.24	6.38	0.58	0.34
5	1.44	1	23.04	5.76	5.42	-0.38	0.14
6	0.04	8	4.84	-0.44	5.74	-0.06	0

$SS_X = 22.8$
$SP = 7.2$
$\beta_1 = 7.2/22.8 = 0.32$
$\bar{Y} = 5.8$
$\bar{X} = 6.2$
$\beta_0 = 5.8 - (0.32 \times 6.2) = 3.82$
$\hat{Y}_i = 3.82 + (0.32 \times X_i)$
$SS_{\text{tot}} = 42.8$
$SS_{\text{reg}} = 2.33$

$SS_{\text{res}} = 42.8 - 2.33 = 40.47$
$df_1 = 1$
$df_2 = 5 - 1 - 1 = 3$
$MS_{\text{reg}} = 2.33/1 = 2.33$
$MS_{\text{res}} = 40.47/3 = 13.49$
$F = 2.33/13.49 = 0.17$
$F_{\text{crit}} = 34.12$
Fail to reject because $0.17 < 34.12$

Question #7

Test the model fit at an α of 0.05.

X_i	$(X_i - \bar{X})^2$	Y_i	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$(\hat{Y}_i - \bar{Y})$	$(\hat{Y}_i - \bar{Y})^2$
4	1	9	9	-3	5.96	-0.04	0
1	16	2	16	16	5.84	-0.16	0.03
8	9	4	4	-6	6.12	0.12	0.01
2	9	8	4	-6	5.88	-0.12	0.01
9	16	6	0	0	6.16	0.16	0.03
6	1	7	1	1	6.04	0.04	0

$SS_X = 52$
$SP = 2$
$\beta_1 = 2/52 = 0.04$
$\bar{Y} = 6$
$\bar{X} = 5$
$\beta_0 = 6 - (0.04 \times 5) = 5.8$
$\hat{Y}_i = 5.8 + (0.04 \times X_i)$
$SS_{\text{tot}} = 34$
$SS_{\text{reg}} = 0.08$

$SS_{\text{res}} = 34 - 0.08 = 33.92$
$df_1 = 1$
$df_2 = 6 - 1 - 1 = 4$
$MS_{\text{reg}} = 0.08/1 = 0.08$
$MS_{\text{res}} = 33.92/4 = 8.48$
$F = 0.08/8.48 = 0.01$
$F_{\text{crit}} = 7.71$
Fail to reject because $0.01 < 7.71$