Week #1

1. What is the difference between descriptive and inferential statistics?

2. Give an example of:

   (a) A population and a sample from that population
   (b) A parameter and its corresponding statistic
   (c) A descriptive research question and an inferential reframing of that question
   (d) A research question with an independent variable, a dependent variable, and a confounding variable

   For each, explain why your example is appropriate.

3. Why is a confounding variable so called? What does it prevent us from doing?

Summing notation

$X_1 =$ the first 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

$X = 3\ 4\ 6\ 10\ 3$
$Y = 9\ 10\ 7\ 7\ 1$

$\Sigma(X_i) = 26$
$[\Sigma(X_i)]^2 = 676$
$\Sigma(X_i - Y_i) = -8$
$\Sigma(X_iY_i) = 182$
$\Sigma(X_i^2) = 170$
$[\Sigma(X_i - Y_i)]^2 = 64$
$\Sigma(X_i)\Sigma(Y_i) = 884$

Question #2

$X = 2\ 8\ 6\ 2\ 10$
$Y = 10\ 2\ 9\ 5\ 6$

$\Sigma(X_i) =$
$[\Sigma(X_i)]^2 =$
$\Sigma(X_i - Y_i) =$
$\Sigma(X_iY_i) =$
$\Sigma(X_i^2) =$
$[\Sigma(X_i - Y_i)]^2 =$
$\Sigma(X_i)\Sigma(Y_i) =$

Question #3

$X = 2\ 9\ 4\ 4\ 7$
$Y = 7\ 2\ 3\ 6\ 7$

$\Sigma(X_i) =$
$[\Sigma(X_i)]^2 =$
$\Sigma(X_i - Y_i) =$
$\Sigma(X_iY_i) =$
$\Sigma(X_i^2) =$
$[\Sigma(X_i - Y_i)]^2 =$
$\Sigma(X_i)\Sigma(Y_i) =$

Question #4

$X = 6\ 1\ 3\ 3\ 9$
$Y = 3\ 8\ 10\ 10\ 1$

$\Sigma(X_i) =$
$[\Sigma(X_i)]^2 =$
$\Sigma(X_i - Y_i) =$
$\Sigma(X_iY_i) =$
$\Sigma(X_i^2) =$
$[\Sigma(X_i - Y_i)]^2 =$
$\Sigma(X_i)\Sigma(Y_i) =$

Question #5

$X = 3\ 7\ 10\ 3\ 2$
$Y = 8\ 6\ 9\ 10\ 2$

$\Sigma(X_i) =$
$[\Sigma(X_i)]^2 =$
$\Sigma(X_i - Y_i) =$
$\Sigma(X_iY_i) =$
$\Sigma(X_i^2) =$
$[\Sigma(X_i - Y_i)]^2 =$
$\Sigma(X_i)\Sigma(Y_i) =$

Question #6

$X = 7\ 10\ 3\ 4\ 9$
$Y = 10\ 10\ 8\ 6\ 1$

$\Sigma(X_i) =$
$[\Sigma(X_i)]^2 =$
$\Sigma(X_i - Y_i) =$
$\Sigma(X_iY_i) =$
$\Sigma(X_i^2) =$
$[\Sigma(X_i - Y_i)]^2 =$
$\Sigma(X_i)\Sigma(Y_i) =$

Week #2

1. Give an example of:

   (a) A discrete variable
   (b) A continuous variable
   (c) A nominal variable
   (d) An ordinal variable
   (e) An interval variable
   (f) A ratio variable

   For each, explain why your example is appropriate.

2. Draw a well-designed bar graph and a poorly-designed one, and explain why the first is better

3. Why might we prefer:

   (a) Relative frequency over frequency?
   (b) Cumulative relative frequency over relative frequency?

Frequency table construction

For the following data sets, create a frequency table containing the frequency (freq), cumulative frequency (cfreq), relative frequency (rfreq), and cumulative relative frequency (crfreq) of each value.

freq = number equal to score
cfreq = number less than or equal to score
rfreq = frequency divided by n
crfreq = cumulative frequency divided by n

Question #1

2, 2, 3, 5, 2, 5, 5, 4, 4, 1, 2, 1

score	freq	cfreq	rfreq	crfreq
1	2	2	0.17	0.17
2	4	6	0.33	0.50
3	1	7	0.08	0.58
4	2	9	0.17	0.75
5	3	12	0.25	1.00

Question #2

1, 4, 3, 1, 5, 5, 1, 5, 3, 3, 3, 2

Question #3

1, 5, 2, 2, 4, 4, 1, 2, 3, 4, 3, 3

Question #4

3, 1, 2, 2, 5, 2, 4, 5, 5, 1, 4, 2

Question #5

2, 4, 5, 2, 1, 4, 3, 5, 5, 1, 2, 3

Question #6

4, 5, 2, 2, 5, 5, 5, 4, 3, 1, 4, 5

Question #7

5, 2, 1, 1, 2, 4, 2, 5, 1, 3, 1, 2

Question #8

3, 2, 4, 4, 2, 4, 2, 5, 4, 4, 3, 1

Question #9

2, 1, 2, 2, 3, 1, 2, 2, 4, 5, 1, 1

Question #10

3, 2, 3, 4, 1, 2, 2, 2, 4, 3, 4, 3

Question #11

2, 1, 3, 1, 1, 5, 1, 2, 5, 1, 1, 3

Question #12

1, 5, 5, 2, 1, 1, 1, 4, 1, 1, 2, 5

Question #13

4, 2, 2, 1, 5, 1, 3, 4, 5, 1, 4, 5

Question #14

2, 4, 5, 3, 5, 3, 5, 3, 3, 2, 5, 1

Question #15

4, 1, 5, 4, 2, 5, 5, 2, 4, 5, 1, 4

Question #16

4, 2, 3, 2, 5, 2, 1, 5, 5, 1, 2, 3

Question #17

1, 5, 3, 4, 3, 3, 2, 1, 4, 1, 3, 1

Question #18

5, 4, 5, 1, 1, 3, 2, 3, 2, 2, 4, 4

Interval construction

For each scenario, find the lower limit (LL), midpoint (MP), and upper limit (UL) of the first five intervals.

- LL_1 = the greatest multiple of the width less than or equal to the lowest score, or (if that does not exist) 0
- $UL_i = LL_i + \text{width} - 1$
- $MP_i = (LL_i + UL_i)/2$

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.

Question #3

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

Question #4

The lowest value in the data set is 1, and the desired interval width is 30.

Question #5

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

Question #6

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

Question #7

The lowest value in the data set is 40, and the desired interval width is 50.

Question #8

The lowest value in the data set is 21, and the desired interval width is 24.

Question #9

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

Question #10

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

Question #11

The lowest value in the data set is 1, and the desired interval width is 14.

Question #12

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

Question #13

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

Question #14

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

Week #3

1. Why is the mean more sensitive to extreme scores than the median?

2. Give a scenario in which:

    (a) The median would be a better estimate than the mean
    (b) The mode would be a better estimate than the median/mean
    (c) It would be crucial to report the variance along with the mean

    For each, justify your example.

3. What are we squaring in a sum of squares, and why do we do it?

    $$\Sigma[(X_i - \bar{X})^2]$$

4. In what sense is a variance a mean?

    $$\frac{\Sigma[(X_i - \bar{X})^2]}{n - 1}$$

5. Why do we use $n - 1$ instead of n in the denominator of sample variance?

6. When would variance be equal to zero? Explain with reference the formula

7. Is it possible to have a negative variance? Justify your answer with reference to the formula

8. Why do we use standard deviation instead of variance?

9. Why is the standard deviation so named?

10. 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

11. Draw a positively skewed distribution and identify the approximate location of the mean, median, and mode. Why do they take the order they do?

Median & IQR calculation

For each frequency table, calculate the median (P_{50}) and interquartile range (P_{25} and P_{75}). The equation for the 50th percentile is:

$$P_{50} = LL + W \left[\frac{0.5(n) - cumF}{fm} \right]$$

LL is the least score with a crfreq ≥ 0.5, W is the interval width (1 if the data is ungrouped), n is the sample size (the greatest cfreq), $cumF$ is the crfreq of the row below the one containing LL (0 if there is no such row), and fm is the freq of the row containing LL.

Note: the frequency table must be in *descending* order

Question #1

score	freq	cfreq	rfreq	crfreq
62	3	30	0.10	1.00
61	10	27	0.33	0.90
60	13	17	0.43	0.57
59	4	4	0.13	0.13

$P_{25} = 59.5 + 1 \times [(0.25 \times 30 - 4)/13] = 59.77$
$P_{50} = 59.5 + 1 \times [(0.5 \times 30 - 4)/13] = 60.35$
$P_{75} = 60.5 + 1 \times [(0.75 \times 30 - 17)/10] = 61.05$

Question #2

score	freq	cfreq	rfreq	crfreq
23	1	22	0.05	1.00
22	4	21	0.18	0.95
21	11	17	0.50	0.77
20	4	6	0.18	0.27
19	2	2	0.09	0.09

Question #3

score	freq	cfreq	rfreq	crfreq
69	1	24	0.04	1.00
68	4	23	0.17	0.96
67	11	19	0.46	0.79
66	8	8	0.33	0.33

Question #4

score	freq	cfreq	rfreq	crfreq
12	1	15	0.07	1.00
11	3	14	0.20	0.93
10	4	11	0.27	0.73
9	6	7	0.40	0.47
8	1	1	0.07	0.07

Question #5

score	freq	cfreq	rfreq	crfreq
8	6	21	0.29	1.00
7	8	15	0.38	0.71
6	3	7	0.14	0.33
5	4	4	0.19	0.19

Question #6

score	freq	cfreq	rfreq	crfreq
47	1	23	0.04	1.00
46	1	22	0.04	0.96
45	5	21	0.22	0.91
44	10	16	0.43	0.70
43	6	6	0.26	0.26

Question #7

score	freq	cfreq	rfreq	crfreq
13	1	14	0.07	1.00
12	2	13	0.14	0.93
11	9	11	0.64	0.79
10	1	2	0.07	0.14
8	1	1	0.07	0.07

Standard deviation calculation

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

Question #1

4, 1, 10, 7, 3

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
4	-1	1
1	-4	16
10	5	25
7	2	4
3	-2	4

$\bar{X} = 5$ $\quad\quad\quad\quad\quad\quad\quad\quad s^2 = 50/4 = 12.5$

$SS = 50$ $\quad\quad\quad\quad\quad\quad\quad\quad s = \sqrt{12.5} = 3.54$

$df = 5 - 1 = 4$

Question #2

2, 7, 5, 10, 6

Question #3

4, 7, 5, 3, 1

Question #4

4, 7, 8, 6, 5

Question #5

3, 4, 8, 9, 6

Question #6

7, 9, 1, 3, 5

Week #4

1. Last week we saw that

   (a) If you subtract the same number from every score, the new mean is the old mean minus that number
   (b) If you divide every score by the same number, the new standard deviation is the old standard deviation divided by that number

   So if we subtract the mean from every number and divide every number by the standard deviation, what is the new mean and standard deviation?

2. Why do we prefer standard scores over raw scores?

3. What does the area under the normal curve tell us about numbers below?

4. On the first number line below, draw bars representing the frequency distribution of the following data:

   4, 6, 4, 3, 8

   Then, subtract the mean from each score and draw the new distribution on the second number line. Then, divide those by the standard deviation and draw them on the third number line. What does each step do to the location, spread, and shape of the distribution?
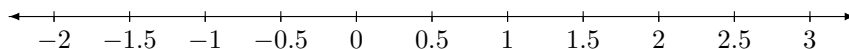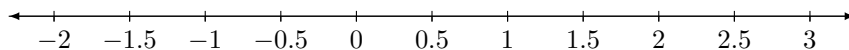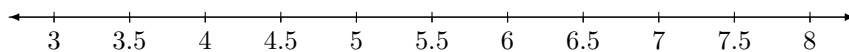
Z-scores

To get the area between z_1 and z_2, subtract the area above the lesser of the two minus the area above the greater of the two. The area below $z = 1$ - (the area above it), the area below $-z$ = the area above z, the area above $-z = 0.5 +$ (the area between z and the mean).

Calculate the area between the following pairs of z scores:

z_1	z_2	Area between z_1 and z_2
-0.47	-0.02	0.68 - 0.51 = 0.17
-0.26	0.19	0.6 - 0.42 = 0.18
0.13	-0.01	0.5 - 0.45 = 0.06
0.79	-0.65	0.74 - 0.21 = 0.53
-0.6	0.62	0.73 - 0.27 = 0.46
0.75	0.31	
0.83	0.55	
0.27	-0.8	
0.21	0.39	
-0.88	-0.19	
-0.61	0.56	
-0.66	0.22	
0.29	0.47	
-0.28	0.03	
0.43	0.97	
-0.08	0.46	
0.32	-0.95	
0.81	-0.12	
-0.31	0.33	
0.4	0.23	
0.68	-0.13	
-0.62	0.53	
0.16	-0.2	
-0.77	-0.57	
-0.53	-0.87	
-0.32	-0.83	
-0.97	-0.45	
-0.34	-0.11	
0.49	0.1	
-0.42	-0.3	

z	Area between mean and z	Area above z
0.01	0.4960	0.0040
0.02	0.4920	0.0080
0.03	0.4880	0.0120
0.08	0.4681	0.0319
0.10	0.4602	0.0398
0.11	0.4562	0.0438
0.12	0.4522	0.0478
0.13	0.4483	0.0517
0.13	0.4483	0.0517
0.16	0.4364	0.0636
0.19	0.4247	0.0753
0.20	0.4207	0.0793
0.21	0.4168	0.0832
0.22	0.4129	0.0871
0.23	0.4090	0.0910
0.26	0.3974	0.1026
0.27	0.3936	0.1064
0.28	0.3897	0.1103
0.29	0.3859	0.1141
0.30	0.3821	0.1179
0.31	0.3783	0.1217
0.31	0.3783	0.1217
0.32	0.3745	0.1255
0.32	0.3745	0.1255
0.33	0.3707	0.1293
0.34	0.3669	0.1331
0.39	0.3483	0.1517
0.40	0.3446	0.1554
0.42	0.3372	0.1628
0.43	0.3336	0.1664
0.45	0.3264	0.1736
0.46	0.3228	0.1772
0.47	0.3192	0.1808
0.49	0.3121	0.1879
0.53	0.2981	0.2019
0.55	0.2912	0.2088
0.56	0.2877	0.2123
0.57	0.2843	0.2157
0.60	0.2743	0.2257
0.61	0.2709	0.2291
0.62	0.2676	0.2324
0.62	0.2676	0.2324
0.65	0.2578	0.2422
0.66	0.2546	0.2454
0.68	0.2483	0.2517
0.75	0.2266	0.2734
0.77	0.2206	0.2794
0.79	0.2148	0.2852
0.80	0.2119	0.2881
0.81	0.2090	0.2910
0.83	0.2033	0.2967
0.83	0.2033	0.2967
0.87	0.1922	0.3078
0.88	0.1894	0.3106
0.95	0.1711	0.3289
0.97	0.1660	0.3340

Week #5

1. What is the difference between standard error and standard deviation?

2. Why is standard error so named? What does it measure?

3. What is the difference between standard error and standard error of the mean?

4. How do changes in n and σ affect $\sigma_{\bar{X}}$? Explain with reference to the formula

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

5. What is the difference between a sampling distribution and a sampling distribution of the mean?

6. Distinguish between the population distribution, the sample distribution, and the sampling distribution, using an example

7. Distinguish between the population mean, the sample mean, and the mean of the sampling distribution of the mean, using an example

8. Imagine a bowl with five bingo chips in it, numbered 1 through 5. For this population:

   (a) Calculate the population mean

   (b) List all 25 possible samples of size 2

$$1, 1$$
$$1, 2$$
$$\vdots$$
$$5, 5$$

   (c) Calculate the mean for every sample

   (d) Count the frequency of each value of the sample mean

   (e) Calculate the mean of the sampling distribution of the mean

   (f) Draw the frequency distribution as a bar plot

   (g) Calculate the probability of drawing a sample with:
   - A mean $>= 4$
   - A mean $<= 2.5$
   - An error of at least 1.5

Week #6

1. What are the steps in hypothesis testing? Give an example and relate each step to the concepts we've learned so far in the course (population, sample, parameter, statistic, sampling distribution of the mean, null and alternate hypotheses, critical values, etc.)

2. Draw a null distribution and illustrate the relationship between the critical values, the rejection region, and α

3. Draw an alternate distribution and illustrate the relationship between the critical values, the rejection region, β, and power

4. How are α and β each affected by change in the critical values, effect size, and standard error?

5. How do we control Type I and Type II error, respectively? Why don't we make $\alpha = 0$?

6. 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. Here's an example to get you started:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

7. Why do we need hypothesis testing? Why can't we just draw a sample from the population and estimate the parameter by the sample statistic?

8. Give an example of a finding which is statistically significant but practically insignificant. When might this occur?

9. What is the goal of a confidence interval?

10. Give an example of a research question where a nondirectional test would be appropriate, then reframe it so that a directional test would be appropriate

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 6 with a mean of 5. The population standard deviation is known to be 2.03. Test $H_0 : \mu = 6$ at an α of 0.1, state your decision, and calculate a 90% confidence interval.

$\sigma_{\bar{X}} = 2.03/\sqrt{6} = 0.83$

$z_{\text{obs}} = (5 - 6)/0.83 = -1.21$

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

Fail to reject because $1.64 > -1.21 > -1.64$

$z_{90} = 1.64$

$CI_{90} = 5 \pm (0.83 \times 1.64) = [3.64, \ 6.36]$

Question #2

Researchers draw a sample of 6 with a mean of 6.33. The population standard deviation is known to be 3.81. Test $H_0 :$ $\mu = 6$ at an α of 0.05, state your decision, and calculate a 95% confidence interval.

Question #3

Researchers draw a sample of 6 with a mean of 4.83. The population standard deviation is known to be 4.23. Test $H_0 :$ $\mu = 4$ at an α of 0.05, state your decision, and calculate a 95% confidence interval.

Question #4

Researchers draw a sample of 8 with a mean of 6.5. The population standard deviation is known to be 1.04. Test $H_0 :$ $\mu = 3$ at an α of 0.1, state your decision, and calculate a 90% confidence interval.

Question #5

Researchers draw a sample of 6 with a mean of 6.33. The population standard deviation is known to be 3.74. Test $H_0 :$ $\mu = 10$ at an α of 0.1, state your decision, and calculate a 90% confidence interval.

Question #6

Researchers draw a sample of 8 with a mean of 6.88. The population standard deviation is known to be 4.75. Test $H_0 :$ $\mu = 3$ at an α of 0.01, state your decision, and calculate a 99% confidence interval.

Question #7

Researchers draw a sample of 10 with a mean of 4.6. The population standard deviation is known to be 2.59. Test $H_0 :$ $\mu = 2$ at an α of 0.1, state your decision, and calculate a 90% confidence interval.

Question #8

Researchers draw a sample of 7 with a mean of 6.71. The population standard deviation is known to be 1.83. Test $H_0 :$ $\mu = 8$ at an α of 0.05, state your decision, and calculate a 95% confidence interval.

Question #9

Researchers draw a sample of 6 with a mean of 4.17. The population standard deviation is known to be 1.1. Test $H_0 :$ $\mu = 3$ at an α of 0.01, state your decision, and calculate a 99% confidence interval.

Question #10

Researchers draw a sample of 8 with a mean of 4.5. The population standard deviation is known to be 2.23. Test $H_0 :$ $\mu = 5$ at an α of 0.05, state your decision, and calculate a 95% confidence interval.

Question #11

Researchers draw a sample of 6 with a mean of 4.17. The population standard deviation is known to be 1. Test $H_0 :$ $\mu = 6$ at an α of 0.1, state your decision, and calculate a 90% confidence interval.

Week #7

1. When should we use a t test instead of a z test?

2. What is the expected value of t under the null hypothesis?

3. Why can't we calculate a t statistic with a sample of $n = 1$?

4. Give an example of a research question for which a one-sample t-test would be appropriate, and explain why it is 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

df	α 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

You draw a sample of 5 with a mean of 6.4 and a standard deviation of 3.05. Test $H_0 : \mu = 7$ at an α of 0.01, state your decision, then calculate a 99% confidence interval.

$$s_{\bar{X}} = 3.05/\sqrt{5} = 1.36$$
$$t_{\text{tobs}} = (6.4 - 7)/1.36 = -0.44$$
$$t_{\text{crit}} = \pm 4.6$$
Fail to reject because $4.6 > -0.44 > -4.6$
$$t_{99} = 4.6$$
$$CI_{99} = 6.4 \pm (1.36 \times 4.6) = [0.12,\ 12.68]$$

Question #2

You draw a sample of 5 with a mean of 7.2 and a standard deviation of 3.35. Test $H_0 : \mu = 5$ at an α of 0.1, state your decision, then calculate a 90% confidence interval.

Question #3

You draw a sample of 5 with a mean of 6.2 and a standard deviation of 2.17. Test $H_0 : \mu = 6$ at an α of 0.1, state your decision, then calculate a 90% confidence interval.

Question #4

You draw a sample of 6 with a mean of 4.5 and a standard deviation of 3.21. Test $H_0 : \mu = 1$ at an α of 0.01, state your decision, then calculate a 99% confidence interval.

Question #5

You draw a sample of 5 with a mean of 6 and a standard deviation of 3.39. Test $H_0 : \mu = 10$ at an α of 0.05, state your decision, then calculate a 95% confidence interval.

Question #6

You draw a sample of 6 with a mean of 7.67 and a standard deviation of 3.27. Test $H_0 : \mu = 1$ at an α of 0.01, state your decision, then calculate a 99% confidence interval.

Question #7

You draw a sample of 7 with a mean of 4.57 and a standard deviation of 3.26. Test $H_0 : \mu = 2$ at an α of 0.1, state your decision, then calculate a 90% confidence interval.

Question #8

You draw a sample of 6 with a mean of 5.5 and a standard deviation of 2.43. Test $H_0 : \mu = 7$ at an α of 0.01, state your decision, then calculate a 99% confidence interval.

Question #9

You draw a sample of 5 with a mean of 2.8 and a standard deviation of 1.48. Test $H_0 : \mu = 7$ at an α of 0.05, state your decision, then calculate a 95% confidence interval.

Question #10

You draw a sample of 6 with a mean of 3.83 and a standard deviation of 2.04. Test $H_0 : \mu = 5$ at an α of 0.1, state your decision, then calculate a 90% confidence interval.

Week #8

1. Give an example of a research question for which you would use an independent t-test, and explain why it is appropriate

2. What is $s_{\bar{X}_1 - \bar{X}_2}$ and what does it measure?

3. What is homogeneity of variance? State it symbolically

4. When might we employ a Welch correction?

5. What is the reasoning behind using pooled variance? Why not simply take the mean of s_1 and s_2?

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

df	α		
	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 7 with a mean of 6.71 and a variance of 12.24, and another sample of 6 with a mean of 5.67 and a variance of 5.87. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.1, state the error, then calculate a 90% confidence interval.

$$df_1 = 7 - 1 = 6$$
$$df_2 = 6 - 1 = 5$$
$$SS_1 = 12.24 \times 6 = 73.44$$
$$SS_2 = 5.87 \times 5 = 29.35$$
$$df_{\text{tot}} = 6 + 5 = 11$$
$$s_p^2 = (73.44 + 29.35)/11 = 9.34$$
$$s_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{(9.34/7) + (9.34/6)} = 1.7$$
$$t_{\text{obs}}(11) = (6.71 - 5.67)/1.7 = 0.61$$
$$t_{\text{crit}} = \pm 1.8$$
Fail to reject because $1.8 > 0.61 > -1.8$
$$t_{90} = 1.8$$
$$CI_{90} = (6.71 - 5.67) \pm (1.7 \times 1.8) = [-2.01, \ 4.09]$$

Question #2

Researchers draw one sample of 7 with a mean of 6.29 and a variance of 12.24, and another sample of 7 with a mean of 5.71 and a variance of 10.57. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.1, state the error, then calculate a 90% confidence interval.

Question #3

Researchers draw one sample of 6 with a mean of 4.83 and a variance of 4.57, and another sample of 7 with a mean of 6.29 and a variance of 5.57. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.1, state the error, then calculate a 90% confidence interval.

Question #4

Researchers draw one sample of 6 with a mean of 7.17 and a variance of 10.97, and another sample of 6 with a mean of 5.5 and a variance of 9.1. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.05, state the error, then calculate a 95% confidence interval.

Question #5

Researchers draw one sample of 7 with a mean of 5.71 and a variance of 11.57, and another sample of 6 with a mean of 4.17 and a variance of 1.37. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.1, state the error, then calculate a 90% confidence interval.

Question #6

Researchers draw one sample of 6 with a mean of 7.83 and a variance of 5.77, and another sample of 6 with a mean of 5.33 and a variance of 12.27. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.05, state the error, then calculate a 95% confidence interval.

Question #7

Researchers draw one sample of 6 with a mean of 4.67 and a variance of 12.67, and another sample of 6 with a mean of 3.33 and a variance of 7.47. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.01, state the error, then calculate a 99% confidence interval.

Question #8

Researchers draw one sample of 7 with a mean of 6.71 and a variance of 5.9, and another sample of 6 with a mean of 4.17 and a variance of 12.57. Test $H_0 : \mu_1 = \mu_2$ at an α of 0.05, state the error, then calculate a 95% confidence interval.

Week #9

1. Give an example of a research design in which you would use a dependent t-test, and explain why it is appropriate

2. How does using a dependent design rather than an independent design affect power? Why?

3. Give an example of a research question which could be addressed by ANOVA but not a single t-test

4. What do the numerator and denominator of the F statistic represent? Why is F is equal to 1 under the null hypothesis?

$$F = \frac{MS_B}{MS_W}$$

$MS_B = $ "mean squares between"

$MS_W = $ "mean squares within"

5. 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

$df = n - 1$

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

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

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

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

$CI_y = \bar{D} \pm s_{\bar{D}} \times t_y$

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

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

Pre	Post	D_i	$D_i - \bar{D}$	$(D_i - \bar{D})^2$
5	5	0	2	4
7	4	3	5	25
1	7	-6	-4	16
9	10	-1	1	1
3	9	-6	-4	16

$\bar{D} = -2$

$\Sigma(D_i - \bar{D})^2 = 62$

$df = 4$

$s_D = \sqrt{62/4} = 3.94$

$s_{\bar{D}} = 3.94/\sqrt{5} = 1.76$

$t_{\text{obs}}(4) = -2/1.76 = -1.14$

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

Fail to reject because $2.78 > -1.14 > -2.78$

$t_{95} = 2.78$

$CI_{95} = -2 \pm (1.76 \times 2.78) = [-6.89,\ 2.89]$

Question #2

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

Pre	Post
2	2
8	8
4	7
9	4
7	9

Question #3

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

Pre	Post
6	2
5	10
9	7
7	6
8	5

Question #4

Test $H_0 : \mu_{\bar{D}} = 0$ at an α of 0.1, state the decision/error, then calculate a 90% confidence interval.

Pre	Post
10	9
9	5
1	7
6	4
2	8
8	3

Question #5

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

Pre	Post
7	4
9	8
3	5
1	6
5	7

Week #10

1. Give an example of two variables which would have a correlation of close to 1

2. Give an example of two variables which would have a correlation of close to -1

3. Give an example of two variables which would have a correlation of close to 0

4. Draw a a relationship which cannot be accurately described by Pearson's r

5. Why doesn't correlation imply causation? Does causation imply correlation?

6. Why does $df = n - 2$ when looking up critical r values?

7. Rank each set of scores as though you were calculating Spearman's rho:

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

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	4	-1	-3	1	9	3
6	5	1	-2	1	4	-2
8	9	3	2	9	4	6
2	10	-3	3	9	9	-9

$\bar{X} = 5$

$\bar{Y} = 7$

$SS_X = 20$

$SS_Y = 26$

$SP = -2$

$r_{XY} = -2/\sqrt{20 \times 26} = -0.09$

$r_{\text{crit}} = \pm 0.8$

Fail to reject because $0.8 > -0.09 > -0.8$

Question #2

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

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

$\bar{X} = 7$

$\bar{Y} = 6$

$SS_X = 10$

$SS_Y = 26$

$SP = -1$

$r_{XY} = -1/\sqrt{10 \times 26} = -0.06$

$r_{\text{crit}} = \pm 0.9$

Fail to reject because $0.9 > -0.06 > -0.9$

Question #3

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

X_i	Y_i
8	5
6	3
10	9
4	7

Question #4

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

X_i	Y_i
10	3
5	8
4	6
9	4
7	9

Question #5

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

X_i	Y_i
7	6
9	3
3	2
1	9

Question #6

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

X_i	Y_i
10	6
3	7
4	2
6	5
7	10

Question #7

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

X_i	Y_i
4	10
8	2
2	9
3	3
7	5
6	7

Question #8

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

X_i	Y_i
5	1
8	9
6	7
2	8
4	5

Question #9

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

X_i	Y_i
1	10
2	2
9	1
4	7

Question #10

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

X_i	Y_i
1	3
5	5
10	6
6	1
3	10

Week #11

1. What does the regression line minimise? Draw a picture

2. How many regression lines are possible for a given data set?

3. Give an example of restriction of range

4. Explain the difference between a univariate outlier and a regression outlier, and draw a picture

5. Draw an example of how an influential outlier might affect the regression line

6. What is a residual and what does it tell us about 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	Y_i	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$\hat{Y}_i$	$(\hat{Y}_i - \bar{Y})$	$(\hat{Y}_i - \bar{Y})^2$
6	7	0	4	0	5.00	0.00	0.00
9	2	9	9	-9	5.23	0.23	0.05
2	3	16	4	8	4.69	-0.31	0.09
7	8	1	9	3	5.08	0.08	0.01

$SS_X = 26$

$SP = 2$

$\beta_1 = 2/26 = 0.08$

$\bar{Y} = 5$

$\bar{X} = 6$

$\beta_0 = 5 - (0.08 \times 6) = 4.54$

$\hat{Y}_i = 4.54 + (0.08 \times X_i)$

$SS_{\text{tot}} = 26$

$SS_{\text{reg}} = 0.15$

$SS_{\text{res}} = 26 - 0.15 = 25.85$

$df_1 = 1$

$df_2 = 4 - 1 - 1 = 2$

$MS_{\text{reg}} = 0.15/1 = 0.15$

$MS_{\text{res}} = 25.85/2 = 12.92$

$F = 0.15/12.92 = 0.01$

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

Fail to reject because $0.01 < 18.51$

Question #2

Test the model fit at an α of 0.01.

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

Question #6

Test the model fit at an α of 0.01.

X_i	Y_i	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
7	1	1	25	-5
2	9	16	9	-12
6	7	0	1	0
10	5	16	1	-4
5	8	1	4	-2

Question #3

Test the model fit at an α of 0.01.

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

Question #7

Test the model fit at an α of 0.05.

X_i	Y_i	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
9	10	4	16	8
5	2	4	16	8
7	6	0	0	0
3	5	16	1	4
10	4	9	4	-6
8	9	1	9	3

Question #4

Test the model fit at an α of 0.05.

X_i	Y_i	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
4	4	4	1	2
6	10	0	25	0
5	6	1	1	-1
7	3	1	4	-2
8	2	4	9	-6

Question #8

Test the model fit at an α of 0.05.

X_i	Y_i	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
5	10	0	25	0
10	2	25	9	-15
2	4	9	1	3
7	3	4	4	-4
1	6	16	1	-4

Question #5

Test the model fit at an α of 0.01.

X_i	Y_i	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
6	3	0	1	-0
8	2	4	4	-4
9	4	9	0	0
1	7	25	9	-15

Question #9

Test the model fit at an α of 0.05.

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