DATA 606 FALL 2022 - Final Exam

Please put your answers in the `Final_Exam_Answers.Rmd` file and submit either the PDF or HTML file.
DO NOT POST YOUR EXAM ON RPUBS OR GITHUB! DO NOT SHARE.

Part I

1. A professor was curious about her students' grade point averages (GPAs). She took a random sample of 15 students and found a mean GPA of 3.01 with a standard deviation of 0.534. Which of the following formulas gives a 99% confidence interval for the mean GPA of the professor's students?

   A) $3.01 \pm 2.947(0.534/\sqrt{15})$
   B) $3.01 \pm 2.977(0.534/\sqrt{15})$
   C) $3.01 \pm 2.576(0.534/\sqrt{15})$
   D) $3.01 \pm 2.947(0.534/\sqrt{14})$
   E) $3.01 \pm 2.977(0.534/\sqrt{14})$

2. Absorption rates into the body are important considerations when manufacturing a generic version of a brand-name drug. A pharmacist read that the absorption rate into the body of a new generic drug (G) is the same as its brand-name counterpart (B). She has a researcher friend of hers run a small experiment to test $H_0 : \mu_G - \mu_B = 0$ against the alternative $H_A : \mu_G - \mu_B \neq 0$. Which of the following would be a Type I error?

   A) Deciding that the absorption rates are different, when in fact they are not.
   B) Deciding that the absorption rates are different, when in fact they are.
   C) Deciding that the absorption rates are the same, when in fact they are not.
   D) Deciding that the absorption rates are the same, when in fact they are.
   E) The researcher cannot make a Type I error, since he has run an experiment.

3. At one vehicle inspection station, 13 of 52 trucks and 11 of 88 cars failed the emissions test. Assuming these vehicles were representative of the cars and trucks in that area, what is the standard error of the difference in the percentages of all cars and trucks that are not in compliance with air quality regulations?

   A) 0.025
   B) 0.032
   C) 0.049
   D) 0.070
   E) 0.095

4. At one SAT test site students taking the test for a second time volunteered to inhale supplemental oxygen for 10 minutes before the test. In fact, some received oxygen, but others (randomly assigned) were given just normal air. Test results showed that 42 of 66 students who breathed oxygen improved their SAT scores, compared to only 35 of 63 students who did not get the oxygen. Which procedure should we use to see if there is evidence that breathing extra oxygen can help test- takers think more clearly?

   A) 1-proportion z-test
   B) 2-proportion z-test
   C) 1-sample t-test
   D) 2-sample t-test
   E) matched pairs t-test

5. We have calculated a 95% confidence interval and would prefer for our next confidence interval to have a smaller margin of error without losing any confidence. In order to do this, we can

   I. change the z* value to a smaller number.

   II. take a larger sample.
   III. take a smaller sample.

   A) I only
   B) II only
   C) III only
   D) I and II
   E) I and III

6. We have calculated a confidence interval based on a sample of size n = 100. Now we want to get a better estimate with a margin of error that is only one-fourth as large. How large does our new sample need to be?

   A) 25
   B) 50
   C) 200
   D) 400
   E) 1600

7. A P-value indicates

   A) the probability that the null hypothesis is true.
   B) the probability that the alternative hypothesis is true.
   C) the probability the null is true given the observed statistic.
   D) the probability of the observed statistic given that the null hypothesis is true.
   E) the probability of the observed statistic given that the alternative hypothesis is true.

8. Suppose that a manufacturer is testing one of its machines to make sure that the machine is producing more than 97% good parts (H_0 : p = 0.97 and H_A : p > 0.97). The test results in a P-value of 0.122. Unknown to the manufacturer, the machine is actually producing 99% good parts. What probably happens as a result of the testing?

   A) They correctly fail to reject H_0.
   B) They correctly reject H_0 .
   C) They reject H_0 , making a Type I error.
   D) They fail to reject H_0 , making a Type I error. E) They fail to reject H_0 , making a Type II error.

9. When the correlation coefficient, r, is close to one:

   A) there is no relationship between the two variables
   B) there is a strong linear relationship between the two variables
   C) it is impossible to tell if there is a relationship between the two variables
   D) the slope of the regression line will be close to one

10. The intercept in linear regression represents:

   A) the strength of the relationship between x and y
   B) the expected x value when y is zero
   C) the expected y value when x is zero
   D) a population parameter

Part II

Consider the three datasets, each with two columns (x and y), provided below.

```
data1 <- data.frame(x = c(55.3846, 51.5385, 46.1538, 42.8205, 40.7692, 38.7179, 35.641,
    33.0769, 28.9744, 26.1538, 23.0769, 22.3077, 22.3077, 23.3333, 25.8974, 29.4872,
    32.8205, 35.3846, 40.2564, 44.1026, 46.6667, 50, 53.0769, 56.6667, 59.2308, 61.2821,
    61.5385, 61.7949, 57.4359, 54.8718, 52.5641, 48.2051, 49.4872, 51.0256, 45.3846,
    42.8205, 38.7179, 35.1282, 32.5641, 30, 33.5897, 36.6667, 38.2051, 29.7436, 29.7436,
    30, 32.0513, 35.8974, 41.0256, 44.1026, 47.1795, 49.4872, 51.5385, 53.5897, 55.1282,
    56.6667, 59.2308, 62.3077, 64.8718, 67.9487, 70.5128, 71.5385, 71.5385, 69.4872,
    46.9231, 48.2051, 50, 53.0769, 55.3846, 56.6667, 56.1538, 53.8462, 51.2821, 50,
    47.9487, 29.7436, 29.7436, 31.2821, 57.9487, 61.7949, 64.8718, 68.4615, 70.7692,
    72.0513, 73.8462, 75.1282, 76.6667, 77.6923, 79.7436, 81.7949, 83.3333, 85.1282,
    86.4103, 87.9487, 89.4872, 93.3333, 95.3846, 98.2051, 56.6667, 59.2308, 60.7692,
    63.0769, 64.1026, 64.359, 74.359, 71.2821, 67.9487, 65.8974, 63.0769, 61.2821,
    58.7179, 55.1282, 52.3077, 49.7436, 47.4359, 44.8718, 48.7179, 51.2821, 54.1026,
    56.1538, 52.0513, 48.7179, 47.1795, 46.1538, 50.5128, 53.8462, 57.4359, 60, 64.1026,
    66.9231, 71.2821, 74.359, 78.2051, 67.9487, 68.4615, 68.2051, 37.6923, 39.4872,
    91.2821, 50, 47.9487, 44.1026), y = c(97.1795, 96.0256, 94.4872, 91.4103, 88.3333,
    84.8718, 79.8718, 77.5641, 74.4872, 71.4103, 66.4103, 61.7949, 57.1795, 52.9487,
    51.0256, 51.0256, 51.0256, 51.4103, 51.4103, 52.9487, 54.1026, 55.2564, 55.641,
    56.0256, 57.9487, 62.1795, 66.4103, 69.1026, 55.2564, 49.8718, 46.0256, 38.3333,
    42.1795, 44.1026, 36.4103, 32.5641, 31.4103, 30.2564, 32.1795, 36.7949, 41.4103,
    45.641, 49.1026, 36.0256, 32.1795, 29.1026, 26.7949, 25.2564, 25.2564, 25.641,
    28.718, 31.4103, 34.8718, 37.5641, 40.641, 42.1795, 44.4872, 46.0256, 46.7949,
    47.9487, 53.718, 60.641, 64.4872, 69.4872, 79.8718, 84.1026, 85.2564, 85.2564,
    86.0256, 86.0256, 82.9487, 80.641, 78.718, 78.718, 77.5641, 59.8718, 62.1795,
    62.5641, 99.4872, 99.1026, 97.5641, 94.1026, 91.0256, 86.4103, 83.3333, 79.1026,
    75.2564, 71.4103, 66.7949, 60.2564, 55.2564, 51.4103, 47.5641, 46.0256, 42.5641,
    39.8718, 36.7949, 33.718, 40.641, 38.3333, 33.718, 29.1026, 25.2564, 24.1026,
    22.9487, 22.9487, 22.1795, 20.2564, 19.1026, 19.1026, 18.3333, 18.3333, 18.3333,
    17.5641, 16.0256, 13.718, 14.8718, 14.8718, 14.8718, 14.1026, 12.5641, 11.0256,
    9.8718, 6.0256, 9.4872, 10.2564, 10.2564, 10.641, 10.641, 10.641, 10.641, 10.641,
    10.641, 8.718, 5.2564, 2.9487, 25.7692, 25.3846, 41.5385, 95.7692, 95, 92.6923))
```

```r
data2 <- data.frame(x = c(52.8720214902, 59.0141444945, 56.3751090389, 37.8391996844,
    39.8853734176, 44.0774051326, 31.4970254888, 25.9525965456, 36.4347239148, 24.997482691,
    32.2462751352, 18.1094722949, 21.7960992263, 24.150491069, 28.6031956015, 20.9319996761,
    35.4355325251, 44.3773828236, 46.313692406, 46.3860739231, 42.4954400876, 44.3005150328,
    46.9882833458, 52.4215231623, 64.090998189, 63.0168717563, 67.3025779226, 67.2259568904,
    50.309988715, 45.4126389249, 54.0206071815, 49.8420222483, 50.7132995657, 55.0455549106,
    37.5437012597, 39.7958978066, 28.5233910215, 31.3996460012, 39.2415652162, 27.7140565712,
    32.9607035213, 35.4047362397, 51.6422264225, 34.207620843, 36.5234892198, 28.4192021281,
    26.1604969417, 40.435327397, 44.544955314, 40.1237272008, 42.1848064503, 40.0026592892,
    57.1252310509, 60.375744968, 55.1058961252, 61.7372710245, 55.7412443246, 58.2964084158,
    57.0425482213, 76.7756112073, 71.8870772602, 73.7883730663, 75.1396847343, 64.5963189058,
    46.9095386323, 46.9051031994, 49.8069877356, 48.8632418189, 66.7481899024, 62.3597662841,
    47.3748964954, 47.4464667165, 53.385280634, 49.4527077733, 54.8026825363, 37.0882451339,
    32.3759619238, 37.4770138847, 52.4066488967, 50.2970432655, 61.5633243835, 61.0865603919,
    59.2364904011, 66.1540050863, 67.2313347767, 73.7099068702, 76.5804513576, 78.8765811639,
    77.4490021363, 86.8537073476, 79.806531301, 76.9882734745, 79.4793575017, 82.3041595613,
    90.700691247, 95.4434878077, 94.9974880471, 91.517847246, 62.2926421901, 59.2034756073,
    65.2563927704, 62.4447424979, 72.4624062907, 69.9194262552, 77.7628590782, 72.7784101997,
    65.198316667, 69.4598588916, 59.3150555205, 51.487713355, 57.1124475509, 50.2244178157,
    50.4338611907, 56.3803054046, 54.8635045999, 45.4998576493, 49.9367453057, 53.8098729164,
    59.8526782804, 48.1756567942, 50.7417190667, 58.4597612975, 50.1145975193, 45.4203820755,
    51.3532681954, 53.091704424, 58.9431665444, 59.3109904219, 68.2221986595, 62.5117719862,
    69.3800563617, 82.8002478704, 85.1603321777, 67.095847697, 62.307750573, 59.9704150458,
    39.9116351739, 39.5965009783, 95.5934163965, 52.772288113, 53.1786144535, 37.6270534612),
    y = c(97.3432226575, 93.5748748733, 96.30514763, 94.359443503, 90.6346620817,
        84.1258533023, 67.9125621568, 73.0331803541, 62.2940454639, 75.4415616626,
        70.1105868383, 53.7081103324, 49.4995295248, 47.1885335266, 44.2237227, 51.6462354868,
        34.9610381566, 57.6847369215, 55.6855917703, 51.0021645381, 56.9036815315,
        58.1329592781, 55.4206367313, 47.867005488, 68.6509850107, 68.2252589114,
        64.4134627173, 62.4764105301, 47.2678818553, 53.8707976895, 47.2698972057,
        50.5765522936, 48.7722867825, 45.9862727459, 32.8840244657, 28.4432191156,
        39.7966662481, 39.2652632229, 33.3614166846, 42.2500516253, 35.8231112903,
        31.9391621952, 50.8383406788, 33.5938054385, 35.6710483891, 21.3157482481,
        23.0422314726, 27.6400711421, 27.3622426221, 30.0731619402, 28.5736747331,
        27.4875100027, 42.7390191388, 41.3564579961, 46.1490510561, 38.5848440052,
        46.6535769763, 41.5831977346, 42.418757739, 55.2545987009, 59.2669926152,
        58.6208121164, 54.2445521781, 65.952512925, 85.6022611586, 81.5656424136,
        82.085025592, 83.1583574877, 84.0048341677, 90.2436599196, 83.4715555182,
        82.622124014, 78.8711362558, 79.9306028004, 75.3639849893, 60.857227495,
        68.5108204552, 63.7886855744, 96.4254150057, 99.6441791736, 89.8641388605,
        89.2243556475, 91.4811005766, 84.1119370252, 85.055640808, 78.2706108976,
        72.8915336827, 71.5625276804, 72.4473094857, 65.5433240194, 52.7231797883,
        54.2027848022, 49.1410154455, 46.9909036215, 41.4667030063, 36.1897016148,
        35.0953011066, 36.8902631263, 40.0295333141, 41.2706776667, 32.8541041237,
        35.4149992261, 27.6242638254, 29.3110927191, 23.4913735562, 26.9005353695,
        33.0271717859, 30.1970315053, 12.3543136666, 20.8436106875, 15.3217851726,
        19.1709698793, 18.5348708515, 14.7884597101, 14.009611915, 3.98542629287,
        21.5900969196, 17.7664144512, 9.03994992133, 19.5212153543, 17.8384056087,
        14.1486628546, 1.74146171252, 5.37240899542, 0.30387242055, 19.2900304453,
        11.6128777918, 10.5721363739, 1.13388036584, 7.49123248279, 1.48813233299,
        18.5293762368, 15.0183695439, 0.601490941965, 5.74997471945, 7.9133104735,
        32.1440226145, 32.1361317349, 33.2341293953, 99.6134716833, 98.6038611855,
        93.060672952))
```

```
data3 <- data.frame(x = c(32.3311102266, 53.4214628807, 63.92020226, 70.2895057187,
    34.1188302357, 67.6707164012, 53.2591294055, 63.5149808671, 67.9805388133, 67.3724659005,
    15.5607495229, 71.7907676942, 70.2425464362, 64.9374355444, 62.2135245453, 67.2694004772,
    40.5701970446, 74.7411813341, 71.7683189223, 76.1669198143, 65.6236574431, 50.8506336394,
    33.0240700249, 39.7063261674, 45.5964849542, 42.9680469104, 52.4944067819, 46.0822757831,
    74.2477082092, 64.5682641863, 74.0216939058, 62.3911805626, 74.189036683, 28.2367819396,
    75.7719387944, 75.8552294691, 65.9708570175, 21.7780404779, 67.7597962473, 78.6171953363,
    68.5077081898, 74.8850211598, 66.4549036599, 77.3178020985, 58.9124603193, 57.617447817,
    76.0882257967, 57.4660505497, 79.4283834934, 76.3565221496, 64.4050752632, 40.6350418091,
    43.9498645857, 30.9962205791, 68.2307689907, 72.0463894612, 46.5927679682, 49.2572183396,
    42.7817612539, 65.475952195, 71.9650826544, 32.1464623358, 31.8384976954, 31.0052582572,
    80.4708943189, 71.9641671122, 78.0794214417, 41.6775957748, 65.953595185, 62.9344593731,
    64.3737979844, 72.5093283599, 30.0522898741, 28.0033242354, 75.4012268619, 38.9800154218,
    65.2199135479, 73.0539899616, 34.3983616372, 43.4904501336, 55.138737967, 43.6843934333,
    35.9036097344, 45.3780188805, 39.7774828713, 38.6644611569, 39.0440366877, 91.6399614428,
    47.4881326771, 44.5902125769, 39.0896145478, 42.2293783752, 37.0003871448, 39.0520864793,
    37.4884147432, 69.3595594592, 43.542775926, 39.8112302539, 70.0689259404, 70.0405435824,
    39.505789079, 62.5168908529, 72.1399254065, 45.2515760666, 42.0633045627, 36.3556951539,
    30.3918276596, 36.4490038543, 40.467576002, 81.7246168002, 48.8231974964, 35.6205617651,
    50.5839631148, 61.8564651063, 39.3237560262, 42.1856791429, 30.8469189898, 29.3462004281,
    82.1105579783, 38.3020058088, 56.5841530218, 33.3393742865, 78.7742390407, 27.9226442446,
    71.6978651182, 74.1383313856, 32.579020066, 59.83218542, 35.0306285457, 74.3001198284,
    63.2501970628, 34.1730737648, 40.9138319319, 62.8332930874, 42.4767923803, 52.0334562787,
    48.9070429644, 59.8518383854, 48.9604602016, 46.8448551673, 39.9630215796, 66.704943997),
    y = c(61.411101248, 26.1868803879, 30.8321939163, 82.5336485877, 45.7345513203,
        37.110947969, 97.4757710964, 25.1000785788, 80.9571652197, 29.720400203,
        80.0656402858, 71.0654666627, 24.1095975542, 81.5542049945, 21.4758389969,
        18.7089683725, 79.3729634752, 21.1016372041, 20.0110618423, 75.9361704048,
        15.5828033531, 13.9876016304, 24.4678303872, 84.2752871038, 9.76334884943,
        17.9454583961, 16.0511142003, 23.1104578154, 20.314187812, 83.6396338956,
        76.1282745076, 5.62307076073, 68.1335832223, 56.1395964513, 69.8292300322,
        62.5170442862, 72.7448559954, 6.61662530728, 72.4212015285, 52.5752573142,
        15.4569189652, 25.4166063231, 19.8366286542, 48.3983464352, 75.6677562173,
        8.19480060319, 59.6799300235, 1.50441817488, 45.2107942872, 10.4182411281,
        78.5841760758, 73.3947503698, 75.9587156671, 71.694404938, 80.8725016628,
        12.9180067349, 84.9723827774, 81.8814032306, 12.9911884302, 14.2745856444,
        17.7102359443, 43.4817094425, 71.8121653901, 40.682503007, 49.5021483467,
        41.8742826668, 93.1333167652, 30.2012640846, 31.1474060835, 31.9163906992,
        28.8625834061, 39.5401302526, 96.6175423534, 46.6721919544, 88.6390766207,
        87.322160691, 84.6829549336, 29.3808085571, 59.5444469033, 40.782542065,
        30.7257603575, 32.8230098696, 91.1118630801, 29.1692166544, 43.75581895,
        33.3172384774, 84.6760108316, 79.4066030605, 85.3899333808, 22.0340116412,
        70.4661940802, 19.9140684075, 60.264279248, 70.6525028457, 60.8144048511,
        65.5213545959, 62.4603112824, 65.3348328092, 7.59346560899, 77.1438066024,
        74.8516272173, 66.4847322418, 0.0151193251552, 70.0034213192, 2.33519661206,
        6.0058486497, 42.75961287, 50.5462690659, 60.0275120878, 6.03754484635, 76.6353305783,
        57.2860155789, 71.8066161014, 71.7927431642, 59.3008196656, 66.0348978235,
        37.3416401041, 42.1487418312, 1.21055166293, 60.0177857932, 70.512514809,
        0.5091067352, 35.4841012146, 25.9868781844, 10.8681445111, 49.1739189791,
        1.80811559665, 69.1525081443, 12.5366493416, 42.4770945921, 65.9524861966,
        25.6936743092, 38.5590195509, 62.5108942269, 56.7312899691, 64.5666620298,
        74.2877488252, 72.9583909677, 72.6295257275, 36.7917136918, 42.9449148487,
        32.0150954299))
```

For each column, calculate (to four decimal places):

a. The mean (for x and y separately).

b. The median (for x and y separately).

c. The standard deviation (for x and y separately).

For each x and y pair, calculate (also to two decimal places):

d. The correlation.

e. Linear regression equation.

f. R-Squared.

g. For each pair, is it appropriate to estimate a linear regression model? Why or why not? Be specific as to why for each pair and include appropriate plots!

h. Explain why it is important to include appropriate visualizations when analyzing data. Include any visualization(s) you create.