STAT 305 D Exam 1

Show all your work.

1. (20 points) Caustic stress corrosion cracking of iron and steel has been studied because of failures around rivets in steel boilers and failures of steam rotors. A new teflon coating may reduce the corrosion behind this cracking. 10 steel bars were taken from manufacturer A, which uses the teflon coating, and 10 were taken from manufacturer B, which does not. Constant load stress corrosion tests (with constant load and constant stress) were applied for the same length of time to each bar. The length of the longest crack in μm was measured for each.

   a. (3 points) Identify the sample (or samples).

      - Sample 1: the 10 bars from manufacturer A.
      - Sample 2: the 10 bars from manufacturer B.

   b. (3 points) Identify the population (or populations).

      - Population 1: All the steel bars that could have been selected from manufacturer A.
      - Population 2: All the steel bars that could have been selected from manufactuer B.

   c. (3 points) Identify and classify all the variables.

      - Blocking (or concomitant) variable: manufacturer (A or B). Or, you could have said teflon (absent or present).
      - Response (or concomitant) variable: length of longest crack

   d. (3 points) Is this study an experimental study or an observational study?

      This is an observational study because the experimenters did not actively apply any treatment. Teflon could have been a treatment, but it was set by the separate manufacturers. As a result, and the experimental conditions are not constant with the level of teflon (absent or present).

   e. (4 points) Suppose the teflon-coated bars corrode and crack less for both steel and iron bars. Can we say that the teflon *causes* this reduction in corrosion and cracking? Why or why not?

      No: the experimenters did not set the levels of teflon themselves. The effect of teflon may be correlated with the manufacturer,m an experimental condition.

   f. (4 points) Suppose the teflon-coated bars corroded and cracked less than the ones that did not receive the teflon. Can we say that the teflon prevents corrosion and cracking? Why or why not?

No: another possibility is that manufacturer A is just better than manufacturer B, regardless of the teflon. We can't distinguish between the two effects because the experimenters did not set the level of teflon themselves, and the level of teflon was perfectly correlated with the choice of manufacturer.

2. (20 points)

   a. (10 points) Using the table of random digits below, select a simple random sample of 10 steel bars from a shipment of 100 from manufacturer A (see question 1). Also, select a simple random sample of 10 steel bars from a shipment of 100 from manufacturer B, continuing in the table of random digits from where you left off from the steel bars. Carefully describe how you did this.

27252	37875	53679	01889	35714	63534	63791	76342	47717	73684
93259	74585	11863	78985	03881	46567	93696	93521	54970	37601
84068	43759	75814	32261	12728	09636	22336	75629	01017	45503
68582	97054	28251	63787	57285	18854	35006	16343	51867	67979
60646	11298	19680	10087	66391	70853	24423	73007	74958	29020

   - A bars: first, I assign a 2-digit index to each steel bar in the shipment: 00, 01, . . ., 99. Next, I move along the top row of the table of random digits from left to right, selecting bars 27, 25, 23, 78, 75, 53, 67, 90, 18, and 89 for the study.
   - B bars: first, I assign a 2-digit index to each iron bar in the shipment: 00, 01, . . ., 99. Next, I continue where I left off in the random number table, moving left to right and selecting 35, 71, 46, 34, 63, 79, 17, 42, 47, and 77 for the study.

   b. (10 points) Suppose you want to vary the temperature (high or low) and see how the bars respond to the stress tests. Using a different table of random digits (below), randomize the 10 bars from manufacturer A to the two levels of temperature (5 high, 5 low). Then, do the same with the 10 bars from manufacturer B, continuing in the table of random digits from where you left off from the steel bars. Carefully describe how you did this.

Random Digits

12159	66144	05091	13446	45653	13684	66024	91410	51351	22772
30156	90519	95785	47544	66735	35754	11088	67310	19720	08379
59069	01722	53338	41942	65118	71236	01932	70343	25812	62275
54107	58081	82470	59407	13475	95872	16268	78436	39251	64247
99681	81295	06315	28212	45029	57701	96327	85436	33614	29070

   - First, I assign the 10 A bars in the sample an index from 0 to 9. Then, I move along the top row of the table table from left to right, selecting bars 1, 2, 5, 9, and 6 for the high temperature group. The other bars will be subjected to a low temperature.

- First, I assign the 10 B bars in the sample an index from 0 to 9. Then, I move along the top row of the table table from left to right, continuing where I left off from the A bars. I select bars 6, 1, 4, 0, and 5 for the high temperature group. The other bars will be in the low temperature group.

3. (20 points) Revisit the cars data from class.

mpg	wt
21.00	2.62
21.00	2.88
22.80	2.32
21.40	3.21
18.70	3.44
18.10	3.46
14.30	3.57
24.40	3.19
22.80	3.15
19.20	3.44
17.80	3.44
16.40	4.07
17.30	3.73
15.20	3.78
10.40	5.25
10.40	5.42
14.70	5.34
32.40	2.20
30.40	1.61
33.90	1.83
21.50	2.46
15.50	3.52
15.20	3.44
13.30	3.84
19.20	3.85
27.30	1.94
26.00	2.14
30.40	1.51
15.80	3.17
19.70	2.77
15.00	3.57
21.40	2.78

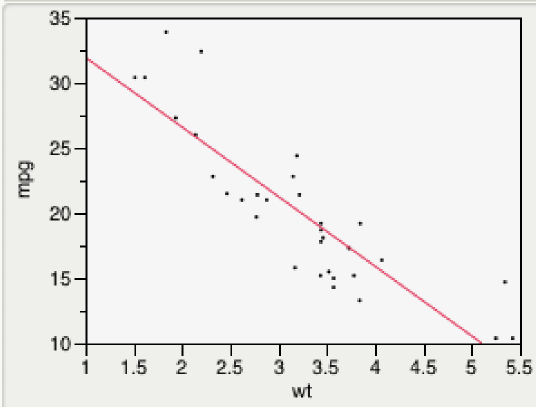
Remember:

- mpg is the fuel economy of the cars in miles per gallon.

- wt is the weight of the cars in tons.

Below, I fit a regression line of mpg on wt.

Bivariate Fit of mpg By wt



Linear Fit

Linear Fit

mpg = 37.285126 - 5.3444716*wt

Summary of Fit

RSquare	0.752833
RSquare Adj	0.744594
Root Mean Square Error	3.045882
Mean of Response	20.09062
Observations (or Sum Wgts)	32

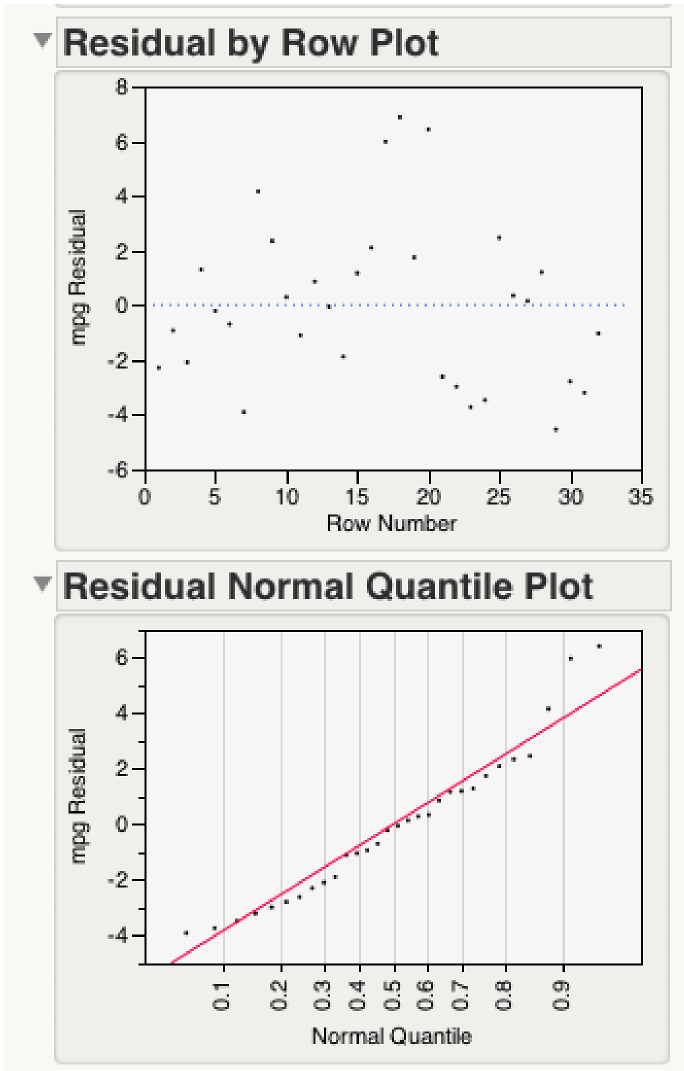
Lack Of Fit

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	1	847.7252	847.725	91.3753
Error	30	278.3219	9.277	**Prob > F**
C. Total	31	1126.0472		<.0001*

Parameter Estimates

Term	**Estimate**	**Std Error**	**t Ratio**	**Prob>|t|**
Intercept	37.285126	1.877627	19.86	<.0001*
wt	-5.344472	0.559101	-9.56	<.0001*



a. (4 points) Identify and interpret the slope.

The slope is -5.344 mpg/wt. On average, the fuel economy of the cars decreases by 5.344 miles per gallon with each ton of weight added..

b. (4 points) Identify and interpret the intercept.

The intercept is 37.385 mg/L. The model predicts that on average, a car weighing 0 tons will have a fuel economy of 37.285 mpg.

c. (4 points) What is problematic about your above interpretation of the intercept in practice?

The intercept predicts the fuel economy of a weightless car, which does not exist.

d. (4 points) Based on the residual plot, comment on the validity of the model.

. Since there is no apparent pattern in the residuals, the model appears valid.

e. (4 points) Based on the normal quantile (normal QQ) plot, do the residuals look bell-shaped (normally-distributed)?

Since the points in the normal QQ plot appear as a straight line, the residuals look normally distributed.

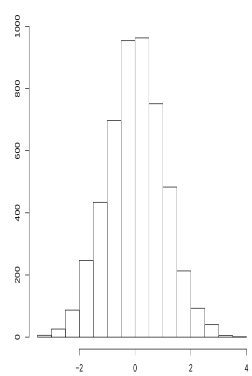
4. (20 points)

a. (10 points) What is the difference between a histogram and a bar plot?

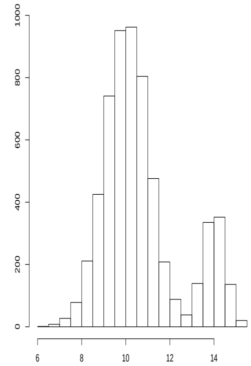
A histogram displays continuous numerical data by dividing the data into equally-sized intervals before plotting. A bar plot displays discrete or categorical data.

b. (10 points) Identify the following distributional shapes.

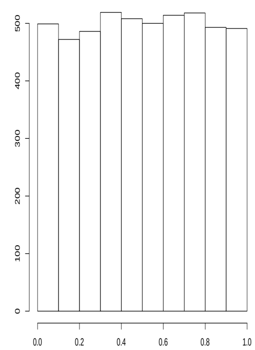
i. (2.5 points) bell-shaped



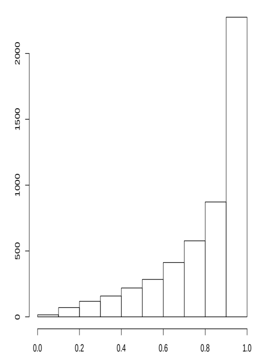
ii. (2.5 points) bimodal, asymmetric



iii. (2.5 points) uniform



iv. (2.5 points) skewed left



5. (20 points) One study reported on a study of strength properties of

high-performance concrete obtained by using superplasticizers and certain binders. The compressive strength of such concrete had previously been investigated, but not much was known about flexural strength (a measure of ability to resist failure in bending). Below is part of the flexural strength data in MegaPascals (MPa):

7.0 7.4 7.7 7.9 8 8.1 8.7 9.0 9.7 11.3 11.8 14

a. (10 points) Find $Q(0.25)$ and $Q(0.75)$ of the data.

Data	7.000	7.400	7.700	7.900	8.000	8.100	8.700	9.000	9.700	11.300	11.800	14.000
i	1.000	2.000	3.000	4.000	5.000	6.000	7.000	8.000	9.000	10.000	11.000	12.000
$\frac{i-.5}{12}$	0.042	0.125	0.208	0.292	0.375	0.458	0.542	0.625	0.708	0.792	0.875	0.958

For $Q(0.25)$:

$$\begin{aligned} i' &= np + 0.5 \\ &= 12 \cdot 0.25 + 0.5 \\ &= 3.5 \end{aligned}$$

Hence:

$$\begin{aligned} Q(0.25) &= (\lceil i' \rceil - i')x_{\lfloor i' \rfloor} + (i' - \lfloor i' \rfloor)x_{\lceil i' \rceil} \\ &= (\lceil 3.5 \rceil - 3.5)x_{\lfloor 3.5 \rfloor} + (3.5 - \lfloor 3.5 \rfloor)x_{\lceil 3.5 \rceil} \\ &= (4 - 3.5)x_3 + (3.5 - 3)x_4 \\ &= 0.5 \cdot 7.7 + 0.5 \cdot 7.9 \\ &= 7.8 \end{aligned}$$

For $Q(0.75)$:

$$\begin{aligned} i' &= np + 0.5 \\ &= 12 \cdot 0.75 + 0.5 \\ &= 9.5 \end{aligned}$$

Hence:

$$\begin{aligned} Q(0.75) &= (\lceil i' \rceil - i')x_{\lfloor i' \rfloor} + (i' - \lfloor i' \rfloor)x_{\lceil i' \rceil} \\ &= (\lceil 9.5 \rceil - 9.5)x_{\lfloor 9.5 \rfloor} + (9.5 - \lfloor 9.5 \rfloor)x_{\lceil 9.5 \rceil} \\ &= (10 - 9.5)x_9 + (9.5 - 9)x_{10} \\ &= 0.5 \cdot 9.7 + 0.5 \cdot 11.3 \\ &= 10.5 \end{aligned}$$

b. (10 points) Make a boxplot of the data. Is the distribution symmetric? Are there any outliers?

- The median, $Q(0.5)$, is the mean of the two middle data points, $(x_6 + x_7)/2 = 8.4$.
- 1.5 IQR $= 1.5(10.5 - 7.8) = 4.05$
- $Q(0.25) - 1.5IQR = 7.8 - 4.05 = 3.75$
- $Q(0.75) + 1.5IQR = 10.5 + 4.05 = 14.55$
- There are no points below $Q(0.25) - 1.5IQR$ or above $Q(0.75) + 1.5IQR$, so there are no outliers. The boxplot is shown below. The distribution looks skewed up (skewed right), not symmetric.