STAT 305 Section B Spring 2012
Exam I Version A
Show your work wherever possible.

1. (20 Points). In an effort to improve gas mileage of its vehicles, Fast-Cars Co. is running a study to determine if the type of rim a tire is placed on will improve gas mileage in the Model D cars produced at the Fast-Cars Co. Detroit plant. Two types of metal were chosen for the rim: aluminum and steel. Of all the Model D cars produced at the plant, the investigators randomly select a group of 50 cars for study. Then, they put aluminum-rimmed wheels on half the cars and steel-rimmed wheels on the other half. Each car consumes 8 tanks of gas, and the rate of fuel consumption in miles per gallon is documented.

    a. (4 points). What is the sample used for the study? What is the population corresponding to the sample?

       Sample: the 50 randomly selected Model D cars produced at the Fast-Cars Co. Detroit plant.

       Population: all Model D Fast-Cars Co. cars produced at the Detroit plant.

    b. (4 points) Is this study an experimental or observational study? Why?

       It's an experimental study because the investigators themselves impose the wheel type on the cars.

    c. (4 points) Name the response variable and treatment variable. What are the levels of the treatment variable?

       Response variable: the rate of fuel consumption of the 8 tanks of gas in miles per gallon.

       Treatment variable: wheel type, with two levels: aluminum-rimmed and steel-rimmed

    d. (4 points) What are the treatment (experimental) groups? How many members of the sample are in each group?

       Aluminum-rimmed wheel group: 25 cars

       Steel-rimmed wheel group: 25 cars

e. (4 points) Suppose we find that the aluminum-wheeled cars have a significantly higher miles-per-gallon rating than the steel-wheeled cars in our sample. Assuming that the sample is representative of the population, can we say that aluminum wheels are better for fuel consumption than steel wheels? Why or why not?

Since the study is an experimental study, we are allowed to conclude that aluminum wheels are better for fuel consumption than steel wheels.

2. (20 points). You work for a company that manufactures computer processors, and you want to design an experiment to compare the speeds of the processors in a particular batch at different two different temperatures (high and low).

a. (10 points) From your batch of 50 processors, you want to take a simple random sample of 12 for your experiment. Use the random number table below to generate the sample. **Be clear about how you are indexing the processors. Start at the upper left corner of the table.**

I give each processor a unique index from 01 to 50 (NOTE: "01", NOT "1"). I move along the random number table above and I pick the following processors:

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

27, 25, 23,18,35,46,34,17,42,47,36,45

b. (10 points). Now, you want to randomly allocate 6 of the 12 processors to the high temperature treatment group and the other 6 to the low-temperature treatment group. Using the random number table below, carry out this randomization. **Be clear about how you are indexing the processors. Start at the upper left corner of the table.**

I take my sample of 12 processors from before and re-index them like this:

Index from Part (a)	27	25	23	18	35	46	34	17	42	47	36	45
New Index	01	02	03	04	05	06	07	08	09	10	11	12

I move along the random number table above and I pick the following processors for the high-temperature group (according to the new index):

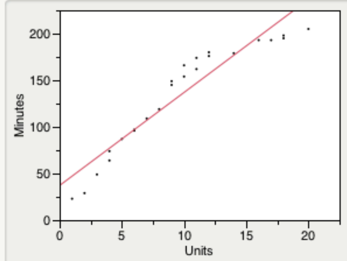
97437	52922	80739	59178	50628	61017	51652	40915	94696	67843
58009	20681	98823	50979	01237	70152	13711	73916	87902	84759
77211	70110	93803	60135	22881	13423	30999	07104	27400	25414
54256	84591	65302	99257	92970	28924	36632	54044	91798	78018
36493	69330	94069	39544	14050	03476	25804	49350	92525	87941

High-temperature group: 10, 09, 06, 01, 11, 04

I put the rest of the processors in the low-temperature group: 02, 03, 05, 07, 08, 12

3. (20 points). A company that markets and repairs small computers studied the relationship between the length of a service call and the number of electronic units that needed repair. For each of 20 calls, the company recorded the number of units needing repair and the length of the call. Below is a JMP report corresponding the fit of a simple linear regression model of call time (in minutes) on the number of components to be repaired (units).

Bivariate Fit of Minutes By Units



Linear Fit

Linear Fit

Minutes = 37.212729 + 9.9695043*Units

Summary of Fit

RSquare	0.896592
RSquare Adj	0.891891
Root Mean Square Error	18.75343
Mean of Response	134
Observations (or Sum Wgts)	24

Lack Of Fit

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	67084.794	67084.8	190.7492
Error	22	7737.206	351.7	Prob > F
C. Total	23	74822.000		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	37.212729	7.985252	4.66	0.0001*
Units	9.9695043	0.721842	13.81	<.0001*

a. (4 points) What call time does the fitted model predict at 0 units? Does your answer make sense in the context of the problem? Why or why not?

The fitted equation is Minutes = 37.212729 + 9.9695043 * Units + e. So at Units = 0, the predicted value for Minutes is 37.212729 + 9.9695043 * 0 = 37.212729 minutes.

The answer doesn't make sense in context: a service call about a computer with no defective units should take no time at all. In fact, at 0 units, the computer is working perfectly, so it doesn't even make sense to make a service call at all.

Alternatively, the answer doesn't make sense because it's an instance of extrapolation: we don't have data for calls at 0 units, so we don't know what happens there.

b. (4 points) What is the name of the term in the fitted regression equation that the quantity in (a)?

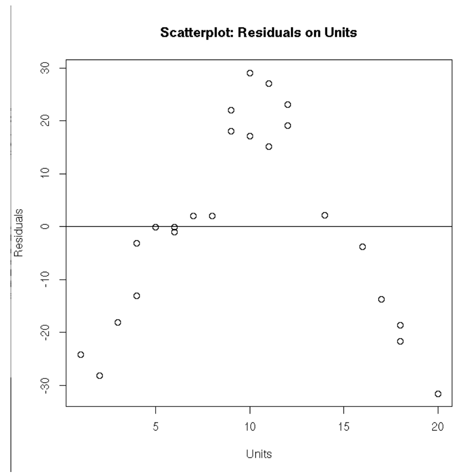
The quantity in (a) is the intercept estimate, b_0.

c. (4 points) State the slope of the estimated regression equation Be sure include the units of the slope. What does this slope mean in the context of the problem? (I'm looking for an interpretation of the form, "On average, we expect…per unit increase…"

The slope is 9.9695043 minutes per unit.

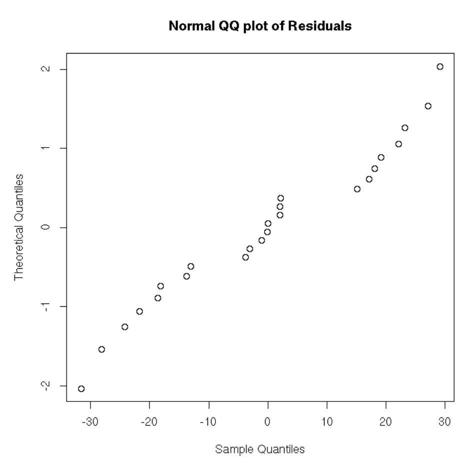
Interpretation: On average, we expect a call to last 9.9695043 more minute for every increase in one hardware unit needing repair.

d. (4 points) Based on the residual plot below, does the model that we used seem valid? Why or why not?



The model is not valid because the residuals show an obvious (non-cloud) pattern when plotted against Units.

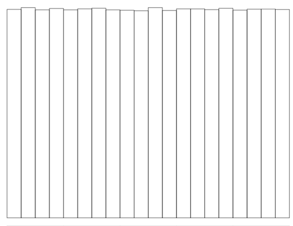
e. Below is a normal quantile-quantile plot of the residuals. What can you say about the distributional shape of the residuals? Why?



The residuals seem roughly normally distributed since their normal QQ plot shows roughly a straight line.

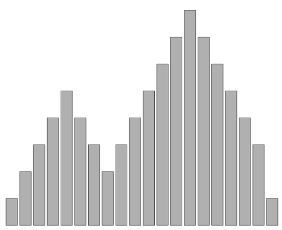
4. (20 points). Describe the following (rough) distributional shapes:

a. (4 points)



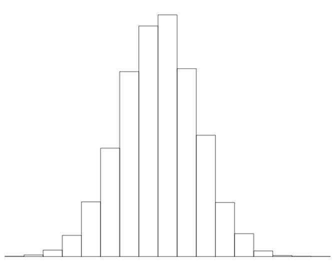
Uniform, symmetric, multi-modal

b. (4 points)



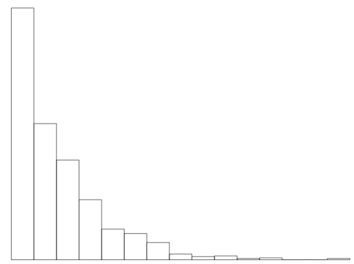
Asymmetric, bimodal

c. (4 points)



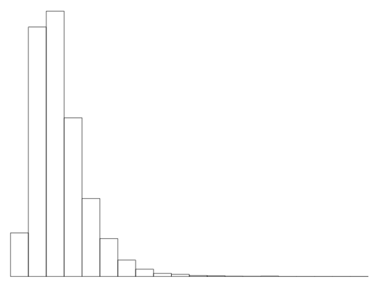
Symmetric, unimodal, bell-shaped

d. (4 points)



Truncated, unimodal, right-skewed, asymmetric

e. (4 points)



Mound-shaped, right-skewed, asymmetric

5. (20 points). The values listed below are the lifetimes (in number of 24mm deep holes drilled in 1045 steel before tool failure) for 12, 8mm drills.

2, 42, 47, 60, 82, 90, 90, 102, 111, 114, 115, 120

a. (10 points) Calculate Q(0.125) and Q(0.25)

First, I calculate:

Sorted values ($Q(p_i)$)	2	42	47	60	82	90	90	102	111	114	115	120
Index i	1	2	3	4	5	6	7	8	9	10	11	12
p_i = (i-.5)/12	0.0417	0.125	0.208	0.291	---	---	---	---	---	---	---	---

So Q(0.125) = $Q(p_2)$ = 42.

For Q(0.25), I use the following:

Let i' = n * (.25) + .5 = 12*.25 + .5 = 3.5

Using the interpolation formula:

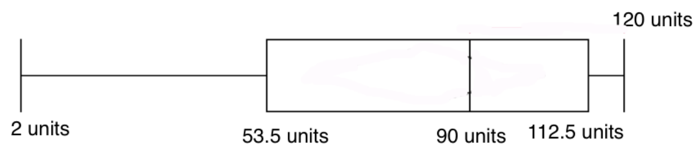
$$\begin{aligned} Q(0.25) &= (\lceil i' \rceil - i')x_{\lfloor i' \rfloor} + (i' - \lfloor i' \rfloor)x_{\lceil i' \rceil} \\ &= (\lceil 3.5 \rceil - 3.5)x_{\lfloor 3.5 \rfloor} + (3.5 - \lfloor 3.5 \rfloor)x_{\lceil 3.5 \rceil} \\ &= (4 - 3.5)x_3 + (3.5 - 3)x_4 \\ &= 0.5 \cdot 47 + 0.5 \cdot 60 \\ &= 53.5 \end{aligned}$$

b. Suppose I tell you that Q(0.5) = 90 and Q(0.75) = 112.5. Draw a boxplot of the data.

IQR = Q(0.75) – Q(0.25) = 112.5 - 53.5 = 59
1.5 * IQR = 59 * 1.5 = 88.5

L = Q(0.25) – 1.5 * IQR = 53.5-88.5 = -35
U = Q(0.75) + 1.5 * IQR = 112.5+88.5 = 201

No data points are below L or above U, so we have no outliers. Hence, the box plot looks like:


120 units
2 units
53.5 units
90 units
112.5 units