STA 110B Spring 1999

Midterm Exam 1
March 9, 1999

Name:	**Section:**

I understand and agree to abide by the Duke honor code,

Signed:

Instructions

This is a closed-book exam, however, one 8.5 by 11 inch "crib sheet" is permitted. You may use a calculator if you find it useful. Show your work in the space provided, but be concise. Correct but unsubstantiated answers will receive no credit.

Point assignments for each of the 3 problems are given in parentheses in the table below. You have 1 hour and 15 minutes total; plan accordingly. You must hand the exam in at 12:10pm, no extra time will be given. Good luck!

		Page 1	**Page 2**	**Page 3**
1.	(50)			
2.	(50)			
Total	(100)			

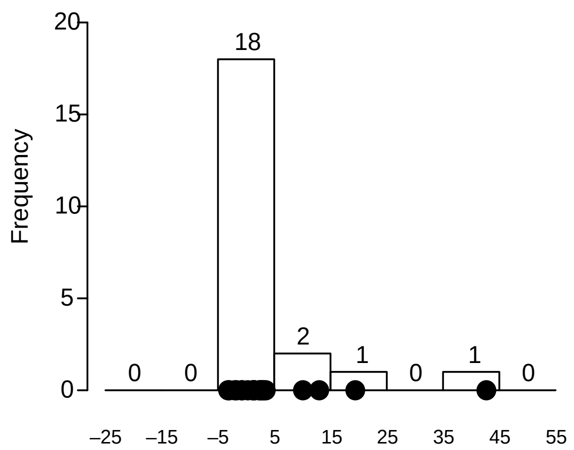
1) The clothing manufacturer Levi-Strauss' quality control department collects weekly data on waste in its manufacturing process. They define waste to be the fraction of cloth remaining when their patterns are drawn by people minus the fraction remaining when they are set by computer, measured as a percentage. Negative values are possible, they indicate that the plant employees beat the computer in controlling waste. The data corresponding to one of the Albuquerque plant's suppliers is given below.

Percentage Waste

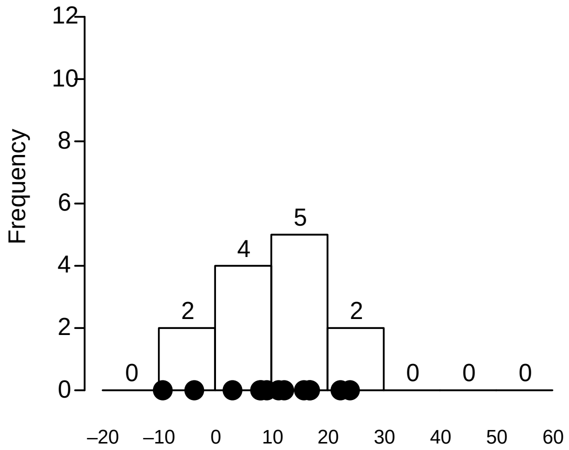
-3.2	-3.0	-2.0	-1.7	-0.8	-0.7	0.3	1.2	1.3	1.4	1.5
2.4	2.4	2.7	2.8	3.0	3.2	3.5	10.1	13.0	19.4	42.7

a) (**5 points**) Consider the following 4 possible histograms of the data. Identify, by circling, all legitimate histograms of the data.

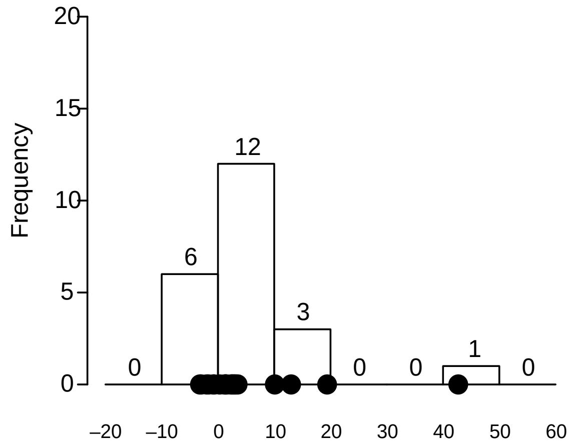
(A)



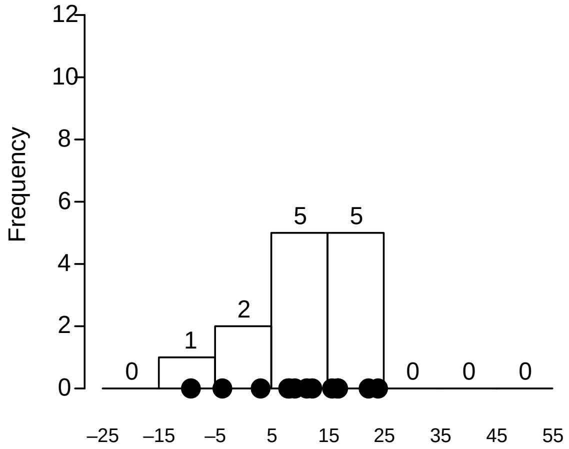
(B)



(C)



(D)



1) Continued. These questions continue our analysis of the Levi-Strauss data, which is repeated here.

Percentage Waste

-3.2	-3.0	-2.0	-1.7	-0.8	-0.7	0.3	1.2	1.3	1.4	1.5
2.4	2.4	2.7	2.8	3.0	3.2	3.5	10.1	13.0	19.4	42.7

b) (**5 points**) Calculate the median of the data.

c) (**5 points**) Calculate the first quartile of the data.

d) (**5 points**) Calculate the third quartile of the data.

e) (**5 points**) What is the inter–quartile range (IQR) of the data.

f) (**5 points**) Identify outliers in the data (judge outliers as points more than 1.5*IQR below the first quartile or more than 1.5*IQR above the third quartile).

1) Continued. These questions continue our analysis of the Levi-Strauss data, which is repeated here.

Percentage Waste

-3.2	-3.0	-2.0	-1.7	-0.8	-0.7	0.3	1.2	1.3	1.4	1.5
2.4	2.4	2.7	2.8	3.0	3.2	3.5	10.1	13.0	19.4	42.7

The sum of the data is 99.5, the some of the squares of the data points minus their mean, $\Sigma(X_i - \bar{X})^2 = 2113.479$.

g) (**5 points**) Calculate the sample mean of the data.

h) (**5 points**) Calculate the sample standard deviation of the data.

i) (**5 points**) Calculate the mean and standard deviation of the data when measured on the decimal scale (1% = 0.01) instead of the percentage scale.

j) (**5 points**) In words, **briefly!** (in 2 or 3 sentences) summarize/describe the data. How strong is evidence that the workers are more wasteful than the computer?

2) According to the Statistical Abstract of the United States 1996, the proportion, π, of the U.S. voting population who voted for the Democratic presidential candidate in 1988 was 0.47. Broken down by gender, this fraction was 44% of men and 50% of women.

Suppose that a pollster polls 100 voters from this population with the aim of estimating π.

a) (**4 points**) Let X denote the number of voters in the sample who voted for the Democratic candidate. What distribution does X have? What are its parameter values?

b) (**4 points**) Let P denote the proportion of voters in the sample who voted for the Democratic candidate. What is the expected value of P?

c) (**4 points**) What is the variance of P?

d) (**4 points**) What is the mean squared error for using P to estimate the proportion of voters in the population voting for the Democratic candidate?

2) Presidential Poll, Continued. The proportion, π, of the U.S. voting population who voted for the Democratic presidential candidate in 1988 was 0.47. Broken down by gender, this fraction was 44% of men and 50% of women.

Suppose that a pollster polls 100 voters from this population with the aim of estimating π.

e) (**9 points**) What is the probability that this poll would correctly predict the winner? Write this probability in terms of the random variable P and justify any approximations you might use in your calculation.

f) (**9 points**) Assume that half of the voting population is male and half is female. A Democratic voter is drawn at random from the voting population, what is the chance that this voter is female?

2) Presidential Poll, Continued. The proportion, π, of the U.S. voting population who voted for the Democratic presidential candidate in 1988 was 0.47. Broken down by gender, this fraction was 44% of men and 50% of women.

Suppose that another pollster polls 1000 *female* voters from this population with the aim of estimating the proportion of *all* voters who voted for the Democratic candidate, π.

g) (**4 points**) Let P_f denote the proportion of voters in this sample who voted for the Democratic candidate candidate. What is the expected value of P_f?

h) (**4 points**) What is the variance of P_f?

i) (**4 points**) What is the mean squared error for using P_f to estimate π, the proportion of voters in the population voting for the Democratic party?

j) (**4 points**) What is the relative efficiency of P (the proportion defined in part b) to P_f for estimating π?