

Student name:

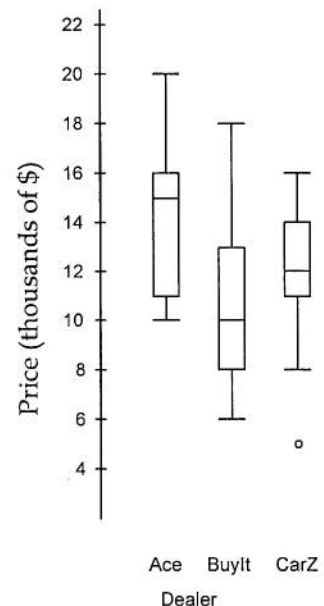
MATH 120: Elementary Statistics  
Section 1

Test 1  
February 21, 2019

**Instructions:**

- This is a regular “closed-book” test, and is to be taken without the use of notes, books, or other reference materials. Collaboration or group work is not permitted.
- Cell-phone usage of any kind is prohibited for the entire duration of the test. This also applies to any restroom breaks taken during the test.
- The time limit for taking this test is 80 minutes from the scheduled start time. Please turn in your test promptly when time is called to avoid late penalties.
- This test adds up to 50 points.

1. [4 pts] The boxplots show prices of used cars, in thousands of dollars, advertised for sale at three different car dealers. Write a few sentences comparing prices at these three dealerships. Be sure to include the context and units in your discussion.



Used car prices seem to be generally lowest at BuyIt, and highest at Ace. At BuyIt, the median price of around \$10,000 is less than the cheapest car at Ace, and below the Q1 price at CarZ. The variability of prices seems lower at CarZ than at the other 2 dealerships. CarZ also offers the lowest overall price, with a low outlier around \$5000. The IQR of prices at CarZ is around \$3000, while it is closer to \$5000 at the other 2 dealers. Ace sells the most expensive car, with a price of around \$20,000.

Grade:

- 1 pt = reasonable comparison of medians, in complete sentences
- 1 pt = reasonable comparison of variability/spread
- 1 pt = reasonable comments on other features
- 1 pt = include context & units

2. [4 pts] In their October 2003 issue, *Consumer Reports* evaluated the price and performance of 23 models of cell phones. Computer output gives these summaries for the prices:

Min	Q1	Median	Q3	Max	Mean	SD
15	30	50	110	200	71.75	52.08

- (a) Were any of the prices outliers? Justify your claim.  
 (b) What statistics would best summarize the center and spread of these data?

(a) I will use the boxplot fence criterion to determine whether there are outliers.  $IQR = Q_3 - Q_1 = 110 - 30 = 80$

Lower fence =  $Q_1 - 1.5 IQR = 30 - 120 = -90$

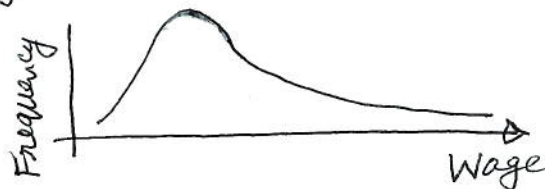
Upper fence =  $Q_3 + 1.5 IQR = 110 + 120 = 230$

Answer: There are no outliers, since all data values lie within the boxplot fences.

(b) The mean is significantly larger than the median, suggesting a skewed distribution. Therefore, the median and IQR would be best for summarizing center and spread.

3. [4 pts] According to the U.S. Bureau of Labor Statistics, in 2009 the mean hourly wage for Chief Executives was \$93 and the median was \$78.28. Is this wage distribution likely to be symmetric, skewed left, or skewed right? Show a plausible sketch (or histogram profile) for the distribution.

Since the mean is larger than the median, this distribution is likely to be skewed right.



(a) = 2.5 pt. (b) = 1.5 pts  
 For (a): 1.5 pt = Boxplot fence calculations, 1 pt = correct answer  
 For (b): 1 pt = answer, 0.5 pt = reason

2 pt →

2 pt →  
 (Sketch)

(a) = (b) = 2 points each  
 For each: 1 pt = calculation step  
 1 pt = answer

4. [4 pts] The table below compares the daily weather forecast in a city with the actual weather for 1 year

		Forecast		Total
		rain	no rain	
Actual	rain	102	31	133
	no rain	46	186	232
	Total	148	217	365

- (a) For what percent of days was the forecast correct?  
 (b) On what percent of days did it actually rain?

(a) The forecast was correct for: 102 days of rain + 186 days of no rain  
 Percent correct =  $\frac{102+186}{365} = 0.789 \Rightarrow \boxed{78.9\% \text{ of days}}$

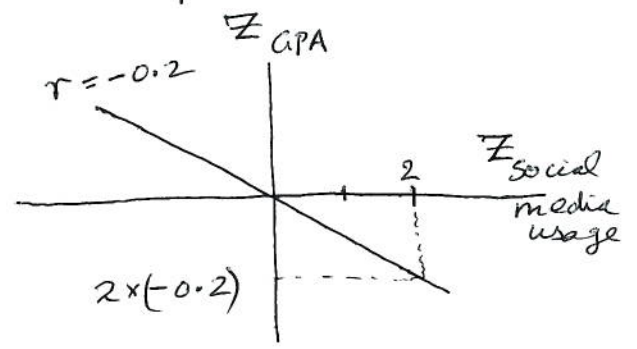
(b) Actual rain = 133 days out of 365  
 Percent days of actual rain =  $\frac{133}{365} \times 100 = \boxed{36.44\%}$

5. [4 pts.] In a psychology experiment, researchers studied the association between social media usage and academic performance among college students. Social media usage was measured in average number of hours per week, and academic performance was measured by GPA. The association was roughly linear, with a correlation of  $r = -0.2$ . The researchers decided to construct a linear regression model to predict GPA from social media usage.

- (a) Identify the explanatory variable and the response variable, including their units.  
 (b) Suppose a student's social media usage is 2 standard deviations higher than the mean usage. What does this model predict about the student's GPA?

(a) Explanatory: social media usage, in average number of hours per week  
 Response: GPA; has no units

(b) For a student whose social media usage is 2 SD above the mean usage, the predicted GPA is 0.4 standard deviations below the mean GPA.



Grade: (a) = (b) = 2 points  
 For (a): 1 pt = Explanatory; 1 pt = response: 50/50 split between answer/units.  
 For (b) 1 pt = steps/reason/logic; 1 pt = correct answer, with clarity of statement being important

6. [10 pts] Now that cigarette smoking has been clearly tied to lung cancer, researchers are exploring possible links to other diseases. An article in the *American Journal of Public Health* gives data on smoking rates and coronary heart disease (CHD) in 21 countries. The mean cigarette consumption in these countries was 2148 cigarettes per adult per year, with a standard deviation of 809 cigarettes per adult per year. The mean CHD rate was 144.9 deaths per 100,000 citizens, with standard deviation of 66.5 deaths per 100,000. The association between cigarette consumption and CHD rates was found to be approximately linear, positive, and with correlation  $r = 0.73$ .

(a) Construct a linear regression model to predict CHD death rates from cigarette consumption rate.

(b) Interpret its slope in this particular application context.

(c) Find  $R^2$  and interpret its meaning.

(a) Let  $x =$  cigarette consumption rate (units = # per adult per year)

$y =$  rate of CHD deaths (units = # per 100,000 citizens)

Linear regression eqn. has the form:  $\hat{y} = b_0 + b_1x$ , where  $b_1 = r \frac{S_y}{S_x}$

In this problem we have:  $\bar{x} = 2148$ ,  $S_x = 809$

$\bar{y} = 144.9$ ,  $S_y = 66.5$

$r = 0.73$

$$b_1 = (0.73) \frac{66.5}{809} \Rightarrow b_1 = 0.06 \frac{\text{deaths per 100,000}}{\text{\# cigs. per adult per year}}$$

The equation then becomes:  $\hat{y} = b_0 + 0.06x$

Plug in  $(\bar{x}, \bar{y})$  to find  $b_0$ :  $144.9 = b_0 + 0.06(2148)$

$\Rightarrow b_0 = 16.02$  CHD deaths per 100,000

Thus, the linear regression model is:

$$\text{CHD death rate} = 16.02 + 0.06(\text{cigarette consumption rate})$$

(b) Interpretation of slope: For each 1 unit increase in cigarette consumption rate per adult per year, the predicted rate of CHD deaths goes up by 0.06 per 100,000.

(c)  $R^2 = (0.73)^2 \times 100 = 53.29\%$

Interpretation: The variability in cigarette consumption rates accounts for about 53.29% of the variability in CHD death rates.

Grade 3: (a) = 5 points. (b) = 2 points. (c) = 3 points.

For (a): 2 pt = correct work on  $b_1$ ; 2 pt = correct work on  $b_0$

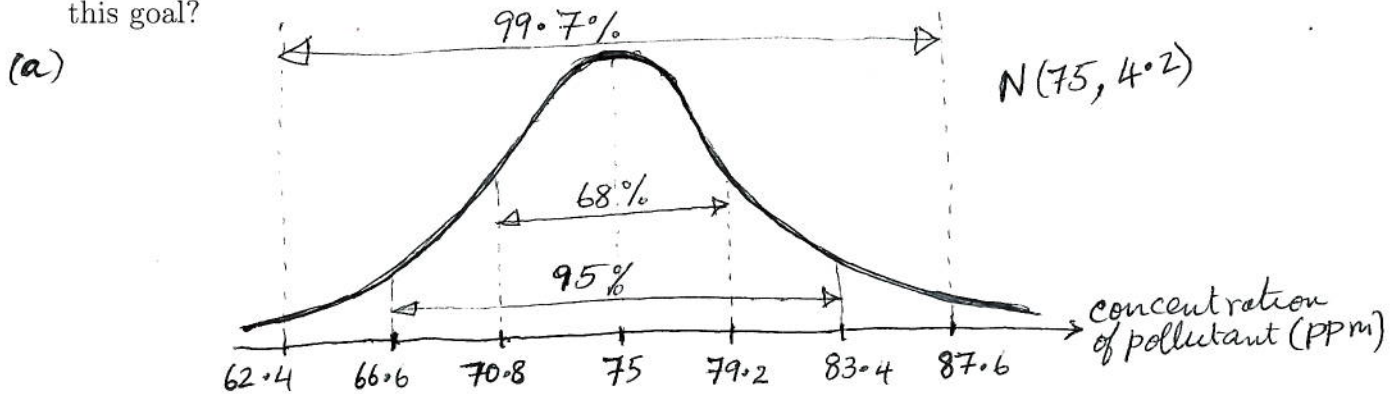
1 pt = correct regression eqn. - 0.5 pt if mixing units

For (c): 1 pt = compute  $R^2$  Page 4 of 6

2 pt = Interpret meaning

- 1.5 pt if regression variables reversed

7. [10 pts] A company's manufacturing process uses 500 gallons of water at a time, which is later treated to remove a chemical pollutant, before the water is released into a nearby lake. Local laws allow a maximum of 80 ppm (parts per million) of the chemical in the treated water. The output from the company's treatment process follows a normal model, with mean pollutant concentration of 75 ppm and standard deviation 4.2 ppm.
- (a) Sketch and label the normal model, showing key numerical values and percentages of the distribution of data.
- (b) In what percent of the cases does the treated water exceed the legal maximum limit of 80 ppm?
- (c) Management wants to ensure that the 80 ppm limit is only exceeded 1% of the time. What standard deviation must the treatment process have in order to achieve this goal?



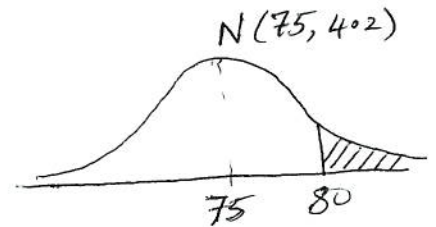
- (b) The shaded area in the sketch shows the % of times the treated water exceeds 80 ppm

$$Z = \frac{80 - 75}{4.2} = 1.19$$

From the Z-table, area to the left = 0.8830

$$\text{Shaded area} = 1 - 0.8830 = 0.117$$

Answer: The legal limit of 80 ppm is exceeded in about 11.7% of the cases



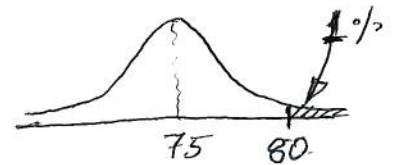
- (c) We want the shaded area to be 1%.

From Z-table, area of 0.99 occurs at  $Z = 2.33$

[We must look up 0.99 because the table shows areas in the left-tail - not in the right tail]

$$\text{Find SD: } 2.33 = \frac{80 - 75}{SD} \Rightarrow SD = \frac{5}{2.33} = 2.146$$

Answer: The SD must be 2.146 PPM



Grade: (a) = 3.5, (b) = 3 pts, (c) = 3.5 pts

For (a): 1 pt = correct num. values on axis; 1.5 pt = show 68-95-99.7 values, 1 pt = clarity of sketch

For (b) 1 pt. each for: correct Z-score, correct table lookup, subtract from 1.

For (c): Know/show what we want to find, 1.5 pt = get correct Z-score; 1.5 pt = compute correct SD

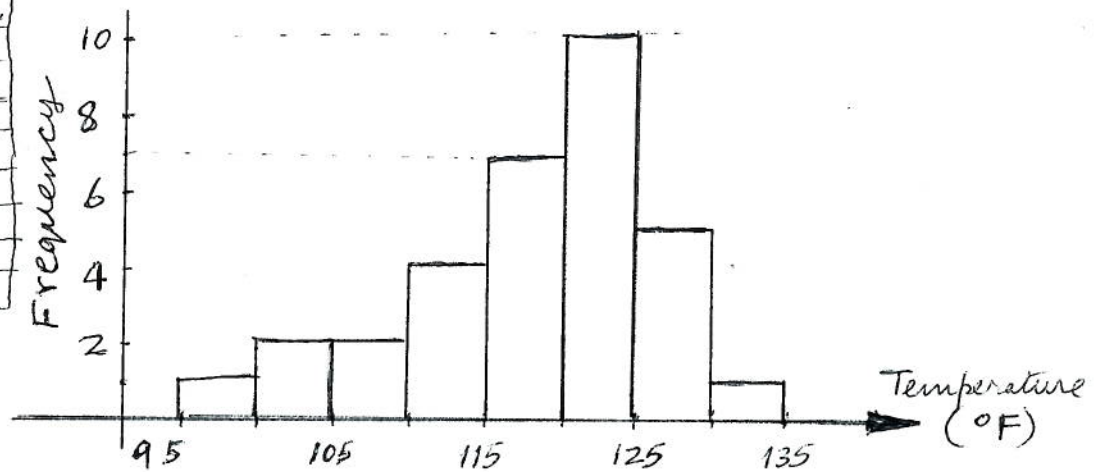
8. [10 pts] The highest temperature ever recorded (in °F) in 32 different states in the U.S. is shown below, in ascending order:

98 102 104 106 109 110 112 114  
 114 115 115 116 116 118 118 118  
 120 120 120 120 121 121 122 122  
 123 124 125 126 126 128 129 134

- (a) Make a suitably scaled, neat, histogram for these data. Show detailed steps.  
 (b) Compute suitable summary statistics for the center and spread of the distribution. Feel free to use your calculator short-cut keys for this part.  
 (c) If we were to convert all temperatures to °C, what would the new values of your center and spread statistics be? [FYI:  $^{\circ}\text{C} = \frac{5}{9}(^{\circ}\text{F} - 32)$ ]

(a) Make a frequency table. I'll use a scale of 5°F

Interval	Freq
95-100-	1
100-105-	2
105-110-	2
110-115-	4
115-120-	7
120-125-	10
125-130-	5
130-135-	1



(b) Since the distribution is skewed, I'll use the median and IQR to summarize the center and spread. Given the 32 numbers, arranged in groups of 8, it is relatively easy to find the median & quartiles by inspection:

$$Q_1 = 114, \text{ Med} = \frac{118 + 120}{2} = 119, Q_3 = \frac{(122 + 123)}{2} = 122.5$$

$$\text{IQR} = 122.5 - 114 = 8.5$$

Therefore, the center (median) = 119°F, the spread (IQR) = 8.5°F

(c) New median =  $\frac{5}{9}(119 - 32) = 48.3^{\circ}\text{C}$

$$\text{New IQR} = \frac{5}{9} \times 8.5 = 4.72^{\circ}\text{C}$$

Reasoning: Median is affected by both subtraction + multiplication  
 IQR is not affected by subtraction, but only by mult.

Grade: (a) = 4 points, (b) = 3 points, (c) = 3 points

For (a): 0.5 pt = reasonable scale (<10); 1 pt = freq. table; 1.5 pt = histogram, 1 pt = complete & correct axes labels.

For (b): 1 pt = choose median/IQR (End of test) for summary stats  
 2 pt = compute them correctly, with units.

For (c): 1.5 pt each for correct center & spread  
 For each, 1 pt = answer, 0.5 pt = steps/reasons.