Math 153 - Test 1

February 21, 2013

Name ___key___

Score _____

Show all work to receive full credit. Supply explanations where necessary. You may use your calculator for all statistical computations.

1. (9 points) Refer to the snowfall data on the attached sheet.

    (a) Find the percentile for the value 13.1.

    $$\frac{26}{44} \times 100 = 59.\overline{09} \approx 59$$

    $\Rightarrow$ 59TH percentile

    (b) Find the value of the 85th percentile.

    $$85 = \frac{L}{44} \times 100$$

    $$\Rightarrow L = \frac{85 \cdot 44}{100} = 37.4 \rightarrow 38$$

    $P_{86} = 38^{TH}$ value

    $= 18.9$

    (c) Find the value of the 25th percentile.

    $$25 = \frac{L}{44} \times 100$$

    $$\Rightarrow L = \frac{25 \cdot 44}{100} = 11$$

    $$P_{25} = \frac{11^{TH} \text{ value} + 12^{TH} \text{ value}}{2}$$

    $$= \frac{4.3 + 5.0}{2} = 4.65$$

2. (4 points) The prices for a gallon of gasoline have mean \$3.74 and standard deviation \$0.108. Suppose you recently paid \$3.95 per gallon. Compute the z-score for your price and determine whether the price you paid is unusual.

    $$z = \frac{3.95 - 3.74}{0.108} = 1.9\overline{4}$$

    Because $z < 2$, the price is NOT unusual.

3. (2 points) When should you use coefficients of variation to compare the spread in different sets of data?

    When the means of the two data sets are significantly different.

1

4. (12 points) Refer to the snowfall data on the attached sheet.

(a) Compute the quartiles.

$$Q_1 = \frac{11^{TH} + 12^{TH}}{2} = \underline{4.65} \quad (\text{AS IN PROBLEM } \#1c)$$

$$Q_2 = \frac{22^{ND} + 23^{RD}}{2} = \underline{10} \qquad Q_3 = \frac{33^{RD} + 34^{TH}}{2} = \underline{16.15}$$

(b) Find the IQR (interquartile range).

$$16.15 - 4.65 = \underline{11.5}$$

(c) Find the mean and sample standard deviation. You may use your calculator.

FROM CALCULATOR... $\overline{X} = 11.345$

$S = 8.316$

(d) Determine whether the snowfall data include any unusual values.

$$\overline{X} - 2s \approx -5.3$$

$$\overline{X} + 2s \approx 28.0$$

29.6 & 34.3

ARE UNUSUAL.

(e) Does the range rule of thumb, $s \approx$ range/4, provide a good approximation for the standard deviation? Justify your answer.

$$\frac{34.3 - 0.4}{4} = 8.475$$

THIS IS A VERY GOOD APPROXIMATION.

(f) Would it be wise to use a stem-and-leaf plot to display the snowfall data? Explain.

No, THE STEMS WOULD HAVE TO GO FROM 0 TO 34, AND MANY STEMS WOULD HAVE NO LEAVES.

5. (4 points) According to Chebyshev's Theorem, at least what percent of the data values from any set will lie within 1.75 standard deviations of the mean?

$$1 - \frac{1}{(1.75)^2} \approx 0.6735$$

$$\Rightarrow \underline{67.35\%}$$

6. (12 points) The frequency distribution shown below gives the salaries (in thousands of dollars) of the employees at a small company.

Salary (thousands of $)	Frequency
12.7–29.0	13
29.1–45.4	8
45.5–61.8	5
61.9–78.2	1
78.3–94.6	2
94.7–111.0	1

(a) What are the class boundaries associated with the last class listed above?

94.65 AND 111.05

(b) What is the class width?

29.1 − 12.7 = 16.4

(c) What are the class midpoints?

20.85, 37.25, 53.65, 70.05, 86.45, 102.85

(d) Use class midpoints to estimate the mean salary.

$$\frac{20.85(13) + 37.25(8) + 53.65(5) + 70.05(1) + 86.45(2) + 102.85(1)}{(13+8+5+1+2+1)}$$

$$= \frac{1183.1}{30} \approx 39.44$$

$ 39,440

(e) Use class midpoints to estimate the median.

THERE ARE 30 VALUES

MEDIAN = $\frac{15^{TH} + 16^{TH}}{2}$ = 37.25 $ 37,250

(f) Do the salaries appear to be normally distributed? Explain.

No, THE SALARIES ARE PEAKED ON

THE LOWER END.

THE DISTRIBUTION IS SKEWED

RIGHT.

3

7. (12 points) A group of 20 dieticians believe that meditation is an effective weight-loss strategy. In order to justify their beliefs, they plan to study a sample of 200 patients. Some patients will be asked to meditate for 30 minutes each day.

(a) Is their study an experiment or an observational study? Briefly explain.

A TREATMENT (MEDITATION) IS BEING APPLIED.

THIS IS AN EXPERIMENT.

(b) Are the weights of patients discrete or continuous?

CONTINUOUS

(c) Identify the level of measurement (nominal, ordinal, interval, ratio) for the weights of patients.

RATIO

(d) The sample of 200 patients is obtained by each dietician randomly selecting 10 volunteers from his or her own practice. Is this sample a simple random sample? Briefly explain.

NOT SIMPLE RANDOM. THE SAMPLE OF 200 COULD NOT COME, FOR EXAMPLE, FROM A SINGLE DIETICIAN.

(e) One of the dieticians, Dr. Wesley Roberts, reported that his ten patients lost 4 pounds per month, on average. Is this number a statistic or a parameter?

IT ONLY REFERS TO HIS SAMPLE. IT IS A STATISTIC.

(f) At the end of their study, the dieticians are convinced that their results are statistically significant. What is the difference between statistical significance and practical significance?

RESULTS ARE STATISTICALLY SIGNIFICANT IF THEY ARE VALID AS MEASURED BY STATISTICAL TESTS.
PRACTICAL SIGNIFICANCE REFERS TO THE EVERY-DAY USE OR VALUE OF THE RESULTS.

4

RESULTS CAN BE STATISTICAL SIGNIFICANT BUT PRACTICALLY WORTHLESS.

8. (10 points) The following histogram shows the distribution of weights of dogs in a local shelter.
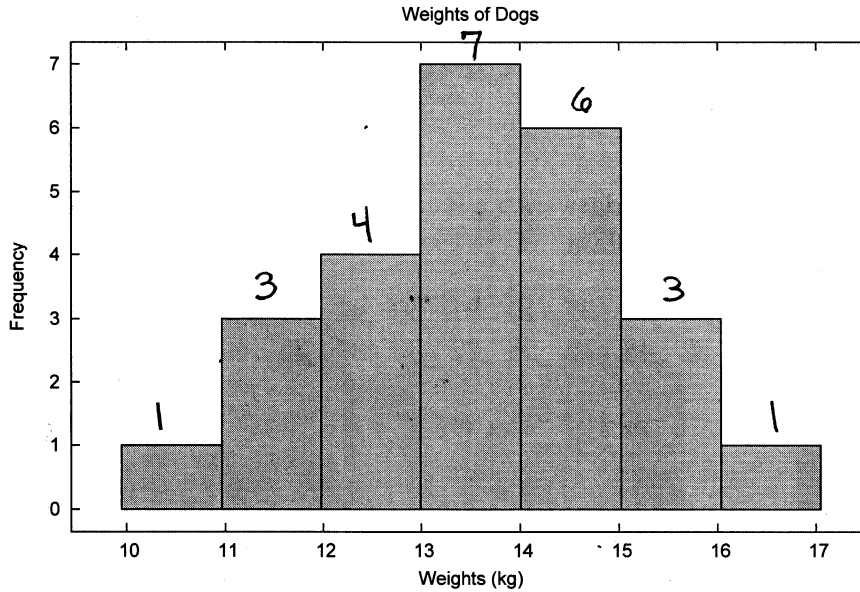
Weights of Dogs

(a) How many dogs are in the sample?

$$1 + 3 + 4 + 7 + 6 + 3 + 1 = 25$$

(b) If the histogram was changed to a relative frequency histogram, what would be the height of the fourth bar?

$$\frac{7}{25} = 0.28 = 28\%$$

(c) Are numbers of dogs discrete or continuous?

DISCRETE

(d) Estimate the median dog weight.

25 WEIGHTS $\Rightarrow$ MEDIAN IS 13TH WEIGHT $\approx$ 13.5 kg

(e) Are the dogs' weights approximately normally distributed? Explain.

YES, THE HISTOGRAM IS ROUGHLY TRACED BY A SYMMETRIC, SMOOTH, BELL-SHAPED CURVE.

9. (3 points) Describe a situation where a pie chart (circle graph) would be the best kind of graphical display. USE A PIE CHART TO SHOW PORTIONS OF A WHOLE. FOR EXAMPLE, SAY WE MAKE A GRAPH SHOWING PORTIONS OF STUDENTS GETTING A's, B's, C's, ... OUT OF THE ENTIRE CLASS.

10. (5 points) Organize the following data into a stem-and-leaf plot. Are the data approximately normally distributed? Explain.

$$47 \; 32 \; 48 \; 51 \; 10 \; 27 \; 50 \; 21 \; 24 \; 12$$
$$21 \; 32 \; 48 \; 12 \; 28 \; 32 \; 36 \; 37 \; 38 \; 49$$

THE DATA ARE

ROUGHLY

NORMAL.

THE LEAVES ARE TRACED

ROUGHLY BY A

BELL-SHAPED CURVE.

Stems	Leaves
1	0 2 2
2	1 1 4 7 8
3	2 2 2 6 7 8
4	7 8 8 9
5	0 1

3 | 2 means 32

11. (8 points) Julio scored 180 on a math test with mean 157.9 and standard deviation 27.3. Mary scored 43 on a physics test with mean 38.1 and standard deviation 6.2.

(a) Compute the corresponding z scores. Who scored better and why?

MATH JULIO $z = \dfrac{180 - 157.9}{27.3} \approx 0.81$

PHYSICS MARY $z = \dfrac{43 - 38.1}{6.2} \approx 0.78$

JULIO SCORED BETTER -- FARTHER ABOVE MEAN.

(b) Compute the coefficients of variation for the tests. Which tests (math or physics) had greater variation?

MATH CV $= \dfrac{27.3}{157.9} \approx 17.3\%$

PHYSICS CV $= \dfrac{6.2}{38.1} \approx 16.3\%$

MATH TESTS HAVE GREATER VARIATION.

(c) What would be an unusually high score on the math test?

$$\overline{X} + 2s = 157.9 + 2(27.3) = 212.5$$

ANYTHING > 212.5

6

12. (6 points) A sample of PSC students is obtained as described. Identify the type of sampling (random, systematic, convenience, stratified, cluster).

(a) Students are selected as they walk in the main door.

CONVENIENCE

(b) Students are separated into groups according to age, then 20 students are selected from each age group.

STRATIFIED

(c) A complete list of students is compiled and every 150th name is selected.

SYSTEMATIC

(d) Students are separated into groups according to last initial. Ten letters are chosen at random and all students with that initial are selected.

CLUSTER

(e) Student ID numbers are selected at random by using a computer.

RANDOM

(f) Students are grouped according to which high school they attended. Two students are selected from each high school.

STRATIFIED

13. (4 points) Give an example of a random sample that is not a simple random sample.

100 PEOPLE ARE SELECTED BY CHOOSING

2 AT RANDOM FROM EACH STATE.

NOT EVERY SAMPLE OF 100 IS EQUALLY LIKELY

BECAUSE, FOR EXAMPLE, IT IS IMPOSSIBLE

TO GET 100 NEW YORKERS.

Intentionally blank

14. (9 points) Consider the snowfall data on the attached sheet.

(a) Construct a frequency distribution for the data.

For Part (b)

Inches of Snowfall	Number of Years	Rel Freq
0.0 - 4.9	11	11/44 = 25%
5.0 - 9.9	10	10/44 = 22.7%
10.0 - 14.9	9	9/44 = 21.0%
15.0 - 19.9	8	8/44 = 18.2%
20.0 - 24.9	2	2/44 = 4.5%
25.0 - 29.9	3	3/44 = 6.8%
30.0 - 34.9	1	1/44 = 2.3%

(b) On graph paper, construct the corresponding relative frequency histogram.

Using class midpoints...

See attached sheet.

(c) On graph paper, construct the boxplot for the data. Do not consider outliers.

5-number summary: Min = 0.4, Q_1 = 4.65,

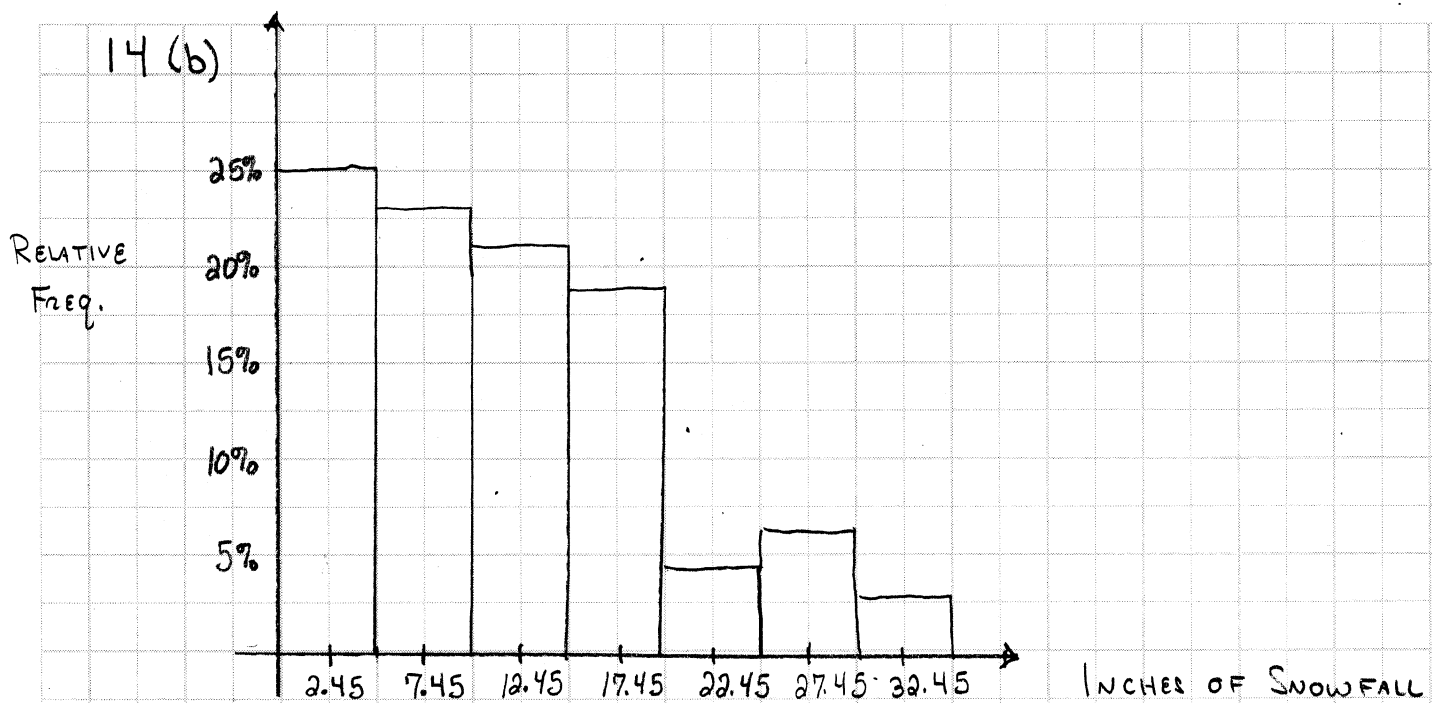Median = 10, Q_3 = 16.15,

Max = 34.3

9

January Snowfall at O'Hare Airport (Inches)

1960-2005 (Excluding 1997 and 1998)

0.4	0.5	1.5	1.6	2.0	3.0	3.2	3.5	3.5	3.7
4.3	5.0	5.4	5.6	5.9	6.2	6.9	7.2	7.4	7.6
9.5	10.0	10.0	10.4	11.1	11.7	13.1	13.6	14.2	14.6
15.2	15.5	15.5	16.8	17.2	17.3	18.6	18.9	21.9	22.9
25.1	27.8	29.6	34.3						

14 (b)

Relative Freq.

A histogram with the vertical axis labeled "Relative Freq." showing percentages: 25%, 20%, 15%, 10%, 5%. The horizontal axis is labeled "Inches of Snowfall" with markings: 2.45, 7.45, 12.45, 17.45, 22.45, 27.45, 32.45.

Bars (left to right):
- 2.45–7.45: ~25%
- 7.45–12.45: ~23%
- 12.45–17.45: ~21%
- 17.45–22.45: ~19%
- 22.45–27.45: ~5%
- 27.45–32.45: ~6%
- 32.45–: ~3%

14 (c)

A box-and-whisker plot with values labeled: 0.4, 4.65, 10, 16.15, 34.3

Number line below: 0, 5, 10, 15, 20, 25, 30, 35