STATISTICS: TEST 2A (SOLUTIONS)

Caveat Lector: Write in pencil; I will deduct points if you use pen. Write all answers on your own paper. Turn off your cell phone; look only at your own test; if you have questions, direct them to the professor.

I have given you some formulas on the back; other formulas you must know, as discussed in class. Please use the formulas for sample variance and sample standard deviation, not population variance and population standard deviation. Note that I advertise a sale on formulas at the bottom of the page.

1. In Sunday's *Rocky Mount Telegram*, Regina Griffin Realty ("Agent for the People") advertised the following prices on real estate for sale, or recently sold:

\$19,000	\$74,900	\$49,900	\$180,900
\$26,900	\$32,900	\$65,000	\$45,000

Calculate (a) the median, (b) the mean, (c) the mode, (d) the midrange, (e) the variance, (f) the standard deviation, (g) the five number summary. Then (h) make a boxplot of the data, and (i) comment on the shape of the distribution.

Solutions:

(a) The median is \$47,450 (the number in the middle).

(b) The mean is

$$\overline{x} = \frac{19,000 + 74,900 + \cdots + 45,000}{8} \approx 61,812.5.$$

(c) There is no mode; all the numbers appear equally often.

(d) The midrange is

$$\text{midrange} = \frac{\max + \min}{2} = \frac{180,900 + 19,000}{2} = 99,950.$$

(e) The variance is

$$\begin{aligned} s^2 &= \frac{\sum (\overline{x} - x)^2}{n-1} \\ &= \frac{(61,812.5 - 19000)^2 + (61,812.5 - 74900)^2 + \cdots + (561,812.5 - 74900)^2}{7} \\ &\approx 2,667,938,392.9 \end{aligned}$$

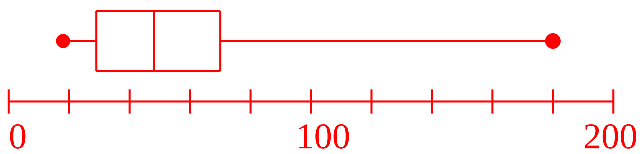
(f) The standard deviation is

$$s = \sqrt{s^2} \approx 51,652.1$$

(g) The five-number summary is

min: \$19,000 Q1: \$29,900 med: \$47,450 Q3: \$69,950 max: \$180,900

(h) The boxplot is



(i) The data has positive skew.

2. Using the data from problem 1, calculate the percentile rank *and* the z-scores for each of the following prices:

(a) \$19,000

(b) \$65,000

Solution to (a):

Percentile rank is (rounding down always!)

$$\frac{\#\text{items below } x + 0.5}{\text{total } \#\text{items}} = \frac{0.5}{8} = 6\text{th percentile}$$

z-score is

$$z = \frac{x - \overline{x}}{s} = \frac{19,000 - 61,812.5}{51,652.089918} \approx -0.83$$

Solution to (b):

Percentile rank is $6.5/8 = 81$st percentile.

z-score is

$$z = \frac{65,000 - 61,812.5}{51,652.089918} \approx 0.06$$

3. Why might one expect the answer to 2(a) to be the 0th percentile, *and* why isn't that the answer?

Solution: Because 19,000 is the lowest value, it should lie in the 0th percentile. That isn't the answer because we have much fewer than 100 numbers, so the calculation is only an approximation.

4. Use Chebyshev's Theorem to answer the following.

(a) Determine the proportion of data values that lie within two standard deviations of the mean.

(b) What range of home prices falls within two standard deviations of the mean?

Solution to (a):

The proportion of data values that lie within two standard deviations of the mean is

$$1 - \frac{1}{k^2} = 1 - \frac{1}{4} = \frac{3}{4} = 75\%$$

Solution to (b):

The range of home prices that fall within two standard deviations of the mean is (strictly translating the problem into mathematics!)

$$\overline{x} \pm 2 \cdot s \approx \$165,116.7, -\$41,491.7$$

5. The G-8 is an annual gathering of some of the world's most economically influential nations: Canada, France, Germany, Italy, Japan, Russia, the United Kingdom, and the United States. The average poverty rate in these countries is 12.8%, with a standard deviation of 3.6. The average unemployment rate in these countries is 7.25%, with a standard deviation of 2.59. Which is more variable among the G-8 countries: the poverty rate, or the unemployment rate?

Solution:

Because we are comparing two different things, we must compute the Coefficient of Variation for each. For the poverty rate,

$$\text{CV} = \frac{3.6}{12.8} \approx 0.28.$$

For the unemployment rate,

$$\text{CV} = \frac{2.59}{7.25} \approx 0.36.$$

6. Explain why *the range* is not a good measure of variation of data.

Solution: The range is not a good measure of data because it only measures how far apart the largest and smallest values are. It does not measure how the data clumps near or spreads away from the mean.

Formulas and facts that you may or may not need.

$$\text{percentile} = \frac{(\#\text{ items below } x) + 0.5}{n} \qquad \text{Count} = \frac{n \cdot \text{percentile}}{100}$$

$$z = \frac{x - \overline{x}}{s} \qquad \text{CVar} = \frac{s}{\overline{x}}$$

(If you need a formula not listed here, you may buy it for a 5% penalty.)

Chebyshev's Theorem: The proportion of data values that lie within k standard deviations of the mean is

$$1 - \frac{1}{k^2}.$$