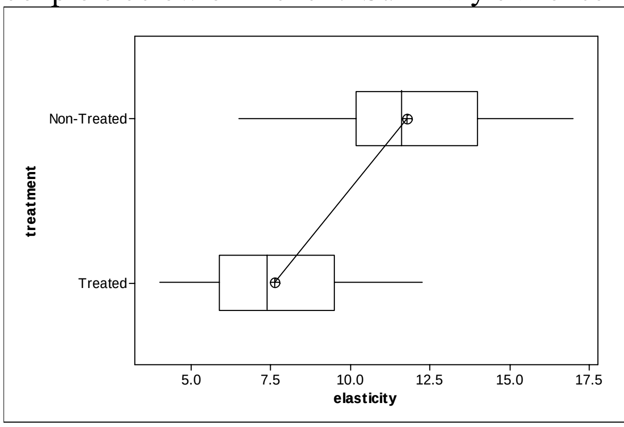
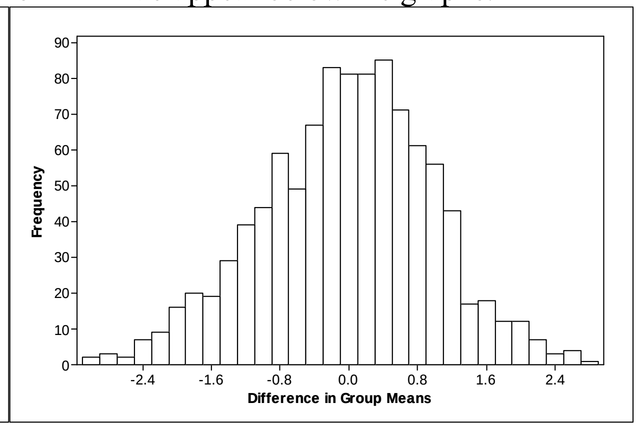
Stat 221 Exam 2C October 31, 2008

1. (28 pts) Student researchers investigated whether balsa wood is less elastic after it has been immersed in water. They took 44 pieces of balsa wood and randomly assigned half to be immersed in water and the other half not to be. They measured the elasticity by seeing how far (in inches) the piece of wood would project a dime into the air. Their results are displayed in the boxplots below on the left. Summary statistics from Minitab appear below the graphs:





Variable	treatment	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
elasticity	Non-Treated	22	11.784	2.674	6.500	10.188	11.625	14.000	17.000
	Treated	22	7.625	2.332	4.000	5.875	7.375	9.500	12.250

To investigate whether the difference between the two groups is statistically significant, the students simulated a randomization test. They randomly assigned the same 44 measurements between the two groups, calculated the difference in group means, and repeated that 1000 times. They obtained the histogram above on the right.

a) (3 pts) Without doing any calculations, make an educated guess for the value of the standard deviation in this histogram (on the right). (Do not bother to explain.)

The standard deviation is close to 1.0, certainly between 0.8 and 1.2. (This can be seen from where the curvature changes on the histogram.)

b) (3 pts) Without doing any calculations, make an educated guess for the value of the mean in this histogram (on the right). (Do not bother to explain.)

The mean is very close to 0.

c) (6 pts) Use the histogram to produce an approximate p-value. Explain (briefly) how you decide on this value.

The difference in group means for the actual experiment (as seen in the output below the graphs) is 11.784 – 7.625 = 4.159. Looking at the histogram, we see that a difference as extreme as this never happened in the 1000 repetitions of random assignment. The approximate p-value is therefore 0/1000 = 0.

d) (6 pts) Complete this sentence: This p-value estimates the probability that …

… the difference in group means would have been 4.159 or more by chance (random assignment) alone, if in fact there was really no effect of the immersion.

e) (10 pts) Summarize your conclusion from this study, addressing whether the data provide strong evidence that immersion in water causes balsa wood to become less elastic. Also explain the reasoning process behind your conclusion. Be sure to address the issue of causation as well as the issue of statistical significance.

The data from the study provide very strong evidence that immersion in water does indeed cause less elasticity. The result is highly statistically significant, because such an extreme result never happened by chance alone in 1000 repetitions. A cause-and-effect conclusion is warranted because this was a randomized experiment.

2. (18 pts) Suppose that the proportion of households in San Luis Obispo County that give candy to trick-or-treaters on Halloween is .75.

a) (4 pts) Is this number a parameter or a statistic? (Do not bother to explain.) What symbol is used to represent it?

This is a parameter, because it refers to the entire population. Its symbol is π.

b) (10 pts) If you visit a random sample of 200 households in SLO County, what is the (approximate) probability that 65% or less will give candy? Show the details of your calculations.

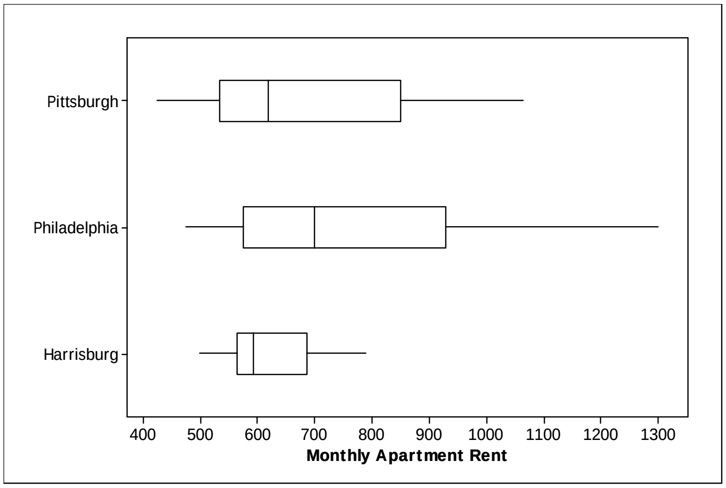
The Central Limit Theorem says that the sample proportion (who give candy) follows a normal distribution with mean .75 and standard deviation $\sqrt{\frac{.75\times(1-.75)}{200}} \approx .0306$. The z-score is therefore (.65 - .75) / .0306 ≈ -3.27. The normal table reveals the probability to the left of this to be about .0005.

c) (4 pts) How would this probability change if the sample size were 80 instead of 200? (Circle your answer. Do not bother to explain.)

Larger   No change   Smaller

With a smaller sample size, it would be less likely for the sample proportion (who give candy) to be close to the population proportion of .75, and therefore more likely for the sample proportion to be less than .65.

3. (24 pts, 4 pts each) The following graph displays the monthly rent (in dollars) for some studio and one-bedroom apartments in 3 Pennsylvania cities, randomly selected from the website `rent.com` in July 2007:



a) Does this study involve random *sampling* or random *assignment*? (Circle your answer.)

random sampling   random assignment

b) Which city has the *least variability* in apartment rents? (Circle your answer.)

Pittsburgh   Philadelphia   Harrisburg

c) Which has the *most variability* in apartment rents? (Circle your answer.)

Pittsburgh   Philadelphia   Harrisburg

d) Which city tends to have the *least expensive* apartment rents? (Circle your answer.)

Pittsburgh   Philadelphia   Harrisburg

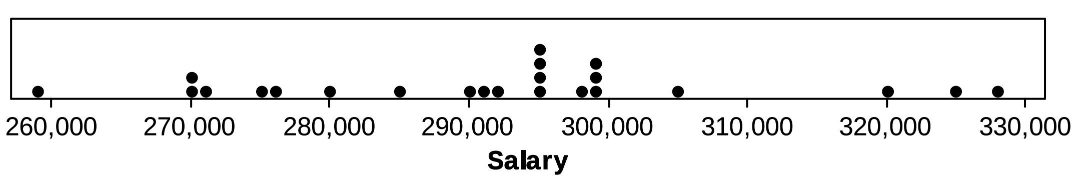
e) Which city tends to have the *most expensive* apartment rents? (Circle your answer.)

Pittsburgh   Philadelphia   Harrisburg

f) Based on these graphs, what do you suspect is the shape for all three distributions? (Circle your answer.)

symmetric   skewed to the right   skewed to the left?

4. (13 pts) The following dotplot displays the salaries of the 23 Presidents of California State University campuses for the 2007-08 academic year:



Consider the following Minitab output:

Variable	N	Minimum	Q1	Median	Q3	Maximum	Mean	StdDev
salary	23	258680	276055	295000	299000	328209	291822	17669

a) (6 pts) The maximum salary here, \$328,209, belongs to Cal Poly's President Baker. Calculate the *z*-score for President Baker's salary, and decide whether his salary is more than two standard deviations above the mean.

The *z*-score for President Baker is: $(328{,}209 - 291{,}822) / 17{,}669 \approx 2.06$. His salary is (slightly) more than two standard deviations above the mean.

b) (7 pts) Determine whether or not his salary qualifies as an outlier, according to the 1.5×IQR rule.

The interquartile range is $Q3 - Q1 = 299{,}000 - 276{,}055 = 22{,}945$. Taking 1.5×IQR gives 34,417.5. Adding this to Q3 gives 333,417.5. President Baker's salary is not greater than this, so it is not an outlier.

5. (17 pts)

a) (3 pts) Which of the following sets of six quiz scores has the *largest* standard deviation? (Circle your answer; do not bother to explain or perform any calculations.)

(A) 0, 0, 0, 10, 10, 10 (B) 5, 5, 6, 6, 7, 7 (C) 0, 3, 5, 6, 8, 10

(A) has the largest standard deviation, because all of the scores are far from the center.

b) (3 pts) Which of these (same) sets of six quiz scores has the *smallest* standard deviation? (Circle your answer; do not bother to explain or perform any calculations.)

(A) 0, 0, 0, 10, 10, 10 (B) 5, 5, 6, 6, 7, 7 (C) 0, 3, 5, 6, 8, 10

(B) has the smallest standard deviation, because all of the scores are near the center.

c) (5 pts) Create a hypothetical example of 5 quiz scores (integers from 0 to 10, with repeats allowed) with the property that 4 of them (i.e., 80%) are smaller than the mean. (You do not have to actually calculate the mean.)

One example that works is: 0, 0, 0, 0, 10 (mean = 2).

d) (6 pts) Suppose that I calculate the median score on an exam to be 78. Then I decide to add 2 points to every student's score. How would this change affect the median, if at all? (Be as specific as possible.) Explain briefly.

All of the scores would increase by 2 points, so the middle score(s) would increase by 2 points, so the median would increase by 2 points to 80.