Statistics 1000 – Summer I 2009 Name:____SOLUTIONS_________________
June 15, 2009

Test # 1

Please complete the following problems. Be sure to ask me if you have any questions or anything is unclear. Partial credit will be given, so **please be sure to show all of your work**. Good luck!

1. (5 pts) Define a sample and a population.

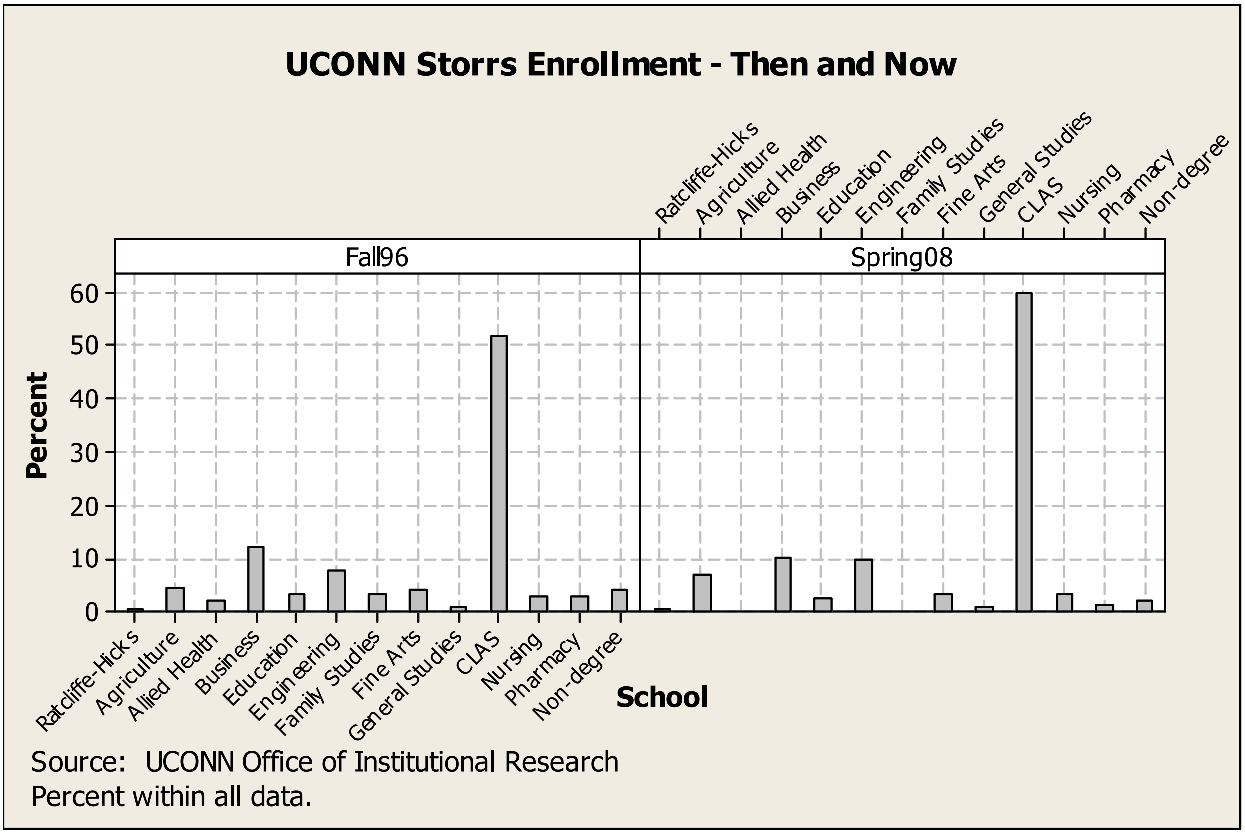
   A population is <u>all</u> possible units of interest, while a sample is a subset of that.

2. (5 pts) List the two types of variables, along with a graphical and a numerical method you can use to summarize each of them.

   <u>Qualitative</u> – These are variables whose measurements denote groups. They can be summarized graphically with a pie chart or bar graph. A numerical summary would be a table of frequencies for each variable group.

   <u>Quantitative</u> – These are variables whose values denote number measurements. They can be summarized graphically with a dotplot, stem-and-leaf plot, histogram, or boxplot. Numerical summaries that could be used are the five-number summary or the mean and standard deviation.

3. (6 pts) The UCONN office of institutional research provides statistics on the enrollment composition of Undergraduates at UCONN's Storrs campus. Two bar charts of the undergraduate enrollment by school are given below. The first shows enrollment percentages in the Fall of 2006, and the second shows enrollment percentages in the Spring of 2008[1].



Interpret this bar chart. Describe the school enrollment of UCONN undergraduates for each of the two years, and compare the two.

From the graph, we see that CLAS has the highest enrollment, and this increased over time, although that may be due to absorbing the Family Studies program, rather than a real increase in enrollment. Business had the next largest, though the enrollment dropped over time. The smallest schools are the Ratcliffe-Hicks and the General Studies programs.

[1] By Spring 2008, Allied Health had become part of the School of Agriculture and Family Studies had become part of CLAS.

4. (10 pts) The following data represent the LSAT scores of some students in a Connecticut college in 2005. Use these data to answer the following questions:

157	145	126	153	130	131	141
139	151	142	127	142	137	138
151	132	138	137	122	141	133
141	136	122				

a. Find the mean and standard deviation of these data.

Using my calculator, I got $\bar{x} = 138$ and $s = 9.302$

b. Find the five-number summary of these data.

Min	Q1	Median	Q3	Max
122	131.5	138	142	157

c. Are there any outliers? (Use the 1.5 IQR rule here)

No, there are not any outliers.

$IQR = Q_3 - Q_1 = 142 - 131.5 = 10.5$

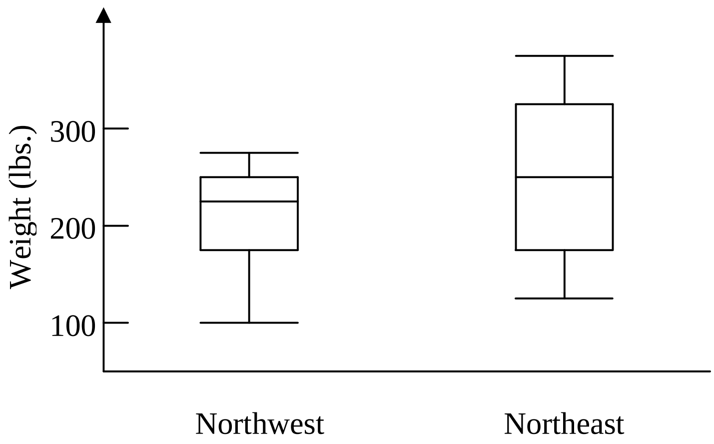
$Q_1 - 1.5IQR = 131.5 - 1.5(10.5) = 115.75$
$Q_3 + 1.5IQR = 142 + 1.5(10.5) = 157.75$

Since no observations were below 115.75 or above 157.75, there are no outliers.

5. (5 pts) The 2004 Red Sox payroll was quite expensive. After looking at the distribution of player's salaries, I found that the mean salary was $4,243,283 and the median salary was $3,087,500. What, if anything, can this tell us about the distribution of salaries for the 2004 Red Sox?

Since the mean is larger than the median, it tells me that the distribution of salaries is skewed to the right.

6. (5 pts) The weights of two populations of white-tail deer have been measured, one in the Northwest, and the other in the Northeast[2]. Boxplots of the weights for the two groups are given below:

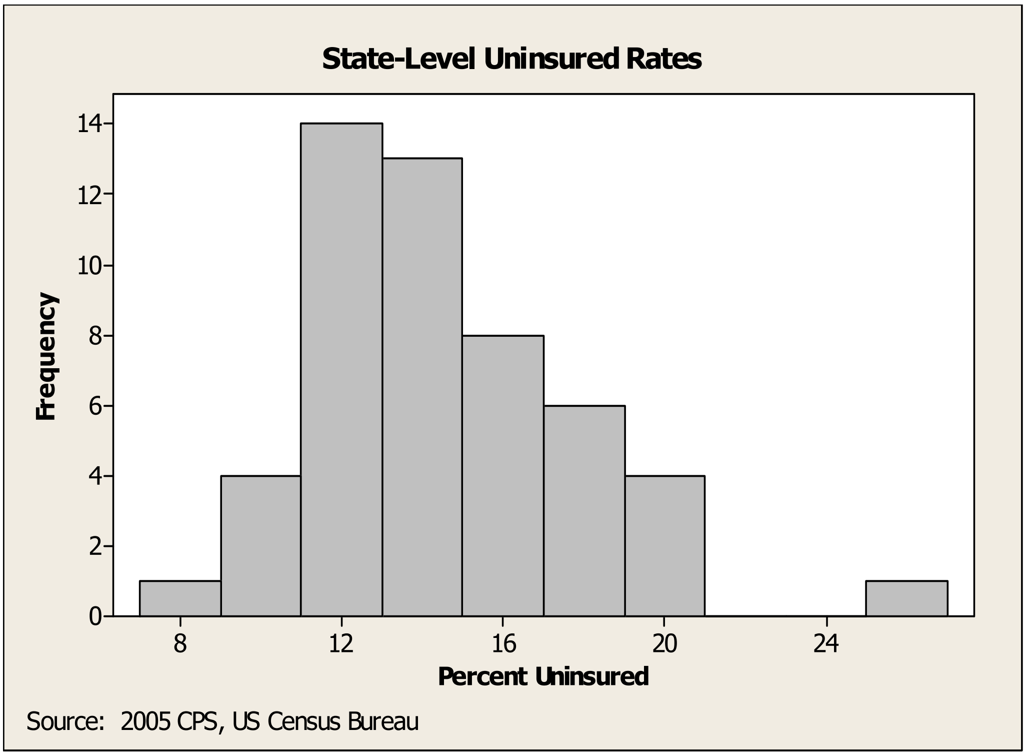


Use the boxplots to compare the distributions of weights for the two groups of deer.

While the means are similar (around 220 lbs), the weights of the NE dear are more variable. The deer weights in the NW are skewed to the left, while those in the NE are more symmetric.

[2] This is fictional data. Don't plan your deer hunting trip based on this!

7. (5 pts) Every fall, the US Census Bureau reports on health insurance coverage in the United States. A histogram of the percentage of people without health insurance among all 50 states and the District of Columbia is given below:



Based on the histogram above, describe the distribution of these data. (Consider such things as the center, skewness, spread, and presence of outliers.)

The center of these data is at about 14% uninsured. The data are skewed to the right, with a possible outlier at 25% uninsured. The spread of the data is from about 8% to 25%, including the outlier.

8. (6 pts) Suppose that the number of traffic stops per day on Eagleville Road is a mound-shaped distributed with a mean of 12 and a standard deviation of 2. One day the UCONN police made 19 traffic stops.

   a. Calculate the Z-score corresponding to x = 19 traffic stops.

   $$Z = \frac{x - \mu}{\sigma} = \frac{19 - 12}{2} = \frac{7}{2} = 3.5$$

   b. Would you consider the observation x = 19 to be an outlier? Please justify your response.

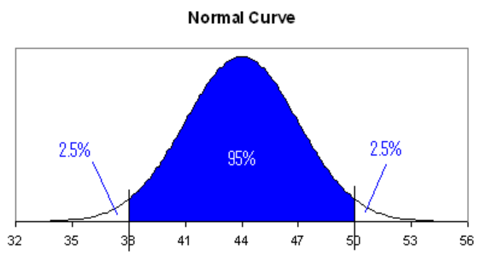
   Since the Z score calculated above in part (a) is more than 3, I think that this day with 19 traffic stops is an outlier.

9. (6 pts) Suppose the life span of a given brand of auto batteries is bell-shaped, with a mean of 44 months and a standard deviation of 3 months.

   a. Find the percentage of batteries that will have a life span of 38 to 50 months.

   Note that 38 is two standard deviations below the mean of 44, and 50 is two standard deviations above the mean. Using the Empirical Rule, there must be about 95% of batteries with lifespans between 38 and 50 months.

   b. Find the percentage of batteries that will have a life span of more than 50 months.



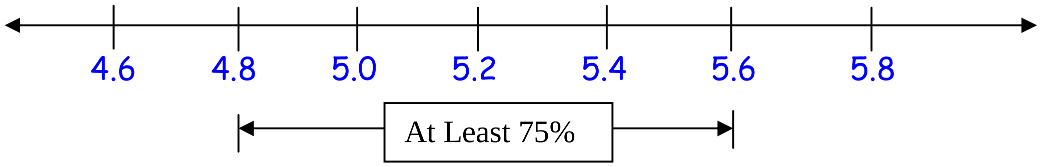
Since 95% of the data are within 38 and 50, there is 5% left over.

That leaves 2.5% on each side.

So, about 2.5% of batteries have a lifespan of more than 50 months.

10. (4 pts) Every July, the famed Tour de France bike race circles France. For a certain stage that is 170 km (106 miles!), the tour organizers suppose that the finishing times (in hours) have a mean of 5.2 hours and a standard deviation of 0.2 hours. If we don't know anything about the shape of the distribution, what percentage of the riders will finish between 4.8 and 5.6 hours?

First, we must recognize that 4.8 to 5.6 hours is +/- 2 standard deviations from the mean. Since we don't know anything about the shape of the distribution, we use Chebyshev's Rule. This states that at least 75% of the data are within 2 standard deviations.

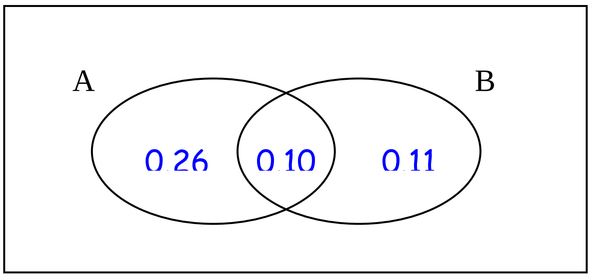


11. (7 pts) Use the given information to answer the questions below. I've given a Venn diagram in case you'd like to use one.

$P(A) = 0.36$

$P(B) = 0.21$

$P(A \cap B) = 0.10$



a. Find $P(A \cup B)$.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.36 + 0.21 - 0.10 = 0.47$$

or

$$P(A \cup B) = 0.26 + 0.10 + 0.11 = 0.47$$

b. Find $P(B^C)$.

$$P(B^C) = 1 - P(B) = 1 - 0.21 = 0.79$$

c. Find $P(B \mid A)$.

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{0.10}{0.36} = 0.278$$

12. (10 pts) Two fair 6-sided dice are tossed, and we record the values shown on the top of each of the two dice. Define the following two events:

A = the event that the sum of the two dice is in an even number.
B = the event that the sum of the two dice is less than or equal to 4.

	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

a. Write down the sample points that are in the event A.

A = {(1,1), (1,3) , (1,5), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5)
(4,2), (4,4), (4,6), (5,1), (5,3), (5,5), (6,2), (6,4), (6,6)}

b. Which sample points are in the event $A \cap B$?

The events in B are B = {(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)}
Then $A \cap B = \{(1,1), (1,3), (3,1), (2,2)\}$

c. Find the probability $P(A \cup B)$.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{18}{36} + \frac{6}{36} - \frac{4}{36} = \frac{20}{36} = \frac{5}{9} = 0.556$$

d. Are the events A and B mutually exclusive? Justify your answer.

No, they are not mutually exclusive. We see above that they overlap.

13. (5 pts) Suppose that 15% of people have both cable television and a cell phone, 35% have cable television, and 55% of people have a cell phone.

Are the events of having cable television and having a cell phone independent? (Show this mathematically.)

We must check if $P(A \cap B) \stackrel{?}{=} P(A) \times P(B)$. We have $P(A \cap B) = 0.15$, and $P(A) \times P(B) = 0.35 \times 0.55 = 0.1925$. These are not equal, so the two events are not independent.

14. (11 pts) You are overseeing training for a large company. You compile the following table of probabilities describing the education level and training level of employees.

Education Level		Training Level 1	Training Level 2	Training Level 3
	A	0.11	0.05	0.07
	B	0.03	0.18	0.16
	C	0.20	0.08	0.12

a. Fill in the missing probability in the table. It is 0.18, so things add to 1.

b. Find the probability that an employee has completed education level B.

$P(B) = 0.03 + 0.18 + 0.16 = 0.37$

c. What is the probability that an employee has completed education level C and training level 2?

$P(C \text{ and Level } 2) = 0.08$

d. Find the probability that a person completed level 3 training, given that they had achieved education level B.

$$P(\text{ Level } 3 \mid B) = \frac{P(\text{Level } 3 \text{ and } B)}{P(B)} = \frac{0.16}{0.37} = 0.432$$

Or, you could restrict the sample space. We know that we are in level B. Now we have:

Education Level		Training Level 1	Training Level 2	Training Level 3
	B	0.03	0.18	0.16

So, now the probability of level 3 is $\frac{0.16}{0.37} = 0.432$.

15. (5 pts) Two events A and B are independent, with $P(A) = 0.25$ and $P(B) = 0.12$. Are A and B mutually exclusive? Justify your answer.

Since A and B are independent, we know that $P(A \cap B) = P(A) \times P(B)$.
Now, $P(A \cap B) = P(A) \times P(B) = 0.25 \times 0.12 = 0.03$.

Since the probability of the intersection is not zero, there is something in the intersection! So A and B can not be mutually exclusive.

16. (5 pts) Aragorn is trying to lead his army to victory over the dark lord Sauron. He can plan a campaign to fight one major battle or three small battles. He believes that he has probability 0.6 of winning the large battle and probability 0.8 of winning a small battle. Victories or defeats in the small battles are independent. Aragorn must win either the large battle or all three small battles to win the campaign. Which strategy should he choose? Gandalf requires that you justify your answer.

We have two possible strategies:

<u>Strategy 1: win 1 large battle.</u>

$P(\text{win the war}) = 0.6$

<u>Strategy 2: win 3 small battles.</u>

$$\begin{aligned} P(\text{win the war}) &= P(\text{win small battle}) \times P(\text{win small battle}) \times P(\text{win small battle}) \\ &= 0.8 \times 0.8 \times 0.8 = 0.512 \end{aligned}$$

I was able to just multiply those probabilities because I was told the small battles are independent.

So, Aragorn's best strategy is to fight one large battle rather than three small ones.