Midterm 2 Review

1. Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections, see plasma proteins as foreign bodies, leading to a high risk of clotting. It is believed that 2% of the population suffer from this disease.

   The test for lupus is very accurate if the person actually has lupus, however is very inaccurate if the person does not. More specifically, the test is 98% accurate if a person actually has the disease, meaning that the probability of a positive test result when a person actually has lupus is 0.98. The test is 74% accurate if a person does not have the disease, meaning that the probability of a negative test result when a person does not actually have lupus is 0.74.

   Dr. Gregory House's team presents him a case where the patient tested positive for lupus. However Dr. House claims that "It's never lupus," even though the test result is positive. What do you think? (Hint: What is the probability that a person who tested positive actually has lupus?)

2. In a classroom with 24 students, 7 students are wearing jeans, 4 students are wearing shorts, 8 students are wearing skirts, and the rest are wearing leggings. If we randomly select 3 students without replacement, what is the probability that one of the selected students is wearing leggings and the other two are wearing jeans?

3. 20% of students at a college own a car, 35% own a bike and 5% own both a bike and a car.

    (a) Make a contingency table or a venn diagram summarizing this information.

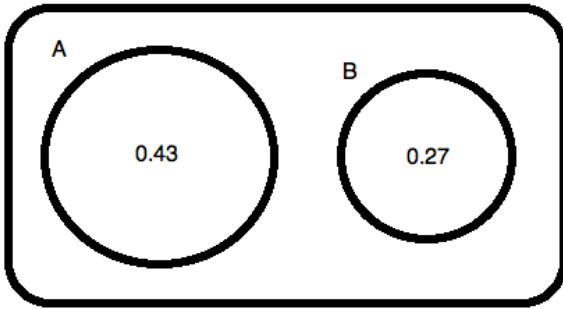   (b) What percent of students at this college own a car or a bike?

    (c) What percent of students at this college do not own a car or a bike?

    (d) What percent of students own a car but not a bike?

    (e) What percent of students own a bike but not a car?

    (f) Are owning a car and a bike independent?

4. Based on the figure below which of the following statements is (are) true.

(A) A and B are mutually exclusive (disjoint).
(B) A and B are independent because they cannot happen at the same time.
(C) P(A or B) = 0.43 * 0.27 = 0.1161
(D) P(A and B) = 0.43 + 0.27 - (0.43 * 0.27) = 0.5839
(E) If A doesn't happen B must happen.
(F) A and B are dependent.

5. Answer the following questions.

   (a) If a fair coin is tossed many times and the last eight tosses are all heads, then the chance that the next toss will be heads is somewhat less than 50%. True or false and justify.

   (b) A coin is tossed and you win a prize if there are more than 60% heads. Would you be surprised if the coin landed on heads more than 60% of the time in 10 tosses? in 100 tosses? Which is more surprising? Which is better in order to win the prize: 10 tosses or 100 tosses? Explain.

   (c) A coin is tossed and you win a prize if there are more than 40% heads. Which is better: 10 tosses or 100 tosses? Explain.

   (d) A coin is tossed and you win a prize if there are between 40% and 60% heads. Which is better: 10 tosses or 100 tosses? Explain.

6. Determine if the below statements are true of false and explain your reasoning.

(a) Drawing a face card (jack, queen, or king) and drawing a red card from a full deck of playing cards are mutually exclusive events

(b) Drawing a face card and drawing an ace from a full deck of playing cards are mutually exclusive events.

7. For commercial flights in 2009, approximately 80% of flights arrived on time, according to the Bureau of Transportation Statistics.

(a) If you randomly select 3 flights and assume they are independent, what is the probability that two or three will arrive on time?

(b) What is the probability that at least one of the 3 flights will be late?

(c) If you randomly select 150 flights and assume they are independent, what is the probability that 114 or fewer arrive on time?

8. Uncle Frank loves to fish. After years of fishing he has discovered the fish in the lake he goes to have normally distributed weight with a mean of 20 oz. and a standard deviation of 6 oz. Uncle Frank gets really excited if he catches one of the heaviest (top) 10% of fish in the lake. What is the cut off for the heaviest 10% of fish in the lake?

(a) 27.68

(b) 12.32

(c) 30.58

(d) 24.38

9. Suppose that the proportion of the adult population who jog is 0.15. Based on this information, which of the following statements is (are) true?

   (A) If we were to take 60 random samples of 40 adults and plotted the proportion of joggers, the resulting distribution would be <u>right skewed</u>.

   (B) If we were to take 600 random samples of 40 adults and plotted the proportion of joggers, the resulting distribution would be <u>symmetric</u>.

   (C) A random sample of 80 adults where 22% are joggers would be considered unusual.

   (D) A random sample of 160 where 22% are joggers would be considered unusual.

   (E) Increasing the sample size from 80 to 160 would cut the standard deviation of the sample proportion in half.

10. In the 2010 Midterm Election 35% of potential voters in California actually voted. (Note that this 35% is a rough estimate but we will take it as a given for the purpose of this question.)

   (a) Which of the following does <u>not</u> need to be satisfied in order to be able to use the Central Limit Theorem to describe the sample proportion of potential Californian voters who actually voted in the 2010 Midterm Election.

      (A) The sample needs to be random.

      (B) The sample size needs to be less than 10% of all potential Californian voters.

      (C) There should be at least 10 "successes" and 10 "failures" in the sample.

      (D) The sample size needs to be greater than or equal to 30.

   (b) Which of the below is true about the shape, center and spread of the distribution of the sample proportion of potential Californian voters who actually voted in random samples of size 100. Make sure to check for the correct notation.

      (A) $\hat{p} \sim N\left(mean = 0.35, SD = \sqrt{\frac{0.35*0.65}{100}} = 0.0477\right)$

      (B) $\bar{x} \sim N\left(mean = 0.35, SD = \sqrt{\frac{0.35*0.65}{100}} = 0.0477\right)$

      (C) $\hat{p} \sim N\left(mean = 0.35, SD = \frac{0.35}{\sqrt{100}} = 0.035\right)$

      (D) $\bar{x} \sim N\left(mean = 0.35, SD = \frac{0.35}{\sqrt{100}} = 0.035\right)$

   (c) Calculate the probability that in a random sample of 100 potential Californian voters majority (more than 50%) has voted.

11. A poll of 1,073 randomly selected adults found that 784 of the respondents wanted Congress to pass a stimulus bill to rescue the economy.

   (a) Report a 95% confidence interval for the population proportion percentage in favor of a stimulus bill. (The conditions for using the CLT have been met)

   (b) Interpret your 95% confidence interval. Interpret what 95% means.

   (c) Report a 80% confidence interval for the population proportion percentage in favor of a stimulus bill. (The conditions for using the CLT have been met)

12. Many professional polls use a 95% confidence interval and report a margin of error of 0.03 (3 percentage points). Find the sample size needed in a survey of a large population if the margin of error for a 95% confidence interval must be 0.03.

13. Your states BigBucks Lottery prize has reached $1 Billion and you decide to play. You have to pick five numbers between 1 and 60 and youll win if your numbers match those drawn by the state. You decide to pick your lucky numbers based on the following random digits.

   43680 98750 13092 76561 58712

   Which numbers do you pick?

   (a) 43, 68, 09, 87, 50
   (b) 4,3,6,8,9
   (c) 43, 09, 50, 13, 27
   (d) 43, 98, 13, 76, 58

14. A migraine is a common type of headache for which patients frequently use acupuncture. To determine whether acupuncture relieves migraine pain, researchers conducted a randomized controlled study where 100 adults suffering from migraine pain were randomly assigned to treatment and control groups (50 in each). Patients in the treatment group received twice-weekly treatment for 12 weeks with an acupuncture program that was specifically designed to treat migraine. Patients in the control group received sham (i.e. fake) acupuncture (needle insertion at nonaccupoint locations). At the end of the study patients were asked whether or not they experienced improvement in migraine headache. The below contingency table summarizes the results of the study.

	No Improvement	Improvement	Total
Control	22	28	50
Treatment	15	35	50
Total	37	63	100

In order to test if the difference between the improvement rates in the treatment and control groups is significant, a randomization test was conducted. Below is step-by-step instructions for how the randomization test was run:

- Write "yes" on 63 index cards and "no" on 37 index cards to indicate whether or not each patient (represented by each card) has shown improvement.

- Shuffle the cards and deal them into two groups of size 50: one pile represents the control group and the other represents the treatment group.

- Count how many in each group showed improvement.

- Repeat 100 times

The histogram shows the distribution of the number of patients in the treatment group that experienced improvement (denoted as nTreatYes) in 100 simulations. Based on this histogram, does it appear that acupuncture is an effective treatment for migraine, or could the results we have seen be simply due to chance?