Stat 1 Final Fall 2010 (50 points total) Name_____

1. Does taking vitamins prevent colon cancer? A study assigned 864 subjects at random to four groups. One group took beta-carotene, another took vitamins C and E, a third took all three, and the fourth group took only a dummy pill (placebo). After four years, there was no significant difference among the groups in the formation of cancer-related polyps in the colon.

   a) (4 pts) Earlier studies of 1000's of people had shown that people who choose to eat lots of vegetables containing these vitamins (including beta-carotene) were less likely to have colon cancer. Explain why the new study is more trustworthy, despite the relatively small sample size.

   b) (1 pt) Suppose that the new study reports a P-value of 0.45. The best interpretation of this p-value is _____

      (i) There is a .45 probability that the null hypothesis is true.
      (ii) There is a .45 probability that the null hypothesis is false.
      (iii) There is a .45 probability of finding a difference in polyps as large or larger than they did, assuming there is no difference between the placebo and the other treatments.
      (iv) There is a .45 probability of finding a difference in polyps as small or smaller than they did, assuming there is no difference between the placebo and the other treatments.
      (v) There is a .45 probability of finding the difference in polyps they did, assuming there is a difference between the placebo and the other treatments.

   c) (1 pt) The best summary of the results (including the P-value) is _____

      (i) All of the treatments are effective in preventing cancer.
      (ii) The treatments differ in preventing cancer.
      (iii) The observed differences in polyp formation were consistent with there being no differences between the four treatments.
      (iv) The observed differences in polyp formation were too large to be due to chance alone.

2. (2 pts) If three people each roll a six-sided die, what is the probability that at least one person will roll a 6?

3. The following table lists the ranks and gender of the faculty members at Purdue University.

	Female	Male	Total
Assistant Professor	151	254	405
Associate Professor	154	397	551
Full Professor	99	642	741
Total	404	1293	1697

   Imagine choosing a faculty member at random. Let A be the event the person selected is female and let B be the event the person selected is a full professor.

   a) (3 pts) Find the marginal and intersection probabilities $P(A)$, $P(B)$ and $P(A \cap B)$.
   b) (2 pts) Find the conditional probabilities $P(A \mid B)$ and $P(B \mid A)$.
   c) (2 pts) Are the events A and B independent? Explain how you know.

4. Professor Everson leaves for a holiday and asks his neighbor to water his plants. If the plants are not watered, there is a 0.8 probability they will die. If they are watered, there is still a 0.1 probability they will die. Professor Everson estimates there is a 0.95 probability that his neighbor will remember to water the plants. Assume these probabilities are correct.

   a) (2 pts) What is the probability the plants will die?

   b) (2 pts) Professor Everson returns and finds his plants have died. What is the conditional probability that his neighbor watered the plants?

5. Suppose that Swarthmore women's heights vary according to a Normal distribution with mean $\mu = 64$ inches and standard deviation $\sigma = 2.5$ inches.

   a) (2 pts) If I choose one woman at random, what is the probability I'll choose someone taller than 65 inches?

   b) (2 pts) If I choose a simple random sample of $n = 4$ women, what is the probability that all four are taller than 65 inches? If you did not answer part a, make up a value to use for answering this part.

   c) (2 pts) If I choose a simple random sample of $n = 4$ women, what is the probability that the *average* height will be over 65 inches?

   d) (2 pts) Suppose the average height for my SRS of $n = 4$ is $\bar{x} = 65.6$ inches. Use this to find a 95% CI for the mean height of all Swarthmore women (still assuming $\sigma = 2.5$).

   e) (3 pts) State hypotheses for testing the claim that the mean height is $\mu = 64$ inches. Does the sample from part d provide evidence at $\alpha = 0.05$ to reject the claim? Report the P-value and explain your conclusions.

   f) (2 pts) Explain qualitatively how the CI and P-value/significance in parts d and e would change if you saw the same average value ($\bar{x} = 65.6$) from a SRS of $n = 16$ (instead of $n = 4$).

6. It has been estimated that 66% of all teenagers have a TV set in their room. You take a SRS of $n = 100$ teens in the Philadelphia area and find that 50 have TV sets in their room.

   a) (3 pts) Construct a 95% confidence interval based on this sample, and explain what parameter it is estimating.

   b) (3 pts) State hypotheses for testing whether $p = 0.66$ in the Philadelpha area. Compute the P-value and explain your conclusion.

   d) (1 pt) The variation from sample to sample when the poll is repeated is described by the standard deviation of the sampling distribution. We would like this variation to be small, so that repeated polls give almost the same result. To reduce the standard deviation, we could _____

      (i) use an SRS of size less than 100.

      (ii) use an SRS of size greater than 100.

      (iii) use a confidence level less than 95%.

      (iv) use a confidence level greater than 95%.

      (v) Both (ii) and (iii).

      (vi) Both (i) and (iv).

7. (3 pts) A friend offers to play a game. You flip two coins and if they both land Heads, your friend will pay you \$2. Otherwise you must pay your friend \$1. Does this game favor you or your friend? Answer by finding the expected value of the amount you would win.

8. In our student survey this semester the correlation between height and shoe length was very close to $r = 0.75$. The average shoe length was about 11 inches, with a standard deviation of about 1 inch. The average height is about 68 inches, with a standard deviation of 4 inches. ...

   a) (1 pt) If I measured in cm instead of inches (1 inch = 2.54 cm) the correlation would be _____

      (i) larger. (ii) smaller. (iii) the same. (iv) we can't tell.

   b) (1 pt) What percent of the variability in heights is explained by the linear regression on shoe length?

   c) (2 pts) Find the least squares prediction equation for predicting a student's height in inches (y) from shoe length in inches (x).

   d) (2 pts) What height would you predict for a student with a 13 inch shoe?

   e) (2 pts) What shoe length would you predict for a student who is 74 inches tall?