SOLUTIONS

Midterm Exam #1
22S:101, Fall 2008

There are 18 questions: 5 true-false (worth 3 points each), 11 multiple choice (worth 5 points each), and two problems (worth 30 points total). For the true-false and multiple choice questions, clearly circle the best answer. For the two problems, circle your final answer and show all your work if you wish to receive partial credit for an incorrect answer. You may use a calculator and your formula sheet. In addition, the first two pages of the binomial distribution table and the standard normal distribution table are attached for your use. Good luck!

True-False questions (3 pts. each).

1. True or False (circle one): If a data set is symmetric around its median, its mean and its median are equal.
2. True or False (circle one): If a dataset is large enough, its histogram will be approximately bell-shaped.
3. True or False (circle one): It is possible for 99% of the observations in a data set to be smaller than the mean of the data set.
4. True or False (circle one): As the relative risk of a disease increases, the odds ratio for that disease decreases.
5. True or False (circle one): The binomial distribution with $n = 10$ and $p = 0.9$ is skewed to the right.

For the remaining two questions on this page, use the following information: In the U.S., 15% of the population have Rh negative blood. Suppose that a random sample of size 9 is taken from the U.S. population.

1. What is the probability that at most 3 of the people in the sample have Rh negative blood?

   (a) 0.0339
   (b) 0.1408
   (c) 0.3333
   (d) 0.8592
   (e) 0.9661

$X \sim \text{bin}(9, 0.15)$

$P(X \leq 3) = 0.2316 + 0.3679 + 0.2597 + 0.1069$
$= 0.9661$

2. What is the probability that an even number of people in the sample have Rh negative blood, given that at most 3 of the people in the sample have Rh negative blood?

   (a) 0.2053
   (b) 0.4913
   (c) 0.5000
   (d) 0.5085
   (e) 0.5202

$$\frac{P(X \text{ is even} \cap X \leq 3)}{P(X \leq 3)} = \frac{P(X=0)+P(X=2)}{0.9661}$$

$$= \frac{0.2316 + 0.2597}{0.9661}$$

$$= 0.5085$$

3. A data set consists of five numbers. The person who collected this data lost it, but remembers that all five numbers were equal to each other. What was the standard deviation of the data?

   (a) 0 (zero spread)
   (b) 1
   (c) 5
   (d) 25
   (e) impossible to determine from the information given

4. Suppose a data set has mean 1.46 and standard deviation 1.80. If we transform this data set by multiplying each observation by 100 and subtracting 100 from the result, what would be the mean and standard deviation, respectively, of the transformed data?

   (a) 46 and 80
   (b) 46 and 180
   (c) 46 and 1800
   (d) 146 and 80
   (e) 146 and 180
   (f) 146 and 1800

$$Y = 100X - 100$$
$$\bar{Y} = 100\bar{X} - 100 = 100(1.46) - 100 = 46$$
$$S_Y = 100\, S_X = 100(1.80) = 180$$

5. Suppose that for a certain diagnostic screening test, the sensitivity was 0.60, the specificity was 0.70, and the prevalence of the disease was 0.20. What is the predictive value of a *negative* test?

   (a) 0.125
   (b) 0.300
   (c) 0.333
   (d) 0.700
   (e) 0.875

$$P(D_2 | T^-) = \frac{P(T^- | D_2)\, P(D_2)}{P(T^- | D_2)\, P(D_2) + P(T^- | D_1)\, P(D_1)}$$
$$= \frac{(0.7)(0.8)}{(0.7)(0.8) + (0.4)(0.2)}$$
$$= \frac{0.56}{0.64} = 0.875$$

6. For the diagnostic screening test described in the previous problem, what is the false positive rate?

   (a) 0.30
   (b) 0.40
   (c) 0.60
   (d) 0.70
   (e) 0.80

$$1 - P(T^- / D_2) = 1 - 0.7 = 0.3$$

7. A certain data set consists of 200 observations, and has a sample mean of 100 and a sample standard deviation of 10. According to Chebyshev's Inequality, the interval $(80, 120)$ contains at least how many observations?

(a) 40
(b) 75
(c) 100
(d) 150
(e) 175

$(80, 120) = (\bar{X} - 2S, \bar{X} + 2S)$, which contains at least 75% of the observations. 75% of 200 = 150

8. What is the interquartile range of the following data set?

1, 3, 5, 7, 9, 11, 16, 16, 20, 26, 27, 29

(a) 6
(b) 16
(c) 17
(d) 18
(e) 28

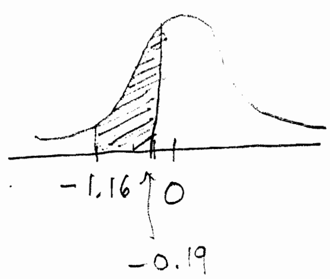
$\frac{12(25)}{100} = 3$; average of $X_{(3)}$ and $X_{(4)}$ is 6

$\frac{12(75)}{100} = 9$; average of $X_{(9)}$ and $X_{(10)}$ is 23

$23 - 6 = 17$

9. Suppose Z has a standard normal distribution. What is $P(-1.16 < Z < -0.19)$?

(a) 0.075
(b) 0.302
(c) 0.425
(d) 0.452
(e) 0.548



$= P(Z > 0.19) - P(Z > 1.16)$
$= 0.425 - .123$
$= 0.302$

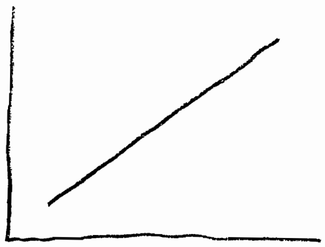
10. Let $\bar{Z}$ be the mean of a random sample of size 16 from a standard normal distribution. What is $P(-1.16 < \bar{Z} < -0.19)$?

(a) 0.001
(b) 0.176
(c) 0.224
(d) 0.760
(e) none of the above

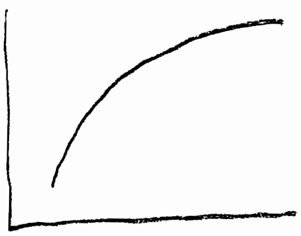
$P\left(\frac{-1.16}{1/\sqrt{16}} < \frac{\bar{Z}}{1/\sqrt{16}} < \frac{-0.19}{1/\sqrt{16}}\right)$
$= P(-4.64 < Z < -0.76)$
$= P(Z > 0.76) - P(Z > 4.64)$
$= 0.224 - 0.000$
$= 0.224$

11. Suppose that a very large random sample is taken from a population of people whose hemoglobin levels are normally distributed (with some mean μ and some standard deviation σ). If a cumulative frequency polygon was constructed from the hemoglobin levels of the people in the sample, which of the following curves it would most closely resemble?

(a)



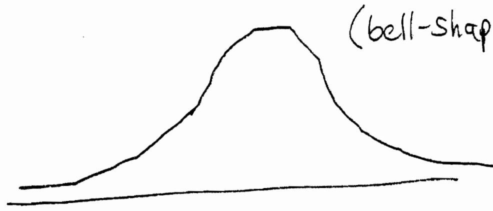
(b)



(c)

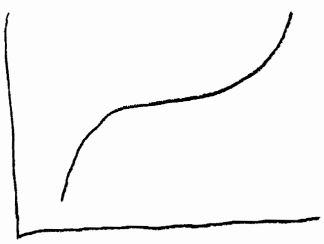
frequency polygon would look like the following:

(bell-shaped)



so cumulative frequency polygon would resemble (e).

(d)



(e)

NAME:

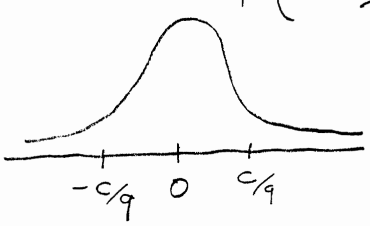
1. (20 pts.) Resting heart rate was measured for a very large group of subjects. The subjects then drank four ounces of coffee within a five-minute period. Fifteen minutes later their heart rates were measured again. Let X represent the *change* in heart rate ("after-coffee" rate minus "before-coffee" rate) for a randomly selected person from this group. Assume that X has a normal distribution with a mean of 5.4 beats per minute and standard deviation 9 beats per minute.

   (a) Find the probability that the "after-coffee" heart rate of the randomly selected person is higher than their "before-coffee" rate. (Hint: First determine how to write this as a probability involving X.)

$$P(\text{after-coffee rate} > \text{before-coffee rate})$$
$$= P(X > 0)$$
$$= P\left(\frac{X-5.4}{9} > \frac{0-5.4}{9}\right)$$
$$= P(Z > -0.6)$$
$$= 1 - P(Z > 0.6)$$
$$= 1 - 0.274 = \boxed{0.726}$$

   (b) Find a number c such that $P(5.4 - c < X < 5.4 + c) = 0.404$.

$$P\left(\frac{5.4-c-5.4}{9} < \frac{X-5.4}{9} < \frac{5.4+c-5.4}{9}\right) = 0.404$$
$$P\left(-\frac{c}{9} < Z < \frac{c}{9}\right) = 0.404$$



$$P\left(0 < Z < \frac{c}{9}\right) = 0.202$$
$$P\left(Z > \frac{c}{9}\right) = 0.5 - 0.202 = 0.298$$
$$\frac{c}{9} = .53$$
$$c = 9(.53) = \boxed{4.77}$$

(c) Find $P(\overline{X} > 1.0)$ where $\overline{X}$ is the mean of a random sample of size 4 from the same group of subjects.

$$\overline{X} \sim N\left(5.4, \frac{9^2}{4}\right) \text{ exactly, by CLT}$$

$$P(\overline{X} > 1.0) = P\left(\frac{\overline{X} - 5.4}{9/2} > \frac{1 - 5.4}{9/2}\right)$$

$$= P(Z > -0.98)$$

$$= 1 - P(Z > 0.98)$$

$$= 1 - .164$$

$$= \boxed{.836}$$

2. (10 pts.) A discrete random variable, X, has a probability distribution given by the following function:

$$P(X = x) = \frac{1}{10} \quad \text{for } x = 1, 2, \ldots, 10.$$

Define the following events, consisting of subsets of the possible outcomes of X:

$$A = \{1, 2, 3, 4, 5\}, \quad B = \{3, 4, 5, 6, 7, 8\}.$$

(a) Are A and B mutually exclusive? Justify your answer.

No, since $P(A \cap B) = P(X = 3, 4, \text{or } 5) = \frac{3}{10} \neq 0$

(b) Are A and B independent? Justify your answer.

Yes, since $P(A) = \frac{5}{10}$, $P(B) = \frac{6}{10}$, and

$P(A \cap B) = P(X = 3, 4, \text{or } 5) = \frac{3}{10} = P(A) \cdot P(B)$.