STAT100 Elementary Statistics and Probability

Exam 1, Sample Test, Summer 2014 **Solution**

Show all work clearly and in order, and circle your final answers. Justify your answers algebraically whenever possible. Calculators are allowed in this exam. You have 80 minutes to take this 104 point exam. If you get more than 100 points, your grade will be 100.

1. (16 points) The following measurements of the diameters (in feet) of Indian mounds in southern Wisconsin were gathered by examining reports in the *Wisconsin Archeologist.*

   22 24 24 30 22 20 28 30 24 34 36 15 37

   (a) (8 points) Calculate the mean and median.

   Solution: The total number of data is $n = 13$. Compute the mean,
   $\bar{x} = \dfrac{22+24+24+30+22+20+28+30+24+34+36+15+37}{13} = \dfrac{346}{13} \approx$ 26.62.**(2 points)**
   To compute the median, first, order the set
   15 20 22 22 24 24 24 28 30 30 34 36 37. **(4 points)**
   Next, compute $np = 13 \times .5 = 6.5$, where $p = .5$ meaning 50th-percentile. Therefore, median is the 7th element: 24. **(8 points)**

   (b) (8 points) Calculate the sample range and the interquartile range.

   Solution: The sample range is: $37 - 15 = 22$. **(2 points)**
   To calculate interquartile range, we need to compute Q_1 and Q_3.
   For Q_1, $p = .25$, $np = 13 \times .25 = 3.25$. Therefore, 1st quantile is the 4th element: 22. **(4 points)**
   For Q_3, $p = .75$, $np = 13 \times .75 = 9.75$. Therefore, 1st quantile is the 10th element: 30. **(6 points)**
   Finally, the interquartile is: $Q_3 - Q_1 = 30 - 22 = 8$. **(8 points)**

2. (24 points) Consider the following data set.

x	1	2	7	4	6
y	5	4	2	3	2

   (a) (10 points) Compute the correlation coefficient between x and y. Do they have a strong linear relation?

Solution: (6 points for the table, 2 points for the formula, 2 points for strong linear relation.)

	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	$(y_i - \bar{y})^2$
	1	5	-3	1.8	9	-5.4	3.24
	2	4	-2	.8	4	-1.6	.64
	7	2	3	-1.2	9	-3.6	1.44
	4	3	0	-.2	0	0	.04
	6	2	2	-1.2	4	-2.4	1.44
$\sum$	20 $\bar{x} = 4$	16 $\bar{y} = 3.2$	0	0	$S_{xx} = 26$	$S_{xy} = -13$	$S_{yy} = 6.8$

$r = \dfrac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \dfrac{-13}{\sqrt{26}\sqrt{6.8}} \approx -0.978$. As $|r|$ is close to 1, x and y have strong linear relation.

(b) (10 points) Find the equations of the least squares fitted line.

Solution: (4 points for $\hat{\beta}_0$ and $\hat{\beta}_1$ each. 2 points for the equation.)
$\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{-13}{26} = -.5$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 3.2 - (-.5) \cdot 4 = 5.2$.
Therefore, the least square fitted line reads $y = -.5x + 5.2$.

(c) (4 points) Using the fitted line, predict the value of y when $x = 5$.

Solution: Plug in $x = 5$ to the fitted line, $y = -.5 \cdot 5 + 5.2 = 2.7$. **(4 points)**

3. (20 points) For two events A and B, the following probabilities are given.

$$P(A) = .4 \qquad P(B) = .25 \qquad P(A|B) = .7$$

Use the appropriate laws of probability to calculate

(a) (4 points) $P(\bar{A})$.

Solution: Use law of complement: $P(\bar{A}) = 1 - P(A) = .6$ **(4 points)**.

(b) (4 points) $P(AB)$.

Solution: Use definition of conditional probability: $P(AB) = P(B)P(A|B) = .175$ **(4 points)**.

(c) (4 points) $P(A \cup B)$.

Solution: Use addition law: $P(A \cup B) = P(A) + P(B) - P(AB) = .475$ **(4 points)**.

(d) (4 points) $P(A\bar{B})$.

Solution: Note that $(A\bar{B}) \cup (AB) = A$, and $(A\bar{B})(AB) = \emptyset$. Therefore, use addition law: $P(A\bar{B}) = P(A) - P(AB) = .225$ **(4 points)**.

(Full credit if using Venn diagram to get the answer.)

(e) (4 points) Are the two events independent? Why or why not?

Solution: They are not independent as $P(A) \neq P(A|B)$.

(1 point for not independent. Other verification such that $P(AB) \neq P(A)P(B)$ are also accepted.)

4. (20 points) X is a discrete random variable. Its distribution is given as below.

x_i	0	1	2	3
$f(x_i)$	27/64	27/64	9/64	1/64

(a) (4 points) Find $\mathbb{E}X$.

Solution: From the following table, we get $\mathbb{E}X = 3/4$ **(4 points)**.

	x_i	$f(x_i)$	$x_i f(x_i)$	x_i^2	$x_i^2 f(x_i)$
	0	27/64	0	0	0
	1	27/64	27/64	1	27/64
	2	9/64	9/32	4	9/16
	3	1/64	3/64	9	9/64
$\sum$		1	$\mu = 3/4$		$\mathbb{E}X^2 = 9/8$

If one observes that $X \sim \text{Binom}(3, 1/4)$, then $\mathbb{E}X = 3 \cdot 1/4 = 3/4$.

(b) (4 points) Find $\text{Var}(X)$ and $\text{sd}(X)$.

Solution: From the table above, $\text{Var}(X) = \mathbb{E}X^2 - \mu^2 = 9/16$ **(2 points)**. Therefore, $\text{sd}(X) = \sqrt{\text{Var}(X)} = 3/4$ **(4 points)**.

If one observes that $X \sim \text{Binom}(3, 1/4)$, then $\text{Var}(X) = 3 \cdot 1/4 \cdot 3/4 = 9/16$.

(c) (6 points) Suppose $Y = 4X - 3$. Find $\mathbb{E}Y$ and $\text{Var}(Y)$.

Solution: $\mathbb{E}Y = \mathbb{E}(4X - 3) = 4\mathbb{E}X - 3 = 4 \cdot 3/4 - 3 = 0$. **(3 points)**

$\text{Var}(Y) = \text{Var}(4X - 3) = 4^2\text{Var}(X) = 16 \cdot 9/16 = 9$. **(6 points)**

(d) (6 points) [*] The *skewness* of a random variable X is defined by

$\text{Skew}(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$, where $\mu = \mathbb{E}X$ and $\sigma = sd(X)$.

Find Skew(X).

Solution: $\text{Skew}(X) = \mathbb{E}\left[\frac{(X-\mu)^3}{\sigma^3}\right] = \frac{1}{\sigma^3}\mathbb{E}(X-\mu)^3$. Use the following table to compute the centered third moment.

	x_i	$f(x_i)$	$x_i - \mu$	$(x_i-\mu)^3$	$(x_i-\mu)^3 f(x_i)$
	0	27/64	-3/4	-27/64	-729/4096
	1	27/64	1/4	1/64	27/4096
	2	9/64	5/4	125/64	1125/4096
	3	1/64	9/4	729/64	729/4096
$\sum$		1			$\mathbb{E}(X-\mu)^3 = 9/32$

Therefore, $Skew(X) = \frac{9/32}{(3/4)^3} = \frac{2}{3}$ **(6 points)**.

5. (24 points) An urn contains two green balls and three red balls.

(a) (12 points) Suppose two balls will be drawn at random one after another and *without* replacement (i.e., the first ball drawn is *not* returned to the urn before te second one is drawn). Find the probabilities of the events

$$A = \{\text{Red ball appears in the first draw}\}$$
$$B = \{\text{Red ball appears in the second draw}\}$$

Solution: The first ball is drawn randomly out of 2 greens and 3 reds, so the probability of drawing a red is $P(A) = \frac{3}{5} = .6$ **(3 points)**.

For the second ball, there are 2 cases.

Case #1: The first ball drawn is red (i.e. given A). There are 2 greens and 2 reds left. The probability of drawing a red is $P(B|A) = \frac{2}{4} = .5$ **(6 points)**.

Case #2: The first ball drawn is green (i.e. given $\bar{A}$). There are 1 greens and 3 reds left. The probability of drawing a red is $P(B|\bar{A}) = \frac{3}{4} = .75$ **(9 points)**.

Using rule of total probability, we get

$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) = .5 \cdot .6 + .75 \cdot .4 = .6$ **(12 points)**.

(No point for claiming $P(B) = P(A) = .6$)

(b) (12 points) Suppose four balls will be drawn at random *with* replacement (i.e. The ball drawn will be returned to the urn before the next one is drawn). Find the probability that at least two red balls are drawn.

Solution: This is a Bernoulli trail. Let X be a random variable representing the number of red ball drawn. Then X has the following binomial distribution:

$$X \sim \text{Binom}(4, .6). \quad \text{(2 points for } n\text{, 2 points for } p\text{)}$$

Compute the distribution of X as the following:

$$P(X = 0) = \binom{4}{0}(.6)^0(.4)^4 = .0256$$
$$P(X = 1) = \binom{4}{1}(.6)^1(.4)^3 = .1536$$
$$P(X = 2) = \binom{4}{2}(.6)^2(.4)^2 = .3456$$
$$P(X = 3) = \binom{4}{3}(.6)^3(.4)^1 = .3456$$
$$P(X = 4) = \binom{4}{4}(.6)^4(.4)^0 = .1296$$

(6 points for the distribution. Computing only for those entries needed to solve the problem is accepted.)

Finally, at least two red balls are drawn means $X \geq 2$.

$P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4) = .3456 + .3456 + .1296 = .8208$

(2 points. Full credit if using $P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$.)