# STAT100 Elementary Statistics and Probability

**Exam 1, Monday, July 28, 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 106 point exam. If you get more than 100 points, your grade will be 100.

1. (16 points) The city of Madison regularly checks the quality of water at swimming beaches located on area lakes. Fifteen times of concentration fecal coliforms, in number of colony forming unites (CFU) per 100 ml of water, was measured during the summer at one beach.

180 1600 90 140 50 260 400 90 380 110 10 60 20 340 80

(a) (6 points) Calculate the sample mean and median.

> **Solution:** The total number of data is $n = 15$. Compute the mean, $\bar{x} = \frac{180+1600+90+140+50+260+400+90+380+110+10+60+20+340+80}{15} = 254$. **(2 points)**
> To compute the median, first, order the set
> 10 20 50 60 80 90 90 110 140 180 260 340 380 400 1600. **(4 points)**
> Next, compute $np = 15 \times .5 = 7.5$, where $p = .5$ meaning 50th-percentile. Therefore, median is the 8th element: 110. **(6 points)**

(b) (6 points) One day, the water quality was bad - the reading was 1600 CFU - and the beach was closed. Drop this value and calculate the sample mean and median for the days where water quality was suitable for swimming.

> **Solution:** If we drop 1600, the total number of data is $n = 14$. Compute the mean, $\bar{x} = \frac{180+90+140+50+260+400+90+380+110+10+60+20+340+80}{14} \approx 157.86$. **(2 points)**
> To compute the median, first, order the set
> 10 20 50 60 80 90 90 110 140 180 260 340 380 400. **(4 points)**
> Next, compute $np = 45 \times .5 = 7$, where $p = .5$ meaning 50th-percentile. Therefore, median is the 7th element: 90. **(6 points)**

(c) (4 points) Which one is more robust: sample mean or median? Explain why.

> **Solution:** Sample median is more robust. **(2 points)** Note that the difference of sample median by dropping one data point is $110 - 90 = 20$, while the difference of sample mean is $254 - 157.86 = 93.14$. It is far more than 20. **(4 points)**

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

| $x$ | 2 | 3 | 5 | 7 | 8 |
|---|---|---|---|---|---|
| $y$ | 9 | 7 | 5 | 3 | 3 |

(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.)

| | $x_i$ | $y_i$ | $x_i - \bar{x}$ | $y_i - \bar{y}$ | $(x_i - \bar{x})^2$ | $(x_i - \bar{x})(y_i - \bar{y})$ | $(y_i - \bar{y})^2$ |
|---|---|---|---|---|---|---|---|
| | 2 | 9 | -3 | 3.6 | 9 | -10.8 | 12.96 |
| | 3 | 7 | -2 | 1.6 | 4 | -3.2 | 2.56 |
| | 5 | 5 | 0 | -.4 | 0 | 0 | .16 |
| | 7 | 3 | 2 | -2.4 | 4 | -4.8 | 5.76 |
| | 8 | 3 | 3 | -2.4 | 9 | -7.2 | 5.76 |
| $\sum$ | 25<br>$\bar{x} = 5$ | 27<br>$\bar{y} = 5.4$ | 0 | 0 | $S_{xx} = 26$ | $S_{xy} = -26$ | $S_{yy} = 27.2$ |

$r = \dfrac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \dfrac{-26}{\sqrt{26}\sqrt{27.2}} \approx -0.978$. 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.)
$\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{-26}{26} = -1$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 5.4 - (-1) \cdot 5 = 10.4$.
Therefore, the least square fitted line reads $y = -x + 10.4$.

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

**Solution:** Plug in $x = 4$ to the fitted line, $y = -4 + 10.4 = 6.4$. **(4 points)**

3. (20 points) A table containing probabilities of two events $A$ and $B$ is given in the right.

| | $A$ | $\bar{A}$ | |
|---|---|---|---|
| $B$ | .12 | | .4 |
| $\bar{B}$ | | | |
| | | .7 | |

(a) (4 points) Fill in the missing entries of the table.

> **Solution:** Here is the full table **(4 points)**.
>
> | | $A$ | $\bar{A}$ | |
> |---|---|---|---|
> | $B$ | .12 | .28 | .4 |
> | $\bar{B}$ | .18 | .42 | .6 |
> | | .3 | .7 | 1 |

(b) (4 points) Find the probability $P(A\bar{B})$.

> **Solution:** $P(A\bar{B}) = .18$ **(4 points)**.

(c) (4 points) Find the probability $P(A|\bar{B})$.

> **Solution:** $P(A|\bar{B}) = \dfrac{P(A\bar{B})}{P(\bar{B})} = \dfrac{.18}{.6} = .3$ **(4 points)**.

(d) (4 points) Are $A$ and $\bar{B}$ disjoint? Explain Why.

> **Solution:** No **(2 points)**. It is because $P(AB) = .12 \neq 0$ **(4 points)**.

(e) (4 points) Are $A$ and $\bar{B}$ independent? Explain Why.

> **Solution:** Yes **(2 points)**. It is because $P(A|\bar{B}) = P(A) = .3$ **(4 points)**. (Full credit if one checks $P(A\bar{B}) = P(A)P(\bar{B})$.)

4. (24 points) In a shipment of 12 room air conditioners, there are 3 with defective thermostats. Two air conditioners will be selected at random and inspected one after another. Denote events

$$A = \{\text{The first one is defective}\},$$
$$B = \{\text{The second one is defective}\}.$$

Express the following events using in set notations and find their probabilities.

(a) (4 points) The first is defective.

**Solution:** $P(A) = \frac{3}{12} = .25$ **(4 points)**

(b) (4 points) The second is good given that the first is defective.
*Hint: if the first one is defective, then there are 11 left, with 2 defective ones.*

**Solution:** $P(\bar{B}|A) = \frac{11-2}{11} \approx .818.$**(4 points)**

(c) (4 points) The first is defective and the second is good.

**Solution:** $P(A\bar{B}) = P(A)P(\bar{B}|A) = \frac{3}{12} \cdot \frac{9}{11} = \frac{9}{44} \approx .205$ **(4 points)**.

(d) (6 points) The first is defective given the second is good.

**Solution:** Use Bayes formula to compute $P(A|\bar{B})$:

$$P(A|\bar{B}) = \frac{P(A\bar{B})}{P(A)P(\bar{B}|A) + P(\bar{A})P(\bar{B}|\bar{A})} = \frac{\frac{9}{44}}{\frac{3}{12} \cdot \frac{9}{11} + \frac{9}{12} \cdot \frac{8}{11}} = \frac{3}{11} \approx .273.$$

**(6 points)**. (Full credit if claiming $P(\bar{B}) = \frac{9}{12} = .75$.)

(e) (6 points) Exactly one is defective.

**Solution:** The event is $(A\bar{B}) \cup (\bar{A}B)$ **(2 points)**. As $A\bar{B}$ and $\bar{A}B$ are disjoint,

$$P((A\bar{B}) \cup (\bar{A}B)) = P(A\bar{B}) + P(\bar{A}B) = \frac{9}{44} + \frac{9}{44} = \frac{9}{22} \approx .409,$$

where $P(\bar{A}B) = P(\bar{A})P(B|\bar{A}) = \frac{9}{12} \cdot \frac{3}{11} = \frac{9}{44}$ **(6 points)**.

**Remark:** Full credit if one solves the problem correctly using tree diagram.

5. (12 points) $X$ is a discrete random variable. Its distribution is given as below.

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

(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 | 1/64 | 0 | 0 | 0 |
| | 1 | 9/64 | 9/64 | 1 | 9/64 |
| | 2 | 27/64 | 27/32 | 4 | 27/16 |
| | 3 | 27/64 | 81/64 | 9 | 243/64 |
| $\sum$ | | 1 | $\mu = 9/4$ | | $\mathbb{E}X^2 = 45/8$ |

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

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

**Solution:** From the table above, $\text{Var}(X) = \mathbb{E}X^2 - \mu^2 = 9/16$ **(3 points)**. Therefore, $\text{sd}(X) = \sqrt{\text{Var}(X)} = 3/4$ **(4 points)**.
If one observes that $X \sim \text{Binom}(3, 3/4)$, then $\text{Var}(X) = 3 \cdot 3/4 \cdot 1/4 = 9/16$.

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

**Solution:** $\mathbb{E}Y = \mathbb{E}(4X + 3) = 4\mathbb{E}X + 3 = 4 \cdot 9/4 + 3 = 12$. **(2 points)**
$\text{Var}(Y) = \text{Var}(4X + 3) = 4^2\text{Var}(X) = 16 \cdot 9/16 = 9$. **(4 points)**

6. (10 points) Suppose 15% of the trees in a forest have severe leaf damage from air pollution. If 5 trees are selected at random, find the probability that no more than two have severe leaf damage.

**Solution:** This is a Bernoulli trail. Let $X$ be a random variable representing the number of selected trees with sever leaf damage. Then $X$ has the following binomial distribution **(4 points)**:

$$X \sim \text{Binom}(5, .15). \quad \text{(2 points for } n\text{, 2 points for } p\text{)}$$

Compute the distribution of $X$ as the following:

$$P(X = 0) = \binom{5}{0}(.15)^0(.85)^5 \approx .444$$
$$P(X = 1) = \binom{5}{1}(.15)^1(.85)^4 \approx .392$$
$$P(X = 2) = \binom{5}{2}(.15)^2(.85)^3 \approx .138 \quad \textbf{(8 points)}$$

Finally, no more than two have severe leaf damage means $X \leq 2$.

$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = .444 + .392 + .138 = .974$ **(10 points)**. (Full credit if using $P(X \leq 2) = 1 - P(X = 3) - P(X = 4) - P(X = 5)$.)