STAT 2507 Solution-Assignment # 2 (Chapter 3 and 4) Fall 2013

Total of marks=100.

Part I. Lab questions. Use only the blanks left to answer lab questions.

1. The Minitab data set 'sat' has the mean *verbal* and *math* scores of the Scholastic Aptitude Test (SAT) for the academic years 1967 to 1981. Retrieve the 'sat' data set (by choosing **file** → **Open Worksheet** (you need to click on "look in Minitab Sample Data folder" icon) and then choosing sat).

   (a) [3] Fit a L-S regression line using *year* to predict the *verbal* scores (verbal score is called the response variable). What is the equation of the L-S regression line?

   Verbal=7373-3.51 Year.

   (b) [3] Fit a L-S regression line using *year* to predict the *math* scores (response variable). What is the equation of the L-S regression line? **Math=4955-2.27 Year.**

   (c) [3] Fit a L-S regression line using *verbal* scores to predict the *math* scores (response variable). What is the equation of the L-S regression line? **Math=193+0.643 Verbal.**

   (d) [4] What is the predicted score for *math* in 1970? **483.1.** In 1983? **453.59.** In 2000? **415.** Which of these seem to make sense? **For the years in the range of data, 1970 to 1983.**

   (e) [3] Use your answer to part (c) to predict the *math* score if the *verbal* score is 450. Answer: **482.35.** Can you predict the *verbal* score if the *math* score is 480? Answer: **Not based on part (c).** Why? **Equation in part (c) is a model to predict y (Math), based on x (Verbal).**

2. (Probability as relative frequency)

   You need to use the Filliping a Fair coin applet to answer this question. This applet is available at (http://www.stat.tamu.edu/ west/ph/probsim3.html).

   (a) [1] Flip the coin 10 times what is $\frac{\#\ of\ Heads\ observed}{total\ \#\ of\ flips}$?**4/10=0.4 , # of heads is random number**

   (b) [1] (press Reset) Flip the coin 100 times. What is $\frac{\#\ of\ Heads\ observed}{total\ \#\ of\ flips}$? **44/100=0.44, # of heads is random number**

(c) [1] (Press Reset) By pressing Flip for $n = 1000$ once, flip the coin 1000 times. What is $\frac{\#\ of\ Heads\ observed}{total\ \#\ of\ flips}$? **497/1000, # of heads is random number**

(d) [1] Compare the relative frequencies in part (a), part (b) and part (c). Which one is closer to 0.5?**part (c)**

(e) [1] Suppose that you could flip it infinitely many times, then what the value of $\frac{\#\ of\ Heads\ observed}{total\ \#\ of\ flips}$ would be? **1/2**

3. (Conditional probability and Independence) the link (http://www.probandstats2e.nelson.com/student/2ce%20Applets/countDice.html) gives an experiment of tossing two fair dice (one green and one red).
   Let C be the event that the green die shows a number less than or equal to 2.
   Let D be the event that the red die shows a number less than or equal to 3.

   (a) [2] (Press Reset) Obtain $P(C)$= **12/36**

   (b) [2] (Press Reset) Obtain $P(D)$= **18/36**

   (c) [2] (Press Reset) By clicking on the simple events in which the green die has a number at most 2 and the red one has a number at most 3, obtain $P(C \cap D)$= **6/36**

   (d) [2] (Use parts (b) and (c)) If you know that the red die showed a number less than or equal to 3, then obtain the probability that the green shows a number less than or equal to 2, by clicking on the relevant simple events in the applet?
   I.e., what is $P(C|D) =$ **6/18**

   (e) [2] What is your conclusion about the relation between the events C and D? **Independent as** $P(C|D) = P(C)$

4. (Mean and variance of random variables using Minitab)
   The values (1 to 10) in the first column of the Open Office file are the values that random variable x takes (copy and paste them in C1 of your Minitab worksheet) and the second column contains $p(x)$ (copy and paste this column from Open Office file in C2 of your Minitab worksheet)

   (a) [2] Obtain the mean of x, i.e., $\mu = E(x)$ by $Calc \to Calculator \to sum(C1 * C2)$, and store the result in variable C3. What is μ? **1.998047**

   (b) by clicking $Calc \to Calculator \to sum((C1 * *2) * C2) - (sum(C1 * C2)) * *2$. obtain $Var(x) = \sigma^2$. What is σ^2 [2] ?**1.962887** What is σ? [1] **1.401031**

(c) [2] Calculate the interval $(\mu - 2\sigma, \mu + 2\sigma)$? **(-0.800109, 4.800109)**

(d) [2] What values of x are inside the interval in part (c)? **1, 2, 3 and 4**

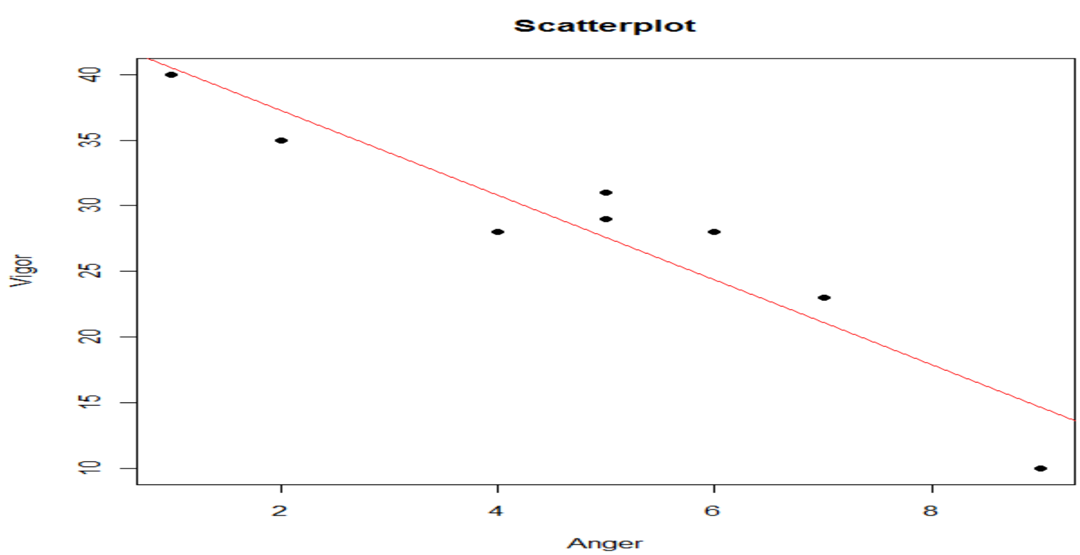
ALSO do the following questions:

1. As part of study of the psychobiological correlates of success in athletes, the following measurements are obtain from members of the U.S. Olympic wrestling team.

Anger x	6	7	5	2	4	5	1	9
Vigor y	28	23	31	35	28	29	40	10

(a) [4] Plot the scatter diagram.

Solution:



(b) [4] Calculate r.

Solution:

$$r = \frac{\sum(x_i y_i) - \sum x_i \sum y_i / n}{\sqrt{(\sum x_i^2 - (\sum x_i)^2/n)(\sum y_i^2 - (\sum y_i)^2/n)}} = \frac{941 - (39 * 224)/8}{\sqrt{(237 - 39^2/8)(6824 - 224^2/8)}} = -0.9387$$

(c) [4] Obtain the least squares line.

Solution:

$b = r\frac{s_y}{s_x} = -0.9387(3.4316) = -3.22$, $a = \bar{y} - b\bar{x} = 224/8 + 3.22(39/8) = 43.70$, so regression line is Vigor=43.70-3.22Anger

(d) [3] Predict the vigor score y when the anger score is $x = 8$.

Solution:

Vigor=43.70-3.22(8)=17.94

2. [8] Suppose that a random variable X follows the distribution given by:

a	-2	-1	0	1	2
$P(X = a)$	1/8	1/4	1/4	?	1/8

Let $Y = X^2$. Find $P(Y \geq 1)$ and $E(Y)$.

Solution:

$P(X = 1) = 1 - (1/8 + 1/4 + 1/4 + 1/8) = 1/4$

a^2	4	1	0	1	4
$P(X = a)$	1/8	1/4	1/4	1/4	1/8

$P(Y \geq 1) = P(Y = 4) + P(Y = 1) = 1/8 + 1/4 + 1/4 + 1/8 = 3/4$

$E(Y) = 4(1/8) + 1(1/4) + 0(1/4) + 1(1/4) + 4(1/8) = 12/8 = 3/2$

3. [10] If 5 balls are thrown at random into 10 boxes, what is the probability that no box will receive more than one ball?

Solution: There are 10^5 possible outcomes in the sample space, i.e., 10 boxes for each one of the 5 balls. If the 5 balls are to be thrown into different boxes (this is equivalent to choosing 5 boxes, for the five balls, out of 10 boxes), there are C_5^{10} ways to do that. So, the probability is $\frac{C_5^{10}}{10^5}$.

4. A survey has revealed that 75% of all college students study. It is also known that 85% of all students who study will graduate, while only 35% of those students who do not study will graduate.

   (a) [6] If a student is randomly selected, what is the probability that he or she will graduate?

   (b) [6] A randomly selected student is observed to graduate. What is the probability that this student studied?

Solution: (a) Let $G = \{student\ graduates\}$;

$H = \{student\ doesnt\ graduate\}$;

$S = \{student\ studies\}$;
$T = \{student\ doesnt\ study\}$
$P(G) = P(G|S)P(S) + P(G|T)P(T) = 0.85 \times 0.75 + 0.35 \times (1 - 0.75) = 0.725$
(b) Using Bayes' rule: $P(S|G) = \dfrac{P(G|S)P(S)}{P(G|S)P(S) + P(G|T)P(T)} = \dfrac{0.85 \times 0.75}{0.725} = 0.879.$

5. For events A and B we have

$$P(A) = 0.3,\ P(B) = 0.8,\ P(A \cup B) = 0.9$$

   (a) [12] Find $P(A|B)$, $P(A' \cap B)$ and $P(B' \cup A')$.

   (b) [3] Are A and B independent? Why?

   Solution: (a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) \rightarrow P(A \cap B) = 0.3 + 0.8 - 0.9 = 0.2$
   $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.8} = 0.25$
   $P(A' \cap B) = P(B) - P(A \cap B) = 0.8 - 0.2 = 0.6$
   $P(B' \cup A') = 1 - P(A \cap B) = 1 - 0.2 = 0.8$
   (b) No there are not independent, since
   $P(A \cap B) \neq P(A)P(B), i.e.,\ 0.2 \neq (0.3)(0.8)$