Math 201: Statistics

November 10, 2011

KEY

Fall 2012, EXAM 1
100 Minutes

Instructions: There are six questions and one bonus question in this exam. Please inspect the exam and make sure you have all the questions. You may only use your calculator and one side of a hand-written A4-size sheet of paper. Do all your work on the paper provided. **You may not exchange any kind of material with another student.**

Remember: ***You must show your work to get proper credit.***

Academic Honesty Code: Koç University Academic Honesty Code stipulates that "copying from others or providing answers or information, written or oral, to others is cheating." By taking this exam, you are assuming full responsibility for observing the Academic Honesty Code.

NAME: _____

LECTURE DAY&TIME : _____

1	/15
2	/15
3	/15
4	/20
5	/15
6	/20
Bonus	/10
Total:	/100

Question 1

The time it takes a waiter to take the order (=yemek siparisi) of the customers in a restaurant are recorded (in minutes) as follows:

3.6 4.9 4.3 5.6 4.7 5.2 4.9 5.3 4.4 4.2 3.7 6.3 4.1

a) Draw a relative frequency histogram of the above data by using class intervals with length 0.5 minutes.

b) Is the histogram left-skewed, symmetric or right-skewed? Why?

c) Write an interval that contains at least 75% of the data by using Chebyshev's rule.

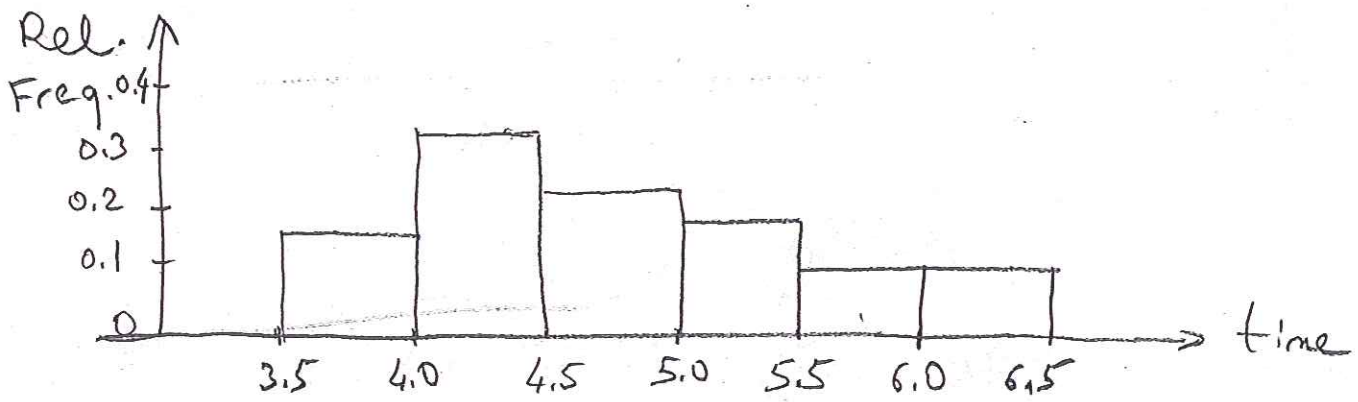
d) Find the 90th percentile of the order times.

3.6 3.7 4.1 4.2 4.3 4.4 4.7 4.9 4.9 5.2 5.3 5.6 6.3

a)

Class	3.5-4.0	4.0-4.5	4.5-5.0	5.0-5.5	5.5-6.0	6.0-6.5
Freq.	2	4	3	2	1	1
Rel Freq	0.15	0.31	0.23	0.15	0.08	0.08

$n = 13$



b) Right-skewed because it has a long tail to the right: the relative frequency of the right-most two intervals (classes) is very low.

c) By Chebyshev's rule $\bar{x} \mp 2s$ will contain at least 75% of the data.

$$\bar{x} = \frac{3.6 + 3.7 + \cdots + 6.3}{13} = 4.71$$

$$s^2 = \frac{(3.6)^2 + \cdots + (6.3)^2 - (13)(4.71)^2}{n-1} = \frac{295.24 - 288.39}{12} = 0.57$$

$\Rightarrow s = 0.76$

$\Rightarrow \bar{x} \mp 2s \equiv (3.19, 6.23)$

d) $(0.90)13 = 11.7 \nearrow 12^{th}$ observation

$\Rightarrow$ 5.6 is 90th percentile

Question 2

The numbers of phone calls made per week by a marketing specialist and relative frequencies are displayed in the table below:

Phone calls	10	11	12	13	14	15
Relative Frequency	0.16	0.26	0.22	0.04	0.18	0.14

a) If the frequency of 12 is 11, what is the frequency of 14 in the data set?

b) Find the mean number of phone calls.

c) Find the interquartile range (IQR).

a) $\frac{11}{n} = 0.22 \Rightarrow n = 50 \Rightarrow$ freq. of $14 = 50(0.18)$
$= 9$

b) $\bar{x} = 10(0.16) + 11(0.26) + 12(0.22) + 13(0.04)$
$+ 14(0.18) + 15(0.14)$
$= 12.24$ phone calls/week

c) $Q_L \equiv 25^{th}$ percentile must be 11 because 16% of the observations is 10 and when the data are ordered, the next 26% of the observations is 11. Since 15% + 26% is greater than 25%, $Q_L = 11$.

Similarly, consider top 25%.

Since $0.14 + 0.18 > 0.25$ (and $0.14 < 0.25$)

we have $Q_U = 14$

$\Rightarrow$ IQR $= 14 - 11 = 3$ phone calls/week

Question 3

A random variable can have possible values of 1, 3, and 6. Its ~~standard deviation~~ variance and expected value are both 3. Find the probability distribution function (pdf) of this variable.

$\mu = 3$

$\sigma = 3$

x	1	3	6
$p(x)$	a	b	c

$$3 = \mu = E(X) = \sum x\,p(x) = 1.a + 3.b + 6.c$$

$$\boxed{a + 3b + 6c = 3}$$

$$\sigma^2 = V(X) = E(X^2) - \mu^2$$

$$3 = 1^2.a + 3^2.b + 6^2.c - (3)^2$$

$$\boxed{12 = a + 9b + 36c}$$

Also $\boxed{a + b + c = 1}$

Solving the three equations in three unknowns, we get

$$a = p(1) = 0.3$$

$$b = p(2) = 0.5$$

$$c = p(3) = 0.2$$

Question 4

A box has 10 blue, 20 red, 30 white, and 40 green balls in it. Each ball has the number 0 or 1 painted on it. 4 of the blue, 5 of the red, 21 of the white, and 20 of the green balls are "1". If one ball is drawn at random from the box,

a) what is the probability that it is red and "1"?

b) what is the probability that it is white or "0"?

c) are the events 'green' and '1' independent?

d) and we are told that a 1 is painted on it, what is the probability that it is blue?

a) 0.05 (5 out of 100)

b) $P(W) + P(0) - P(W,0) = 0.3 + 0.5 - 0.09$

$= 0.71$

c) $P(\text{green}) = 0.4$ $P(1) = 0.5$ (50 out of 100)

$P(\text{green} \cap 1) = 0.2$ $\left(= \frac{20}{100}\right)$

since $P(\text{green}), P(1) = P(\text{green} \cap 1)$,

they are independent.

d) $P(\text{blue} \mid "1") = \frac{P(\text{blue} \cap "1")}{P("1")} = \frac{0.4/100}{50/100}$

$= 0.08$

Q4 → 10 blue, 20 red, 30 white, 40 green balls in a box.

Question 5

a) If 4 balls are drawn at random from the box in question 4 without replacement, what is the probability that they each have a different color?

b) If 10 balls are drawn at random from the box in question 4, what is the probability that more than 7 are green? (Use your calculator to get an exact solution)

a) $E = \{4 \text{ balls drawn have diff colors}\}$

$$P(E) = \frac{\binom{10}{1}\binom{20}{1}\binom{30}{1}\binom{40}{1}}{\binom{100}{4}} = \frac{10 \cdot 20 \cdot 30 \cdot 40}{100 \cdot 99 \cdot 98 \cdot 97} \cdot 4! \cong 0.06$$

b) X = nbr of green balls drawn, [Assuming balls are drawn with replacement → $\frac{40}{100}$] $X \sim Bin(10, 0.4)$

$$P(X > 7) = P(X=8) + P(X=9) + P(X=10)$$
$$= \binom{10}{8}(0.4)^8(0.6)^2 + \binom{10}{9}(0.4)^9(0.6) + (0.4)^{10}$$
$$\cong 0.0106 + 0.00157 + 0.000105$$
$$= 0.012275$$

If balls drawn without replacement:

$$P(X>7) = \frac{\binom{40}{8}\binom{60}{2}}{\binom{100}{10}} + \frac{\binom{40}{9}\binom{60}{1}}{\binom{100}{10}} + \frac{\binom{40}{10}}{\binom{100}{10}}$$

BONUS

Find the numerical value of the expression below using the binomial tables.

$\sum_{i=4}^{8} C_{20,i} 0.3^i 0.7^{20-i} + \sum_{i=8}^{15} C_{15,i} 0.6^i 0.4^{15-i}$

$= P(8 \geqslant X \geqslant 4) + P(Y \geqslant 8)$, where $X \sim Bin(20, 0.3)$ and $Y \sim Bin(15, 0.6)$

$= P(X \leqslant 8) - P(X \leqslant 3) + [1 - P(Y \leqslant 7)]$

$= [0.887 - 0.107] + 1 - 0.213$ $[1 - P(Y \leqslant 7)]$

$= 1.567$

Question 6

Find the regression line (least square fitted line) for the points (1,1), (2,2), (3,3), (4,3). What is the sum of squares of vertical deviations for this line and the given bivariate data? Do the average of the x-coordinates and the average of the y-coordinates satisfy the equation of the regression line?

i	x_i	y_i	$x_i y_i$	x_i^2
1	1	1	1	1
2	2	2	4	4
3	3	3	9	9
4	4	3	12	16
Sum	10	9	26	30

$\Rightarrow \bar{x} = \frac{10}{4} = 2.5$

$\bar{y} = \frac{9}{4} = 2.25$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{26 - \frac{(10)(9)}{4}}{30 - \frac{(10)^2}{4}} = \frac{3.5}{5} = 0.7$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2.25 - (0.7)(2.5) = 0.5$$

Regression line $\hat{y} = 0.5 + 0.7x$

$$\sum_{i=1}^{4} (y_i - \hat{y})^2 = \left(1 - (0.5 + (0.7)(1))\right)^2 + \left(2 - (0.5 + (0.7)(2))\right)^2 + \left(3 - (0.5 + (0.7)(3))\right)^2 + \left(3 - (0.5 + (0.7)(4))\right)^2 = 0.3$$

Yes, $\bar{y}$ and $\bar{x}$ satisfy the regression line, because $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, that is,

$$\bar{y} = \hat{\beta}_1 \bar{x} + \hat{\beta}_0$$