Spring 2012 Midterm #1
Closed book & notes; only a single-sided and handwritten A4 formula sheet and a calculator allowed; 90 minutes. No questions accepted!

Instructions: There are six pages (one cover and five pages with questions) in this exam. Please inspect the exam and make sure you have all 6 pages. You may only use your calculator and your formula sheet. Do all your work on these pages. If you use the back of a page, make sure to indicate that. **You may not exchange any kind of material with another student.** *You might get one bonus point for filling the front cover properly!*

Remember: ***You must show all 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 and SURNAME: KEY **SIGNATURE:** _____

INSTRUCTOR: _____ **LECTURE TIME:** _____

1	/20
2	/20
3	/20
4	/20
5	/20
Total:	/100

Show all your work to be eligible for partial credit and round your **final** answers to **two decimal** places in **all of the exam questions**.

Question 1 (20 points) The following data show the *percentage change* in Dow Jones Industrial Average (ABD'de bir borsa endeksi) at the end of the year, compared to its beginning. The data correspond to years 1997-2008.

21.7	21.1	25.5	-6.2	-6	-16.2	21	1.9	-0.6	16.3	7.2	-31.4

Find

(a) Mean (b) Standard deviation (Hint: $\sum x^2 = 3651.69$) (c) Median

(d) 25th percentile (or first quartile) (e) IQR (Interquartile range) (f) Range

Answer the following:

(g) What is the percentage of observations that are within 1 standard deviation of the mean?

a) $\bar{x} = \frac{\Sigma x}{n} = \frac{54.3}{12} = 4.525 \cong 4.53$

b) $s = \sqrt{s^2} = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{3651.69 - \frac{(54.3)^2}{12}}{11}} \cong \sqrt{309.63} \cong 17.59$

c) $-31.4, -16.2, -6.2, -6, -0.6, 1.9, 7.2, 16.3, 21, 21.1, 21.7, 25.5$

$\text{median} = \frac{1.9+7.2}{2} = 4.55$

d) $12(0.25) = 3$. So 1st quartile is the average of 3rd and 4th: $\frac{-6.2+(-6)}{2} = -6.1$

e) $IQR = Q_3 - Q_1$ $\quad Q_3 \rightarrow 12(0.75) = 9 \Rightarrow Q_3 = \frac{9^{th}+10^{th}}{2} = \frac{21+21.1}{2} = 21.05$

$Q_1 = -6.1$ by part (d)

$= 21.05 - (-6.1)$

$= 27.15$

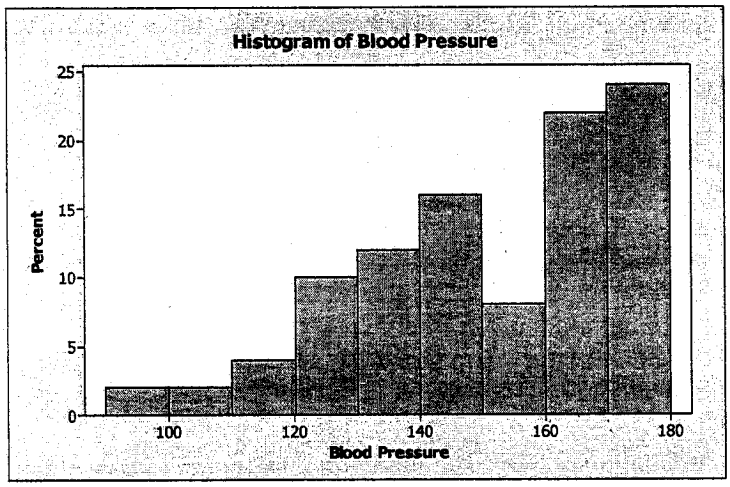
f) $\text{Range} = 25.5 - (-31.4) = 56.9$

g) $4.53 \mp 17.59 = [-13.06, 22.12]$

There are 9 observations between -13.06 and 22.12, so

$\frac{9}{12} = 75\%$

Question 2 (15 points): The following is a histogram of blood pressure measurements of a random sample of college students, ranging from 90 to 180 mmHg (millilitres of mercury).



(a) (2 pts) Which one is the modal class?

[170,180] contains the largest relative frequency.

(b) (3 pts) In which class is the 70th percentile of the blood pressure data?

The 70th percentile is in the class [160,170] because [170,180] contains about 23% of the observations and [160,170] contains more than 20%. Then top 30% is reached in class [160,170].

The mean of the data set is 150, the standard deviation is 21,

(c) (2 pts) Is the median likely to be smaller, about the same, or larger than the mean? Why?

Median is likely to be larger than the mean because the histogram is left-skewed.

(d) (3 pts) Apply Chebyshev's rule to estimate the percentage of observations between 108 and 192

$150 + 2 \times 21 = 192$ and $150 - 2 \times 21 = 108$.
So, according to Chebyshev's rule, at least $\frac{3}{4} = 75\%$ of the measurements lie in the interval [108,192]

(e) (3 pts) Apply empirical rule to estimate the percentage of observations between 108 and 192.

Approximately 95% of the measurements lie in the interval [108, 192] according to the empirical rule.

(f) (2 pts) Which estimate would be more appropriate for this data set, Chebyshev or empirical rule estimate? Why?

Chebyshev's rule would be more appropriate because the histogram is not bell-shaped, not even symmetrical. But empirical estimate is also good in this case.

Question 3: **(20 points)**

How many different linear arrangements (i.e., orderings) are there of the letters "A,B,C,D,E" for which

(a) A and B are next to each other?

(AB)CDE ⇒ treating AB as one, there are 4! orderings,)2
and for each of them AB & BA yield a different arrangement so $4! \times 2 = 48$)1

(b) A is the first and B is the last letter?

A 3 2 1 B → $3! = 6$

(c) E is not last in the arrangement?

Total # of arrangements = $5! = 120$)2
# of arrangement w/ E in last spot = $4! = 24$)2
" " " " " not in " " = $120 - 24 = 96$)1

(d) Skyler, a first grade student made an arrangement randomly using the above letters, "A,B,C,D,E". Let Q be the event that "A is the first and B is the last letter in his arrangement" and R be the event that "E is not last in his arrangement". Find the probabilities of these events.

$$P(Q) = \frac{6}{120} = \frac{1}{20}$$

$$P(R) = \frac{96}{120} = \frac{4}{5}$$

(e) Are events Q and R defined in part (d) mutually exclusive? Explain your answer.

$Q \cap R = R$ (since if A is first, B is last, we should have E not in last spot).

So $P(Q \cap R) = P(Q) = \frac{1}{20} \neq 0$

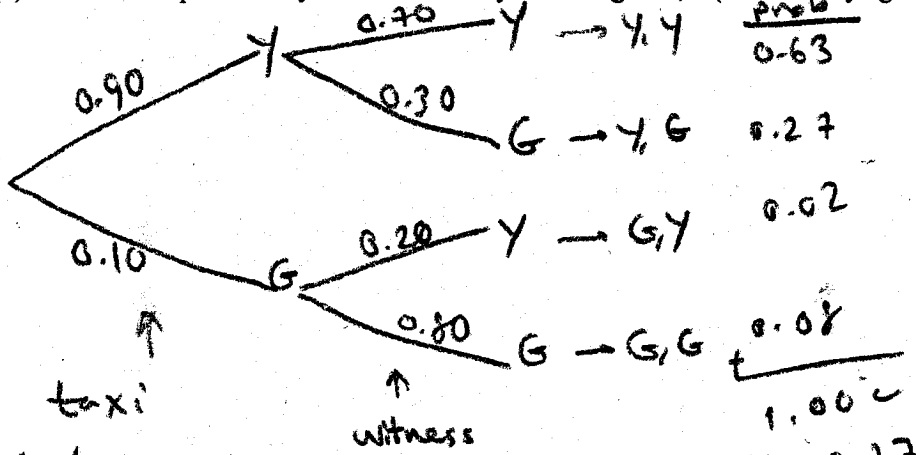
Hence Q and R are not mutually exclusive as prob. of their intersection is not equal to 0.

Question 4: (20 points)
In a small town, 90% of the taxis are yellow and 10% are green. One day a taxi is involved in a hit-and-run accident late at night. There is an eyewitness one block away, who testifies that the perpetrating (suç işleyen) taxi was green. The judge orders a test of the eyewitness' (görgü şahidi) abilities, conducted at a similar distance and a similar time of night. The results show that

P("witness says green" | "taxi is green") = 0.8;

P("witness says yellow" | "taxi is yellow") = 0.7.

(a) What is the probability that witness says taxi is green, P(witness says green)?



P(witness says green) = P(Y,G) + P(G,G) = 0.27 + 0.08 = 0.35

(b) Evaluate P("taxi is green" | "witness says green"). Based on this calculation, what is your deduction (çıkarım) about the reliability (güvenilirlik) of the witness?

P("taxi is green" | "witness says green") =

$= \frac{P(\text{taxi green and witness green})}{P(\text{witness green})} = \frac{P(G,G)}{P(\text{witness green})}$

$\frac{0.08}{0.35} = 0.2286$

(So reliability of the witness is very low)

(c) Let event A be "witness says taxi is green" and B be the event that "taxi is yellow". Are A and B independent?

$P(A \cap B) \stackrel{?}{=} P(A)\,P(B)$

$A \cap B$ = "witness green" AND "taxi is yellow"

$\Rightarrow P(A \cap B) = P(Y,G) = 0.27$

$P(A) = 0.35$

$P(B) = 0.90$

$\Rightarrow 0.27 \stackrel{?}{\neq} \underbrace{(0.35)(0.90)}_{0.315}$ so A & B are dependent

Question 5 (20 points): In a study of the relationship between age (X) and blood pressure (Y), data are obtained from 15 randomly selected people of older ages. The data points are shown below:

Age	43	56	48	61	67	70	40	45	59	55	72	53	56	50	65
Pressure	128	135	120	143	141	152	110	118	125	130	135	124	134	127	125

Use the following summary information:
$\sum x_i = 840$, $\sum y_i = 1947$, $\sum x_i^2 = 48384$, $\sum x_i y_i = 110140$, and $SS_{yy} = 1582.4$

(a) Find the equation of the regression line.

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} \quad \begin{matrix} \rightarrow SS_{xy} \\ \rightarrow SS_{xx} \end{matrix} = \frac{110140 - \frac{(840)(1947)}{15}}{48384 - \frac{(840)^2}{15}}$$

$$= \frac{1108}{1344} \cong 0.82$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{1947}{15} - (0.82)\frac{840}{15} \cong 83.88$$

$$\Rightarrow \boxed{\hat{y} = 83.88 + 0.82x}$$

(b) Interpret the slope and the intercept in the context of this problem.

Slope: If age increases by 1 year, blood pressure is expected to increase 0.82 units

Intercept: A newborn baby would have 83.88 blood pressure. This is an unreliable estimate, since $x=0$ is out of the range of the data set.

(c) Calculate the correlation coefficient.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}\, SS_{yy}}} = \frac{1108}{\sqrt{(1344)(1582.4)}} \cong 0.76$$

(d) Predict the blood pressure for a 55 year-old person.

$$\hat{y} = 83.88 + 0.82(55) = 128.98$$

(e) Interpret the relationship between age and blood pressure based on the correlation coefficient.

There is a moderate positive linear relationship between age and blood pressure. As age increases, blood pressure also increases.