Math 211: Statistics for Sciences

December 12, 2012

Fall 2012 Midterm #2

Closed book & notes; only a single-sided and handwritten A4 formula sheet and a calculator allowed; 100 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.**

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 and SURNAME: KEY

SIGNATURE: ____________________

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

Question 1 (20 Points):

Charles Dickens just arrived in Istanbul and lives quite a distance from his office. The time it takes him to drive to the office every morning is independent of the other days and normally distributed, with a mean of 53 minutes, and a standard deviation of 23 minutes. Jane Austen, who works in the same office, lives closer. The time it takes her to drive to the office every morning is also independent of the other days and normally distributed, with a mean of 21 minutes, and a standard deviation of 8 minutes. (Hint: Let X be the random variable representing the amount of time it takes Charles to drive to work, and let Y be a random variable representing the amount of time it takes Jane to drive to work. Assume that X and Y are independent).

$X \sim N(53, 23^2)$ and $Y \sim N(21, 8^2)$

(a) What is the probability that for the next three mornings, the total driving time for Charles is more than 200 minutes?

Let X_i be time for next day $i=1,2,3$. Then total driving time is $T = X_1 + X_2 + X_3 \sim N(3\times 53, 3\times 23^2) = N(159, 1587)$

so $P(T > 200) = P\left(Z > \frac{200-159}{\sqrt{1587}}\right) = P\left(Z > \frac{41}{39.8371}\right)$

$= P(Z > 1.03) = 1 - 0.8485 = 0.1515$

(b) What is the probability that on any given morning, the combined driving time for Jane and Charles exceeds 100 minutes?

Let $V = X+Y$, so we want $P(V > 100)$

$V \sim N(53+21, 23^2+8^2) = N(74, 593)$

so $P(V > 100) = P\left(Z > \frac{100-74}{\sqrt{593}}\right) = P\left(Z > \frac{26}{24.3516}\right)$

$= P(Z > 1.07) = 1 - 0.8577 = 0.1423$

(c) Charles leaves his home at 07:00, and Jane leaves her home at 07:30. What is the probability that Charles will arrive before Jane?

$P(X < Y+30) = P(X - Y < 30)$. let $U = X - Y$, then

$U \sim N(53-21, 23^2+8^2)$

$= N(32, 593)$

so $P(U < 30) = P\left(Z < \frac{30-32}{\sqrt{593}}\right) = P\left(Z < \frac{-2}{24.3516}\right)$

$= P(Z < -0.08) = 0.4681$

Question 2 (20 Points):

Babür Khan likes baklava. The probabilities for the number of plates (tabak) of baklava, denoted X, he eats every day are given below. (Assume that the amount of baklava he eats each day is independent of the other days.)

$E(X) = \mu_X = \sum x\,p(x) = 0(0.05) + 1(0.15) + 2(0.25) + 3(0.55) = 2.30$

x	0	1	2	3
$p(x)$	0.05	0.15	0.25	0.55

$E(X^2) = \sum x^2 p(x) = 0^2(0.05) + 1^2(0.15) + 2^2(0.25) + 3^2(0.55) = 6.1$

(a) Doctors tell him that if he eats more than 100 plates of baklava in the next month (31 days), he will get diabetes (şeker hastalığı). What is the probability that he will get diabetes by the end of next month?

Let X_i be the amount of baklava B.Khan eats on day i, so $Y = \sum_{i=1}^{31} X_i$ is the total amount of baklava he eats in the next month.

First observe that $P(X_i \leq 3) = 1$ and so $P(Y \leq 93) = 1$, $\Rightarrow$ $\underline{P(Y > 100) = 0}$

OR for $P(Y > 100)$, we can use normal approximation, since $n = 31 > 30$.

By CLT, $Y \overset{approx}{\sim} N(\mu_Y, \sigma_Y^2)$ where $\mu_Y = 31\mu_X = 31(2.30) = 71.3$

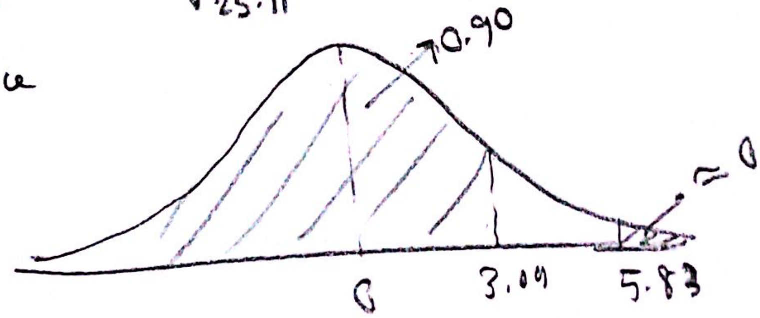
and $\sigma_Y^2 = 31(\sigma_X^2)$ where $\sigma_X^2 = E(X^2) - (\mu_X)^2 = 6.1 - (2.30)^2 = 0.81$

so $\sigma_Y^2 = 31(0.81) = 25.11$, so $Y \overset{approx}{\sim} N(71.3, 25.11)$

so $P(Y > 100) = P(Y \geq 100.5)$ (since Y is discrete, we need continuity correction.)

$\cong P\left(Z \geq \frac{100.5 - 71.3}{\sqrt{25.11}}\right) = P\left(Z \geq \frac{29.2}{5.011}\right) = P(Z \geq 5.83) \cong 0$

since



(b) His wife tells him that if he eats more than 70 plates of baklava next month (31 days), she will leave him. What is the probability that she leaves him at the end of the next month?

$P(Y > 70) = P(Y \geq 70.5) = P\left(Z \geq \frac{70.5 - 71.3}{\sqrt{25.11}}\right)$

$= P\left(Z \geq \frac{-0.80}{5.011}\right) = P(Z \geq -0.16) = 1 - 0.4364$

$= 0.5636$

Question 3 (20 Points):

(a) A random sample of 49 New York state employees (çalışanlar) is selected and the sample mean annual (yıllık) salary is \$21,543, with a standard deviation of \$3,000. Estimate with 95% confidence interval the true mean salary of all New York State employees.

$n = 49 > 30$ is large so by CLT, 95% CI for μ is

$\bar{x} = 21{,}543$

$s = 3000$

$\alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

$z_{0.025} = 1.96$

$$\bar{x} \mp z_{0.025}\left(\frac{s}{\sqrt{n}}\right)$$

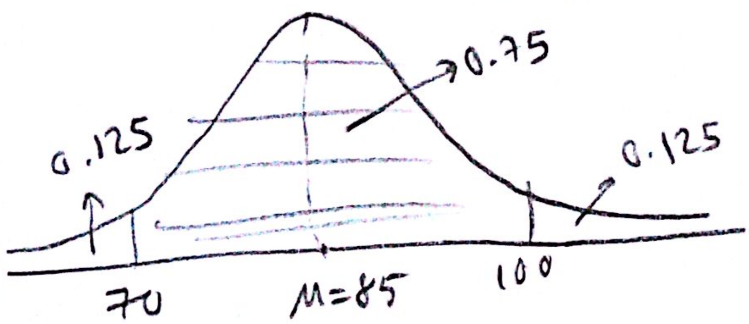
$$= 21{,}543 \mp (1.96)\left(\frac{3000}{\sqrt{49}}\right)$$

$$= 21{,}543 \mp 840$$

$$= (20703,\ 22383)$$

(b) It is estimated that 75% of all 20 year old-men have weights ranging from 70 and 100 kg. Assuming the weight distribution can be modeled with a normal curve and that 70 and 100 are equidistant (eş uzaklıkta) from the average weight, μ, of all 20 year old-men, calculate the population standard deviation, σ, of all 20 year old-men.

$X \sim N(\mu, \sigma^2)$ 70 & 100 equidistant from μ means that $\mu = \frac{70+100}{2} = 85$ kg



so $P(X < 100) = 0.75 + 0.125 = 0.875$

$$\Rightarrow P\left(Z < \frac{100-85}{\sigma}\right) = 0.875$$

$$\Rightarrow P\left(Z < \frac{15}{\sigma}\right) = 0.875 \quad \Rightarrow \frac{15}{\sigma} = 1.15$$

$$\Rightarrow \sigma = \frac{15}{1.15} = \underline{13.04}$$

Question 4 (20 Points):

The random variable X can have the values 0 or 1, and the random variable Y can have the values 0, 1, or 2.

(a) If the joint probability distribution of X and Y is given below, What is A and B if $E(X) = 0.4$? You must show your work!!

		Y		
		0	1	2
X	0	0.1	A	0.15
	1	0.2	0.1	B

(1) $\sum p(x,y) = 0.1 + A + 0.15 + 0.2 + 0.1 + B = 1$

$\Rightarrow A + B = 0.45$

and

x	$p(x)$
0	$A + 0.35$
1	$B + 0.30$

$\Rightarrow E(X) = 0(A + 0.35) + 1(B + 0.30) = 0.40$

$\Rightarrow B = 0.10$

& $A = 0.35$

Answer parts (b)-(e) according to the below joint pdf.

		Y		
		0	1	2
X	0	0.1	0.4	0.1
	1	0.2	0.1	0.1

(b) Find the marginal distributions of X and Y.

x	$p(x)$
0	0.6
1	0.4

y	0	1	2
$p(y)$	0.3	0.5	0.2

(c) Find Cov(X, Y).

$Cov(X,Y) = E(XY) - E(X)E(Y)$ where $E(X) = 0(0.6) + 1(0.4) = 0.4$

$E(Y) = 0(0.3) + 1(0.5) + 2(0.2) = 0.9$ and let $U = XY$, then

u	0	1	2
$p(u)$	0.80	0.1	0.1

so $E(U) = 0(0.80) + 1(0.10) + 2(0.10) = 0.3$

$\Rightarrow Cov(X,Y) = 0.3 - (0.4)(0.9) = \underline{-0.06}$

(d) Are X and Y independent? Explain!

No, X & Y are not independent, since $Cov(X,Y) \neq 0$.
(If they were indep., $Cov(X,Y)$ would have been zero).

(e) Find the variance of $2X + 3Y$, Var($2X + 3Y$).

$Var(2X+3Y) = 4Var(X) + 9Var(Y) + 12\,Cov(X,Y)$ where

$Var(X) = E(X^2) - (\mu_x)^2$ with $E(X^2) = 0^2(0.6) + 1^2(0.4) = 0.4$

$= 0.4 - (0.4)^2 = 0.24$

and

$Var(Y) = E(Y^2) - (\mu_y)^2$ with $E(Y^2) = 0^2(0.3) + 1^2(0.5) + 2^2(0.2) = 1.3$

$= 1.3 - (0.9)^2 = 0.49$

So $Var(2X+3Y) = 4(0.24) + 9(0.49) + 12(-0.06)$

$= \underline{4.65}$

Question 5 (20 Points):

Suppose a random sample of size n drawn from the uniform pdf, $f_X(x, \theta) = 1/\theta, 0 \le x \le \theta$.

(a) Let $\hat{\theta}_1 = 2\bar{X}$, where $\bar{X}$ is the sample mean. Is $\hat{\theta}_1$ an unbiased estimator for θ? If not, construct an unbiased estimator for θ based on $\hat{\theta}_1$.

$E(\hat{\theta}_1) = E(2\bar{X}) = 2E(\bar{X}) = 2\mu_X$ where

$\mu_X = E(X) = \int_0^\theta x\left(\frac{1}{\theta}\right)dx = \frac{1}{\theta}\left(\frac{x^2}{2}\right)\Big|_0^\theta = \frac{\theta^2}{2\theta} = \frac{\theta}{2}$

so $E(\hat{\theta}_1) = 2\left(\frac{\theta}{2}\right) = \theta \Rightarrow \hat{\theta}_1$ is unbiased for θ

(b) Let $\hat{\theta}_2 = X_{\max}$, where $X_{\max}$ is the largest order statistic. Is $\hat{\theta}_2$ an unbiased estimator for θ? If not, construct an unbiased estimator for θ based on $\hat{\theta}_2$.

To find $E(\hat{\theta}_2)$, we need pdf of X_{max}, $f_{X_{max}}(x) = n\,(F_X(x))^{n-1} f_X(x)$

where $F_X(x) = \int_0^x \frac{1}{\theta}dt = \frac{x}{\theta}$ so $f_{X_{max}}(x) = n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta} = \frac{n}{\theta^n}x^{n-1}$

for $0 \le x \le \theta$

so $E(\hat{\theta}_2) = E(X_{max}) = \int_0^\theta x\left(\frac{n}{\theta^n}x^{n-1}\right)dx = \frac{n}{\theta^n}\int_0^\theta x^n dx$

$= \frac{n}{\theta^n}\left(\frac{x^{n+1}}{n+1}\right)\Big|_0^\theta = \frac{n}{n+1}\frac{\theta^{n+1}}{\theta} = \frac{n}{n+1}\theta \Rightarrow \hat{\theta}_2$ is biased for θ

But $\hat{\theta}_3 = \frac{n+1}{n}\hat{\theta}_2$ is unbiased for θ, since $E(\hat{\theta}_3) = \frac{n+1}{n}E(\hat{\theta}_2)$

$= \left(\frac{n+1}{n}\right)\left(\frac{n}{n+1}\right)\theta = \theta.$

(c) Let $\hat{\theta}_4$ be another unbiased estimator with variance, $\text{Var}(\hat{\theta}_4) = \theta^2/(n^2 + n + 1)$. Which one of the estimators, $\hat{\theta}_1$ in part (a) or $\hat{\theta}_4$ is more efficient?

$Var(\hat{\theta}_1) = Var(2\bar{X}) = 4\,Var(\bar{X}) = 4\frac{\sigma_X^2}{n}$ where $\sigma_X^2 = E(X^2) - (\mu_X)^2$

so $E(X^2) = \int_0^\theta x^2\frac{1}{\theta}dx = \frac{x^3}{3\theta}\Big|_0^\theta = \frac{\theta^2}{3} \Rightarrow \sigma_X^2 = \frac{\theta^2}{3} - \left(\frac{\theta}{2}\right)^2 = \frac{\theta^2}{12}$

$$\left.\begin{aligned} &\text{so } Var(\hat{\theta}_1) = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n} \\ &\text{and } Var(\hat{\theta}_4) = \frac{\theta^2}{n^2+n+1} \end{aligned}\right\} \Rightarrow Var(\hat{\theta}_4) < Var(\hat{\theta}_1) \text{ since}$$

$\frac{\theta^2}{n^2+n+1} < \frac{\theta^2}{3n} \Leftrightarrow 3n < n^2+n+1 \Leftrightarrow (n-1)^2 \ge 0$

so $\hat{\theta}_4$ is more efficient.

(d) Let $Y = X^2$ where X is the uniform random variable defined above. Find the pdf of Y.

$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F_X(\sqrt{y}) = \frac{\sqrt{y}}{\theta}$

for $0 \le y \le \theta^2$

so $f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left(\frac{\sqrt{y}}{\theta}\right) = \frac{1}{2\theta\sqrt{y}}$, $0 \le y \le \theta^2$.