

Problem Set 7 – Answers

1)

- a. Use the rules of expected value to show that

$$\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$$

- b. Use part (a) along with the rules of variance and standard deviation to show that

$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

when “a” and “c” have the same sign.

- c. What happens if “a” and “c” have opposite signs?

SOLUTION

a. $\begin{aligned}\text{Cov}(aX + b, cY + d) &= E[(aX + b)(cY + d)] - E(aX + b) \cdot E(cY + d) \\ &= E[acXY + adX + bcY + bd] - (aE(X) + b)(cE(Y) + d) \\ &= acE(XY) - acE(X)E(Y) = ac\text{Cov}(X, Y)\end{aligned}$

b. $\begin{aligned}\text{Corr}(aX + b, cY + d) &= \frac{\text{Cov}(aX + b, cY + d)}{\sqrt{\text{Var}(aX + b)}\sqrt{\text{Var}(cY + d)}} = \frac{ac\text{Cov}(X, Y)}{|a| \cdot |c| \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}} \\ &= \text{Corr}(X, Y) \text{ when } a \text{ and } c \text{ have the same signs.}\end{aligned}$

c. When a and c differ in sign, $\text{Corr}(aX + b, cY + d) = -\text{Corr}(X, Y)$.

2)

A quality control plan calls for randomly selecting three items from the daily production (assumed large) of a certain machine and observing the number of defectives. However, the proportion p of defectives produced by the machine varies from day to day and is assumed to have a uniform distribution on the interval $(0, 1)$. For a randomly chosen day, find the unconditional probability that exactly two defectives are observed in the sample.

SOLUTION

Let $Y = \#$ of defectives in a random selection of three items. Conditioned on p , we have

$$P(Y = y | p) = \binom{3}{y} p^y (1-p)^{3-y}, y = 0, 1, 2, 3.$$

We are given that the proportion of defectives follows a uniform distribution on $(0, 1)$, so the unconditional probability that $Y = 2$ can be found by

$$\begin{aligned} P(Y = 2) &= \int_0^1 P(Y = 2, p) dp = \int_0^1 P(Y = 2 | p) f(p) dp \\ &= \int_0^1 3p^2 (1-p) dp = 3 \int_0^1 (p^2 - p^3) dp = 3 \left(\frac{p^3}{3} - \frac{p^4}{4} \right) \Big|_{p=0}^1 = 3 * \frac{1}{12} = \frac{1}{4} \end{aligned}$$

3)

The National Fire Incident Reporting Service stated that, among residential fires, 73% are in family homes, 20% are in apartments, and 7% are in other types of dwellings. If four residential fires are independently reported on a single day, what is the probability that two are in family homes, one is in an apartment, and one is in another type of dwelling?

SOLUTION

$Y_1 = \text{number of family home fires}$

$Y_2 = \text{number of apartment fires}$

$Y_3 = \text{number of fires in other types of dwellings}$

$n = 4, p_1 = 0.73, p_2 = 0.2, p_3 = 0.07$

$$P(Y_1 = 2, Y_2 = 1, Y_3 = 1) = \frac{4!}{2! * 1! * 1!} (0.73)^2 * (0.2) * (0.07) = 0.08953$$

4)

The weights of a population of mice fed on a certain diet since birth are assumed to be normally distributed with $\mu = 100$ and $\sigma = 20$ (measurement in grams). Suppose that a random sample of $n = 4$ mice is taken from this population. Find the probability that

exactly two weigh between 80 and 100 grams and exactly one weighs more than 100 grams.

SOLUTION

Let $Y_1 = \#$ of mice weighing between 80 and 100 grams, and let $Y_2 = \#$ weighing over 100 grams. Thus, with X having a normal distribution with $\mu = 100$ g. and $\sigma = 20$ g.,

$$p_1 = P(80 \leq X \leq 100) = P(-1 \leq Z \leq 0) = .3413$$

$$p_2 = P(X > 100) = P(Z > 0) = .5$$

$$P(Y_1 = 2, Y_2 = 1) = \frac{4!}{2!1!} (.3413)^2 (.5)^1 (.1587)^1 = .1109.$$

5)

A statistics class takes two exams X (Exam 1) and Y (Exam 2) where the scores follow a bivariate normal distribution with parameters:

- $\mu_X = 70$ and $\mu_Y = 60$ are the marginal means
- $\sigma_X = 10$ and $\sigma_Y = 15$ are the marginal standard deviations
- $\rho = 0.6$ is the correlation coefficient

Suppose we select a student at random. What is the probability that

- a. the student scores over 75 on Exam 2?
- b. the student scores over 75 on Exam 2, given that the student scored $X=80$ on Exam 1?
- c. the sum of his / her Exam 1 and Exam 2 scores is over 150?
- d. the student did better on Exam 1 than Exam 2?

SOLUTION

- a. the student scores over 75 on Exam 2?

Note that $Y \sim N(60, 15^2)$, so the probability that the student scores over 75 on Exam 2 is

$$\begin{aligned} P(Y > 75) &= P\left(Z > \frac{75 - 60}{15}\right) \\ &= P(Z > 1) \\ &= 0.1587 \end{aligned}$$

- b. the student scores over 75 on Exam 2, given that the student scored $X=80$ on Exam 1?

Note that $(Y|X = 80) \sim N(\mu_*, \sigma_*^2)$ where
 $\mu_* = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X) = 60 + (0.6)(15/10)(80 - 70) = 69$
 $\sigma_*^2 = \sigma_Y^2(1 - \rho^2) = 15^2(1 - 0.6^2) = 144$

If a student scored $X = 80$ on Exam 1, the probability that the student scores over 75 on Exam 2 is

$$\begin{aligned} P(Y > 75|X = 80) &= P\left(Z > \frac{75 - 69}{12}\right) \\ &= P(Z > 0.5) \\ &= 0.3085 \end{aligned}$$

- c. the sum of his / her Exam 1 and Exam 2 scores is over 150?

Note that $(X + Y) \sim N(\mu_*, \sigma_*^2)$ where
 $\mu_* = \mu_X + \mu_Y = 70 + 60 = 130$
 $\sigma_*^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y = 10^2 + 15^2 + 2(0.6)(10)(15) = 505$

The probability that the sum of Exam 1 and Exam 2 is above 150 is

$$\begin{aligned} P(X + Y > 150) &= P\left(Z > \frac{150 - 130}{\sqrt{505}}\right) \\ &= P(Z > 0.8899883) \\ &= 0.1867 \end{aligned}$$

- d. the student did better on Exam 1 than Exam 2?

Note that $(X - Y) \sim N(\mu_*, \sigma_*^2)$ where
 $\mu_* = \mu_X - \mu_Y = 70 - 60 = 10$
 $\sigma_*^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y = 10^2 + 15^2 - 2(0.6)(10)(15) = 145$

The probability that the student did better on Exam 1 than Exam 2 is

$$\begin{aligned} P(X > Y) &= P(X - Y > 0) \\ &= P\left(Z > \frac{0 - 10}{\sqrt{145}}\right) \\ &= P(Z > -0.83) = P(Z < 0.83) = 1 - P(Z > 0.83) = 1 - 0.2033 = 0.7967 \end{aligned}$$

- 6) Let X = height (inches) and Y = weight (lbs) for an American male. Suppose X and Y have a bivariate normal distribution, the mean and sd of heights are 70 in and 3 in, the mean and sd of weights are 170 lbs and 20 lbs, and $\rho = .9$.
- Determine the distribution of Y given $X = 68$, i.e., the weight distribution for 5'8" American males.
 - Determine the distribution of Y given $X = 70$, i.e., the weight distribution for 5'10" American males. In what ways is this distribution similar to that of part (a), and how are they different?
 - Calculate $P(Y < 180 | X = 72)$, the probability that a 6-foot-tall American male weighs less than 180 lb.

SOLUTION

As stated in the section, $Y|X=x$ is normal with mean $\mu_2 + \rho\sigma_2\left(\frac{x - \mu_1}{\sigma_1}\right)$ and variance $(1 - \rho^2)\sigma_2^2$.

- Substitute into the above expressions: mean = $170 + .9(20)\left(\frac{68 - 70}{3}\right) = 158$ lbs, variance = $(1 - .9^2)(20)^2 = 76$, sd = 8.72 lbs. That is, the weight distribution of 5'8" tall American males is $N(158, 8.72)$.

- Similarly, $x = 70$ returns mean = 170 lbs and sd = 8.72 lbs, so the weight distribution of 5'10" tall American males is $N(170, 8.72)$. These two conditional distributions are both normal and both have standard deviation equal to 8.72 lbs, but the average weight differs by height.

- Plug in $x = 72$ as above to get $Y|X=72 \sim N(182, 8.72)$. Thus, $P(Y < 180 | X = 72) = \Phi\left(\frac{180 - 182}{8.72}\right) = \Phi(-0.23) = .4090$.

- 7) For a few years, the SAT consisted of three components: writing, critical reading, and mathematics. Let W = SAT Writing score and X = SAT Critical Reading score for a randomly selected student. According to the College Board, in 2012 W had mean 488 and standard deviation 114, while X had mean 496 and standard deviation 114. Suppose X and W have a bivariate normal distribution with $\text{Corr}(X, W) = .5$.

- An English department plans to use $X + W$, a student's total score on the nonmath sections of the SAT, to help determine admission. Determine the distribution of $X + W$.
- Calculate $P(X + W > 1200)$.

- c. Suppose the English department wishes to admit only those students who score in the top 10% on this Critical Reading + Writing criterion. What combined score separates the top 10% of students from the rest?

SOLUTION

- a. Since X and W are bivariate normal, $X + W$ has a (univariate) normal distribution, with mean $E(X+W) = E(X) + E(W) = 496 + 488 = 984$ and variance given by $V(X+W) = V(X) + V(W) + 2 \text{Cov}(X, W) = V(X) + V(W) + 2 \text{SD}(X) \text{SD}(W) \text{Corr}(X, W) = 114^2 + 114^2 + 2(114)(114)(.5) = 38,988$. Equivalently, $\text{SD}(X+W) = \sqrt{38,988} = 197.45$. That is, $X+W \sim N(984, 197.45)$.
- b. $P(X+W > 1200) = 1 - \Phi\left(\frac{1200-984}{197.45}\right) = 1 - \Phi(1.09) = .1379$.
- c. We're looking for the 90th percentile of the $N(984, 197.45)$ distribution:
 $.9 = \Phi\left(\frac{x-984}{197.45}\right) \Rightarrow \frac{x-984}{197.45} = 1.28 \Rightarrow x = 984 + 1.28(197.45) = 1237$.
-

- 8) Measurements for the length and width of a rectangular plastic covers for CDs are rounded to nearest mm (so they are discrete). Let X denote the length and Y denote the width. The possible value of X are 129, 130, and 131 mm. The possible values of Y are 15 and 16 mm. We show the probability for each pair in the following table:

		y	
		15	16
x	129	0.12	0.08
	130	0.42	0.28
	131	0.06	0.04

- Find the conditional probability that CD cover has a length of 130mm given the width is 15mm.
- What is the conditional probability that X has a maximum width of 130mm given $Y = 16$?
- Find $E(X|Y = 15)$.
- Find variance $V(X|Y = 15)$ and $\sigma_{X|Y=15}$.

SOLUTION

- i) Find the conditional probability that CD cover has a length of 130mm given the width is 15mm.

$$P_{X|Y}(130|15) = \frac{p(130, 15)}{p_Y(15)} = \frac{0.42}{0.12 + 0.42 + 0.06} = 0.7 \quad (1)$$

- ii) What is the conditional probability that X has a maximum width of 130mm given $Y = 16$?

$$\begin{aligned} p(X \leq 130|Y = 16) &= \frac{p(129, 16)}{p_Y(16)} + \frac{p(130, 16)}{p_Y(16)} = \frac{0.08}{0.4} + \frac{0.28}{0.4} \\ &= 0.9 \end{aligned} \quad (2)$$

- iii) Find $E(X|Y = 15)$.

$$E(X|Y = 15) = \mu_{X|Y=15} = \sum_{x \in D_X} x \cdot p_{X|Y}(x|15) \quad (3)$$

$$\mu_{X|Y=15} = 129p_{X|Y}(129|15) + 130p_{X|Y}(130|15) + 131p_{X|Y}(131|15) \quad (4)$$

$$\mu_{X|Y=15} = 129 * \left(\frac{0.12}{0.6}\right) + 130 * \left(\frac{0.42}{0.6}\right) + 131 * \left(\frac{0.06}{0.6}\right) = 129.9 \quad (5)$$

- iv) Find variance $V(X|Y = 15)$ and $\sigma_{X|Y=15}$.

First, we need to calculate the expected value of X^2 given $Y = 15$, then we can use the shortcut formula

$$\begin{aligned} E(X^2|Y = 15) &= 129^2 p_{X|Y}(129|15) + 130^2 p_{X|Y}(130|15) \\ &\quad + 131^2 p_{X|Y}(131|15) = 16,874.3 \end{aligned} \quad (6)$$

$$V(X|Y = 15) = \sigma_{(X|Y = 15)}^2 = E(X^2|Y = 15) - \mu_{X|Y=15}^2 \quad (7)$$

$$\sigma_{(X|Y = 15)}^2 = 16,874.3 - 16,874.01 = 0.29 \quad (8)$$

$$\sigma_{(X|Y = 15)} = \sqrt{\sigma_{(X|Y = 15)}^2} = 0.539 \quad (9)$$