Problem set 7

A. Farris | Statistics 103 | UC Davis
Winter 2021

Instructions: Here is our seventh set of practice problems.

Practice problems

(1) Consider the following joint distribution of X and Y:

X\Y	0	1
0	0.20	0.60
1	0.05	0.15

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

(b) What is the conditional distribution of X given that Y equals 0?

(c) What is the conditional distribution of Y given that X equals 1?

(d) Are X and Y independent?

answer: (a) $p_X(0) = .8$, $p_X(1) = .2$; $p_Y(0) = .25$, $p_Y(1) = .75$; *(b)* $p(0|Y = 0) = .8$, $p(1|Y = 0) = .2$; *(c)* $p(0|X = 1) = .25$, $p(1|X = 1) = .75$; *(d) Yes.*

(2) For the joint distribution in Problem (2), find the correlation between X and Y.

answer: 0

(3) For the joint distribution

X\Y	-1	0	1
-1	0	0	.3
0	0	.4	0
1	.3	0	0

find the correlation.

answer: -1

(4) Suppose X and Y are the random variables with the joint PMF:

X\Y	1	2	3
1	0.1	0.3	0.1
2	0.2	0	0.1
3	0	0.1	0.1

(a) Compute $\mathrm{Cov}(X, Y)$.

(b) Compute $\mathrm{Var}(X)$

(c) Compute Var(Y)

(d) Compute Var($X+Y$).

answer: 0.1*;* 0.61*;* 0.6*,* 1.41

(5) In a large population of individuals, wage income has a mean of \$ 20,000 per year and a standard deviation of \$ 5,000 per year. Asset income has a mean of \$ 3,000 per year and a standard deviation of \$ 2,500 per year. The correlation between wage income and asset income is 0.4. Total income is the sum of wage and asset income. Find the mean and the standard deviation of total income for this population.

answer: \$ 23,000; \$ 6422.62

(6) Suppose that $E[X] = 2$, $E[Y] = 3$ and $\text{Cov}(X, Y) = 1$. Find $E[XY]$.

answer: 7

(7) Suppose that $SD(X) = 4$, $SD(Y) = 2$, and $SD(X + Y) = 5$. Do you have enough information to determine the correlation between X and Y? If so, what is it?

answer: Yes; $^5/_{16}$

(8) Suppose that X is a random variable with PDF given by

$$p_X(x) = \begin{cases} \frac{4}{3}x^{1/3} & \text{if } 0 < x < 1 \\ 0 & \text{if otherwise} \end{cases}$$

Which prediction would you use, to minimize an absolute error cost?

answer: The median, $0.5^{3/4}$

(9) A probability model for a price specifies a random variable X with CDF

$$F_X(x) = \frac{x^3}{1000}$$

for $0 \leq x \leq 10$, mean 7.5, and variance 3.75. What is the best prediction for X under squared error cost? Under absolute error cost?

answer: 7.5; $500^{1/3} \approx 7.937$

(10) Suppose that a probability model for the number of transactions X that will be completed by a firm in a quarter is characterized as first randomly selecting a random value of Q, and then, given this, sampling a random variable from a Binomial$(200, p)$ distribution where $p = Q$. If you observe $Q = 0.25$, what would your prediction (using squared error cost) be for X?

answer: 50

(11) Suppose X and Y are the random variables with the joint PMF:

X\Y	-1	0	1
-1	0.1	0.2	0
0	0.2	0	0.1
1	0.1	0.1	0.2

What would your prediction be for Y if you were to observe $X = -1$, $X = 0$, or $X = 1$? (Use squared error cost)

answer: respectively: $-1/3$, $-1/3$, $1/4$

(12) Suppose X and Y are the random variables with the joint PMF:

X\Y	-1	0	1
-1	0.1	0.2	0
0	0.2	0	0.1
1	0.1	a	b

Suppose that your competitor, using the probability model above, predicts that the value of Y will be 0 given that $X = 1$. What probabilities do they assign to a and to b? (Assume that they are minimizing squared error cost)

answer: respectively: 0.2, 0.1

(13) Suppose X and Y are the random variables with the joint PMF:

X\Y	1	2	3
6	0.5	0.1	0.1
12	0.1	0.1	0.1

What would your prediction be for X if you were to observe $Y = 1$, $Y = 2$, or $X = 3$? (Use squared error cost)

answer: respectively: 7, 9, 9

(14) Suppose that $X_1, X_2, X_3 \dots$ are IID with mean μ and SD σ. What is the variance of $\overline{X}_n - \mu$? Is this large for large n, small for large n, or neither?

answer: σ^2/n*; this gets small for large* n.

(15) (Approximately) what are the probabilities that $\overline{X}_{50} \leq 9.5$ and that $\overline{X}_{100} \leq 79.5$, if $X_1, X_2, X_3 \dots$ are independent and identically distributed with mean 80 and SD 10?

answer: respectively, ≈ 0.3618 *and* ≈ 0.3085

(16) You sample 40 random variables, independently, from the same probability distribution. If the probability distribution has mean 1200 and SD 50, with approximately

what probability will your sample mean be within 10 of 1200? What would this be if you sampled 60 of them instead?

answer: ≈ 0.7941; ≈ 0.8787