Lecture 2

Chapters R.2-R.4

R.2 Discrete Random Variables and Expectations

Random variable – any variable whose value cannot be predicted exactly.

Discrete random variable – one that has a specific set of possible values.

Continuous random variable – one that can take on a continuous range of values.

Exercise 1: Give 3 examples of discrete random variables and 3 examples of continuous random variables.

Discrete
1. roll of dice
2. pick a card from a deck
3. flip a coin

Continuous
1. real GDP
2. Height
3. Tempurature

Experiment – any process of observation or measurement.

Outcomes – the mutually exclusive different ways that an experiment can turn out.

Value – the result of an experimental outcome.

Population – the set of all unique values for the random variable.

Probability – the chance with which a specific outcome occurs.

Exercise 2: suppose you are playing a game of monopoly. You roll two dice, and the sum tells you how far to move. Fill in the table of possible outcomes.

Dice1

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Dice2

← outcome

values

outcome

values

total outcomes 36

Exercise 3: From the table of outcomes, fill each value, its frequency, and its probability.

Value of X:	2	3	4	5	6	7	8	9	10	11	12
Frequency:	1	2	3	4	5	6	5	4	3	2	1
Probability:	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Expected value of a discrete random variable – the weighted average of all its possible value.

Formula – Expected value of a discrete random variable.

$$E(X) = p_1 x_1 + p_2 x_2 + \dots + p_n x_n = \sum_{i=1}^{n} p_i x_i$$

Exercise 4: From the dice experiment above, what is the expected value?

$$E(X) = \frac{1}{36}(2) + \frac{2}{36}(3) + \frac{3}{36}(4) + \frac{4}{36}(5) + \frac{5}{36}(6) + \frac{6}{36}(7)$$

$$+ \frac{5}{36}(8) + \frac{4}{36}(9) + \frac{3}{36}(10) + \frac{2}{36}(11) + \frac{1}{36}(12)$$

$$E(X) = \frac{2}{36} + \frac{6}{36} + \frac{12}{36} + \frac{20}{36} + \frac{30}{36} + \frac{42}{36} + \frac{40}{36} + \frac{36}{36} + \frac{30}{36} + \frac{22}{36} + \frac{12}{36}$$

$$E(X) = \frac{257}{36}$$

$$\boxed{E(X) = 7}$$

also called population mean
use notation μ_X

The expected value of a function of a discrete random variable we simply calculate the function, then the expected value. $E\{g(X)\} = g(x_1)p_1 + \dots + g(x_n)p_n = \sum_{i=1}^{n} g(x_i)p_i$

Exercise 5: What is the expected value of X^2 from the dice throw?

$$E(X^2) = 2^2\left(\frac{1}{36}\right) + 3^2\left(\frac{2}{36}\right) + 4^2\left(\frac{3}{36}\right) + 5^2\left(\frac{4}{36}\right) + 6^2\left(\frac{5}{36}\right) + 7^2\left(\frac{6}{36}\right)$$

$$+ 8^2\left(\frac{5}{36}\right) + 9^2\left(\frac{4}{36}\right) + 10^2\left(\frac{3}{36}\right) + 11^2\left(\frac{2}{36}\right) + 12^2\left(\frac{1}{36}\right)$$

$$E(X^2) = 4\left(\frac{1}{36}\right) + 9\left(\frac{2}{36}\right) + 16\left(\frac{3}{36}\right) + 25\left(\frac{4}{36}\right) + 36\left(\frac{5}{36}\right) + 49\left(\frac{6}{36}\right)$$
$$+ 64\left(\frac{5}{36}\right) + 81\left(\frac{4}{36}\right) + 100\left(\frac{3}{36}\right) + 121\left(\frac{2}{36}\right) + 144\left(\frac{1}{36}\right)$$

$$E(X^2) = 0.11 + 0.50 + 1.33 + 2.78 + 5.00 + 8.17 + 8.89$$
$$+ 9.00 + 8.83 + 6.72 + 4.00$$

$$\boxed{E(X^2) = 54.83}$$

Rules of Expected Value:

1. $E(X + Y) = E(X) + E(Y)$
2. $E(bX) = bE(X)$, where b is a constant.
3. $E(b) = b$, where b is a constant.

Exercise 6: Suppose Y = 2+3X. Let X be the sum of dice from our original example (Exercise 2 and 3). What is the expected value of Y?

$E(Y) = E(2+3X)$ Use Rule 1

$E(Y) = E(2) + E(3X)$ Use Rule 2

$= E(2) + 3E(X)$ Use Rule 3

$= 2 + 3E(X)$ Plug in $E(X) = 7$

$= 2 + 3(7) = 2 + 21$

$$\boxed{E(Y) = 23}$$

Variance – the measure of distance of dispersion of a random variable's probability distribution.

Formula – Population variance.

$$Var(X) = \sigma_X^2 = E\{(X - \mu_X)^2\}$$
$$= (x_1 - \mu_X)^2 P_1 + (x_2 - \mu_X)^2 P_2 + \ldots + (x_n - \mu_X)^2 P_n$$
$$= \sum_{i=1}^{n} (x_i - \boxed{\mu_X})^2$$

this is a constant so no i

Exercise 7: What is the population variance of the dice throw from the previous exercises? $\mu_X = 7$

$$Var(X) = (2-7)^2 \cdot \frac{1}{36} + (3-7)^2 \cdot \frac{2}{36} + (4-7)^2 \cdot \frac{3}{36} + (5-7)^2 \cdot \frac{4}{36}$$

$$+ (6-7)^2 \cdot \frac{5}{36} + (7-7)^2 \cdot \frac{6}{36} + (8-7)^2 \cdot \frac{5}{36} + (9-7)^2 \cdot \frac{4}{36}$$

$$+ (10-7)^2 \cdot \frac{3}{36} + (11-7)^2 \cdot \frac{2}{36} + (12-7)^2 \cdot \frac{1}{36}$$

$$Var(X) = (-5)^2 \cdot \frac{1}{36} + (-4)^2 \cdot \frac{2}{36} + (-3)^2 \cdot \frac{3}{36} + (-2)^2 \cdot \frac{4}{36}$$

$$+ (-1)^2 \cdot \frac{5}{36} + (0)^2 \cdot \frac{6}{36} + (1)^2 \cdot \frac{5}{36} + (2)^2 \cdot \frac{4}{36}$$

$$+ (3)^2 \cdot \frac{3}{36} + (4)^2 \cdot \frac{2}{36} + (5)^2 \cdot \frac{1}{36}$$

$$Var(X) = 25 \cdot \frac{1}{36} + 16 \cdot \frac{2}{36} + 9 \cdot \frac{3}{36} + 4 \cdot \frac{4}{36} + 1 \cdot \frac{5}{36} + 0 \cdot \frac{6}{36}$$

$$+ 1 \cdot \frac{5}{36} + 4 \cdot \frac{4}{36} + 9 \cdot \frac{3}{36} + 16 \cdot \frac{2}{36} + 25 \cdot \frac{1}{36}$$

$$Var(X) = 0.69 + 0.89 + 0.75 + 0.44 + 0.14 + 0 + 0.14 + 0.44$$

$$+ 0.75 + 0.89 + 0.69$$

$$\boxed{Var(X) = 5.83}$$

Formula – population variance

$$Var(X) = E(X^2) - \mu_X^2$$

Exercise 8: Use the above formula to re-solve for the population variance of the dice throw.

$E(X^2) = 54.83$ $\qquad$ $\mu_X = 7$

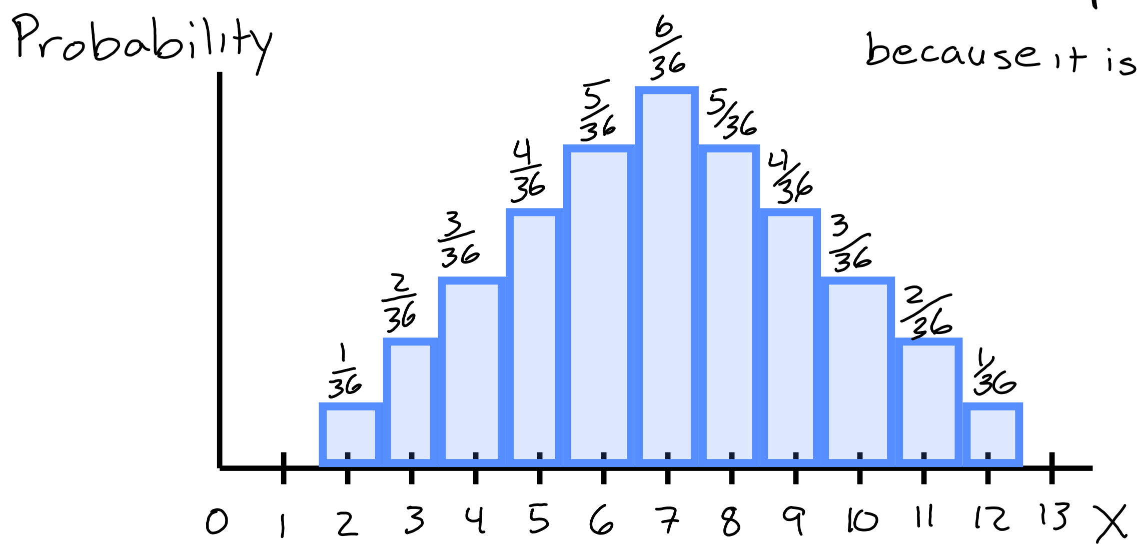
$$Var(X) = 54.83 - 7^2$$

$$Var(X) = 54.83 - 49$$

$$\boxed{Var(X) = 5.83}$$

R.3 Continuous Random Variables

Exercise 9: Draw a histogram of the probabilities from the dice throw.



Each value of X has a distinct probability because it is discrete

Exercise 10: Suppose we want to guess what the temperature in a classroom will be tomorrow. If the classroom is always between 55 and 75 with equal probability, what is the probability that it will be exactly 69?

Since the variable is continuous there is an infinite range of outcomes. 55, 55.00001, 55.000...02, ...

$$Pr(X=69) = \frac{1}{\infty} = 0$$

we can only do probability of CRV for a range of values

Exercise 11: What is the probability of it being between 59 and 60?

There are 20, 1° ranges from 55–75. So 59 to 60 is 1, 1° range.

$$\boxed{Pr(59 \le X \le 60) = \frac{1}{20} = 0.05}$$

Exercise 12: What is the probability of it being between 65 and 70?

$$\boxed{Pr(65 \le X \le 70) = \frac{5}{20} = 0.25}$$

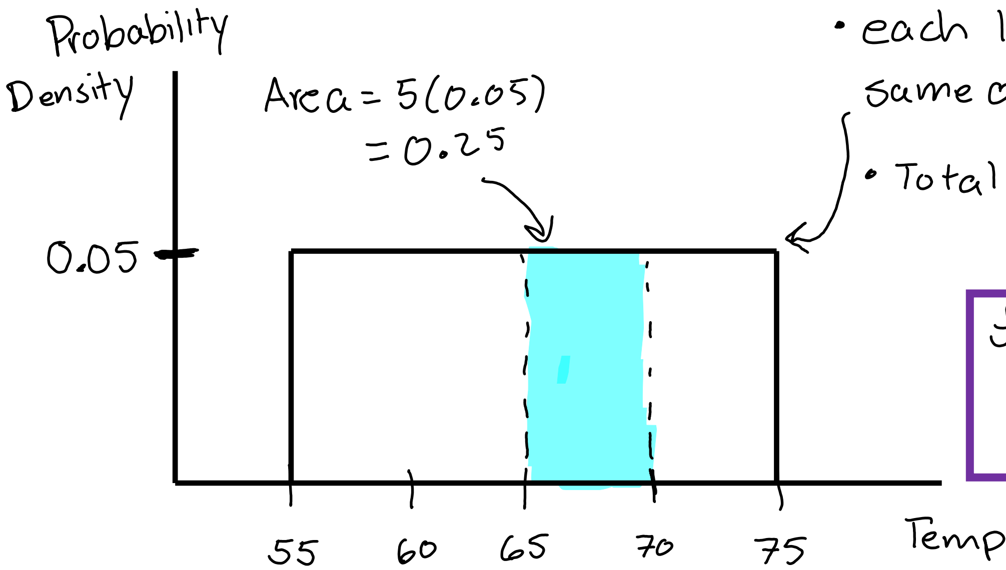
Probability density – The height at any point on the graph of probabilities.

Probability density function – A function that represents the probability of a continuous random variable, X/

Rules for probability density functions:

1. The probability of any point must be between 0 and 1.
2. The area of the probability density function must equal 1.

Exercise 13: Draw the probability density and write the probability density function for exercise 12.



- each 1 degree range has the same 0.05 probability
- Total area = 0.05(20) = 1

$$\boxed{\begin{array}{l} f(X) = 0.05 \text{ if } 55 \le X \le 75 \\ f(X) = 0 \text{ if } X < 55 \text{ or } X > 75 \end{array}}$$

R.4 Population Covariance and Correlation

Population covariance – expected value of the product of the deviation of two random variables from their means.

Formula – Population covariance:

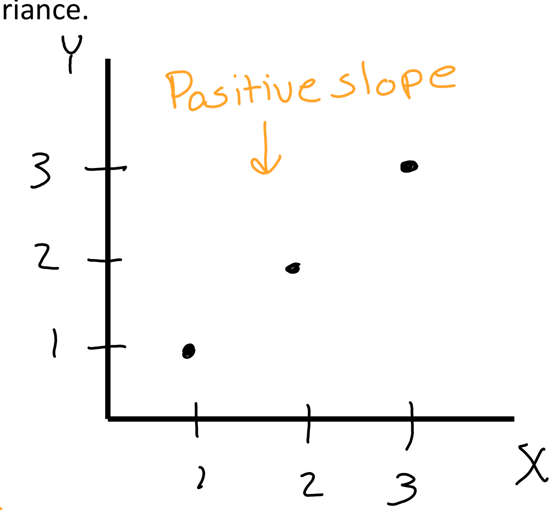
$$\operatorname{cov}(X,Y) = \sigma_{XY} = E\{(X-\mu_X)(Y-\mu_Y)\} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_X)(Y_i-\mu_Y)$$

Exercise 14: Plot the points of the random variables and find their covariance.

X	Y
1	1
2	2
3	3

$$\mu_X = \frac{1}{3}[1+2+3] = \frac{1}{3}[6] = 2$$

$$\mu_X = \frac{1}{3}[1+2+3] = 2$$



$$\sigma_{X,Y} = \frac{1}{3}[(1-2)(1-2)+(2-2)(2-2)+(3-2)(3-2)]$$

$$= \frac{1}{3}[(-1)(-1)+(0)(0)+(1)(1)]$$

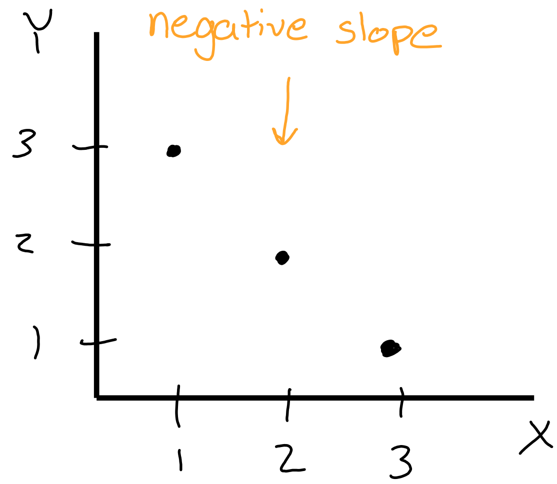
$$= \frac{1}{3}[1+0+1] = \frac{1}{3}\cdot 2 = \boxed{\frac{2}{3}}$$

positive

Exercise 15: Plot the following points and find the covariance.

X	Y
1	3
2	2
3	1

$\mu_X = \mu_Y = 2$



$$\sigma_{X,Y} = \frac{1}{3}\left[(1-2)(3-2)+(2-2)(2-2)+(3-2)(1-2)\right]$$

$$= \frac{1}{3}\left[(-1)(1)+(0)(0)+(1)(-1)\right]$$

$$= \frac{1}{3}\left[-1+0+-1\right] = \frac{1}{3}(-2) = \boxed{\frac{-2}{3}}$$

← negative

Exercise 16: Find the covariance of the following points.

X	Y
1	300
2	200
3	100

$$\mu_Y = \frac{1}{3}\left[300+200+100\right] = \frac{1}{3}\left[600\right] = 200$$

$$\sigma_{X,Y} = \frac{1}{3}\left[(1-2)(300-200)+(2-2)(200-200)+(3-2)(100-200)\right]$$

$$\sigma_{XY} = \frac{1}{3}\left[(-1)(100)+(0)(0)+(1)(-100)\right]$$

$$= \frac{1}{3}\left[-100+0+(-100)\right] = \frac{1}{3}\left[-200\right] = \boxed{\frac{-200}{3}}$$

Y scaled by 100 and so did covariance

Correlation coefficient – a unitless measure of the association of two variables.

Formula – correlation coefficient $\rho_{XY} = \sigma_{XY} / \sqrt{\sigma_X^2 \; \sigma_Y^2}$

Exercise 17: Find the correlation coefficient from exercise 15.

$$\sigma_X^2 = \frac{1}{3}\left[(1-2)^2+(2-2)^2+(3-2)^2\right] = \frac{1}{3}\left[(-1)^2+0^2+1^2\right] = \frac{1}{3}\left[1+0+1\right] = \frac{2}{3}$$

$$\sigma_Y^2 = \frac{1}{3}\left[(3-2)^2+(2-2)^2+(1-2)^2\right] = \frac{1}{3}\left[1^2+0^2+(-1)^2\right] = \frac{2}{3}$$

$$\rho_{X,Y} = \frac{-\frac{2}{3}}{\sqrt{\frac{2}{3}\cdot\frac{2}{3}}} = \frac{-\frac{2}{3}}{\sqrt{\frac{2^2}{3^2}}} = \frac{-\frac{2}{3}}{\frac{2}{3}} = \boxed{-1}$$

Exercise 18: Find the correlation coefficient from exercise 16.

$$\sigma_Y^2 = \frac{1}{3}\left[(100-200)^2 + (200-200)^2 + (300-200)^2\right] = \frac{1}{3}\left[(-100)^2 + 0 + 100^2\right]$$

$$= \frac{1}{3}\left[10{,}000 + 10{,}000\right] = \frac{20{,}000}{3}$$

$$\rho_{X,Y} = \frac{-\frac{200}{3}}{\sqrt{\frac{2}{3} \cdot \frac{20{,}000}{3}}} = \frac{-\frac{200}{3}}{\sqrt{\frac{40{,}000}{9}}} = \frac{-\frac{200}{3}}{\frac{200}{3}} = \boxed{-1}$$

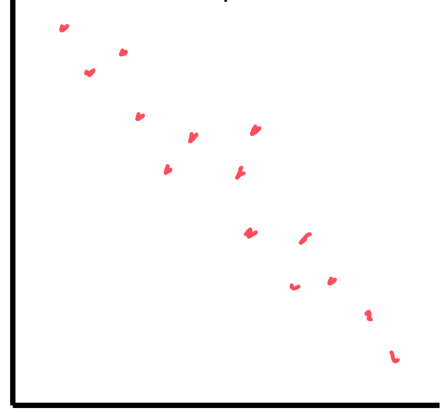
even though covariance scaled by 100, correlation stays the same.

Exercise 19: Label the following plots with their correlation coefficient.

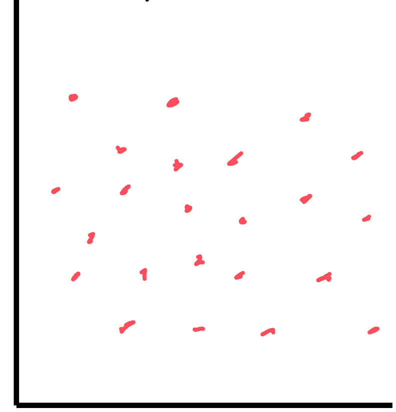
$(\rho = -1,\ -1 < \rho < 0,\ \rho = 0,\ 0 < \rho < 1,\ \rho = 1)$

$\rho = 1$

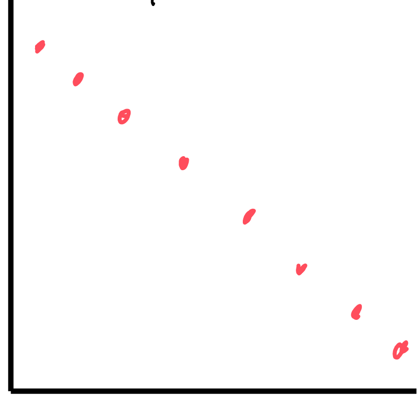
$-1 < \rho < 0$



$\rho = 0$



$\rho = -1$



$0 < \rho < 1$