Recitation 2

Exercise 1

Consider the data

x	y
3	8
4	13
5	12
6	14
7	16

1. Sketch a scatterplot.
   Answer

2. Would you expect a positive association, a negative association or no association between x and y?
   Answer
   Looking at the points we expect to have a positive association between x and y. Indeed they could be interpolated by a positive line.

3. Compute the correlation coefficient, r.
   Answer
   $\bar{x} = 5$, $\bar{y} = 12.6$

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x}) \times (y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
3	8	-2	-4.6	9.2	4	21.16
4	13	-1	0.4	-0.4	1	0.16
5	12	0	-0.6	0	0	0.36
6	14	1	1.4	1.4	1	1.96
7	16	2	3.4	6.8	4	11.56
				17	**10**	**35.2**

It follows that $r = 17/\sqrt{(10 \times 35.2)} = 0.91$.

Exercise 2

An instructor of Statistics collected data from one of her classes in Spring 2016 to investigate the relationship between Study time per week (number of hours) to predict the final grade. For the 8 students in her class the data were as shown in the table.

Student	Study Time	Grade
1	14	26
2	25	30
3	15	20
4	5	18
5	10	23
6	12	25
7	5	21
8	21	28

1. Identify the response variable and the explanatory variable.
   Answer
   Response variable: Grade. Explanatory variable: Study Time.

2. Construct a scatterplot.
   Answer

3. Find and interpret the correlation coefficient.
   Answer
   $\bar{x} = 13.375$, $\bar{y} = 23.875$

Student	Study Time	Grade	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x}) \times (y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
1	14	26	0.62	2.12	1.31	0.38	4.49
2	25	30	11.62	6.12	71.11	135.02	37.45
3	15	20	1.62	-3.88	-6.29	2.62	15.05
4	5	18	-8.38	-5.88	49.27	70.22	34.57
5	10	23	-3.38	-0.88	2.97	11.42	0.77
6	12	25	-1.38	1.12	-1.55	1.90	1.25
7	5	21	-8.38	-2.88	24.13	70.22	8.29
8	21	28	7.62	4.12	31.39	58.06	16.97
					172.34	**349.84**	**118.84**

It follows that $r = 172.34/\sqrt{(349.84 \times 118.84)} = 0.8452$. It means that Study Time and Grade are highly positive correlated. When the Study Time increases, the Grade increases too.

4. If all grades increase by one, does the correlation coefficient change?
   Answer It will not change.

5. Verify your answer in part 4.

Exercise 3

You flip a coin three times.

1. Use a tree diagram to show the possible outcome patterns. How many outcomes are in the sample space?
   Answer

First Coin	Second Coin	Third Coin
T	T	T
T	T	H
T	H	T
T	H	H
H	T	T
H	T	H
H	H	T
H	H	H

   There are $2^3 = 8$ outcomes.

2. Using the sample space constructed in part 1, find the probability (i) to have at least 2 heads; (ii) to have at least one tail.
   Answer
   (i)$4/8 = 0.5$; (ii)$7/8$

Exercise 4

A die is rolled.

1. List the possible outcomes in the sample space. **Answer**
   $S = 1, 2, 3, 4, 5, 6$.

2. What is the probabilty of getting a number which is even? **Answer**
   $3/6 = 0.5$.

3. What is the probabilty of getting a number which is greater than 4?
   Answer
   $2/6 = 1/3 = 0.333$

4. What is the probabilty of getting a number which is less than 3? What is its complement?
   Answer
   $2/6 = 1/3 = 0.333$. Its complement is given by 1-0.333=0.667.

Exercise 5

Two dice are rolled.

1. Construct the sample space. How many outcomes are there?
   Answer

	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

   There are $6^2 = 36$ possible outcomes.

2. Find the probability of rolling a sum of 7.
   Answer
   $6/36 = 1/6 = 0.166667$

3. Find the probability of getting a total of at least 10.
   Answer
   $6/36 = 1/6 = 0.166667$

4. Find the probability of getting a odd number as the sum. **Answer**
   $18/36 = 0.5$

Exercise 6

Based on records of automobile accidents, it has been reported the counts who survived (S) and died (D) according to whether they wore a seat belt (Y=yes, N=no). The data are presented in the following contingency table.

Wore Seat Belt	**Survived (S)**	**Died (D)**
Yes (Y)	410	5
No(N)	160	15

1. What is the sample space of possible outcomes for a randomly selected individual involved in an auto accident?
   Answer
   $18/36 = 0.5$

2. Compute (i) P(D), (ii) P(N).
   Answer
   (i) P(D)=20/590=0.03; (ii) P(N)=175/590=0.30

3. Compute the probability that an individual did not wear a seat belt and survived.
   Answer
   $P(N \cap S)$=160/590=0.27

4. Based on part 1, what would the answer to part 3 have been if the events N and S were independent? So, are N and S independent, and if not, what does that mean in the context of these data?
   Answer
   If events N and S were independent, then $P(N \cap S)$=P(N)$\times$ P(S)=0.3 $\times$ 570/590=0.3 $\times$ 0.97=0.291. So N and S are not independent. This indicates that chance of surviving depends on seat belt use since 0.291 is not equal to 0.27.

5. Compute the probability that the individual survived, given that the person (i) wore and (ii) did not wear a seat belt. Interpret the results.
   Answer
   $P(S \mid Y) = P(S \cap Y)/P(Y) = (410/590)/(415/590) = 0.99$
   $P(S \mid N) = P(S \cap N)/P(N) = (160/590)/(175/590) = 0.91$

   Once again, it means that the events are not independent. Wearing or not the seat belt influences the chance of surviving.

6. Are the events of dying and wearing a seat belt independent? Justify your answer.
   Answer
   They are not independent since neither $P(S \mid Y)$ nor $P(S \mid N)$ equals $P(S)$. Specifically 0.99 and 0.91 are different from 0.97.

Exercise 7

It has been asked to 1000 people whether they were happy in their marriages.The poll reported that 46% were men, while 54% were women. Among men, 40% declared to be very happy, 51% pretty happy, while 9% not too happy. Among women, 43% declared to be very happy, 49% pretty happy, while 8% not too happy.

1. Set up a contingency table that cross classifies gender by level of happiness.
   Answer

Gender	Very Happy	Pretty Happy	Not Too Happy	Total
Male	184	235	41	**460**
Female	232	265	43	**540**

2. Compute the probability that a married adult is very happy.
   Answer
   P(VH)=(184+232)/1000=0.42

3. Compute the probability that a married adult is very happy, (i) given that their gender is male and (ii) given that their gender is female.
   Answer
   $P(VH \mid M) = P(VH \cap M)/P(M) = (184/1000)/(460/1000) = 0.4$

   $P(VH \mid F) = P(VH \cap F)/P(F) = (232/1000)/(540/1000) = 0.43$

4. For these subjects, are the events being very happy and being a male independent?
   Answer
   No, since neither $P(VH \mid M) = 0.40$ nor $P(VH \mid F) = 0.43$ are equal to P(VH)=0.42.

Exercise 8

A wheat farmer living in Pennsylvania finds that his annual profit is \$ 80 if the summer weather is typical, \$ 50 if the weather is unusually dry, and \$ 20 if there is a severe storm that destroys much of his crop. Weather bureau records indicate that the probability is 0.70 of typical weather, 0.20 of unusually dry weather, and p of a severe storm. Let X be the farmer's profit.

1. Construct the probability distribution of X and find p.
   Answer
   $p = 1 - 0.70 - 0.20 = 0.10$.

x_i	$P(x_i)$
80	0.70
50	0.20
20	0.10

2. Find the mean of the probability distribution of X.
   Answer
   $E(X) = 80 \times 0.70 + 50 \times 0.20 + 20 \times 0.10 = 68$

3. Find the variance of the probability distribution of X.
   Answer

x_i	$P(x_i)$	$x_i - E(X)$	$p(x_i) \times (x_i - E(X))^2$
80	0.70	12	100.8
50	0.20	-18	64.8
20	0.10	-48	230.4

The variance is 396.

Exercise 9

In a population the vehicle speed distribution is well approximated by a Normal curve with mean 50 and standard deviation 15. Compute the probability that

1. a randomly selected vehicle speed is greater than 73;
   Answer
   $P(X > 73) = P(\frac{X-50}{15} > \frac{73-50}{15}) = P(Z > 1.53) = P(Z < -1.53) = \Phi(-1.53) = 0.063$

2. a randomly selected vehicle speed is between 40 and 73;
   Answer
   $P(40 < X < 73) = P(\frac{40-50}{15} < \frac{X-50}{15} < \frac{85-50}{15}) = P(-0.67 < Z < 1.53) = \Phi(1.53) - \Phi(-0.67) = 0.9370 - 0.2525 = 0.6845$
   $\Phi(1.53) = 1 - \Phi(-1.53) = 1 - 0.063 = 0.9370$
   $\Phi(-0.67) = 1 - \Phi(0.67) = 0.2525$

3. a randomly selected vehicle speed is less then 85.
   Answer
   $P(X < 85) = P(\frac{X-50}{15} < \frac{85-50}{15}) = P(Z < 2.33) = 1 - \Phi(-2.33) = 1 - 0.0099 = 0.9901$

Exercise 10

At one private college, a minimum IQ score of 125 is necessary to be considered for admission. IQ scores can be well approximated with a Normal distribution with mean of 100 and standard deviation of 15.

1. Compute the probability of being admitted.
   Answer
   $P(X > 125) = P(\frac{X-100}{15} > \frac{125-100}{15}) = P(Z > 1.67) = P(Z < -1.67) = \Phi(-1.67) = 0.0475$

2. Compute the probability that a randomly selected IQ score is between 120 and 145.
   Answer

$P(120 < X < 145) = P(\frac{120-100}{15} < \frac{X-100}{15} < \frac{145-100}{15}) = P(1.33 < Z < 3) = \Phi(3) - \Phi(1.33) = 0.9987 - 0.9082 = 0.0905$

3. Compute the probability that a randomly selected IQ score is less than 125.
   Answer
   $P(X < 125) = P(\frac{X-100}{15} < \frac{125-100}{15}) = P(Z < 1.67) = 0.9525$

4. Compute the probability that a randomly selected IQ score is less than 90.
   Answer
   $P(X < 90) = P(\frac{X-100}{15} < \frac{90-100}{15}) = P(Z < -0.67) = 0.2514$

Esercizio 11

Let the random variable Z follow a standard Normal distribution. Compute the probabilities below:

1. P(-2 $<$ Z $<$ -1)
   Answer
   $\Phi(-1)$=1-Φ(1)=1-0.8413=0.1587
   $\Phi(-2)$=1-Φ(2)=1-0.9772=0.0228
   P(-2 $<$ Z $<$ -1)=Φ(-1)-Φ(-2)=0.1587-0.0228=0.1359

2. P(Z $>$ 1.52)
   Answer
   P(Z $>$ 1.52)=1-Φ(1.52)=1-0.9357=0.0643

3. P(-2 $<$ Z $<$ 0.89)
   Answer
   P(-2 $<$ Z $<$ 0.89)=Φ(0.89)-Φ(-2)=0.8133-0.0228=0.7905

4. P(0 $<$ Z $<$ 2.15)
   Answer

5. P(0 $<$ Z $<$ 2.15)=Φ(2.15)-Φ(0)=0.9842-0.5=0.4842