Review 2

Chapter 2.4 & 3.5:

- Tabular and graphical methods: Cross-tabulation (qualitative and quantitative data) and scatter diagram (only quantitative data).
- Numerical Method: Covariance and Correlation Coefficient.

Chapter 4:

- Multiple step Experiments, permutations, and combinations.
- Event, addition law and mutually exclusive events
- Bayes' theorem

Equations:

$$s_{xz} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})}{n-1}, r_{xz} = \frac{s_{xz}}{s_x s_z} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(z_i - \bar{z})^2}}.$$

Example 1:

(a)

The following data are for 30 observations on two qualitative variables x and y. The categories for x are $\mathrm{A}, \mathrm{B},$ and C; the categories for y are 1 and 2.

Index	1	2	3	4	5	6	7	8	9	10
x	A	B	B	C	B	C	B	C	A	B
y	1	1	1	2	1	2	1	2	1	1
Index	11	12	13	14	15	16	17	18	19	20
x	A	B	C	C	C	B	C	B	C	B
y	1	1	2	2	2	2	1	1	1	1
Index	21	22	23	24	25	26	27	28	29	30
x	C	B	C	A	B	C	C	A	B	B
y	2	1	2	1	1	2	2	1	1	2

(i) Develop a cross-tabulation for the data, with *x* in columns and *y* in rows.

(ii) What is the relationship, if any, between *x* and *y*?

(b) For the following data,

X	2	4	6	8	10
Y	-5	-7	-9	-11	-13

Compute and interpret the sample covariance and the sample correlation coefficient.

[Solution:]

(a)

(i)

Category	A	B	C	Total
1	5	11	2	18
2	0	2	10	12
Total	5	13	12	30

(ii)

Category A values for x are always associated with category 1 values for y. Category B values for x are usually associated with category 1 values for y. Category C values for x are usually associated with category 2 values for y.

(b)

Since $\bar{x}=6, \bar{y}=-9$,

$$s_{xy}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}=\frac{(2-6)(-5+9)+\cdots+(10-6)(-13+9)}{5-1}$$

$$=-10.$$

Also, since

$$s_x^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}=\frac{(2-6)^2+\cdots+(10-6)^2}{5-1}=10$$

And

$$s_y^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}=\frac{(-5+9)^2+\cdots+(-13+9)^2}{5-1}=10.$$

Thus

$$r_{xy}=\frac{s_{xy}}{s_x s_y}=\frac{-10}{\sqrt{10}\sqrt{10}}=-1.$$

The covariance indicates the two variables are negatively correlated. Further, the correlation function indicates there is perfectly linear correlation between the two variables with negative slope.

Example 2:

Consider a sample with the following data

27	25	20	15	30	34	28	25

Compute

(a) the coefficient of variation,

(b) the $77th$ percentile and median.

(c) Show the box plot.

(d) Compute the z-scores for the data $27, 15,$ and 20.

(e) Consider the other sample of data

55	51	41	31	61	69	57	51

Find the sample correlation for the two samples of data.

[Solution:]

(a)

$$\overline{x} = \frac{\sum_{i=1}^{8} x_i}{8} = \frac{27 + 25 + \cdots + 28 + 25}{8} = 25.5,$$

$$s = \sqrt{\frac{\sum_{i=1}^{8}(x_i - \overline{x})^2}{8-1}} = \sqrt{\frac{(27-25.5)^2 + \cdots + (25-25.5)^2}{7}} = \sqrt{34.57} = 5.88,$$

and thus

$$CV = \frac{s}{\overline{x}} \cdot 100 = \frac{5.88}{25.5} \cdot 100 = 23.06.$$

(b)

1. The data are

15	20	25	25	27	28	30	34

2. The indexes are

$$8 \cdot \frac{77}{100} = 6.16$$

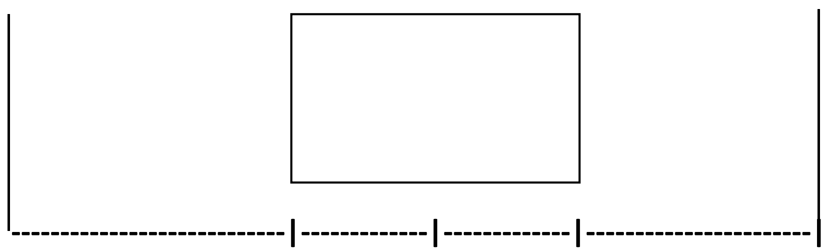
and

$$8 \cdot \frac{50}{100} = 4.$$

Thus, the $77th$ percentile is 30 and the median is

$$\frac{25 + 27}{2} = 26.$$

(c)



12.75 22.5 26 29 38.75

(d)

$$27: z = \frac{27 - 25.5}{5.88} = 0.255,$$
$$15: z = \frac{15 - 25.5}{5.88} = -1.786,$$
$$20: z = \frac{20 - 25.5}{5.88} = -0.935.$$

(e)

Since $\bar{x} = 25.5, \bar{y} = 52$,

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$
$$= \frac{(27 - 25.5)(55 - 52) + \cdots + (25 - 25.5)(51 - 51)}{8 - 1}$$
$$= 69.14$$

Also, since

$$s_x = s = 5.88$$

and

$$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{(55 - 52)^2 + \cdots + (51 - 52)^2}{8 - 1} = 138.29.$$

Thus

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{69.14}{5.88\sqrt{138.29}} = 1.$$

Example 3:

Of the 20 students (12 boys and 8 girls) in the sixth grade class at some school, 5 students were absent on Thursday. How many different combinations that 2 of the absent students were girls?

[Solution:]

$$\binom{12}{3}\binom{8}{2} = \frac{12!}{9!\,3!} \cdot \frac{8!}{6!\,2!} = 6160.$$

Example 4:

You are given the following information on events $A, B,$ and C:

$$P(A) = 0.3, P(B) = 0.3, P(C) = 0.7, P(A \cup B) = 0.5, P(B \cap C) = 0.21.$$

(a) Compute $P(A^c \cap B^c)$.

(b) Compute $P(A|B)$.

(c) Are A and B mutually exclusive? Explain.

(d) Are B and C independent? Explain.

[Solution:]

(a)

$$P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - 0.5 = 0.5.$$

(b)

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.1}{0.3} = \frac{1}{3} = 0.333.$$

since

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.3 + 0.3 - 0.5 = 0.1.$$

(c)

Since $P(A \cap B) = 0.1 \neq 0$, A and B are not mutually exclusive.

(d)

Since $P(B \cap C) = 0.21 = 0.3 \cdot 0.7 = P(B) \cdot P(C)$, B and C are independent.

Example 5:

In a city, 35% of the residents live in houses and 65% of the residents live in apartments. Of the people who live in houses, 50% own their own business. Of the people who live in apartments, 10% own their own business. If a person owns his or her own business, find the probability that he or she lives in a house.

[Solution:]

A: the residents live in houses

A^c: the residents live in apartments

B: a person owns his or her own business

$$P(A) = 0.35, P(A^c) = 0.65, P(B|A) = 0.5, P(B|A^c) = 0.1.$$

By Bayes' theorem,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)} = \frac{0.35 \cdot 0.5}{0.35 \cdot 0.5 + 0.65 \cdot 0.1}$$
$$= 0.73$$