Jinan University
Faculty of Business Administration
Tripoli - Lebanon

Statistics

Preparatory Entrance Exam

Version 1

Statistics

Jinan University
Faculty of Business Administrations

Contents

Chapter 1

Distrubutions in one variable

1.1 Statistical Data

The statistical data are represented as follow:

- In the case of discrete variables $\{(x_i, n_i) : 1 \leq i \leq p\}$ with $x_1 < x_2 < \cdots < x_p$;
- In the case of continuous variables $\{([a_i, a_{i+1}[, n_i) : 1 \leq i \leq p\}$

$\sum_{i=1}^{p} n_i = n.$

1.2 Graphical Representation

1.2.1 Discrete Variable

Bare chart

Frequency polygon

Curve of cumulative frequency

If $x < x_1$, $F(x) = 0$.
If $x_1 \leq x < x_2$, $F(x) = p_1$.

If $x_2 \leq x < x_3$, $F(x) = p_1 + p_2$.
...
If $x_n \leq x$, $F(x) = p_1 + ... + p_n = 1$.

1.2.2 Continuous Variable

Histogram

Frequency Polygon

Cumulative Curve

1.3 Important Measures

1.3.1 Measure of central tendency

MODE

- The mode of a set of data is the value that has the greatest frequency
- The modal class is the class having the greatest frequency.

MEDIANE

- If X is discrete variable have N value: $v_1 \leq v_2 \leq ... \leq v_N$. The median m_e corresponds to the value at middle of the distribution when arranged in ascending order.
  If $N = 2k + 1$, $m_e = v_{k+1}$
  If $N = 2k$, $m_e = \dfrac{v_k + v_{k+1}}{2}$.

- If X is a continuous variable, we can obtain the median m_e by using this formula: $\frac{m_e - a_i}{a_{i+1} - a_i} = \frac{0,5 - p_{i-1}}{p_i - p_{i-1}}$, so $m_e = a_i + \frac{0,5 - p_{i-1}}{p_i - p_{i-1}}(a_{i+1} - a_i)$

MEAN

- If X is discrete variable of the distridution $\{(x_i, n_i) : 1 \leq i \leq p\}$, we called **mean** the number: $\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i$.
- If X is a continuous variable of the distibution, $\{([a_i, a_{i+1}[, n_i) : 1 \leq i \leq p\}$. $\overline{x} = \frac{1}{n}\sum_{i=1}^{p} n_i x_i = \sum_{i=1}^{p} f_i x_i$. where x_i is the center of the class $[a_i, a_{i+1}[$ ($x_i = \frac{a_i + a_{i+1}}{2}$).

1.3.2 Measures of variability

VARIANCE AND STANDARD DEVIATION

The **variance** is positif number: $V = \overline{x^2} - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{p} n_i x_i^2 - \overline{x}^2 = \sum_{i=1}^{p} f_i x_i^2 - \overline{x}^2$ also $V = \frac{1}{n}\sum_{i=1}^{p} n_i (x_i - \overline{x})^2 = \sum_{i=1}^{p} (f_i x_i - \overline{x})^2$. The standard deviation is $\sigma = \sqrt{V}$.

1.4 Exercises

Consider the following statistical table:

x_i	[155;160[	[160;165[	[165;170[	[170;175[	[175;180[
n_i	4	6	12	5	3

Find the mean, the median, the modal class, the variance and the standard deviation.

Answers:

the mean is 167, the median is 167.08 and the modal class is [165;170[.

Chapter 2

Distributions in two variables

2.1 Data presentation

Adistribution in two variables is a set defined on the same sample of size n of two variables X and Y.

• Data presentation:

	1	$\cdots$	i	$\cdots$	n
X	x_1	$\cdots$	x_i	$\cdots$	x_n
Y	y_1	$\cdots$	y_i	$\cdots$	y_n

2.2 Important Values

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$
$$V(X) = \sigma_X^2 = \overline{X^2} - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{X}^2$$
$$V(Y) = \sigma_Y^2 = \overline{Y^2} - \overline{Y}^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \overline{Y}^2$$
$$Cov(X,Y) = \overline{XY} - \overline{X}.\overline{Y}$$
$$r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

2.3 Scatter Plot

The diagram of dispersion is a graph that associates to every observation a point in a rectangular system of coordinates. The set of points obtained form the so-called data points or the scatter plot.

2.4 Linear adjustment

We called **the regression line** or **the Least Squares line** of y with respect of x the following line: $(\Delta) : y = ax + b$. (Δ) gives a **linear adjustment** of the data points.
Where $a = \dfrac{Cov(X,Y)}{V(X)}$ and $b = \overline{Y} - a\overline{X}$.

2.5 Exercises

Consider the following statistical table:

x_i	480	450	480	540	570	420	390	520	470	480
y_i	22	18	20	24	24	22	14	22	18	16

1. Find $\overline{X}$, $\overline{Y}$, $V(X)$, $V(Y)$, σ_X and σ_Y.

2. Find $cov(X.Y)$ and r.

Answer:
$\overline{X} = 480$, $\overline{Y} = 20$, $V(X) = 2600$, $V(Y) = 10.4$, $cov(X,Y) = 118$ and $r = 0.72$.