

DESCRIPTIVE STATISTICS

Exercises

EXERCISE 1

The following table shows the glycemia (mg/dL) of 500 older adults grouped in 5 classes having the same width:

CLASS INTERVAL	MIDDLE POINT	FREQUENCY		CUMULATIVE FREQUENCY	
		f	p%	F	P%
65- 75	70	75	15	75	15
75- 85	80	100	20	175	35
85- 95	90	150	30	225	65
95- 105	100	125	25	450	90
105- 115	110	50	10	500	100

- Calculate mean and variance
- Identify the modal class
- Represent the data in a Galton Ogive and identify the glycemie value exceeded by only 5% of these older adults
- Find the class containing the 50th percentile

SOLUTION 1

a.

To calculate the mean and the variance, it is necessary to consider the middle point as the representative value of each class.

$$\bar{x} = \frac{\sum_{i=1}^k x_i \cdot f_i}{n} = \frac{(70 \cdot 75) + (80 \cdot 100) + (90 \cdot 150) + (100 \cdot 125) + (110 \cdot 50)}{500} = \frac{44750}{500} = 89.5 \text{ mg/dL}$$

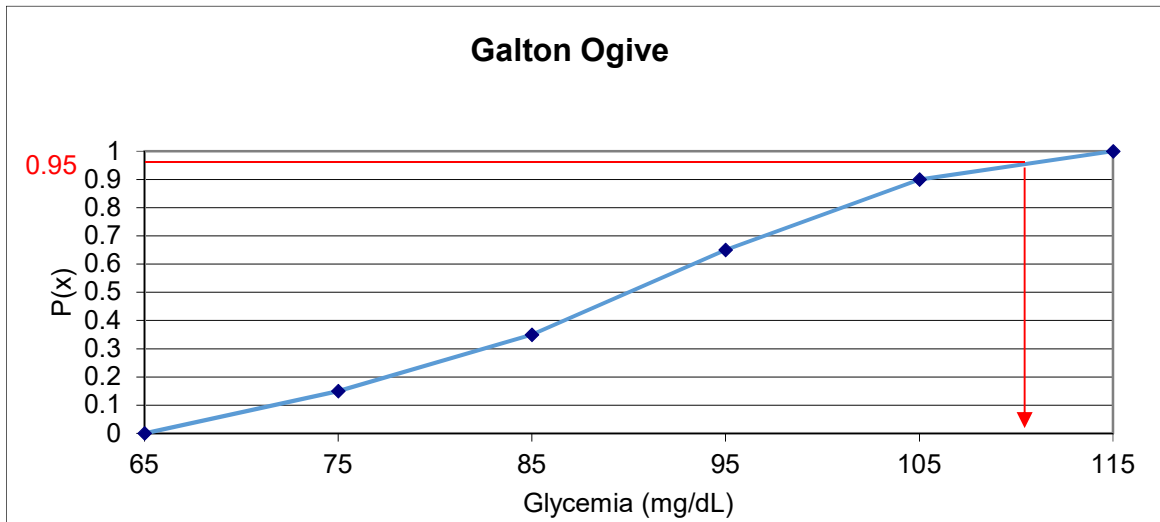
$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{(n - 1)} = \frac{(70 - 89.5)^2 \cdot 75 + (80 - 89.5)^2 \cdot 100 + (90 - 89.5)^2 \cdot 150 + (100 - 89.5)^2 \cdot 125 + (110 - 89.5)^2 \cdot 50}{499} = \frac{28518.75 + 9025 + 37.5 + 13781.25 + 21012.5}{499} = \frac{72375}{499} = 145.04(\text{mg/dL})^2$$

$$s = \sqrt{s^2} = \sqrt{145.04} = 12.04 \text{ mg/dL}$$

b.

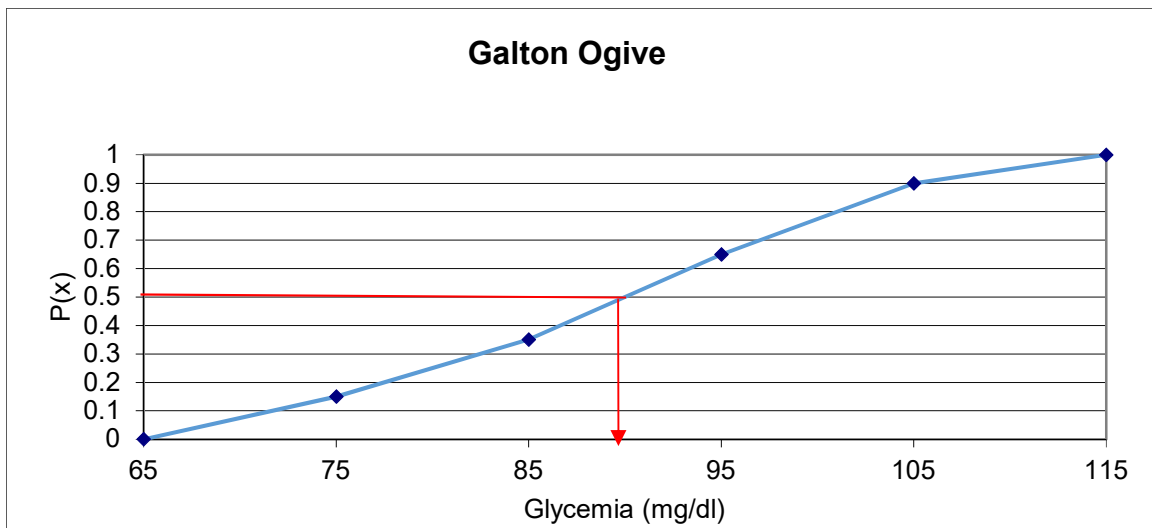
The modal class is 85-|95 mg/dL

c.



110 mg/dL is the glycemic value exceeded by only 5% of older adults

d.



The class containing the 50th percentile is 85-|95. Consequently, the median is 90 mg/dL

EXERCISE 2

The following table shows the absolute frequencies of the hemoglobin blood concentration (g/dL) categorized into 5 classes for 300 patients:

	Blood concentration of Hb (g/dL)					Total
	12 (11.5, 12.5]	13 (12.5, 13.5]	14 (13.5, 14.5]	15 (14.5, 15.5]	16 (15.5, 16.5]	
Females	18	65	14	2	1	100
Males	2	40	71	58	29	200
Total	20	105	85	60	30	300

- What is the proportion of patients with Hb > 14.5 g/dL?
- What is the proportion of females with Hb > 14.5 g/dL?

- c. What is the proportion of males with Hb > 14.5 g/dL?
 d. What is the proportion of females among patients with Hb < 12.5 g/dL?

SOLUTION 2

a.

$$p = \frac{60 + 30}{300} = \frac{90}{300} = 0.3$$

b.

$$p = \frac{2 + 1}{100} = \frac{3}{100} = 0.03$$

c.

$$p = \frac{58 + 29}{200} = \frac{87}{200} = 0.435$$

d.

$$p = \frac{18}{20} = 0.9$$

EXERCISE 3

Five men with obesity have been visited in the same day. The following table shows their weights (kg):

Patient ID	Weight (kg)
1	120
2	147
3	132
4	128
5	138

- a. Calculate mean and standard deviation

The scale was later discovered to have been calibrated badly and that all measurements were wrong overestimated by 5 kg.

- b. Calculate mean and standard deviation
 c. Calculate mean and standard deviation in hg
 d. Calculate the coefficient of variation of the weight both in kg and hg

SOLUTION 3

a.

$$\bar{x} = \frac{120 + 147 + 132 + 128 + 138}{5} = \frac{665}{5} = 133 \text{ kg}$$

$$s = \sqrt{\frac{(120 - 133)^2 + (147 - 133)^2 + (132 - 133)^2 + (128 - 133)^2 + (138 - 133)^2}{4}}$$

$$= \sqrt{\frac{169 + 196 + 1 + 25 + 25}{4}} = \sqrt{\frac{416}{4}} = \sqrt{104} = 10.2 \text{ kg}$$

b.

$$\bar{x} = 133 - 5 = 128 \text{ kg}$$

$$s = 10.2 \text{ kg} \rightarrow \text{remain unchanged}$$

c.

1 kg = 10 hg. So,

$$\bar{x} = 128 \cdot 10 = 1280 \text{ hg}$$

$$s = 10.2 \cdot 10 = 102 \text{ hg}$$

d.

$$CV = \frac{s}{\bar{x}} = \frac{10.2}{128} = \frac{1.02}{12.8} = 12.55$$

E

EXERCISE 4

A sample is composed by 120 males and 80 females. The following table shows their age in years with the percentage distribution by gender:

Age (years)	Males (%)	Females (%)
0-19	10	20
20-29	10	20
30-49	30	30
50-89	50	30
Total	100	100

- How many people are < 20 years old?
- What is the percentage of individuals that are ≥ 50 years old?
- How many males are ≥ 30 years old?
- Find the modal classes for males and females separately and for the total sample
- Identify the median of the total sample

SOLUTION 4

Age (years)	Males		Females	
	p%	f	p%	f
0-19	10	12	20	16
20-29	10	12	20	16
30-49	30	36	30	24
50-89	50	60	30	24
Total	100	120	100	80

a.

28 subjects are < 20 years old

b.

$$\frac{60 + 24}{120 + 80} = 0.42 \rightarrow 42\%$$

c.

96 males are ≥ 30 years old

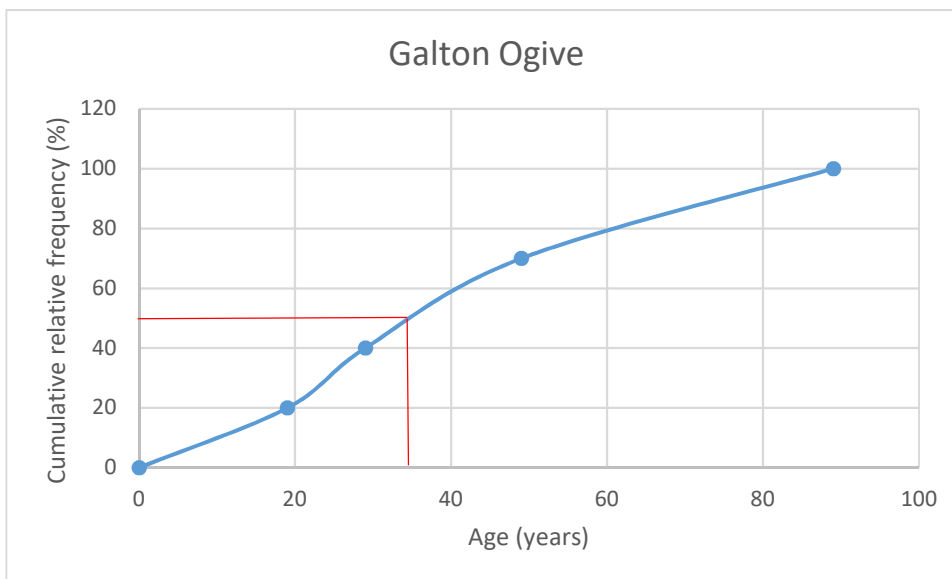
d.

Modal class for males: 50-89 years

Modal class for females: 30-49 and 50-89 years \rightarrow bimodal distribution

Modal class overall: 50-89 years

e.



The median age of the total sample is 35 years.

EXERCISE 5

The following table shows the distribution of frequencies of the attitude towards smoking observed in a group of young people.

Smoking habit	Age			
	[16, 18]	[18, 22]	[22, 25]	[25, 30]
Yes	7	8	21	30
No	16	18	9	10

- Calculate the mean age of the smokers and non-smokers
- Identify the median age class of the smokers and non-smokers
- Identify the modal age class of the smokers and non-smokers

SOLUTION 6

Age			
[16, 18]	[18, 22]	[22, 25]	[25, 30]

Smoking habit (class middle point)	17	20.5	24	28
Yes	7	8	21	30
No	16	18	9	10

a.

$$\bar{x}_S = \frac{(17 \cdot 7) + (20.5 \cdot 8) + (24 \cdot 21) + (28 \cdot 30)}{66} = \frac{119 + 164 + 504 + 840}{66} = \frac{1627}{66} = 24.65$$

$$\bar{x}_{NS} = \frac{(17 \cdot 16) + (20.5 \cdot 18) + (24 \cdot 9) + (28 \cdot 10)}{53} = \frac{272 + 369 + 216 + 280}{53} = \frac{1137}{53} = 21.45$$

b.

Adding the cumulative frequencies into the table:

Smoking habit (class middle point)	Age				Total
	[16, 18]]18, 22]]22, 25]]25, 30]	
	17	20.5	24	28	
Yes	7 (11%)	8 (23%)	21 (55%)	30 (100%)	66
No	16 (30%)	18 (64%)	9 (81%)	10 (100%)	53

Median age class for smokers:]22, 25]

Median age class for non-smokers:]18, 22]

c.

Modal class for smokers:]25, 30] years

Modal class for non-smokers:]18, 22] years