EXAM OF STATISTICS
(DESCRIPTIVE STATISTICS AND REGRESSION)

Pharmacy/Biotechnology 1st year — Version B — **October, 25 2021**

Duration: 1 hour.

(5 pts.) 1. The table below shows the number of daily sugary drinks drunk by a sample of 16-years-old people.

Drinks	n_i	f_i	N_i	F_i
0		0.1		
1			48	
2				0.725
3	24			
4				0.975
5			120	

(a) Complete the table explaining how.

(b) Plot the cumulative frequency polygon.

(c) Are there outliers?

(d) Study the normality of the distribution.

(e) If another sample of 18-years-old people has a mean 2.1 drinks and a variance 1.5 drinks^2, in which distribution is more representative the mean?

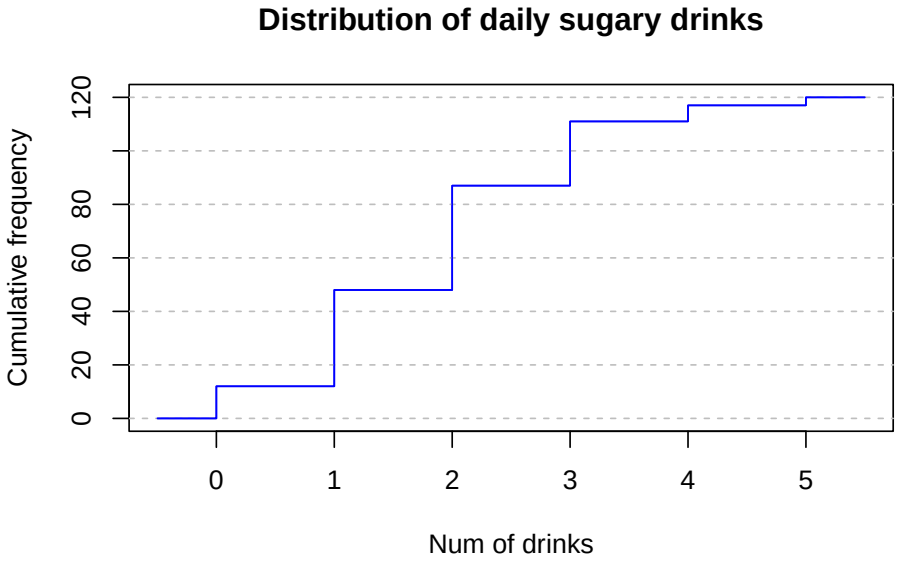
(f) Who consumes a higher relative amount of sugary drinks, a 16-years-old who consumes 3 drinks a day or a 18-years-old who consumes 4?

Use the following sums for the computations: $\sum x_i = 225$ drinks, $\sum x_i^2 = 579$ drinks^2, $\sum (x_i - \bar{x})^3 = 80.16$ drinks^3 and $\sum (x_i - \bar{x})^4 = 616.32$ drinks^4.

Solution

(a)

Drinks	*Abs.Freq.*	*Rel.Freq.*	*Cum.Abs.Freq.*	*Cum.Rel.Freq.*
0	12	0.100	12	0.100
1	36	0.300	48	0.400
2	39	0.325	87	0.725
3	24	0.200	111	0.925
4	6	0.050	117	0.975
5	3	0.025	120	1.000



(b)

(c) Quartiles: $Q_1 = 1$ drinks, $Q_2 = 2$ drinks, $Q_3 = 3$ drinks
$IQR = 2$ drinks
Fences: $f_1 = -2$ drinks and $f_2 = 6$ drinks.

(d) $\bar{x} = 1.875$ drinks, $s^2 = 1.3094$ drinks^2, $s = 1.1443$ drinks, $g_1 = 0.4458$ and $g_2 = -0.0043$. As the coefficient of skewness and the coefficient of kurtosis are between -2 and 2 we can assume that the sample comes from a normal population.

(e) Let Y be the daily sugary drinks drunk by 18-year-old people. Then, $cv_x = 0.6103$ and $cv_y = 0.5832$. As the coefficient of variation of 18-year-old is a little bit smaller than the one of 16-year-old, the mean of the 18-year-old is a little bit more representative.

(f) Standard score for 16-year-old: $z(3) = 0.9832$
Standard score for 18-year-old: $z(4) = 1.5513$
As the standard score of 4 for a 18-year-old is greater than the standard score of 3 for a 16-years-old, 4 drinks is relatively higher.

(4 pts.) 2. The rowan is a species of tree that grows at different altitudes. In order to study how the rowan adapts to different habitats, we have collected a sample of branches of 12 trees at different altitudes in Scotland. In the laboratory, the respiration rate of each branch was observed during the night. The following table shows the altitude (in meters) of each branch and the respiration rate (in nl of O_2 per hour per mg of weight).

Altitude	90	230	240	260	330	400	410	550	590	610	700	790
Respiration rate	110	200	130	150	180	160	230	180	230	260	320	370

(a) Is there a linear relationship between altitude and respiration rate of rowan. How is this relationship?

(b) How much increases the respiration rate per each increment of 100 meters in the altitude?

(c) What respiration rate is expected for a rowan at 500 meters of altitude? And for a rowan at the sea level?

(d) Are these predictions reliable?

Use the following sums for the computations (X=Altitude and Y=Respiration rate): $\sum x_i = 5200$ m, $\sum y_i = 2520$ nl/(mg· h), $\sum x_i^2 = 2760000$ (m)^2, $\sum y_i^2 = 594600$ nl/(mg· h)^2 and $\sum x_i y_j = 1253400$ m· nl/(mg· h).

Solution

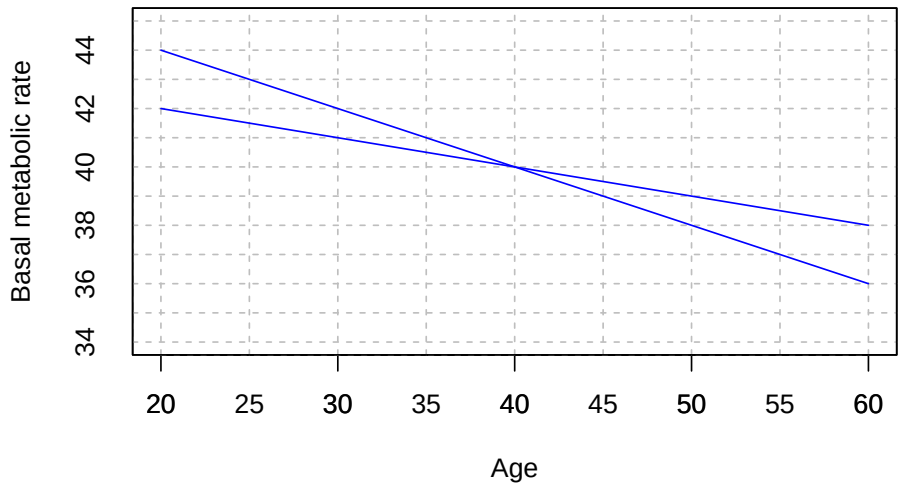
(a) $\bar{x} = 433.3333$ m, $s_x^2 = 42222.2222$ (m)^2,
$\bar{y} = 210$ nl/(mg· h), $s_y^2 = 5450$ nl/(mg· h)^2,
$s_{xy} = 13450$ m · nl/(mg· h).
As the covariance is positive, there is a direct linear relation between the altitude and the respiration rate.

(b) The respiration rate increases $b_{yx} = 0.3186$ nl/(mg·h) per meter, or what is the same, 31.8553 nl/(mg·h) per 100 meters.

(c) Regression line of the respiration rate on the altitude: $y = 71.9605 + 0.3186x$.
Predictions: $y(500) = 231.2368$ nl/(mg· h) and $y(0) = 71.9605$ nl/(mg· h).

(d) $r^2 = 0.7862$. As the coefficient of determination is not far from 1, the regression line fits well, but the sample size is too small to have reliable predictions. In addition, the prediction for the sea level is less reliable because it falls outside the range of values of the sample.

(1 pts.) 3. The relationship between basal metabolic rate and age is being studied in a sample of healthy men and the following regression lines have been obtained



(a) Compute the means of the basal metabolic rate and the age.

(b) How is the fit of the two lines?

Solution

Let X be the age and Y the basal metabolic rate.

(a) $\bar{x} = 40$ and $\bar{y} = 40$.

(b) $b_{yx} = -0.1$, $b_{xy} = -5$ and $r^2 = 0.5$, thus the fit of the regression lines moderate.