# EXAM OF STATISTICS
## (DESCRIPTIVE STATISTICS AND REGRESSION)

**Pharmacy/Biotechnology 1st year**          **Version B**          **January 17, 2022**

Name:                                    DNI:                    Group:

**Duration**: 1 hour.

(5 pts.) 1. The table below shows the age $X$ (in years) and the maximum blood pressure $Y$ (in mmHg) of a group of 5 girls and a group of 5 women.

|                | Girls |     |     |     |     | Women |     |     |     |     |
| -------------- | ----- | --- | --- | --- | --- | ----- | --- | --- | --- | --- |
| Age            | 7     | 8   | 10  | 8   | 9   | 55    | 41  | 70  | 34  | 61  |
| Blood pressure | 105   | 110 | 98  | 97  | 103 | 145   | 121 | 154 | 121 | 155 |

(a) Is there a linear relation between the pressure and the age of women?

(b) How much changes the blood pressure for every year that increases the age in women?

(c) In which group the linear regression model explains a higher percentage of the variability of the blood pressure, in girls or in women? Below is the statistics summary of the linear regression model for girls.

### Model equation

Blood.Pressure = 120.6923 + -2.1538 Age

### Model coefficients

| Coefficient | Estimation | Std.Error | t-statistic | p-value    |
| ----------- | ---------- | --------- | ----------- | ---------- |
| Intercept   | 120.6923   | 20.2198   | 5.969016    | 0.00940878 |
| Age         | -2.153846  | 2.389573  | -0.9013519  | 0.4338353  |

### Model goodness of fit

| $R^2$     | $R^2$ ajusted | F-statistic | p-value   |
| --------- | ------------- | ----------- | --------- |
| 0.2131014 | -0.04919815   | 0.8124352   | 0.4338353 |

(d) What is the expected blood pressure for a 55-year-old woman? And for a 9-year-old girl? Are these predictions reliable? Which one is more reliable?

(e) Explain why the prediction for a 55-year-old woman is not 145 mmHg as appears in the data table.

Use the following sums for the computations:
Girls: $\sum x_i = 42$ years, $\sum y_i = 513$ mmHg, $\sum x_i^2 = 358$ years$^2$, $\sum y_i^2 = 52747$ mmHg$^2$ and $\sum x_i y_j = 4298$ years·mmHg.
Women: $\sum x_i = 261$ years, $\sum y_i = 696$ mmHg, $\sum x_i^2 = 14483$ years$^2$, $\sum y_i^2 = 98048$ mmHg$^2$ and $\sum x_i y_j = 37285$ years·mmHg.

(5 pts.) 2. A study tries to determine the relation between the skin color, quantified in the Von Luschan scale, and suffering a particular skin disease. The table below summarizes the number of healthy and sick individuals in each interval of the Von Luschan scale:

| Skin color (Von Luschan) | Healthy | Sick |
|:---:|---:|---:|
| $0-6$ | 79 | 25 |
| $6-12$ | 266 | 40 |
| $12-18$ | 193 | 21 |
| $18-24$ | 188 | 12 |
| $24-30$ | 117 | 3 |
| $30-36$ | 56 | 0 |

(a) Compute the Von Luschan value such that 30% of individuals are above that value.

(b) In which group is the mean more representative, in healthy or in sick persons?

(c) Which distribution is more symmetric, the healthy or the sick persons distribution? And, which one is more peaked?

(d) Which individual has a relatively higher Von Luschan value, a healthy person with a score 15 or a sick person with a score 10?

(e) Taking the standard scores of healthy and sick persons, in which distribution is the mean of the standard scores more representative?

Use the following sums for the computations:

Healthy: $\sum x_i n_i = 14481$, $\sum x_i^2 n_i = 294867$, $\sum (x_i - \bar{x})^3 n_i = 169417.62$ and $\sum (x_i - \bar{x})^4 n_i = 9325428.37$.

Sick: $\sum x_i n_i = 1083$, $\sum x_i^2 n_i = 15669$, $\sum (x_i - \bar{x})^3 n_i = 15887.66$ and $\sum (x_i - \bar{x})^4 n_i = 440770.94$.