Template: Detailed Solutions for Exercises 3

Prepared by: sarra harik

October 2025

Calculation Rules and Notation (Clear and Concise)

Notation: n_{ij} frequency in row i and column j. $N = \sum_{i,j} n_{ij}$ total sample size.

Midpoints: For grouped class $[a, b[$ we use midpoint $m = \dfrac{a+b}{2}$.

Marginal totals: $n_{X_j} = \sum_i n_{ij}$ (column totals), $n_{Y_i} = \sum_j n_{ij}$ (row totals).

Sample mean (grouped): $\bar{X} = \dfrac{1}{N} \sum_j n_{X_j} x_j, \quad \bar{Y} = \dfrac{1}{N} \sum_i n_{Y_i} y_i.$

Second moment: $E[X^2] = \dfrac{1}{N} \sum_j n_{X_j} x_j^2, \quad E[Y^2] = \dfrac{1}{N} \sum_i n_{Y_i} y_i^2.$

Variance: $\mathrm{Var}(X) = E[X^2] - \bar{X}^2, \quad s_X = \sqrt{\mathrm{Var}(X)}.$

Joint moment: $E[XY] = \dfrac{1}{N} \sum_{i,j} n_{ij} x_j y_i.$

Covariance: $\mathrm{Cov}(X, Y) = E[XY] - \bar{X}\bar{Y}.$

Pearson correlation: $r = \dfrac{\mathrm{Cov}(X, Y)}{s_X s_Y}.$

Least-squares linear regression of Y on X:

$$b = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}, \qquad a = \bar{Y} - b\bar{X}, \qquad \hat{Y} = a + bX.$$

Given Frequency Table (Exercise 3)

Sleeping time Y / Age X	$[5,7[$	$[7,9[$	$[9,11[$	$[11,15[$
$[1,3[$	0	0	2	36
$[3,11[$	0	3	12	26
$[11,19[$	2	8	35	16
$[19,31[$	0	26	22	3
$[31,59[$	22	15	6	0
Column totals n_{X_j}	24	52	77	81
Row totals n_{Y_i}	38	41	61	51
43				

Table 1: Frequency table reconstructed from the provided data. Total $N = 234$.

Step-by-step Solution — Marginal Means and Variances

1. Midpoints

$$x_1 = 6, \quad x_2 = 8, \quad x_3 = 10, \quad x_4 = 13,$$
$$y_1 = 2, \quad y_2 = 7, \quad y_3 = 15, \quad y_4 = 25, \quad y_5 = 45.$$

2. Mean of X

Compute $\sum_j n_{X_j} x_j$:

$$6 \times 24 = 144,\ 8 \times 52 = 416,\ 10 \times 77 = 770,\ 13 \times 81 = 1053.$$

Sum: $144 + 416 + 770 + 1053 = 2383$. Then

$$\bar{X} = \frac{2383}{234} = 10.183760683760683.$$

Compute $E[X^2]$: use $6^2 \times 24 = 864$, $8^2 \times 52 = 3328$, $10^2 \times 77 = 7700$, $13^2 \times 81 = 13689$. Sum $= 25581$. Thus

$$E[X^2] = \frac{25581}{234} = 109.32051282051282.$$

Variance and sd:

$$\mathrm{Var}(X) = 109.32051282051282 - (10.183760683760683)^2 = 5.611531156402961,$$

$$s_X = \sqrt{5.611531156402961} = 2.3688670617835355.$$

3. Mean of Y

Compute $\sum_i n_{Y_i} y_i$:

$$2 \times 38 = 76,\ 7 \times 41 = 287,\ 15 \times 61 = 915,\ 25 \times 51 = 1275,\ 45 \times 43 = 1935.$$

Sum = 4488. Then
$$\bar{Y} = \frac{4488}{234} = 19.17948717948718.$$

Compute $E[Y^2]$ with $2^2 \times 38 = 152$, $7^2 \times 41 = 2009$, $15^2 \times 61 = 13725$, $25^2 \times 51 = 31875$, $45^2 \times 43 = 87175$. Sum = 134936. Then
$$E[Y^2] = \frac{134936}{234} = 576.6166666666666.$$

Variance and sd:
$$\mathrm{Var}(Y) = 576.6166666666666 - (19.17948717948718)^2 = 208.36692776604088,$$
$$s_Y = \sqrt{208.36692776604088} = 14.43500930911058.$$

Step-by-step Solution — Covariance and Correlation

1. Compute $E[XY]$

Compute each nonzero cell contribution $n_{ij}x_jy_i$ and sum (displayed as a compact table in the final PDF for clarity). The computed total is
$$\sum_{i,j} n_{ij}x_jy_i = 174696.$$

Thus
$$E[XY] = \frac{174696}{234} = 746.0.$$

2. Covariance

$\mathrm{Cov}(X,Y) = E[XY] - \bar{X}\bar{Y} = 746.0 - (10.183760683760683)(19.17948717948718) = -26.1953758492219$

3. Pearson correlation

$$r = \frac{-26.19537584922199}{(2.3688670617835355)(14.43500930911058)} \approx -0.7660672150548864.$$

Regression — Predicting Y from X

Slope and intercept:
$$b = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \frac{-26.19537584922199}{5.611531156402961} \approx -4.6681333702178796,$$

$a = \bar{Y} - b\bar{X} = 19.17948717948718 - (-4.6681333702178796)(10.183760683760683) \approx 66.71864026166327$

Regression equation:
$$\boxed{\hat{Y} = 66.71864026166327 - 4.6681333702178796\,X.}$$

Extrapolation example

Predict for $X = 66$:
$$\hat{Y}(66) = 66.71864026166327 - 4.6681333702178796 \times 66 \approx -241.37816217271677.$$