

Statistics Quiz (1) (Oct. 7)

Answers and Computations

卷：A

I. Multiple Choice (7 pts for Q1 & Q2 ; 8 pts each for Q3 & Q4)

1. (B) Negative linear relationship.
2. (C) 2.83
3. (A) 2
4. (D) the sample space

II. Fill in the Blanks (10 pts each)

5. Chebyshev's Theorem
6. population parameter

III. Short Answer (10 pts each)

7. Statistics is the discipline (art and science) that concerns the collection, organization, analysis, interpretation, and presentation of data.
8. **Definition:** $z_i = \frac{x_i - \bar{x}}{S}$, where x_i is the i th data value, $\bar{x}$ is the sample mean, and S is the standard deviation.

Interpretation: z-score denotes the number of standard deviations a data value x_i is from the mean (If you write this, you get 5 points).

Interpretation: A data value greater (less) than the sample mean will have a z-score greater (less) than zero ; A data value equal to the sample mean will have a z-score of zero (If you write it like this, you'll get 3 points at most!).

IV. Computation (10 pts each)

9. The table (sample size $n = 944$) is:

Age (years)	16–20	21–24	25–27	28 or older
Count	191	467	244	42

Event notation. Let

$$\begin{aligned} A &= \{\text{teenager expects independence at 16–20}\}, \\ B &= \{\text{teenager expects independence at 21–24}\}, \\ C &= \{\text{teenager expects independence at 25–27}\}, \\ D &= \{\text{teenager expects independence at 28 or older}\}. \end{aligned}$$

Then A, B, C, D form a partition of the sample space.

(a) Compute $P(A), P(B), P(C), P(D)$.

$$P(A) = \frac{191}{944} \approx 0.2023, P(B) = \frac{467}{944} \approx 0.4947, P(C) = \frac{244}{944} \approx 0.2585, P(D) = \frac{42}{944} \approx 0.0445.$$

(b) Probability of being financially independent before age 25. Let

$$E = \{\text{teenager expects independence at before 25}\} = A \cup B.$$

Then

$$P(E) = P(A) + P(B) = \frac{191 + 467}{944} = \frac{658}{944} \approx 0.6970.$$

(c) Interpretation (“unrealistic?”).

(chap4.2 ex20): The probability of being financially independent before age 25, 0.6970, seems high given the general economic conditions. The teenagers who responded to this survey may have had unrealistic expectations about becoming financially independent at a relatively young age.

V. Bonus Question (+20 pts)

Given paired observations

$$(x, y) = (4, 50), (8, \text{NA}), (6, 48), (11, 40), (3, 60), (16, 32),$$

discard the pair with NA. The effective data are

$$(x, y) = (4, 50), (6, 48), (11, 40), (3, 60), (16, 32), n = 5.$$

Step 1: Compute summaries.

$$\bar{x} = \frac{48 - 8}{5} = 8, \bar{y} = \frac{230}{5} = 46,$$

$$\sum x_i^2 = 502 - 8^2 = 438, \sum y_i^2 = 11028,$$

$$\sum x_i y_i = 4 \cdot 50 + 6 \cdot 48 + 11 \cdot 40 + 3 \cdot 60 + 16 \cdot 32 = 1620.$$

Step 2: Sums of squares and cross-product. Using deviation form,

$$S_{xx} = \frac{\sum x_i^2 - n\bar{x}^2}{n - 1} = \frac{438 - 5 \cdot 8^2}{5 - 1} = 29.5,$$

$$S_{yy} = \frac{\sum y_i^2 - n\bar{y}^2}{n - 1} = \frac{11028 - 5 \cdot 46^2}{5 - 1} = 112,$$

$$S_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n - 1} = \frac{1620 - 5 \cdot 8 \cdot 46}{5 - 1} = -55.$$

Step 3: Sample correlation coefficient.

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-55}{\sqrt{29.5 \cdot 112}} \approx -0.9568$$

Interpretation. $r \approx -0.96$ indicates a very strong negative linear association: as x increases, y tends to decrease strongly and consistently.