

115-1 Statistics (I)
Quiz 1 Solution — Version A
October 13, 2026

0. Oath. Copy the prescribed oath at the top of the first answer-sheet page, inserting the student's own name as required by the examination paper.

I Multiple Choice

1. **(C) Pie chart.** Favorite pet is a categorical variable. A pie chart displays the proportions of the categories in this single sample. The frequencies are Dog: 6, Gerbil: 2, Cat: 3, Fish: 3, and Lizard: 1 (total: 15).
2. **(A) Skewed to the right.** Housing prices and salaries typically have a few very high values, producing a long right tail.
3. **(C) Mean.** The mean uses every numerical value, so an extreme value can substantially change it. The range measures dispersion rather than location.
4. **(C) Chebyshev's theorem.** For any data set with a defined, positive standard deviation, at least $1 - 1/k^2$ of the observations lie within k standard deviations of the mean, where $k > 1$. No bell-shaped distribution is required.

II Fill-in-the-Blank

5. **Average annual percentage growth:** 3.6155% (**approximately** 3.616%). The equivalent annual growth factor is the geometric mean:

$$\begin{aligned}\bar{g} &= (1.0313 \times 1.0429 \times 1.0343)^{1/3} = (1.112433887011)^{1/3} \\ &\approx 1.036155022881, \\ r &= (\bar{g} - 1) \times 100\% \approx \boxed{3.6155\%}.\end{aligned}$$

Here “percentage growth” is interpreted as the average annual growth rate. The cumulative growth over all three years is a different quantity: $(1.112433887011 - 1) \times 100\% \approx 11.2434\%$.

6. **First blank: standardized.**

Second blank: $\boxed{z_i = \frac{x_i - \bar{x}}{s}}$. Here $\bar{x}$ is the sample mean and $s > 0$ is the sample standard deviation. The z -score gives the signed number of sample standard deviations by which x_i differs from $\bar{x}$.

III Short Answer

7. **Simpson's paradox.** An association or comparison observed in each of several separate groups can disappear or reverse when the groups are combined. This can occur because the groups have different compositions or weights, and the grouping variable is associated with the variables being compared.

For a group g , let p_{Ag} and p_{Bg} denote the success proportions for categories A and B in that group. Let w_{Ag} and w_{Bg} be the proportions of all category A and all category B observations, respectively, that belong to group g , with $w_{Ag}, w_{Bg} \geq 0$. Their overall proportions are weighted averages:

$$p_A = \sum_g w_{Ag} p_{Ag}, \quad p_B = \sum_g w_{Bg} p_{Bg}, \quad \sum_g w_{Ag} = \sum_g w_{Bg} = 1.$$

Even if $p_{Ag} > p_{Bg}$ in every group, $p_A < p_B$ may occur when w_{Ag} and w_{Bg} differ. Thus an overall comparison can conceal the within-group relationship; relevant group composition must be considered when interpreting an association.

8. **Covariance.** Covariance measures the direction of the linear association between two variables using the sum of products of their deviations from their respective means, divided by $n - 1$ for a sample or by N for a finite population.

For paired sample observations (x_i, y_i) , $i = 1, \dots, n$, with $n > 1$, the sample covariance is

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

For a population of N paired values, the corresponding descriptive population covariance is

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y).$$

Interpretation:

- $s_{xy} > 0$: a positive linear association in the sample; larger values of x tend to accompany larger values of y .
- $s_{xy} < 0$: a negative linear association in the sample; larger values of x tend to accompany smaller values of y .
- $s_{xy} = 0$: the sample sum of products of deviations is zero, indicating no linear association as measured by sample covariance. A nonlinear relationship may still exist. This sample result does not establish population independence or zero population covariance. Even zero population covariance does not imply independence.

Covariance has units equal to the product of the units of x and y . Its magnitude depends on the measurement scales, so it is not a standardized measure of association strength.

IV Calculation

9. Apple iPads in Schools. Throughout this question, percentages are recorded numerically in percentage-point units; for example, $x_i = 15$ represents 15%. The $n = 18$ observations, sorted in ascending order, are

12, 15, 15, 18, 18, 19, 20, 21, 22,
22, 24, 24, 26, 26, 27, 29, 42, 64.

(a) First and third quartiles. Use the percentile position $L_p = (p/100)(n + 1)$ and linearly interpolate between the adjacent ordered observations $x_{(j)}$.

$$\begin{aligned} L_{25} &= \frac{25}{100}(18 + 1) = 4.7500, \\ Q_1 &= x_{(4)} + 0.7500(x_{(5)} - x_{(4)}) = 18 + 0.7500(18 - 18) = \boxed{18.0000}, \\ L_{75} &= \frac{75}{100}(18 + 1) = 14.2500, \\ Q_3 &= x_{(14)} + 0.2500(x_{(15)} - x_{(14)}) = 26 + 0.2500(27 - 26) = \boxed{26.2500}. \end{aligned}$$

Thus the first and third quartiles represent student percentages of 18.0000% and 26.2500%, respectively.

(b) Outliers. Using the 1.5 IQR rule: The interquartile range is $\text{IQR} = Q_3 - Q_1 = 26.2500 - 18.0000 = 8.2500$. The 1.5 IQR fences are

$$\begin{aligned} \text{Lower fence} &= Q_1 - 1.5 \text{IQR} = 18.0000 - 1.5(8.2500) = 5.6250, \\ \text{Upper fence} &= Q_3 + 1.5 \text{IQR} = 26.2500 + 1.5(8.2500) = 38.6250. \end{aligned}$$

No observation is below 5.6250. The observations 42 and 64 exceed 38.6250. Therefore, under the 1.5 IQR rule, the outliers are $\boxed{42 \text{ and } 64}$, representing 42% and 64% of students.

Alternatively, using the z -score rule $|z_i| > 3$ and the supplied $\bar{x} = 24.67$ and $s = 11.87$:

$$\begin{aligned} z_{\text{for } x=12} &= \frac{12 - 24.67}{11.87} \approx -1.0674, \\ z_{\text{for } x=42} &= \frac{42 - 24.67}{11.87} \approx 1.4600, \\ z_{\text{for } x=64} &= \frac{64 - 24.67}{11.87} \approx 3.3134. \end{aligned}$$

The corresponding cutoffs are

$$\bar{x} - 3s = 24.67 - 3(11.87) = -10.9400, \quad \bar{x} + 3s = 24.67 + 3(11.87) = 60.2800.$$

Only 64 lies outside these cutoffs. Thus this rule identifies $\boxed{64}$, representing 64% of students, as the only outlier. The two rules use different criteria and can give different results; the question does not specify a rule.

V Bonus Calculation

10. Semester grade point average. Use credit hours w_i as weights and grade points x_i as values.

Course	Grade	Grade points (x_i)	Credits (w_i)	$w_i x_i$
Calculus	C	2	5	10
Biology	A	4	4	16
English	D	1	3	3
Music	B	3	2	6
P.E.	A	4	1	4
Total			15	39

$$\text{GPA} = \frac{\sum_{i=1}^5 w_i x_i}{\sum_{i=1}^5 w_i} = \frac{5(2) + 4(4) + 3(1) + 2(3) + 1(4)}{5 + 4 + 3 + 2 + 1} = \frac{39}{15} = \boxed{2.6000}.$$