115-1 Statistics (I)

Midterm Examination Solution — Version A

October 27, 2026

0. Oath. Copy the prescribed oath at the top of the first page of the answer sheet, replacing the name placeholder with the student's own name, as required on the examination paper.

I Multiple Choice

1. (B) The data are categorical; neither summary is meaningful; use a bar chart or a pie chart instead. The numerals 1–4 are labels for nominal categories, not quantitative measurements. The regions have no inherent order, so the cumulative frequency of the assigned codes and their numerical mean do not have meaningful interpretations.

2. (B) The classical method requires equally likely outcomes; the relative frequency estimate is 0.40. The outcomes "buy" and "not buy" need not be equally likely. From the observed customers, the relative frequency estimate is $40/100 = 0.4000$.

3. (A) A and B cannot be mutually exclusive. Since $P(A \cup B) \leq 1$,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq 0.6 + 0.5 - 1 = 0.1000 > 0.$$

Therefore $A \cap B$ cannot be empty. Independence and an exact intersection probability are not implied by the given marginals.

4. (B) $f(x) = (3 - x)/6$ for $x = 0, 1, 2$. Every probability is nonnegative, and

$$f(0) + f(1) + f(2) = \frac{3}{6} + \frac{2}{6} + \frac{1}{6} = 1.$$

Option (A) contains a negative value; the sums for (C) and (D) are $9/8$ and $3/2$, respectively.

II Fill-in-the-Blank

5. First blank: linear.

Second blank: causation.

The sample correlation coefficient measures *linear* association, not necessarily causation. Values near $+1$ or -1 indicate strong positive or negative linear association, respectively.

6. First blank: function.

Second blank: empirical.

A random variable is a function from experimental outcomes to real numbers. A discrete probability distribution derived using observed relative frequencies is an *empirical discrete distribution*.

III Short Answer

7-1. Interval scale versus ratio scale. Both scales are quantitative and have meaningful differences measured in fixed units. An interval scale has no *absolute (true) zero*; its zero may be an arbitrary reference point, so ratios of values are not meaningful (for example, 20°C is not twice as hot as 10°C).

A ratio scale has a *true zero* representing the absence of the measured quantity. Consequently, ratios are meaningful (for example, a weight of 20 kg is twice a weight of 10 kg). Thus the key additional property of the ratio scale is the meaningfulness of ratios, made possible by a true zero.

7-2. Portfolio expected return and risk. Let $r = ax + by$, where $b = 1 - a$ and $a, b \geq 0$.

(a) By linearity of expectation and the variance formula for a linear combination,

$$\begin{aligned} \mathrm{E}(r) &= a\,\mathrm{E}(x) + b\,\mathrm{E}(y), \\ \mathrm{Var}(r) &= a^2\,\mathrm{Var}(x) + b^2\,\mathrm{Var}(y) + 2ab\sigma_{xy}. \end{aligned}$$

(b) Portfolio risk is $\sigma_r = \sqrt{\mathrm{Var}(r)}$. Diversification can make it smaller than the standard deviation of either fund because the returns do not necessarily move together. For positive portfolio weights, a *negative* covariance σ_{xy} makes the term $2ab\sigma_{xy}$ negative and reduces the portfolio variance most strongly. In the limiting case of perfect negative correlation,

$$\sigma_{xy} = -\sigma_x\sigma_y \quad \Longrightarrow \quad \mathrm{Var}(r) = (a\sigma_x - b\sigma_y)^2,$$

which can be zero if $a\sigma_x = b\sigma_y$. Thus negative covariance provides the strongest possible diversification benefit.

IV Calculation

8. On-Time Performance of Airlines. Let A, B, C denote the respective airlines; O is on time and L is late.

(a) The joint probabilities are $P(i \cap O) = P(i)P(O \mid i)$ and $P(i \cap L) = P(i)[1 - P(O \mid i)]$, for $i = A, B, C$. Hence

Airline	O	L	Marginal P(airline)
A	$0.25(0.80) = 0.2000$	$0.25(0.20) = 0.0500$	0.2500
B	$0.35(0.70) = 0.2450$	$0.35(0.30) = 0.1050$	0.3500
C	$0.40(0.85) = 0.3400$	$0.40(0.15) = 0.0600$	0.4000
Marginal	0.7850	0.2150	1.0000

(b) The probability of an on-time arrival is

$$P(O) = 0.2000 + 0.2450 + 0.3400 = \boxed{0.7850}.$$

(c) Before any arrival information is known, **airline C** is most likely, with $P(C) = 0.4000$. After a late-arrival announcement, Bayes' formula gives

$$\begin{aligned} P(A \mid L) &= \frac{0.0500}{0.2150} \approx 0.2326, \\ P(B \mid L) &= \frac{0.1050}{0.2150} \approx 0.4884, \\ P(C \mid L) &= \frac{0.0600}{0.2150} \approx 0.2791. \end{aligned}$$

Thus, given a late arrival, **airline B** is most likely, with posterior probability $\boxed{0.4884}$.

(d) The events C and O are **not independent**, because

$$P(C \cap O) = 0.3400 \neq P(C)P(O) = 0.4000(0.7850) = 0.3140.$$

Equivalently, $P(O \mid C) = 0.8500 \neq P(O) = 0.7850$.

9. Printer Manufacturing Costs. Costs x, y, and $t = x + y$ are measured in dollars per printer.

(a) The joint distribution and its marginal probabilities are

	Direct labor cost y			
Parts cost x	40	45	50	$P(x)$
80	0.2000	0.2500	0.0500	0.5000
90	0.1000	0.1500	0.2500	0.5000
$P(y)$	0.3000	0.4000	0.3000	1.0000

Therefore,

$$\begin{aligned}
\mathrm{E}(x) &= 80(0.5) + 90(0.5) = \boxed{85.0000}, \\
\mathrm{Var}(x) &= (80-85)^2(0.5) + (90-85)^2(0.5) = \boxed{25.0000}, \\
\mathrm{E}(y) &= 40(0.3) + 45(0.4) + 50(0.3) = \boxed{45.0000}, \\
\mathrm{Var}(y) &= (40-45)^2(0.3) + (45-45)^2(0.4) \\
&\quad + (50-45)^2(0.3) = \boxed{15.0000}.
\end{aligned}$$

The expectations are in dollars; the variances are in squared dollars.

(b) The five values of $t = x + y$ arise by combining joint outcomes that have the same sum:

t (dollars)	120	125	130	135	140
$P(t)$	0.2000	0.2500	0.1500	0.1500	0.2500

For example, $P(t = 130) = P(x = 80, y = 50) + P(x = 90, y = 40) = 0.05 + 0.10 = 0.15$.

$$\begin{aligned}
\mathrm{E}(t) &= 120(0.20) + 125(0.25) + 130(0.15) \\
&\quad + 135(0.15) + 140(0.25) = \boxed{130.0000}, \\
\mathrm{Var}(t) &= (120-130)^2(0.20) + (125-130)^2(0.25) \\
&\quad + (130-130)^2(0.15) + (135-130)^2(0.15) \\
&\quad + (140-130)^2(0.25) = \boxed{55.0000}.
\end{aligned}$$

As a check, $\mathrm{E}(t) = \mathrm{E}(x) + \mathrm{E}(y) = 85 + 45 = 130$.

(c) Using $\mathrm{Var}(x + y) = \mathrm{Var}(x) + \mathrm{Var}(y) + 2\sigma_{xy}$,

$$\begin{aligned}
\sigma_{xy} &= \frac{55.0000 - 25.0000 - 15.0000}{2} = \boxed{7.5000}, \\
\rho_{xy} &= \frac{\sigma_{xy}}{\sqrt{\mathrm{Var}(x)}\sqrt{\mathrm{Var}(y)}} = \frac{7.5000}{\sqrt{25.0000}\sqrt{15.0000}} \approx \boxed{0.3873}.
\end{aligned}$$

The positive covariance and positive correlation indicate a positive, but not particularly strong, linear association: higher parts costs tend to accompany higher direct labor costs. The covariance has units of dollars squared, whereas correlation is unitless.

(d) The costs x and y are **not independent**. For example,

$$P(x = 80, y = 40) = 0.2000 \neq P(x = 80)P(y = 40) = 0.5000(0.3000) = 0.1500.$$

V Bonus Proofs

10. Independence and mutually exclusive events. Assume $P(A) > 0$ and $P(B) > 0$.

(a) Suppose A and B are independent. Then $P(A \cap B) = P(A)P(B)$. Since $A \cap B$ and $A \cap B^c$ are disjoint and their union is A,

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)[1 - P(B)] \\ &= P(A)P(B^c). \end{aligned}$$

Hence A **and** B^c **are independent**.

(b) Suppose instead that A and B are mutually exclusive. Then $A \cap B = \varnothing$, so

$$P(A \cap B) = 0 \neq P(A)P(B) > 0.$$

Thus A **and** B **are dependent**. Moreover, because $P(B) > 0$,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \boxed{0}.$$

Once B is known to have occurred, A cannot occur.