National Chengchi University, 115-1 Statistics (I)

Midterm Exam (Exam A), October 27, 2026 – Solutions and Grading Notes

Scope: Chapters 1 to 5.4 (Anderson et al., 14/e). Regular questions: 100 points. Bonus: 20 points. Total: 120.

Numerical answers are reported to 4 decimal places where applicable. Exact fractions are also acceptable.

Answer key (summary)

No.	Type / Topic	Answer	Points
1	MC, categorical data (Ch2.1)	(B)	5
2	MC, assigning probabilities (Ch4.1)	(B)	5
3	MC, mutually exclusive vs. independent (Ch4.3–4.4)	(A)	5
4	MC, valid discrete probability distribution (Ch5.2)	(B)	5
5	Fill-in (Ch3.5)	**linear**; **causation**	5
6	Fill-in (Ch5.1–5.2)	**function**; **empirical**	5
7	Short answer, interval vs. ratio scale (Ch1.2)	see below	10
8	Short answer, portfolio risk (Ch5.4)	see below	10
9	Calculation, airlines (Ch4.4–4.5)	$P(O) = 0.7850$; most likely given late: B, 0.4884; not independent	25
10	Calculation, printer costs (Ch5.2–5.4)	$E(t) = 130$, $\mathrm{Var}(t) = 55$, $\sigma_{xy} = 7.5$, $\rho_{xy} = 0.3873$; not independent	25
11	Bonus, proofs (Ch4.3–4.4)	see below	20

(I) Multiple choice (20%, 5% each)

1. Answer: (B). The region codes are only labels (nominal scale), so the data are categorical even though they are recorded as numbers. For categorical data, only counting the observations in each category (frequency, relative frequency, percent frequency) is meaningful; the appropriate displays are a bar chart or a pie chart.

- (A) is wrong: a numerical code does not make the data quantitative; arithmetic on the codes (such as an average) is meaningless.
- (C) is wrong: a cumulative frequency distribution requires an ordering of the values (quantitative data); "North < Central < South < East" has no meaning, so the cumulative distribution is not meaningful either.
- (D) is wrong: a histogram is for quantitative data, and the codes are not measured on an interval scale.

2. Answer: (B). The classical method assigns $1/n$ to each outcome *only when* the n experimental outcomes are equally likely. Nothing supports the assumption that "buy" and "not buy" are equally likely; in fact the data suggest the opposite. With data from 100 repetitions of the experiment, the relative frequency method is appropriate and gives $P(\text{buy}) = 40/100 = 0.40$.

- (A) is wrong: the classical method is not "always" applicable; the equally likely condition is essential.
- (C) is wrong: the two-outcome sample space is fine; the error is in the method of assigning probabilities, not in the list of outcomes.
- (D) is wrong: the two basic requirements ($0 \le P(E_i) \le 1$, $\sum P(E_i) = 1$) are necessary but not sufficient; many assignments satisfy them, but only the one supported by the data (or by equally likely outcomes) is justified.

3. Answer: (A). By the addition law, $P(A\cup B) = P(A)+P(B)-P(A\cap B) = 1.1-P(A\cap B) \le 1$, so $P(A \cap B) \ge 0.1 > 0$. Hence A and B have sample points in common and cannot be mutually exclusive.

- (B) is wrong: independence is a property that must be checked ($P(A \cap B) = P(A)P(B)$); it is not implied by the given information.
- (C) and (D) hold only if A and B are independent ($P(A \cap B) = 0.6 \times 0.5 = 0.3$ and then $P(A \cup B) = 0.8$); they are not guaranteed. All we know is $0.1 \le P(A \cap B) \le 0.5$.

4. Answer: (B). A valid discrete probability distribution requires $f(x) \ge 0$ for all x and $\sum f(x) = 1$.

- (A): $f(1) = -0.5 < 0$. Invalid (negative probability).
- (B): $f(0) = 3/6$, $f(1) = 2/6$, $f(2) = 1/6$; all nonnegative and the sum is $6/6 = 1$. **Valid.**
- (C): $f(1) = 2/8$, $f(2) = 3/8$, $f(3) = 4/8$; the sum is $9/8 \ne 1$. Invalid.
- (D): $f(1) = 1/4$, $f(2) = 2/4$, $f(3) = 3/4$; the sum is $6/4 \ne 1$. Invalid.

Grading notes. Each MC question: 5 points for the correct letter, 0 otherwise. No partial credit.

(II) Fill in the blanks (10%, 5% each; 2.5% per blank)

5. First blank: **linear**. Second blank: **causation**.
The correlation coefficient measures the strength of the *linear* association between x and y; a high correlation does not mean that changes in one variable *cause* changes in the other (lecture notes Chapter 3, p.17; textbook Section 3.5).
Acceptable alternatives: blank 1: "linear" (also "linear relationship"); blank 2: "causation" (also "a causal relationship", "cause and effect", "causality").

6. First blank: **function**. Second blank: **empirical**.
A random variable is a function that maps each experimental outcome to a real number (lecture notes Chapter 5, p.1). When the relative frequency method is used with a large data set treated as the population, the resulting distribution is called an *empirical* discrete distribution (lecture notes Chapter 5, p.3; textbook Section 5.2).
Acceptable alternatives: blank 1: "function" (also "mapping", "real-valued function"); blank 2: "empirical" (also "empirical discrete"). Not acceptable for blank 2: "relative frequency" alone (that is the method, not the name of the distribution).

Grading notes. 2.5 points per blank. Minor spelling slips that leave the word unambiguous (e.g. "causasion") may receive 1.5 points; a different statistical term receives 0.

(III) Short answer (20%, 10% each)

7. Interval scale vs. ratio scale (model answer).

- **Interval scale:** the data have all the properties of ordinal data (they can be ranked), and in addition the interval (difference) between two values is expressed in a fixed unit of measure, so differences are meaningful. Interval data are always numerical. The zero point is arbitrary: a value of 0 does *not* mean that none of the characteristic is present. Example: SAT scores (textbook), or temperature in degrees Celsius (0°C does not mean "no temperature").
- **Ratio scale:** the data have all the properties of interval data, and in addition the *ratio* of two values is meaningful. This requires an inherent (absolute) zero: the value 0 indicates that nothing exists for the variable at the zero point. Examples: distance, height, weight, time, cost (a cost of \$0 means the item is free; a \$30,000 car costs twice as much as a \$15,000 car).

- **The difference:** the ratio scale has a meaningful (absolute) zero, so statements such as "twice as much" are valid; on an interval scale only differences (not ratios) are meaningful because the zero is arbitrary (e.g. 20°C is not "twice as warm" as 10°C, but an SAT score of 620 is 70 points higher than 550).

Grading notes. 4 points: states that on the ratio scale the ratio of two values is meaningful, whereas on the interval scale only differences (fixed unit) are meaningful. 4 points: explains the role of zero (absolute/inherent zero meaning "nothing exists" for ratio data; arbitrary zero for interval data). 2 points: correct, clear statement that interval data have the ordinal properties plus a fixed unit of measure, or a correct illustrative example for each scale. An answer that only gives examples without stating the properties receives at most 4 points.

8. Portfolio return $r = ax + by$, $a + b = 1$ (model answer).

(a). (5%) By the formulas for a linear combination of two random variables (textbook equations 5.8 and 5.9),

$$E(r) = aE(x) + bE(y), \qquad \mathrm{Var}(r) = a^2\mathrm{Var}(x) + b^2\mathrm{Var}(y) + 2ab\,\sigma_{xy}.$$

(b). (5%) Because $0 < a, b < 1$, we have $a^2 < a$ and $b^2 < b$, so the first two terms are *shrunken* versions of the individual variances (for instance, with $a = b = 0.5$ and $\mathrm{Var}(x) = \mathrm{Var}(y) = V$, the first two terms add to only $V/2$). The third term $2ab\,\sigma_{xy}$ is negative whenever the covariance is negative, which reduces $\mathrm{Var}(r)$ further. Therefore $\mathrm{Var}(r)$, and hence the standard deviation of the portfolio, can be smaller than both $\mathrm{Var}(x)$ and $\mathrm{Var}(y)$ (even when $\sigma_{xy} = 0$ this is possible; with $\sigma_{xy} < 0$ the reduction is larger). The risk reduction is strongest when σ_{xy} is **negative** (the more negative, i.e. the closer ρ_{xy} is to -1, the stronger): when x tends to be above its mean, y tends to be below its mean, so the two returns offset each other and the portfolio return fluctuates less. (In the extreme case $\mathrm{Var}(x) = \mathrm{Var}(y) = V$, $\sigma_{xy} = -V$, $a = b = 0.5$ one gets $\mathrm{Var}(r) = 0$.)

Grading notes. (a) 2.5 points for each correct formula; deduct 1 point if the cross term is written as $ab\,\sigma_{xy}$ or $2\sigma_{xy}$. (b) 2 points for the argument that the weights a^2, b^2 shrink the individual variances (or any correct argument that the sum can be below both variances); 2 points for identifying negative covariance as the sign that reduces risk most; 1 point for the interpretation (returns move in opposite directions and offset each other). An answer that merely says "diversification reduces risk" without using the formula receives at most 2 points for (b).

(IV) Calculation (50%, 25% each)

9. On-Time Performance of Airlines. Prior probabilities: $P(A) = 0.25$, $P(B) = 0.35$, $P(C) = 0.40$. Conditional probabilities: $P(O \mid A) = 0.80$, $P(O \mid B) = 0.70$, $P(O \mid C) = 0.85$; hence $P(L \mid A) = 0.20$, $P(L \mid B) = 0.30$, $P(L \mid C) = 0.15$.

(a). (5%) Joint probabilities by the multiplication law, $P(\text{airline} \cap O) = P(\text{airline})\,P(O \mid \text{airline})$:

	O (on time)	L (late)	Total
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
Total	0.7850	0.2150	1.0000

(b). (5%) $P(O) = 0.2000 + 0.2450 + 0.3400 = \mathbf{0.7850}$ (equivalently $P(L) = 0.2150$).

(c). (10%) Before any information, the most likely airline is C (prior probability 0.40, the largest of $0.25, 0.35, 0.40$).

Given that the flight is late, by Bayes' theorem $P(\text{airline} \mid L) = P(\text{airline} \cap L)/P(L)$ with $P(L) = 0.2150$:

$$P(A \mid L) = \frac{0.0500}{0.2150} = 0.2326, \quad P(B \mid L) = \frac{0.1050}{0.2150} = \mathbf{0.4884}, \quad P(C \mid L) = \frac{0.0600}{0.2150} = 0.2791.$$

Given the late arrival, the most likely airline is B, with posterior probability **0.4884**. (The information "late" moves the most likely airline from C to B, because B has the worst on-time rate.)

(d). (5%) Let C denote "the flight belongs to airline C". $P(C \cap O) = 0.3400$ while $P(C)P(O) = 0.40 \times 0.7850 = 0.3140$. Since $0.3400 \neq 0.3140$, the events are **not independent** (dependent). Equivalently, $P(O \mid C) = 0.85 \neq P(O) = 0.7850$, or $P(C \mid O) = 0.3400/0.7850 = 0.4331 \neq P(C) = 0.40$.

Grading notes. (a) 4 points for the six joint probabilities (deduct 1 point per wrong cell, minimum 0); 1 point for the marginal probabilities. (b) 5 points for 0.7850 (3 points if the method is right but a joint probability from (a) was wrong and carried through consistently). (c) 2 points for identifying C as the prior most likely airline; 2 points for $P(L) = 0.2150$ or the correct Bayes set-up; 4 points for the three posterior probabilities (at least $P(B \mid L) = 0.4884$ together with the comparison against the other two airlines); 2 points for the conclusion "B". (d) 3 points for a correct numerical comparison ($P(C \cap O)$ vs. $P(C)P(O)$, or $P(O \mid C)$ vs. $P(O)$); 2 points for the conclusion "not independent". Consistent carry-through errors are penalized only once.

10. Printer Manufacturing Costs. Joint distribution $f(x, y)$ (rows: x = parts cost; columns: y = direct labor cost):

	$y = 40$	$y = 45$	$y = 50$	$f(x)$
$x = 80$	0.20	0.25	0.05	0.50
$x = 90$	0.10	0.15	0.25	0.50
$f(y)$	0.30	0.40	0.30	1.00

(a). (10%) Marginal distribution of x: $f(80) = 0.50$, $f(90) = 0.50$.

$$E(x) = 80(0.5) + 90(0.5) = \mathbf{85},$$
$$\text{Var}(x) = (80 - 85)^2(0.5) + (90 - 85)^2(0.5) = 12.5 + 12.5 = \mathbf{25} \quad (\sigma_x = 5).$$

Marginal distribution of y: $f(40) = 0.30$, $f(45) = 0.40$, $f(50) = 0.30$.

$$E(y) = 40(0.3) + 45(0.4) + 50(0.3) = 12 + 18 + 15 = \mathbf{45},$$

$$\begin{aligned}\text{Var}(y) &= (40 - 45)^2(0.3) + (45 - 45)^2(0.4) + (50 - 45)^2(0.3) \\ &= 7.5 + 0 + 7.5 = \mathbf{15} \quad (\sigma_y = \sqrt{15} = 3.8730).\end{aligned}$$

(b). (5%) $t = x + y$ takes the values $120, 125, 130, 135, 140$:

t	120	125	130	135	140
$f(t)$	0.20	0.25	$0.05 + 0.10 = 0.15$	0.15	0.25

$$\begin{aligned}E(t) &= 120(0.20) + 125(0.25) + 130(0.15) + 135(0.15) + 140(0.25) \\ &= 24 + 31.25 + 19.5 + 20.25 + 35 = \mathbf{130},\end{aligned}$$
$$\begin{aligned}\text{Var}(t) &= (120 - 130)^2(0.20) + (125 - 130)^2(0.25) + (130 - 130)^2(0.15) \\ &\quad + (135 - 130)^2(0.15) + (140 - 130)^2(0.25) \\ &= 20 + 6.25 + 0 + 3.75 + 25 = \mathbf{55} \quad (\sigma_t = \sqrt{55} = 7.4162).\end{aligned}$$

(Check: $E(t) = E(x) + E(y) = 85 + 45 = 130$.)

(c). (5%) By equation (5.6),

$$\sigma_{xy} = \frac{\text{Var}(x+y) - \text{Var}(x) - \text{Var}(y)}{2} = \frac{55 - 25 - 15}{2} = \mathbf{7.5}.$$

(Direct check with the definition:

$$\begin{aligned}\sum (x-85)(y-45)f(x,y) &= (-5)(-5)(0.20) + (-5)(0)(0.25) + (-5)(5)(0.05) \\ &\quad + (5)(-5)(0.10) + (5)(0)(0.15) + (5)(5)(0.25) \\ &= 5 + 0 - 1.25 - 2.5 + 0 + 6.25 = 7.5.)\end{aligned}$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{7.5}{5 \times 3.8730} = \mathbf{0.3873}.$$

Interpretation: the covariance is positive, so parts cost and direct labor cost tend to move in the same direction (printers with the more expensive parts also tend to need more labor); the correlation 0.3873 indicates a positive but moderately weak linear relationship.

(d). (5%) x and y are **not independent**: for example $f(80, 40) = 0.20$ while $f_x(80) f_y(40) = 0.50 \times 0.30 = 0.15 \neq 0.20$ (any one cell suffices; e.g. $f(80, 50) = 0.05 \neq 0.15$). Equivalently, $\sigma_{xy} = 7.5 \neq 0$ (independent random variables would have zero covariance), or $P(y = 40 \mid x = 80) = 0.20/0.50 = 0.40 \neq P(y = 40) = 0.30$.

Grading notes. (a) 2 points for the two marginal distributions; 2 points each for $E(x)$, $\text{Var}(x)$, $E(y)$, $\text{Var}(y)$ (1 point if the method is right but arithmetic is wrong). (b) 3 points for the distribution of t (deduct 1 point for each wrong probability, minimum 0); 1 point for $E(t)$; 1 point for $\text{Var}(t)$. (c) 2 points for $\sigma_{xy} = 7.5$ (also full credit if computed directly from the definition); 2 points for $\rho_{xy} = 0.3873$; 1 point for the interpretation (positive, moderately weak relationship). (d) 3 points for a valid numerical check; 2 points for the conclusion. Carry-through errors from (a)/(b) are penalized only once.

(V) Bonus (20%)

11. Let $P(A) > 0$ and $P(B) > 0$.

(a). (10%) **Claim:** if A and B are independent (i.e. $P(A \cap B) = P(A)P(B)$), then A and B^c are independent, i.e. $P(A \cap B^c) = P(A)P(B^c)$.

Proof. Every sample point of A is either in B or in B^c, so

$$A = (A \cap B) \cup (A \cap B^c),$$

and the two events $A \cap B$ and $A \cap B^c$ are mutually exclusive (a sample point cannot be in both B and B^c). By the addition law for mutually exclusive events,

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

Therefore, using the independence of A and B, i.e. $P(A \cap B) = P(A)P(B)$, and the complement rule $P(B^c) = 1 - P(B)$,

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)\,[1 - P(B)] = P(A)P(B^c).$$

Hence A and B^c are independent. □

(Remark: applying the result twice, with the roles of the events exchanged, also shows that A^c and B^c are independent.)

(b). (10%) **Claim:** if A and B are mutually exclusive (and $P(A) > 0$, $P(B) > 0$), then A and B are dependent.

Proof. Mutually exclusive means $A \cap B$ contains no sample points, so $P(A \cap B) = 0$. If A and B were independent we would need $P(A \cap B) = P(A)P(B)$; but $P(A)P(B) > 0$ because both probabilities are positive, so $P(A)P(B) \neq 0 = P(A \cap B)$. Thus the independence condition

fails and A and B are dependent. $\square$

Moreover,

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

Interpretation: once B is known to have occurred, A cannot occur; knowing B changes the probability of A from $P(A) > 0$ down to 0, which is exactly why mutually exclusive events (with positive probabilities) are dependent.

Grading notes. (a) 1 point for stating what must be shown ($P(A \cap B^c) = P(A)P(B^c)$); 3 points for the key idea of writing A as the union of the mutually exclusive events $A \cap B$ and $A \cap B^c$ (no hint is given on the exam, so this step carries the most weight); 2 points for $P(A) = P(A \cap B) + P(A \cap B^c)$ (addition law); 2 points for substituting $P(A \cap B) = P(A)P(B)$; 2 points for completing the algebra to $P(A)P(B^c)$ with $P(B^c) = 1 - P(B)$ and stating the conclusion. A numerical example alone receives 0. A proof that assumes $P(A \mid B^c) = P(A)$ without justification receives at most 3 points. An equivalent proof via conditional probabilities, $P(B^c \mid A) = 1 - P(B \mid A) = 1 - P(B) = P(B^c)$, receives full credit provided each step is justified. (b) 3 points for $P(A \cap B) = 0$; 3 points for the contradiction with $P(A)P(B) > 0$ (the positivity of both probabilities must be used); 2 points for $P(A \mid B) = 0$; 2 points for the interpretation.