

Exam 1 — Solution Key (150 pts)

Stat 515 (Fall 2026)

Instructor: Yen-Yi Ho

Rounding convention on this exam: probabilities to four decimal places, all other numerical answers to two IF NEEDED — exact values and counts are written as they are (10, not 10.00; 27,405, not 27,405.00). Accept answers that differ only in the last digit from table rounding.

1. (10 pts) School district homework survey components.

a) (5 pts) Identifications.

Target population: All students enrolled in grades 6–8 in the school district — the group about which the district wants to draw a conclusion.

Sampled population: Students in grades 6–8 who appear on the enrollment roster as of the first day of the semester, whose mailing address on file is current, and whose household actually returns the questionnaire.

Operationally this is the 613 respondents.

Sampling unit: The individual student ID number drawn from the roster (equivalently, the household address to which one questionnaire is mailed).

Observation unit: The individual student who completes the questionnaire.

Response: The number of minutes per day the student spends on homework.

b) (5 pts) Why target and sampled populations may differ.

Three distinct mechanisms separate the two:

1. Frame (coverage) error. The roster is frozen at the first day of the semester. Students who enrolled after that date have zero chance of selection, and students who have since withdrawn are still eligible for selection. The list therefore does not match the grade 6–8 population at the time of the study.

2. Delivery error. The questionnaire is mailed to the address on file. Families that have moved, or whose address is recorded incorrectly, never receive the survey, so they are effectively outside the sampled population.

3. Nonresponse. Only 613 of 900 questionnaires were returned, a response rate of $613/900 \approx 68.1\%$. The 287 nonrespondents are excluded, and

nonresponse is very unlikely to be random here: students who do little or no homework, or whose families are less engaged with school, are plausibly less likely to reply. This makes the sample mean a biased estimate of the target-population mean, and the bias is expected to be upward.

Answer: The sampled population is "students on the first-day roster, reachable at the address on file, who chose to respond" — a narrower and self-selected subset of the target population.

2. (12 pts) Short questions.

1) (3 pts) Which population shows more variability?

$\text{Var}(\text{Population 1}) = 20$ so $\text{SD}_1 = \sqrt{20} \approx 4.47$

$\text{Var}(\text{Population 2}) = 40$ so $\text{SD}_2 = \sqrt{40} \approx 6.32$

Variance measures the average squared distance of observations from their mean, so a larger variance means the observations are spread farther from the center.

Answer: Population 2, because its variance (40) is larger than that of Population 1 (20).

2) (6 pts) Shape of the histogram, and mean vs. median.

The histogram has its tall bars on the right (near 496–500) and a long thin tail extending to the left (488–494).

Answer: The histogram is SKEWED LEFT, and its mean would be SMALLER THAN its median. (3 pts each.)

Reasoning: the few extreme small values in the left tail pull the mean down, while the median, being a positional measure, is resistant to them.

3) (3 pts) Best fit for a Normal distribution with mean 2 and SD 1.

A $N(2, 1)$ curve is symmetric, centered at 2, with essentially all of its mass inside $2 \pm 3(1) = (-1, 5)$.

C1 is symmetric but centered at 0. C3 is strongly right-skewed (exponential-looking), not symmetric at all. C4 is symmetric but centered near 3 and spread over roughly $(-6, 12)$, implying $\text{SD} \approx 3$. C2 is symmetric, centered at about 2, and spans roughly $(-1.1, 5.5)$, i.e. about 6 standard deviations wide with $\text{SD} \approx 1$.

Answer: C2

3. (12 pts) Standard normal percentiles.

In every part, the strategy is the same: convert the statement into a CUMULATIVE area (area to the left), then read the z-value out of the body of the standard normal table.

a) (2 pts) Find z_0 with $P(Z > z_0) = 0.017$.

$$P(Z > z_0) = 0.017 \quad \text{so} \quad P(Z \leq z_0) = 1 - 0.017 = 0.9830$$

Search the body of the table for 0.9830. It appears at row 2.1, column .02.

$$z_0 \approx 2.12 \quad (\text{exact value } 2.1201)$$

Answer: $z_0 \approx 2.12$

b) (4 pts) Find z_1 with $P(|Z| > z_1) = 0.017$.

The event $|Z| > z_1$ means $Z < -z_1$ or $Z > z_1$ — both tails. By the symmetry of the standard normal curve the two tails have equal area, so each one carries half of 0.017.

$$P(Z > z_1) = 0.017 / 2 = 0.0085$$

$$P(Z \leq z_1) = 1 - 0.0085 = 0.9915$$

Search the body of the table for 0.9915. It appears at row 2.3, column .09.

$$z_1 \approx 2.39 \quad (\text{exact value } 2.3867)$$

Answer: $z_1 \approx 2.39$

Common error: dividing after the lookup instead of before it, or using 0.9830 again. Note $z_1 > z_0$ — splitting the same total area between two tails pushes each cutoff farther out.

c) (6 pts) Placement exam: $X \sim N(\mu = 500, \sigma = 100)$. Find x_0 with $P(X > x_0) = 0.017$.

Standardize. The z-score that cuts off an upper tail of 0.017 is exactly the z_0 found in part a), so back-transform it to the X scale.

$$P(X > x_0) = P(Z > (x_0 - 500)/100) = 0.017$$

$$(x_0 - 500)/100 = z_0 = 2.12$$

$$x_0 = 500 + 2.12(100)$$

$$= 500 + 212 = 712$$

Answer: $x_0 = 712$. Only about 1.7% of students score above 712 — a score 2.12 standard deviations above the mean.

4. (6 pts) Selecting a sample.

Order does not matter (the 4 students form an unordered group), and no student can be chosen twice, so this is a combination.

$$\begin{aligned}\text{Number of samples} &= C(30, 4) = 30! / [4! \cdot 26!] \\ &= (30 \cdot 29 \cdot 28 \cdot 27) / (4 \cdot 3 \cdot 2 \cdot 1) \\ &= 657,720 / 24 = 27,405\end{aligned}$$

Because every possible sample is equally likely, each one has probability 1 divided by the number of samples.

$$P(\text{a particular sample}) = 1 / 27,405 \approx 0.0000365$$

Answer: 27,405 different samples; each has probability exactly 1/27,405.

The question asks for an exact fraction because this probability rounds to 0.0000 at four decimal places. As a decimal it is about 3.65×10^{-5} .

5. (8 pts) Visually impaired students.

Let X = hours of sleep, $X \sim N(\mu = 9.06, \sigma = 2.11)$. We want $P(8 < X < 10)$.

$$z_1 = (8 - 9.06) / 2.11 = -1.06 / 2.11 = -0.50$$

$$z_2 = (10 - 9.06) / 2.11 = 0.94 / 2.11 = 0.45$$

$$P(8 < X < 10) = P(-0.50 < Z < 0.45)$$

$$= P(Z < 0.45) - P(Z < -0.50)$$

$$= 0.6736 - 0.3085$$

$$= 0.3651$$

Answer: $P(8 < X < 10) \approx 0.365$, so about 36.5% of these students sleep between 8 and 10 hours on a typical day.

6. (15 pts) Discrete probability distribution.

x	10	12	18	20
p(x)	0.2	0.3	0.1	0.4

a) (5 pts) $E(X)$.

$$E(X) = \sum x \cdot p(x)$$

$$= 10(0.2) + 12(0.3) + 18(0.1) + 20(0.4)$$

$$= 2.0 + 3.6 + 1.8 + 8.0$$

$$= 15.4$$

$$\text{Answer: } E(X) = 15.4$$

b) (5 pts) $\text{Var}(X)$.

$$E(X^2) = 10^2(0.2) + 12^2(0.3) + 18^2(0.1) + 20^2(0.4)$$

$$= 100(0.2) + 144(0.3) + 324(0.1) + 400(0.4)$$

$$= 20 + 43.2 + 32.4 + 160 = 255.6$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = 255.6 - (15.4)^2$$

$$= 255.6 - 237.16 = 18.44$$

$$\text{SD}(X) = \sqrt{18.44} \approx 4.29$$

$$\text{Answer: } \text{Var}(X) = 18.44 \text{ (standard deviation } \approx 4.29)$$

c) (5 pts) $P(X < 15)$.

The values of X strictly less than 15 are $x = 10$ and $x = 12$.

$$P(X < 15) = p(10) + p(12) = 0.2 + 0.3 = 0.5$$

$$\text{Answer: } P(X < 15) = 0.5$$

7. (25 pts) Elderly wheelchair user study.

Home Features	Injurious Fall(s)	No Falls	Totals
All 5	2	7	9
At least 1, but not all	26	162	188
None	20	89	109
Totals	48	258	306

a) (5 pts) **P(injurious fall).**

$$P(\text{Fall}) = 48 / 306 = 0.1569$$

Answer: ≈ 0.157

b) (5 pts) **P(all five features OR no falls).**

These two events overlap — 7 users had all five features and no falls — so use the Addition Rule.

$$\begin{aligned} P(\text{All 5} \cup \text{No Falls}) &= P(\text{All 5}) + P(\text{No Falls}) - P(\text{All 5} \cap \text{No Falls}) \\ &= 9/306 + 258/306 - 7/306 \\ &= 260/306 = 0.8497 \end{aligned}$$

Answer: ≈ 0.850

c) (5 pts) **P(no falls AND none of the features).**

This is a single cell of the table — the intersection, read directly.

$$P(\text{No Falls} \cap \text{None}) = 89 / 306 = 0.2908$$

Answer: ≈ 0.291

d) (5 pts) **P(injurious fall | no features installed).**

Conditioning on "None" restricts attention to that row, whose total is 109.

$$\begin{aligned} P(\text{Fall} \mid \text{None}) &= P(\text{Fall} \cap \text{None}) / P(\text{None}) \\ &= (20/306) / (109/306) = 20/109 = 0.1835 \end{aligned}$$

Answer: ≈ 0.183

e) (5 pts) **Are "no features installed" and "injurious fall" independent?**

Two events are independent if and only if $P(A \mid B) = P(A)$, equivalently $P(A \cap B) = P(A) \cdot P(B)$.

$$P(\text{Fall} \mid \text{None}) = 20/109 = 0.1835$$

$$P(\text{Fall}) = 48/306 = 0.1569$$

$0.1835 \neq 0.1569$, so the condition fails.

Equivalent check with the multiplication rule:

$$P(\text{Fall} \cap \text{None}) = 20/306 = 0.0654$$

$$P(\text{Fall}) \cdot P(\text{None}) = (48/306)(109/306) = 0.1569 \times 0.3562 = 0.0559$$

$0.0654 \neq 0.0559$ — the same conclusion.

Answer: NOT independent. Knowing that a user had none of the five modifications raises the probability of an injurious fall from about 15.7% to about 18.4%, so the two events are associated.

8. (12 pts) Use of laughing gas.

Each of the 5 dentists independently uses nitrous oxide with probability 0.60, so $X \sim \text{Binomial}(n = 5, p = 0.6)$.

a) (4 pts) Expected number.

$$E(X) = n \cdot p = 5(0.6) = 3$$

Answer: $E(X) = 3$ dentists

b) (4 pts) Standard deviation.

$$\text{Var}(X) = n \cdot p \cdot (1 - p) = 5(0.6)(0.4) = 1.2$$

$$\text{SD}(X) = \sqrt{1.2} \approx 1.0954$$

Answer: $\text{SD}(X) \approx 1.10$ dentists

c) (4 pts) $P(\text{exactly 2 use laughing gas})$.

$$P(X = 2) = C(5, 2) \cdot (0.6)^2 \cdot (0.4)^3$$

$$C(5, 2) = 10$$

$$= 10 \times 0.36 \times 0.064$$

$$= 0.2304$$

Answer: $P(X = 2) \approx 0.2304$

9. (16 pts) Cracks in interstate highway.

Let X = number of cracks requiring repair. X follows a Poisson distribution with mean 2 cracks per mile.

a) (5 pts) $P(\text{no crack in one mile})$. Here $\lambda = 2$.

$$P(X = 0) = e^{(-2)} \cdot 2^0 / 0! = e^{(-2)}$$

$$= 0.1353$$

Answer: ≈ 0.135

b) (5 pts) Average number of cracks in 5 miles.

The Poisson mean scales with the length of the interval.

$$\lambda_5 = 2 \text{ cracks/mile} \times 5 \text{ miles} = 10 \text{ cracks}$$

Answer: 10 cracks on average (the variance is also 10)

c) (6 pts) $P(\text{at least two cracks in 5 miles})$. Now $\lambda = 10$.

Use the complement, since "at least 2" is everything except 0 and 1.

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1)$$

$$P(X = 0) = e^{-10} \cdot 10^0 / 0! = e^{-10} = 0.0000454$$

$$P(X = 1) = e^{-10} \cdot 10^1 / 1! = 10 e^{-10} = 0.000454$$

$$P(X \geq 2) = 1 - 0.0000454 - 0.000454$$

$$= 1 - 0.000499 = 0.9995$$

Answer: ≈ 0.9995 — it is nearly certain that at least two cracks require repair over 5 miles.

10. (12 pts) Playing cards.

a) (3 pts) Number of 5-card hands.

Order is irrelevant, so use combinations.

$$\begin{aligned}C(52, 5) &= 52! / (5! \cdot 47!) \\&= (52 \times 51 \times 50 \times 49 \times 48) / (5 \times 4 \times 3 \times 2 \times 1) \\&= 311,875,200 / 120 = 2,598,960\end{aligned}$$

Answer: 2,598,960 hands

b) (3 pts) Hands with exactly 3 Aces.

Choose 3 of the 4 Aces, then 2 of the 48 non-Aces, and multiply (Fundamental Theorem of Counting). "Exactly" 3 means the other two cards must come from the 48 non-Aces, not from all 49 remaining cards.

$$C(4, 3) \times C(48, 2) = 4 \times 1,128 = 4,512$$

Answer: 4,512 hands

c) (3 pts) P(exactly 3 Aces).

$$P = 4,512 / 2,598,960 = 94/54,145 \approx 0.001736$$

Answer: ≈ 0.00174 , i.e. about 1 hand in 576

d) (3 pts) Number of flushes (all five cards the same suit).

Pick the suit (4 ways), then choose 5 of the 13 cards in that suit.

$$4 \times C(13, 5) = 4 \times 1,287 = 5,148$$

Answer: 5,148 hands (probability $5,148/2,598,960 \approx 0.00198$)

11. (12 pts) Software quality control.

Let A, B, C be the events that the project came from Team A, B, C. These three events partition the sample space.

Team	P(Team)	P(Bug Team)	P(Team $\cap$ Bug)
A	0.50	0.04	0.020
B	0.30	0.06	0.018
C	0.20	0.03	0.006
Total	1.00	—	0.044

a) (6 pts) Law of Total Probability.

$$\begin{aligned}P(\text{Bug}) &= P(A)P(\text{Bug}|A) + P(B)P(\text{Bug}|B) + P(C)P(\text{Bug}|C) \\&= (0.50)(0.04) + (0.30)(0.06) + (0.20)(0.03) \\&= 0.020 + 0.018 + 0.006 \\&= 0.044\end{aligned}$$

Answer: $P(\text{Bug}) = 0.044$, i.e. 4.4% of all projects contain a critical bug.

b) (6 pts) Bayes' Rule for $P(B | \text{Bug})$.

$$\begin{aligned}P(B | \text{Bug}) &= P(B) \cdot P(\text{Bug} | B) / P(\text{Bug}) \\&= (0.30)(0.06) / 0.044 \\&= 0.018 / 0.044 \\&= 0.4091\end{aligned}$$

Answer: $P(B | \text{Bug}) \approx 0.409$

Interpretation: Team B produces only 30% of projects, but because its bug rate is the highest, it accounts for about 41% of the buggy ones. (For completeness, $P(A | \text{Bug}) \approx 0.455$ and $P(C | \text{Bug}) \approx 0.136$; the three posteriors sum to 1.)

12. (10 pts) Diagnostic testing.

Let D = "has the disease" and + = "tests positive".

$$P(D) = 0.01 \quad P(\text{no } D) = 0.99$$

$$P(+ | D) = 0.95 \quad P(+ | \text{no } D) = 0.03$$

a) (4 pts) $P(\text{positive test})$.

$$\begin{aligned}
P(+) &= P(D)P(+|D) + P(\text{no } D)P(+|\text{no } D) \\
&= (0.01)(0.95) + (0.99)(0.03) \\
&= 0.0095 + 0.0297 \\
&= 0.0392
\end{aligned}$$

Answer: $P(+) = 0.0392$, i.e. about 3.9% of people tested will get a positive result.

b) (4 pts) Bayes' formula for $P(D | +)$.

$$\begin{aligned}
P(D | +) &= P(D) \cdot P(+ | D) / P(+) \\
&= (0.01)(0.95) / 0.0392 \\
&= 0.0095 / 0.0392 \\
&= 0.2423
\end{aligned}$$

Answer: $P(D | +) \approx 0.242$, or about 24%.

c) (2 pts) Why is this so much smaller than 95%?

Because the disease is rare. Out of 10,000 people tested, only 100 actually have the disease and about 95 of them test positive — but 9,900 are disease-free, and 3% of them, about 297 people, test positive anyway. The false positives (297) outnumber the true positives (95) by roughly three to one, so $95/(95 + 297) \approx 24\%$.

The sensitivity $P(+ | D)$ answers "given the disease, how often is the test positive?" The quantity asked for here, $P(D | +)$, reverses the conditioning, and with a low base rate the two are very far apart. Confusing them is the base-rate fallacy.

Answer: The low prevalence (1%) means false positives greatly outnumber true positives, so a positive result still leaves the person more likely to be healthy than sick.

Point Distribution Summary

Question	Topic	Source	Points
1	Survey components: target/sampled population, units	Exam 1 · HW1 Q1	10
2	Variability, histogram shape, normal shape ID	Exam 1	12
3	Standard normal percentiles: one-tail, two-tail, back-transform	Exam 1 (extended)	12
4	Combinations and equally likely samples	Exam 1	6
5	Normal probability between two values	Exam 1	8
6	Discrete distribution: mean, variance, probability	Exam 1	15
7	Contingency table, conditional probability, independence	Exam 1	25
8	Binomial: mean, SD, exact probability	Exam 1	12
9	Poisson: zero count, scaling, complement rule	Exam 1	16
10	Combinations with playing cards	HW3 P5	12
11	Total Probability Rule and Bayes' Rule	HW3 P3	12
12	Bayes' formula, base-rate reasoning	HW2 Q3	10
	TOTAL		150