Answers - Conditional probability and the Normal distribution

Year: Year 13

Answer to Question 1

Method: for a conditional probability use $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$; two events are independent exactly when $P(A \cap B) = P(A)P(B)$, or equivalently when $P(A \mid B) = P(A)$.

a. Divide the probability of the intersection by the probability of the condition. $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{0.3}{0.5} = 0.6$ and $P(B \mid A) = \dfrac{0.3}{0.6} = 0.5$. So $P(A \mid B) = 0.6$ and $P(B \mid A) = 0.5$.

b. Subtract the intersection so that it is not counted twice. Addition law: $P(A \cup B) = 0.6 + 0.5 - 0.3 = 0.8$. So $P(A \cup B) = 0.8$.

c. Compare the conditional probability with the unconditional one. $P(A \mid B) = 0.6 = P(A)$ (equivalently $P(A)P(B) = 0.6 \times 0.5 = 0.3 = P(A \cap B)$), so the events **are independent**. Knowing that B has happened does not change the probability of A.

Answer to Question 2

Method: for a conditional probability use $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$; two events are independent exactly when $P(A \cap B) = P(A)P(B)$, or equivalently when $P(A \mid B) = P(A)$.

a. Rearrange the addition law $P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$. $P(X \cap Y) = 0.4 + 0.35 - 0.61 = 0.14$. So $P(X \cap Y) = 0.14$.

b. Use the definition of conditional probability with Y as the condition. $P(X \mid Y) = \dfrac{0.14}{0.35} = 0.4$, which equals $P(X) = 0.4$. The two probabilities are equal.

c. Equal conditional and unconditional probabilities is the test for independence. Since $P(X \mid Y) = P(X)$ the events are **independent**. Check: $0.4 \times 0.35 = 0.14$. So knowing that Y happened tells us nothing about X.

Answer to Question 3

a. On the second pick there are only 9 counters left, so the second-stage fractions have denominator 9. First pick: R $\frac{6}{10}$, Y $\frac{4}{10}$. After a red: R $\frac{5}{9}$, Y $\frac{4}{9}$. After a yellow: R

$\frac{6}{9}$, Y $\frac{3}{9}$. Each pair of branches from the same point adds up to 1.

b. Multiply along the branches R then R. $P(RR) = \frac{6}{10} \times \frac{5}{9} = \frac{30}{90} = \frac{1}{3}$. So the probability that both are red is $\frac{1}{3}$.

c. There are two paths: red then yellow, or yellow then red. Add them. $P(RY) + P(YR) = \frac{6}{10} \times \frac{4}{9} + \frac{4}{10} \times \frac{6}{9} = \frac{24}{90} + \frac{24}{90} = \frac{48}{90} = \frac{8}{15}$. So the probability of different colours is $\frac{8}{15}$.

d. This is read straight from the second branch after Y. After a yellow has gone, 9 counters remain and 6 are red, so $P = \frac{6}{9} = \frac{2}{3}$. So the answer is $\frac{2}{3}$.

Answer to Question 4

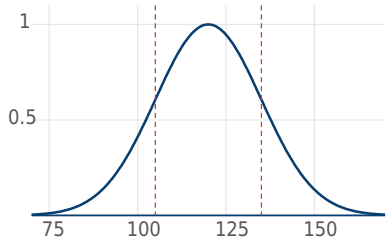
Method: standardise with $Z = \frac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. Standardise: $z = \frac{100 - 120}{15} = -1.333$. $P(X$. So about 9.12% of values are below 100.

b. Standardise: $z = \frac{150 - 120}{15} = 2$, then use the upper tail. $P(X > 150) = P(Z > 2) = 1 - 0.97725 = 0.0228$. So $P(X > 150) = 0.0228$.

c. Standardise both limits. 105 and 135 are one standard deviation either side of the mean, so $P(105$. This is the well-known result that about 68% of values lie within one standard deviation of the mean.

d. Think about what probability means for a continuous variable. X is continuous, so the probability of any single exact value is the area of a zero-width strip, which is 0. So for a continuous model we only find probabilities of intervals.



Answer to Question 5

Method: standardise with $Z = \frac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. The area to the left of a is 0.95, so use the inverse Normal with 0.95. Inverse Normal gives $z = 1.6449$, so $a = 120 + 1.6449 \times 15 = 144.7$. So $a = 144.7$ to 1 decimal place.

b. An area of 0.9 to the right means an area of 0.1 to the left, so a is below the mean. $P(X$, so $z = -1.2816$ and $a = 120 - 1.2816 \times 15 = 100.8$. So $a = 100.8$.
c. Use the inverse Normal with area 0.75; a is the upper quartile of X. $z = 0.6745$, so $a = 120 + 0.6745 \times 15 = 130.1$. So $a = 130.1$.

Answer to Question 6

Method: standardise with $Z = \dfrac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. Use the inverse Normal on the standard Normal distribution with area 0.2 to the left. $z = -0.8416$. The value is negative because 0.2 is less than 0.5.
b. Standardise 40 with the unknown mean and set it equal to the z-value from part a). $\dfrac{40 - \mu}{8} = -0.8416$, so $40 - \mu = -6.733$ and $\mu = 46.73$. So the mean is $\mu = 46.73$ to 2 decimal places.
c. Now that μ is known, standardise 50 and use the upper tail. $P(X > 50) = P\left(Z > \dfrac{50 - 46.73}{8}\right) = P(Z > 0.409) = 0.341$. So $P(X > 50) \approx 0.341$.

Answer to Question 7

a. The first stage is the bus, the second stage is Jaya's arrival; each pair of branches adds up to 1. First branches: on time 0.8, late 0.2. From each, branches: before bell 0.95 / after bell 0.05 (on time) and 0.4 / 0.6 (late). Writing all four outcomes at the ends of the branches helps for parts b) and c).
b. Add the two paths that end with before the bell. $P(\text{before bell}) = 0.8 \times 0.95 + 0.2 \times 0.4 = 0.76 + 0.08 = 0.84$. So the probability that Jaya arrives before the bell is 0.84.
c. Use $P(T \mid B) = \dfrac{P(T \cap B)}{P(B)}$ with the path on time then before the bell. $P(\text{on time} \mid \text{before bell}) = \dfrac{0.76}{0.84} = \dfrac{19}{21} = 0.905$. So it is very likely that the bus was on time.
d. Independent days means the probabilities multiply. $0.8^5 = 0.328$ (3 s.f.). So the bus is on time all week only about one week in three.

Answer to Question 8

Method: for a conditional probability use $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$; two events are independent exactly when $P(A \cap B) = P(A)P(B)$, or equivalently when

$P(A \mid B) = P(A)$.

a. Use the total of the canteen column out of the grand total.
$P(\text{canteen}) = \dfrac{70}{120} = \dfrac{7}{12} = 0.583$. So the probability is $\frac{7}{12} \approx 0.583$.

b. Given canteen, we restrict ourselves to the 70 canteen students. The denominator is the canteen column: $\dfrac{42}{70} = 0.6$. So 60% of canteen users are in Year 13.

c. Given Year 12, we restrict ourselves to the 60 Year 12 students. The denominator is the Year 12 row: $\dfrac{28}{60} = \dfrac{7}{15} = 0.467$. So $P(\text{canteen} \mid \text{Year 12}) \approx 0.467$.

d. Compare the conditional probability with the overall probability.
$P(\text{canteen} \mid \text{Year 12}) = 0.467 \neq P(\text{canteen}) = 0.583$. Equivalently $P(\text{canteen})P(\text{Y12}) = \frac{7}{12} \times \frac{1}{2} = 0.292 \neq \frac{28}{120} = 0.233$. The events are **not independent**. Year 12 students are less likely to use the canteen.

Answer to Question 9

Method: for a conditional probability use $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$; two events are independent exactly when $P(A \cap B) = P(A)P(B)$, or equivalently when $P(A \mid B) = P(A)$.

a. $P(U) = 0.3 \times 0.9 + 0.7 \times 0.2 = 0.27 + 0.14 = 0.41$.

b. $P(R \mid U) = \dfrac{0.27}{0.41} = 0.659$.

c. $P(U^c) = 0.59$ and $P(R \cap U^c) = 0.3 \times 0.1 = 0.03$, so $P(R \mid U^c) = \dfrac{0.03}{0.59} = 0.0508$.

d. The probability of rain rises from 0.3 to 0.659 when he carries an umbrella and falls to 0.051 when he does not, so it is a useful but not perfect predictor.

So taking an umbrella makes rain more than twice as likely, while leaving it at home makes rain very unlikely.

Answer to Question 10

Method: standardise with $Z = \dfrac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. Standardise: $z = \dfrac{185 - 172}{7} = 1.857$, then use the upper tail.
$P(H > 185) = P(Z > 1.857) = 0.0316$. So about 3.2% of students are taller than 185 cm.

b. Standardise both limits: $\dfrac{165 - 172}{7} = -1$ and $\dfrac{180 - 172}{7} = 1.143$. $P(165$. So

about 71.5% of students are between 165 cm and 180 cm.

c. The tallest 5% means an area of 0.95 to the left of the cut-off. $z = 1.6449$, so $h = 172 + 1.6449 \times 7 = 183.5$ cm. So a student needs to be at least 183.5 cm tall.

d. Multiply the proportion by the number of students. $0.05 \times 240 = 12$ students. So about 12 students are invited.

Answer to Question 11

Method: standardise with $Z = \dfrac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. Standardise: $z = \dfrac{2.9 - 3.00}{0.08} = -1.25$. $P(T$. So about 10.6% of bags are rejected.

b. A bag is accepted with probability $1 - 0.1056$; the 5 bags are independent, so multiply. $(1 - 0.1056)^5 = 0.8944^5 = 0.572$. So the probability that no bag in the box is rejected is 0.572.

c. Now the area below 2.9 must be 0.01, so find z by the inverse Normal and rearrange for μ. $P(T$ gives $z = -2.3263$. So $\dfrac{2.9 - \mu}{0.08} = -2.3263$, $\mu = 2.9 + 0.1861 = 3.086$ g. So the factory must raise the mean to about 3.086 g.

Answer to Question 12

a. Add the true positives and the false positives. $P(+) = 0.02 \times 0.9 + 0.98 \times 0.07 = 0.018 + 0.0686 = 0.0866$. So about 8.7% of people test positive.

b. Divide the true positives by all the positives. $P(C \mid +) = \dfrac{0.018}{0.0866} = 0.208$. So only about 21% of people who test positive have the condition.

c. A negative result for someone without the condition has probability 0.98×0.93. $P(-) = 0.9134$ and $P(C^c \cap -) = 0.98 \times 0.93 = 0.9114$, so the probability is $\dfrac{0.9114}{0.9134} = 0.998$. So a negative result is correct 99.8% of the time.

d. Compare part b) with the claim in the newspaper. Only about 21% of positive results are true positives because the condition is rare, so the statement is false. A negative result, however, is very reliable (99.8%). This is why a positive screening result is always followed by a second test.

Answer to Question 13

a. The second probability depends on the first result: after a goal it is 0.8. $0.8 \times 0.8 = 0.64$. So the probability of scoring both is 0.64.

b. There are two paths with exactly one goal; find each one and add. Score then miss: $0.8 \times 0.2 = 0.16$. Miss then score: $0.2 \times 0.6 = 0.12$. Total 0.28. So the probability of exactly one goal is 0.28.
c. Use conditional probability: the path score then miss divided by the total from part b). $\dfrac{0.16}{0.28} = \dfrac{4}{7} = 0.571$. So the probability is $\frac{4}{7} \approx 0.571$.
d. First find the probability of scoring and of missing the second penalty, then use them as the first stage of a new tree. $P(S_2) = 0.64 + 0.12 = 0.76$ and $P(M_2) = 0.24$. $P(S_3) = 0.76 \times 0.8 + 0.24 \times 0.6 = 0.608 + 0.144 = 0.752$. So the probability that she scores her third penalty is 0.752.

Answer to Question 14

Method: standardise with $Z = \dfrac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. Standardise 53 and 63: $z = -1.4$ and $z = 0.6$. The medium grade is what remains. $P(M$. $P(M > 63) = P(Z > 0.6) = 0.2743$. Medium: $1 - 0.0808 - 0.2743 = 0.6449$. So about 8.1% are small, 64.5% medium and 27.4% large.
b. The 6 eggs are independent, so raise the probability of medium to the power 6. $0.6449^6 = 0.0719$. So the probability that all six are medium is about 0.0719.
c. The smallest 10% means an area of 0.1 to the left of the cut-off. $z = -1.2816$, so the cut-off is $60 - 1.2816 \times 5 = 53.6$ g. So eggs below about 53.6 g are rejected.

Answer to Question 15

a. First branches A 0.5, B 0.3, C 0.2; from each, defective (0.03, 0.04, 0.06) and good.
b. $P(D) = 0.5 \times 0.03 + 0.3 \times 0.04 + 0.2 \times 0.06 = 0.015 + 0.012 + 0.012 = 0.039$.
c. $P(A \mid D) = \dfrac{0.015}{0.039} = \dfrac{5}{13} = 0.385$ and $P(C \mid D) = \dfrac{0.012}{0.039} = \dfrac{4}{13} = 0.308$.
d. Machine A produces 38.5% of the defective cases, more than C (30.8%), because it makes many more cases. A high defect rate does not mean the largest number of defects.

Answer to Question 16

Method: standardise with $Z = \dfrac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. Turn each percentage into an area to the left and use the inverse Normal. $P(T$

gives $z = -1.2816$ so $35 - \mu = -1.2816\sigma$. $P(T > 50) = 0.15$ gives $P(T$ and $z = 1.0364$, so $50 - \mu = 1.0364\sigma$. These are two simultaneous equations in μ and σ.

b. Subtract the equations to eliminate μ. Subtract: $15 = (1.0364 + 1.2816)\sigma = 2.318\sigma$, so $\sigma = 6.47$. Then $\mu = 35 + 1.2816 \times 6.47 = 43.29$. So $\mu \approx 43.29$ minutes and $\sigma \approx 6.47$ minutes.

c. Standardise 30 and 45 with the values from part b). $P(30$. So about 58% of journeys take between 30 and 45 minutes.

d. The time allowed must have 99% of journeys below it. $z = 2.3263$, so $t = 43.29 + 2.3263 \times 6.47 = 58.3$ minutes. So the manager should allow about 58 minutes.

Answer to Question 17

a. Start with the intersection, then subtract it from each total. Both: 12. Badminton only $32 - 12 = 20$. Squash only $28 - 12 = 16$. Neither $60 - (20 + 12 + 16) = 12$. The four regions add up to 60.

b. The condition tells us which total to divide by. $P(B \mid S) = \dfrac{12}{28} = \dfrac{3}{7} = 0.429$. $P(S \mid B) = \dfrac{12}{32} = \dfrac{3}{8} = 0.375$. So the two conditional probabilities are different.

c. Read the numbers from the Venn diagram. $P(\text{neither}) = \dfrac{12}{60} = 0.2$. Exactly one sport: $\dfrac{20 + 16}{60} = \dfrac{36}{60} = 0.6$. So 20% play neither sport and 60% play exactly one.

d. Compare $P(B \cap S)$ with $P(B)P(S)$. $P(B)P(S) = \dfrac{32}{60} \times \dfrac{28}{60} = 0.249$ but $P(B \cap S) = \dfrac{12}{60} = 0.2$. Since these differ, the events are **not independent**. Members who play squash are less likely to play badminton.

Answer to Question 18

a. $X \sim B(200, 0.9)$. $np = 180$ and $np(1 - p) = 200 \times 0.9 \times 0.1 = 18$, so $X \approx Y \sim N(180, 18)$.

b. $P(X > 190) = P(X \geq 191) \approx P(Y > 190.5) = P\left(Z > \dfrac{10.5}{\sqrt{18}}\right) = P(Z > 2.475) = 0.0067$.

c. The estimate 0.0067 is about twice the exact value 0.0035. The binomial distribution is skewed because $p = 0.9$ is far from 0.5, so the symmetric Normal curve overestimates the upper tail.

d. The exact probability 0.0035 is below 0.5%, so 200 tickets is acceptable (the Normal estimate would wrongly suggest otherwise).

Answer to Question 19

a. The class boundaries are at $\mu - \sigma$, μ, $\mu + \sigma$. So $P(R$, $P(215$, $P(255$, $P(R > 295) = 0.1587$. Multiplying by 200 gives 31.7, 68.3, 68.3, 31.7.
b. The observed values 31, 69, 64, 36 are all within 5 of the expected values, so the model fits well.
c. $z = 2.3263$ so $r = 255 + 2.3263 \times 40 = 348$ ms.
d. Waiting times are usually positively skewed with a long right tail, not symmetrical, so a Normal curve would not fit.

Answer to Question 20

Method: standardise with $Z = \dfrac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. Standardise: $z = \dfrac{420 - 500}{50} = -1.6$. $P(L$. So about 5.5% of batteries are faulty.
b. A battery is rejected if it is faulty and caught, or good and wrongly rejected; add the two paths.
$P(\text{rejected}) = 0.0548 \times 0.95 + 0.9452 \times 0.02 = 0.05206 + 0.01890 = 0.07096$. So about 7.1% of batteries are rejected.
c. Divide the faulty-and-rejected path by the total from part b).
$P(\text{faulty} \mid \text{rejected}) = \dfrac{0.05206}{0.07096} = 0.734$. So about 73% of rejected batteries are really faulty.
d. This is the complement of the rejection probability for a faulty battery.
$1 - 0.95 = 0.05$. So 5% of faulty batteries get through quality control.

Answer to Question 21

a. Use the addition law with $P(A \cap B) = P(A)P(B)$. Independent:
$P(A \cup B) = P(A) + P(B) - P(A)P(B)$, so $0.8 = 0.5 + P(B) - 0.5P(B)$,
$0.3 = 0.5P(B)$, $P(B) = 0.6$. So $P(B) = 0.6$.
b. First find the intersection from $P(A \cap B) = P(B \mid A)P(A)$.
$P(A \cap B) = 0.2 \times 0.5 = 0.1$, so $0.8 = 0.5 + P(B) - 0.1$ and $P(B) = 0.4$. So $P(B) = 0.4$.
c. Mutually exclusive means $P(A \cap B) = 0$. Mutually exclusive:
$P(A \cup B) = P(A) + P(B)$, so $P(B) = 0.3$. Then $P(A \cap B) = 0$ but $P(A)P(B) = 0.15 \neq 0$, so they are not independent. Two events with non-zero probabilities can never be both mutually exclusive and independent.

Answer to Question 22

Method: standardise with $Z = \dfrac{X - \mu}{\sigma}$, sketch the curve to decide which tail you need, then use the Normal distribution functions of your calculator. For an unknown value, use the inverse Normal function and rearrange $x = \mu + z\sigma$.

a. Write the two percentages as areas to the left: 0.2 below 38 and 0.9 below 78. $z_1 = -0.8416$ and $z_2 = 1.2816$. Then $38 - \mu = -0.8416\sigma$ and $78 - \mu = 1.2816\sigma$. Subtracting: $40 = 2.1232\sigma$, $\sigma = 18.84$, $\mu = 38 + 0.8416 \times 18.84 = 53.86$. So $\mu \approx 53.86$ and $\sigma \approx 18.84$.

b. If 70% pass, then 30% score below the pass mark. $P(M$ gives $z = -0.5244$, so $m = 53.86 - 0.5244 \times 18.84 = 43.98 \approx 44$. So the pass mark is 44.

c. Each student passes with probability 0.7 independently, so use $B(3, 0.7)$: $\binom{3}{2} = 3$ arrangements. Binomial: $3 \times 0.7^2 \times 0.3 = 0.441$. So the probability that exactly two pass is 0.441.