Statistics Revision Solutions

1 (i) $H \sim N(200,\ 12^2)$ and $W \sim N(175,\ 9^2)$

$P(172.5 < W < 174) = 0.0652$ (shown)

(ii) $H_1 + H_2 - 2W \sim N(50,\ 144(2) + 4(81) = 612\ )$

$P\left(H_1 + H_2 - 2W > 0\right) = 0.9784$ (shown)

(iii) $H + W \sim N(375,\ 225)$

$P(H + W > A) > 0.95 \Rightarrow P(H + W < A) < 0.05$

$\therefore\ A < invNorm(0.05,\ 375,\ \sqrt{225}) = 350.38$ and $A_{\max} = 350$ (shown)

2(i) $M \sim N(80,\ 12^2)$ and $W \sim N(56,\ 6^2)$

$T = M_1 + M_2 + M_3 + W_1 + W_2 + W_3 + W_4 \sim N(80 \times 3 + 56 \times 4 = 464,\ 12^2 \times 3 + 6^2 \times 4 = 576)$

$P(T > 500) = 0.0668$ (shown)

(ii) $T' = \left(M_1 + M_2 + M_3\right) - \left(W_1 + W_2 + W_3 + W_4\right) \sim N(16,\ 576)$

$P(T' > 0) = 0.748$ (shown)

3. $P(X > b) = 0.35 \Rightarrow P(X < b) = 0.65$

$b = invNorm(0.65,\ 130,\ 11) = 134.2 \approx 134$ (shown)

$P(X < a) = 0.2 \rightarrow a = invNorm(0.2,\ 130,\ 11) = 120.7 \approx 121$ (shown)

$\therefore\ (a, b) = (121,\ 134)$ (shown)

(i) $\overline{X} \sim N(130,\ \frac{11^2}{2})$

$P(121 < \overline{X} < 134) = 0.573$ (shown)

(ii) $Y \sim B(10,\ 0.45)$

$P(Y \leq 3) = 0.266$ (shown)

4(i) $S \sim N(10,\ 1.5^2)$ and $T \sim N(18,\ 4^2)$

$\therefore\ P(S < 10) \bullet P(T < 15) = 0.1133$ (shown)

(ii) $T - S \sim N(8,\ 18.25)$

$P(T - S > 5) = 0.7587$ (shown)

(iii) $S - \frac{1}{2}T \sim N(1,\ 1.5^2 + \frac{1}{4} \times 16 = 6.25)$

$P\left(S - \frac{1}{2}T > 0\right) = 0.6554$ (shown)

5. $X \sim N(2, 1^2)$ and $Y \sim N(3, 4^2)$

$X_1 + X_2 - 2Y \sim N(-2,\ 66)$

$P(|X_1 + X_2 - 2Y| < 2) = P(-2 < X_1 + X_2 - 2Y < 2) = 0.189$ (shown)

$P_X \sim N(10,\ 5^2)$ and $P_Y \sim N(24,\ (4 \times 8)^2 = 1024)$

$T = \sum_{i=1}^{10} P_{X_i} + \sum_{j=1}^{5} P_{Y_j} \sim N(100 + 120 = 220,\ 250 + 5120 = 5370)$

$P(T > 180) = 0.707$ (shown)

6(i) $X \sim N(320,\ 50)$

$P(X < m) = 15P(X > m) = 15[1 - P(X < m)]$

$\Rightarrow 16P(X < m) = 15$ or $P(X < m) = \frac{15}{16}$

$\therefore\ m = invNorm(\frac{15}{16},\ 320,\ \sqrt{50}) = 330.84 \approx 331$ (shown)

(ii) Let Y be the rv denoting the number of oranges (out of 60) each having mass more than m grams.

Then $Y \sim B(60, \frac{1}{16})$

Since $np = \frac{60}{16} = 3.75 < 5$ and $p = \frac{1}{16} < 0.1$, $Y \sim P_o(3.75)$ approximately.

$P($ more than 56 oranges have mass less than m grams$)$

$= P($ less than 4 oranges have mass more than m grams$) = P(Y < 4) = P(Y \leq 3)$

$= 0.484$ (shown)

7(i)

Possible digit combination	Probability (including permutation)
6 9 9	$\frac{3!}{2!}\left(\frac{1}{10}\right)^3 = \frac{3}{1000}$
7 8 9	$3!\left(\frac{1}{10}\right)^3 = \frac{6}{1000}$
7 9 9	$\frac{3!}{2!}\left(\frac{1}{10}\right)^3 = \frac{3}{1000}$
8 8 8	$\left(\frac{1}{10}\right)^3 = \frac{1}{1000}$
8 9 8	$\frac{3!}{2!}\left(\frac{1}{10}\right)^3 = \frac{3}{1000}$
8 9 9	$\frac{3!}{2!}\left(\frac{1}{10}\right)^3 = \frac{3}{1000}$
9 9 9	$\left(\frac{1}{10}\right)^3 = \frac{1}{1000}$

Total $= \frac{20}{1000} = \frac{1}{50}$ (shown)

(ii) P(3 numbers are identical | sum of 3 numbers is at least 24)

$$= \frac{\left(\frac{1}{1000} + \frac{1}{1000}\right)}{\frac{1}{50}} = \frac{1}{10} \text{ (shown)}$$

(Note that **888** and **999** are the only two combinations which satisfy the numerator, ie the 3 numbers are identical and their sum is at least 24)

8 (i) Number of ways to seat 11 children (without any restrictions) in a circle $= (11-1)! = 10!$

Firstly, number of ways to arrange the 5 boys $= 5!$

Secondly, number of ways to slot the 6 girls between the 5 boys $= {}^6C_5 \times 5! = 6 \times 5!$

$\therefore$ Required probability $= \frac{6 \times 5! \times 5!}{10!} = \frac{1}{42}$ (shown)

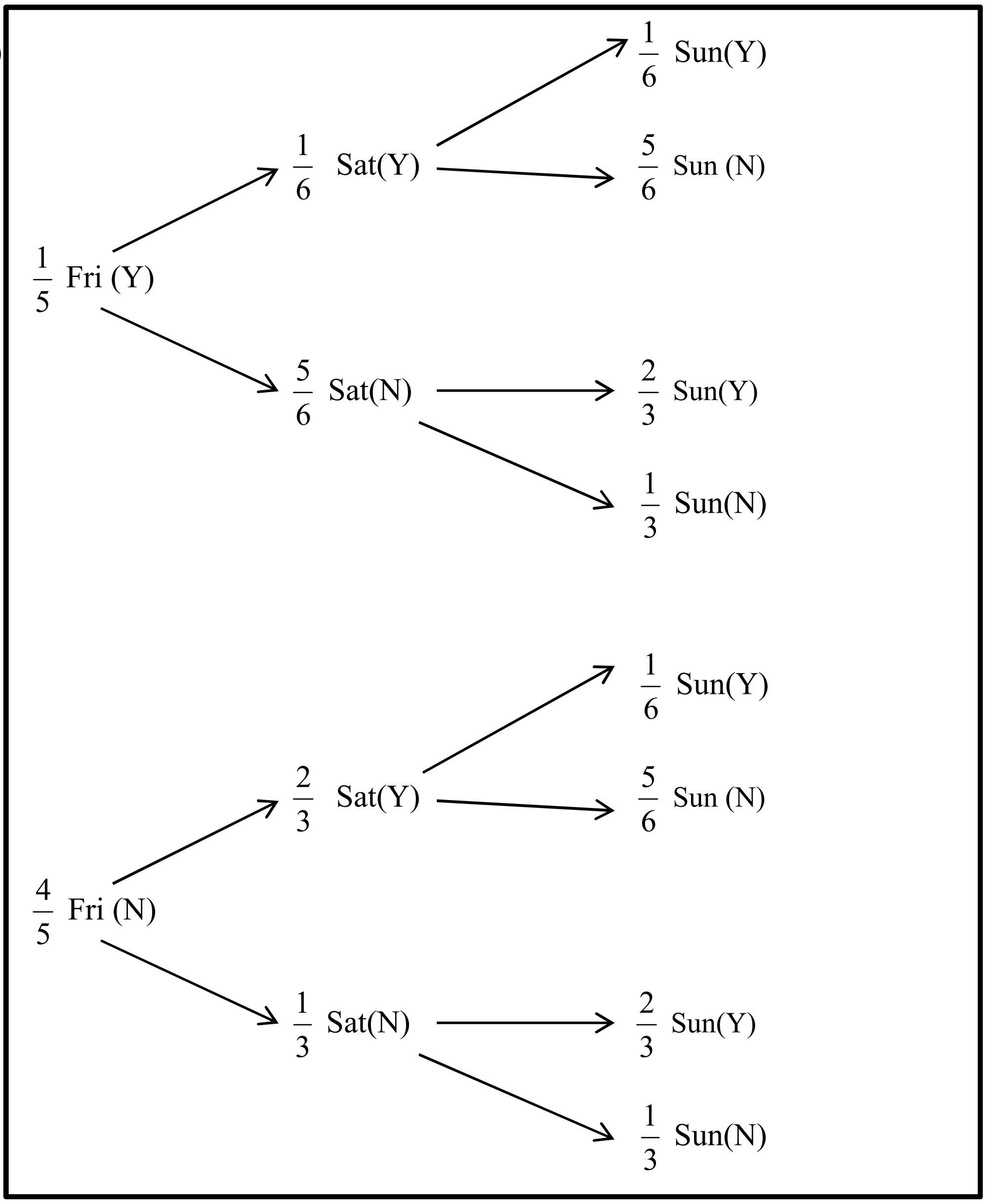
(ii) Grouping J, C and D together as a single unit, number of ways to arrange the children such that J sat between C and D $= (9-1)! \times 2! = 8! \times 2!$ (note that C and D can switch sides)

Number of ways to arrange the children such that J, C and D simply sat together

$= (9-1)! \times 3! = 8! \times 3!$ (note that J, C and D can be freely permutated within the unit)

$\therefore$ Required probability $= \dfrac{8! \times 2}{8! \times 3} = \dfrac{1}{3}$ (shown)

9 (i)



(Y): go on that day (N): not going on that day

P(A goes on Sun) $= \left(\frac{1}{5}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{5}\right)\left(\frac{5}{6}\right)\left(\frac{2}{3}\right) + \left(\frac{4}{5}\right)\left(\frac{2}{3}\right)\left(\frac{1}{6}\right) + \left(\frac{4}{5}\right)\left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{23}{60}$ (shown)

(ii) P(A will go on Fri but not on Sun) $= \left(\frac{1}{5}\right)\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) + \left(\frac{1}{5}\right)\left(\frac{5}{6}\right)\left(\frac{1}{3}\right) = \frac{1}{12}$

P (A will go on Fri) $= \frac{1}{5}$

Hence, P (A will not go on Sun | A will go on Fri) $= \frac{1}{12} \div \frac{1}{5} = \frac{5}{12}$ (shown)

(iii) P(A goes on both Fri and Sun) $= \left(\frac{1}{5}\right)\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{5}\right)\left(\frac{5}{6}\right)\left(\frac{2}{3}\right) = \frac{7}{60}$

Hence, P(A goes on Fri | A goes on Sun) $= \frac{7}{60} \div \frac{23}{60} = \frac{7}{23}$ (shown)

10 (i) Unbiased estimate of population mean $= \frac{249}{100} = 2.49$ (shown)

Unbiased estimate of population variance $= \frac{1}{99}\left[779 - \frac{249^2}{100}\right] = 1.61$ (shown)

(ii) Since $n = 100$ is large, by CLT, $\overline{X} \sim N\left(2.49, \frac{1.61}{100}\right)$ approx

$P(\overline{X} < 2.7) = 0.951$ (shown)

(iii) Since $n = 150$ is large, by CLT,

$T = X_1 + X_2 + ... + X_{150} \sim N(2.49 \times 150 = 373.5,\ 1.61 \times 150 = 241.5)$ approx

$P(T < 370) = 0.411$ (shown)

11. $X \sim N(\mu,\ \sigma^2)$

$P(X < 1) = 0.07 \rightarrow P\left(Z < \frac{1-\mu}{\sigma}\right) = 0.07$

$$\Rightarrow \frac{1-\mu}{\sigma} = invNorm(0.07) = -1.476 ----------(1)$$

$P(X > 1.6) = 0.05 \rightarrow P(X < 1.6) = 0.95 \rightarrow P\left(Z < \frac{1.6-\mu}{\sigma}\right) = 0.95$

$$\Rightarrow \frac{1.6-\mu}{\sigma} = invNorm(0.95) = 1.645 ----------(2)$$

Solving (1) and (2) gives $\sigma = 0.192,\ \mu = 1.284kg$ (shown)

P(exactly 1 large, exactly 1 small, exactly 1 between large and small)

$= 3! \times (0.05)(0.07)(1-0.05-0.07) = 0.0185$ (shown)

12 (i) To test: $H_0 : \mu = 57.5 \qquad H_1 : \mu < 57.5$

$\bar{x} = 56.9,\ \sigma = 2.8,\quad n = 50$

Using the **Z –test**, by the GC, $p = 0.0649 > 0.02$

$\therefore$ H_0 is not rejected and there is **insufficient** evidence at the 2% level to suggest that the company is overstating the mass of coffee per jar. (shown)

(ii) Test statistic $Z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} < invNorm(0.02) = -2.0537$

$$\frac{\bar{x} - 57.5}{\dfrac{2.8}{\sqrt{50}}} < -2.0537 \Rightarrow \bar{x} < 56.7 \text{ (shown)}$$

(iii) Since $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$, then $\bar{X} - \mu \sim N\left(0, \dfrac{\sigma^2}{n}\right)$

$$P\left(\left|\bar{X} - \mu\right| < 0.5\right) = P\left(-0.5 < \bar{X} - \mu < 0.5\right) = P\left(-\frac{0.5}{\left(\dfrac{2.8}{\sqrt{n}}\right)} < Z < \frac{0.5}{\left(\dfrac{2.8}{\sqrt{n}}\right)}\right) = 0.98$$

The above can be re-interpreted as $P\left(Z < -\dfrac{0.5}{\left(\dfrac{2.8}{\sqrt{n}}\right)}\right) = \dfrac{1-0.98}{2} = 0.01$

$$-\frac{0.5}{\left(\dfrac{2.8}{\sqrt{n}}\right)} = invNorm(0.01) \Rightarrow n = 170 \text{ (shown)}$$

(iv) Sample Variance $s^2 = 2.8^2$

Hence unbiased estimate of the population variance $= \dfrac{n}{n-1}s^2 = \dfrac{50}{49}(2.8)^2 = 8$

Test statistic $Z = \dfrac{\overline{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} = -1.5$ (where $\hat{\sigma} = \sqrt{8}$) (shown)

13(i) Since $(\overline{x}, \overline{y})$ lie on the regression line and it is given that $\overline{x} = 1600$,

Then sample mean of weights $\overline{y} = 0.09(1600) - 90 = 54$ (shown)

(ii) Sample variance for heights $= \dfrac{1}{n}\left[\sum (x-\overline{x})^2\right] = 120^2 \Rightarrow \sum (x-\overline{x})^2 = 120^2 n$ ------ (1)

Sample variance for weights $= \dfrac{1}{n}\left[\sum (y-\overline{y})^2\right] = 12^2 \Rightarrow \sum (y-\overline{y})^2 = 12^2 n$ ------ (2)

Also, $b = \dfrac{\sum (x-\overline{x})(y-\overline{y})}{\sum (x-\overline{x})^2} = 0.09$

Substituting (1) into the above gives $\sum (x-\overline{x})(y-\overline{y}) = 0.09(120^2 n)$ ------(3)

Substituting (2) and (3) into $d = \dfrac{\sum (x-\overline{x})(y-\overline{y})}{\sum (y-\overline{y})^2}$ gives $d = \dfrac{0.09(120^2 n)}{12^2 n} = 9$

Equation of line of regression of x on y is $x = c + dy = c + 9y$

Since (1600. 54) lies on the line, $c = 1600 - 9(54) = 1114$

and hence $x = 1114 + 9y$ (shown)

(iii) $r^2 = bd = 0.81 \Rightarrow r = 0.9$ (shown)

A strong value for the coefficient r denotes a high degree of similarity (coincidence) between both regression lines.

(iv) $y = 0.09(1500) - 90 = 45kg$ (shown)

14 (a) $\sum (x-30) = \dfrac{1125}{10} = 112.5,$ $\sum (x-30)^2 = 179671(0.1)^2 = 1796.71$

$\sum (x-30)(y-50) = 13927(0.1) = 1392.7$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{1392.7 - \frac{(112.5)(188)}{10}}{\sqrt{\left(1796.71 - \frac{112.5^2}{10}\right)\left(5226 - \frac{188^2}{10}\right)}} = -0.762 \text{ (shown)}$$

It appears that there is a strong negative linear association between the scores on a fitness test and the weights of the participating students.

(b) $L = a + bW$ ------(1)

$$b = \frac{S_{LW}}{S_{WW}} = \frac{\sum LW - \frac{\sum L \sum W}{n}}{\sum W^2 - \frac{\left(\sum W\right)^2}{n}} = 0.77$$

$$\overline{L} = \frac{400.20}{20} = 20.01, \quad \overline{W} = \frac{176.00}{20} = 8.8$$

Substituting the above quantities into (1), $a = \overline{L} - b\overline{W} = 20.01 - 0.77(8.8) = 13.234$

Hence, equation of regression line is given by $L = 13.234 + 0.77W$ (shown)

15. $\overline{X} \sim N\left(4, \frac{3.6}{n}\right)$

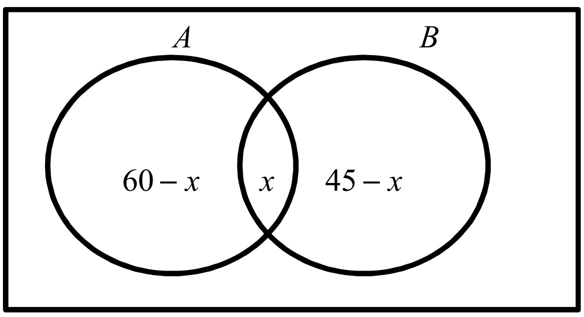
$$P\left(\overline{X} > 4.1\right) < 0.01 \rightarrow P\left(\overline{X} < 4.1\right) > 0.99$$

$$P\left(Z < \frac{4.1-4}{\sqrt{\frac{3.6}{n}}}\right) > 0.99 \Rightarrow \frac{4.1-4}{\sqrt{\frac{3.6}{n}}} > invNorm(0.99) = 2.326$$

Solving gives $n > 1947.6 \Rightarrow n \geq 1948$ (shown)

Yes, central limit theorem (CLT) was used in approximating $\overline{X}$ to a normal distribution based on the fact that the sample size (number of observations) n was large. (shown)

16(a)



(i) $P(A) = \frac{60}{125} = \frac{12}{25}$ (shown)

(ii) Number of boys that are shortsighted $= 65 - 20 = 45$ (this is represented by $(A \cup B)'$)

$$(60 - x) + (x) + (45 - x) + 45 = 125 \Rightarrow x = 25$$

$$A \cup B = 60 + (45 - 25) = 80$$

$$\therefore \; P(A \cup B) = \frac{80}{125} = \frac{16}{25} \text{ (shown)}$$

(iii) $P(A \cap B) = \frac{25}{125} = \frac{1}{5}$ (shown)

(iv) $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\left(\frac{1}{5}\right)}{\left(\frac{45}{125}\right)} = \dfrac{5}{9}$ (shown)

(b) (i) $\left(\frac{45}{125}\right)\left(\frac{44}{124}\right)\left(\frac{43}{123}\right) = \frac{1419}{31775}$ (shown)

(ii) $\left(\frac{3!}{2!}\right)\left(\frac{45}{125}\right)\left(\frac{80}{124}\right)\left(\frac{79}{123}\right) = \frac{2844}{6355}$ (shown)

(iii) $(3!)\left(\frac{25}{125}\right)\left(\frac{35}{124}\right)\left(\frac{20}{123}\right) = \frac{70}{1271}$ (shown)

shortsighted girl | non-shortsighted girl | shortsighted boy

17(i) Let X be the random variable denoting the number of days a student is late from Monday to Wednesday.

Then $X \sim B(3,\ 0.09)$

Required probability $= P(X = 1) \bullet (0.009) = 0.000239$ (shown)

(ii) $P($ student is not late on 5 days$) = (1 - 0.009)^5 = 0.956$

Let Y be the random variable denoting the number of students (out of a class of 25) who are not late for a stretch of 5 days.

Then $Y \sim B(25,\ 0.956)$ and $P(Y = 25) = 0.323$ (shown)

(iii) $P($ at least 1 latecomer in a class of 25 in a given school week of 5 days)

$= 1 - 0.323 = 0.677$

Let Y' be the random variable denoting the number of weeks (out of 80 weeks) where there is at least one latecomer.

Then $Y' \sim B(80,\ 0.677)$

Since $np = 54.16 > 5$ and $nq = 25.84 > 5$,

$Y' \sim N(54.16,\ 17.494)$ approx.

$P(Y' > 50) = P(Y' > 50.5) = 0.809$ (continuity correction has to be used) (shown)

18 (a) $X \sim N(30,\ 0.2^2)$

$T = X_1 + X_2 + X_3 + X_4 + X_5 \sim N(150,\ 5 \times 0.2^2 = 0.2)$

$P(150.5 < T < 151.5) = 0.1314$ (shown)

(b) $Y \sim N(10,\ 0.15^2)$

$S \sim N(6 \times 10 = 60,\ 6 \times 0.15^2 = 0.135)$, $W \sim N(2 \times 30 = 60,\ 4 \times 0.2^2 = 0.16)$

and $S - W \sim N(0, 0.295)$

$P(S < W + 1) = P(S - W < 1) = 0.9672$ (shown)

(c) $\bar{L} \sim N\left(30, \dfrac{0.2^2}{25} = 1.6 \times 10^{-3}\right) \to \bar{L} - 30 \sim N(0,\ 1.6 \times 10^{-3})$

$P\left(\left|\bar{L} - 30\right| \geq a\right) = 0.02 \to P\left(\bar{L} - 30 \leq -a\right) + P\left(\bar{L} - 30 \geq a\right) = 0.02$

Since the distribution of $\left(\bar{L} - 30\right)$ is centred at zero,

$P\left(\bar{L} - 30 \leq -a\right) = 0.01$

$$P\left(Z \leq \frac{-a}{\sqrt{1.6 \times 10^{-3}}}\right) = 0.01$$

$P(Z \leq -25a) = 0.01$

$\therefore\ -25a = invNorm(0.01) = -2.3263$

$\Rightarrow a = 0.093$ (shown)