# Scholarship Statistics and Modelling (93201)

## Evidence Statement

**QUESTION ONE**

**Tasks: Q1 (a) (i) & Q1 (a) (ii)**

**Evidence (i):**

Let $x$ = number of trays of green
and $y$ = number of trays of gold

$x + y \leq 2000$
$x \leq 1200$
$y \geq 200$
$2x \leq 5y$

Objective function is: Profit = $12mx + 13my$ where m is a multiple constant

Optimum point is $x = 0$ and $y = 2000$

**So 0 trays of green and 2000 trays of gold should be packed daily to maximise the profit.**

**Judgement:**

S + P: Optimal point correct along with correct objective function and all constraints correct.
S: Optimal point correct.
P: Constraints (3 out of 4) correctly formed.

**Evidence (ii):**

1. It's highly unlikely that all 2000 gold would be sold. We know that we will sell some green due to demand. In fact if $x = 600$ and $y = 1400$, we would not have to sell about 50 gold for this to be a more profitable combination.
2. Objective function is almost parallel to constant boundary line $x + y = 2000$. Small differences of $x$ and $y$ from the optimal value don't affect the profit significantly.

**Judgement:**

O: Evidence implies that some green will be sold, so profit will be close to the maximum and more likely occur – no P awarded for this question part. No credit for (1200, 800) point.

**Task: Q1 (b)**

**Evidence:**

Let $x$ = number of "all cardboard" trays
Let $y$ = number of "cardboard/wood" trays
Let $z$ = number of "wood" trays

Cost: $x + 2y + 4z = 5000$
Total: $x + y + z = 2000$
Limits: 10% of 2000 $\leq x \leq$ 30% of 2000

Get: $y = 3000 - 3z$ and $x = 2z - 1000$

Now for $y$ to be the largest $z$ must be a minimum and $x$ must be the smallest.

**So Shedz's daily requirement is $x = 200$, $z = 600$ and $y = 1200$**

**Judgement:**

S: Correct answers for $x$, $y$ and $z$.
P: Equations (2 out of 3) correctly formed.

Score S if answer is such that not all budget spent but equations for total and limits hold.

**QUESTION TWO**

**Tasks: Q2 (a) (i) and Q2 (a) (ii)**

**Evidence (i):**

$\lambda = np$ and at 1% $\lambda = 0.5$ and at 2% $\lambda = 1.0$

Or using Binomial, Pr (x $\leq$ 1) = 0.9106 for 1% and Pr(x $\leq$ 1) = 0.7358 for 2%

Pr (at least one load is accepted) = 1 – pr (no loads are accepted)
= 1 – (1 – 0.9106) × (1 -0.7358)
= 1 – 0.0894 × 0.2642
= 1 – 0.024
**= 0.976**

**Judgement:**

(i) S: Probability calculated correctly.
P: Poisson distribution and $\lambda$ correctly identified or Binomial distribution and $\pi$ correctly identified.

**Evidence (ii):**

p = 0.01 and $\lambda = np = 0.01n$

From Poisson tables, find $\lambda$ where pr($x$ = 0) + pr ($x$ = 1) = 0.94
This gives $\lambda = 0.4$ so $n$ = 0.4/0.01 **= 40**

or solve for $\lambda$ : $e^{-\lambda} + \lambda e^{-\lambda} = 0.94$ which gives n = **39**

or use binomial probabilities where $n$ = 40 gives probability = 0.939 and $n$ = 39 gives probability = 0.942 with $n$ = **40** being slightly closer.
Note: $n \times 0.01 \times 0.99^{n-1} + 0.99^{n} = 0.94$ and solve for n.

**Judgement:**

(ii) O: n correctly calculated.
P: Identification of correct method or P: Not a whole number.

**Tasks: Q2 (b) (i) & Q2 (b) (ii)**

**Evidence (i):**

Select a pallet at random from 1 to 20. With chosen pallet, select four levels out of the thirty (use random numbers from 1 to 30). Sample every tray in the chosen four levels to give a 6 x 4 = 24 tray sample as required.

**Evidence (ii):**

Compared to simple random sampling:

Advantage – Trays occur naturally in groups of 6, easy to get 24.
Disadvantage – sample may not be representative of population i.e. pick up defective pallet – last one packed.

**Judgement:**

S: Both (i) and in (ii) one of an advantage or a disadvantage correct consistent with answer to (i).
P: Either (i) or in (ii) one of an advantage or a disadvantage correct consistent with answer to (i).

Note: Must use random somewhere in answer.
Can't have 600 clusters (20 x 30) as answer.

**QUESTION THREE**

**Standards: 90643, 90646**

**Task: Q3 (a)**

**Evidence:**

Let $W = K_1 + \ldots\ldots\ldots\ldots\ldots + K_{36} + T$
So $E(W) = E(K_1) + \ldots\ldots\ldots + E(K_{36}) + E(T)$

So $E(W) = 36 \times 105.6 + 499.7 =$ **4301.3 g**

$Var(W) = var(K_1) + \ldots\ldots\ldots\ldots\ldots + var(K_{36}) + var(T)$ by **assuming the weights of the kiwifruit are independent and the weight of the tray is independent of the weight of the fruit.**

So $var(W) = 36 \times (4.2)^2 + (19.7)^2 = 1023.13$
So $\sigma(W) =$ **31.99 g**

**Judgement:**

S: Two of mean, standard deviation or assumption correct.
P: One of mean, standard deviation or assumption correct.

Note: For assumption keywords are “independent” and “weight”. Accept “weights are independent”.

**Task: Q3 (b)**

**Evidence (i):**

$\mu_{\bar{x}} = \mu = 4301.3$ g and $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{31.99}{\sqrt{6}} = 13.06$g

Given that W is normal implies that $\overline{W}$ is normal.

So 95% limits are $4301.3 \pm 1.96 \times 13.06 = 4301.3 \pm 25.6$
So limits are **4275.7 g to 4326.9 g**

**Evidence (ii):**

$\bar{x} = \dfrac{26}{6} = 4.3333\text{kg} = 4333.3\text{g}$
or $pr(\bar{x} > 4333.3) = pr(z > 2.45) = 0.0071$ which lies in the upper 2.5% tail.

**So grower’s weight of packed trays is (outside upper limit so overweight/ different from the population mean).**

**Judgement:**

S + P: Both (i) and (ii) correct.
S: (i) correct.
2P: Limits for total calculated instead of mean and 26kg compared to it correctly.
P: Either (ii) correct / conditionally correct or $\pm 25.6$g correct in (i).

**Task: Q3(c)**

**Evidence:**

Let $n$ = sample size so then:

$\Pr(4301.3 - 5 < \overline{w} < 4301.3 + 5) / 0.95 = 0.89$

$\Pr(4296.3 < \overline{w} < 4306.3) = 0.8455$

Now $\overline{w}$ is normal as $w$ is normal in (b).

So pr $\left( \dfrac{-5}{31.99/\sqrt{n}} < z < \dfrac{5}{31.99/\sqrt{n}} \right) = 0.8455$

So from the inverse normal: $\dfrac{5}{31.99/\sqrt{n}} = 1.424$ which gives $\boldsymbol{n = 83}$

**Judgement:**

O: Sample size n correctly calculated.
P: Conditional probability clearly indicated in answer.

**QUESTION FOUR**

**Task: Q4 (a)**

**Evidence:**

We have the following features of the relationship:

1. Three outliers
2. Ignoring outliers we have a moderate / reasonably strong (linear) relationship.
3. The residual plot shows the effect of the three outliers with some positive upward trend indicated as $W$ increases in value (some evidence of a pattern).
4. Ignoring outliers, for each 1g increase in weight, the average width of mark increases by 0.3mm.
5. Ignoring outliers, there is a suggestion of non-constant scatter with slightly more variation in width of mark for fruit weighing between 98 g and 103 g than for fruit weighing between 106 g and 111 g.
6. Excluding the outliers, the fitted line to the data shows that the variation in the width of the mark is 85% explained by the variation in the weight.

**Judgement:**

S: At least three features correctly stated.
P: One or two features correctly stated.

**Task: Q4 (b)**

**Evidence:**

Choose $S = 0.2724W - 24.293$ as it fits the data best excluding the outliers (highest $R^2$ value)

When $W = 89$g, $S$ is predicted as **– 0.05 mm**
When $W = 115$g, $S$ is predicted as **7.0 mm**

When $W = 89$ the prediction isn't possible as line no longer applies. When $W = 115$, the prediction is good as we are just outside the data range and the correlation is high and positive.

**Judgement:**

$S + P$: Both predictions and validity statements correct with choice of line justified.
$S$: Both predictions plus one validity statement correct with choice of line justified.
$P + P$: Both predictions correct
$P$: One prediction correct.

Note: Wrong choice of line score as 2P maximum so stop at two predictions correct. That is for W = 89g, S = 1.52mm and for W = 115g S = 5.29mm. Score MEI for no units.

**Task: Q4(c)**

**Evidence:**

Look at other factors influencing the size of the mark like position of vine in orchard, fertilizer used, position of fruit on vine and proximity of shelter belts.

**Judgement:**

O: Improvement given with at least two distinct factors – no P awarded for this question part.

Note: Investigate outliers separately counts as one improvement but "take more samples" doesn't count. Also "other factors affecting the size of the mark should be investigated" isn't sufficient.

**QUESTION FIVE**

**Task: Q5 (a)**

**Evidence:**

**Essay Points:**

1. Day shift mean is 1.38 lower and the number of defectives display less spread and so are more consistent.
2. Both median (50% higher) and range (2 more) are higher for the night shift.
3. Overall trends appear to be steady for the day shift and upwards for the night shift.
4. Upwards trend appears to start at about 9.30pm on the night shift whereas there is no discernable trend for the day shift.
5. The number of defectives appears to fluctuate in a similar fashion for both shifts.
6. Fitting a trend line by eye appears to show an approximate upwards gradient of 3/32 = 0.1 for the night shift and a slight negative gradient of –1/32 = –0.03 for the day shift.
7. As regards the quality requirement, 2.5 % of the sample of 72 corresponds to 1.8 defectives. The day shift has a mean of 1.75 which is less than 1.8 so the quality requirement is being met. For the night shift, the mean 3.13 > 1.8 so the quality requirement isn't being met.

   Notes: 1. The mean is being taken as a measure of overall quality.
   2. Can combine any two distinct measures for points 1 and 2.

**Judgement:**

2S: five points – must include point 7 otherwise score S + P
S + P: four points.
S: three points
2P: two points
P: one point

**Task: Q5 (b)**

**Evidence:**

**Comparison of Shifts:**

1. Select at random samples (of at least 30) of "two tray" samples from each shift over several days.
2. Record the number of defectives found in each "two tray" sample.
3. Calculate the mean and standard deviation of the number of defectives separately from each shift.
4. Construct a confidence interval for the difference in population means: $\mu_{night} - \mu_{day}$
5. If 0 lies within that interval conclude no significant difference in the mean numbers of defectives from each shift otherwise conclude a significant difference.

**Assumptions:**

1. Both sets of observations are representative samples from day and night shifts generally. Assume sampling randomly ensures this.
2. The process of identification of defectives is consistent.

**Judgement:**

O: Points 1 to 5 implied plus one assumption.
P: Any three points – could include one assumption as a point.

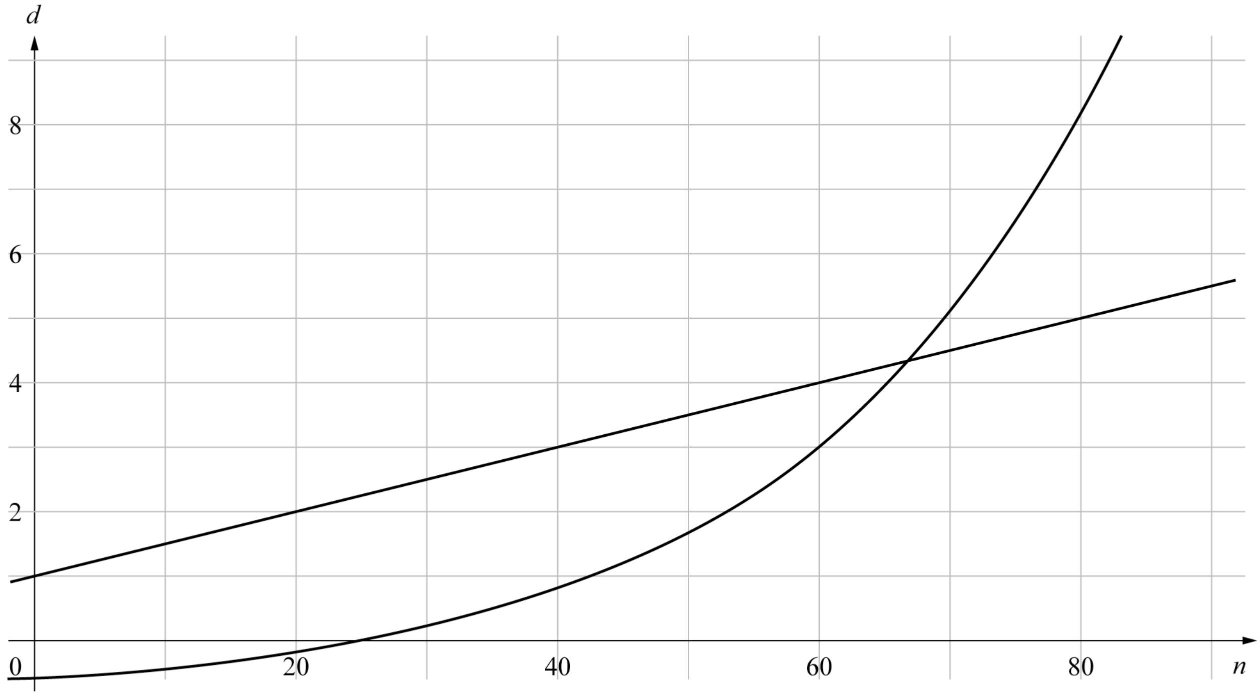
Note: Accept comparison of two CI's, one for the day and the other for the night. Look for overlap.
Assumption contained within the answer is acceptable.

**QUESTION SIX**

**Task: Q6 (a) (i)**

**Evidence:**

For a range of $n$ values: $0 < n < 80$ plot $d$ versus $n$ for $d = 0.05n + 1$ and $d = 0.37e^{0.04n} - 1$




**Task: Q6 (a) (ii)**

Curve crosses line at $n = 67$. At $n = 65$, step 1 gives $d > 4.25$ for rejection and $d < 3.98$ for acceptance so re-sampling would occur at $d = 4$.

***n* = 70** is when a decision to accept or reject must occur. Then step 1 with $d > 4.5$ for rejection and step 2 with $d < 5.08$ for acceptance. However step 1 is applied first so $\boldsymbol{d_{\min} = 5}$.

**Judgement:**

S + P: (i) correct and both n and d correct.
S: (i) correct and both n = 66/67 and d = 5.
2P: (i) correct and d correct only.
P: (i) correct only.
Note: If graph incorrect can score up to 2P with n = 66/67 or 70 and d =5.
Accept graph if only first quadrant shown with position of n (n =24.86) intercept clear.
Accept graph if intercept of 1 shown and others in approx right places i.e n = 25 for axis and n = 65 to 70 for line and curve intercept.

**Task: Q6 (b)**

**Evidence:**

Needs to show n = 20 followed by n = 5.
Number of defective fruit at $n = 25$ is the sum T of two binomial random variables X and Y where X is Bin (n=20, $\pi = 0.04$) and Y is Bin ($n = 5$, $\pi = 0.04$).
T = 0 to be accepted at n =25.
Pr (accepted at n=25) = Pr (T = 0) = Pr(X= 0).Pr(Y= 0) = 0.4420x 0.8154 = **0.3604**

**Judgement:**

O: Correct probability with evidence of correct method.
P: Evidence of correct method like a probability tree or pr(X=0) worked out.
Note: Using Bin (n = 25, $\pi = 0.04$, x = 0) scores P. Write RAWW.

**Task: Q6(c)**

**Evidence:**

Compute a confidence interval for a proportion: $p \pm z\sigma_p$

So we get $\frac{2}{20} \pm 1.96\sqrt{\frac{0.1 \times 0.9}{20}}$ = 0.1 $\pm$ 0.13 i.e. (– 0.03 to 0.23)

**As 2.5 %( 0.025) lies within this 95% confidence interval for the true proportion of defective fruit we would not conclude that the true proportion of defective fruit is different from 2.5%.**

**Judgement:**

S: Correct justification with a calculation.
P: Idea of calculating a confidence interval or pr(x $\geq$ 2 = 0.08824) using a hypothesis test idea.

Note: pr(x $\geq$ 2) = 1 – $(0.975)^{20}$ - 20(0.025) $(0.975)^{19}$ with n = 20 and $\pi$ = 0.025.