Assessment Schedule – 2008

Scholarship Statistics and Modelling (93201)

Evidence Statement

General Principles:

1. Ignore incorrect answers if alongside correct answers.
2. Ignore minor copying errors e.g. σ = \$9.73 in Q1 (a).

QUESTION ONE

Tasks Q1 (a)

Evidence:

E = \$4
Z = 2.576
σ = \$9.53

So $n = \left(\frac{z\sigma}{E}\right)^2 = \left(\frac{2.576 \times 9.53}{4}\right)^2 = \mathbf{38}$

Judgement:

S: Value of n correct
P: Correct working leading to n = 37
P: Correct working with E = \$2 get n = 151
P: n > 38

Note: Answer rounded upwards to ensure E $\leq$ \$4
MEI: n $\geq$ 38
Wrong if n $\leq$ 38

Tasks Q1 (b)

Evidence:

$\bar{x}$ = \$56.24

Now to construct a 95% confidence interval for the mean:

$\mu = 56.24 \pm 1.96 \times \frac{9.53}{\sqrt{38}}$

So $\mu = 56.24 \pm 3.03$

So \$53.21 < μ < \$59.27

Now 4800/80 = \$60 which is just outside the upper limit of the 95% confidence interval. Yes there is enough evidence to suggest that the mean amount spent per person has changed (decreased).

Judgement:

S: Correct calculation + justification
S: If 60 is used instead of 56.24 in interval and 56.24 is compared to interval (\$56.97 < μ < \$63.03).
P: Correct confidence interval.
P: Interpretation consistent with incorrect interval calculation (i.e. using n = 80 but not CI for μ_1 - μ_2).

Note: "carried error" if n incorrect
Alternative answers acceptable e.g. hypothesis test
CI for μ_1 - μ_2 not accepted

Tasks Q1 (c)

Evidence:

- Select a random sample of amounts spent by lunch and dinner diners separately.
- Calculate the mean and standard deviation of these amounts.
- Construct a confidence interval for $\mu_1 - \mu_2$ where μ_1 = mean expenditure of the population of lunch diners and μ_2 = mean expenditure of the population of dinner diners, and
- If zero lies within the calculated confidence interval we conclude that there isn't enough evidence to suggest that the means are different **OR** if zero doesn't lie in the confidence interval we conclude that there is enough evidence to suggest that the means are different.

Assumptions:

- We assume samples are independent which means that the amount spent at lunch is independent of the amount spent at dinner. (Distributions are independent).
- Both sample standard deviations are good approximations of the population standard deviation.
- Random samples of amounts spent selected on one day form a good representative sample of the respective population of amounts spent on lunch and dinner.

Judgement:

O: Four points plus at least one assumption
O: Random missing from first dot point but all else correct.
P: Any three points – could include an assumption as a point.

Tasks Q2 (a)

Evidence:

- Identify each guest with a number from a listing on the guest database of size N.
- Select a starting point at random.
- Choose every kth person until the required sample size n is obtained.
- If required, cycle back to the start when the end of the listing is reached **OR** k = next integer $\geq \frac{N}{n}$

An advantage is that this sampling method is quick and cheap as you don't have to select random numbers for each sample member selection.

A disadvantage is that if there is some periodic variation in the population then your sample could be biased by being non-representative of the population.

Judgement

S: Four points plus one of advantage or disadvantage.
P: Three points could include one of advantage or disadvantage as a point.

Note: Advantages and disadvantages must be justified and only count as one point.

Tasks Q2 (b)

Evidence:

E = 3%
Z = 1.96
p = 0.78
n = 50

So construct a 95% confidence interval: Population proportion = $0.78 \pm 1.96 \times \sqrt{\frac{0.78 \times 0.22}{50}}$

So based on the pilot: Population proportion = 0.78 ± 0.115

Possible Sample Sizes (two of the following)

1. $n = \left(\frac{1.96}{0.03}\right)^2 \times 0.78 \times 0.22 = \underline{733}$ based on a prior estimate of the proportion as 0.78 which is the point estimate from the pilot sample.
2. $n = \left(\frac{1.96}{0.03}\right)^2 \times 0.665 \text{ x } 0.335 = \underline{951}$ based on a prior estimate of the proportion as 0.67 which is the lowest point of the 95% confidence interval closest to 0.5 (worst case scenario).
3. $n = \left(\frac{1.96}{0.03}\right)^2 \times 0.895 \text{ x } 0.105 = \underline{402}$ based on a prior estimate of the proportion as 0.89 which is the upper limit of the 95% confidence interval (best case scenario).

Judgement

O: Two sample size calculations with associated explanations for differences.
P: One sample size with associated explanation OR two sample sizes with no associated explanations.
P: Two sample size calculations with associated explanations for differences with E = 1.5%.

Note: If confidence interval rounded differently allow "carried error".
If one out with n value, score P.
Can use "variability of sample proportion" as part of an explanation.

Tasks Q2 (c)

Evidence:

n = 56
π = 0.05
Use Poisson approximation to the Binomial with λ = 56 x 0.05 = 2.8
So calculating cumulative probabilities we get:

x	Prob($\leq x$)
4	0.8477
5	0.9349
6	0.9756

If we use Binomial directly with n = 56 and π = 0.05 we get:

x	Prob($\leq x$)
4	0.8526
5	0.9398
6	0.9788

We would expect **5 incorrectly completed questionnaires** 95% of the time.

Judgement

Accept either 5(closest to 0.95) or 6 (ensures at least 0.95)
S: Correct answer with evidence.
S: Can use an interval approach for the sample proportion with upper limit of 0.107×56 to get 6 (rounded up).
P: Identification of probability distribution with correct parameters.
Note: Can use a Normal distribution approach with μ = 2.8 and variance σ^2 = 2.8.

Tasks Q3 (a)

Evidence:

Actual Trend
- The actual % occupancy rates rose steadily for 3.5 years from January 2003 then they started to drop.
- Overall the actual % occupancy rate went below 65% in Q3 2003.

Seasonal Patterns

- The actual % occupancy rates peak in either the first or the second quarter. Q1 for 2003, 2004, 2006 and 2007. Q2 for 2005.
- The actual % occupancy rates trough in either the second or the third quarter. Q2 for 2004. Q3 for 2003, 2005, 2006 and 2007.

Deseasonalised

- The deseasonalised % occupancy rates follow an upwards trend until the second quarter in 2006 after which they start to drop.
- Overall the deseasonalised % occupancy rates fluctuated between 67% and 80%.

Judgement

S: Three distinct points about % occupancy rates, one about each of actual trend, seasonal patterns and deseasonalised % occupancy rates.
P: One or two points about occupancy rates.

Tasks Q3 (b) (i) and Q3 (b) (ii)

Evidence (i):

Extrapolating the deseasonalised graph we get 70% corresponding to t = 21.

$$S\,(Q1) = \frac{2.5+3.5+1.7+1.6+3.2}{5} = 2.5$$

Forecast = 70 + 2.5 = **72.5%** (Range 70.5% to 74.5%)

Note: Can have any reasonable value in the range 68 to 72% when extrapolating

Evidence (ii):

When t = 21, y = 0.4496(21) + 69.679 = 79.1

Forecast = 79.1 + 2.5 = **81.6%**

Judgement

S: Both (i) and (ii) correct.
P: One of (i) or (ii) correct.
P: Incorrectly calculated S everything else right.

Tasks Q3(c)

Evidence:

The best forecast to use is 71.5% as its calculation is based on the most recent trend which is downwards. The other forecast is based on the trend equation which gives an overall upwards trend as earlier data has been given more weight in its calculation so it doesn't take into account the recent downwards trend over the last 4 to 5 quarters so it's likely to be an overestimate of the occupancy rate.

Judgement

O: All three points (underlined) are clearly explained
P: Two points.

Note: Can have "carried error" from Q3 (b).

Tasks Q4

Evidence:

Points for essay:

Section A

1. Occupants of deluxe rooms spend more on average with greater variability than those occupants in standard rooms.
2. For the deluxe rooms there is one outlier at (2,180).
3. The spending on the minibar in the deluxe rooms is highly positively (linear) correlated with the number of nights (excluding the outlier).
4. Spending on the minibar in the deluxe rooms is 94.83% explained by the number of nights (excluding the outlier).
5. The spending on the minibar in the deluxe rooms increases by $8.85 on average per night.
6. The relationship between the total minibar expenditure by the number of nights for the standard room is better modelled by a non-linear relationship compared to a linear relationship.
7. For the standard rooms, the amount of the increase in spending on the minibar decreases as the number of nights increase.

Section B

8. Prediction: Deluxe Room $y = 8.85(7) + 5.53 = \$67.48$ – line better fits points overall or $R^2 = 0.9483 > 0.3277$.
9. Prediction: Standard Room $y = 5.58\ln 7 + 5.23 = \$16.09$ – curve better fits points overall or $R^2 = 0.9274 > 0.7674$.
10. Validity: Both predictions are very good with high correlation (fit) to the line for the deluxe rooms and a close fit of the data points, with only a small amount of random scatter, to the curve for the standard room. In both cases $x = 7$ is within the data range.
11. Other factors, two required, to consider for improving the model would be number staying in each room, amount of time spent in hotel room, investigation of the reasons for the outlier at (2,180), types of snacks and drinks available on the minibar and purpose of stay, business versus pleasure and gender (subgroups).

Judgement

O + 2S: Any four points from section A and all four points from section B.
2S + P: Any four points from section A and three points from section B.
2S: Any four points in section A and one from 8 or 9 and one from 10 or 11.
S + P: Any five points.
S: Any four points.
3P: Any three points.
2P: Any two points.
P: Any one point.

Note: R^2 comparisons okay for justifications in B8 and B9 but NOT for validity in B10.
Any other good comments about deluxe or standard rooms are acceptable as points in section A.
Outlier must be identified in some way.
Must have both highly and positively to score A3.
Excess B points can be converted to A points.

Tasks Q5 (a), (b) and (c)

Evidence (a):

n = 100, p = 0.3 so μ = np = 30 and $\sigma = \sqrt{np(1-p)} = \sqrt{100\times0.3\times0.7} = 4.5826$
So pr(x > 24.5) continuity correction = pr (z > -1.201) = 0.5 + 0.3851 = **0.8851** (accept 0.8849).

Note: Exact answer using binomial distribution directly with graphic calculator is **0.8864**

Evidence (b):

Pr (10^{th} phone call results in 5^{th} booking) = a x pr (4 bookings from 9 phone calls)
= a × ${}^9C_4 a^4(1-a)^5$
= ${}^9C_4 a^5(1-a)^5$ or $126a^5(1-a)^5$ (Either acceptable)

Evidence (c):

Use a tree diagram

Pr (no booking occurs) = 0.6 × 0.3 + 0.3 × 0.51 + 0.1 × 0.657 = **0.3987**
So percentage of enquires that result in no booking is **39.87%.**

Judgement

2S: All three correct in (a), (b) and (c).
S + P: Any two correct out of (a), (b) or (c).
S: Any one correct out of (a), (b) and (c).
2P: Identification of correct method in at least two of (a), (b) or (c).
P: Identification of correct method in one of (a), (b) or (c).

Note: MEI accept 0.3987 in (c)
Must have continuity correction in 5(a)
Accept differences in answers due to rounding.

Tasks Q5 (d)

Evidence:

Pr (phone communication given a booking) = $\dfrac{pr(\text{phone} \cap \text{booking})}{pr(\text{booking})}$

So $0.75 = \dfrac{0.6a}{0.6a+0.3a^2+0.1a^3}$

$\dfrac{3}{4} = \dfrac{6}{6+3a+a^2}$ which gives $a^2 + 3a - 2 = 0$ which gives **a = 0.5616** (accept 0.562 or 0.56).

We reject the negative solution by noting that 0 < a < 1.

Judgement

O: Correct method and answer.
P: Identification of conditional probability with some <u>relevant context</u>.

Note: If give both solutions with no choice score P.

Tasks Q6 (a)

Evidence:

Let x = number of standard rooms
Let y = number of deluxe rooms
The constraints are:

- $x + y \leq 65$
- $45\,000x + 60\,000y \leq 3\,000\,000$ when simplified gives $3x + 4y \leq 200$
- when $0 < x \leq 10$ $y \leq 5$, $10 < x \leq 20$ $y \leq 10$, $20 < x \leq 30$ $y \leq 15$, $30 < x \leq 40$ $y \leq 20$ and $40 < x \leq 50$ $y \leq 25$.

The profit function is $P = 1.0x + 1.4y$
To ensure maximum profit, $x = 40$ and $y = 20$ that is **40 standard rooms and 20 deluxe rooms** should be built.

Judgement

S: Correct optimal point.
P: All three constraints correct OR any two constraints correct plus the correct profit function.
Note: Can use graph as evidence for the constraints.
Profit function can take the form $P = 1.0ax + 1.4ay$ where a is a constant.

Tasks Q6 (b)

Evidence:

Increase x to get the following table noting that $x + y = 65$:

x	*y*	**P**
41	24	74.6
42	23	74.2
43	22	73.8
44	21	73.4
45	20	73

Take (41, 24), as it gives the maximum profit P and it gives a cost ($000) = 45(41) + 60(24) = $3,285,000 so smallest required increase in the budget is **$285,000**.

Judgement

S: Correct method and answer.
P: Identification of correct procedure with all the three points in the note below.

Note:

- Increase x with $40 < x \leq 50$ and $y \leq 25$.
- Must lie on the line $x + y = 65$.
- Find the minimum cost that ensures maximum profit.

Tasks Q6(c)

Evidence:

Let p% be the required occupancy rate for standard rooms so profit function becomes: P = 60% of $1.4y + p\%$ of $1.0x$

So P = $0.84y + 0.01px$ so to give more than one solution this profit line has to be parallel to either

(a) $3x + 4y = 200$ from (40, 20) to (60, 5) or

(b) $x + y = 65$ from (60, 5) to (65, 0).

In (a) $x = \dfrac{200-4y}{3}$ which has to be an integer.
So solution points are **(40, 20), (44, 17), (48, 14), (52, 11), (56, 8) and (60, 5).**

The occupancy rate p% is given by: $\dfrac{0.84}{0.01p} = \dfrac{4}{3}$ which gives p = 63.
So the required occupancy rate for standard rooms = **63%** (accept 0.63).

In (b) $y = 65 - x$ so solution points are **(60, 5), (61, 4), (62, 3), (63, 2), (64, 1) and (65, 0).**

The occupancy rate is given by: $\dfrac{0.84}{0.01p} = \dfrac{1}{1}$ which gives $p = 84$.
So the required occupancy rate for standard rooms = **84%** (accept 0.84).

Judgement

O: Correct method and answer for either correct solution
P: Correct method and answer for either points or occupancy rate.

Note: Must clearly identify all points either as a list or state in a set format.

Specifications: In 2009 we will specify all those points which will NOT be accepted.