# Assessment Schedule – 2013

# Scholarship Statistics (93201)

## Evidence Statement

**General Principles**:

1. Ignore incorrect answers if alongside correct answers. The exception is contradictory statements.
2. Ignore minor copying errors.

**QUESTION ONE**

**Tasks Q1 (a)**

**Evidence:**

**Graph 1:**

- Both breeds milk for fewer days as they age; Jersey's 228 days @ two years to 208 days @ ten years while Ayrshire's 234 days @ two years to 215 days @ ten years.
- The decline is more rapid for cows older than six years.
- Ayrshire cows produce milk for about seven days more each year than Jersey cows.
- For nine year old cows, Ayrshire milk for about ten more days than Jersey cows.

**Graph 2:**

- Both breeds of cow produce a greater volume of milk each year as they age from two to six years. It peaks at an average of 4250 litres per cow for Ayrshire and an average of 3400 litres per cow for jersey. Then annual production declines.
- Ayshire cows produce about 600 L more milk per year than Jersey cows.

**Graph 3:**

- The milk fat percentage is reasonably constant at 5.7% for Jersey and 4.4% for Ayrshire over all ages of cows with a slight decline for ten year olds.
- Jersey cow's milk contains about 1.2% more milk fat than Ayrshire cows.

**Note:**

1. Any other reasonable distinct comparison is acceptable.

**Judgement:**

**S**: THREE distinct comparisons / contrasts of milk production with at least one from each graph.
**P**: ONE or TWO distinct comparisons/contrasts of milk production.

**Tasks Q1 (b) (i)**

**Evidence:**

- Relationship is moderate.
- A greater volume of milk tends to contain a greater amount of protein.
- For cows producing a high volume of milk there appears to be a plateau of approximately 380g protein.
- The volume of milk produced ranges from six to nine litres while the weight of protein in the milk ranges from 260 to 380g.
- As the amount of milk increases, the variability of the weight of protein in the milk decreases.
- As the amount of milk increases by a litre, the amount of protein increases by 34 grams on average.

**Note:**

1. Must have context in the positive relationship statement.

**Judgement**

**S:** Three distinct points about the relationship.
**P**: One or two points about the relationship.

**Task Q1 (b) (ii)**

**Evidence**:
- For 7.5 litres choose $y = -7.2 \times 7.5^2 + 142 \times 7.5 - 320 = 340$g or the line: $y = 34.3 \times 7.5 + 80 = 337$ g.
  For 9.2 litres choose $y = -7.2 \times 9.2^2 + 142 \times 9.2 - 320 = 377$ g.
- For the first prediction there is a lot of scatter in the graph so the prediction is not precise and it would be expected to be in the range 320 to 370 grams. However, in the case of the second prediction it would be expected that the calculated answer to be precise given that there is little scatter in the graph.
- Factors that need to be considered include:
  - Type of food supplied (different food input might be expected to affect the chemical composition of the milk).
  - Health of the cow.
  - Good representative sample of all Jersey cows in NZ (e.g. spread of ages).
  - Time since calved (milk composition changes from newborn to older offspring).
  - Environment. This is influenced by the geographical area of NZ where cows are farmed (differing climates would be expected to affect this rate).
  - Season of the year.
  - Breed of Jersey cow
  - Results based on one particular day.

**Note:**
1. Each dot point has two parts.
2. Good representative sample isn't an acceptable factor as it must have a consequence.
3. Cannot have "age" on its own as a factor.

**Judgement**:
**O**: FIVE or SIX parts correct.
**S**: THREE or FOUR parts correct.
**P**: ONE or TWO parts correct.

**QUESTION TWO**

**Tasks Q2 (a) (i)**

**Evidence:**
**Figure 1:**
- The difference in mean volume of milk produced is 1.69 litres. Group B cows tend to produce more milk than Group A cows after the treatment.
- There is a difference in ranges between Groups A (86 litres to 95 litres approx.) and Group B (89 litres to 98 litres approx.). Group B cows tend to produce more milk overall than Group A cows after the treatment.

**Figure 2:**
- A difference in the mean volume of 1.69 litres or more is unlikely to happen by chance alone since the tail proportion is 0.004.
- This sample provides evidence (tail proportion = 0.004) that the new product results in greater milk production than the current product.

**Note:**
1. One statement with evidence from each figure is required. Any two comments from a figure can be regarded as a single statement, eg a difference of 1.69 litres in the mean leads to a tail proportion of 0.004.
2. From figure one need to say more than just "the mean difference is 1.69 litres".

**Tasks Q2 (a) (ii)**

**Evidence:**
- Cows were randomly assigned to the two groups. This should balance factors like age that might affect milk production.
- Only 27 in each group which is a small sample. Small samples have a greater chance of imbalance between the two groups.
- The method of selection, of 54 cows to start with, needs to be considered before the findings could be generalised to all cows (or all cows of a breed).
- Confirming that cows were keep as one herd with the same grazing conditions.
- Testing whether the effectiveness of treatment varied with time.

**Note:**
1. Blinding with appropriate evidence is acceptable.
2. Method of selection relates to getting a good representative sample.

**Judgement**
**O:** Two points are correct for (i) and two experimental design factors are identified and discussed in (ii).
**S:** Three points. Two points are correct for (i) and one experimental design factor is identified and discussed in (ii) or vice versa.
**P:** One or two points or factors are correct.

**Tasks Q2 (b) (i)**

**Evidence:**
- The mean weight of a Jersey cow is likely to be between 452 and 464 kg whereas the mean weight of an Ayrshire cow is likely to be between 536 and 559 kg.
- The weight of an Ayrshire cow is more varied than Jersey cow weight. This is evidenced by the size of the "box" or by the comparative values of both the interquartile ranges and the standard deviations.
- Both breeds are expected to have weights close to a Normal Distribution. In this sample the Ayrshire cow weights are slightly left skewed (mean < median) whereas the Jersey cow weights display a symmetrical distribution (mean and median are almost equal).

**Note:**
1. The difference in distributions may just be the result of a small sample size.
2. One example of central measures, spread and distribution is required.
3. Allow variation in the interpretation of the confidence interval.
4. Don't allow comparisons of just sample means.
5. Comparisons are not accepted if figures are just quoted i.e. range is 10, range is 15.
6. Allow bootstrap confidence intervals don't overlap so Ayrshire cows weigh more than Jersey cows on average.
7. Distributions: Jersey more "normal like" and Ayrshire more "rectangular".

**Judgement**
**S**: Three comparisons correct.
**P:** One or Two comparisons correct.

**Tasks Q2 (b) (ii)**

**Evidence:**
The mean weight of cows from each breed is estimated to be between 452 and 559 kg from the bootstrap confidence intervals.
The largest difference is 559 – 452 = 107 kg and the smallest difference is 536 – 464 = 72 kg.
The difference in mean weight is between 72 and 107 kg.
Hence the difference in food requirements is between 3% of 72 and 3% of 107 kg so 2.16 kg and 3.21 kg.
So Ayrshire cows need on average between 2.16kg and 3.21 kg of dry matter more per day than Jersey cows.

The validity of using separate confidence intervals is affected by the inflated variability in the weights. A bootstrap confidence interval for the differences should be used. The calculated interval is larger than it should be.

**Note:**
1. Score P if the difference of the mid interval values is multiplied by 3%.
2. Ignore units in answer.
3. Confidence interval for difference in means can be calculated as 95% (2.32 kg to 3.10 kg) or 99% (2.18 kg to 3.22 kg).
4. It's correct to refer to "difference in the means" rather than "mean difference".

**Judgement**
**S**: Interval estimate for the difference is reasonable with an appropriate validity answer OR
**S**: Confidence interval for difference in means is calculated correctly and interpreted in full (clearly articulated) OR
**S**: Confidence interval for difference in means is calculated correctly and clearly linked to an appropriate validity comment.
**P**: Difference estimate is reasonable or a reasonable validity comment or an incorrect confidence interval for the difference in means is interpreted correctly.

**QUESTION THREE**

**Tasks Q3 (a)**

**Evidence:**
Panel size $n$ =15 with number correctly identifying sample of new yoghurt $x$ = 11.
Tail probability pr ($x \geq 11$) = 0.0015 + 0.0003 + 0.000029 + $0^+$ = 0.0018.
As this probability is highly unlikely to be due to chance we conclude that the new yoghurt is distinguishable.

The Binomial model is applicable here as we have a fixed number of trials (15 tasters), two outcomes at each trial (new yoghurt is either distinguishable or not), tasters operate independently and probability of guessing sample of new yoghurt is fixed at 1/3.

**Note:**
1. Can work out pr (x=11) as 0.0015 and make comment about its value being highly unlikely.
2. The confidence interval approach is incorrect as n is small.
3. Binomial can be justified with at least three conditions in context.

**Judgement**
**S:** An appropriate conclusion is provided with correct evidence along with Binomial model being justified.
**P:** An appropriate conclusion is provided with correct evidence OR the Binomial model is partly justified (2 aspects).

**Tasks Q3 (b)**

**Evidence:**
Let $x$ = weight of an ice cream tub.
So pr (500g $\leq x \leq$ 520) = pr (–3.714 < $z$ < 2.0) = 0.4999 + 0.4772 = 0.9771 assuming a normal model.
So 97.71 % of current production is within acceptable limits and thus meeting acceptable requirements.

**Note:**
1. Can calculate 513 +/– 1.96 × 3.5 and get 513 +/– 7, ie 506.1 to 519.9. Compare this to specification limits of 500 g to 520 g.
2. Tail probabilities (23/1000) giving chance of being "out-of-spec" can be calculated instead leading to an unacceptable argument of 2.29% being outside the specification limits.
3. Continuity correction incorrect.
4. Score P is reference is made to a confidence interval.

**Judgement**
**S:** Correct answer with appropriate method.
**P:** Some indication of correct method.

**Tasks Q3(c)**

**Evidence:**
Use Poisson with mean = 6. If $x$ = number of bags that need topping up in an hour we require $x$ so that pr ($\leq x$) $\geq$ 0.9.
We keep adding the Poisson probabilities under mean = 6 until 0.9 is passed (in fact we stop at 0.9161).
So pr ($x \leq 9$) = 0.9161.
Nine bags require topping up at least 90% of the time.
So smallest number of operators required is five.

We use the Poisson here as the number of under-filled bags per hour occurs randomly and independently. Events don't occur simultaneously. Also the number of under-filled bags is proportional to the time interval.

**Note:**
1. Answer can be expressed to three decimal places as 0.103.
2. Can use Normal approximation to Poisson with mean 6 and standard deviation 2.45 to get limit 6 + 1.28 × 2.45 = 9.14, so nine bags require topping up at least 90% of the time. Note 10% tail corresponds to z = 1.28.

**Judgement**

**O:** Correct answer plus justification, of three points (must include the proportional justification), for using the Poisson distribution.

**S:** Correct answer.

**P:** Some evidence of correct calculation, eg choice of Poisson with part calculation or some justification for using Poisson distribution (at least three of the four points).

**QUESTION FOUR**

**Task Q4 (a)**

**Evidence:**

**Trend Graph 5**

- Between winter 2003 and autumn 2007 overall milk production was steady at 14 000 million litres per year.
- Between winter 2008 and autumn 2009, annual production rose from 14 000 million litres to 18 000 million litres.
- Milk production peaks at 18,000 million litres per annum in Year 6.

**Seasonal Graph 6**

- Seasonal milk production peaked in spring 2004 at over 6 000 million litres.
- Seasonal milk production troughed in autumn 2005 at 1 500 million litres.
- Generally milk production peaked in spring at between 5 000 and 6 000 million litres.
- The troughs that occurred annually in autumn were about 2 000 million litres.
- The fluctuations in production had a high variability between winter 2003 and autumn 2006. After that the fluctuations became fairly constant between 3 000 and 5 000 million litres up to autumn 2009.
- The seasonal variations have decreased between winter 2008 and autumn 2011. (2 000 to 5 000) millions of litres in winter 2008 down to (4 000 to 4 500) millions of litres in autumn 2011.

**Note:**

1. Ignore units if implied elsewhere.
2. Can refer to season number.

**Judgement**

**S:** THREE points in total with at least ONE from each of graphs 5 and 6.

**P:** One or two points in total.

**Task Q4 (b)**

**Evidence:**

Extrapolating Graph 5 to milking year 10 using milking years 5,6,7 and 8 we get about 18 000 million litres.
If we take the proportion for autumn from Table 2 we get 18 000 × 4 661 / 17 139 = 4 900 million litres OR using an additive model we get a seasonal for Autumn as +376. Forecast = 4500 + 376 = 4 876 (4 900 million litres).

Validity of prediction is affected by the most recent fluctuating milking year production estimates of seasonal variation figures. Also validity is affected by predicting two years ahead of the given data.

**Note:**

1. Can work from Table 3 to get estimate for milking year 10. Estimate is 16,200 million litres. So answer will be 4,400 million litres.
2. Can use trend line calculated from last four years: $y = 800x + 14\,746$.
3. Units not required in prediction.
4. Can use a variety of methods to get prediction. Some evidence of how forecast was calculated needs to be shown.

**Judgement**

**S:** A reasonable prediction plus a suitable validity comment.

**P**: A reasonable prediction or a suitable validity comment.

**Task Q4(c)**

**Evidence:**
Deflate all prices to 1998 dollars using the CPI (Consumer Price Index) values.

| Year | Milk Fat Price $ per kg | Average Price $000 per hectare |
|---|---|---|
| 1998 | 32.0 | 11.08 |
| 2002 | 39.7 | 12.78 |
| 2006 | 50.2 | 22.49 |
| 2010 | 46.1 | 21.40 |

OR Deflate all prices to base year dollars using the CPI (Consumer Price Index) values.

| Year | Milk Fat Price $ per kg | Average Price $000 per hectare |
|---|---|---|
| 1998 | 38.32 | 13.27 |
| 2002 | 47.60 | 15.30 |
| 2006 | 60.17 | 26.94 |
| 2010 | 55.25 | 25.63 |

- Overall both sets of prices have increased in "real terms".
- The greatest increase in "real terms" happened between 2002 and 2006.
- There was a drop in "real terms" for the prices between 2006 and 2010.
- The overall greatest increase over the 12 years was in the average price per hectare.

**Note:**
1. Must do some correction with the data. Just percentages not acceptable.
2. Can deflate to any years' dollars even to base year $ with CPI = 1000 (see above).
3. Must imply that "real terms" is being discussed.
4. Score N if discussion is based on incorrect calculations.
5. Maximum S if "real terms" is omitted completely in answer.

**Judgement**
**O:** Apply deflation process or perform suitable calculation and make three distinct discussion points.
**S**: Apply deflation process and make one discussion point.
**P**: Some indication of correct method.

**QUESTION FIVE**

**Tasks Q5**

**Evidence: (Distinct feature given in bold).**
**Milk yield is reduced in lame cows.**
- When cows are lame the milk production potential is reduced.
- Mean reduction in milk yield was 360 kg within a range of 160 to 550 kg OR
- Table 5 gives yield losses before and after diagnosis.
- Relates well as impact of lameness is indicated clearly and would infer that the stated range is a confidence interval.

**High yielding lame cows produce more milk than cows that were never lame.**
- Graph 8 supports a higher milk yield of a lame cow only over the early months of lactation since birth.
- A 95% confidence interval is provided with 342 extra kg of milk as the mid-point (95% CI 135 to 549 kg). Milk yield for both a lame and a non-lame cow reduce over time (shown by Graph 8). Peaks after two months from birth.
- The potential of milk production in high yielding cows is lost due to lameness.

**Incidence of Lameness**

- Incidence of lameness is shown by Graph 7.
- Graph 7 shows the number of lame cows peaking at the 3rd month in milk then drops away to the 12th month in milk OR
- Seventy per cent of cows become lame at least once.
- Identifies the month where there is likely to be a significant drop in milk production due to lameness.

**Detecting Lameness**

- The extent of lameness isn't always detected so treatment can be applied.
- In the study of 53 herds, farmers underestimated the prevalence of lameness by concluding 5% lame as opposed to 25% by the researcher OR
- Table 5 gives yield losses before and after diagnosis.
- Treatment is either delayed or not applied when needed thus milk production is affected.

**Note:**

1. No more than three points allowed for credit per feature.
2. Look for feature that is implied and then mark on the basis of the dot points under that particular feature.
3. Any stated feature must have evidence in the report to support it.
4. If more than three features are identified then best three are taken.

**Judgement**

**O +2S:** Nine or eight points over three distinct features.
**O:** Seven points over three distinct features
**2S:** Six points over two or three distinct features
**S+P:** Five points over two or three distinct features
**S:** Four points over two or three distinct features**.**
**3P:** Three points
**2P:** Two points
**P:** One point
**N**: No points.

## Scoring for each Question

Each question part within a question is scored as:

**N** = No meaningful work, insufficient or incorrect answer.
**P** = partially correct to a predetermined level.
**S** = totally correct to a scholarship level.
**O** = totally correct to an outstanding level.

The codes are put together for each question and then converted to a mark out of eight according to the following table:

| Mark | Codes |
|---|---|
| 8 | O + 2S, O + S + P, O + S + 2P, O + 2S + P |
| 7 | O, O + P, O + 2P, O + S, 3S |
| 6 | 2S, 2S + P, 2S + 2P |
| 5 | S + P, S + 2P, S + 3P |
| 4 | S |
| 3 | 3P |
| 2 | 2P |
| 1 | P |
| 0 | N |

The marks for each question are totalled to give an overall mark. Best possible overall mark is 40.