Solutions to the 2005 exam paper

These solutions were devised by two 2005 students. The answers are designed to aid revision, are by no means definitive, may contain errors, and do not contain the level of detail needed for the exam. They are not endorsed or verified by the OU.

Q1a
Finland: population increased but live births decreased.
Northern Ireland: population increased and so did number of live births.

Q1b

UK	1.13
Cyprus	1.11
Finland	1.07
Irish Republic	1.55
Italy	0.93
Luxembourg	1.21
Netherlands	1.26
Total	1.07

Q1c
Rounding errors (more detail needed to gain 2 marks).

Q1d
Need to draw a time series graph with key, axes labelled etc.

Q1e
The overall population amongst the seven countries appears to increase from 2000-2002, whilst the majority of the live birth rates (LBR) in the countries decreases. Little change in LBR is apparent in Italy, UK and Finland, whilst Cyprus shows a dramatic decrease in LBR between 2001-2002. The Irish Republic seems to buck the trend and is the only country to illustrate an increase in LBR over the period.

Q2
Negative relationship.
Roughly linear, with a great deal of variability / scatter.

Q3
Median = 16.5
Q_U = 30
Q_L = 11
IQR = 19
IQR x 1.5 = 28.5, and 28.5 + Q_U = 58.5, so upper adjacent value = 39.

Q4a
Herring

Q4b
Total catch, in order to calculate angles.

Q4c
Bar Chart since data is discrete.

Q5
Exponential distributions have mode at 0.
Data represented in diagram have a higher mode.

Q6a
Left-skew

Q6b

X^2
Transformations with powers above 1 expand high values relative to low values, thus reducing left-skewness.

Q6c
Mean below 7.
Median on boxplot is 7. Left-skew data have mean smaller than median.

Q7
For valid p.m.f.s, sum of probabilities = 1.
Sum of these probabilities = 1.2, hence not a valid p.m.f.

Q8
E(X) = 0.5
V(X) = 0.45

Q9a
$P(X \leq 10) - P(X \leq 5)$
$= (1 - \frac{9}{100}) - (1 - \frac{9}{25})$
= 0.91 - 0.64
= 0.27

Q9b
$0.95 = 1-9/{q^{0.95}}_2$
$-0.05 = -9/{q^{0.95}}_2$
$20 = {q^{0.95}}_2/9$
$180 = {q^{0.95}}_2$
$q^{0.95} = 13.42$
I don't know if this is right, but i put this in the exam!

Q10a
Discrete uniform on integers 1 - 120.
Data are discrete.
No reason to suspect that the fault is more likely to occur at any particular pylon rather than any other pylon.

Q10b
Mean = 60.5
Variance = 1200

Q11a
A: Geometric (0.1)
B: B (600, 0.005)
C: Poisson (75)
D: Exponential (2.5)

Q11b
B: This one caused some debate in the OUSA M248 conference. I thought this would be N (3, 2.985), but it was pointed out that as np is only 3, the Poisson approximation for rare events should be used instead, which is Poisson (3).
C: N (75, 75)

Q11c
Using N (3, 2.985), I calculated
P(X>2)
$= 1 - P(X \leq 2)$

= 1 - P(Y $\le$ 2.5)

= 1 - P $\left(Z \le \frac{2.5-3}{\sqrt{2.985}}\right)$

$\approx$ 1 - P (Z $\le$ - 0.29)

= 1 - (1 - 0.6141)

= 0.6141 $\approx$ 0.614

Using Poisson (3) gives

1 - SIGMA [from i = 0 to 2] P(X = i) i.e. 1 - e^(-3)(1 + 3 + 3²/2!) = 0.577.

<u>Q12a</u>

Mean = standard deviation, so variance = mean squared.

E(U) = E(X) - E(Y) = -1

Variance (U) = $3^2 + 4^2$ = 9 + 16 = 25

<u>Q12b</u>

V(3Y) = 3^2 V(X) = 3^2 x 16 = 144

<u>Q13</u>

$\hat{p}$ = 30 / 80 = 0.375

z = 0.995-quantile = 2.576

(p^-, p^+) = (0.375 $\pm$ 2.576 $\sqrt{0.375x0.625/80}$)

$\approx$ (0.2356, 0.5144)

<u>Q14a</u>

$T_{50} \approx$ N (86 x 50, 11.5^2 x 50) so $T_{50} \approx$ N (4300, 6612.5)

Assume normal distribution.

<u>Q14b</u>

P (T_{50} > 4500)

$\approx$ P (z > $\frac{4500-4300}{6612.5}$)

$\approx$ P (z > 0.03)

= 1 - 0.9931

= 0.0069

$\approx$ 0.007

<u>Q15</u>

Transformation is increasing.

Approximate confidence interval is

(68.7 x 2.54, 69.1 x 2.54)

$\approx$ (174.5, 175.5)

<u>Q16a</u>

1.9^2 / 1.3^2 $\approx$ 2.14, i.e. less than about 3.

S_P^2 = $\frac{16x3.61+17x1.69}{17+18-2}$ (remember to use sd **squared**, not sd)

$\approx$ 2.62

<u>Q16b</u>

(0.086, 2.314)

<u>Q16c</u>

Confidence interval only includes positive values.

Mean weight for patients using diet with exercise is different from that for patients using only the diet.

Appears that patients who use diet and exercise lose more weight.
(More detail on meaning of plausible ranges needed for full marks)

Q17a
H_0: $m = 500$ H_1: $m \neq 500$,

where m is the population mean volume of drink (in ml) put into the cartons by the new filling machine.

The significance level is 5%.

Q17b
Since the confidence interval does not include the value 500, there is moderate evidence against the null hypothesis at the 5% significance level.
There is moderate evidence that the population mean volume of drink put into the cartons by the new filling machine is greater than 500 ml.

Q18a
H_0: m = 0 H_1: m $\neq$ 0,
where m is the population median difference in heart rates before and after the running programme.

Q18b
29.5

Q18c
$E(W_+) = 18$
$V(W_+) = 51$

Q18d
$z = \frac{29.5-18}{\sqrt{51}} \approx 1.61$
$P(|z| \geq 1.61)$
$= 2 \times (1 - 0.9463)$
$= 0.1074$
≈ 0.107
Since p > 0.10, there is little evidence against the null hypothesis.
There is little evidence to suggest that the population median difference in heart rates before and after the running programme is not 0.

Q18e
After the ties are excluded, the sample size is only 8, so the approximation is likely to be poor. Also, there are only 6 distinct differences.

Q19
1: Results
2: Introduction
3: Results
4: Methods
5: Discussion
6: Methods

Q20a
$D = \frac{98}{446} - \frac{77}{481} \approx 0.0596$

$\hat{p} = \frac{98+77}{446+481} \approx 0.1888$

N (0, 0.1888 (1-0.1888) $\left(\frac{1}{446} + \frac{1}{481}\right)$, which is N (0, 0.000662)

P (|Y| $\geq$D) = P (|Z| $\geq$ $\frac{0.0596}{\sqrt{0.000662}}$)
$\approx$ P (|Y| $\geq$2.32)
$\approx$ 0.020

<u>Q20b</u>
Moderate evidence against H_0.
Moderate evidence to suggest that the proportions are not equal.
Data suggest that proportion of male deaths at the hospital that are attributable to smoking is higher than that for females.

<u>Q21a</u>
Power = P (z $\geq$ 1.645 - $\frac{5}{12/\sqrt{15}}$)
$\approx$ P (z $\geq$ 0.03)
$\approx$ 0.488

<u>Q21b</u>
The test is not very strong. The researcher may not reject the null hypothesis, even though it is false. This is a Type II error.

<u>Q22a</u>
Exponential.
Modelling the time between successive events that are random in time is a typical application of an exponential distribution.

<u>Q22b</u>
Mean = 2.2
M.L.E. = $\frac{1}{2.2}$ $\approx$0.4545

<u>Q22c</u>
$\hat{l}$ is biased.
Since there is one value that is much higher than the others, and since the sample size is small, this is likely to be a problem.

<u>Q22d</u>
L(theta) = 1-exp(-6theta x 13.2)

<u>Q23a</u>
$\frac{0.0152}{\sqrt{7.0172/\sqrt{7379.886}}}$ ~ t(70)
So 0.4929 ~ t(70)
0.9-quantile of t(70) = 1.294 so p > 0.10

<u>Q23b</u>
There is little evidence against the null hypothesis. I conclude that the milling temperature (degrees C) and steel elongation (%) are unrelated.

<u>Q24</u>
These questions are often ambiguous.
Either, residuals increase in magnitude as they move across the fitted values, hence variance is not constant, so LRM is not OK.
Or, residuals are scattered about zero in a random, unpatterned fashion, hence LRM is OK.

Q25

Spearman correlation coefficient = -1
If the Pearson correlation coefficient = -1, then the points must lie exactly on a straight line with negative slope, so the Spearman correlation coefficient must also be -1.

Q26a

(i) 132 / 662 $\approx$ 0.1994
(ii) 180 / 196 $\approx$ 0.9184
(iii) 16 / 132 $\approx$ 0.1212

Q26b

(i) Expected value = 132 x 196 / 662 $\approx$ 39.082
Contribution to test statistic is $(16 - 39.082)^2$ / 39.082 $\approx$ 13.63
(ii) Degrees of freedom = 1
0.995-quantile of $\chi^2(1) = 7.88$

p value < 0.005

Q26c

Strong evidence of an association between location and outcome.
Since actual number of fatal accidents at intersections lower than expected number, data suggest fatal accidents are less likely to occur at intersections than at non-intersections.