M248 Exam Solutions 2004

The following solutions have been created by Chris Fredericks and Judith White (M248 students 2005). The answers have been designed to aid revision, are by no means definitive, and may contain errors.

Part I

Question 1.

(a) $\dfrac{10102 \times 1000}{34500} \approx 293$ (to the nearest whole number)

(b) Method 1: “All public schools” – “Nursery and primary”
Method 2: “Comp” + “Grammar” + “Modern” + “Other”

(c) Method 1: 9507 – 5952 = 3555 (in thousands)
Method 2: 1313 + 673 + 1164 + 403 = 3553 (in thousands)

The discrepancy is due to rounding errors. Method 1 is more likely to be accurate as it involves fewer data values and hence less rounding has taken place.

(d)

Year	1970/71	1980/81	1990/91	1994/95	2000/01	2001/02
No. of Pupils (1000s)	3555	4546	3473	3655	3917	3949

(e)

Single diagram of a line plot needs to be produced showing;

- Total number of pupils at public sector secondary schools
- Comp
- Grammar
- Modern
- Other

for the years stated.

(NB, Ensure that the key used is clear and that the axes are labeled. Since the periods are not all the same length, concentrate more on the overall pattern, rather than scale)

(f) Significant increase in the total number of pupils at public sector secondary schools from 1970/01 – 1980/01 (3,555,000 – 4,546,000). In 1990/91, the number of pupils is approximately the same as that in 1970.01 again. From 1990/01 – 2001/02 the total number of pupils increases steadily again.

(g) The number of pupils in comprehensive schools has increased significantly, whilst the number of pupils in the three types of other schools has slowly decreased.

Question 2.

(a) Left-skew

(b) No. For left-skew data, the mean is less than the median, but the question states that the mean (38.6) is greater than the median (approx 36).

(NB, this type of question ALWAYS comes up, so annotate your HB accordingly)

Question 3.

Median = 10.5

Upper Quartile = 21

Question 4.

Illustrates a positive relationship, with longer wave periods associated with higher waves. Appears broadly linear but with much scatter.

Question 5.

(a) 196/639 = 0.3067

(b) 128/232 = 0.5517

(c) 236/443 = 0.5327

Part II

Question 6.

The mean and standard deviation of an exponential distribution are both equal to 1/λ. Since the mean is 7.2 and the standard deviation is 2.9, the model would not be suitable.

(NB, Remember that you were given the sample mean and *variance*, so need to adjust the data accordingly)

Question 7.

$$E(X) = 2\times 0.5 + 3\times 0.4 + 4\times 01 = 2.6$$
$$V(X) = \left(2^2 \times 0.5 + 3^2 \times 0.4 + 4^2 \times 01\right) - 2.6^2 = 0.44$$

Question 8.

B(6,0.4) has mean 6 x 0.4 = 2.4
B(6,0.6) has mean 6 x 0.6 = 3.6

Figure 3(a) is symmetric about x=2.6, so this represents B(6,0.4), and by deduction, figure 3(b) is B(6,0.6).

(NB, a larger *p* gives a smaller probability of small values and a larger probability of large values.)

Question 9.

(a) E(U)=E(2X)=2E(X)=2 x 3 = 6 For a Poisson distribution, $\mu=\sigma^2$

V(U) = V(2X) = 2^2V(X) = 4V(X) = 4 x 3 = 12

(b)
V(W)=V(2X)+V(Y)=V(U)+V(Y)= 12+5=17

Question 10.

(a) $1-e^{-0.05t}$ t ≥0

(b)
$$P(10 < t < 20) = F(20) - F(10)$$
$$= (1-e^{-0.05\times 20}) - (1-e^{-0.05\times 10})$$
$$\approx 0.6321 - 0.3935$$
$$= 0.2386$$
(c)
$$F(q_{0.99}) = 1-e^{-0.05\times q_{0.99}} = 0.99$$
$$so; e^{-0.05\times q_{0.99}} = 0.01$$
$$hence; q_{0.99} = \frac{\ln(0.01)}{-0.05}$$
$$\approx 92.103$$

Hence, 1% of intervals between arrivals are approximately longer than 92 seconds.

Question 11.

$\sum p(x) \neq 1$ and $p(x) < 0$ when $x = 4$

Question 12.

(a)

$$\frac{x-a}{b-a} = \frac{10-8}{20-8} = \frac{1}{6}$$

(b)

$$X \approx B(10, \frac{1}{6})$$
$$P(X = 3) = \binom{10}{3} \frac{1}{6}^{3} \frac{5}{6}^{7}$$
$$P(X = 3) = 0.155$$

(c)

$$\mu = \frac{a = b}{2} = 14$$
$$\sigma^2 = \frac{(b-a)^2}{12} = 12$$

(d)

(i)

$$T_n \approx N(n\mu, n\sigma^2)$$
$$T_n \approx N(560, 480)$$

(ii)

$$P(T_{40} > 600) = P(z > \frac{600-560}{\sqrt{480}})$$
$$\approx P(z > 1.83)$$
$$= 1 - 0.9664$$
$$= 0.0336$$

Part III

Question 13.

(a)

A is $X \approx Poisson(6)$

B is $T \approx M\left(\frac{1}{10}\right)$ (i.e. exponential)

C is $X \approx B\left(20, \frac{1}{10}\right)$

D is $X \approx G\left(\frac{1}{10}\right)$ (i.e. geometric)

(b)

E could be Poisson(μ). Amount of money rounded to discrete values. Range starts at zero and is unbounded.

F could be Poisson(μ). Distribution is discrete and events occur at random. It has an unbounded range {0,1,2,3,…}; one mode, which can take any value within the range (depending on the value of the parameter μ). Could equally be Exponential, i.e. anything right-skew.

Question 14.

$\theta = f(\mu)$, where $f(\mu) = \dfrac{1}{0.00354\mu}$.

Hence, f is a decreasing function.

So,

$$(\theta^{-}, \theta^{+}) = \left(f(\mu^{+}), f(\mu^{-})\right)$$

$$= (34.03, 38.17)$$

Question 15.

$$(\mu^{-}, \mu^{+}) = \left(\overline{x} \pm t\frac{s}{\sqrt{n}}\right)$$

$$= 41.50 \pm 1.960 \times \frac{2.618}{\sqrt{8}}$$

$$= (39.69, 43.31)$$

Question 16.

(a)

$$(d^-, d^+) = \left(\hat{d} \pm z\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}} \right)$$

$\hat{p}_1 = \frac{48}{71}$

$\hat{p}_2 = \frac{22}{52}$

$\hat{d} = \hat{p}_1 - \hat{p}_2 = \frac{467}{1846} = 0.2530$

So,

$$(d^-, d^+) = (0.080, 0.426)$$

(b)

In percentage terms, a plausible range of values (at the 95% confidence level) stretches from 8% to about 43%. The plausible range does not include zero or any negative values: we conclude that the proportion of children with dental caries in the area with fluoridated water is lower.

Part IV

Question 17.

(a)

$$P\left(z \geq 1.645 - \frac{10}{20\big/\sqrt{20}}\right) = 1.645 - 2.236$$
$$P(z \geq -0.591)$$
$$\approx 0.7224$$

(b)

Power is quite high (72%), so the psychologists procedure has a reasonably good chance of finding the true mean.

Question 18.

(a)

Advantage:
Makes some use of the size of the differences. More powerful test.

Disadvantage:
Must assume that differences are modeled by a symmetric distribution. Null distribution of test is found by assuming that absolute differences with a particular rank is just as likely to be associated with a positive difference as a negative difference.

(NB, Perhaps worth making a few notes in the HB re advantages and disadvantages on non-parametric tests)

(b)

(i)

$$E_{W+} = \frac{n(n+1)}{4} = 76.5$$
$$V_{W+} = \frac{n(n+1)(2n+1)}{24} = 446.25$$

(ii)

$$z = \frac{W_+ - E(W_+)}{SD(W_+)}$$
$$z = \frac{119.5 - 76.5}{\sqrt{446.25}} \approx 2.036$$
$$P(Z \leq 2.036) = 1 - \Phi(2.04) = 0.0207$$

Since this is a two-sided test, the required p-value is 0.041. Therefore, there is weak evidence against H_0 (the null hypothesis). We can assume that the population median *m* is equal to zero.

Question 19.

i	**O_i**	**E_i**	**O_i - E_i**	**$(O_i – E_i)^2$/ E_i**
A	40	30	10	3.33
B	6	15	-9	5.40
C	14	15	-1	0.07

$$\sum \frac{(O_i - E_i)}{E_i} = 8.797$$

$$\chi^2 (k - p - 1) = 2$$

Using tables for $\chi^2(2)$, $0.01<p<0.025$

Therefore, there is moderate evidence against the null distribution, which may suggest that the ecological theory is not a great model for the data.

Question 20.

If you look at a particular quantile of the t-distribution, the values decrease as the degrees of freedom increase, but (and this is the crucial bit) they are always larger than the value of the corresponding quantile of the standard normal distribution (the z-quantile). Hence, a z-distribution will always be shorter.

Question 21.

- Difference between proportions (A minus B) 0.083, with 95% CI (-0.078,0.243).
- 31 hospitalized in group A, 24 in group B.
- Calculated approximate p value, p=0.31.

Part V

Question 22.

(a)

$$L(p)=\left(e^{-p}\right)^{3}\times\left(e^{-2p}\right)\times\left(\frac{1296e^{-6p}p^{5}}{120}\right)$$

$$L(p)=\frac{54}{5}p^{5}e^{-11p}$$

(b)

Approximate position of $\hat{p}$ is at the peak of the curve. Hence the approximate value of $\hat{p}$ is 0.55.

(NB, The precise value can be found using calculus, i.e. by differentiating L(p)and solving for p to get 6/11)

Question 23.

(a)

Exponential, M(λ) is $\hat{\lambda}=\frac{1}{\overline{X}}$

so MLE of λ is 30/77≈0.3896

(b)

The estimator is biased for λ.

Question 24.

For a general random sample of size n, from a population with unknown mean θ, the least squares estimate of θ is the value $\hat{\theta}$ that minimises,

$$\sum_{i=1}^{n}(x_i-\hat{\theta})^2$$

so

$$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$$

Hence,

$\hat{\theta}$=0.1325

Question 25.

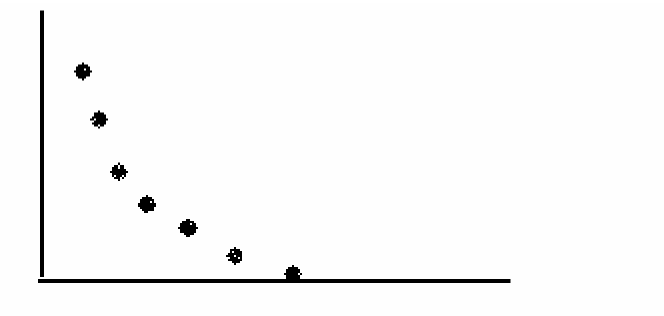
Plot shows no particular pattern, and points seem to be randomly scattered around zero. Therefore, the assumption of zero mean and constant variance holds.

Alternatively;

There is a bit of a pattern if you consider the point at (48.85, -7ish) as an outlier: small values have negative residuals, medium values have positive residuals and large values have negative residuals, so it could be argued that there is a pattern and so a curve could provide a better fit to the data than a straight line.

(NB, Making a case for either argument with good justification would probably get you the mark)

Question 26.



(NB, Any graph showing points monotonically decreasing in a curved fashion so that R_S= -1 and R> -1 hold)

Question 27.

(a)

Scatterplot displays a clear upward trend (positive correlation). A straight line could be fitted against the scatterplot quite well (i.e. it is linear) and hence fitting a linear regression model would be adequate for the data.

(b)

The slope parameter β of the linear regression model represents the average number of tuna caught (in thousands of tonnes) per hooks used (in 100 millions).

(c)

A 100 million increase in hooks used corresponds to an increase in the catch of 14.27 thousand tonnes, that is β.

(d)

(Use HB Unit D2:10)

$$= 28.51 \pm 2.831\sqrt{17.286}\sqrt{\frac{(1-0.7037)^2}{2.343}+\frac{1}{23}}$$

t (21) for 0.995 is 2.831

$$\approx 28.51 \pm 3.3488$$

$$= (25.16, 31.86)$$

(e)

If a large number of samples of size n, were drawn independently from the population, and a 99% CI calculated on each occasion, then approximately 99% of these intervals would contain the true mean. That is for every 100 million hooks used the mean number of tuna caught would 99% of the time lie in the interval (25.16, 31.86)