MAS1403/ACE2013 2007/08 exam – some brief solutions

Question A1 – data summaries

(a)

$$\begin{aligned} \bar{x} &= £37,000 \\ \text{median} &= £40,000 \\ s &= £12,165 \end{aligned}$$

(b) Box–and–whisker plot: vertical lines at 26.5, 40 and 45.25 to form the "box" – horizontal line cutting through this box extending from 15 to 53 to for the "whisker".

Question A2 – Binomial distribution

(a)

$$\begin{aligned} X &: \text{No. put on hold} \\ X &\sim \text{Bin}(15, 0.3) \end{aligned}$$

(b)

$$\begin{aligned} E(X) &= 4.5\text{customers} \\ Var(X) &= 3.15 \qquad \text{and so} \\ \text{s.d.} &= \sqrt{3.15} = 1.77\text{customers} \end{aligned}$$

(c)

$$\begin{aligned} \Pr(X = 5) &= {}^{15}\text{C}_5 0.3^5 0.7^{10} \\ &= 0.206 \end{aligned}$$

(d)

$$\begin{aligned} T &: \text{Time between sales} \\ T &\sim \exp(2) \end{aligned}$$

Thus,

$$\begin{aligned} \Pr(T < \frac{1}{4}) &= 1 - e^{-2\times\frac{1}{4}} \\ &= 0.393 \end{aligned}$$

Question A3 – Normal distribution

$$\begin{aligned} \Pr(X > 1060) &= \Pr(Z > \frac{1060 - 1000}{25}) \\ &= 1 - \Pr(Z \leq 2.4) \\ &= 1 - 0.9918 \\ &= 0.0082 = 0.82\% \end{aligned}$$

Question A4 – Correlation

$$\begin{aligned} S_{XY} &= 1338 \\ S_{XX} &= 29.26 \\ S_{YY} &= 75950 \\ r &= \frac{1338}{\sqrt{29.26 \times 75950}} \\ &= 0.898 \end{aligned}$$

Question A5 – Confidence interval/test for two means

(a) 95% CI given by:

$$\begin{aligned} \bar{x} &\pm t \times \sqrt{s^2/n} \\ 385 &\pm 2.064 \times \sqrt{98.5^2/25} \\ 385 &\pm 40.6608 \end{aligned}$$

giving (£344.34, £425.66). The interval includes zero – casts doubt on the validity of the headline.

(b)

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2 \end{aligned}$$

$$\begin{aligned} t &= \frac{|385 - 435|}{87.96 \times \sqrt{\frac{1}{25} + \frac{1}{20}}} \\ &= 1.895 \end{aligned}$$

Comparing to t tables on row ∞ ($25 + 20 - 2 = 43$, so we use the ∞ row), we get a p–value **between 5% and 10%**.

- slight evidence against H_0
- Retain H_0
- Insufficient evidence to suggest a real difference in the average guest bills of the two hotels

Question B6 – Goodness–of–fit and EMV

(a)

$$\bar{x} \;=\; 1.22 \text{ complaints}$$

(b) Appropriate probability model – Poisson! (X is unbounded and positively skewed)

(c) Using

$$\Pr(X = r) \;=\; \frac{e^{-\lambda} \times \lambda^r}{r!},$$

where $\lambda = 1.22$, we get:

No. of complaints	Bonus	$\Pr(X = r)$
0	1000	0.295
1	750	0.360
2	200	0.220
3	–500	0.089
4+	–1500	0.036

(d)

$$\begin{aligned} EMV(\text{Bonus scheme}) &= (0.295 \times 1000) + (0.360 \times 750) + \ldots \\ &= \pounds 510.50 \end{aligned}$$

(e)

$$\begin{aligned} H_0 &: \text{Complaints follow a Poisson distribution} \\ H_1 &: \text{Complaints do not follow a Poisson distribution} \end{aligned}$$

After pooling categories 3 & 4, we get:

No. of complaints	O	E	$\frac{(O-E)^2}{E}$
0	12	29.5	10.381
1	61	36.0	17.36
2	20	22.0	0.182
3+	7	12.5	2.42
			30.343

Thus, $X^2 = 30.343$. From tables, on $\nu = 4 - 1 - 1 = 2$ we see that the p–value is **smaller than 1%**.

- Strong evidence to reject H_0
- Reject H_0 in favour of H_1

– The no. of complaints does not conform to a Poisson distribution!

Since we have rejected the Poisson distribution, this would cast doubt on our EMV calculation in part (d).

Question B7 – Linear programming

(a)

$$\begin{aligned} x &: \text{No. of pizzas to make} \\ y &: \text{No. of pasta dishes to make} \end{aligned}$$

$$\begin{aligned} 3x + 9y &\leq 4500 \\ 8x + 4y &\leq 4000 \\ x &\geq 200 \\ y &\geq 100 \end{aligned}$$

$$P = 4x + 5y$$

(b) Graph – went through this in detail in tutorials

(c) Max. profit is obtained when $x = 300$ and $y = 400$ giving £3200
Min. profit is obtained when $x = 200$ and $y = 100$ giving £1300

(d) Need to solve the simultaneous equations:

$$\begin{aligned} 3x + 9y &= 4500 \\ 8x + 4y &= 4000 \end{aligned}$$

Question B8 – Time series/forecasting

(a) Time series plot. Comments: general decreasing linear trend, with seasonality.

(b) Moving averages:

	Morning	Lunchtime	Afternoon
2005	*	2690	2473.33
2006	2246.67	2053.33	1826.67
2007	1630	1390	*

(c)

$$\begin{aligned} S_{TY} &= -6006.67 \\ S_{TT} &= 28 \end{aligned}$$

$$\begin{aligned} \hat{\beta} &= -214.52 \\ \hat{\alpha} &= 3116.89 \end{aligned}$$

giving $Y = 3116.89 - 214.52T + \epsilon$.

(d) Mean of seasonal effects = 11, giving *adjusted* seasonal effects:

$$\begin{aligned} \text{Morning} &: -580.8 \\ \text{Lunchtime} &: -76.9 \\ \text{Afternoon} &: +657.6 \end{aligned}$$

(e) 10am–12 noon in 2008 represents time point $T = 10$, giving:

$$\begin{aligned} Y &= 3116.89 - 214.52 \times 10 \\ &= 971.69 \end{aligned}$$

Adjusting for seasonality gives 971.69 – 580.8 = 390.89, or about 391 visitors. Thus, the family *should* close Borwood House.

Question B9 – Graphical/numerical summaries and Normal distribution

(a) Simple percentages: 3.3, 4.9, 5.7, 9.8, 14.6, 34.1, 16.3, 11.4.

(b) Relative percentage frequency polygons – sample sizes very different! We want to make relative comparisons.

(c) Graphs – plot midpoints versus percentages. Comments: Students: symmetric (Normal distribution?), career–breakers: positively skewed. On average career–breakers spend more, but have a greater spread of values.

(d)

$$\bar{x}_{\text{students}} \approx £2340.61$$

Approximation since we assume each student takes the midpoint of their class interval.

(e)

$$\begin{aligned} \Pr(X \geq 3500) &= \Pr\left(Z \geq \frac{3500 - 2340.61}{588.37}\right) \\ &= 1 - \Pr(Z < 1.97) \\ &= 1 - 0.9756 \\ &= 0.0244 \end{aligned}$$

This calculation seems valid since student expenditure seems Normally distributed

(f) No – career–breakers group is almost certainly not Normal