Probability and Statistics 1 2006–07

Progress Test

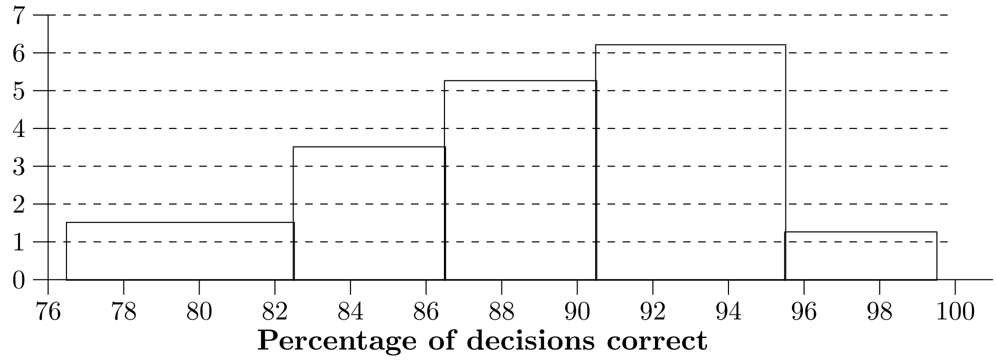
16 January 2007, 2.30 – 4.00 pm

Answer all 6 questions.
The total number of marks available is 50.

1. A panel reviews videos of 80 football matches and assigns to each match a percentage value indicating the proportion of decisions which the referee judged correctly. Figure 1 (below) is a histogram illustrating the data set. Note that the interval widths are 6, 4, 4, 5 and 4.

(i) Calculate the number of matches in each of the five categories. **[2 marks]**

(ii) a) On graph paper draw a cumulative frequency diagram.
b) Read off the median and quartiles of the data set.
c) Comment on the symmetry or skewness of the data set. **[5 marks]**



2. 18 office workers, 8 male and 10 female, are given training to improve their times for running the 400 metres. The percentage improvements in their times are as follows:

Male	11	7	13	8	32	13	22	6		
Female	15	11	19	21	16	9	14	13	17	20

(i) Display the data on a back-to-back stem-and-leaf diagram. **[2 marks]**

(ii) Calculate the median and quartiles for each data set and determine whether there are any outliers present. **[5 marks]**

(iii) Draw a box-and-whisker plot and compare the two samples in terms of location and spread. **[4 marks]**

3. A group of scientists is asked to comment on the threat to global civilisation in the next 1000 years represented by three possible catastrophes:

W: global warming by at least 7°C
S: eruption of a supervolcano
D: an incurable disease

The scientists agree that, taking all other circumstances into account, the probabilities of the catastrophes are $\mathbb{P}(W) = \frac{3}{4}$, $\mathbb{P}(S) = 0.2$ and $\mathbb{P}(D) = 0.4$. However, they are not independent: if S occurs, the chance of an incurable disease will be 50%, and if W occurs the chance of a supervolcano is $\frac{2}{9}$.

(i) Assuming the scientists' predictions are correct, calculate $\mathbb{P}(S \cap D)$ and $\mathbb{P}(S \cap W)$. **[2 marks]**

(ii) Express in set-theoretic notation the event that the planet will be affected by global warming and an incurable disease but that there will be no supervolcano eruption. **[1 mark]**

(iii) Express in words the event $W \cap (S \cup D)^c$. **[1 mark]**

(iv) The scientists claim that the chance that none of the three catastrophes takes place is just one in ten. Prove that the event in (ii) has probability $\frac{11}{60}$. **[3 marks]**

4. A family always takes a summer holiday in Spain, Italy or Greece. The probability that they enjoy their holiday is 60% if the holiday is in Spain, 70% if in Greece, 40% if in Italy. When the family enjoys a holiday, they go back to the same country in the following year, except that they never visit one country for more than two years in a row. After spending two years in a country, or after a holiday they didn't enjoy, they choose one of the other two countries (probability $\frac{1}{2}$ each). In 2003 and 2004 the family went on holiday to Spain.

(i) Draw a tree diagram illustrating the family's possible holiday destinations in 2005 and 2006. **[2 marks]**

(ii) What is the probability that the family takes a holiday in Greece in 2006? **[1 mark]**

(iii) Given that the family visits Italy in 2006, calculate
a) the probability that the family went to Italy in 2005;
b) the probability that the family visits Spain in 2007. **[5 marks]**

5. A stallholder in a market has a box of T-shirts to sell. There are at least 20 in each size (small, medium, large and extra large). From experience the stallholder knows that 15% of customers buy small T-shirts, 20% medium, 35% large and 30% extra large.

(i) Find the probability that the sizes of the T-shirts bought by the first two customers are different. **[3 marks]**

(ii) Out of the first 20 T-shirts sold, what is the probability that at least 5 are large? **[3 marks]**

(iii) Write down the distribution of the number of T-shirts which are sold until (and including) the first medium-sized T-shirt. (Give the name of the distribution and its expectation.) **[2 marks]**

6. A continuous random variable, X, has cumulative distribution function, F_X, given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.02x^2 & \text{if } 0 \le x < 5 \\ 1 - 0.02(10-x)^2 & \text{if } 5 \le x < 10 \\ 1 & \text{if } x \ge 10 \end{cases}$$

(i) Calculate and sketch the density function, f_X, of X. **[3 marks]**

(ii) Find the expectation and variance of X. **[4 marks]**

(iii) Calculate $\mathbb{P}(4 < X < 8)$. **[2 marks]**

Formulae

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Distribution	Notation	Mean	Variance	$p(x)$	Domain
Binomial	$\mathrm{Bin}(k, \theta)$	$k\theta$	$k\theta(1-\theta)$	$\binom{k}{x}\theta^x (1-\theta)^{k-x}$	$x = 0, 1, \ldots, k$
Poisson	$\mathrm{Pois}(\lambda)$	λ	λ	$e^{-\lambda} \frac{\lambda^x}{x!}$	$x = 0, 1, \ldots$
Geometric	$\mathrm{Geom}(\theta)$	θ^{-1}	$\theta^{-2}(1-\theta)$	$\theta (1-\theta)^{x-1}$	$x = 1, 2, \ldots$

Probability and Statistics 1 2006–07

Solutions to Progress Test

1. (i) The table below shows the calculations. $\boxed{2}$

Interval	Width	Frequency density	Frequency
76.5–82.5	6	1.5	9
82.5–86.5	4	3.5	14
86.5–90.5	4	5.25	21
90.5–95.5	5	6.2	31
95.5–99.5	4	1.25	5

(ii) a) Cumulative frequency diagram. $\boxed{3}$
b) Median and quartiles are obtained by reading off the values on the x-axis corresponding to values 0.25, 0.5 and 0.75 (or 20, 40 and 60) on the y-axis. The values which should be obtained are LQ 85.6, Median 89.7, UQ 93.1 (or not far off). $\boxed{1}$
c) The histogram suggests a left skew, as does the comparison between mean and median. $\boxed{1}$

2. (i) Interval widths of 5 seem suitable

```
Male        Female
 876 | 00 | 9
 331 | 10 | 134
     | 10 | 5679
   2 | 20 | 01
     | 20 |
   2 | 30 |
```

$\boxed{2}$

(ii) Males: 8 observations; LQ is 7.5, median 12, UQ 17.5. Outliers: the highest value is not quite 1.5 times the IQR above the UQ, so it can't be considered an outlier. $\boxed{3}$
Females: quartiles are 13 and 19, median 15.5. No outliers. $\boxed{2}$

(iii) Box plot. $\boxed{2}$
Location: on the whole the males show less improvement than the females, but there are exceptions. $\boxed{1}$
Spread: the male data are more spread out than the female data. $\boxed{1}$

3. (i) $\mathbb{P}(S \cap D) = \mathbb{P}(S)\,\mathbb{P}(D \mid S) = 0.2 \times \frac{1}{2} = 0.1.$ $\boxed{1}$
$\mathbb{P}(S \cap W) = \mathbb{P}(W)\,\mathbb{P}(S \mid W) = \frac{3}{4} \times \frac{2}{9} = \frac{1}{6}.$ $\boxed{1}$

(ii) $W \cap D \cap S^c.$ $\boxed{1}$

(iii) The world will be affected by global warming but there will be neither a killer disease nor a supervolcano. $\boxed{1}$

(iv) Let q represent the probability of the event in (ii). Then $0.9 = \mathbb{P}(S \cup W \cup D) = 0.2 + q + (\frac{7}{12} - q) + (\frac{3}{10} - q) = \frac{13}{12} - q$. Hence $q = \frac{13}{12} - 0.9 = \frac{11}{60}$. $\boxed{3}$

4. (i) Tree diagram. $\boxed{2}$

(ii) $\mathbb{P}(G_6) = \mathbb{P}(I_5 \cap G_6) + \mathbb{P}(G_5 \cap G_6) = 0.5 \times 0.3 + 0.5 \times 0.7 = 0.5.$ $\boxed{1}$

(iii) a) $\mathbb{P}(I_5 \mid I_6) = \mathbb{P}(I_5 \cap I_6)/\mathbb{P}(I_6) = \frac{0.5 \times 0.4}{0.5 \times 0.4 + 0.5 \times 0.15} = \frac{8}{11}.$ $\boxed{2}$
b) We have

$$\begin{aligned}
\mathbb{P}(S_7 \mid I_6) &= \frac{\mathbb{P}(I_5 \cap I_6 \cap S_7) + \mathbb{P}(G_5 \cap I_6 \cap S_7)}{\mathbb{P}(I_6)} \\
&= \frac{0.5 \times 0.4 \times 0.5 + 0.5 \times 0.15 \times 0.3}{0.275} = \frac{49}{110} = 0.4454. \quad \boxed{3}
\end{aligned}$$

5. (i) The probability that they are the same is $0.15^2 + 0.2^2 + 0.35^2 + 0.3^2 = 0.0225 + 0.04 + 0.1225 + 0.09 = 0.275$. $\boxed{2}$

Hence the probability that they are different is 0.725. $\boxed{1}$

(ii) Let X be the number of L T-shirts sold, so that $X \sim \text{Bin}(20, 0.35)$. We require $\mathbb{P}(X \geq 5) = 1 - [p(0) + p(1) + p(2) + p(3) + p(4)] = 1 - [0.0002 + 0.0020 + 0.0100 + 0.0323 + 0.0738] = 0.8818$. $\boxed{3}$

(iii) Geometric distribution with mean 5. $\boxed{2}$

6. (i) $f(x) = 0.04x$ for $0 < x < 5$ or $0.04(10 - x)$ for $5 < x < 10$. $\boxed{2}$

Sketch (triangular). $\boxed{1}$

(ii) By symmetry, $\mathbb{E}(X) = 5$. $\boxed{1}$ $\mathbb{E}(X^2) = \int_0^5 0.04x^3\,dx + \int_5^{10} 0.04x^2(10 - x)\,dx = 6.25 + \frac{4}{30}(1000 - 125) - (100 - 6.25) = 29.167$. $\boxed{2}$

To obtain the variance, subtract 25: $\text{Var}\,(X) = 4.167$. $\boxed{1}$

(iii) $\mathbb{P}(4 < X < 8) = F(8) - F(4) = 0.92 - 0.32 = 0.6$. $\boxed{2}$