# Year 2 Double Maths (Statistics) Holiday Assignment

## Due in first lesson September 2017

As part of this assignment you are required to use your understanding of statistics to investigate the statement.
Do you agree with the statement "Young people made a significant contribution to the result of the Election in June?" Give evidence.

**Now do the following past paper questions.**
**Full exam level presentation showing all working and explanations.**

### A. Past paper work

1. Each of the 25 students on a computer course recorded the number of minutes $x$, to the nearest minute, spent surfing the internet during a given day. The results are summarised below.

$$\Sigma x = 1075, \ \Sigma x^2 = 46\,625.$$

(*a*) Find $\mu$ and $\sigma$ for these data.

**(3)**

Two other students surfed the internet on the same day for 35 and 51 minutes respectively.

(*b*) Without further calculation, explain the effect on the mean of including these two students.

**(2)**

4. The discrete random variable $X$ has the probability function shown in the table below.

| $x$ | $-2$ | $-1$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|
| P($X = x$) | 0.1 | $\alpha$ | 0.3 | 0.2 | 0.1 | 0.1 |

Find
(*a*) $\alpha$

**(1)**

(*b*) P($-1 < X \leq 2$),

**(2)**

(*c*) F($-0.4$),

**(1)**

(*d*) E($3X + 4$),

**(4)**

(*e*) Var($2X + 3$).

**(4)**

**5.** A market researcher asked 100 adults which of the three newspapers *A*, *B*, *C* they read. The results showed that 30 read *A*, 26 read *B*, 21 read *C*, 5 read both *A* and *B*, 7 read both *B* and *C*, 6 read both *C* and *A* and 2 read all three.

(*a*) Draw a Venn diagram to represent these data.

**(6)**

One of the adults is then selected at random.

Find the probability that she reads

(*b*) at least one of the newspapers,

**(2)**

(*c*) only *A*,

**(1)**

(*d*) only one of the newspapers,

**(2)**

(*e*) *A* given that she reads only one newspaper.

**(2)**

**6.** Three swimmers Alan, Diane and Gopal record the number of lengths of the swimming pool they swim during each practice session over several weeks. The stem and leaf diagram below shows the results for Alan.

| Lengths | | 2 $\vert$ 0 means 20 |
|---|---|---|
| 2 | 0 1 2 2 | (4) |
| 2 | 5 5 6 7 7 8 9 | (7) |
| 3 | 0 1 2 2 4 | (5) |
| 3 | 5 6 6 7 9 | (5) |
| 4 | 0 1 3 3 3 3 3 4 4 4 | (10) |
| 4 | 5 5 6 6 6 7 7 8 8 9 9 9 | (12) |
| 5 | 0 0 0 | (3) |

(*a*) Find the three quartiles for Alan's results.

**(4)**

The table below summarises the results for Diane and Gopal.

| | Diane | Gopal |
|---|---|---|
| Smallest value | 35 | 25 |
| Lower quartile | 37 | 34 |
| Median | 42 | 42 |
| Upper quartile | 53 | 50 |
| Largest value | 65 | 57 |

(*b*) Using the same scale and on the same sheet of graph paper draw box plots to represent the data for Alan, Diane and Gopal.

**(8)**

(*c*) Compare and contrast the three box plots.

**(4)**

**A Answers**
1. (a) 43, 4  (b) one value is 8 below the mean, the other is 8 above, therefore the mean is unchanged
4. (a) 0.2 (b) 0.6 (c) 0.3 (d) 4.9 (e) 8.04
5. (a) correct diagram (b) 0.61 (c) 0.21 (d) 0.47 (e) awrt 0.45
6. (a) 30, 42, 46  (b) correct box plots (c) Alan negative skew, Diane positive, Gopal symmetric. All have same median. All have same IQR. Diane tends to swim more lengths than the other two.

## B.  Past paper work

**2.**   A meteorologist measured the number of hours of sunshine, to the nearest hour, each day for 100 days. The results are summarised in the table below.

| Hours of sunshine | Days |
| --- | --- |
| 1 | 16 |
| 2–4 | 32 |
| 5–6 | 28 |
| 7 | 12 |
| 8 | 9 |
| 9–11 | 2 |
| 12 | 1 |

(*a*) On graph paper, draw a histogram to represent these data.

**(5)**

(*b*) Calculate an estimate of the number of days that had between 6 and 9 hours of sunshine.

**(2)**

**3.** A discrete random variable $X$ has the probability function shown in the table below.

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X = x)$ | $\frac{1}{3}$ | $a$ | $\frac{2}{3} - a$ |

(*a*) Given that $E(X) = \frac{5}{6}$, find $a$.

**(3)**

(*b*) Find the exact value of Var $(X)$.

**(3)**

(*c*) Find the exact value of $P(X \le 15)$.

**(1)**

---

**4.**     A contractor bids for two building projects. He estimates that the probability of winning the first project is 0.5, the probability of winning the second is 0.3 and the probability of winning both projects is 0.2.

(*a*) Find the probability that he does not win either project.

**(3)**

(*b*) Find the probability that he wins exactly one project.

**(2)**

(*c*) Given that he does not win the first project, find the probability that he wins the second.

**(2)**

(*d*) By calculation, determine whether or not winning the first contract and winning the second contract are independent events.

**(3)**

**6.**     Hospital records show the number of babies born in a year. The number of babies delivered by 15 male doctors is summarised by the stem and leaf diagram below.

|         Babies | (4 | 5 means 45) | Totals |
|---|---|---|---|

| Babies |   |  | Totals |
|---|---|---|---|
| 0 |  |  | (0) |
| 1 | 9 |  | (1) |
| 2 | 1 6 7 7 |  | (4) |
| 3 | 2 2 3 4 8 |  | (5) |
| 4 | 5 |  | (1) |

| 5 | 1 | (1) |
|---|---|-----|
| 6 | 0 | (1) |
| 7 |   | (0) |
| 8 | 6 7 | (2) |

(*a*) Find the median and inter-quartile range of these data.

**(3)**

(*b*) Given that there are no outliers, draw a box plot on graph paper to represent these data. Start your scale at the origin.

**(4)**

(*c*) Calculate the mean and standard deviation of these data.

**(5)**

The records also contain the number of babies delivered by 10 female doctors.

| 34 | 30 | 20 | 15 | 6 |
|----|----|----|----|---|
| 32 | 26 | 19 | 11 | 4 |

The quartiles are 11, 19.5 and 30.

(*d*) Using the same scale as in part (*b*) and on the same graph paper draw a box plot for the data for the 10 female doctors.

**(3)**

(*e*) Compare and contrast the box plots for the data for male and female doctors.

**(2)**

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+

**B Answers**

**2. 28$^1/_3$ days**
**3. a) $^1/_2$, b) $^{17}/_{36}$  c)$^5/_6$**
**4. a) 0.4, b) 0.4 c) 0.2 d) events are independent**
**6. a) 33, 24 c) 41.2, 20.7 e) male median > female median, IQR male > IQR female,**
**range male > range female, male positive skew, female almost symmetrical**
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+

**C. Past paper work**

**1.** An unbiased die has faces numbered 1 to 6 inclusive. The die is rolled and the number that appears on the uppermost face is recorded.

(*a*) State the probability of not recording a 6 in one roll of the die.

**(1)**

The die is thrown until a 6 is recorded.

(*b*) Find the probability that a 6 occurs for the first time on the third roll of the die.

**(3)**

---

**2.** Statistical models can be used to describe real world problems.  Explain the process involved in the formulation of a statistical model.

**(4)**

---

**3.** For the events *A* and *B*,

(*a*) explain in words the meaning of the term $P(B \mid A)$,

**(2)**

(*b*) sketch a Venn diagram to illustrate the relationship $P(B \mid A) = 0$.

**(2)**

Three companies operate a bus service along a busy main road. Amber buses run 50% of the service and 2% of their buses are more than 5 minutes late. Blunder buses run 30% of the service and 10% of their buses are more than 5 minutes late. Clipper buses run the remainder of the service and only 1% of their buses run more than 5 minutes late.

Jean is waiting for a bus on the main road.

(*c*) Find the probability that the first bus to arrive is an Amber bus that is more than 5 minutes late. **(2)**

Let *A, B* and *C* denote the events that Jean catches an Amber bus, a Blunder bus and a Clipper bus respectively. Let *L* denote the event that Jean catches a bus that is more than 5 minutes late.

(*d*) Draw a Venn diagram to represent the events *A, B*, C and *L*. Calculate the probabilities associated with each region and write them in the appropriate places on the Venn diagram.

**(4)**

(*e*) Find the probability that Jean catches a bus that is more than 5 minutes late.

**(2)**

**4.** A discrete random variable $X$ takes only positive integer values. It has a cumulative distribution function $F(x) = P(X \le x)$ defined in the table below.

| $X$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $F(x)$ | 0.1 | 0.2 | 0.25 | 0.4 | 0.5 | 0.6 | 0.75 | 1 |

(*a*) Determine the probability function, $P(X = x)$, of $X$.

**(3)**

(*b*) Calculate $E(X)$ and show that $Var(X) = 5.76$.

**(6)**

(*c*) Given that $Y = 2X + 3$, find the mean and variance of $Y$.

**(3)**


**6.** The labelling on bags of garden compost indicates that the bags weigh 20 kg. The weights of a random sample of 50 bags are summarised in the table below.

| Weight in kg | Frequency |
|---|---|
| 14.6 – 14.8 | 1 |
| 14.8 – 18.0 | 0 |
| 18.0 – 18.5 | 5 |
| 18.5 – 20.0 | 6 |
| 20.0 – 20.2 | 22 |
| 20.2 – 20.4 | 15 |
| 20.4 – 21.0 | 1 |

(*a*) On graph paper, draw a histogram of these data.

**(4)**

(*b*) Using the coding $y = 10$(weight in kg – 14), find an estimate for the mean and     standard deviation of the weight of a bag of compost.

**(6)**

$$[\text{Use } \Sigma fy^2 = 171\,503.75]$$

(*c*) Using linear interpolation, estimate the median.

**(2)**

The company that produces the bags of compost wants to improve the accuracy of the labelling. The company decides to put the average weight in kg on each bag.

(*d*) Write down which of these averages you would recommend the company to     use. Give a reason for your answer.

**(2)**

**+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++**

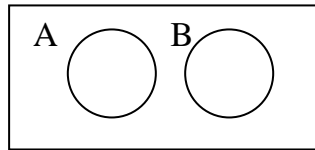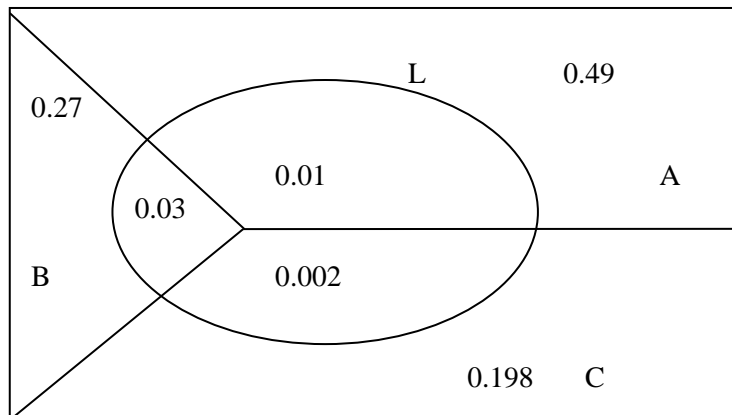## C Answers

1.  a) $\dfrac{5}{6}$  b) $\dfrac{25}{216}$

2.  Observe real world problem
    Collect data
    Compare observed against expected outcomes
    Refine model if necessary

3.  a) Probability of B given A has occurred
    b)



    c) 0.01
    d)



    e) 0.042

4.  a)

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| P($X=x$) | 0.1 | 0.1 | 0.05 | 0.15 | 0.1 | 0.1 | 0.15 | 0.25 |

    b) 5.2  c) 13.4  23.04

6.    a) Frequency densities – 5, 0, 10, 4, 110, 75, 1.7    graph
      b) 19.8    0.963    c) 20.1
      d) Median – data skewed;   or mean – fewer complaints

+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
+

## D. Past paper work

**1.** The total amount of time a secretary spent on the telephone in a working day was recorded to the nearest minute. The data collected over 40 days are summarised in the table below.

| Time (mins) | 90–139 | 140–149 | 150–159 | 160–169 | 170–179 | 180–229 |
|---|---|---|---|---|---|---|
| No. of days | 8 | 10 | 10 | 4 | 4 | 4 |

Draw a histogram to illustrate these data

**(4)**

**2.** A car dealer offers purchasers a three year warranty on a new car. He sells two models, the Zippy and the Nifty. For the first 50 cars sold of each model the number of claims under the warranty is shown in the table below.

|  | Claim | No claim |
|---|---|---|
| Zippy | 35 | 15 |
| Nifty | 40 | 10 |

One of the purchasers is chosen at random. Let $A$ be the event that no claim is made by the purchaser under the warranty and $B$ the event that the car purchased is a Nifty.

(*a*)  Find P($A \cap B$).

**(2)**

(*b*)  Find P($A'$).

**(2)**

Given that the purchaser chosen does not make a claim under the warranty,

(*c*)  find the probability that the car purchased is a Zippy.

**(2)**

(*d*) Show that making a claim is not independent of the make of the car purchased.

Comment on this result.

**(3)**

**4.** A restaurant owner is concerned about the amount of time customers have to wait before being served. He collects data on the waiting times, to the nearest minute, of 20 customers. These data are listed below.

(*a*) Find the median and inter-quartile range of the waiting times.

**(5)**

An outlier is an observation that falls either 1.5 × (inter-quartile range) above the upper quartile or 1.5 × (inter-quartile range) below the lower quartile.

(*b*) Draw a boxplot to represent these data, clearly indicating any outliers.

**(7)**

(*c*) Find the mean of these data.

**(2)**

(*d*) Comment on the skewness of these data. Justify your answer.

**(2)**

---

**5.** The discrete random variable $X$ has probability function

$$P(X = x) = \begin{cases} k(2 - x), & x = 0, 1, 2, \\ k(x - 2), & x = 3, \\ 0, & \text{otherwise,} \end{cases}$$

where $k$ is a positive constant.

(*a*) Show that $k = 0.25$.

**(2)**

(*b*) Find E($X$) and show that E($X^2$) = 2.5.

**(4)**

(*c*) Find Var($3X - 2$).

**(3)**

Two independent observations $X_1$ and $X_2$ are made of $X$.

(*d*) Show that P($X_1 + X_2 = 5$) = 0.

**(1)**

(*e*) Find the complete probability function for $X_1 + X_2$.

**(3)**

(*f*)   Find  $P(1.3 \le X_1 + X_2 \le 3.2)$.

**(3)**

**++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++**

**D  Answers**

**2. a) 0.1  b) 0.75  c) 0.6**
**4. a) Q2 = 16, Q1 = 15, Q3 = 16.5, IQR = 1.5   c) 16.1**
   **d) almost symmetrical / slight negative skew.**
   **Mean(16.1) ≈ median(16), and Q3 – Q2 (0.5) ≈  Q2 – Q1 (1.0)**
**5. b) 1  c) 13.5**
   **e)**

| y | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| P(Y=y) | 0.25 | 0.25 | 0.0625 | 0.25 | 0.125 | 0 | 0.0625 |

   **f) 0.3125**

**++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++**