

IGCSE Statistics: CRICKET PLAYER SELECTION (2)

NAME: *Dr. French* DATE: *May 2012*

Use the batting statistics for the previous year (see **Table 1**) to analyze the performance of AUGUSTUS WILLOW (player A) and HOWARD ZATT (player B). *Only one* should be selected to play for England, *you* must decide which one!



Instructions

1. Fill in **Table 1** and then use **Graphs 1 and 2** to plot frequency density plots ('histograms') for players A and B. Note the *axis scales are the same* to enable a *fair visual comparison*. Note the *modal score*, which is the *peak* of the histogram.

REMEMBER: Frequency density = $\frac{\text{Frequency (i.e. number of innings)}}{\text{range of score}}$

2. Fill in **Table 2** and then use **Graph 3** to plot a cumulative frequency curve. Use a *red* pen for player A and a *blue* pen for player B.

By finding where the 'S' shaped curves cut the 25%, 50% and 75% horizontal lines, estimate the **Lower Quartile (LQ), Median and Upper Quartile (UQ)** batting scores for each player.

3. Use the LQ, median and UQ results above to plot a **Box and Whisker diagram** for each player in **Graph 4** (which should be underneath the cumulative frequency diagram in **Graph 3**).

4. Using the information you have found out about each player, *justify which one should be selected*.

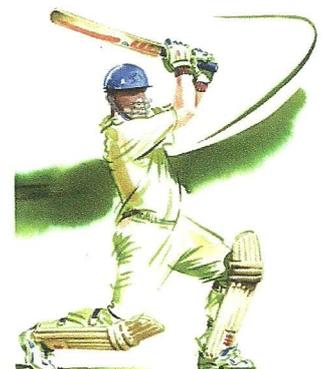


Table 1: Batting statistics for players A and B.

$$\text{Frequency density} = \frac{\text{Frequency (i.e. number of innings)}}{\text{range of score}}$$

| x= Runs scored per innings | Number of innings. Player A | Frequency density. Player A | Number of innings. Player B | Frequency density. Player B |
|----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| $0 < x \leq 10$ | 5 | $5/10 = 0.5$ | 13 | $13/10 = 1.3$ |
| $10 < x \leq 20$ | 11 | $11/10 = 1.1$ | 15 | $15/10 = 1.5$ |
| $20 < x \leq 30$ | 12 | $12/10 = 1.2$ | 15 | $15/10 = 1.5$ |
| $30 < x \leq 40$ | 15 | $15/10 = 1.5$ | 14 | $14/10 = 1.4$ |
| $40 < x \leq 60$ | 12 | $12/20 = 0.6$ | 8 | $8/20 = 0.4$ |
| $60 < x \leq 100$ | 10 11 | $11/40 = 0.275$ | 4 | $4/40 = 0.1$ |
| $100 < x \leq 150$ | 4 | $4/50 = 0.08$ | 1 | $1/50 = 0.02$ |
| TOTAL INNINGS | 70 | | 70 | |

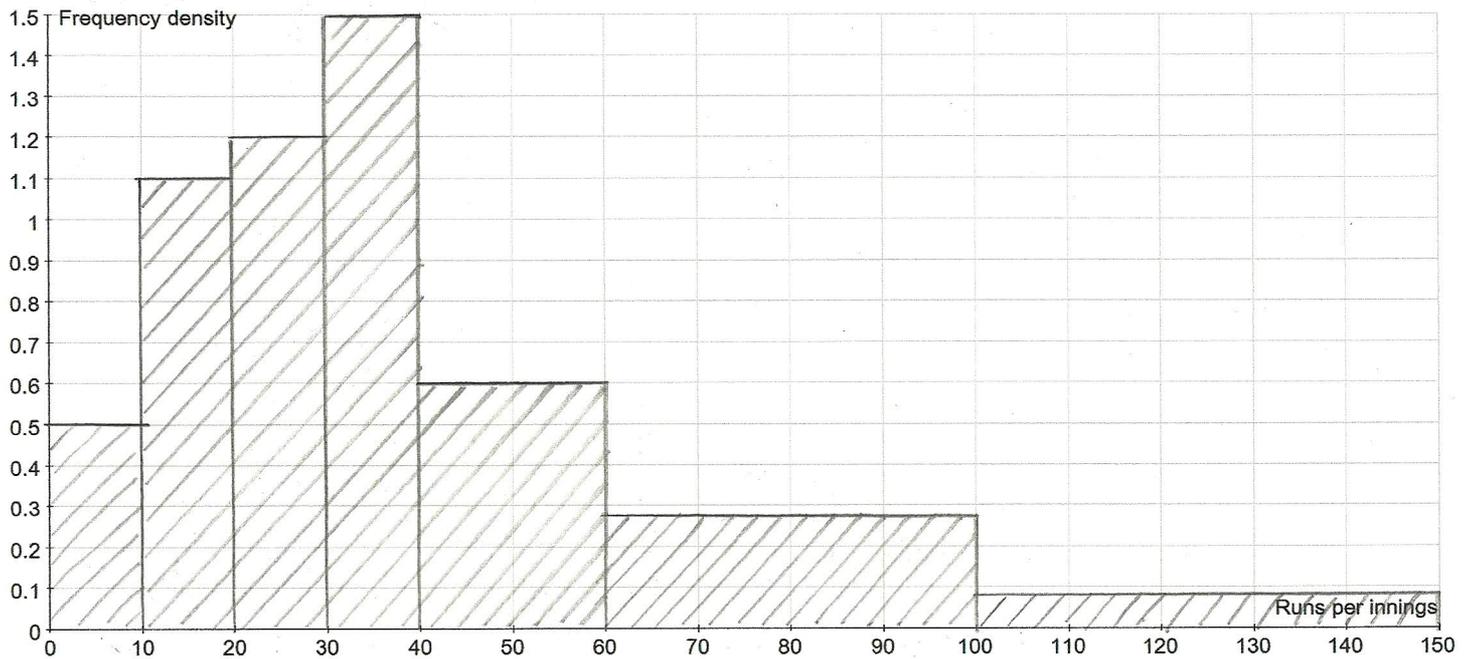
Min x = 0 Max x = 150 Min Frequency density = 0 Max frequency density = 1.5

Table 2: Cumulative frequency of innings vs runs scored.

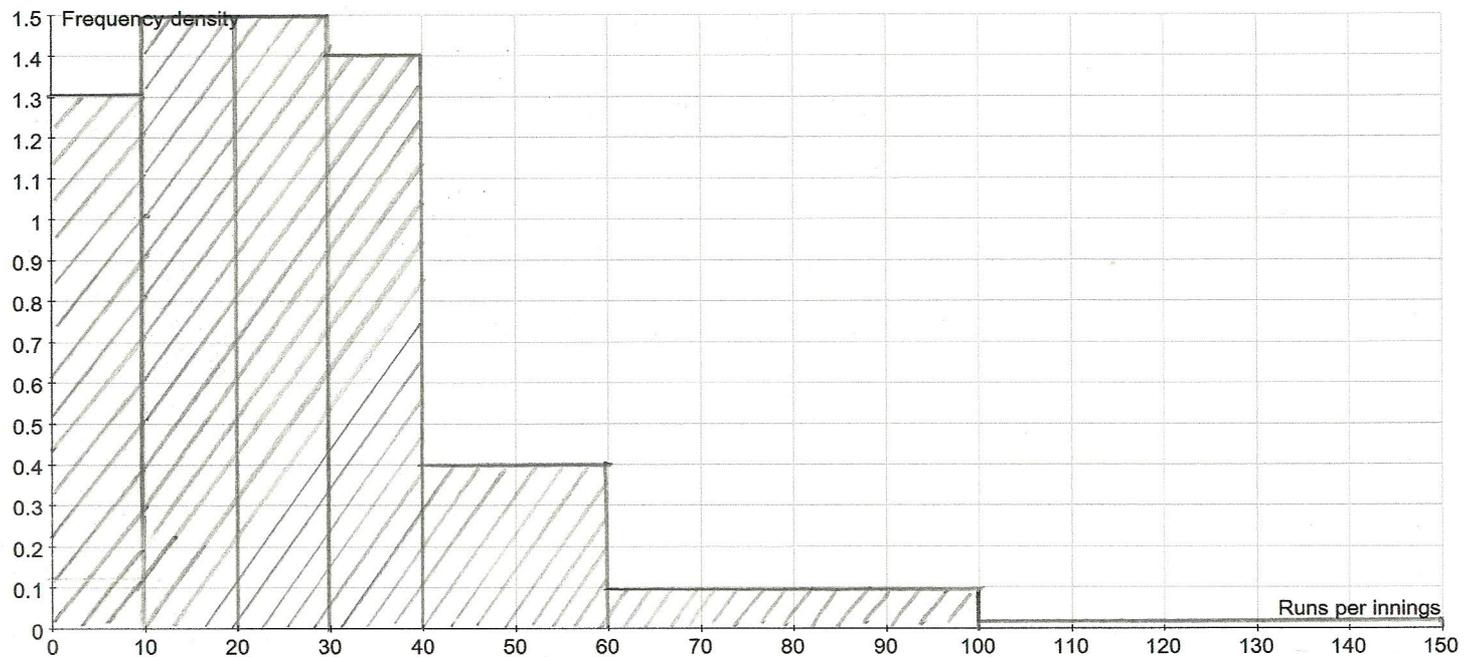
Since the total number of innings may be different for players A and B, divide the CUMULATIVE FREQUENCY N_A (or indeed N_B) by the total number of innings (see Table 1) and then multiply by 100 to get a percentage.

| M Maximum runs scored per innings | N_A Number of innings where score is \leq maximum runs scored. PLAYER A | CA (%) Divide N_A by Total number of innings, then multiply by 100 TOTAL INNINGS = 70 | N_B Number of innings where score is \leq maximum runs scored. PLAYER B | C_B (%) Divide N_B by Total number of innings, then multiply by 100 TOTAL INNINGS = 70 |
|--------------------------------------|--|--|--|---|
| 10 | 5 | $100 \times 5 / 70 = 7\%$ | 13 | $100 \times 13 / 70 = 19\%$ |
| 20 | 16 | $100 \times 16 / 70 = 23\%$ | 28 | $100 \times 28 / 70 = 40\%$ |
| 30 | 28 | $100 \times 28 / 70 = 40\%$ | 43 | $100 \times 43 / 70 = 61\%$ |
| 40 | 43 | $100 \times 43 / 70 = 61\%$ | 57 | $100 \times 57 / 70 = 81\%$ |
| 60 | 55 | $100 \times 55 / 70 = 79\%$ | 65 | $100 \times 65 / 70 = 93\%$ |
| 100 | 66 | $100 \times 66 / 70 = 94\%$ | 69 | $100 \times 69 / 70 = 99\%$ |
| 150 | 70 | $100 \times 70 / 70 = 100\%$ | 70 | $100 \times 70 / 70 = 100\%$ |

Graph 1. Frequency density plot for player A



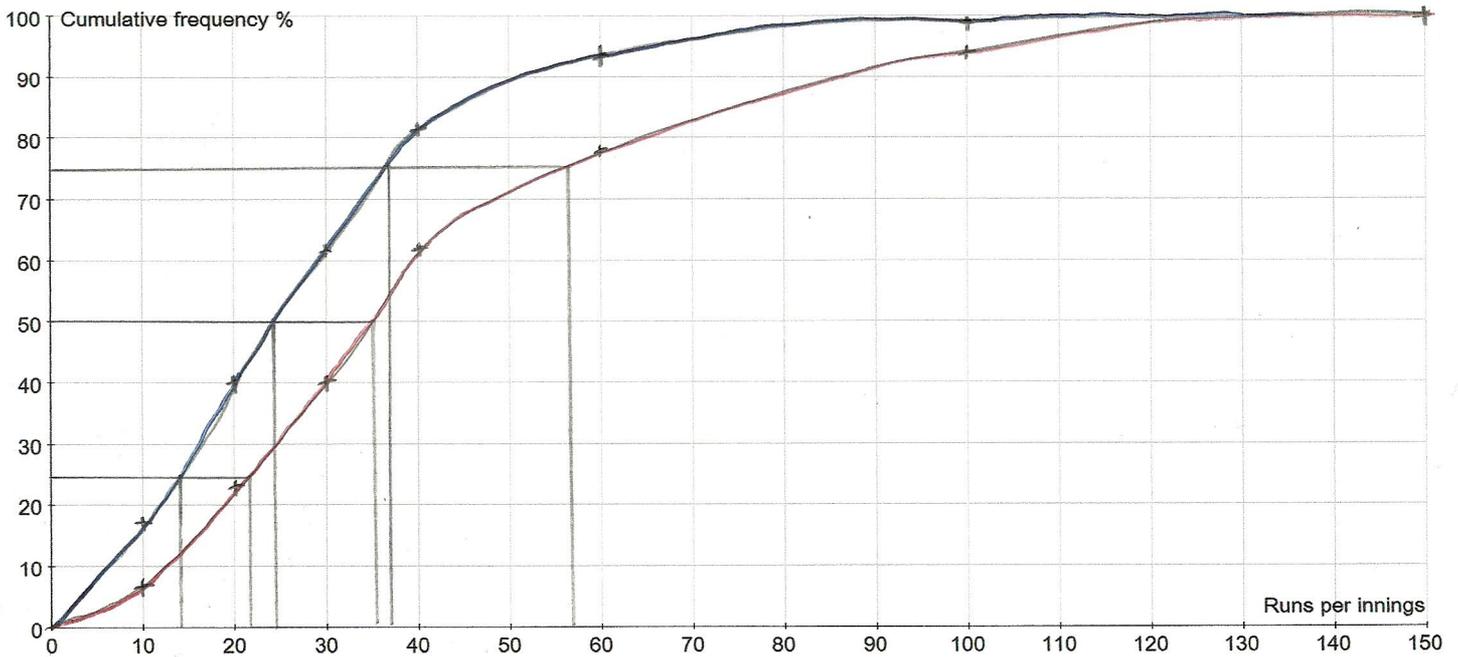
Graph 2. Frequency density plot for player B



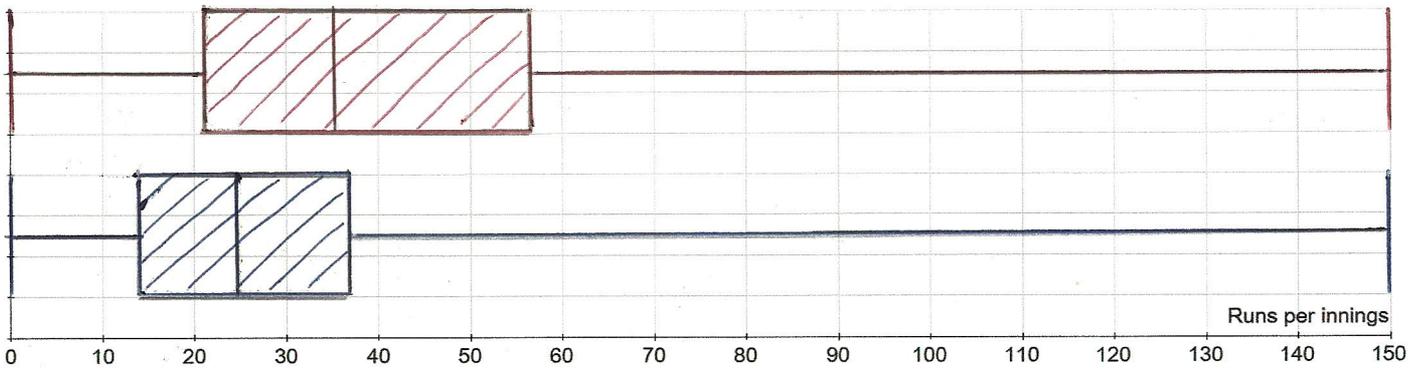
MATHS HEALTH WARNING! NOTE THESE TWO GRAPHS *USE THE SAME SCALE*. Graphs comparing two or more quantities (e.g. stock market price of a company vs time) can be *misleading* unless they both use the **same scales** in *both* horizontal and vertical axes.

Did player B have a *different* total number of innings than player A? What does that tell you about the total *areas* of the above graphs? What could we do to make the comparison fairer if they were different? (HINT: Imagine player B's statistics were for *ten* years and player A just one. Assume they both play equally regularly!)

Graph 3: Cumulative frequency plot for players A and B. (Use a red pen for A and a blue pen for B).



Graph 4: Box and whisker plot for players A and B.



PLAYER SELECTED: A (AUGUSTUS WILLOW)

Reason:

HIGHER MEDIAN SCORE (35 COMPARED TO 25)
 MORE LIKELY TO OCCASIONALLY SCORE MUCH HIGHER
 RUNS (HIGHER UQ AT 57 COMPARED TO 37)
 ALSO LA IS ALSO HIGHER, SO POOR PERFORMANCES
 OF AUGUSTUS ARE LESS POOR THAN HOWARD.