

IGCSE Statistics: CRICKET PLAYER SELECTION

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

Use the batting statistics for the previous year (see **Table 1**) to analyze the performance of RUFUS THWACK (player A) and IVOR NOTADUCK (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*.

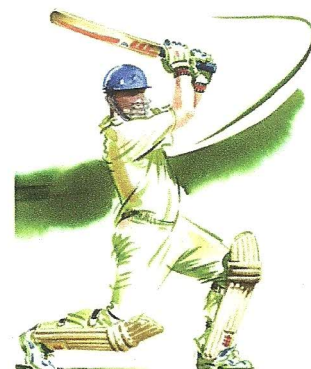


Table 1: Batting statistics for players A and B.

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

x= Runs scored per innings	Number of innings. Player A	Frequency density. Player A	Number of innings. Player B	Frequency density. Player B
$0 < x \leq 10$	10	$10/10 = 1.00$	2	$2/10 = 0.2$
$10 < x \leq 20$	5	$5/10 = 0.5$	5	$5/10 = 0.5$
$20 < x \leq 30$	12	$12/10 = 1.2$	15	$15/10 = 1.5$
$30 < x \leq 40$	15	$15/10 = 1.5$	14	$14/10 = 1.4$
$40 < x \leq 60$	4	$4/20 = 0.2$	8	$8/20 = 0.4$
$60 < x \leq 100$	2	$2/40 = 0.05$	6	$6/40 = 0.15$
$100 < x \leq 150$	2	$2/50 = 0.04$	10	$10/50 = 0.2$
TOTAL INNINGS	50		60	

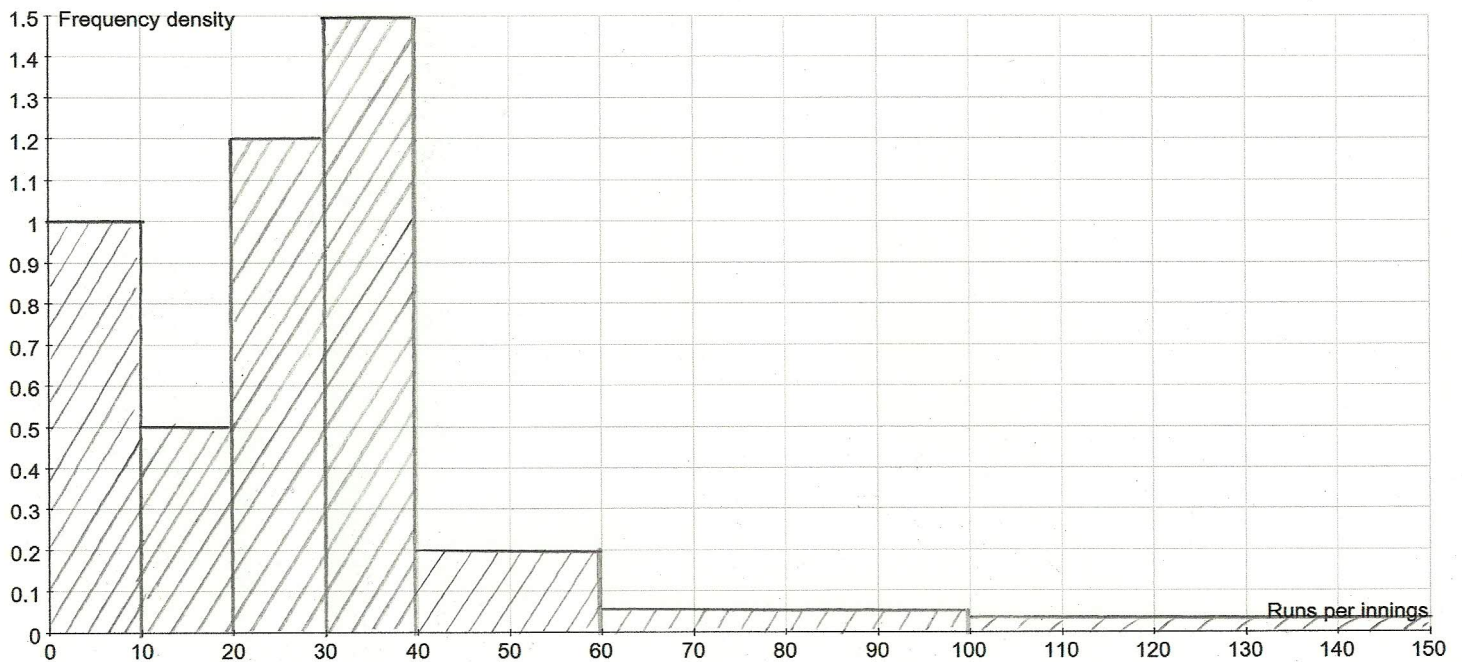
Min x = 0 Max x = 150 Min Frequency density = 0 Max frequency density = 1.5

Table 2: Cumulative frequency of innings vs runs scored.

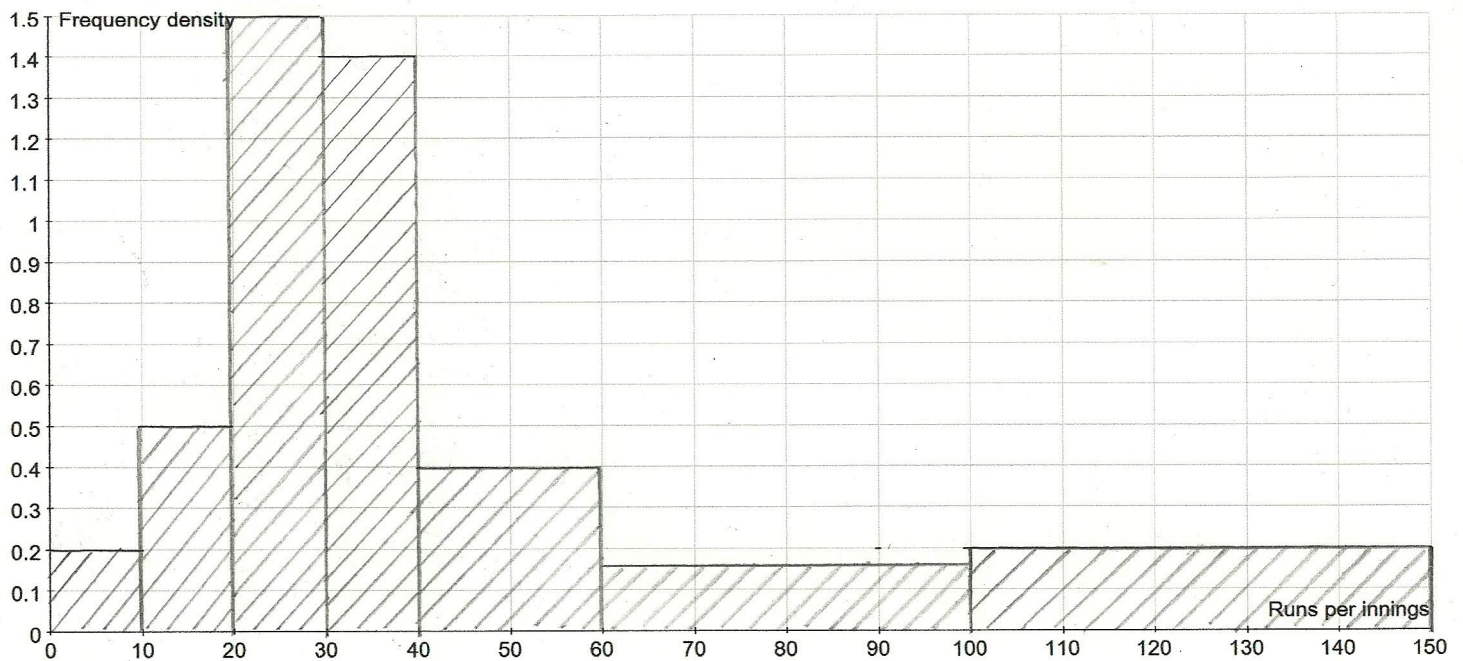
Since the total number of innings may be different for players A and B, divide the CUMULATIVE FREQUENCY N_A (or indeed N_B) by the total number of innings (see Table 1) and then multiply by 100 to get a percentage.

M Maximum runs scored per innings	N_A Number of innings where score is \leq maximum runs scored. PLAYER A	C_A (%) Divide N_A by Total number of innings, then multiply by 100 TOTAL INNINGS = 50	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 = 60
10	10	$100 \times 10/50 = 20\%$	2	$100 \times 2/60 = 3\%$
20	15	$100 \times 15/50 = 30\%$	7	$100 \times 7/60 = 12\%$
30	27	$100 \times 27/50 = 48\%$	22	$100 \times 22/60 = 37\%$
40	42	$100 \times 42/50 = 84\%$	36	$100 \times 36/60 = 60\%$
60	46	$100 \times 46/50 = 92\%$	44	$100 \times 44/60 = 73\%$
100	48	$100 \times 48/50 = 96\%$	50	$100 \times 50/60 = 83\%$
150	50	$100 \times 50/50 = 100\%$	60	$100 \times 60/60 = 100\%$

Graph 1. Frequency density plot for player A



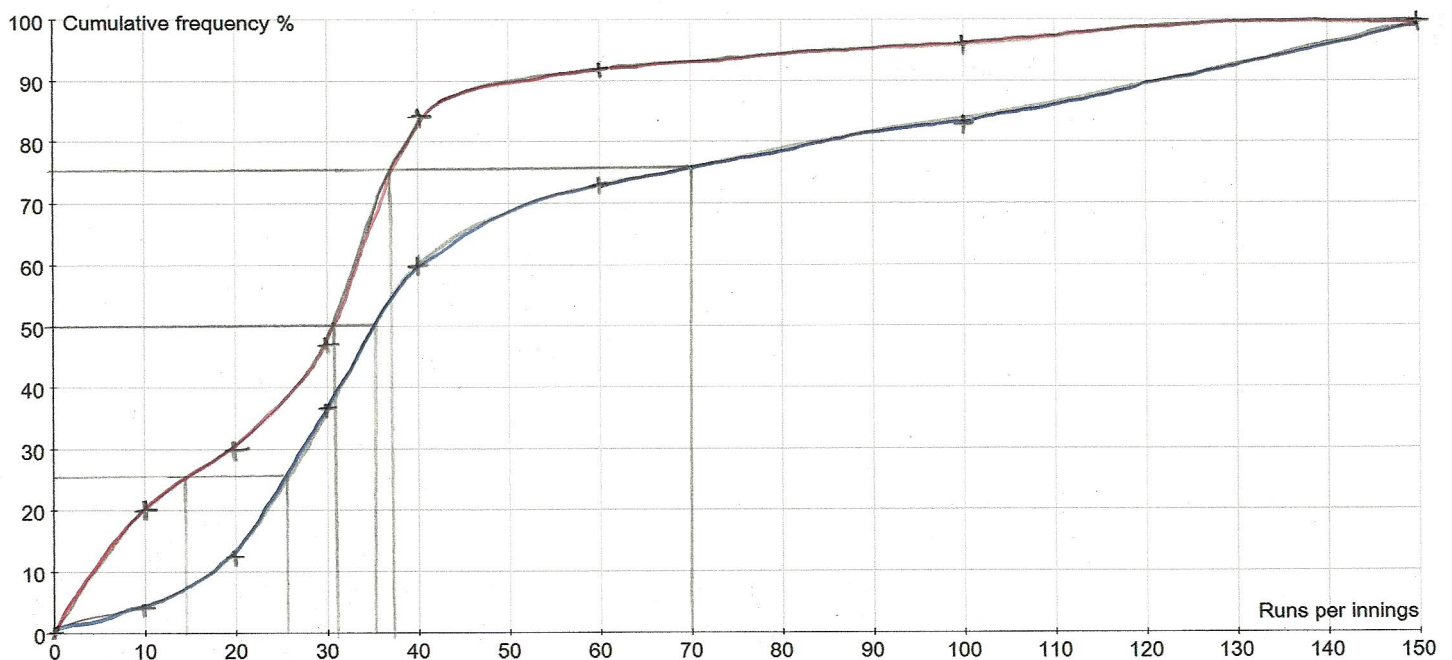
Graph 2. Frequency density plot for player B



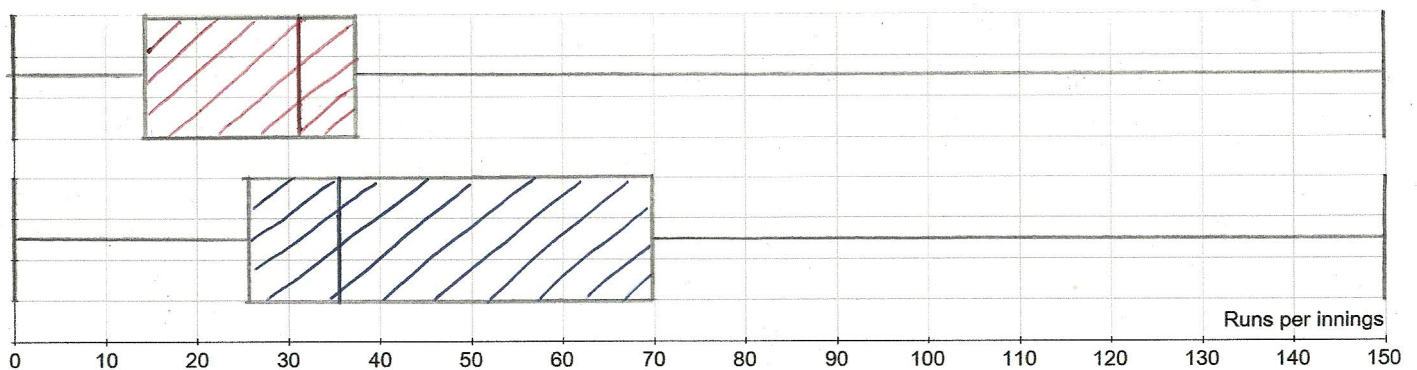
MATHS HEALTH WARNING! NOTE THESE TWO GRAPHS *USE THE SAME SCALE*. Graphs comparing two or more quantities (e.g. stock market price of a company vs time) can be *misleading* unless they both use the **same scales** in *both* horizontal and vertical axes.

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? (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: B (NOR NOTADUCK)

Reason:

MEDIAN SCORE IS SLIGHTLY HIGHER (35 COMPARED TO 31) BUT LOWER QUARTILE IS SIGNIFICANTLY HIGHER, SO IS UPPER QUARTILE. RUFUS HAS A TENDANCY TO SCORE LOW NUMBERS OF RUNS WHEREAS NOR OCCASIONALLY WILL SCORE VERY HIGHLY

I.E. IN THIS CASE IT IS THE INTER-QUARTILE RANGE RATHER THAN THE MEDIAN SCORE WHICH IS THE CLINCHER!