

## GREAT RUN STATISTICS

### CALCULATOR NEEDED

Two ten-mile charity road races occurred, race A and race B.

Your task is to perform a statistical analysis using the data below to *investigate the differences between the races*.

You will need to (using the blank graphs on the following pages)

1. Complete table 1 and hence plot **histograms** for the each race
2. From table 1, estimate the **mean** times for each race
3. Complete table 2 and hence plot **cumulative frequency graphs** for each race
4. From the cumulative frequency graphs, determine *estimates* of **lower quartiles**, **median** and **upper quartile** times, and hence plot a **box and whisker** for each race
5. Comment on your findings, using the data

**Table 1a (RACE A)**

Race time interval (minutes)	Number of athletes (frequency)	Frequency density (divide frequency by time interval)
$40 \leq t < 60$	300	15
$60 \leq t < 80$	2,500	125
$80 \leq t < 90$	4,000	400
$90 \leq t < 100$	5,000	500
$100 \leq t < 120$	5,000	250
$120 \leq t < 160$	7,500	187.5
$160 \leq t \leq 200$	2,000	50

Total athletes: **26,300**

**Table 2a (RACE A)**

Race time interval (minutes)	Number of athletes (cumulative frequency)	Cumulative frequency as a % of total frequency
$t \leq 60$	300	1.1
$t \leq 80$	2,800	10.6
$t \leq 90$	6,800	25.9
$t \leq 100$	11,800	44.9
$t \leq 120$	16,800	63.4
$t \leq 160$	24,300	92.4
$t \leq 200$	26,300	100

Total athletes: **26,300**

**Table 1b (RACE B)**

Race time interval (minutes)	Number of athletes (frequency)	Frequency density (divide frequency by time interval)
$40 \leq t < 60$	500	25
$60 \leq t < 80$	800	40
$80 \leq t < 90$	700	70
$90 \leq t < 100$	2,000	200
$100 \leq t < 120$	3,000	150
$120 \leq t < 160$	8,000	200
$160 \leq t \leq 200$	4,500	112.5

Total athletes: **19,500**

**Table 2b (RACE B)**

Race time interval (minutes)	Number of athletes (cumulative frequency)	Cumulative frequency as a % of total frequency
$t \leq 60$	500	2.6
$t \leq 80$	1,300	6.7
$t \leq 90$	2,000	10.3
$t \leq 100$	4,000	20.5
$t \leq 120$	7,000	35.9
$t \leq 160$	15,000	77.0
$t \leq 200$	19,500	100

Total athletes: **19,500**

**ESTIMATE MEAN TIMES FOR:**

RACE A: 
$$\bar{t}_A = \frac{50 \times 300 + 70 \times 2500 + 85 \times 4000 + 95 \times 5000 + 110 \times 5000 + 140 \times 7500 + 180 \times 2000}{26,300}$$
  

$$= \boxed{112.7} \text{ minutes}$$

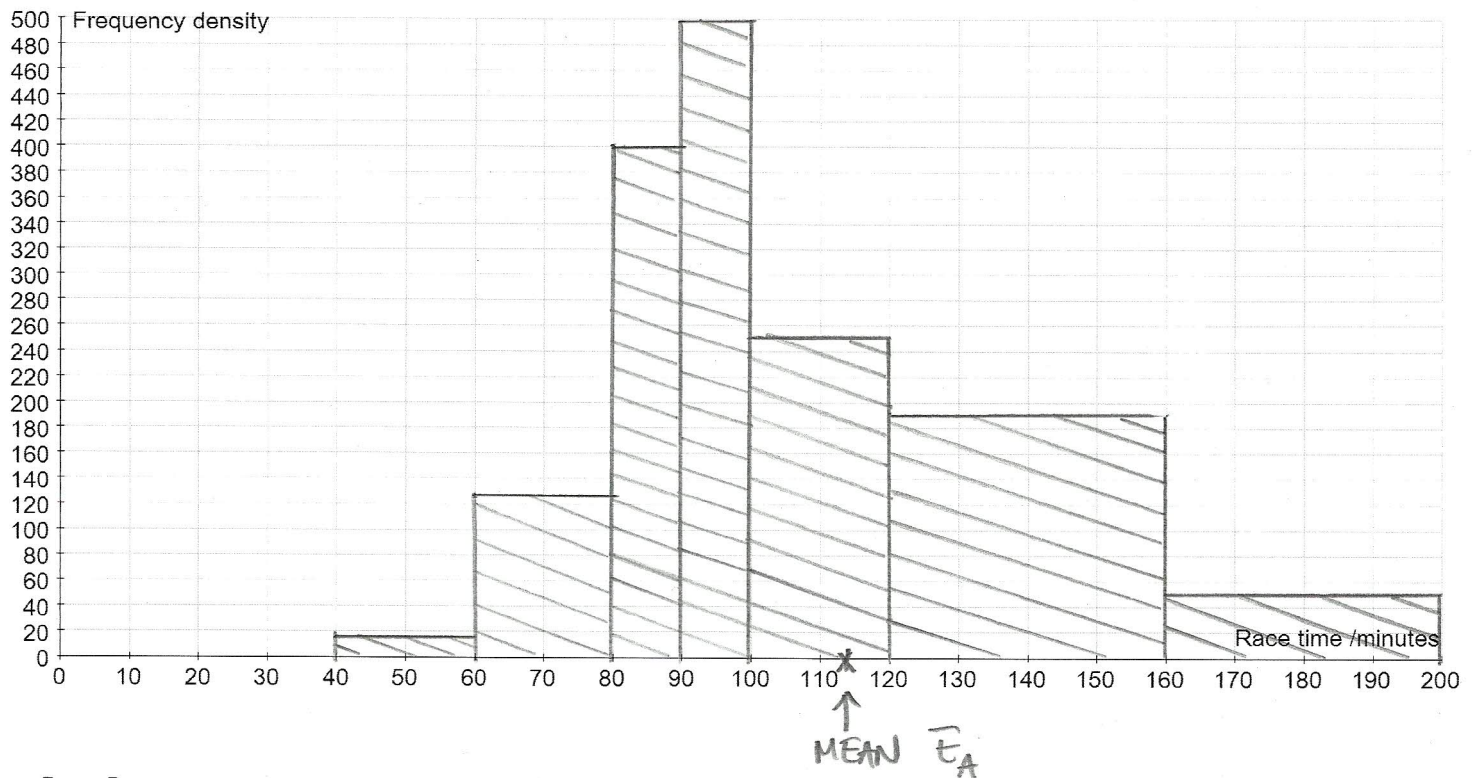
RACE B: 
$$\bar{t}_B = \frac{50 \times 500 + 70 \times 800 + 85 \times 700 + 95 \times 2000 + 110 \times 3000 + 140 \times 8000 + 180 \times 4500}{19,500}$$
  

$$= \boxed{132.8} \text{ minutes}$$

## Histograms

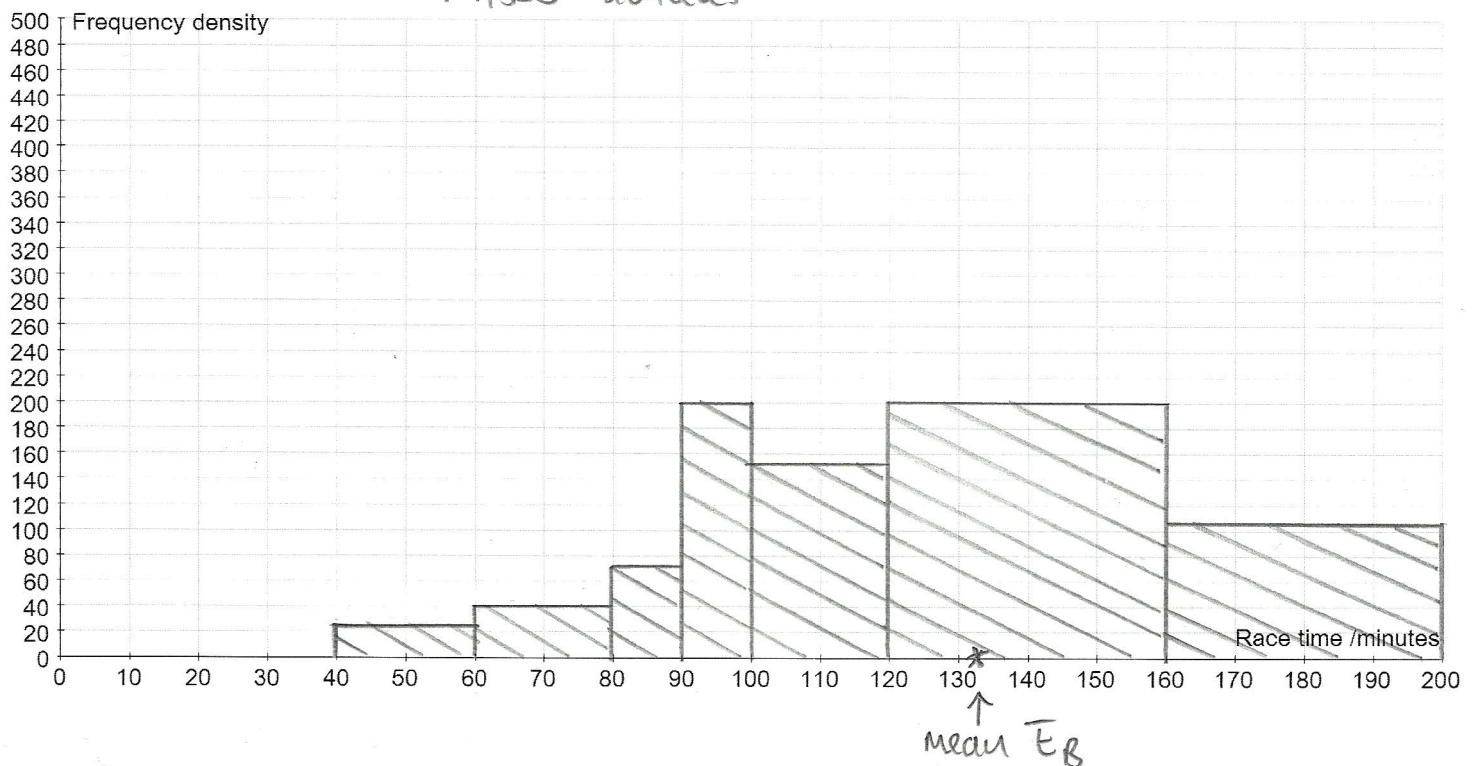
Race A

26,300 athletes



Race B

19,500 athletes



What is the obvious 'statistical health warning' here? Could you suggest a mathematical change which would enable a more *fair* comparison between the histograms?

Races A and B do not have the same # of athletes

(26,300 ; 19,500).  $\therefore$  areas of histograms A and B

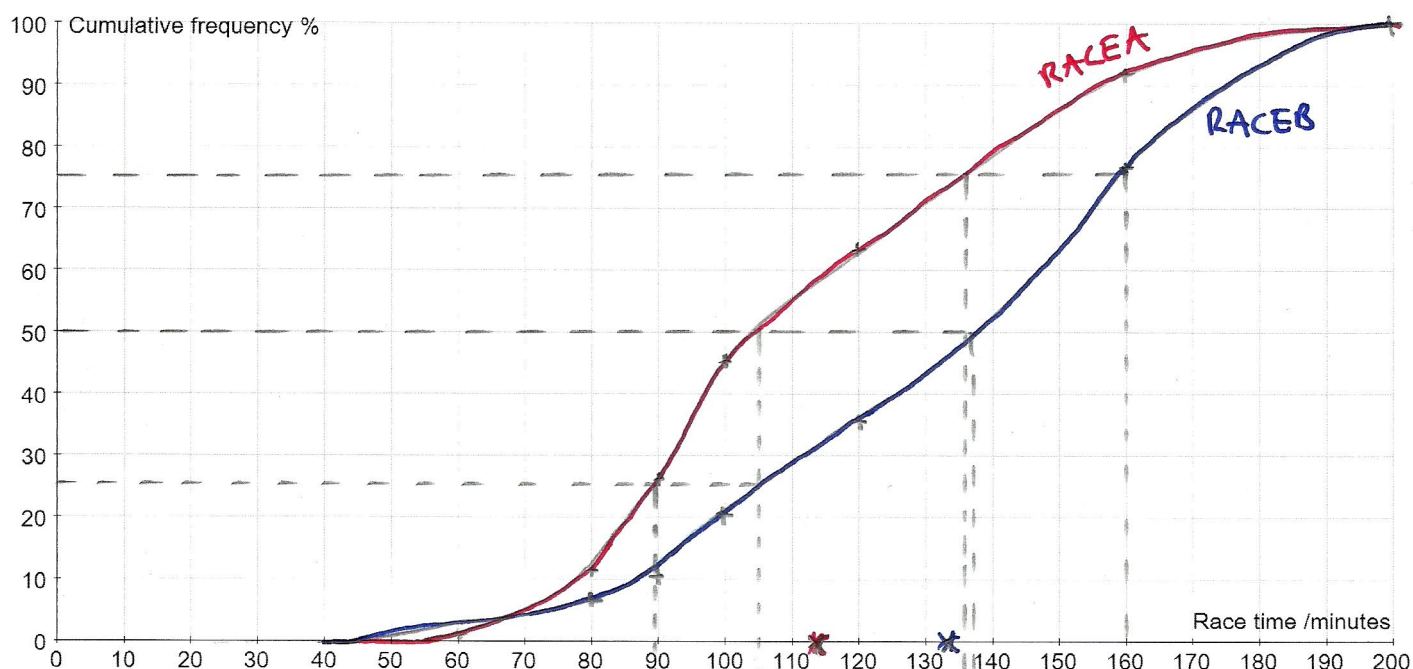
will not be the same, since area = total frequency.

To make a fair comparison NORMALIZE i.e. divide each frequency density by the # athletes.

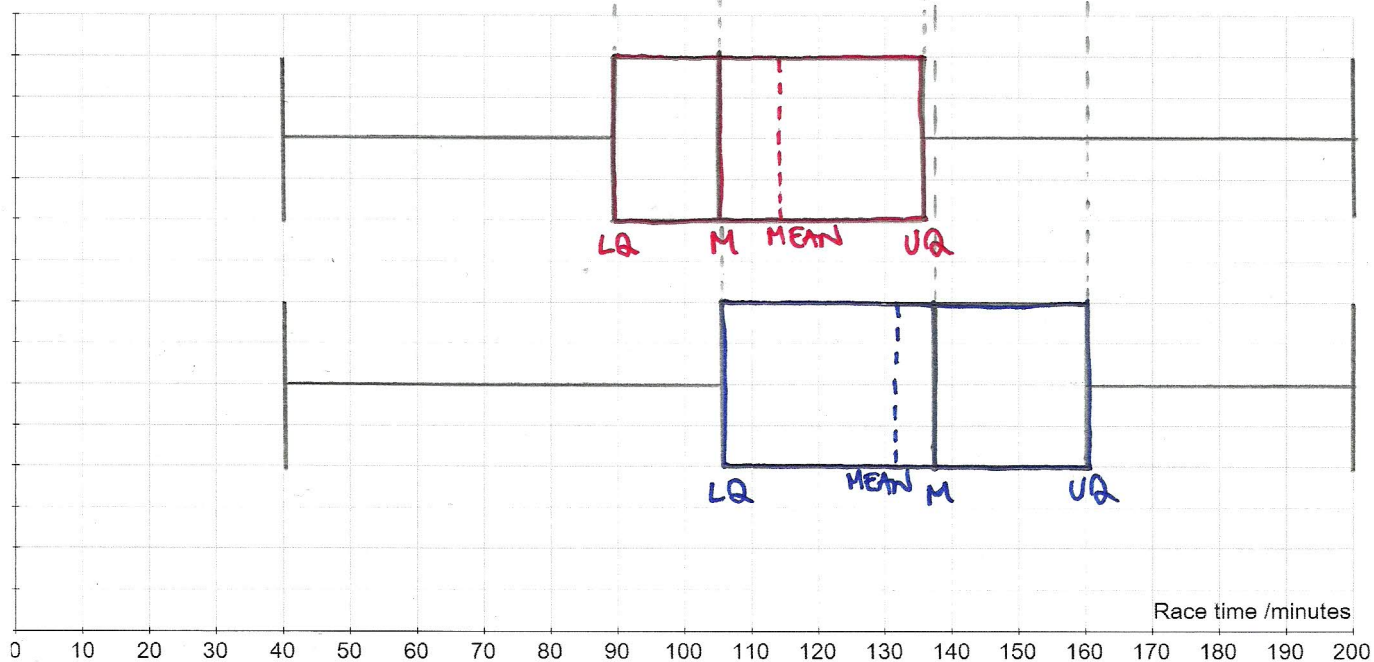


## Cumulative frequency graph

Use a red pen for race A and a blue pen for race B



Box and whisker plot. Separate races A and B vertically by a few squares



**RACE A**

LQ = 90  
 MEDIAN = 105  
 UQ = 136  
 IQR = 46  
 MEAN = 112.7

**RACE B**

LQ = 105  
 MEDIAN = 138  
 UQ = 160  
 IQR = 55  
 MEAN = 132.8

Describe, using the analysis you have performed, how the races differ. Use numbers and facts, not opinions!

- RACE A (26,300 ATHLETES) WAS 35% BETTER ATTENDED THAN RACE B (19,500 ATHLETES).
- OVERALL, A GREATER PROPORTION OF ATHLETES FROM RACE B RAN SLOWER THAN IN RACE A I.E. THE DISTRIBUTION (X HISTOGRAM) IS SKEWED TO LONGER TIMES
- THE AVERAGE TIME FOR RACE A WAS  $\approx 33$  MINUTES FASTER THAN RACE B. (105 mins : 138 mins, USING MEDIAN)
- THE MEDIAN TIMES FOR BOTH RACES DIFFER FROM THE MEAN - INDICATING A SKEW TO THE HISTOGRAMS
- THE IQR IS SIMILAR FOR BOTH RACES  
→ HISTOGRAM IS NEEDED TO SEE THE SKEW  
YOU WOULD FIND IT HARD TO SPOT IT JUST FROM THE CUMULATIVE FREQUENCY GRAPH