

2023 STATISTICS

Questions ○ Memos ○ Diagnostic Report



THE
ANSWER
SERIES *Your Key to Exam Success*

STATISTICS (70%): DBE NOVEMBER 2023

QUESTION 1 84%

Truck drivers travel a certain distance and have a rest before travelling further. A driver kept record of the distance he travelled (in km) on 8 trips and the amount of time he rested (in minutes) before he continued his journey. The information is given in the table below.

Distance travelled (in km) (x)	180	200	400	600	170	350	270	300
Amount of rest time (in minutes) (y)	20	25	55	120	15	50	40	45

- 1.1 Determine the equation of the least squares regression **90%** line for the data. (3)
- 1.2 If a truck driver travelled 550 km, predict the amount of **84%** time (in minutes) that he should rest before continuing his journey. (2)
- 1.3 Write down the correlation coefficient for the data. **87%** (1)

MEMOS

- 1.1 The equation: $y = \mathbf{A} + \mathbf{B}x$
 $\mathbf{A} = -23,8461\dots \approx -23,85$
 $\mathbf{B} = 0,2270\dots \approx 0,23$
 $\therefore y = -23,85 + 0,23x <$



1.2 Subst. $x = 550$: $y = -23,85 + 0,23(550)$
 $= \mathbf{102,65 \text{ minutes} <$

1.3 The correlation coefficient, $\mathbf{r} = 0,9828\dots$
 $\approx \mathbf{0,98 <$

Common Errors and Misconceptions

- (a) When writing the equation in **Q1.1** a few candidates **interchanged the values of a and b**. Some calculated the values of a and b correctly but **did not write the** required **equation** or they swapped the values of a and b in the equation. Another common error in this question was that candidates **failed to round off** their answers for **A** and **B** correctly **to two decimal places**.
- (b) In **Q1.2** some candidates **substituted 550 for y** instead of substituting 550 **for x**.
- (c) Some candidates **rounded off** the value of **r** to **1** instead of **0,98** when answering **Q1.3**. Candidates should have been aware that **0,98** and **1** have **very different meanings when analysing correlation coefficients**.

QUESTION 1 (cont.)

1.4 Interpret your answer to QUESTION 1.3.

67%

1.5 At each stop, the truck driver spent money buying food and other refreshments. The amount spent (in rands) is given in the table below.

100	150	130	200	50	180	200	190
-----	-----	-----	-----	----	-----	-----	-----

1.5.1 Calculate the mean amount of money he spent at each shop.

1.5.2 Calculate the standard deviation for the data.

1.5.3 At how many stops did the driver spend an amount that was less than one standard deviation below the mean?

(1)

(2)

(1)

(2)

[12]

MEMOS

1.4 There is a very strong positive correlation between the distance travelled and the amount of rest time. ◀

1.5.1 The mean, $\bar{x} = \frac{1200}{8} = \text{R}150$ ◀

1.5.2 The standard deviation, $\sigma = 50,4975\dots$
 $\approx 50,50$ ◀

1.5.3 Amount $< \bar{x} - 1\sigma$

\therefore Amount $< 150 - 50,50$

\therefore Amount $< 99,50$

\therefore **Only 1 stop was less (when he spent R50)** ◀

Common Errors and Misconceptions

- (d) When answering **Q1.4** some candidates either wrote 'strong' or 'positive' as their answer. Both were unacceptable as they **only described one attribute** of the correct answer.
- (e) In **Q1.5.2** some candidates **did not round off** their answer correctly **to two decimal places**.
- (f) Many candidates calculated the standard deviation interval correctly in **Q1.5.3** but **failed to indicate the number** of stops that were **less than one standard deviation below the mean**. They were not awarded full marks because they **did not answer the question completely**.



QUESTION 1: Suggestions for Improvement



- (a) Teachers should **link** the equation of the least squares regression line ($y = \mathbf{A} + \mathbf{B}x$) with the equation of the **straight line** and emphasise that '**A**' refers to the ***y*-intercept** and '**B**' refers to the ***gradient***.
- (b) Learners **must be careful** not to interchange **the x- and y-values** in the table when they input these into the calculator.
- (c) When determining the equation of the least squares regression line, it is advisable that learners **write down the values of a and b** and then **write down the equation** of the regression line. In this way, they can get the CA mark for the equation.
- (d) When teaching *Statistics*, the **focus** should **not only** be the **calculations**. Teachers should also **pay attention to the meaning** of the **different concepts**, e.g. ***mean, standard deviation, correlation coefficient, etc.*** The values obtained in the calculations should then become more meaningful for learners.
- (e) The understanding of statistical **terminology** is developed by using these terms frequently in the class. **The use of diagrams** when explaining the **concepts of standard deviation** and **deviation intervals from the mean** should **help learners in understanding these concepts**.
- (f) **Practice calculator skills** with learners. Learners **should be familiar with** what the **symbols** on the calculator represent, for example, σx represents ***population standard deviation*** and r represents ***correlation coefficient***.
- (g) Learners should be able to **use the values** of their calculations to **make predictions and comments** about the data. **Time should be devoted** to **interpretation questions**.



QUESTION 2 50%

At a certain school, the staff committee wanted to determine how many glasses of water the staff members drank during a school day. All staff members present on a specific day were interviewed. The information is shown in the table below.

NUMBER OF GLASSES OF WATER DRANK PER DAY	NUMBER OF STAFF MEMBERS	CUMULATIVE FREQUENCY
$0 \leq x < 2$	5	
$2 \leq x < 4$	15	
$4 \leq x < 6$	13	
$6 \leq x < 8$	5	
$8 \leq x < 10$	2	

2.1 Complete the cumulative frequency column provided in the table above. (2) **90%**

2.2 How many staff members were interviewed? (1) **95%**

2.3 How many staff members drank fewer than 6 glasses of water during a school day? (1) **58%**



MEMOS

2.1

NUMBER OF GLASSES OF WATER DRANK PER DAY	NUMBER OF STAFF MEMBERS	CUMULATIVE FREQUENCY
$0 \leq x < 2$	5	5
$2 \leq x < 4$	15	20
$4 \leq x < 6$	13	33
$6 \leq x < 8$	5	38
$8 \leq x < 10$	2	40

2.2 40 staff members were interviewed ◀

2.3 33 staff members drank fewer than 6 glasses of water ◀

Common Errors and Misconceptions

- (a) A few candidates calculated the midpoint of the class intervals when answering **Q2.1**. These candidates were attempting to calculate the **estimated mean** of the data instead of the **cumulative frequencies**.
- (b) In **Q2.3** many candidates were unable to differentiate between **frequency** and **cumulative frequency**. They gave the answer as 13, the frequency of the interval $4 \leq x < 6$, instead of the correct answer of 33.

QUESTION 2 (cont.)

18% 2.4 The staff committee observed that k staff members were absent on the day of the interviews. It was found that half of these k staff members drank from 0 to fewer than 2 (that is $0 \leq x < 2$) glasses of water per day, while the remainder of them drank from 4 to fewer than 6 (that is $4 \leq x < 6$) glasses of water per day. When these k staff members are included in the data, the estimated mean is 4 glasses of water per staff member per day.

How many staff members were absent on the day of the interviews?



(4)

[8]

Common Errors and Misconceptions

(c) In response to **Q2.4** reading for meaning proved to be **a challenge** for many candidates. **Some** of the candidates who were able to understand the question made careless errors when setting up the equation. They left out the **brackets** and wrote $5 \times 13 + \frac{k}{2}$ instead of $5 \times \left(13 + \frac{k}{2}\right)$. Some candidates failed to realise that the **total number of teachers** was now **$40 + k$** . Other candidates **incorrectly subtracted the frequencies**, i.e. $13 - 5$, and **arrived at the correct answer** of 8. They were not awarded any marks for this attempt. A number of candidates only wrote down 8, the correct answer. These candidates were **only awarded one mark as** there was **no working to substantiate their answer**.

MEMOS

2.4

NUMBER OF GLASSES OF WATER DRANK PER DAY	MIDPOINT	NUMBER OF STAFF MEMBERS
$0 \leq x < 2$	1	$5 + \frac{1}{2}k$
$2 \leq x < 4$	3	15
$4 \leq x < 6$	5	$13 + \frac{1}{2}k$
$6 \leq x < 8$	7	5
$8 \leq x < 10$	9	2
		$40 + k$

Estimated mean = 4 when k staff members are added

Total no. of glasses of water

$$= \left[\left(5 + \frac{1}{2}k\right) \times 1 \right] + (15 \times 3) + \left[\left(13 + \frac{1}{2}k\right) \times 5 \right] + (5 \times 7) + (2 \times 9)$$

$$\therefore \text{Estimated mean} = \frac{5 + \frac{1}{2}k + 45 + 65 + \frac{1}{2}k + 35 + 18}{40 + k} = 4$$

$$\therefore 168 + 3k = 160 + 4k$$

$$\therefore 8 = k$$

\therefore 8 staff members were absent <



THE ANSWER
SERIES *Your Key to Exam Success*

QUESTION 2: Suggestions for Improvement



- (a) Learners need to understand **the difference** in the meaning of the concepts: **frequency** and **cumulative frequency**.
- (b) **Reading for understanding** is a **fundamental requirement** in the *Data Handling* section. **This skill needs to be developed** in classroom activities.
- (c) While the **calculator** is a **useful tool** in answering many questions in the *Data Handling* section, **teachers cannot overlook performing the calculations manually**. This **assists learners** in **understanding** the calculation but moreover, it provides learners with the necessary skills to deal with questions that include variables.