SHARE ONLINE REVISION

General Certificate of Education Advanced Level

STATISTICS 6046/1

PAPER 1

ZIMSEC 2021 SOLUTION 3 hours

Additional materials:
Answer paper
Graph paper
List of formulae MF7
Electronic calculator (Non-programmable)

Time 3 hours

INSTRUCTIONS TO CANDIDATES

Write your name in the spaces provided on the answer sheet/answer booklet.

Answer ***all*** questions.

If a numerical answer cannot be given exactly, and the accuracy required is not specified in the question, then in the case of an angle it should be given correct to the nearest degree, and in other cases it should be given to 2 significant figures.

INFORMATION TO CANDIDATES

The number of marks is given in brackets [ ] at the end of each question or part of question.

The total number of marks for this paper is 120.

The use of a scientific calculator is expected, where appropriate.

You are reminded of the need for clear presentation in your answers.

This question paper consists of 7 printed pages and 1 blank page.

1.

a) State the central limit theorem. [2]

b) 120 bags of mealie-meal of a particular brand are weighed and the mean mass is found to be 750g with standard deviation of 3.9g.

Find a 95% confidence interval for the mean mass of bags of mealie-meal of this brand. [3]

SOLUTION

a) **$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, provided that the sample size taken from a non-normal, n, is large ($n \geq 30$) and σ is unknown.**

b) Let X be the number of mealie-meal bags

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$\mu = 750 \;\; and \;\; \hat{\sigma}^2 = \sqrt{\frac{3.9^2}{120-1}}$

Now, 95% confidence interval for the mean mass of bags of mealie-meal of this brand is;

$$\bar{x} \pm 1.96\left(\frac{\hat{\sigma}}{\sqrt{n-1}}\right)$$

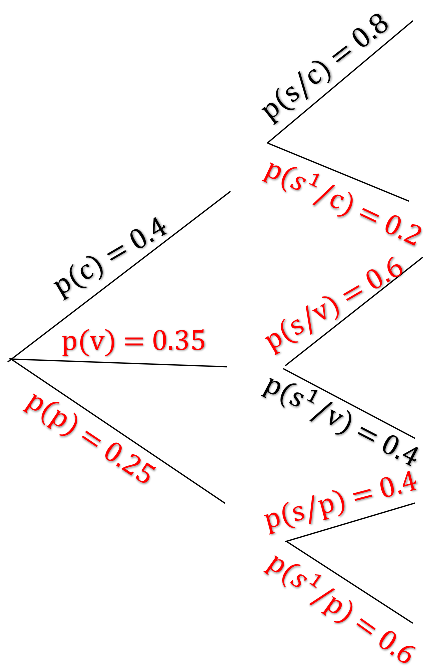
$= 750 \pm 1.96\left(\frac{3.9}{\sqrt{120-1}}\right)$

$= (749.2992757; 750.7007243)$

= (749.30g; 750.70g).

2. A street ice cream vendor sells 3 types of ice cream of flavours, chocolate, vanilla and peach. Of the sells, 40% is chocolate, 35% is vanilla, and 25% is peach. Sales are by cone or cup. The percentage of cone sales for chocolate, vanilla and peach are 80%, 60% and 40% respectively. Let C, V and P represent the sales of chocolate, vanilla and peach flavours respectively.

a) Copy and complete the tree diagram to illustrate the relevant probabilities of a randomly selected sale of ice cream.



[3]

b) Find the probability of selling ice by cone. [2]

SOLUTION

a) On diagram.

b) Probability of selling an ice cream cone is;

P(selling an ice cream cone) $= [(p(c) \times p(s\backslash c) + p(v) \times p(s\backslash v) + p(s\backslash p)]$

$= [(0.4 \times 0.8) + (0.35 \times 0.6) + (0.25 \times 0.4)]$

$\mathbf{= 0.63}$

3. A committee of 5 people is to be chosen from 3 men and 6 women.

   Calculate the

   a) number of ways the committee can be chosen, [2]

   b) probability of choosing a committee with at least 2 men. [3]

SOLUTION

a) 5 people are to be chosen from a total of (3+6)= 9 people

   Then, number of ways the committee can be chosen is;

   $^{9}_{5}C = 126$ **ways.**

b) Choosing at least 2 men, means men must be 2 or more in a committee of 5 people.

   Hence, $({}^{3}_{2}C \times {}^{6}_{3}C) + ({}^{3}_{3}C \times {}^{6}_{2}C) = 60 + 15 = 75$

   Now, $n(E) = 75$ and $n(S) = 126$

   So, $P(E) = \frac{n(E)}{n(S)}$

   $= \frac{75}{126}$

   $= \frac{25}{42}$

   $= \mathbf{0.595}$ **(3 s.f)**

4. In manufacturing a certain metal tube, there are 3 distinct processes X,Y and Z. The mean time of each process and the standard deviation of these times are given in the table below.

	Mean times (in seconds)	standard deviation
Process X	30	3
Process Y	10	1
Process Z	20	2

During the course of its manufacturing, each metal tube is subjected to process X and Z only once and to process Y three times. Assuming that the times spend on each stage are independently and normally distributed,

Find the

a) probability that metal tube would take less than 75 seconds to manufacture, [3]
b) percentage of metal tubes which would take more than 87 seconds to manufacture. [3]

SOLUTION

a) For process X and Z i.e. $(\mathrm{X}+\mathrm{Z})$

$\mu = \mathrm{E(X+Z)} = \mathrm{E(X)} + \mathrm{E(Z)} = 30 + 20 = 50$

$\sigma^2 = \mathrm{Var(X)} + \mathrm{Var(Z)} = 3^2 + 2^2 = 13$

Now, $\mathbf{X+Z \sim N(50, 13)}$

Also, for Y, $\mathrm{E(3Y)} = 3 \times 10 = 30$ and $\mathrm{Var(3Y)} = 3^2(1) = 9$ hence, $\mathbf{Y \sim N(30, 9)}$

So, $\mathbf{X+Y+Z \sim \boldsymbol{N}(80, 22)}$

Let X+Y+Z be M

So, $\mathrm{P(M < 75)}\ P\left(z < \frac{75-80}{\sqrt{22}}\right) = P(z < -1.066)$

$= 1 - \emptyset(1.066)$

$= 1 - 0.8568$

$= 0.1432$

$= \mathbf{0.143}$

b) $P(M > 87) = P\left(M > \frac{87-80}{\sqrt{22}}\right) = P(z > 1.492)$

$= 1 - \emptyset(1.492)$

$= 1 - 0.9322$

$= 0.0678$

Hence, the percentage of metal tubes which would take more than 87 seconds to manufacture is $(\mathbf{0.0678 \times 100})$ = 6.78%

5.

a) Give an example of each of the following components of a time series:

i) trend

ii) cyclical

iii) seasonal

iv) irregular/random. [4]

b) State one advantage of using a 5-point moving average compared to a 3-point moving average. [2]

SOLUTION

a)

i) Changes in population growth, income, wealth.

ii) Influences in the economy exerted by trade unions, world organization which are reflected in changes in time series levels.

iii) Production of a certain farm produce.

iv) Natural disaster i.e. flood, draught.

b) Fast when large amounts of data are involved and will also smooth out the trend more effectively.

6.

a) Define the following terms giving examples :

i) measures of location

ii) measures of dispersion [4]

b) The following data represents the number of goals scored over a season by 30 players of a local football club.

17 21 35 64 2 0 20 44 52 81 47 21 14 10 0

2 1 1 3 1 5 2 1 18 0 1 19 1 2 0

i) Construct a grouped frequency distribution table with equal class intervals.

ii) Deduce the shape of the distribution of goals. [3]

SOLUTION

a)

i) Measures of location are statistical methods used to show representative value in the data, for example mean, mode and median.

ii) Measures of dispersions are methods used to measure variability or spread of data in the data set, for example variance and standard deviation.

b) i)

Goals	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
Frequency	16	5	3	1	2	1	1	0	1

ii) Bell shaped.

7. Data about employment of males and females in a town are shown in the below.

Sex	Unemployment	Employed	Total
Male	412	206	618
Female	305	358	663
Total	717	564	1 281

A person from the town is chosen at random. Let A be the event, 'the person is male' and let B be the event, 'the person is employed.'

a) Find

i) P(A)

ii) P(A and B)

iii) Are events A and B independent? Justify your answer. [4]

b) Given that the person is unemployed, find the probability that then person is female. [3]

SOLUTION

a)

i) $P(A) = \frac{618}{1\ 281} = \frac{206}{427} = \mathbf{0.48}$

ii) $P(A \text{ and } B) = \frac{206}{1\ 281} = \mathbf{0.16}$

iii) $P(A) \times P(B) = P(A \text{ and } B)$

LHS: $(\frac{618}{1281} \times \frac{564}{1281}) = 0.2124072419$

Hence, $P(A) \times P(B) \neq P(A \text{ and } B)$, so events A and b are not independent.

b) $\text{P(unemployed)} = 1 - P(B) = 1 - \frac{564}{1284} = \frac{717}{1281}$

Now, $\text{P (unemployed | female)} = \frac{\text{P(unemployed and female)}}{\text{P(unemployed)}}$

$= \frac{(305/1281)}{(717/1281)} = \frac{305}{717} = \mathbf{0.425}$

8.

a) Distinguish between correlation and regression. [2]

b) An electric oven was switched on and the temperature of the oven was noted at five minute interval.

Time (x) minutes from switching off oven	0	5	10	15	20	25	30	35	40
Temperature	0.4	1.5	3.4	5.5	7.7	9.7	11.7	13.5	15.5

i) Calculate the summary statistics that would be used to find the product moment correlation coefficient. [2]

ii) Hence, find the product moment correlation coefficient and comment on its value. [4]

SOLUTION

a) Correlation is a single statistic, or data point, whereas regression is the entire equation with all of the data points that are represented with a line.

Correlation shows the relationship between the two variables, while regression allows us to see how one affects the other.

b)

i) $\sum x = \mathbf{180}$, $\sum x^2 = \mathbf{5100}$, $n = \mathbf{9}$, $\sum xy = \mathbf{1964}$, $\sum y = \mathbf{68.9}$, $\sum y^2 = \mathbf{756.99}$

ii) Product moment correlation coefficient (r),

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{n\sum x^2 - (\sum x)^2} \times \sqrt{n\sum y^2 - (\sum y)^2}}$$

$$r = \frac{9(1964)-(180)(68.9)}{\sqrt{9(5100)-(180)^2} \times \sqrt{9(756.99)-(68.9)^2}}$$

$\boldsymbol{r = 0.9987105013}$

There is a strong positive correlation between the temperature of the oven and time.

9. Five independent measures of the diameter of a ball bearing were made using a certain instrument. The results made in millimeters were

8.1; 9.1; 8.9; 8.9; 9.1

Given that true diameter of the ball bearing was 9.0mm,

a) calculate the unbiased estimated of the mean and variance of the measurement error of the instrument, [5]

b) assuming that the measurement errors are independent and normally distributed, find the 90% confidence interval of the mean measurement error.[3]

SOLUTION

a) Let X be the measures of the diameter of a ball bearing.

$X \sim N(\mu, \hat{\sigma}^2)$ with both μ and σ^2 unknown.

$$\hat{\sigma}^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)$$

$$= \frac{1}{5-1}\left(389.65 - \frac{(44.1)^2}{5}\right)$$

$\sigma = \sqrt{0.172}$

$= 0.4147288271$

$= \mathbf{0.415}$

The sample mean, (unbiased estimate of the mean), $\bar{\boldsymbol{x}} = \frac{\sum \boldsymbol{x}}{\boldsymbol{n}} = \frac{\mathbf{44.1}}{\mathbf{5}} = \mathbf{8.82}$

b) Since, n is small, a $t(n-1)$ distribution is used.

$n = 5$, so use a $t(5-1) = t(4)$ distribution.

The 90% confidence limits for the mean measurement error are;

$$\bar{x} \pm t\left(\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

$t = 2.132$

$$\therefore \text{ confidence limts are; } 8.82 \pm 2.132\left(\frac{\sqrt{0.172}}{\sqrt{5}}\right)$$

$= (8.424572907; 9.215427093)$

= (8.4mm; 9.2mm)

10.

a) Define the following terms using exams

i) parameter,

ii) statistic. [4]

b) A teacher recorded the test marks (x), of 15 students. It was found that $\sum x = 820$ and that the standard deviation of x was 6.3.

Calculate the value of

i) $\sum(x - 53)$

ii) $\sum(x - 53)^2$ [4]

SOLUTION

a)

i) A parameter is a measure from a population.

ii) A statistic is a measure from a sample.

b)

i) $\sum(x-53)=\sum x-\sum 53$

$=820-(15)53$

$=\mathbf{25}$

ii) $\sigma^2=\frac{\sum(x-53)^2}{n}-\left(\frac{\sum(x-53)}{n}\right)^2$

$6.3^2=\frac{\sum(x-53)^2}{15}-\left(\frac{25}{15}\right)^2$

$39.69+\left(\frac{25}{15}\right)^2=\frac{\sum(x-53)^2}{15}$

$\left(\frac{38221}{900}\right)^1 15=\sum(x-53)^2$

$\sum(x-53)^2=\frac{38221}{60}$

$\sum(x-53)^2=637.0166667$

$\boldsymbol{\sum(x-53)^2=637.02}$

11.

a) Customers arrive randomly at a supermarket at an average rate of 3.6 per minute.

Assuming that, the arrival form a Poisson distribution,

calculate the probability that,

i) no customer arrives in any particular time,

ii) one or more customers arrive in any 30 seconds. [5]

b) An exponential distribution has probability density function (pdf), f given by

$$f(t) = \frac{1}{5}e^{-\frac{1}{5}t}, \text{where } t \geq 0.$$

Find E(T). [4]

SOLUTION

a)

i) Let X be the number of customers that arrive at a supermarket

X~Po(3.6)

$P(X = 0) = e^{-3.6}\left(\frac{3.6^0}{0!}\right)$

$= \mathbf{0.0273}$

ii) In 30 minutes the value of λ changes

$\lambda = \frac{30(3.6)}{60} = 1.8$

$P(X \geq 2) = 1 - P(X < 2)$

$= 1 - e^{-1.8}(1 + 1.8)$

$= 0.537163113$

$\mathbf{= 0.537}$

b) $\mathrm{E(T)} = \int_0^\infty t.f(t)\, dt$ [*using integration by parts*]

$\frac{1}{5}\int_0^\infty t.e^{-\frac{1}{5}t}\, dt$

let; $u = t$ $\quad \frac{du}{dt} = 1$

$\frac{dv}{dt} = e^{-\frac{1}{5}t}$ $\quad v = -5e^{-\frac{1}{5}t}$

Now, $\frac{1}{5}\int_0^\infty t.e^{-\frac{1}{5}t}\, dt = \frac{1}{5}\left[-5t.e^{-\frac{1}{5}t}\right]_0^\infty - \frac{1}{5}\int_0^\infty 1.(-5)e^{-\frac{1}{5}t}\, dt$

$= \frac{1}{5}(-5)\left[\infty . e^{-\frac{1}{5}(\infty)} - \left(0 . e^{-\frac{1}{5}(0)}\right)\right] - \frac{1}{5}(-5)[(-5)e^{-\frac{1}{5}t}]_0^\infty$

$= (-5)\left[e^{-\frac{1}{5}(\infty)} - e^{-\frac{1}{5}(0)}\right]$

$= (-5)[0 - e^0]$

$= (-5)[-1]$

$\mathbf{= 5}$

12.

a) Masses of bags of oranges are normally distributed with mean 5 kg and standard deviation 0.2 kg. Five hundred bags of oranges are delivered to a supermarket. Find the

i) probability that a randomly delivered bag weighs more than 5.5 kg,

ii) number of bags from a single delivery that would be expected to weigh more than 5.5 kg. [4]

b) 20% of the bolts manufactured by a machine are known to be faulty. A random sample of 200 such bolts is taken.
Use a suitable approximation to find the probability that more than 50 bolts will be faulty. [5]

SOLUTION

a)

i) Let X represent masses of oranges

$X \sim N(5, 0.2^2)$

$P(X > 5) = P\left(z > \frac{5.5-5}{0.2}\right)$

$= P(z > 2.5)$

$= 1 - \emptyset(2.5)$

$= 1 - 0.9938$

$\mathbf{= 0.0062}$

ii) Number of bags from a single delivery that would be expected to weigh more than 5.5 kg is $0.0062 \times 500 = 3.1 \approx 4$ bags

b) Let X bolts manufactured by a machine and n be the number of such bolts.

$\boldsymbol{X \sim B(200, 0.2)}$

Since, $np > 5$ and $nq > 5$, hence we use a normal approximation to the binomial distribution.

So, $X \sim N(np, npq)$ such that $X \sim N(40, 32)$

Using continuity correction, $P(X > 50)$ will be $P(X > 50.5)$

$P(X > 50.5) = P\left(z > \frac{50.5-40}{\sqrt{32}}\right)$

$= P(z > 1.856)$

$= 1 - \emptyset(1.856)$

$= 1 - 0.9682$

$= \mathbf{0.0318}$

13.

a) Most discrete random variables assume whole number values. Write down two examples of discrete data which do not take up whole numbers only. [2]

b) A continuous random variable x has probability density function given by

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

i) Show that $k = \frac{3}{4}$.

ii) Find E(x).

iii) Find P($x > 2.5$). [8]

SOLUTION

a)

i) $(x-1)(3-x) = 3x - x^2 - 3 + x$

$= -x^2 + 4x - 3$

$= k\int_1^3 (-x^2 + 4x - 3)\,dx = 1$

$= k\left[-\frac{1}{3}x^3 + 2x^2 - 3x\right]_1^3 = 1$

$= k\left[\left(-\frac{1}{3}(3)^3 + 2(3)^2 - 3(3)\right) - \left(-\frac{1}{3}(1)^3 + 2(1)^2 - 3(1)\right)\right] = 1$

$= k\left[(0) - \left(-\frac{4}{3}\right)\right] = 1$

$k\left(\frac{4}{3}\right) = 1$

$\therefore \boldsymbol{k = \frac{3}{4}}$ **(shown).**

ii) E(X)

$\frac{3}{4}\int_1^3 x.(-x^2 + 4x - 3)\,dx$

$\frac{3}{4}\int_1^3 (-x^3 + 4x^2 - 3x)\,dx$

$= \frac{3}{4}\left[-\frac{1}{4}x^4 + \frac{4}{3}x^3 - \frac{3}{2}x^2\right]_1^3$

$= \frac{3}{4}\left[\left(-\frac{1}{4}(3)^4 + \frac{4}{3}(3)^3 - \frac{3}{2}(3)^2\right) - \left(-\frac{1}{4}(1)^4 + \frac{4}{3}(1)^3 - \frac{3}{2}(1)^2\right)\right]$

$= \frac{3}{4}\left[\frac{9}{4} - \left(-\frac{5}{12}\right)\right]$

$= \frac{3}{4}\left(\frac{8}{3}\right)$

$= \mathbf{2}$

iii) $P(x > 2.5)$

$$\frac{3}{4}\int_{2.5}^{\infty}(-x^2 + 4x - 3)\,dx$$

$$= \frac{3}{4}\left[-\frac{1}{3}x^3 + 2x^2 - 3x\right]_{2.5}^{\infty}$$

$$= \frac{3}{4}\left[\left(-\frac{1}{3}(\infty)^3 + 2(\infty)^2 - 3(\infty)\right) - \left(-\frac{1}{3}(2.5)^3 + 2(2.5)^2 - 3(2.5)\right)\right]$$

$$= \frac{3}{4}\left[(0) - \left(-\frac{5}{24}\right)\right]$$

$$= \frac{3}{4}\left(\frac{5}{24}\right)$$

$$= \mathbf{\frac{5}{32}}$$

14. Leaners from school X, Y and Z sat for mathematics examination.

The results were as follows:

	School X	School Y	School Z
Passed	76	60	40
Failed	15	25	20

Test at 5% level of significance whether there is an association between results and school. [13]

SOLUTION

	School X	School Y	School Z	**Total**
Passed	76	60	40	**176**
Failed	15	25	20	**60**
Total	**91**	**85**	**60**	**236**

H_0: There is no association between results and school.

H_1: There is association between results and the school.

Expected frequency $= \frac{\text{row total} \times \text{column total}}{\text{grand total}}$

67.86	63.39	44.75
23.14	21.61	15.25

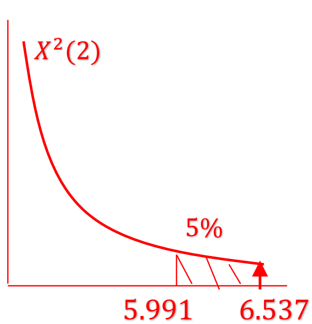
Degrees of freedom, $v = (3-1)(2-1) = 2$

So use $X^2(2)$

Test at 5% level of significance.

Therefore, $X^2_{5\%}(2) = 5.991$, so reject H_0 if $X^2_{cal} > 5.991$

O	E	$\frac{(O-E)^2}{E}$
76	67.86	0.9764161509
15	23.14	2.863422645
60	63.39	0.1812920019
25	21.61	0.5317954651
40	44.75	0.5041899441
20	15.25	1.479508197
$\sum O = 236$	$\sum E = 236$	$\sum \frac{(O-E)^2}{E} = 6.537$



Since, $X^2_{cal} > 5.991$ we reject H_0 and hence conclude that, there is an association between results and school at 5% level of significance.

15. During a handball practice, each member of the team of 40 players attempted to throw a ball between two posts. Each player had four attempts. The number of successes were summarized as follows

Number of success in 4 attempts	0	1	2	3	4
Number of players	14	14	9	2	1

Given that the mean number of successful attempts is 1.075, test at 5% level of significance whether the data follows a binomial distribution. [14]

SOLUTION

$np = 1.075$

$4p = 1.075$

$p = 0.26875$

Let X be the number of successful attempts.

H_0: Data follows a binomial distribution.

H_0: Data does not follow a binomial distribution.

If H_0 is true, then $X \sim B(4, \ 0.26875)$

Calculating expected frequencies

$P(X = 0) = {}^4_0C \times 0.26875^0 \times 0.73125^4 \times 40 = 21.3890625$

$P(X = 1) = {}^4_1C \times 0.26875^1 \times 0.73125^3 \times 40 = 16.81380835$

$P(X = 2) = {}^4_2C \times 0.26875^2 \times 0.73125^2 \times 40 = 9.269150757$

$P(X = 3) = {}^4_3C \times 0.26875^3 \times 0.73125^1 \times 40 = 2.271073975$

$P(X = 4) = {}^4_4C \times 0.26875^4 \times 0.73125^0 \times 40 = 0.2086670532$

Observed (O)	Expected (E)
14	21.3890625
14	16.81380835
9	9.269150757
2	2.271073975
1	0.2086670532

Degrees of freedom, $v = 3 - 2 = 1$

So use, $X^2(1)$

Test at 5% level of significance.

So, $X^2_{5\%}(1) = 3.841$, hence reject H_0 if $X^2_{cal} > 3.841$

Observed (O)	Expected (E)	$\frac{(\lvert O-E\rvert-0.5)^2}{E}$
14	21.3890625	2.218852843
14	16.81380835	0.318411449
12	11.74889179	0.005272592
$\sum O = 40$	$\sum E = 40$	$\sum \frac{(\lvert O-E\rvert-0.5)^2}{E} = 2.542536885$

Since, $X^2_{cal} < 3.841$, we fail to reject H_0 and hence conclude that data follows a binomial distribution at 5% level of significance.

FEEL FREE TO CONTACT ME FOR ANY ADJUSTMENTS, CLARIFICATIONS AND ASSISTANCE!

schingwingwi7@gmail.com / +263784089807

Sharewell B. C

"Concept before anything!", Author

Proverbs 11 vs. 2

[Turn Over

Complied and typed by Sharerwell B. Chingwingwi (Share)

+263784089807

schingwingwi7@gmail.com

NB: If you click any of the links below and it fails to direct you straight to the respective group, then first copy the whole link of any group and, place it in any WhatsApp/Telegram chat/group, then click again and join!

WhatsApp group 1 link: https://chat.whatsapp.com/J3aylX2aHrMAdZF71Ehq5U

WhatsApp group 2 link: https://chat.whatsapp.com/BmwI8owy2CVDBLl1RnJy0w

WhatsApp group 3 link: https://chat.whatsapp.com/ILPsSpcbjLjAQ89xvqnm7c

Telegram group link: t.me/shareALstatistics