ROYAL CIVIL SERVICE COMMISSION
BHUTAN CIVIL SERVICE EXAMINATION (BCSE) 2023
EXAMINATION CATEGORY: TECHNICAL

PAPER III: SUBJECT SPECIALISATION PAPER FOR STATISTICS

Date : October 7, 2023
Total Marks : 100
Writing Time : 150 minutes (2.5 hours)
Reading Time : 15 minutes (prior to writing time)

GENERAL INSTRUCTIONS:

1. Write your Registration Number clearly and correctly on the Answer Booklet.
2. The first 15 minutes is to check the number of pages of Question Paper, printing errors, clarify doubts and to read the instructions. You are NOT permitted to write during this time.
3. This paper consists of **TWO SECTIONS**, namely SECTION A & SECTION B:
   - **SECTION A** has two parts: Part I - 30 Multiple Choice Questions
     Part II - 4 Short Answer Questions
     All questions under SECTION A are **COMPULSORY.**
   - **SECTION B** consists of two Case Studies. Choose only **ONE** case study and answer the questions of your choice.
4. All answers should be written on the Answer Booklet provided to you. Candidates are not allowed to write anything on the question paper. If required, ask for additional Answer Booklet.
5. All answers should be written with correct numbering of Section, Part and Question Number in the Answer Booklet provided to you. Note that any answer written without indicating the Section, Part and Question Number will NOT be evaluated and no marks will be awarded.
6. Begin each Section and Part on a fresh page of the Answer Booklet.
7. You are not permitted to tear off any sheet(s) of the Answer Booklet as well as the Question Paper.
8. Use of any other paper including paper for rough work is not permitted.
9. **You must hand over the Answer Booklet to the Invigilator before leaving the examination hall.**
10. This paper has **12 printed pages**, including this instruction page.

GOOD LUCK

SECTION A

PART I: Multiple Choice Questions (30 marks)

Choose the correct answer and write down the letter of your chosen answer in the Answer Booklet against the question number e.g. 31 (d). Each question carries ONE mark. Any double writing, smudgy answers or writing more than one choice shall not be evaluated.

1. What is the mean of range, median and mode of the given data set?
   6, 5, 3, 5, 9
   a) 6.0
   b) 5.6
   c) 5.33
   d) 5.0

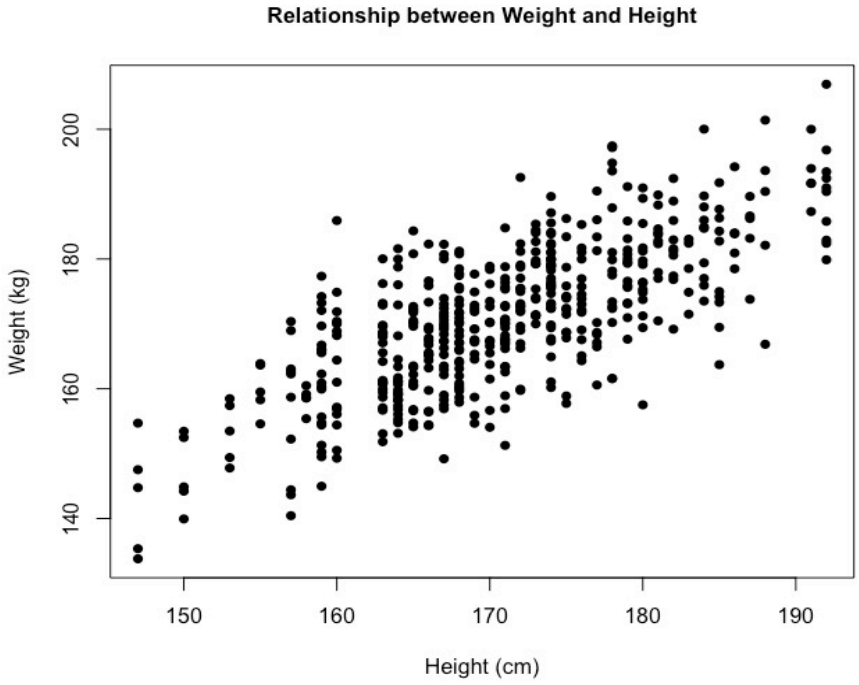
2. What is the value of variance if the sample mean is 50 and the coefficient of variation (CV) is 20%?
   a) 10
   b) 20
   c) 50
   d) 100

3. If two events A and B are independent, then what is P(A|B)?
   a) P(A) * P(B)
   b) P(A)
   c) P(B)
   d) P(A ∩ B)/P(B)

4. The mean weight of the 100 boxers was found to be 75 kg with a standard deviation of 5 kg. If all boxers were asked to increase their weight by 5%, what will be the new mean weight and standard deviation?
   a) 78.75 kg and 8.75 kg
   b) 75 kg and 5.25 kg
   c) 78.75 kg and 5.25 kg
   d) 78.75 kg and 5 kg

5. If $Y \sim Bin(10, 0.2)$, then what is the variance of Y, $Var(Y)$?
   a) 2
   b) 1.6
   c) 0.8
   d) 2.6

The scatter plot provided below describes the relationship between weight and height. Use the plot to answer the **Question 6**.



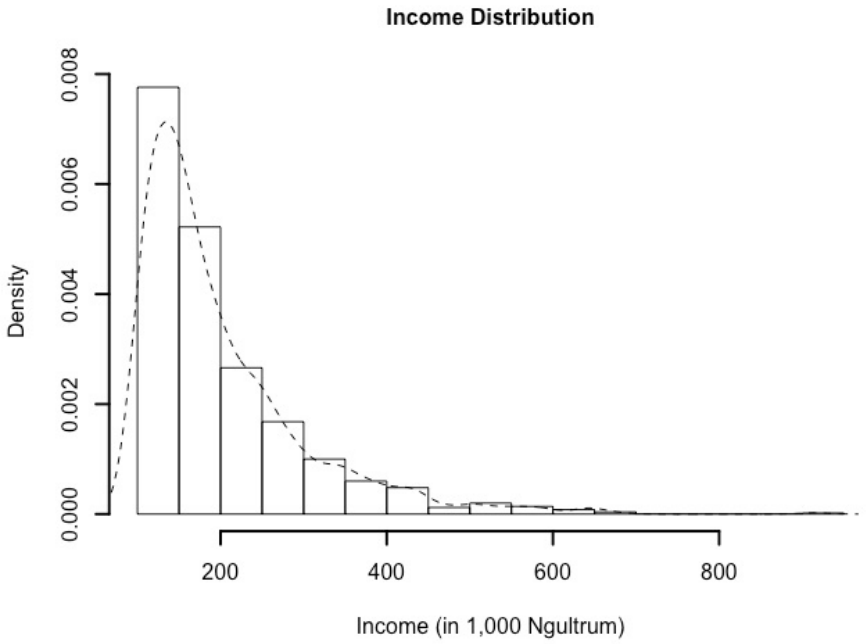
6. The correlation coefficient between weight and height is 0.8. How will you interpret the relationship?
   a) There is a perfect positive relationship
   b) There is no relationship
   c) There is a fairly strong positive relationship
   d) Not able to conclude from the scatter plot

7. The two-sided p-value is always
   a) Square of the one-sided p-value
   b) Half of the one-sided p-value
   c) Twice as big as one-sided p-value
   d) Same as the one-sided p-value

8. If $X_1, X_2, \ldots\ldots, X_n$ are independent and identically distributed (iid) with mean μ and variance $\sigma 2$, the bias of sample mean $\bar{X}$ is
   a) σ
   b) 1
   c) 0
   d) μ

9. The score of the statistical test at Sherubtse College is normally distributed with a population mean of 75 and a variance of 9. Tenzin scored 80. What proportion of scores were below Tenzin's score?
   a) 95.15%
   b) 97.15%
   c) 93.15%
   d) 90.15%

The histogram below provides income distribution. Use the histogram to answer the **Questions 10 & 11.**



10. Which relation satisfy the income distribution?
    a) Mean > Median > Mode
    b) Mode > Median > Mean
    c) Mode > Mean > Median
    d) Median > Mean > Mode

11. What is the robust measure of dispersion for the income distribution?
    a) Standard deviation
    b) Range
    c) Lower quartile
    d) Interquartile range

The summary output of the regression model, which examines the relationship between sleep duration (the response variable) in hours and alcohol consumption (the predictor variable) with two levels: Drinkers and Non-drinkers, is presented below. Use the summary output to answer the **Questions 12, 13 and 14**.

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.95493	0.06728	103.37	<2e-16 ***
Non-Drinkers	1.01267	0.09515	10.64	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

12. What is the estimated mean of 'sleep duration', in hours, for people who do not consume alcohol?
    a) 1.01267
    b) 7.9676
    c) 6.95493
    d) 5.94226

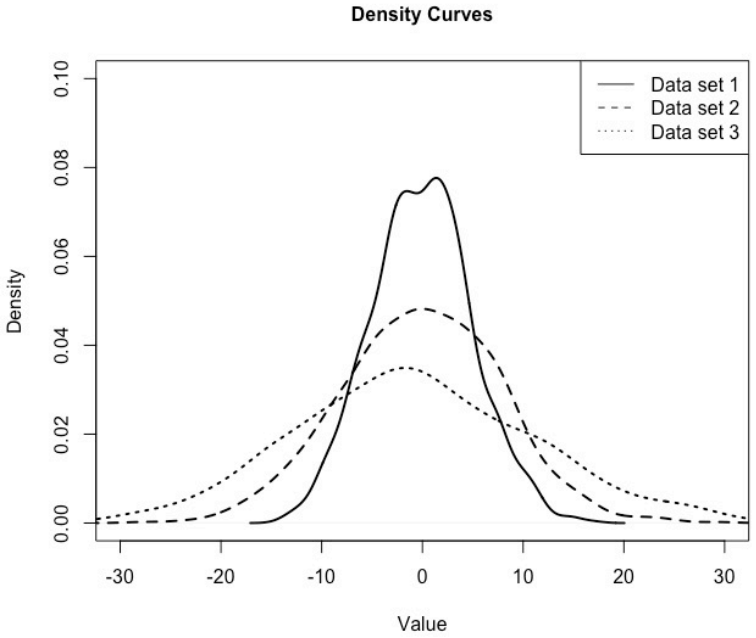
13. What is the estimated mean of '*sleep duration'*, in hours, for people who consume alcohol?
    a) 1.01267
    b) 7.9676
    c) 6.95493
    d) 5.94226

14. What is the estimated difference in mean of 'sleep duration', in hours, between two groups?
    a) 1.01267
    b) 7.9676
    c) 6.95493
    d) 5.94226

15. What is the mode of the data set, if it's median and mean are 10 and 5 respectively?
    a) 20
    b) 12
    c) 30
    d) 5

16. What is the purpose of using differencing in time series analysis?
    a) To remove trends and seasonality from the data
    b) To add noise and randomness to the data
    c) To convert non-stationary data into stationary data
    d) To smooth out the data and make it more predictable.

The plot given below shows densities of three different data sets. Use the plot to answer the **Question 17**.



17. Which statement is TRUE based on the above plot?
    a) Data set 1, Data set 2 and Data set 3 are dispersed equally
    b) Mean of Data set 1 is more than Data set 2 and Data set 3
    c) Data set 2 is more dispersed than Data set 3 and Data set 1
    d) Data set 1, Data set 2 and Data set 3 share same location

18. What is the limiting distribution of following probability mass function?

$$lim_{n\to\infty} f(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

    a) Poisson distribution
    b) Geometric distribution
    c) Normal distribution
    d) Negative binomial distribution

19. In multiple linear regression, the high degree of multicollinearity among predictors will
    a) Reflect inherent effect of the particular predictor on the response variable
    b) Inflate variability of the estimated coefficients
    c) Not contribute in reducing errors sum of squares
    d) Deflate p-value of the predictors

20. If $f(y)$ is the probability density function, then
    a) $\sum_{all\ y} f(y) = 1$
    b) $\int_{-\infty}^{\infty} f(y)dy = 1$
    c) $\sum_{all\ y} f(y) = 0$
    d) $\int_{-\infty}^{\infty} f(y)dy > 1$

21. In a selection of statistical investigators, there were two male and five female applicants. The National Statistics Bureau wants to select two applicants. What is the probability that the two selected applicants are female?
    a) 0.476
    b) 0.576
    c) 0.376
    d) 0.676

22. Suppose Y_1 and Y_2 are two random variables with means 2 and -7 respectively. What is the expected value of $E(3Y_1 - 2Y_2)$?
    a) 8
    b) 20
    c) -8
    d) 25

23. Multi-stage sampling is a sampling method that divides population into groups. One of the advantages of multi-stage sampling is
    a) Survey procedures is carried out only in the selected clusters
    b) Covers whole geographic areas
    c) Yields more precise results as compared to Simple Random Sampling of same size
    d) Data analysis is much simpler

Use the following statistics to answer the **Questions 24 and 25**.
Assume $\bar{y} = 10$ sample variance $s^2 = 25$ and n = 81.

24. What is the standard error of sample mean, SE($\bar{y}$)?
    a) 0.31
    b) 0.06
    c) 2.78
    d) 0.56

25. Assuming normality, what is the margin of error of sample mean, ME($\bar{y}$) at 95% confidence level?
    a) 5.56
    b) 0.12
    c) 0.62
    d) 1.10

26. A weekly mean and standard deviation of study time of Class X students of Paro Central School are 10 hours and 30 minutes respectively. As the Board Examination approaches, they agreed to study two more hours weekly. What is the new weekly mean and standard deviation of study time?
    a) 10 hours and 0.5 hrs
    b) 12 hours and 2.5 hrs
    c) 12 hours and 0.5 hrs
    d) 10 hours and 2.5 hrs

27. Why the method of least-squares is used on time-series data?
    a) De-seasonalising the data
    b) Obtaining trend equation/obtaining best fitting line
    c) Exponentially smoothing the series
    d) Eliminating irregular movements

28. In determining appropriate sample size, what does design effect greater than one indicate?
    a) Design is more efficient
    b) Sample size may be reduced
    c) Same precision obtained as design under the simple random sampling
    d) Design is less efficient

29. If X and Y are independent random variables, what is $Cov(X, Y)$?
    a) 1
    b) -1
    c) 0.5
    d) 0

Suppose an economist developed a first order autoregressive (AR(1)) to forecast annual gross domestic product (GDP) as given below. Use the model to answer the **Question 30**.

$$y_t = \rho y_{t-1} + \epsilon_t, \quad where\ \epsilon_t \sim N(0, \sigma^2)$$

30. The GDP in 2021 is approximately Nu. 187,700 million and if $\rho = 1$. What will be the forecast in 2023?
    a) Nu. 187,700 * 2 million
    b) Nu. 187,700 million ngultrums
    c) Nu. 187,700/2 million ngultrums
    d) Need value of σ^2, so can't forecast

PART II – Short Answer Questions [20 marks]

This part has 4 Short Answer Questions. Answer ALL the questions. Each question carries 5 marks.

1. Assume $X_1, X_2, \ldots\ldots, X_n$ are independent and identically distributed (iid) random variables, and E $(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Write/derive the distribution of $\bar{X}$. (5 marks)

2. A random sample of 50 college students were taken and measured their height. The sample yielded a mean of 155 cm and a standard deviation of 3.65 cm. The approximate 95% confidence interval is 154 cm to 156 cm. Indicate **TRUE** or **FALS**E for the following statements and then provide a short explanation to support your answer.

   a) We are 95% confident that the average height of college students in this sample is between 154 cm and 156 cm. (2.5 marks)

   b) 95% of the college students measure 154 cm to 156 cm in height. (2.5 marks)

3. Find probability that one tail comes up on two tosses of a fair coin? (5 marks)

4. What is an outlier? What are some ways to detect outliers and what should be done? (5 marks)

SECTION B: Case Study [50 marks]

Choose either CASE I OR CASE II from this section. Each case study carries 50 marks. Mark for each sub-question is indicated in the brackets.

CASE I

It is believed that birth weights of babies born in India are normally distributed with a mean 3000 g and a standard deviation of 500 g. Pema Tshewang, the novice researcher thinks that babies in Bhutan have a mean birth weight greater than 3000 g and would like to test this hypothesis. Pema Tshewang took a random sample (iid) of 44 babies from Bhutan, measured their birth weights and observed that the sample mean of these 44 weights $\bar{X}$ is 3275.955 g.

Your role as a statistician, is to help Pema Tshewang, carry out the test in the following:

1. Set up the hypothesis. State whether it is one-sided or two-sided test. Decide the level of significance (α) between 0.01 or 0.05. (10 marks)

2. Calculate test statistic and compare to its sampling distribution under the null hypothesis by calculating the p-value. (20 marks)

3. What do you conclude at the level of significance decided? (5 marks)

4. What would be the conclusion if you have chosen other level of significance? (5 marks)

5. If Pema Tshewang is interested in testing the research hypothesis that the mean birth weight of Bhutanese babies is different from 3000g, what would be the hypothesis and conclusion? (10 marks)

CASE II

Consider a scenario where you are a sales manager for a retail store, and you want to analyze the relationship between the advertising expenditure and the sales of a product. You have collected data for the past 12 months, including the monthly advertising expenditure (in ngultrum) and the corresponding monthly sales (in units) of the product.

A simple linear regression model is used to describe the relationship at $\alpha = 0.05$ and a summary output of the model is provided below (R output):

```
Call:
lm(formula = Sales ~ Advertising_Expenditure, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-13.024   -4.777    2.133    7.212    8.549

Coefficients:
                            Estimate   Std. Error   t value   Pr(>|t|)
(Intercept)              -10.122378       11.240    -0.901      0.389
Advertising_Expenditure    0.020315       0.0013    15.486          ?
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.057 on 10 degrees of freedom
Multiple R-squared:  0.96,    Adjusted R-squared:  0.956
F-statistic: ? on ? and ? DF,  p-value: ?
```

1. What are the estimated coefficients of the regression model? Write the regression equation? (5 marks)

2. Interpret estimated regression coefficients. Does the intercept have meaning interpretation? Explain briefly. (To obtain full mark, mention units while interpretating). (15 marks).

3. How well does the model fit? Explain briefly. (5 marks)

4. What is the predicted value of sales for advertising expenditure Nu. 8,300? (5 marks)

5. Fill in the details for the ANOVA table (Hint: use summary output) (8 marks)

Source	Degrees of freedom	Sum of squares	Mean square	F statistic
Model (Regression)	i	iv	vi	viii
Residual (Error)	ii	v	vii	
Total	iii	20,491.7		

6. State your hypothesis for the regression model? What do you conclude about the relationship between sales and advertising expenditure? (7 marks)

7. What factors do you think could affect 'sales' other than 'Advertising_Expenditure'? List at least five factors with a brief explanation. (5 marks)

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

Critical Values of the F-Distribution: $\alpha = 0.05$

Denom.	Numerator Degrees of Freedom									
d.f.	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.883	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348

TASHI DELEK