NATIONAL TESTING AGENCY (NTA)

Procedure to be adopted for compilation of NORMALIZED SCORES for multi-session Papers in COMMON UNIVERSITY ENTRANCE TEST (UNDERGRADUATE): CUET (UG)- 2022

(Normalization procedure based on Percentile and Equi percentile method)

Need for Normalisation

For several subjects, the CUET (UG)-2022 has been conducted in different shifts. Since the question paper for any subject in different shifts are different and it is quite possible that despite all best possible efforts of maintaining equivalence among various question papers, the difficulty level of these question papers administered in different sessions may not be the same or similar. Some of the candidates might have end up attempting a relatively tougher set of questions when compared to other sets. The candidates who attempted the comparatively tougher examination are likely to get lower marks as compared to those who attempted the easier one, the scores of the students across shifts are not directly comparable. There is a need for normalizing the marks across shiftsto make them amenable to such comparisons.

A Committee headed by Senior Professor from Indian Statistical Institution Delhi and comprising of Senior Professors from Indian Institute of Technology Delhi and University of Delhi had detail deliberations to arrive at methodology.

As already intimated in the Information Bulletin of CUET (UG) - 2022 the raw score of each candidate in each subject will be normalised using the ***Equi-percentile method***. This is done separately for each subject for which examination is held in multiple shifts.

For each subject for which the examination is held in multiple shifts, the raw score for each candidate appearing for the subject will be converted into **NTA *score*** (percentile score & normalized score) in the following three steps.

Step1: Convert raw scores into percentile scores.

The percentiles are calculated separately for each shift.

1. Record the number of candidates who have appeared in a shift. Denote this number by *N*.
2. Sort all the candidates in one shift in decreasing order of their marks.
3. Note the *raw marks* for each candidate. Suppose this is denoted by *T*. Count the number of candidates in that shift whose raw scores are less than *or equal to T*. Denote this number by *m*.
4. The percentile score for this candidate is then calculated as:

$$\mathbf{P} = \frac{m}{N} * 100$$

Illustration:

Suppose that the examination in a certain subject is held in two different shifts, *S*1 and *S*2, say. Consider six candidates A*, B, C, D, E and* F *out* of which *A, B, C* are from shift S1 and the other three are from shift *S*2.

Let the raw marks of the six candidates be x_A, x_B, x_C(shift $S1$ marks) and y_D, y_E, y_F(shift $S2$ marks).

For candidates *A, B, C*, the percentiles are calculated using the total marks obtained by candidates appearing in shift $S1$ (in the same subject) as explained above.

Similarly, for candidates *D, E, F*, the percentiles are calculated using the total marks obtained by candidates appearing in shift $S2$ (in the same subject).

Let the respective percentiles be denoted by $P_A, P_B, P_C, P_D, P_E, P_F$.

The above data is summarised in the following table. The terms in red colour indicate that these are the output of this step.

Shift S1			Shift S2		
Candidate	Raw score	Percentile	Candidate	Raw score	Percentile
A	x_A	P_A	D	y_D	P_D
B	x_B	P_B	E	y_E	P_E
C	x_C	P_C	F	y_F	P_F

Since the calculation for percentiles in any shift depends only on the data from that shift alone, there is a separate table for each shift.

Step2: Pull-back of the percentiles to the marks scale for each session to get Normalised Score.

- The data across all sessions tabulated at the end of step 1 is collated into a single table.
- The columns for the shift-wise raw score should are kept separate.
- All the records are then sorted in decreasing order of the percentiles.

Illustration (Continued):

In the illustrative example given above, suppose that the percentiles of the six candidates satisfy

$$P_E > P_A > P_C = P_F > P_B > P_D.$$

Then the collated table would look as given below:

Candidate	Percentile	Rawscore $S1$	RawScore $S2$
E	P_E	–	y_E
A	P_A	x_A	-
C&F	$P_C = P_F$	x_C	y_F
B	P_B	x_B	-
D	P_D	–	y_D

Candidate *C* from shift $S1$ and candidate *F* from shift $S2$ have the same percentile. The relevant entries under" Raw Score $S1$" and "Raw Score $S2$" are the actual raw scores x_C and y_E respectively. This has the obvious interpretation that marks x_C of shift $S1$ are equivalent to marks y_E of shift $S2$.

Candidates A and B, appearing in shift $S1$, have a blank entry in column "Raw Score $S2$", as there is no corresponding candidate having exactly the same percentile from shift $S2$. Similarly, Candidates D and E, appearing in shift $S2$, have a blank entry in column "Raw Score $S1$", as there is no corresponding candidate having exactly the same percentile from shift $S1$.

- In the remaining part of this *Step 2*, the blank entries in the two "Raw Score" columns are filled up using linear interpolation.
- Consider a record (row) whose entry in the column "Raw Score $S1$" is blank. The blank will be replaced by the interpolated score X which is calculated as:

$$X = x_1 + \frac{x_2 - x_1}{p_2 - p_1}(P - p_1)$$

where

- P is the corresponding entry in "Percentile" column
- x_1 is the first *non-blank* entry BELOW X. i.e. $x_1 < X$ and there is no other non-blank entry in the column between x_1 and X.
- x_2 is the first *non-blank* entry ABOVE X. i.e. $x_2 > X$ and there is no other non-blank entry in the column between X and x_2.
- p_1 is the entry in the "Percentile" column corresponding to x_1 from the column "Raw Score $S1$".
- p_2 is the entry in the "Percentile" column corresponding to x_2 from the column "Raw Score $S1$".

Note that there may be several blank entries between x_1 and x_2.

- All the blank entries in column "Raw Score S1" can now be replaced by the interpolated values.
- The blank entries in column "Raw Score $S2$" are also replaced using a similar procedure.

Illustration (Continued):

The table in the illustrative example would look like the following, where the entries in red indicate the addition output at the end of this step.

Candidate	Percentile	Raw Score $S1$	Raw Score $S2$
E	P_E	X_E	y_E
A	P_A	x_A	Y_A
C&F	P_C=P_F	x_C	y_F
B	P_B	x_B	Y_B
D	P_D	X_D	y_D

Step3: Calculation of the Normalised *Score*

Now for each subject *there is a score assigned to each percentile value and each session.* The ***Normalised score, Z,*** corresponding to a percentile value *P,* is calculated as:

$$Z = \text{Average of } (u_A, u_B, u_C, \ldots\ldots\ldots u_t) = \frac{u_A, u_B, u_C, \ldots\ldots u_t}{t}$$

Where $u_A, u_B, u_C, \ldots\ldots\ldots u_t$ denote the raw scores corresponding to the percentile *P* in each of different sessions

Illustration (Continued):

The **final** table in the *illustrative example* would be as follows:

Candidate	Percentile	Raw Score *S*1	Raw Score *S*2	Normalised Score
E	P_E	X_E	y_E	$(X_E+y_E)/2$
A	P_A	x_A	Y_A	$(x_A+Y_A)/2$
C&F	$P_C=P_F$	x_C	y_F	$(x_C+y_F)/2$
B	P_B	x_B	Y_B	$(x_B+Y_B)/2$
D	P_D	X_D	y_D	$(X_D+y_D)/2$

Step- by- Step Procedure for Converting from Raw Score to Normalized Score: -

Example: Suppose a test was held in 4 phases of examinees as per details given below: -

(Allocation of Days and shifts were done randomly)

Session	Day	Shift	No of Candidates			Marks	
			Absent	Appeared	Total	Highest	Lowest
Session-1	Day-1	Shift-1	3974	28012	31986	200	-40
Session-2	Day-1	Shift-2	6189	32541	38730	194	-36
Session-3	Day-2	Shift-1	6036	41326	47362	188	-36
Session-4	Day-2	Shift-2	9074	40603	49677	200	-40
Total (Session-1 to Session-4)			25273	142482	167755	200	-40

i. Highest Raw Score and Percentile Score: -

Session	Total Candidates Appeared	Highest Raw Score	Candidates who scored EQUAL OR LESS THAN Highest Raw Score	Percentile Score	Remarks
Session-1	28012	200	28012	100.0000000 [(28012/28012) *100]	i.e., All the highest raw scores would be normalized to 100 Percentile Score for their respective session.
Session-2	32541	194	32541	100.0000000 [(32541/32541) *100]	
Session-3	41326	188	41326	100.0000000 [(41326/41326) *100]	
Session-4	40603	200	40603	100.0000000 [(40603/40603) *100]	

ii. Lowest Score and Percentile Score: -

Session	Total Candidates Appeared	Highest Raw Score	Candidates who scored EQUAL OR LESS THAN Highest Raw Score	Percentile Score	Remarks
Session-1	28012	-40	1	0.0035699 [(1/28012) *100]	i.e., Percentile Score of all the lowest raw scores are different i.e., Percentile Score depend on the total number of candidates who have taken the examination for their respective session.
Session-2	32541	-36	1	0.0030730 [(1/32541) *100]	
Session-3	41326	-36	1	0.0024198 [(1/41326) *100]	
Session-4	40603	-40	1	0.0024629 [(1/40603) *100]	

Step1: Convert Raw Score into Percentile Score

The percentile score for this candidate is then calculated as:

$$Percentile\ Score = \frac{No.of\ Candidates\ appeared\ from\ the\ session\ with\ raw\ score\ EQUAL\ TO\ OR\ LESS\ than\ T\ Score}{\text{Total No. of Candidates appeared in the session}}$$

Shift S1			Shift S2		
CANDIDATE	Raw Score S1	Percentile	CANDIDATE	Raw Score S2	Percentile
A20020720	200	100	B20123935	194	100
A20411664	192	99.9691438	B20012622	192	99.99904053
A20018569	190	99.8312554	B20621750	184	99.99712158
A20339879	184	99.7329013	B20298730	180	99.99232422
A20027230	182	99.6075482	B20197060	176	99.97793215
A20074407	180	99.4320538	B20035799	172	99.97025637
A20751862	88	88.6545749	B20175737	114	98.12423242
A20685124	54	77.4200391	B20272584	112	97.99086583
A20397755	52	76.4422845	B20030697	102	97.15708474
A20471411	50	75.4172814	B20656224	100	96.9671093
A20922992	48	74.1377149	B20410215	42	69.2904047
A21004667	40	69.2904047	B20236444	26	67.7672549
A21141123	22	52.1353428	B20128586	14	46.37511514
A21141561	10	35.4807294	B20001667	4	27.08877034
A25465232	8	30.6758464	B20042147	2	23.91003991
A26545946	6	27.5844446	B20051256	0	12.225608
A25656543	2	24.886459	B20481968	-2	11.62116211
A26625216	1	22.3514324	B20091701	-12	10.49825625
A26174451	0	19.9976858	B20549576	-14	10.39474209
A25463225	-1	15.23523	B20098269	-30	10.2430506
A26598636	-6	11.1043613	B20914176	-32	0.03454099
A26596462	-40	0.053034	B20071356	-36	0.053034

Shift S3			Shift S4		
CANDIDATE	Raw Score S1	Percentile	CANDIDATE	Raw Score S2	Percentile
C20150694	188	100	D20479616	200	100
C20087997	186	99.9691438	D20040337	194	99.99904053
C20121991	184	99.8312554	D20568599	184	99.99712158
C20058572	180	99.7329013	D20007708	180	99.99232422
C20076289	174	99.6075482	D20563271	176	99.97793215
C20060310	172	99.4320538	D20265618	172	99.97025637
C20008597	56	78.3389742	D20074767	114	98.12423242
C20241896	54	77.4200391	D20840372	112	97.99086583
C20388248	52	76.4422845	D20563852	102	97.15708474
C20672438	50	75.4172814	D20467385	98	88.6545749
C20430859	36	69.0018201	D20398094	44	68.9660903
C20518247	20	52.1353428	D20428934	40	49.50970986
C20045510	14	38.0919321	D20032939	32	46.37511514
C20361875	10	35.4807294	D20690279	30	27.08877034
C20860609	9	30.6758464	D20840699	28	23.91003991
C20861476	5	27.5844446	D20084751	26	20.88002763
C20512680	3	24.886459	D21237483	24	12.1887008
C20069270	2	22.3514324	D21077463	1	11.53492478

C20355550	0	12.1887008	D20348188	-2	11.0243302
C20549576	-3	11.23523	D20777569	-6	0.03645993
C26598636	-22	0.0539983	D26174451	-38	0.03454099
C26596462	-36	0.053034	D25463225	-40	0.053034

Step2: Pull Back of the percentiles to the marks scale for each session to get Normalized marks.

Candidate	Percentile	Raw Score S1	Raw Score S2	Raw Score S3	Raw Score S4
A20020720 = B20123935 = C20150694 = D20479616	100	200	194	188	200
D20040337	99.9750521	---	---	---	194
C20121991	99.9702544	---	---	184	---
A20411664	99.9691438	192	---	---	---
B20621750	99.9687198	---	184		---
A20751862 = D20467385	88.6545749	88	---	---	98
A21004667 = B20410215	69.2904047	40	42		---
C20430859	69.0018201	---	---	36	---
D20398094	68.9660903	---	---	---	44
B20236444	67.7672549	---	26		---
A21141123 = C20518247	52.1353428	22	---	20	---
D21237483	12.4236038	---	---	---	24
B20051256	12.225608	---	0	---	---
C20355550	12.1887008	---	---	0	---
A26598636	11.1043613	-6	---	---	---
D20348188	11.0243302	---	---	---	-2
B20098269	10.2430506	---	-30	---	---
A26596462 = B20071356 = C26596462 = D25463225	0.053034	-40	-36	-38	-40

In the remaining part of this Step2, the blank entries in the Raw Score Columns are filled up using linear interpolation.

Using interpolation Formulae is:

$$X = x_1 + \frac{x_2 - x_1}{p_2 - p_1}\,(P - p_1)$$

Candidate	Percentile	Raw Score S1	Raw Score S2	Raw Score S3	Raw Score S4
A20020720 = B20123935 = C20150694 = D20479616	100	200	194	188	200
D20040337	99.9750521	194	193.5312	186.0244	184.6453
C20121991	99.9702544	184	192.2868	184.4902	184
A20411664	99.9691438	192	192	184.1366	183.9947
B20621750	99.9687198	184	191.9961	184	183.9927
A20751862 = D20467385	88.6545749	88	98	88	131.6305
A21004667 = B20410215	69.2904047	40	42	40	42
C20430859	69.0018201	36	39.69721	38.96857	36
D20398094	68.9660903	44	39.65971	38.59318	35.9661
B20236444	67.7672549	26	38.40183	26	34.82886
A21141123 = C20518247	52.1353428	22	20	22	18.68243
D21237483	12.4236038	24	-5.09973	0.092685	0.117609
B20051256	12.225608	0	-5.23485	0	0.018478
C20355550	12.1887008	0	-5.26003	-0.55848	0
A26598636	11.1043613	-6	-6	-16.9667	-3.39535
D20348188	11.0243302	-2	-6.24622	-18.1777	-3.64595
B20098269	10.2430506	-30	-8.64987	-30	-6.09235
A26596462 = B20071356 = C26596462 = D25463225	0.053034	-40	-36	-38	-40

Step3: Calculation of Normalized Score:

$$Z = \text{Average of } (u_A, u_B, u_C, \ldots\ldots\ldots u_t) = \frac{u_A, u_B, u_C, \ldots.. u_t}{\text{t}}$$

Candidate	Percentile	Raw Score S1	Raw Score S2	Raw Score S3	Raw Score S4	Normalized Score
A20020720 = B20123935 = C20150694 = D20479616	100	200	194	188	200	195.5
D20040337	99.9750521	194	193.5312	186.0244	184.6453	189.550225
C20121991	99.9702544	184	192.2868	184.4902	184	186.19425
A20411664	99.9691438	192	192	184.1366	183.9947	188.032825
B20621750	99.9687198	184	191.9961	184	183.9927	185.9972
A20751862 = D20467385	88.6545749	88	98	88	131.6305	101.407625
A21004667 = B20410215	69.2904047	40	42	40	42	41
C20430859	69.0018201	36	39.69721	38.96857	36	37.666445
D20398094	68.9660903	44	39.65971	38.59318	35.9661	39.5547475
B20236444	67.7672549	26	38.40183	26	34.82886	31.3076725
A21141123 = C20518247	52.1353428	22	20	22	18.68243	20.6706075
D21237483	12.4236038	24	-5.09973	0.092685	0.117609	4.777641
B20051256	12.225608	0	-5.23485	0	0.018478	-1.304093
C20355550	12.1887008	0	-5.26003	-0.55848	0	-1.4546275
A26598636	11.1043613	-6	-6	-16.9667	-3.39535	-8.0905125
D20348188	11.0243302	-2	-6.24622	-18.1777	-3.64595	-7.5174675

B20098269	10.2430506	-30	-8.64987	-30	-6.09235	-18.685555
A26596462 = B20071356 = C26596462 = D25463225	0.053034	-40	-36	-38	-40	-38.5

The above method is based on the work: “Normalization of marks in multi-session examinations”, Abhay G. Bhatt et al, CURRENT SCIENCE, Vol. 118, No. 1, 10 January, 2020