Program : BScTE, 4th Semester | Date : 15 September 2025, Monday

Semester : Summer | Time : 10.00 am to 1.00 pm, Morning

5

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

ORGANISATION OF ISLAMIC COOPERATION (OIC)

DEPARTMENT OF TECHNICAL AND VOCATIONAL EDUCATION (TVE)

Examination	: Semester Final	Academic Year	: 2023-2024
Course No	: TVE 4237	Full Marks	: 150
Course Title	: Educational Measurement and Statistics	Duration	: 3.00 hours

There are seven (7) questions. Answer Six (6) questions.

Figures in the right margin indicate marks of the questions. The symbols have their usual meanings.

1. a) Define parallel, tau-equivalent and congeneric tests in the context of Classical Test Theory. How do they differ from each other? 5 CO1, CO2,

   b) A national exam uses two test forms claimed to be parallel. However, one form shows consistently higher scores than the other. Identify what kind of test form it is. What risks does this pose for fairness in high-stakes decisions (e.g., university admission)? 10 CO3

   c) Three items measuring math ability are defined as: Item 1: $\lambda = 0.5, \delta = 10, e = +2$; Item 2: $\lambda = 1.0, \delta = 5, e = -3$; Item 3: $\lambda = 1.5, \delta = 0, e = +1$. The student's true ability $\tau = 80$. 10
      (i) Compute each observed score.
      (ii) Discuss why these items are congeneric but not tau-equivalent.

2. a) A polytomous essay question shows average item difficulty (0.65) but a negative item-total correlation. How could this happen, and what revisions might make the item better aligned with the test's construct? 5 CO1, CO2, CO3

   b) In what ways does the concept of item-total correlation for polytomous items provide richer insight into test quality compared to *p-values* for dichotomous items? 5

   c) Why is it not always desirable for all test items to fall within the "ideal" *p-value* range (0.30–0.70)? What might be the risk of designing such a test only with items of moderate difficulty? 5

   d) A teacher administers a 20-item test to 40 students. For *Item 12*, 30 students answered correctly. Among the top 10 scorers, 9 answered correctly, while among the bottom 10 scorers, 4 answered correctly. The mean total score of students who answered *Item 12* correctly is 17, and for those who answered incorrectly it is 12. The standard deviation of the total test scores is 3. Given that $p=0.75$ and $q=0.25$, calculate the item difficulty, discrimination index, and point-biserial correlation for *Item 12*. 10

3. a) Classical Test Theory suggests longer tests often have higher reliability. Discuss the possible situations where increasing the number of items might not improve reliability. 5 CO1, CO2,

   b) A 30-item math test is split into two halves. The correlation between the two halves is 0.70. The variance of the difference scores is 20, and the variance of the total test scores is 110. 10 CO3
      (i) Compute reliability using the Spearman-Brown formula.
      (ii) Compute reliability using Rulon's formula.
      (iii) Analyze why the two estimates may differ and which one is more appropriate in this context.

   c) Two test raters grade 25 essays. Their scores have a correlation of 0.58 (inter-rater reliability). Meanwhile, the test-retest correlation of total essay scores over two weeks is 0.80. 10
      (i) Compare 'these two reliability estimates' mathematically.
      (ii) Explain why they differ, considering the sources of error variance involved.
      (iii) Which reliability measure would you prioritize if the test is used for long-term academic tracking?

4. a) Define construct validity with an example. Classify the construct validity. 5 CO1,

   b) Can a classroom test have high instructional validity but low curricular validity? Give an example and explain the implications. 6 CO3

   c) Suppose a teacher designs a test with strong content and construct validity, but students perceive it as irrelevant and demotivating (low face validity). Discuss whether ignoring face validity undermines the overall effectiveness of the test. 7

d) A university entrance test shows a validity coefficient of $r = .25$ with first-year GPA. However, only admitted students (the top scorers) are included in the analysis. Explain how restriction of range may distort this coefficient, and whether the test could still be more valid than it appears. 7

5 a) What is variance? Explain with example. 5 CO2, CO3

a) Variance is often criticized as a descriptive or interpretive statistic because its calculated value is expressed in squared units of measurement. Explain with an example why this is considered a limitation and how this issue is addressed in statistical practice. 10

b) The mean for the following distribution is 80, and the standard deviation is 12. Assuming a normal distribution, compute the z-scores and determine the rank order for the following students: 10

Student	*Student 1*	*Student 2*	*Student 3*	*Student 4*
Score	68	104	86	62

6 a) Define *percentile rank*, *interquartile range* and *standard deviation*. 5 CO2, CO3

b) Find P_{25} and P_{50} in the following distribution of scores: 115, 112, 110, 108, 106, 104, 100, 100, 98, 96, 96, 94, 93, 91, 90, 88. 10

c) Draw the necessary diagram and indicate the type of distribution (in terms of skewness) to which each of these following sets of data refer: 10

(i) Mean = 78.37, median = 78.37, mode = 78.37
(ii) Mean = 374.3, median = 379.7, mode = 391.3
(iii) Mean = 109.5, median = 107.4, mode = 107.4

7 a) A group of 200 students must be classified into five categories — A, B, C, D, and E — based on their ability. Each category represents an equal range of abilities. Assuming that ability follows a normal distribution, determine how many students should fall into each of the categories A, B, C, D, and E. 15 CO2, CO3

b) The distribution of raw scores in a test is normal with $M = 0$ and $SD = 10$. Find out the raw scores that cut off – *(i)* the top 5% and *(ii)* the bottom 5%. 10

Formulas

$$r = \frac{\sum(X_i - \bar{X}_i)(T - \bar{T})}{\sqrt{\sum(X_i - \bar{X}_i)^2 \sum(T - \bar{T})^2}}$$

$$M = \frac{\sum fX}{N}\ ;\ M_d = \left(\frac{N+1}{2}\right) th\ ;\ M_d = L + \left[\frac{(N/2) - F}{f}\right] \times i\ ;\ SD = \sqrt{\frac{\sum(X-\bar{X})^2}{N}}\ ;\ SD = \sqrt{\frac{\sum fx^2}{N}};\ S^2 = \frac{\sum(X-\bar{X})^2}{N}$$

$$PR = \frac{100}{N}\left[F + \left(\frac{X-L}{i}\right) \times f\right];\ z = \frac{X-M}{\sigma}\ ;\ Xi = \tau + ei\ ;\ Xi = \lambda i.\tau + \delta i + ei\ ;\ D = \frac{U-L}{N};\ r_p b = \frac{M_1 - M_0}{S_t}\ x \sqrt{\frac{pq}{n}}$$

$$p = \frac{\text{Number of students who answered correctly}}{\text{Total number of students who attempted the item}}\ ;\ p = \frac{\text{Mean score on the item}}{\text{Maximum possible score}};$$

$$Reliability = \frac{\sigma^2 T}{\sigma^2 X};\ \sigma^2 E = \sigma^2 X - \sigma^2 T;\ r_{test-retest} = Correlation\ (X1, X2);$$

$$r_{xx'} = \frac{Cov(X_1, X_2)}{\sqrt{Var(X_1).Var(X_2)}} = \frac{\sigma^2 T}{\sqrt{\sigma^2 X_1.\sigma^2 X_2}};\ r_{SB} = \frac{2r_{H1,H2}}{1 + r_{H1,H2}};\ r_{rulon} = 1 - \frac{\sigma_D^2}{\sigma_X^2};\ Q = \frac{Q3 - Q1}{2}$$

Cumulative Standardized Normal Distribution

Area between the Mean and Successive Standard Deviation
Units under the Normal Curve (Proportions out of 10,000)

z(x/σ)	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4383	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	.4987									

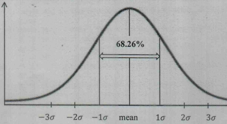

Figure: Area under Normal Curve