

**25UCU305: DISCRETE MATHEMATICS WITH PROBABILITY AND HYPOTHESIS TESTING**

**Year: I (B.Sc., CS, IT, AI-DS, DS &V, BCA, DA, Cyber Security, DevOps, FSWD)**

**Semester: II**

**Unit IV Section A**

1. Define Null Hypothesis ( $H_0$ ) with Example. (K).
2. What is alternate hypothesis? Give an example. (K)
3. State the meaning of degrees of freedom in hypothesis testing. (K)
4. Explain level of significance ( $\alpha$ ) in large sample tests. (C)
5. Define Type I error with its probability notation. (K)
6. What is Type II error? Provide its formula. (K)
7. State the Z-test formula for single population mean. (K)
8. State Z-test statistic for difference between two sample means. (K)
9. A product analyst wants to compare the average session duration between Android and iOS users of a mobile app.

Platform	n	$\bar{x}$ (minutes)	s (minutes)
Android	500	12.8	3.5
iOS	450	14.1	3.1

Assuming large samples, compute the Z-statistic to test whether iOS users have significantly higher session duration than Android users. (Technology Company (Product Analytics)).

A sample of 400 individuals is found to have a mean height of 67.47 inches. Can it be reasonably regarded as a sample from a large population with mean height of 67.39 inches and standard deviation 1.30 inches?

10. A sample of 900 members has a mean 3.5 cms and S.D. 2.61 cms. Is the sample from a large population of mean 3.25 cms and S.D 2.61 cms?
11. A random sample of 1,600 ball bearings has a mean diameter of 50.15 mm and standard deviation 2.6 mm. Can we conclude at 1% level of significance that these bearings come from a large population having mean diameter 50 mm and standard deviation 2.6 mm?
12. A sample of 400 individuals is found to have a mean height of 67.47 inches and standard deviation 1.30 inches. Compute the Z-statistic.
13. A data analyst at a fitness-tracking company wants to understand whether the average daily step count of premium users differs from the company's benchmark value of 10,000 steps. A sample of 500 premium users shows: Sample mean = 9,820 steps Sample standard deviation = 1,450 steps. Use a Z-test for sample mean to calculate the Z-statistic.(Scalar AIML / NxtWave / Google Data Analytics).

**Section B**

14. The mean breaking strength of cables supplied by a manufacturer is 1800 with a S.D of 100. By a new technique in the manufacturing process it is claimed that the breaking strength of the cables has increased. In order to test this, claim a sample of 50 cables is tested. It is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance?
15. An ambulance service claims that it takes on an average 8.9 minutes to reach its destination in emergency calls. To check on this, claim the agency which licenses ambulance services has them timed on 50 emergency calls, getting a mean of 9.3 minutes with a standard deviation of 1.6 minutes. What can they conclude at 5% level of significance?
16. A random sample of 625 bulbs manufactured by a company has an average life of 1,210 hours with a standard deviation of 80 hours. Can we conclude that these bulbs are drawn from a large population having mean life 1,200 hours and standard deviation 80 hours? Compute the Z-statistic and test the hypothesis at 5% level of significance.
17. In a study of the effect of chemical on the labourers in a chemical unit, the following results were obtained on their systolic blood pressures. Compute the pooled variance.

	Males	
	Exposed group	Controlled groups
No	250	55
Mean	117.5	121.6
S.D	10.58	10.82

18. A random sample of 200 villages from Coimbatore district gives mean population per village at 485 with a S.D of 50. Another random sample of the same size from the same district gives the mean population per village at 510 with a S.D of 40. Is the difference between the mean values given by the two samples statistically significant? Justify your answer. Assume  $\sigma$  is not known.
19. Explain the procedure of test of significant difference between the population mean and Sample mean for large sample test.
20. A global streaming platform (similar to Netflix) historically claims that the average daily watch time per user is 92 minutes with a known population standard deviation of 28 minutes. After launching a new recommendation engine, the analytics team wants to verify if the average watch time has changed. They randomly sample 2,500 active users and record the following data:

Parameter	Value
Claimed Population Mean ( $\mu_0$ )	92.0 minutes
Known Population Standard Deviation ( $\sigma$ )	28.0 minutes
Sample Size (n)	2,500 users

Sample Mean ( $\bar{x}$ )	94.6 minutes
Level of Significance ( $\alpha$ )	5% (0.05)

Perform the large-sample Z-test (using the known population  $\sigma = 28$  minutes) to test whether the average daily watch time has significantly changed after the new recommendation engine. (Used in Google Data Analytics Certificate, Microsoft PL-300, IBM Data Science, SCALAR AIML, NXT Wave 2024–2025).

Explain the procedure of test of significant difference between the two-sample means for large sample test.

21. A multinational e-commerce company claims that the average time customers spend on its mobile app is 12 minutes per session, with a population standard deviation of 3.8 minutes. To optimize server capacity, the data analytics team collects a random sample of 1,600 user sessions after a recent app update and finds: Sample mean time = 12.45 minutes Sample standard deviation = 3.9 minutes. Using the known population standard deviation of 3.8 minutes. Perform a hypothesis test at the 5% significance level to determine whether the average session time has changed after the update. (Designed for SCALAR Data Science, SCALAR AIML, NXT Wave, Google Data Analytics Certificate, IBM Data Science Professional Certificate, and Microsoft Power BI Data Analyst exams)
22. A/B test on a shopping website: Version A (control): 1,500 users, average time on site = 4.8 minutes,  $s = 2.1$  min Version B (new design): 1,350 users, average time on site = 5.2 minutes,  $s = 2.3$  min Test at 5% significance whether the new design increases average time spent on the site. (Use two-tailed test first, then state if one-tailed is more appropriate.). (Google Data Analytics Certificate – Course 8 (2024)).
23. A retail chain compares customer satisfaction scores (0–10) between online and in-store purchases. Online:  $n = 2,000$ , mean = 8.45,  $s = 1.6$  In-store:  $n = 1,800$ , mean = 8.30,  $s = 1.7$  Using 5% significance level, is there evidence that online customers are more satisfied? (Microsoft PL-300 Power BI Analyst Exam (2024)).
24. Two machine-learning models were trained on the same dataset. Accuracy on validation set: Model X (TensorFlow): 1,024 samples, mean accuracy = 92.3%,  $s = 3.8\%$  Model Y (PyTorch): 900 samples, mean accuracy = 91.5%,  $s = 4.1\%$  At  $\alpha = 0.01$ , test whether Model X has significantly higher accuracy than Model Y. (IBM Data Science Professional Certificate – Final Project Style).
25. An e-commerce company runs a promotion in two regions: Region East: 2,500 orders, average order value = \$78.40,  $\sigma \approx \$22$  Region West: 2,200 orders, average order value = \$81.20,  $\sigma \approx \$23$  Test at 1% level whether the average order value differs between the regions. (AWS Certified Data Analytics – Specialty (2024)) .
26. An online advertising team is running an A/B test to compare the performance of two ad copies for a new product launch.

	Number of Impressions (Sample Size)	Mean Click-Through Rate (CTR)	Standard Deviation of CTR
Red (Control)	5,000	4.82%	2.1%
Blue (New)	4,800	5.31%	2.3%

Using a large-sample Z-test at 5% level of significance, answer the following:

- (a) State the null and alternative hypotheses to test whether the new Blue ad copy performs significantly better. (b) Compute the Z-statistic (c) State the critical value and decision rule. (d) Make the statistical decision and write the conclusion. (e) Give one practical recommendation to the marketing team based on the result. (Google Data Analytics (2025 Mock Exam))

27. Meta (Facebook) is testing a new feed algorithm to increase user engagement. They randomly assigned users to two groups and measured Average Daily Active

Group	Sample Size (Users)	Mean DAU (minutes)	Standard Deviation (minutes)
Control (Old Algorithm)	3,000	42.6	15.4
Treatment (New Algorithm)	2,700	45.1	16.2

Usage (DAU) in minutes.

Using a large-sample Z-test at 5% significance level ( $\alpha = 0.05$ ), test whether the new algorithm significantly increases daily usage. Answer the following: (a) State the null ( $H_0$ ) and alternative ( $H_1$ ) hypotheses. (b) Compute the Z-statistic (show complete calculations). (c) What is the critical value for this one-tailed test? (d) State your statistical decision and conclusion. (e) What recommendation would you give to the product team? (Meta (Facebook) Data Analyst Mock Test (2024)).

#### Unit IV Section C (10 Marks)

28. Samples of students were drawn from two universities in kgm, means and standard deviations are calculated. Make large sample test to the significance of the difference between the means.

	Sample size	Mean	S.D.
University A	400	55	10
University B	100	57	15

29. The average hourly wage of a sample of 150 workers in a plant A was Rs. 2.56 with a S.D. of Rs. 1.08. The average wages of a sample of 200 workers in a plant B was

2.87 with a S.D of Rs. 1.28. Can an applicant safely assume that the hourly wages paid by plant B are lower than those paid by plant A? use 1% level of significance.

30. The electric light bulbs of manufacturer A, have mean life time of 1600 hours with a S.D of 200 hours, while those of manufacturer B have a mean life time of 1400h with a S.D of 100 hr. If random samples of 125 bulbs from each manufacturer are selected, then what is the probability that the bulbs will have a mean lifetime that is at least (a) 160 and (b) 250h more than B bulbs?
31. Two samples of cables of sizes each 100 are taken in order to test the breaking strength and their mean and standard deviations are given.

	Sample size	Sample Mean	Sample SD
Cable I	100	1925	1905
Cable II	100	40	30

Test whether mean breaking strength of the two cables are equal at 1% level of significance.

32. A food-delivery app tests two recommendation algorithms: Algo 1: 1,600 deliveries, average delivery time = 32.8 minutes,  $s = 8.2$  min Algo 2: 1,400 deliveries, average delivery time = 30.9 minutes,  $s = 7.9$  min Using 5% significance, decide if Algo 2 significantly reduces delivery time. (SCALAR AIML / NXT Wave Certification (2024–2025)).
33. A global e-commerce company claims that the average delivery time for standard shipments in Europe is 48 hours with a known population standard deviation of 12 hours.

To monitor SLA (Service-Level Agreement) compliance, the Data Analytics team randomly selects 1,600 delivery records from last month.

Parameter	Value
Claimed Population Mean ( $\mu_0$ )	48.0 hours
Known Population Standard Deviation ( $\sigma$ )	12.0 hours
Sample Size (n)	1,600
Sample Mean Delivery Time ( $\bar{x}$ )	49.2 hours
Level of Significance ( $\alpha$ ) – Part (a)–(d)	5% (0.05)
Level of Significance – Part (e)	1% (0.01)

Using the large-sample Z-test procedure and the known population  $\sigma = 12$  hours, answer the following. State the null and alternative hypotheses (specify one-tailed or two-tailed). (b) Compute the Z-statistic (show complete working). (c) Test the hypothesis at **5%** level of significance and state the statistical conclusion. (d) Give one clear business recommendation to the operations/supply-chain team. (e) If the company wants to be 99% confident ( $\alpha = 0.01$ ), would the conclusion change? Perform the calculation and justify. (Real-world style used in SCALAR AIML, NXT Wave, Google Data Analytics Professional Certificate, and IBM Data Science Capstone 2024–2025).

- 34.** An international ride-hailing company (similar to Uber) operates in two cities: City A and City B. The data science team wants to determine whether the average ride fare is the same in both cities after a new dynamic-pricing algorithm was rolled out. From the transaction database, they extract two independent random samples:

City	Sample Size (n)	Sample Mean Fare (USD)	Sample Standard Deviation (USD)
City A	1,200 rides	\$18.40	\$5.10
City B	980 rides	\$17.85	\$5.30

Assume the population standard deviations are approximately equal to the sample values (valid for large samples). Perform a two-sample large-sample Z-test at **5%** significance level to decide if there is a significant difference in the average ride fare between the two cities. Compute the following.

- (a) State  $H_0$  and  $H_1$ . (b) Calculate the Z-statistic (c) What is the critical value? (d) State the decision and conclusion. (e) Give one business recommendation based on the result. (Google Data Analytics Professional Certificate (Course 8), Microsoft PL-300, IBM Data Science Professional Certificate, SCALAR AIML/NXT Wave hypothesis testing modules, and university external papers.) .