**Final questions for students**

1. Define and describe population, sample, parameter and statistic.
2. Open the following expression:
3. Define the scales of measurement for all of the following variables: eye color, number of children per family, GDP, electricity consumption, water pressure.
4. Give an example of 5 continuous and discrete variables
5. Give a proper definition of a random and deterministic variables with some examples.
6. You have 10 observations of some variable. Build a summary (frequency) table: 1, 2, 3, 2, 2, 4, 3, 4, 2, 5
7. Describe the methods used to summarize continuous numerical data.
8. Describe the methods used to summarize discrete numerical data.
9. Which graphical forms are used for visualizing the qualitative data?
10. If the data is described in histogram what is the most likely type of the variable
11. P(A) = 0.25, P(B) = 0.5. If the two event are statistically independent what is the value for joint probability.
12. The probability distribution of a variable (collectively exhausting outcomes) is given as below. Find the value of b.

|  |  |
| --- | --- |
| X | P(X) |
| 1 | c |
| 3 | 3c |
| 5 | 9c |
| 7 | 27c |

1. Probability of graduating from university in the UK is 0.7. The joint probability of being a foreign student and graduating is 0.3. If the person graduated, what is the probability that s/he is a foreign student?
2. The probabilities for collectively exhaustive outcomes are given as follows:

|  |  |
| --- | --- |
| X | P(X) |
| 0 | 0.4 |
| 1 | 0.2 |
| 2 | 0.27 |
| 3 | 0.13 |

Find the ratio of cumulative probability of 3 to the cumulative probability of 0.

1. P(A) = 0.4 and P(B) = 0.6. If the two are statistically independent, find the probability of P(A or B).
2. Skewness of the variable is 0 whereas the kurtosis is 3. Comment on symmetricity and tail probability.
3. Calculate mean and variance for the following sample data: 1, 3, 4, 3, 4, 1, 1, 5, 1, 0
4. The means for two variables, X and Y, are 6 and 5 respectively. Find the mean the following combination: 5X + Y
5. The portfolio variance is 0.04. There are 2 assets in the portfolio with equal weights. The covariance of the assets is 0.02. If the first asset has the variance of 0.04, find the variance of the second.
6. E(X) = 3 and var(X) = 5. Find E(X2)
7. The Bernoulli variable has p = 0.2. What is the standard deviation of the variable?
8. On average 20% of the mobile phones tend to be defect. If you bought 5 mobile phones what is the probability that 2 of them will be defect?
9. On average there are 4 car accidents a day. What is the probability that next day there will be less than 3 car accidents?
10. There are 8 balls in the box and 4 of them are green. If I take 5 of them without returning back what is the probability that none of them will be green?
11. Assume you have 4 friends (Nizami, Baxtiyar, Tofig and Zaur) and a car. You have an agreement that, every time you go somewhere in your car they should sit in a different order. What is the (joint) probability that Zaur will be in the left back and Tofiq will be in the middle back seat?
12. The level of water in the electric station dam (reservoir) follows normal distribution with the average of 30 m and variance of 49. If the level drops below 5 m the electric generators will not work. What is the probability of that?
13. X ~ N(3, 16). Cumulative probability of certain X is 20%. Find that value of X
14. X follows normal distribution and has population mean of 10. P(X > 15) is 2.5%. Find the variance of the variable.
15. X follows normal distribution. P(X > 5) = 30%, P(X < 2) = 10%. Find mean and variance of X.
16. Y ~ N(5, 8) and Z ~ N(0, 1). Find the distribution of Y + 10Z
17. Variable has uniform distribution with mean and variance equal to 5 and 4 respectively. If you take a sample of 20 observations from the population, what would be the distribution of the sample means?
18. Population proportion of blue-eyed people is 10%. If in the sample you had 50 people what would be the sampling distribution of sample proportions?
19. Assume you are looking into sampling distribution of sample mean where you do not have population variance, but you know population mean. If you have 30 observations, which distributions you would use to check validity of the sample and what would be your critical value with 95% confidence level?
20. Apply CLT and find distribution of Xave + Yave where Xave is the average of 40 observations from binomial with n = 100 and p = 0.2, and Yave is the average of 50 observations from normal with mean of 5 and variance of 25.
21. Assume you know the population mean of daily forex returns which is 0. You do not know the population variance. You have taken a sample of 49 observations from the population and sample mean and variance are 0.005 (to make it clear - 0.5%) and 0.0004 respectively. Is this sample representative based on the mean value?
22. The sample has a skewness of 0.2 and kurtosis of 3.1. Compare this data to normal distribution.
23. Sample mean of 15 observations was 30. If the population has fairly symmetric distribution and population variance is 25, can we claim that this sample is a good approximate for population when population mean is equal to 33?
24. Sample variance of 15 observations was 30. If the population variance is 25, can we claim that this sample is a good approximate for population when it comes to variance?
25. Assume you are comparing average math test results of Chinese and American students. You pair 50 students, then you find the difference between the results in each pair. Both average and standard deviation of the difference are 5 points. Can we claim that there is no difference in population means with 90% confidence?
26. Assume you are comparing math test results of Chinese and Americans. For them to be equal, the population means and variances should be equal. Sample variance of 25 Chinese students’ math scores was 42 and sample variance of 15 American students’ math scores was 52. Can we claim that the population variances are equal with 90% confidence?
27. Population variance and sample variance for the number of daily equipment failures in some large factory are 25 and 36 respectively. If the sample had 40 observations and sample mean was 70, find 90% interval for population mean.
28. Sample proportion of smokers was 15% in a sample of 200 people. Find 95% confidence interval for population proportion of smokers.
29. The number of items sold in auctions in New-York every day follows binomial distribution with variance of 5. If you observed the sales for November and December and the daily average number of sales was 3.5, build 95% confidence interval for the population mean.
30. Sampling variance of the sample proportions is equal to 0.01. If you had 16 observations in the sample what is population proportion.
31. Sample variance of 8 observations is 9. Find 95% confidence interval for population variance.
32. The population variance of the Cola amount in the bottle is 0.0016 liter-squared. In a sample of 36 observations you found that the sample average was 0.99 liter and sample variance was 0.0009 liter-squared. Find 99% confidence interval for the mean of Cola amount.
33. Test the hypothesis that average consumer eats less than 7kg of meat every months given the average and standard deviation of 50 people in April were 3kg and 4kg respectively using 90% confidence level.
34. For the last 50 days the number of defects in a production line of cars has had the sample average of 50. Assume the population variance is 50. Test the hypothesis that population mean number of defects is below 41.
35. For the last 50 days the number of defects in a production line of cars has had the sample average of 15. Assume the population variance is 25. Find out p-value associated with the hypothesis that population mean number of defects is below 14.
36. The number of workplace incidents for the last 60 months has the average of 2 and variance of 2.25. Test the hypothesis that on average the place has more than 3 incidents a month.
37. The daily average number of packages sent by the post office was the 20000 in April. Assume population variance of 1000000. Test the hypothesis that post office sends 2100 packages a day.
38. The daily average number of packages sent by the post office was the 2000 in April. Assume population variance of 100. Find the p-value for the hypothesis that post office sends 2010 packages a day.
39. Branch manager claims that the annual percentage of fraudulent loans is below 5%. Assuming that for the last 30 years under the same manager’s rule the percentage has been 7%, test the manager’s claim.
40. For the last 40 days on average 1 million people used Baku Metro every day. The standard deviation was 200 000. Using 99% confidence level, test the hypothesis that the population average of Metro users is 1.1 million.
41. Sample standard deviation of TOEFL test scores for 62 Azerbaijani students was 8. Test the hypothesis that population mean is above 110 in case the sample mean is 83.
42. Sample standard deviation of TOEFL test scores for 22 Azerbaijani students was 8. Test the hypothesis that population variance above 70.
43. The population variances of math test scores are 25 and 36 for women and men respectively. If the average of test scores were 78 for 40 women and 71 for 50 men, can we claim that the population averages are equal?
44. Assume you are trying to compare the population means of two groups. You statistically test and see that the population variances are equal. In the first sample you have 30 observations and sample variance is 20. In the second sample you have 40 observations and sample variance is 25. Calculate pooled variance.
45. Pooled variance for two samples is 40. You have 20 observations in the first and 15 observations in the second sample. If the sample averages are 90 and 94 respectively, can we claim that the population means are equal with 5% significance?
46. You are trying to prove that ability to jump higher is genetic (sorry, could not think of something useful). Assume you take 5 twins which were raised in different places and measure how high (in terms of meters) they can jump. The results are shown in the table below. Is the hypothesis in the first sentence valid with 90% confidence?

|  |  |
| --- | --- |
| 1st sibling | 2nd sibling |
| 1.5 | 2 |
| 1.1 | 1.3 |
| 1.2 | 1.8 |
| 1 | 1.6 |
| 2 | 2.1 |

1. Usually the people in Azerbaijan think that if you have the foreign education you get higher salaries. We took two samples: 1) a sample of 30 locally educated people 2) a sample of 20 foreign educated people. The first sample has average salary of 1000 AZN and standard deviation of 200 AZN. The second sample has the average salary of 1500 AZN and standard deviation of 500 AZN. If we assume that the population variances for those groups *are equal*, is there significant salary difference with 90% confidence level?
2. You want to apply pooled-variance t-test for comparison of two sample means. *The first sample* has 10 observations, the average of 10 and standard deviation of 5. *The second sample* has 15 observations, the average of 12 and the standard deviation of 6. Is it right to apply pooled-variance test here?
3. There are two groups of students: 1) state university students 2) private university students. Assume the population variance for their GMAT scores is not the same. You take a sample of 100 students from the first and 225 students from the second group. The sample means and standard deviations are: for group (1) 500 and 100, for group (2) 550 and 45. Are their results statistically different?
4. You are comparing 3 population means. In total you have 50 observations. SST is 2300 whereas SSW is 1300. What is the test conclusion?
5. There are 4 companies which sell spaghetti. A cook claims that the average cooking time is the same for all 4 brands. You are tasked to test this claim. As an experiment you are allowed to cook each brand 20 times. After calculations you get SST equal to 1200 whereas SSA is equal to 1020. What is the result of the testing?
6. Site the assumptions required for ANOVA testing.
7. Compute MSA if you are comparing 3 populations based on the following 3 samples: i) 4, 5, 6; ii) 3, 7, 5; iii) 2, 4, 9.
8. Assume you are trying to test the association between two categorical variables: gender and extroversion. You take a sample of 100 people and the results are described in the following below contingency table. Does gender affect social type of the person?

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Social type | |
| Introvert | Extrovert |
| Gender | Male | 20 | 35 |
| Female | 15 | 30 |

1. Assume you are trying to test the crazy idea like whether the gender of children born is affected by the month of the year. What distribution would you use for testing in this case and what would be your degrees of freedom?
2. Assume there are three general types of food: diet food, normal food, fatty food. You assume the people consume different proportions in winter and summer. You collect information on 100 different people and see what they consume. Based on the results in below table can we claim that the proportion changes between seasons?

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Season | |
| Winter | Summer |
| Food | Diet | 10 | 10 |
| Normal | 30 | 20 |
| Fatty | 20 | 10 |

1. The regression analysis yields the following result: Y = 5 + 3X + e. What is the value of Y when X is 4?
2. The regression and standard errors of the coefficients are estimated as follows. Test the hypothesis that slope coefficient is equal to 0 (assume critical equal to 2).
3. The regression and standard errors of the coefficients are estimated as follows. Test the hypothesis that intercept is equal to 0.
4. Given the following regression, explain the meaning of intercept: ***Exchange rate = 1.3 – 0.007\*Trade\_balance + e***
5. Given the following regression, explain the meaning of slope coefficient (Assume the trade balance is measured in billions of dollars): ***Exchange rate = 1.4 – 0.02\*Trade\_balance + e***