Müəllimin adı: Osmanli Xeyse

Fənnin adı: Statistics

Qrupun nömrəsi: 1043/1062/1025-1026

Mövzu 1: Review of Probability Theory

1. BurgerTown offers many options for customising a burger. There are 3 types of meats and 7 condiments: lettuce, tomatoes, pickles, onions, ketchup, mustard, and special sauce. A burger must include meat, but may include as many or as few condiments as the customer wants. How many different burgers are possible?

2. A small nation issues license plates that consist of just one number (selected from the digits 0 through 9, inclusive) and four letters, selected from a 20-letter alphabet. Repeats are permitted. However, there is one four-letter combination that is not allowed to appear on license plates. How many allowable license plate combinations exist?

3. If a, b, and c are integers randomly chosen from the set of prime numbers greater than 2 and less than 30, what is the probability that ab + c is equal to 23?

4. A student council is to be chosen from a class of 12 students consisting of a president, a vice president, and 3 committee members. How many such councils are possible?

5.How many distinct 4-letter “words” can be made from the name “CHRISTYNA”? (A “word” is any arrangement of 4 letters regardless of whether it can be found in a dictionary.)

Mövzu 2: Review of mathematical Statistics

1.The probability of rain is 1/6 for any given day next week. What is the chance it rains on both Monday and Tuesday?

2.A bag contains 3 red, 2 blue, and 7 white marbles. If a marble is randomly chosen from the bag, what is the probability that it is NOT blue?

3.A bag contains 6 red chips numbered 1–6 respectively and 6 blue chips numbered 1–6 respectively. If 2 chips are to be picked sequentially from the bag of 12 chips, without replacement, what is the probability of picking a red chip and then a blue chip with the same number?

4.A coin with heads on one side and tails on the other has a 1/2 probability of landing on heads. If the coin is flipped 5 times, how many distinct outcomes are possible if the last flip must be heads? Outcomes are distinct if they do not contain exactly the same results in exactly the same order.

5. A 6-sided cube has sides numbered 1 through 6. If the cube is rolled twice, what is the probability that at least one of the rolls will result in a number higher than 4?

Mövzu 3: Introduction to data analysis

1.For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the measurement scale.

a. Number of telephones per household
b. Length (in minutes) of the longest telephone call made in a month
c. Whether someone in the household owns a Wi-Fi-capable cell phone
d. Whether there is a high-speed Internet connection in the household

2. The following information is collected from students upon exiting the campus bookstore during the first week of classes.

a. Amount of time spent shopping in the bookstore
b. Number of textbooks purchased
c. Academic major
d. Gender

Classify each of these variables as categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the measurement scale for each of these variables.

3. For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the measurement scale for each variable.

a. Name of Internet service provider
b. Time in hours spent surfing the Internet per week
c. Number of emails received in a week
d. Number of online purchases made in a month

4. For each of the following variables, determine whether the variable is categorical or numerical. If the variable is numerical, determine whether the variable is discrete or continuous. In addition, determine the measurement scale for each variable.

a. Amount of money spent on clothing in the past month
b. Favourite department store
c. Most likely time period during which shopping for clothing takes place (weekday, weeknight, or weekend)
d. Number of pairs of shoes owned

5.Suppose the following information is collected from Robert Keeler on his application for a home mortgage loan at the Metro County Savings and Loan Association.
a. Monthly payments: $1,927

b. Number of jobs in past 10 years: 1
c. Annual family income: $76,000

d. Marital status: Married
Classify each of the responses by type of data and measurement scale.

Mövzu 4: Organising and Visualising Data

1. The GMAT scores from a sample of 50 applicants to an MBA program indicate that none of the applicants scored below 450. A frequency distribution was formed by choosing class intervals 450 to 499, 500 to 549, and so on, with the last class having an interval from 700 to 749. Two applicants scored in the interval 450 to 499, and 16 applicants scored in the interval 500 to 549.

What percentage of applicants scored below 500?
What percentage of applicants scored between 500 and 549?
What percentage of applicants scored below 550?
What percentage of applicants scored below 750?
2. A set of data has values that vary from 11.6 to 97.8.

a. If these values are grouped into nine classes, indicate the class boundaries.
b. What class interval width did you choose?
c. What are the nine class midpoints?

3. A survey asked 1,264 women who were their most trusted shopping advisers. The survey results were as follows:

a. Construct a bar chart, a pie chart, and a Pareto chart.
b. Which graphical method do you think is best for portraying these data?
c. What conclusions can you reach concerning women's most trusted shopping advisers?

Advertising	**7 %**
Friends/family	**45 %**
Manufacturer websites	**5 %**
News media	**11 %**
Online user reviews	**13 %**
Retail websites	**4 %**
Salespeople	**1 %**
Other	**14 %**

4. The following is a stem-and-leaf display representing the amount of gasoline purchased, in gallons (with leaves in tenths of gallons), for a sample of 25 cars that use a particular service station on the New Jersey Turnpike:

a. Construct an ordered array.
b. Which of these two displays seems to provide more information? Discuss.
c. What amount of gasoline (in gallons) is most likely to be purchased?
d. Is there a concentration of the purchase amounts in the centre of the distribution?

9|147
10|02238
11|125566777
12|223489
13|02

5. The following is a set of data from a sample of n = 11 items:

a. Construct a scatter plot.
b. Is there a relationship between X and Y? Explain.

X: **7 5 8 3 6 0 2 4 9 5 8**
Y: **1 5 4 9 8 0 6 2 7 5 4**

Mövzu 5: Organising Multidimensional Data

1. A study of 1,000 people asked what respondents wanted to grill during barbecue season. The results were as follows:

a. Construct a bar chart, a pie chart, and a Pareto chart.
b. Which graphical method do you think is best for portraying these data?
c. What conclusions can you reach concerning what folks want sizzling on the grill during barbecue season?

Beef	**38 %**
Chicken	**23 %**
Fruit	**1 %**
Hot dogs	**6 %**
Pork	**8 %**
Seafood	**19 %**
Vegetables	**5 %**

2.Construct a stem-and-leaf display, given the follow- ing data from a sample of midterm exam scores in finance:

54 69 98 93 53 74

3. How do histograms and polygons differ in their construction and use? Compare and contrast the bar chart for categorical data with the histogram for numerical data. What are the three different ways to break down the percentages in a contingency table? How can a multidimensional table differ from a two variable contingency table?

4.Why would you construct a summary table?

What is the difference between a time-series plot and a scatter plot?
What type of insights can you gain from a three-way table that are not available in a two-way table?

5. What are the advantages and disadvantages of using a bar chart, a pie chart, and a Pareto chart?

Why is it said that the main feature of a Pareto chart is its ability to separate the “vital few” from the “trivial many”?
What type of insights can you gain from a three-way table that are not available in a two-way table?

6-ci movzu : Numerical Descriptive Masures

1. The following set of data is from a sample of $n = 5$:

7 4 9 8 2

a. Compute the mean, median, and mode.
b. Compute the range, variance, standard deviation, and coefficient of variation.
c. Compute the *Z* scores. Are there any outliers?
d. Describe the shape of the data set.

2. The following is a set of data from a sample of $n = 6$:

7 4 9 7 3 12

a. Compute the first quartile (Q1) , the third quartile (Q3) , and the interquartile range.
b. List the five-number summary.
c. Construct a box-plot and describe its shape.

3. Suppose that the rate of return for a particular stock during the past two years was 20% and -30%. Compute the geometric rate of return per year.

4. The following is a set of data for a population with $N = 10$:

7 5 6 6 6 4 8 6 9 3

a. Compute the population mean.
b. Compute the population standard deviation.

5. The following set of data is from a sample of $n = 7$:

1 2 7 4 9 0 7 3

a. Compute the mean, median, and mode.
b. Compute the range, variance, standard deviation, and coefficient of variation.
c. Compute the *Z* scores. Are there any outliers?
d. Describe the shape of the data set.

7. movzu: Numerical Descriptive Measures for a Population

1. The following is a set of data from a sample of $n = 5$:

7 4 9 8 2

a. Compute the first quartile (Q1) , the third quartile (Q3) , and the interquartile range.
b. List the five-number summary.
c. Construct a box-plot and describe its shape.

2. Suppose that the rate of return for a particular stock during the past two years was 10% and 30%. Compute the geometric rate of return per year. (*Note:* A rate of return of 10% is recorded as 0.10, and a rate of return of 30% is recorded as 0.30.)

3. Consider a population of 1,024 mutual funds that primarily invest in large companies. You have determined that m, the mean one-year total percentage return achieved by all the funds, is 8.20 and that s, the standard deviation, is 2.75.

a. According to the empirical rule, what percentage of these funds are expected to be within -/+1 standard deviation of the mean?

b. According to the empirical rule, what percentage of these funds are expected to be within -/+2 standard deviations of the mean?

c. According to the Chebyshev rule, what percentage of these funds are expected to be within -/+1, -/+2, or -/+3
standard deviations of the mean?

d. According to the Chebyshev rule, at least 93.75% of these funds are expected to have one-year total returns between what two amounts?

4. The following set of data is from a sample of $n = 6$:

7 4 9 7 3 1 2

a. Compute the mean, median, and mode.
b. Compute the range, variance, standard deviation, and coefficient of variation.
c. Compute the *Z* scores. Are there any outliers?
d. Describe the shape of the data set.

5. The following is a set of data from a sample of $n = 11$ items:

X	**7 5 8 3 6 10 12 4 9 15 18**
Y	**21 15 24 9 18 30 36 12 27 45 54**

a. Compute the covariance.

b. Compute the coefficient of correlation.

c. How strong is the relationship between *X* and *Y*? Explain.

8.movzu: Basic Probability

1. A sample of 500 respondents in a large metropolitan area was selected to study consumer behaviour. Among the questions asked was “Do you enjoy shopping for clothing?” Of 240 males, 136 answered yes. Of 260 females, 224 answered yes. Construct a contingency table to evaluate the probabilities. What is the probability that a respondent chosen at random

a. enjoys shopping for clothing?
b. is a female *and* enjoys shopping for clothing?
c. is a female *or* enjoys shopping for clothing?
d. is a male *or* a female?

2. For each of the following, indicate whether the type of probability involved is an example of *a priori* probability, empirical probability, or subjective probability.

a. The next toss of a fair coin will land on heads.
b. Italy will win soccer's World Cup the next time the com- petition is held.
c. The sum of the faces of two dice will be seven.
d. The train taking a commuter to work will be more than 10 minutes late.

3. If P(A *and* B)=0.4 and P(B)=0.8 , find P(A|B)
If P(A)=0.7 and P(B)=0.6 , and A and B are independent , find P(A *and* B)
If P(A)=0.3 and P(B)=0.4 , find P(A *and* B) = 0.2 , are A and B independent ?

4. An urn contains 12 red balls and 8 white balls. One ball is to be selected from the urn.
a. Give an example of a simple event.
b. What is the complement of a red ball?

5. Each year, ratings are compiled concerning the performance of new cars during the first 90 days of use. Suppose that the cars have been categorised according to whether a car needs warranty-related repair (yes or no) and the country in which the company manufacturing a car is based (United States or not United States). Based on the data collected, the probability that the new car needs a warranty repair is 0.04, the probability that the car was manufactured by a U.S.-based company is 0.60, and the probability that the new car needs a warranty repair *and* was manufactured by a U.S.-based company is 0.025. Construct a contingency table to evaluate the probabilities of a warranty-related repair. What is the probability that a new car selected at random

9.movzu:Bayes' Theorem

1. A municipal bond service has three rating categories (*A*, *B*, and *C*). Suppose that in the past year, of the municipal bonds issued throughout the United States, 70% were rated *A*, 20% were rated *B*, and 10% were rated *C*. Of the municipal bonds rated *A*, 50% were issued by cities, 40% by suburbs, and 10% by rural areas. Of the municipal bonds rated *B*, 60% were issued by cities, 20% by suburbs, and 20% by rural areas. Of the municipal bonds rated *C*, 90% were issued by cities, 5% by suburbs, and 5% by rural areas.
a. If a new municipal bond is to be issued by a city, what is the probability that it will receive an *A* rating?
b. What proportion of municipal bonds are issued by cities?
c. What proportion of municipal bonds are issued by suburbs?

2. You would like to make a salad that consists of lettuce, tomato, cucumber, and peppers. You go to the super- market, intending to purchase one variety of each of these ingredients. You discover that there are eight varieties of lettuce, four varieties of tomatoes, three varieties of cucumbers, and three varieties of peppers for sale at the supermarket. If you buy them all, how many different salads can you make?

3. A team is being formed that includes four different people. There are four different positions on the teams. How many different ways are there to assign the four people to the four positions??

4. Laid-off workers who become entrepreneurs because they cannot find meaningful employment with another company are known as *entrepreneurs by necessity*. *The Wall Street Journal* reports that these entrepreneurs by necessity are less likely to grow into large businesses than are *entrepreneurs by choice* (J. Bailey, "Desire—More Than Need—Builds a Business," *The Wall Street Journal*, May 21, 2001, p. B4). This article states that 89% of the entrepreneurs in the United States are entrepreneurs by choice and 11% are entrepreneurs by necessity. Only 2% of entrepreneurs by necessity expect their new business to employ 20 or more people within five years, whereas 14% of entrepreneurs by choice expect to employ at least 20 people within five years.
a. If an entrepreneur is selected at random and that individ- ual expects that his or her new business will employ 20 or more people within five years, what is the probability that this individual is an entrepreneur by choice?
b. Discuss several possible reasons why entrepreneurs by choice are more likely than entrepreneurs by necessity to believe that they will grow their businesses.

5. A student has seven books that she would like to place in her backpack. However, there is room for only four books. Regardless of the arrangement, how many ways are there of placing four books into the backpack?

10.movzu: The Normal Distribution and other Continuous Distributions

1. Given a standardised normal distribution (with a mean of 0 and a standard deviation of 1, as in Table E.2), what is the probability that

a. *Z* is between -1.57 and 1.84?

b. *Z* is less than -1.57 or greater than 1.84?

c. What is the value of *Z* if only 2.5% of all possible *Z* values are larger?

d. Between what two values of *Z* (symmetrically distributed around the mean) will 68.26% of possible *Z* values be contained?

2. Some Internet companies sell a service that will boost a website's traffic by delivering additional unique visitors. Assume that one such company claims it can deliver 1,000 visitors a day. If this amount of website traffic is experienced, then the time between visitors has a mean of 1.44 minutes (or 0.6944 per minute). Assume that your website gets 1,000 visitors a day and that the time between visitors has an exponential distribution. What is the probability that the time between two visitors is
a. less than 1 minute?
b. less than 2 minutes?
c. more than 3 minutes?
d. Do you think it is reasonable to assume that the time between visitors has an exponential distribution?

3. The scheduled commuting time on the Long Island Railroad from Glen Cove to New York City is 65 minutes. Suppose that the actual commuting time is uniformly distributed between 64 and 74 minutes. What is the probability that the commuting time will be

a. less than 70 minutes?

b. between 65 and 70 minutes?

c. greater than 65 minutes?

d. What are the mean and standard deviation of the commuting time?

4. Telephone calls arrive at the information desk of a large computer software company at a rate of 15 per hour.

a. What is the probability that the next call will arrive within 3 minutes (0.05 hour)?

b. What is the probability that the next call will arrive within 15 minutes (0.25 hour)?

c. Suppose the company has just introduced an updated version of one of its software programs, and telephone calls are now arriving at a rate of 25 per hour. Given this information, what are your answers to (a) and (b)?

5. Given a standardised normal distribution (with a mean of 0 and a standard deviation of 1, as in Table E.2), deter- mine the following probabilities:

a. $P(Z > 1.08)$

b. $P(Z < -0.21)$

c. $P(-1.96 < Z < -.021)$

d. What is the value of Z if only 15.87% of all possible Z values are larger?

1.

Each of three balls are randomly placed into one of three bowls. Find the probability distribution for Y = the number of empty bowls.

2.

a Find the probability distribution for Y, the number of errors detected by the auditor.

b Construct a probability histogram for $p(y)$.

c Find the probability that the auditor will detect more than one error.

3.

a If five volunteers are randomly selected, what is the probability that at least one does not have the Rh factor?

b If five volunteers are randomly selected, what is the probability that at most four have the Rh factor?

c What is the smallest number of volunteers who must be selected if we want to be at least 90% certain that we obtain at least five donors with the Rh factor?

4.

a Find the probability that exactly 14 survive.

b Find the probability that at least 10 survive.

c Find the probability that at most 16 survive.

d Find the mean and variance of the number that survive.

5.

a $E(Y)$

b $V(Y)$

6.

Let Y possess a density function

a Find c.

b Find $F(y)$.

c Graph $f(y)$ and $F(y)$.

d Use $F(y)$ in part (b) to find $P(1 \leq Y \leq 2)$.

e Use geometry and the graph for $f(y)$ to calculate $P(1 \leq Y \leq 2)$.

7.

Let the distribution function of a random variable Y be

a Find the density function of Y.

b Find $P(1 \leq Y \leq 3)$.

c Find $P(Y \geq 1.5)$.

d Find $P(Y \geq 1 | Y \leq 3)$.

8.

Y has density function

find the mean and variance of Y.

9.

, Y has distribution function

find the mean and variance of Y.

10.

Weekly CPU time used by an accounting firm has probability density function (measured in hours) given by

a Find the expected value and variance of weekly CPU time.

b The CPU time costs the firm \$200 per hour. Find the expected value and variance of the weekly cost for CPU time.

c Would you expect the weekly cost to exceed \$600 very often? Why?

11.

a within 25 feet of the end of the line?

b within 25 feet of the beginning of the line?

c closer to the beginning of the line than to the end of the line?

12.

The grade point averages (GPAs) of a large population of college students are approximately normally distributed with mean 2.4 and standard deviation .8. What fraction of the students will possess a GPA in excess of 3.0?

13.

The width of bolts of fabric is normally distributed with mean 950 mm (millimeters) and standard deviation 10 mm.

a What is the probability that a randomly chosen bolt has a width of between 947 and 958 mm?

b What is the appropriate value for C such that a randomly chosen bolt has a width less than C with probability .8531?

14.

The magnitude of earthquakes recorded in a region of North America can be modeled as having an exponential distribution with mean 2.4, as measured on the Richter scale. Find the probability that an earthquake striking this region will

a exceed 3.0 on the Richter scale.

b fall between 2.0 and 3.0 on the Richter scale.

15.

If Y has an exponential distribution and $P(Y > 2) = .0821$, what is

a $\beta = E(Y)$?

b $P(Y \leq 1.7)$?

16.

Suppose that a random variable Y has a probability density function given by

Find the value of k that makes $f(y)$ a density function.

Does Y have a χ^2 distribution? If so, how many degrees of freedom?

What are the mean and standard deviation of Y?

17.

If Y has a probability density function given by

obtain $E(Y)$ and $V(Y)$ by inspection.

18.

Identify the distributions of the random variables with the following moment-generating functions:

a $m(t) = (1 - 4t)^{-2}$.

b $m(t) = 1/(1 - 3.2t)$.

c $m(t) = e^{-5t+6t^2}$.

19.

If $\theta_1 < \theta_2$, derive the moment-generating function of a random variable that has a uniform distribution on the interval (θ_1, θ_2).

20.

A random variable Y has the density function

a Find $E(e^{3Y/2})$.

b Find the moment-generating function for Y.

c Find $V(Y)$.

21.

A manufacturer of tires wants to advertise a mileage interval that excludes no more than 10% of the mileage on tires he sells. All he knows is that, for a large number of tires tested, the mean mileage was 25,000 miles, and the standard deviation was 4000 miles. What interval would you suggest?

22.

Let Y have density function

a Find the value of c that makes $f(y)$ a density function.

b Give the mean and variance for Y.

c Give the moment-generating function for Y.

23.

If Z is a standard normal random variable, what is

a $P(Z^2 < 1)$?

b $P(Z^2 < 3.84146)$?

24.

The proportion of time Y that an industrial robot is in operation during a 40-hour week is a random variable with probability density function

a Find $E(Y)$ and $V(Y)$.

b For the robot under study, the profit X for a week is given by $X = 200Y - 60$. Find $E(X)$ and $V(X)$.

c Find an interval in which the profit should lie for at least 75% of the weeks that the robot is in use.

25.

Industrial accidents occur according to a Poisson process with an average of three accidents per month. During the last two months, ten accidents occured. Does this number seem highly improbable if the mean number of accients per month, μ, is still equal to 3? Does it indicate an increase in the mean number of accidents per month?