Statistical Analysis

Master in Statistics and Information Management

Professor *Ana Cristina Costa* (ccosta@novaims.unl.pt)

HOMEWORK

LU2: PROBABILITY DISTRIBUTIONS

I) In the following multiple-choice questions, choose only one option.

1. Which of the following is NOT an assumption of the Binomial distribution?

a) All trials must be independent.

b) Each trial must be classified as a success or a failure.

c) All trials are dependent on each other.

d) The number of successes in the trials is counted.

2. Which one of these variables is a binomial random variable?

a) Time it takes a randomly selected student to complete a multiple choice exam

b) Number of textbooks a randomly selected student bought this term

c) Number of women taller than 68 inches in a random sample of 5 women

d) Number of CDs a randomly selected person owns

3. Find the mean and standard deviation of X~B(n=16, p=1/2):

a) E(X)=16 and sd=2

b) E(X)=8 and sd=2

c) E(X)=8 and sd=1/2

d) E(X)=16 and sd=1/2

4. Which of the following is not a requirement of a binomial distribution?

a) In every trial, the probability of success is the same.

b) The random variable has only two possible outcomes.

c) There are a fixed number of trails.

d) Any outcome is dependent on previous outcomes.

5. Suppose that a quiz consists of 20 True-False questions. A student hasn't studied for the exam and will just randomly guesses at all answers (with True and False equally likely). How would you find the probability that the student will get 8 or fewer answers correct?

a) Find the probability that X=8 in a binomial distribution with n = 20 and p=0.5.

b) Find the area between 0 and 8 in a uniform distribution that goes from 0 to 20.

c) Find the probability that X=8 for a normal distribution with mean of 10 and standard deviation of $\sqrt{5}$

d) Find the cumulative probability for 8 in a binomial distribution with n = 20 and p = 0.5.

6. 5% of workers in the US use public transportation to get to work. You randomly select 250 workers and ask them if they use public transportation to get to work. Find the probability that exactly 16 workers say yes. [*Suggestion: use Excel*]

a) 0.8769

b) 0.0637

c) 0.1923

d) 0

7. The owner of a small convenience store is trying to decide whether to discontinue selling magazines. He suspects that only 5% of the customers buy a magazine and thinks that he might be able to use the display space to sell something more profitable. What is the probability that at least 5 out of a set of 50 customers buy magazines? [*Suggestion: use Excel*]

a) 0.104

b) 0.896

c) 0.066

d) 0.774

8. What should be the value of the parameter *p* of the Binomial distribution with n = 9 so that its standard deviation is equal to its expected value?

a) 0.1

b) 0.2

c) 0.5

d) 0.9

9. It is known that 90% of patients undergoing a given surgery survive. In 10 patients undergoing surgery, how many are expected to survive?

a) none
b) 1
c) 9
d) 10

10. A police department receives an average of 5 requests per hour. The probability of receiving two requests between 10h and 11h is given by the distribution

a) Uniform
b) Binomial
c) Poisson
d) Normal

11. Which of the following is not a condition for the Poisson distribution to be a suitable model?

a) Events occur randomly in space or time
b) Events occur independently
c) Events must occur at a constant rate
d) Events must occur in pairs of two

12. Which of the following is not true about the Poisson distribution?

a) there is only one parameter
b) the Poisson distribution models continuous variables
c) the length of the interval or size of the space affects the mean
d) the mean value can be any positive number

13. The lambda (λ) in the Poisson distribution signifies the

a) standard deviation
b) random variable
c) mean
d) time

14. In Poisson probability distribution, if the value of λ is integer then the distribution will be

a) non-modal
b) unimodal
c) bimodal
d) trimodal

15. The mean and the variance are equal in

a) all probability distributions.
b) the Binomial distribution.
c) the Poisson distribution.
d) the empirical distribution.

16. An average of 4 trains stop at a train station per day. What is the probability that tomorrow 5 trains will stop at the station? [*Suggestion: use Excel*]

a) 0.1563
b) 0.7851
c) 0.2149
d) 0.1755
e) 0.8434

17. If the random variable X has Poisson distribution of variance equal to 0.558, then P (X ≥ 1) is equal to: [*Suggestion: use Excel*]

a) 0.42765
b) 0.28618
c) 0.71544
d) 0.57235
e) 0.32451

18. Suppose the mean score on an exam is 10 with a standard deviation of 5. A score of 30 would have a standardised score (known as z-score) of:

a) 2
b) 3
c) 4
d) 6

19. Let X be a random variable with standard normal distribution. If P(X < z) = 0.1788 then *z* is equal to

[*Suggestion: find the approximate value using a statistical table, and then use Excel to find the exact solution provided here*]

a) −0.92

b) −0.80

c) 0.6

d) 0.80

e) 0.92

20. Let X be a random variable with standard normal distribution. If P(X > z) = 0.9049 then *z* is equal to

[*Suggestion: find the approximate value using a statistical table, and then use Excel to find the exact solution provided here*]

a) −1.31

b) –1.65

c) 0.8289

d) 1.31

e) 1.65

21. Let X be a random variable with standard normal distribution. If P(X > z) = 0.119 then *z* is equal to

[*Suggestion: find the approximate value using a statistical table, and then use Excel to find the exact solution provided here*]

a) −0.55

b) −1.18

c) 0.55

d) 0.883

e) 1.18

22. The exam grades of a set of students were transformed into a Standard Normal distribution. A student has a negative z value, which means that he/she had

a) a negative grade

b) a lower grade than the class average

c) a higher grade than the class average

d) a higher grade than the first quartile of the class grades

23. In the standard normal distribution, the z value corresponding to the 80^{th} percentile is approximately

[*Suggestion: find the approximate value using a statistical table, and then use Excel to find the exact solution provided here*]

a) 0.84

b) 0.2881

c) 0.20

d) 0.80

e) 0.16

24. From previous surveys, it was found that the average time spent by a sales supervisor candidate in a given test is approximately normal with an average value of 60 minutes and a standard deviation of 20 minutes. What percentage of candidates will take less than 60 minutes to complete the test?

a) 10%

b) 20%

c) 50%

d) 60%

e) 80%

25. The IQ values of the individuals of a certain population follow a normal distribution of mean value 100 and standard deviation 16. The probability of an individual having an IQ between 110 and 125 is

[*Suggestion: find the approximate value using a statistical table, and then use Excel to find the exact solution provided here*]

a) 0.2049

b) 0.3251

c) 0.6245

d) 0.7951

e) 0.8820

26. The IQ values of the individuals of a certain population follow a normal distribution of mean value 100 and standard deviation 16. Identify the **false** statement among the following:

a) The probability of an individual having an IQ between 110 and 125 is approximately 0.2049.

b) The 90% percentile of the IQs is 120.5.

c) The 60% mid values of the distribution correspond to IQ values between 86.53 and 113.47.

d) It is not possible to determine the median of the IQ values distribution.

e) The mean IQ of individuals is equal to 100.

27. The height of Portuguese men between the ages of 18 and 70 is a random variable with a normal distribution with a mean value of 1,75 m and a standard deviation of 0,09 m. The 80^{th} percentile of this distribution is approximately equal to

[*Suggestion: find the approximate value using a statistical table, and then use Excel to find the exact solution provided here*]

a) 1,75 m

b) 1,90 m

c) 1,83 m

d) 0,84 m

28. For which of the binomial distributions listed below is the normal distribution not a reasonable approximation?

a) n = 50, p = 0.4

b) n = 40, p = 0.12

c) n = 75, p = 0.11

d) n = 40, p = 0.8

29. Approximately 60% of mathematics students do their homework on time. In a class of 100 students, what is the standard deviation if we assume normality and use the normal distribution as an approximation of the binomial distribution?

a) 4.9

b) 6.0

c) 3.2

d) 24

30. QuenCola, a soft-drink company, knows that it has a 42% market share in one region of the province. QuenCola's marketing department conducts a blind taste test with 100 people at a mall in the region. Use a normal approximation to calculate the probability that fewer than 40 of these people will choose QuenCola.

[*Suggestion: find the approximate value using a statistical table, and then use Excel to find the exact solution provided here*]

a) $P(X \leq 39.5) \cong 0.3062$

b) $P(X \leq 39.5) \cong 0.3427$

c) $P(X \leq 40.5) \cong 0.3806$

d) $P(X \leq 40.5) \cong 0.6194$

II) Solve the following problems.

1. In quality control problems, batches are examined by sampling. Suppose that in a batch of 100 pieces there are 10 defective ones. If you randomly select with replacement 5 pieces, what is the probability that no defective pieces are found in this sample? [*Suggestion: use Excel*]

2. In one region, 20% of farms produce potatoes. A random sample was selected with replacement of 10 farms from that region. Calculate the probability that at least two of the selected farms will be producing potatoes. [*Suggestion: use Excel*]

3. Assuming that the random variable X has a Poisson distribution of mean value 0.558, calculate P (X = 0). [*Suggestion: use Excel*]

4. The hotel facilities of the Good Sleep (GS) chain registered the stay of more than 2 million guests last year. The GS Algarve website, which has an average of approximately seven visits per minute, allows many GS hotel facilities to attract guests. [*Suggestion: use Excel*]

a) Identify the probability distribution of the random variable that represents the number of visits to the website within one minute.

b) Calculate the probability that there will be no website visits within a minute.

c) Calculate the probability of having two or more visits to the website within one minute.

d) Calculate the probability of having one or more visits to the website within 30 seconds.

Suggestion*: solve the following exercises using a statistical table, and then use Excel to find the exact solution provided here*.

5. Let X be a random variable with standard normal distribution. Find:

a) $P(X < 1.2)$

b) $P(X < -1.64)$

c) $P(1.2 < X < 1.33)$

d) $P(-1.7 < X < -1)$

e) $P(X > 1.33)$

f) $P(X > -1)$

g) $P(-1.7 < X < 1.2)$

6. The weight of cement bags distributed by a company follows a normal distribution with mean μ = 20 kg and standard deviation σ = 0.25 kg. Calculate the probability of a randomly chosen cement bag to weight

a) more than 20 kg;

b) less than 20 kg;

c) between 19.8 and 20.2 kg.

7. Let X be a random variable with distribution N(μ, 2).

a) If μ=1, calculate P(X < 1.5).

b) If μ=1, calculate P(0.8 < X < 1.2).

c) Find the value of μ so that P(X < –0.5) = 0.4013.

8. Let X be a random variable with distribution N(–2, 1). Calculate:

a) P(X ≤ 0)

b) P(–2 < X <1)

c) The probability that X does not belong to the interval] –3, –1[.

***Hint**: Compute* P(X ≤ –3) + P(X ≥ –1)*, or alternatively compute* 1–P(–3 < X < –1)*.*

9. Let X be a random variable with distribution N(0, 1).

a) Calculate P(–1.79 < Z < –0.54).

b) Find the value of z so that P(z < Z < 2) = 0.1.

10. Let X_1 be a normal random variable with mean 2 and variance 3, and let X_2 be a normal random variable with mean 1 and variance 4. Assume that X_1 and X_2 are independent. What is the distribution of the linear combination $Y = 2X_1 + 3X_2$?

11. History suggests that scores on the Math portion of the Standard Achievement Test (SAT) are normally distributed with a mean of 529 and a variance of 5732. History also suggests that scores on the Verbal portion of the SAT are normally distributed with a mean of 474 and a variance of 6368. Select two students at random. Let *X* denote the first student's Math score, and let *Y* denote the second student's Verbal score. What is *P*(*X* > *Y*)?

***Hint**: You want to compute* P(X – Y > 0) = P(W > 0)*. Find the distribution of W using the theorem about the linear combination of independent normal variables.*

12. Suppose that the distribution of the weight of a pre-packaged “1-pound bag” of carrots is X~N(1.18, 0.07) and the distribution of the weight of a pre-packaged “3-pound bag” of carrots is Y~N(3.22, 0.09). Selecting

bags at random, find the probability that the sum of three 1-pound bags exceeds the weight of one 3-pound bag.

***Hint**: You want to compute* $P(X_1 + X_2 + X_3 > Y) = P(X_1 + X_2 + X_3 - Y > 0) = P(W > 0)$*, where* $X_i \sim N(1.18, 0.07)$*. Find the distribution of W using the theorem about the linear combination of independent normal variables.*

13. Let $X_1, X_2, \ldots, X_{16}$ be independent random variables with normal distribution N(77, 5) and $\bar{X} = \frac{1}{16}\sum_{i=1}^{16} X_i$. Compute:

a) $P(77 < \bar{X} < 79.5)$.

b) $P(74.2 < \bar{X} < 78.4)$.

***Hint**: First, determine the distribution of the sample mean using the theorem about the linear combination of independent normal variables:* $\bar{X} = \frac{1}{16}X_1 + \frac{1}{16}X_2 + \cdots + \frac{1}{16}X_{16}$.

14. With ingestion of a particular drug for the treatment of depression, 20% of patients report experiencing side effects during the first week of treatment. A random sample of 15 patients was collected. Let X be a random variable that represents the number of patients experiencing such side effects in this sample.

a) Identify the probability distribution of X.

b) On average, how many patients from the sample have experienced side effects?

c) What is the probability that exactly 5 patients from the sample will experience side effects?

d) Calculate the probability that more than 2 patients from the sample will experience side effects.

e) Determine the probability that at least 1 and less than 4 patients from the sample experience side effects.

f) Assuming now that a sample of 100 patients has been collected, calculate the probability of

i. At least 15 patients have experienced side effects.

ii. Between 16 and 45 (inclusive) patients have experienced side effects.

iii. Use the normal approximation to compute these probabilities.

SOLUTIONS

Group I) Multiple choice

1. c
2. c
3. b
4. d
5. d
6. b
7. a
8. a
9. c
10. c
11. d
12. b
13. c
14. c
15. c
16. a
17. a
18. c
19. a
20. a
21. e
22. b
23. a
24. c
25. a
26. d
27. c
28. b
29. a
30. a

Group II) Problems

Question 1) 0.59049

Question 2) 0.62419

Question 3) 0.57235

Question 4
- a) Poisson
- b) 0.0009
- c) 0.9927
- d) 0.9698

Question 5
- a) 0.8849
- b) 0.0505
- c) 0.0233
- d) 0.1141
- e) 0.0918
- f) 0.8413
- g) 0.8403

Question 6
- a) 0.5
- b) 0.5
- c) 0.5762

Question 7
- a) 0.59871
- b) 0.07966
- c) 0

Question 8
- a) 0.9772
- b) 0.4987
- c) 0.3174

Question 8
- a) 0.2579
- b) 1.16

Question 10) $\mathrm{N}(7, \sqrt{48})$

Question 11) 0.6915

Question 12) 0.9830

Question 13
- a) 0.4772
- b) 0.8561

Question 14
- a) X~B(n=15, p=0.2)
- b) 3
- c) 0.1031
- d) 0.602
- e) 0.613
- f)
  - i) Normal approximation = 0.91621; exact value = 0.919556
  - ii) Normal approximation = 0.87076; exact value = 0.871494