BASIC STATISTICS FOR ECONOMISTS, STE101. EXAM SOLUTIONS

Department of statistics
Edgar Bueno and Jonas Bjermo
2023–01–12

Part one. Multiple choice

1. The three approaches for defining a probability are:
   (a) randomization, relative frequency and subjective;
   (b) classical, randomization and subjective;
   (c) **[classical, relative frequency and subjective;]**
   (d) classical, randomization and relative frequency;
   (e) none of the above.

2. Which of the following is **correct** about two events A and B such that $P(A) > 0$ and $P(B) > 0$:
   (a) if A and B are not independent then they are mutually exclusive;
   (b) if A and B are independent then they are mutually exclusive;
   (c) **[if A and B are independent then they are not mutually exclusive;]**
   (d) if A and B are mutually exclusive then they are independent;
   (e) A and B are independent if and only if they are mutually exclusive.

3. Which of the following sentences is **not** correct regarding a continuous random variable X with probability density function —pdf— $f(x)$:
   (a) $f(x)$ can take values larger than one.
   (b) $f(x) \geq 0$ for all values of x;
   (c) $P(a < X < b) = \int_a^b f(x)dx$;
   (d) the area under the curve of $f(x)$ over all values of x must be 1;
   (e) **[$f(x)$ is the probability that the random variable X takes the value x;]**

4. The probability that four students A, B, C and D get a passing grade in an exam of statistics is, respectively, 0.8, 0.7, 0.5 and 0.4. As they do not know each other, their grades can be considered to be independent. What is the probability that all four students pass the exam?
   (a) 0;
   (b) **[0.112;]**
   (c) 0.6;
   (d) 1;
   (e) 2.4.

5. In a card game, the player has three possible outcomes: win, tie or lose. If the player wins (which happens with probability 0.19), he gets two dollars; if the player loses (which happens with probability 0.47), he loses one dollar; in the case of a tie, the player neither wins nor loses any money. What is the variance of the amount of money of the player at the end of one game?

   (a) -0.09;
   (b) 0;
   (c) 1.22;
   (d) 5;
   (e) 5.2;

6. One week before the local elections of a city, a poll is carried out by selecting a random sample of 100 voters. The proportion of individuals in the sample who will vote for Mrs. A is 0.4. A 99% confidence interval for the proportion of individuals who will vote for this candidate in the elections is:

   (a) (0% , 99%);
   (b) (27.4% , 52.6%);
   (c) (30.4% , 49.6%);
   (d) (39.4% , 40.6%);
   (e) (39.5% , 40.5%);

7. One week before the local elections of a city, a candidate, Mrs. B, believes that more than 30% of the voters support her. In order to verify her claim, the campaign has selected a sample of 100 voters. 45 out of the 100 voters in the sample claim that they will vote for Mrs. B. With a significance level of 1%, which of the following is **correct**. (**Hint:** Use the alternative $P > 0.3$):

   (a) the critical value is 2.36 and the test statistic is 3.02, therefore the null hypothesis is rejected;
   (b) the critical value is 2.36 and the test statistic is 3.02, therefore the null hypothesis is not rejected;
   (c) the critical value is 3.02 and the test statistic is 2.36, therefore the null hypothesis is rejected;
   (d) the critical value is 3.02 and the test statistic is 2.36, therefore the null hypothesis is not rejected;
   (e) none of the above.

8. Table 1 summarizes the scores of 170 students in an exam of statistics:

Points	[0 , 40)	[40 , 50)	[50 , 60)	[60 , 70)	[70 , 80)	[80 , 90)	[90 , 100)
Frequency	51	17	22	34	21	17	8

Table 1: Scores of 170 students in an exam of statistics

The teacher of the course wants to test whether the scores in Table 1 can be considered as a random sample from a (truncated) normal distribution. If it was, the probability in each class would be as given in the following table. (**Hint:** This is a goodness of fit test.)

Points	[0 , 40)	[40 , 50)	[50 , 60)	[60 , 70)	[70 , 80)	[80 , 90)	[90 , 100)
Probability	0.33	0.17	0.17	0.14	0.10	0.06	0.03

Having a significance level $\alpha = 0.05$, what is the critical value:

(a) 1.97;

(b) 12.59;

(c) 14.07;

(d) 18.51;

(e) 389.92.

(**Note:** The following information is needed for Exercises 9 and 10.) A researcher has asked the thirteen married men in a small community about the brideprice they had to pay to the bride's family when they got married. The brideprice values (in USD) are

20000 3000 10000 20000 13000 0 31000 20000 63000 8000 3000 12000 4000

9. What is the **median** of the brideprice?

(a) 12000;

(b) 15500;

(c) 15923;

(d) 20000;

(e) 31000.

10. What is the **mode** of the brideprice?

(a) 12000;

(b) 15500;

(c) 15923;

(d) 20000;

(e) 31000.

(**Note:** The following information is needed for Exercises 11 and 12.) The amount of money spent on clothing by students on Stockholm University during 2021 can be modeled by a normal distribution with expected value of 1200 and variance of 40 000. The amount of money spent on course literature can be modeled by a normal distribution with expected value of 800 and variance of 18 000. The covariance between money spent on clothing and course literature is −24 000.

11. What is the probability that one student chosen at random spent more than 2000 on clothing plus course literature?

    (a) 0;
    (b) 0.0967;
    (c) 0.5;
    (d) 0.25;
    (e) 1.

12. A random sample of ten students is selected. What is the probability that the mean expenditure on clothing is between 1150 and 1250?

    (a) 0;
    (b) 0.0010;
    (c) 0.0100;
    (d) 0.1974;
    (e) 0.5708.

Part one. Multiple choice

1. See Section 3.2 in Newbold et. al or in the lecture notes.

2. If A and B are independent, then $P(A)P(B) = P(A \cap B) > 0 = P(\emptyset)$, therefore (c) is correct.

3. See Section 5.1 in Newbold et. al or Section 5 in the lecture notes.

4. Let A, B, C and D be, respectively, the probabilities that students A, B, C and D pass the exam. By independence, we have

$$P(A \cap B \cap C \cap D) = P(A) \cdot P(B) \cdot P(C) \cdot P(D) = 0.8 \cdot 0.7 \cdot 0.5 \cdot 0.4 = 0.112.$$

5. Let X = "Amount of money of the player at the end of one game". We have $P_X(2) = 0.19$, $P_X(-1) = 0.47$ and $P_X(0) = 0.34$. Then

$$\mu_X = \sum_x x P_X(x) = 2 \cdot 0.19 + -1 \cdot 0.47 + 0 \cdot 0.34 = -0.09$$

and

$$\sigma_X^2 = \sum_x x^2 P_X(x) - \mu_X^2 = \left(2^2 \cdot 0.19 + (-1)^2 \cdot 0.47 + 0^2 \cdot 0.34\right) - \left(-0.09^2\right) = 1.22.$$

6. The confidence interval is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.4 \pm 2.5758\sqrt{\frac{0.4(1-0.4)}{100}} = (27.4\% \,, 52.6\%).$$

7. We have
$$\hat{\sigma}^2_{\bar{X}} = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.45(1-0.45)}{100} = 0.002475.$$
Therefore the test statistic is
$$t_{obs} = \frac{\bar{x}_s - \mu_0}{\hat{\sigma}_{\bar{X}}} = \frac{0.45 - 0.3}{0.002475^{0.5}} = 3.02.$$
Regarding the critical value, we have $t_{n-1,\alpha} = t_{99,0.01} = 2.36$.
As $t_{obs} > t_{n-1,\alpha}$ the null hypothesis is rejected.

8. $\chi^2_{K-1,\alpha} = \chi^2_{6,0.05} = 12.59$.

9. $c = (N+1)p/100 = (13+1)50/100 = 7$ then $a = 7$ and $b = 0$. Therefore $\breve{x}_{50} = (1-b)x_{(a)} + bx_{(a+1)} = x_{(7)} = 12000$.

10. The mode is 2000, as this is the most frequently occurring value.

11. Let X = "amount of money spent on clothing by the randomly chosen student", Y = "amount of money spent on course literature by the randomly chosen student" and $Z = X + Y$ "amount of money spent on clothing plus course literature by the randomly chosen student". We know that $X \sim N(1200\,, 40000)$ and $Y \sim N(800\,, 18000)$ and $\sigma_{XY} = -24000$. Then $Z \sim N(\mu_Z\,, \sigma^2_Z)$ with $\mu_Z = \mu_X + \mu_Y = 1200 + 800 = 2000$. Taking into account that the normal distribution is symmetric, we have $P(Z > 2000) = 0.5$.

12. Let X = "amount of money spent on clothing by the randomly chosen student". We know that $X \sim N(1200\,, 40000)$, therefore $\bar{X} \sim N(1200\,, 40000/10)$. Therefore
$$P(1150 < \bar{X} < 1250) = P(-0.79 < Z < 0.79) = 0.78524 - 0.21476 = 0.57048.$$

Part two. Complete solution

13. The owner of an electronic store wants to estimate the expected price paid by customers on mobile phones. In order to do so, she is planning to select a sample of previous sales of mobile phones, looking at the receipts she will collect information on the prices of these phones and, finally, using this information she will compute a confidence interval for the parameter of interest. By experience, she knows that the distribution of prices paid can be adequately modeled by a normal distribution.

    (a) First, she needs to decide the sample size. Assuming a standard deviation of 7000 SEK, what is the smallest sample size needed if she wants to estimate the expectation by a 95% confidence interval with a length no larger than 5000 SEK? (5p.)

    (b) Due to time constrains she could not use the sample size found in (a). Therefore, she decided to select a sample of 16 receipts from previous sales. The prices of the mobile phones (in SEK) are

2000	3000	3000	4000	11000	12000	13000	13000
13000	13000	13000	14000	15000	17000	22000	22000

    Find the sample mean, $\bar{x}$, and the sample variance, S^2_x. (5p.)

    (c) Find a 95% confidence interval for the expected price paid by customers on mobile phones. (5p.)

    (d) If the owner wants to compute a 99% confidence interval, would it be wider or narrower than the one in (c)? (5p.)

Solution

(a). The confidence interval is of the form $\bar{x}_s \pm z_{\alpha/2}\sigma_{\bar{X}}$ with $\sigma_{\bar{X}} = \sigma_X/\sqrt{n}$. Therefore the length is $2z_{\alpha/2}\sigma_{\bar{X}}$ and we have

$$2z_{\alpha/2}\sigma_{\bar{X}} = 2z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}} < 5000 \longrightarrow n > \left(\frac{2z_{\alpha/2}\sigma_X}{5000}\right)^2 = \left(\frac{2 \cdot 1.96 \cdot 7000}{5000}\right)^2 = 30.1 \approx 31.$$

(b).

$$\bar{x} = \frac{1}{n}\sum_s x_i = \frac{1}{16}(2000 + \cdots + 22000) = 11875$$
$$S_x^2 = \frac{1}{n-1}\sum_s (x_i - \bar{x})^2 = \frac{1}{16-1}\left((2000 - 11875)^2 + \cdots + (22000 - 11875)^2\right) = 37\,983\,333.$$

(c). We have $\hat{\sigma}^2_{\bar{X}} = S^2_{x,s}/n = 2\,373\,958$ then

$$\bar{x}_s \pm t_{n-1,\alpha/2}\hat{\sigma}_{\bar{X}} = 11875 \pm 2.131 \cdot 1541 = (8591\,, 15159).$$

(d). Wider.

14. The owner of the electronic store from the previous exercise classifies mobile phones according to its price as follows. If the price is less than 8000 SEK, it is considered as "low price"; if it is between 8000 SEK and 16000 SEK, it is considered as "medium price"; if it is larger than 16000 SEK, it is considered as "high price". She believes that half of the mobile phones sold at her store belong to the "medium price" category, while only one quarter belong to each of the remaining two categories.

    (a) State the hypothesis of interest. (5p.)
    (b) How many phones in the sample belong to each category? (5p.)
    (c) Compute the test statistic and the critical value (using a significance level of 5%). (5p.)
    (d) What is the conclusion regarding the hypothesis? (5p.)

Solution

(a). Let P_l, P_m and P_h be the probabilities of selling a low, medium and high price mobile phone. The hypothesis is

$$H_0 : P_l = 0.25, P_m = 0.5, P_h = 0.25 \qquad \text{vs.} \qquad H_1 : \text{At least one } P_k \neq P_k^0.$$

(b).

Price		
Low	Medium	High
4	9	3

(c). We have $O_l = 4$, $O_m = 9$ and $O_h = 3$ and $E_l = 16 \cdot 0.25 = 4$, $E_m = 16 \cdot 0.5 = 8$ and $E_h = 16 \cdot 0.25 = 4$. Then the test statistic is

$$\chi^2_{obs} = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k} = \frac{(4-4)^2}{4} + \frac{(9-8)^2}{8} + \frac{(3-4)^2}{4} = 0.375.$$

The critical value is

$$\chi^2_{K-1,\alpha} = 5.991.$$

(d). As $\chi^2_{obs} < \chi^2_{K-1,\alpha}$, the null hypothesis is not rejected.