BASIC STATISTICS FOR ECONOMISTS, STE101. REEXAM

Department of statistics
Edgar Bueno
2023–08–15

Time: 14:00 — 19:00
Approved aid: Hand-held calculator with no stored text, data or formulas
Provided aid: Formula Sheet and Probability Distribution Tables, returned after the exam.

Problems 1 — 12: Multiple choice questions (max 60 points):

- A total of 12 multiple choice questions with five alternative answers per question one of which is the correct answer. Mark your answers on the attached **answer form**.
- Marking more than one alternative will result in zero points for that question.
- Each correct answer is worth 5 points.
- Written solutions should not be submitted; only your answers on the answer form will be considered in the assessment and final grading.

Problems 13 — 14: Complete written solutions (max 40 points):

- Use only the provided answer sheets when submitting your solutions and answers.
- For full marks, clear, comprehensive and well-motivated solutions are required. Unclear and unexplained solutions may result in point deductions even if the final answer is correct.
- Check your calculations and solutions before submitting. Careless mistakes may result in unnecessary point deductions.

The maximum total number of points is 60 + 40 = 100. At least 50 points are required to pass (grades A-E). The grading scale is as follows:

Points	0—39	40—49	50—59	60—69	70—79	80—89	90—100
Grade	F	Fx	E	D	C	B	A

NOTE: Fx and F are failing grades that require re-examination. Students who receive the grade Fx or F cannot supplement extra assignments for a higher grade.

Part one. Multiple choice

1. Which of the following is **correct** as an interpretation of the mean, $\bar{x}$:

   (a) It is the most likely value;
   (b) It is the most frequently occurring value;
   (c) It is the center of gravity of the observations;
   (d) It is the middle observation;
   (e) None of the above.

2. A researcher is studying the relation between the continent in which a country is located (Africa, Asia, Europe and Latin America) and its Human Development Index —HDI— (Very high, High, Medium, Low). Using a statistical software, the researcher carries out a test of independence between both variables and obtains a p-value smaller than 0.00001. Considering a significance level $\alpha = 0.01$, which of the following is **correct** regarding the test that has been implemented:

   (a) the continent and the HDI are independent;
   (b) the continent and the HDI are dependent;
   (c) the continent and the HDI are negatively correlated;
   (d) the continent and the HDI are positively correlated;
   (e) the information provided is not enough for making a conclusion.

3. Let $X_1, \cdots, X_n$ be a large random sample from a distribution $f_X(x)$ with expectation μ_X and variance σ_X^2. Let also $\bar{X} = \sum_{i=1}^n X_i/n$ be the sample mean and S_X^2 be the sample variance. Which of the following sentences is **not correct**:

   (a) if $f_X(x)$ is not the normal distribution then $\sqrt{n}(\bar{X} - \mu_X)/\sigma_X$ follows exactly a t distribution;
   (b) if $f_X(x)$ is the normal distribution then $\sqrt{n}(\bar{X} - \mu_X)/\sigma_X$ follows exactly a normal distribution;
   (c) if $f_X(x)$ is the normal distribution then $\sqrt{n}(\bar{X} - \mu_X)/S_X$ follows exactly a t distribution;
   (d) if $f_X(x)$ is not the normal distribution then $\sqrt{n}(\bar{X} - \mu_X)/S_X$ follows approximately a t distribution;
   (e) if $f_X(x)$ is not the normal distribution then $\sqrt{n}(\bar{X} - \mu_X)/S_X$ follows approximately a normal distribution.

4. Let us consider a random experiment with sample space given by the seasons of the year, i.e. $S = \{Spring, Summer, Fall, Winter\}$. Which of the following is a probability on S?

   (a) $P(Spring) = 1/5$; $P(Summer) = 1/5$; $P(Fall) = 1/5$; $P(Winter) = 1/5$;
   (b) $P(Spring) = 1$; $P(Summer) = 3/4$; $P(Fall) = 2/4$; $P(Winter) = 1/4$;
   (c) $P(Spring) = 1/4$; $P(Summer) = 1/4$; $P(Fall) = 1/4$; $P(Winter) = 0$;
   (d) $P(Spring) = 1/4$; $P(Summer) = 2/4$; $P(Fall) = 3/4$; $P(Winter) = 1$;
   (e) None of the above.

5. 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 4000 20000 63000 8000 3000 12000 31000

   What is the **range** of the brideprice?

   (a) -9500;
   (b) 9500;
   (c) 11000;
   (d) 16500;
   (e) 63000.

6. An ice-cream shop offers 10 different flavors. How many combinations of 2 scoops can be made if the order is important and the flavors can be used more than once?

   (a) 20;
   (b) 45;
   (c) 55;
   (d) 90;
   (e) 100.

7. 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 expected amount of money of the player at the end of one game?

   (a) -0.39;
   (b) -0.09;
   (c) 0.00;
   (d) 0.33;
   (e) 1.00.

8. Coffee Inc. is a company that imports two types of coffee to Sweden. The number of sacks of the type *Arabica* imported every month can be described by a normally distributed random variable with expectation 50 and variance 4. The number of sacks of the type *Robusta* imported every month can be described by a normally distributed random variable with expectation 45 and variance 9. The covariance between the number of sacks of both types is 5. What is the probability that the total imports during one month exceed 100 sacks?

   (a) 0.00;
   (b) 0.08;
   (c) 0.15;
   (d) 0.35
   (e) 0.41.

9. A teacher knows by experience that the number of points students get in the final exam can be adequately modeled by a normal distribution. A sample of nine students has been selected and their score in the exam has been measured. The sample mean is $\bar{x}_s = 53.6$ and the sample variance is $S^2_{x,s} = 492.5$. A 90% confidence interval for the expected number of points of the students in the exam is:

(a) $(12.3\,, 94.9)$;
(b) $(39.8\,, 67.4)$;
(c) $(41.4\,, 65.8)$;
(d) $(43.3\,, 63.9)$;
(e) $(44.1\,, 63.1)$.

10. One week before the local elections of a city, a candidate, Mrs. A, 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. 37 out of the 100 voters in the sample claim that they will vote for Mrs. A. The value of the statistic for testing the alternative that the proportion of voters for Mrs. A is larger than 30% is:

(a) 1.45;
(b) 1.66;
(c) 1.98;
(d) 14.50;
(e) 30.03;

11. The teacher of a course in statistics wants to explain the score of students in the final exam (variable *exam*) in terms of the score in a previous home assignment (variable *assignment*) through a linear regression of the form:

$$exam = \beta_0 + \beta_1\, assignment + \epsilon$$

The following table shows the scores of the eight students in the course:

Assignment	42	48	50	50	51	55	59	67
Exam	38	43	57	33	81	50	48	84

The estimated intercept of the regression line of interest is:

(a) -548.7;
(b) - 25.2;
(c) 0.6;
(d) 1.5;
(e) 11.4.

12. The following table 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

The teacher of the course wants to test whether the scores in the previous table 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.

Part two. Complete solution

13. (a) On the first semester of a year, 232 students took the final exam in a course of statistics. 141 students passed the exam. Considering the set of students taking the exam as a random sample, find a 95% confidence interval for the passing rate (i.e. the proportion of students passing the exam). (5p.)

    (b) On the second semester of the year, 261 students took the final exam of the course. 139 students passed the exam. Considering the set of students taking the exam as a random sample, find a 95% confidence interval for the passing rate. (5p.)

    (c) Assuming that the sets of students taking the exam each semester are independent of each other, find a 95% confidence interval for the difference of the passing rate between the first and the second semester of the year. (5p.)

    (d) Assuming that all 493 students taking the exam during the year are independent of each other and that the passing rate is equal to 0.5, find the (approximated) probability that the number of students passing the exam is larger than 280. (**Hint:** What is the distribution of the number of students passing the exam?) (5p.)

14. In order to fit the linear regression that explains "sleeping time" (variable *sleep*, in hours per week) in terms of "working time" (variable *work*, in hours per week) and a dummy variable indicating whether the individual has kids younger than three

$$young\_kid = \begin{cases} 1 & \text{if the individual has kids younger than 3} \\ 0 & \text{otherwise} \end{cases}$$

data on $n = 10$ individuals was collected and it is shown in the following table.

work	young_kid	sleep
34	0	35
29	0	48
43	0	49
41	1	52
32	0	54
29	1	56
44	0	57
16	0	58
59	0	60
6	0	65

The estimated regression is

$$\widehat{sleep} = 57 - 0.12\,work + 1.00\,young\_kid.$$

(a) Interpret the three estimated coefficients. (5p.)

(b) Predict the number of hours sleeping for a person working forty hours per week who has one kid younger than three. (5p.)

(c) Find the ten fitted values and the ten residuals. (5p.)

(d) Calculate the coefficient of determination and the adjusted coefficient of determination. (**Note:** The adjusted coefficient of determination may take an unexpected value.) (5p.)