Stockholm University

# BASIC STATISTICS FOR ECONOMISTS, STE101. EXAM

Department of statistics
Edgar Bueno
2024–01–11

**Time:** 08:00 — 13:00
**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 will result in point deductions even if the final answer is correct.
- Check your calculations and solutions before submitting. Careless mistakes will 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. A salesperson has classified each of her potential customers regarding how likely they are to buy her product. The categories are: high, medium and low. She wants to summarize the data in an adequate chart. Which of the following is a type of chart that is adequate for this situation?

   (a) Bar chart;

   (b) Histogram;

   (c) Box-and-whisker plot;

   (d) Scatter plot;

   (e) Steam-and-leaf display.

2. Which of the following sentences is **not** correct regarding the *cumulative distribution function* of a continuous random variable $X$, $F_X(x)$:

   (a) it is a non-decreasing function;

   (b) it takes values between 0 and 1, i.e. $0 \leq F_X(x) \leq 1$ for all $x$;

   (c) it is a step function;

   (d) $\lim_{x \to -\infty} F_X(x) = 0$;

   (e) $\lim_{x \to \infty} F_X(x) = 1$.

3. In the context of simple linear regression, which of the following is **not** correct?

   (a) the coefficient of determination $R^2$ is equal to the coefficient of correlation between the independent variable $x$ and the dependent variable $y$;

   (b) the coefficient of determination $R^2$ indicates the proportion of variability of the dependent variable $y$ that is explained by the independent variable $x$;

   (c) the intercept $b_0$ indicates the expected value of the dependent variable $y$ when the independent variable $x$ equals zero;

   (d) the slope $b_1$ indicates the expected increment in the dependent variable $y$ associated to a one unit increment in the independent variable $x$;

   (e) the least squares regression is the one that minimizes the *sum of squares error*.

4. 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 positively correlated;

   (b) the continent and the HDI are negatively correlated;

   (c) the continent and the HDI are independent;

   (d) the continent and the HDI are dependent;

   (e) the information provided is not enough for making a conclusion.

5. Consider the experiment of randomly selecting one student from a course of statistics and measuring two variables: $X$ = "number of hours solving exercises during the previous week" and $Y$ = "number of hours solving home assignments during the previous week". The joint probability distribution of $X$ and $Y$ is given in the following table.

| | | $Y$ | | |
|---|---|---|---|---|
| | | 0 | 8 | 16 |
| $X$ | 5 | 0.2 | 0.1 | 0.1 |
| | 8 | 0.1 | 0.1 | 0.1 |
| | 18 | 0.2 | 0.1 | 0 |

Which of the following is **correct**?

(a) $X$ and $Y$ are dependent because their expectations are not equal;

(b) $X$ and $Y$ are dependent because their covariance $\sigma_{XY}$ is not equal to 0;

(c) $X$ and $Y$ are independent because $P(X = 5, Y = 0) = P(X = 5)P(Y = 0)$;

(d) $X$ and $Y$ are independent because one of the joint probabilities is equal to 0;

(e) it is not possible to establish if $X$ and $Y$ are independent with the information provided.

6. 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.

7. A car rental company knows by experience that 10% of the customers rent a *sport utility vehicle* —suv— and that the customers' choice is independent of each other. What is the probability that, out of the next 100 customers, the number of customers renting a suv is larger than eight but at most eleven?

(a) 0.0300;

(b) 0.1322;

(c) 0.3000;

(d) 0.3822;

(e) 0.6178.

8. 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. 45 out of the 100 voters in the sample claim that they will vote for Mrs. A. With a significance level of 1%, which of the following is **correct**. (**Hint:** Use the alternative $P > 0.3$):

   (a) the critical value is 2.33 and the test statistic is 60.61, 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 rejected;

   (c) the critical value is 2.36 and the test statistic is 3.02, therefore the null hypothesis is not rejected;

   (d) the critical value is 3.02 and the test statistic is 2.36, therefore the null hypothesis is rejected;

   (e) the critical value is 3.02 and the test statistic is 2.36, therefore the null hypothesis is not rejected;

9. 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 the candidate of the party B is 0.4. A 99% confidence interval for the proportion of individuals who will vote for this candidate on the elections is:

   (a) $(0\%\,, 99\%)$;

   (b) $(27.4\%\,, 52.6\%)$;

   (c) $(30.4\%\,, 49.6\%)$;

   (d) $(39.4\%\,, 40.6\%)$;

   (e) $(39.5\%\,, 40.5\%)$.

10. The following table shows the scores of the eight students in a course of statistics in the final exam (variable *exam*) and the score in a previous home assignment (variable *assignment*):

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

    Fitting a regression that explains the score in the exam in terms of the score in the assignment yields an intercept $b_0 = -25.2$ and a slope $b_1 = 1.5$. The *sum of squares error* —SSE— is:

   (a) 0;

   (b) 656;

   (c) 1594;

   (d) 1881;

   (e) 2508;

Table 1, which will be used in Exercises 11 and 12, 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

11. What is the approximated mean of the 170 scores given in Table 1?

    (a) 24.3;

    (b) 51.6;

    (c) 52.9;

    (d) 57.8;

    (e) 62.9.

12. 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 |

    What is the value of the test statistic:

    (a) 1.97;

    (b) 12.59;

    (c) 14.07;

    (d) 18.51;

    (e) 389.92.

## Part two. Complete solution

A Swedish sociologist is conducting an experiment about discrimination. She wants to determine if the name of the applicant influences the likelihood of being called to an interview. To this end, she sends out job applications to 100 advertised entry–level jobs in Sweden using fictive Swedish-sounding names (like Karl Andersson). She also sends out job applications to 100 advertised entry–level jobs in Sweden using fictive foreign-sounding first names (like Shady Gamhour). It turns out that 40 out of the 100 applications with Swedish sounding names are called to an interview; whereas 21 out of the 100 applications with foreign sounding names are called to an interview.

13. Considering the samples of applications with Swedish–sounding names and foreign–sounding names as two independent samples from two populations, test the null hypothesis that the proportion of applicants called to an interview is the same in both populations.

    (a) State the hypothesis of interest. (4p.)

    (b) Compute the test statistic and the critical value (using a significance level of 1%). (8p.)

    (c) What is the conclusion regarding the hypothesis? (4p.)

14. Let $X$ = "first name's origin" (Swedish or foreign) and $Y$ = "the applicant is called to an interview" (Yes or No). Test the hypothesis that $X$ and $Y$ are independent.

    (a) Summarize the provided information in a $2 \times 2$ contingency table. (4p.)

    (b) State the hypothesis of interest. (4p.)

    (c) Compute the test statistic and the critical value (using a significance level of 1%). (8p.)

    (d) What is the conclusion regarding the hypothesis? (4p.)

15. Are your conclusions in 13. and 14. consistent with each other? (Yes or No.) Explain why they should be or they should not be consistent. (4p.)