



Università degli Studi di Milano-Bicocca
Foundations of Probability and Statistics

Master degree: Data Science

Name and Surname: _____

Subject: Foundations of Probability and Statistics

Bicocca ID number: _____

Date: 18 January 2022

Max score: 25

Professor: Matteo Borrotti

Time: 80 minutes

Note: The maximum score on the written exam of Foundations of Probability is 25.
There are 12 questions.
The exam lasts for 1 hour and 20 minutes.
The use of a pocket calculator (calcolatrice) is allowed. You cannot use your mobile phone
Please write Name and Surname in all pages.

Question 1 Consider the following distribution. The distribution reports the number of passed exams in the first year of the Data science master degree:

Number of exams	Absolute frequency
0	14
1	41
2	83
3	116
4	56
5	5
Total	315

Please, build the relative cumulative distribution function and the relative plot.

See next page...

Question 2 Please consider the following number of available free beds in seven different hospitals:

212, 320, 180, 410, 374, 215, 538.

Please compute the median.

Question 3 Please consider the following distribution:

Level c_j	Relative frequency
c_1	0
c_2	1
c_3	0
c_4	0
c_5	0
c_6	0

Please, what is the value of the Gini index. Comment the result.

Question 4 If a distribution has Pearson skewness index, γ , equal to 1.3 what can we say about the distribution shape?

See next page...

Domanda 5 If $\text{cov}(x,y) = 1.25$ and if we consider $v = 2.5 + 1 \times x$ and $w = -0.1 + 2 \times y$, $\text{cov}(v,w)$ is equal to:

1. 2.5
2. 4.9
3. none of above

Domanda 6 If we know that $P(A) = \frac{4}{52}$, $P(B) = \frac{12}{52}$ and $P(A \cap B) = \frac{4}{52}$ then $P(A \cup B)$ is:

Domanda 7 In a clinical experimentation, a group of patients is divided in three sub-groups: the first group (A) is composed by 5 patients, the second (B) is composed by 5 patients and the third (C) by 3 patients. For each patient is recorded the reaction time to a particular drug. The results (in minute) are the following (ordered by group): 76, 52, 178, 113, 83, 80, 124, 46, 88, 36, 94, 51, 74. Please, compute the mean for each group and the total variance (hint: recall the concept of total deviance).

See next page...

Domanda 8 We know that the distribution of the sample mean, \bar{y} , is characterized by $E(\bar{y}) = \mu$ and $var(\bar{y}) = ??$. Please, define $var(\bar{y})$.

Domanda 9 Consider X a random variable that describe the weight of a set of coffee packages. From the set, we sample $n = 100$ packages and we obtain that:

$$\sum_{i=1}^{100} x_i = 24800 \quad \sum_{i=1}^{100} (x_i)^2 = 6152900$$

Define and compute the confidence interval of the mean weight with confidence level at 97% ($z_{1-\frac{\alpha}{2}} = 2.17$).

See next page...

Domanda 10 Analyzing a regression model, what does R^2 tell us about the model and how is it defined?

Domanda 11 Please define the ANOVA test.

See next page...

Domanda 12 The following contingency table reports the distribution of $n = 319$ students with respect to two variables: “Type of high school diploma” and “Number of passed exams in the first university year”.

	0-1	2-5	5+
classic diploma	10	67	31
scietific diploma	4	52	36
others	14	65	40

Please, analyze the dipendence (in distribution) between the two variables.

End