
Assignment 2

Introduction to Data Analytics

Prof. Nandan Sudarsanam & Prof. B. Ravindran

1. We use discrete distributions when
 - (a) the outcomes fall in a fixed range
 - (b) the number of possible outcomes is countable
 - (c) the probability values corresponding to the outcomes are discrete
 - (d) the number of iterations of the experiment are fixed

2. Based on a survey, it was found that the probability that a student likes to play football was 0.25 and the probability that a student likes to play cricket is 0.43. It was also found that the probability that a student likes to play both football and cricket is 0.12. What is the probability that a student does not like to play either?
 - (a) 0.32
 - (b) 0.2
 - (c) 0.44
 - (d) 0.56

3. Can the value of a probability density function be greater than one? What about the cumulative distribution function?
 - (a) PDF: yes, CDF: yes
 - (b) PDF: yes, CDF: no
 - (c) PDF: no, CDF: yes
 - (d) PDF: no, CDF: no

4. We are told that for a particular coin, the probability of observing k heads in n tosses is x ($0 \leq k \leq n, 0 < x < 1$) and the probability of observing k heads in $n+1$ tosses is y ($0 < y < 1$). What is the probability, p , of the coin showing up heads in a single toss?
 - (a) $\frac{y \binom{n}{k}}{x \binom{n+1}{k}}$
 - (b) $\frac{x \binom{n+1}{k}}{y \binom{n}{k}}$
 - (c) $\frac{y \binom{n}{k}}{x \binom{n+1}{k}} - 1$

(d) $\frac{x\binom{n+1}{k}-y\binom{n}{k}}{x\binom{n+1}{k}}$

5. The probability that you can hit the bullseye of a dart board in a single throw is 0.6. Given 10 throws, what is the probability of you hitting the bullseye at most 4 times? Also, what is the probability of you hitting the bullseye for the first time on your fourth attempt?
- (a) 0.166, 0.038
(b) 0.054, 0.038
(c) 0.166, 0.064
(d) 0.054, 0.064
6. It is known that 45% of the population support a particular candidate for election. What is the probability that in a poll, more than half of the 600 randomly sampled people express support for that candidate? (Hint: Use the normal approximation to the binomial distribution. In calculating the probability value, first calculate the z-score (https://en.wikipedia.org/wiki/Standard_score) and then use the following z-table: <http://www.stat.ufl.edu/~athienit/Tables/Ztable.pdf>).
- (a) 0.0082
(b) 0.0069
(c) 0.9931
(d) 0.9918
7. It is known that the weight of 1000 g rice packets produced by a certain company follows the normal distribution $\mathcal{N}(1000, 25)$. Given a sample of size 100, what is the distribution of the sample mean?
- (a) $\mathcal{N}(1000, 2.5)$
(b) $\mathcal{N}(1000, 0.5)$
(c) $\mathcal{N}(1000, 0.25)$
(d) $\mathcal{N}(1000, 6.25)$
8. Past experience shows that in a particular class, average scores of students in a particular subject is 69%. After adopting new methods of teaching relying more on multimedia content, we want to test whether the new methods help improve the average scores. Which among the following pair of null and alternate hypotheses do you consider suitable in this scenario?
- (a) null: $\mu = 69$; alternate: $\mu > 69$
(b) null: $\mu \leq 69$; alternate: $\mu > 69$
(c) null: $\mu > 69$; alternate: $\mu \leq 69$
(d) null: $\mu = 69$; alternate: $\mu \neq 69$
9. A soft drinks company wants to introduce a new product. In setting up the manufacturing process it needs to verify that the average amount of sugar in each 500 ml bottle is less than or equal to 55 g. It randomly samples 10 bottles and observes the following values of sugar in the samples: 53.8, 56.1, 54.5, 54.8, 55.2, 55.1, 54.7, 55.8, 55.9, 54.5. Perform the z-test on this data and report the z_{stat} value and p value. Assume known standard deviation of 0.3 g. (Hint: Use the following z-table: <http://www.stat.ufl.edu/~athienit/Tables/Ztable.pdf>).

- (a) $z_{stat} : 0.42$; $p - value : 0.6628$
- (b) $z_{stat} : 0.23$; $p - value : 0.5910$
- (c) $z_{stat} : 0.23$; $p - value : 0.4090$
- (d) $z_{stat} : 0.42$; $p - value : 0.3372$

10. The above soft drinks company wants to ensure that the variance in the amount of sugar in each 500 ml bottle of its new product is within an acceptable limit of 0.3 g. It takes a random sample of 65 bottles from its manufacturing unit and measures the amount of sugar in each bottle. Which hypothesis test should the company use to verify that the variance in the amount of sugar is less than 0.3 g?

- (a) z-test
- (b) t-test
- (c) chi-square test
- (d) proportion z-test