

Sample questions for Decision-Making and Analytical Skills exam

These tasks are sample tasks from the book, but exam tasks will be similar. In case of questions, please check the relevant book chapters first.

The exam will contain some exercises you have to solve using Orange. You should be able to use the widgets we covered on class and interpret the results of the analyses.

1. A construction company has to complete a project no later than three months from now or there will be significant cost overruns. The manager of the construction company believes that there are four possible values for the random variable X , the number of months from now it will take to complete this project: 2, 2.5, 3, and 3.5. The manager currently thinks that the probabilities of these four possibilities are in the ratio 1 to 2 to 4 to 2. That is, $X = 2.5$ is twice as likely as $X = 2$, $X = 3$ is twice as likely as $X = 2.5$, and $X = 3.5$ is half as likely as $X = 3$.
 - a. Find the probability distribution of X .
 - b. What is the probability that this project will be completed in less than three months from now?
 - c. What is the probability that this project will not be completed on time?
 - d. What is the expected completion time (in months) of this project from now?
 - e. How much variability (in months) exists around the expected value you found in part d?
2. Is the number of passengers who show up for a particular commercial airline flight a discrete or a continuous random variable? Is the time between flight arrivals at a major airport a discrete or a continuous random variable? Explain your answers.
3. The probability distribution of the weekly demand for copier paper (in hundreds of reams) used in the duplicating center of a corporation is provided in the file P04_27.xlsx. Assuming that it costs the duplicating center \$5 to purchase a ream of paper, find the mean and standard deviation of the weekly copier paper cost for this corporation.
4. Suppose that you want to estimate the mean monthly food expenditure of all households in your local community. You decide to estimate this population parameter by calling 150 randomly selected residents and asking each individual to report the household's monthly food expenditure. Assume that you use the local phone directory as the frame in selecting the households to be included in your sample. What are some possible sources of error that might arise in your effort to estimate the population mean?
5. Provide an example of when you might want to take a stratified random sample instead of a simple random sample, and explain what the advantages of a stratified sample might be.
6. The file P02_02.xlsx contains data about 211 movies released in 2006 and 2007.
 - a. Recode Distributor so that all distributors except for Paramount Pictures, Buena Vista, Fox Searchlight, Universal, Warner Bros., 20th Century Fox, and Sony Pictures are listed as Other. (Those in Other released fewer than 16 movies.) Similarly, recode Genre so that all genres except for Action, Adventure, Thriller/Suspense, Drama, and Comedy are listed as Other. (Again, those in Other are genres with fewer than 16 movies.)
 - b. Create a crosstabs and an associated column chart for these two recoded variables. Express the counts as percentages so that for any distributor, the percentages add to 100%. Discuss your findings.
7. An automobile manufacturer employs sales representatives who make calls on dealers. The manufacturer wishes to compare the effectiveness of four different call-frequency plans for the sales representatives. Thirty-two representatives are chosen at random from the sales force and randomly assigned to the four call plans (eight per plan). The representatives follow their plans for

6 months, and their sales for the 6-month study period are recorded. These data are listed in the file P19_01.xlsx.

- a. Do the sample data support the hypothesis that at least one of the call plans helps produce a higher average level of sales? Perform an appropriate statistical test and report a p -value.
 - b. If the sample data indicate the existence of mean sales differences across the call plans, which plans produce significantly different average sales levels at the 95% level?
8. The file P08_06.xlsx contains data on repetitive task times for each of two workers. John has been doing this task for months, whereas Fred has just started. Each time listed is the time (in seconds) to perform a routine task on an assembly line. The times shown are in chronological order.
- a. Find a 95% confidence interval for the mean time it takes John to perform the task. Do the same for Fred.
 - b. Do you believe both of the confidence intervals in part a are valid and/or useful? Why or why not? Which of the two workers would you rather have, assuming that task time is the only issue?
9. A manufacturer is interested in determining whether it can claim that the boxes of detergent it sells contain, on average, more than 500 grams of detergent. The firm selects a random sample of 100 boxes and records the amount of detergent (in grams) in each box. The data are provided in the file P09_02.xlsx.
- a. Identify the null and alternative hypotheses for this situation.
 - b. Is there statistical support for the manufacturer's claim? Explain.
10. Suppose a well-known baseball player states that, at this stage of his career, he is a "300 hitter" or better. That is, he claims that he gets a hit in at least 30% of his at-bats. Over the next month of the baseball season, this player has 105 at-bats and gets 33 hits.
- a. Identify the null and alternative hypotheses from the player's point of view.
 - b. Is there enough evidence from this month's data to reject the null hypothesis at the 5% significance level?
 - c. We might raise two issues with this test. First, does the data come from a random sample from some population? Second, what is the relevant population? Discuss these issues. Do you think the test in part b is valid? Is it meaningful?
11. The human resources manager of DataCom, Inc., wants to examine the relationship between annual salaries (Y) and the number of years employees have worked at DataCom (X). These data have been collected for a sample of employees and are given in columns B and C of the file P10_05.xlsx.
- a. Estimate the relationship between Y and X . Interpret the least squares line.
 - b. How well does the estimated simple linear regression equation fit the given data? Provide evidence for your answer.