

STATISTICS WORKSHEET-3

Question No.	Answers
1	B
2	C
3	A
4	A
5	B
6	B
7	B
8	D
9	A

10. What Is Bayes' Theorem?

Ans.

Bayes' theorem is a way to figure out the conditional probability. Conditional probability is the probability of an event happening, given that it has some relationship to one or more other events. For example, your probability of getting a parking space is connected to the time of day you park, where you park, and what conventions are going on at any time. Bayes' theorem is slightly more nuanced. In a nutshell, it gives you the actual probability of an event given information about tests.

"Events" Are different from "tests." For example, there is a test for liver disease, but that's separate from the event of actually having liver disease.

Tests are flawed: just because you have a positive test does not mean you have the disease. Many tests have a high false-positive rate. Rare events tend to have higher false-positive rates than more common events. We're not just talking about medical tests here. For example, spam filtering can have high false-positive rates. Bayes' theorem takes the test results and calculates your real probability that the test has identified the event.

The Formula

Bayes' Theorem (also known as Bayes' rule) is a deceptively simple formula used to calculate conditional probability formal definition for the rule is:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In most cases, you can't just plug numbers into an equation; You have to figure out what your "tests" and "events" are first. For two events, A and B, Bayes' theorem allows you to figure out $p(A|B)$ (the probability that event A happened, given that test B was positive) from $p(B|A)$ (the

probability that test B happened, given that event A happened). It can be a little tricky to wrap your head around as technically you're working backward; you may have to switch your tests and events around, which can get confusing. An example should clarify what I mean by "switch the tests and events around."

11. What is a z-score?

Ans. Z-score also known as the standard score gives us an idea of how far a data point is from the mean. It indicates how many standard deviations an element is from the mean. Hence, Z-Score is measured in terms of standard deviation from the mean. For example, a standard deviation of 2 indicates the value is 2 standard deviations away from the mean. To use a z-score, we need to know the population mean (μ) and also the population standard deviation (σ).

A z-score can be calculated using the following formula:

$$z = (X - \mu) / \sigma$$

where,

z = Z-Score,

X = The value of the element,

μ = The population mean, and

σ = The population standard deviation

Usually, the population means (μ), the population standard deviation (σ), and the observed value (x) are provided in the problem statement, and substituting the same in the above Z-score equation yields the Z-Score value. Depending upon whether the given Z-Score is positive or negative, one makes use of the respective positive Z-Table or negative Z-Table available online or on the back of your statistics textbook in the appendix.

12. What is a t-test?

Ans. The t-test tells you how significant the differences between group means are. It lets you know if those differences in means could have happened by chance. The t-test is usually used when data sets follow a normal distribution but you don't know the population variance.

For example, you might flip a coin 1,000 times and find the number of heads follows a normal distribution for all trials. So you can calculate the sample variance from this data, but the population variance is unknown. Or, a drug company may want to test a new cancer drug to find out if it improves life expectancy. In an experiment, there's always a control group (a group who are given a placebo, or "sugar pill"). So while the control group may show an average life expectancy of +5 years, the group taking the new drug might have a life expectancy of +6 years. It would seem that the drug might work. But it could be due to a fluke. To test this, researchers would use a Student's t-test to find out if the results are repeatable for an entire population.

Q13. What is a percentile?

Ans. A percentile is a comparison score between a particular score and the scores of the rest of a group. It shows the percentage of scores that a particular score surpassed. For example, if you score 75 points on a test, and are ranked in the 85 th percentile, it means that the score of 75 is higher than 85% of the scores.

The percentile rank is calculated using the formula

$$R = (P/100)(N)$$

where P is the desired percentile and N is the number of data points.

Q14. What is ANOVA?

Ans. Analysis of variance, or ANOVA, is a statistical method that separates observed variance data into different components to use for additional tests.

A one-way ANOVA is used for three or more groups of data, to gain information about the relationship between the dependent and independent variables.

The Formula for ANOVA is:

$$F = \text{MSE} / \text{MST}$$

where:

F=ANOVA coefficient

MST=Mean sum of squares due to treatment

MSE=Mean sum of squares due to error

Q15. How can ANOVA help?

Ans. ANOVA helps test three or more variables. It is similar to multiple two-sample t-tests. However, it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups difference by comparing the means of each group and includes spreading out the variance into diverse sources.