Question Type: Statistics

Duration: 20 Minutes
Difficulty: Easy
Domain: Analytics

Problem

Amazon wants to assess the effect of discounts on the number of items shoppers add in their carts. A table summary below shows the mean and variance of cart sizes in the experiment. How would you assess the difference between the two populations - discount vs no discount groups?

Groups	Mean	Variance
No Discount (Control)	5.4	1.2
Discount (Variation)	5.6	3.5

Solution

> How would you assess the difference between the two populations below - discount vs no discount groups?

Groups	Mean	Variance
No Discount (Control)	5.4	1.2
Discount (Variation)	5.6	3.5

[Candidate] Before I propose an analysis, I have questions. Can I assume that the distributions are normal?

[Interviewer] Yes.

[Candidate] Then test I would propose is T-test for two-sample means. The test evaluates whether there is a statistical significance in the difference of two population means.

[Interviewer] Suppose you run the test and the result succeeds. What does this mean?

[Candidate] If the result succeeds, then conclude that, at the significance level 0.05, there is statistical significance in the difference between average number of items in the shopping carts. Discount appears to cause shoppers to add, on average, 3.7% more items compared to shoppers without discounts.

[Interviewer] Suppose that the test fails, what does this mean?

[Candidate] If the test fails, then there is no statistical significance on the effect of discounts on cart sizes.

[Interviewer] Does the failure to reject the null hypothesis mean that null hypothesis is true?

[Candidate] Yes, since p-value is the probability of the observed value or more extreme given that the null hypothesis is true, if the p-value is greater than or equal to the significance level (0.05), then null hypothesis is true.

[Interviewer] Suppose that one of the distributions is normal, but the other one isn't. How would you evaluate the discount effect then?

[Candidate] Hmm, that's a good question. I know that T-tests assume that the distributions are normal. If one of the distributions is not normal, then the statistical power of T-test is diminished. I have two ideas. (1) Apply log-transformation to create normality on both distributions before using a T-test. (2) Or, use a statistical test such as Kruskal-Wallis that does not assume any shape of the distributions.

[Interviewer] Okay, which correction would you use?

[Candidate] I would try both and pick the one with the highest statistical power

[Interviewer] What else do you notice about the table summary?

[Candidate] Well, I see that the variances of the two groups differ. The cart size variance of the discount group is about three times that of the regular group.

[Interviewer] What could be causing the difference in the variances?

[Candidate] It could be the differences in the sample size or discount effect.

[Interviewer] Can you elaborate?

[Candidate] Variance is $\sum_{i=1}^{N}(X_i - M)^2 / N$. As sample size increases, variance decreases; as sample size decreases, variance increases. The experimental group might have less sample size than that of the control group. Or, the other possibility is that the difference in the variances of the item size is due to discount effect.

[Interviewer] Great. How would you evaluate variances using a statistical test?

[Candidate] A statistical test for evaluating variances is the chi-square variance test. The null hypothesis is that the variances of the cart sizes are the same; the alternative is that the variances of the cart sizes differ.

Interviewer Comments

The candidate made a false conclusion that the failure to reject the null hypothesis means that the null hypothesis is true. The hypothesis testing does not assume such as P-value is $P(X >= x| Ho)$, not $P(Ho | X)$. Other than that, the candidate provided sounds responses to the questions asked.

Interviewer Assessment

In the statistics section, a candidate is assessed based on correctness and soundness of statistical methodology, business sense and communication. For each dimension the candidate is rated in the following scale: (5) superior, (4) good, (3) adequate, (2) marginal, (1) not competent.

Assessments	Rating	Comments
Statistical Methodology	4	The candidate offered decent answers to the questions asked. However, there's a slight concern that the candidate may misinterpret fundamentals in statistics as he made a false conclusion on the null hypothesis.
Business Sense	5	The first problem assessed the candidate's business sense. The candidate demonstrated strong business sense as he understood the open-ended business problem and proposed a methodology that can work. He proposed bucketing drivers based on custom indexes to address business question such as which drivers contributed revenue growth on the Facebook games platform.
Communication	5	The candidate's responses were overall solid. He illustrated his points with tables and explained in an easy-to-follow manner.