Statistics Round

Duration: 40 Minutes
Difficulty: Hard
Domains: Analytics

Problem

LinkedIn's marketing team tested two versions of email campaigns, A and B, in San Francisco (SF) and New York City (NYC). Of the 100,000 emails sent, 80% of the emails were version A while the rest were version B. Address the two questions below to guide the marketing team.

#1 - Given that the click-through-rate (CTR) of email A was 15% while that of email B was 30%, which version is more effective?

#2 - In SF, the CTR of email A was 15% while that of email B was 12.5%. In NYC, the CTR of email A was 15% while that of email B was 41.7%. The summary table below shows the number of emails per variation per city. Explain your assessment.

Variations	NYC	SF
A	20,000	60,000
B	12,000	8,000

Solution

> **#1 -** Given that the click-through-rate (CTR) of email A was 15% while that of email B was 30%, which version is more effective?

[Candidate] Before I address your question, I will first organize CTR numbers you presented in a summary table:

Variations	Size	CTR	Count
A	80,000	15%	12,000
B	20,000	30%	6,000

As you mentioned, a total of 100,000 emails were dispatched, 80% being email A while the remaining being email B. Given the 15% CTR for email A and 30% CTR for email B, the counts of conversions were 12,000 and 6,000, A and B respectively.

[Interviewer] Very well. What is your assessment?

[Candidate] Just based on CTR's, email is more effective in conversion. However, the result might be due to chance.

[Interviewer] Can you elaborate?

[Candidate] To assess the CTR difference between the two emails, I would suggest a statistical test.

[Interviewer] Interesting. How would you use a statistical test?

[Candidate] First, I would state the hypothesis:

> **Ho:** The CTR's of emails A and B are the same.
> **Ha:** The CTR's of emails A and B are different.

Second, I would use either a chi-square contingency test or a T-test for two-sample proportions. If p-value of a test is less than the significance level at 0.05, reject the null hypothesis and conclude there is statistical significance in the CTR's of the two emails. Finally, I can conclude that email B brings higher CTR than email A.

[Interviewer] Okay. Do you believe the unequal sample sizes between A and B pose an issue?

[Candidate] Not necessarily. I could foresee how the test's power may decrease as a consequence of the unequal sample size. However, none of the statistical tests I proposed assume that the sample size being equal. Therefore, I do not see an issue.

Interviewer Feedback: In a technical interview, illustrating on a whiteboard or word document demonstrates comprehension of a problem and facilitates brainstorming.

The candidate illustrated her understanding with a summary table. She then assessed the numbers, demonstrating her analytical skills.

She also showed a strong sense in statistical procedure. She was careful not to assess the email conversions without a statistical test. She proposed appropriate statistical tests, followed with logical steps to a conclusion.

> **#2 -** In SF, the CTR of email A was 15% while that of email B was 12.5%. In NYC, the CTR of email A was 15% while that of email B was 41.7%. The summary table below shows the number of emails per variation per city. Explain your assessment.
>
> | Variations | NYC | SF |
> |---|---|---|
> | **A** | 20,000 | 60,000 |
> | **B** | 12,000 | 8,000 |

[Candidate] The change in click-through-rates within each city is a well-known statistical phenomenon called Simpson's Paradox, which states that trend changes or reverses when observations are grouped. In the previous problem, you stated that the CTR of email A was 15% while that of email B was 30%, which suggests that email B produced higher CTR than email A. However, within SF, email A performed better than email B.

[Interviewer] Your observation that the reversal in CTR's is Simpson's Paradox is correct. What else can you tell me about the problem?

[Candidate] As I had done in the previous problem, let me organize the numbers in a summary table, rounded:

	Total		NYC		SF	
Variations	Size	Conversions	Size	Conversions	Size	Conversions
A	80,000 (**80%**)	12,000 (**15%**)	20,000	3,000 (**15%**)	60,000	9000 (**15%**)
B	20,000 (**20%**)	6,000 (**30%**)	12,000	5,000 (**41.7%**)	8,000	1,000 (**12.5%**)

I have generic observations about the table. I see that the distribution of emails are higher for version B than A. I also see that San Francisco received higher allocation of email A than New York while New York received higher allocation of email B than email A.

[Interviewer] Okay, do you see any issues with unequal email allocations between the two cities?

[Candidate] Not at all as the primary metric is proportion, not count.

[Interviewer] Do you have a hypothesis on why CTR's vary?

[Candidate] I presume that the version A email could be a generic email that does not improve CTR in one city over another. Version B might resonate with NYC market; hence, the CTR is at

least 3X higher than that of version A. Perhaps, version B includes jobs in a finance industry that's more prevalent in NYC than SF. I want to emphasize that the observed CTR's occurred due to chance alone.

[Interviewer] What statistical method would you use to validate whether cities affect click-through-rate of emails?

[Candidate] If we disregard the possibility of interaction effect between email variations and cities, then we can use T-test for sample proportions to test whether cities affect email conversions.

First, I would establish the following hypothesis:

$$\textbf{Ho: } CTR_{NYC} = CTR_{SF}$$
$$\textbf{Ha: } CTR_{NYC} \neq CTR_{SF}$$

Next, I would create the following numerical summary pooled across variations by city:

Cities	Calculations	CTR's
NYC	(3,000 + 5,000) / (20,000 + 12,000)	25%
SF	(9,000 + 1,000) / (6,000 + 60,000)	~15%

Next, I would calculate the sample standard error, and calculate the T-statistic:

$$T - Statistic = (CTR_{NYC} - CTR_{SF}) / SE$$

If the corresponding p-value is less than the usual significance level at 0.05, then reject the null hypothesis and conclude that there is statistical significance in the difference between the CTRs in the two cities at the significance level of 0.05.

[Interviewer] Could you have used a one-sample T-test?

[Candidate] Let me think. I believe it is possible to use a one-sample T-test. Instead of comparing the difference of CTR's between NYC and SF, I can compare if the difference of CTR's against 0 such that the hypothesis statements is now the following:

$$\textbf{Ho: } (CTR_{NYC} - CTR_{SF}) = 0$$
$$\textbf{Ha: } (CTR_{NYC} - CTR_{SF}) \neq 0$$

Based on this change, the T-statistic is calculated as follows:

$$T - Statistic = [(CTR_{NYC} - CTR_{SF}) - 0] / SE$$

[Interviewer] Suppose that we do care about the interaction effect between city and variation. How would you evaluate it?

[Candidate] As I think about this problem more, I realize that I could have applied ANOVA with two main effects and one-interaction effect. The model could assess three null hypothesis at the same time:

Hypothesis 1 - There is no difference in CTRs between NYC and SF.
Hypothesis 2 - There is no difference in CTRs between variations A and B.
Hypothesis 3 - There is an interaction effect on CTRs between cities and variations.

The ANOVA test would provde statistical significance of each of the three terms measured. If there is an interaction effect between the city and email variation, then the p-value of the interaction term will be less than the significance level.

[Interviewer] Okay, what is your final recommendation to the marketing team upon learning that there is an interaction effect?

[Candidate] I would explain that the statistical test result suggests that there is an effect on CTR given the variation across cities and emails. I would then advise that the study is expanded on more markets to assess this business hypothesis further. In addition, I would suggest that more variations of the email are tested to assess the following assumption - emails tailored to a city's demography performs better than a generic version across markets.

Interviewer Feedback: The candidate began her analysis with a numerical summary, which shows that she is quite comfortable with analysis of data - a key attribute in data science.

She then successfully spotted a statistical phenomenon called Simpson's Paradox, which many candidates often failed to identify. This indicates that she grasps fundamentals in statistics.

However, I found a few errors in her methodology. When asked how she would assess the CTRs using statistical test, her approach was roundabout. She proposed using T-Test, only measuring the difference between the two cities, neglecting the possibility of interaction effect to measure. She also made a false assessment that one-way T-test is possible, which is not mathematically possible given the sample-sizes of the groups are different.

Nonetheless, she recovered with a better model design, which involved using ANOVA with two main terms + interaction term.

Finally, she concluded with a sound recommendation to the marketing team.

Final 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	In both problems, the candidate's statistical know-how was fairly strong. In problem #1, she devised a statistical framework that addressed the interview question. Her approach of utilizing T-test was appropriate. In problem #2, most of her responses were solid except on a follow-up question involving one-sample T-test. Given that sample-size difference between the two groups, mathematically, the statistical test is not possible.
Product Sense	5	In both questions, she ensured that her responses are grounded in the marketing problem. She devised a numerical summary, making sense of the CTR metrics. This allowed her to suggests approaches that aligned with the problem. Lastly, in her last problem, when asked about her recommendation, she offered sound suggestions. The idea of testing email variations on more than two emails makes sense to assess whether emails focused on target market perform better than a general one.
Communication	5	The candidate ensured that she understood the problems, illustrating a numerical summary and explaining her analysis clearly. Her explanation of statistical were easy to follow and comprehensive. Lastly, she clearly explained recommendations to a marketing team, suggesting that she possesses fluidity in stakeholder engagement.