Nils Bakers Case

Tangina Parvez

Master's in Business Analytics

Hult International Business school

1.**Exploring similarities and differences between regression and t-test approaches to hypothesis testing.**

To compare two means, we use the t-test. On the explanatory variable, it calculates the contrast between two categories. Regression calculates the relation between the change in the explanatory variable and the change in the response variable. When the answer variable is regressed against a binary explanatory variable, regression analysis transforms into a t-test. To look at the variations in average household data versus footprint, t-tests and regression analysis have both been done for the Nils Baker case. Since the inside/outside footprint is the explanatory variable, the Nils Baker data is appropriate for a t-test study rejecting the null hypothesis.

t-Test: Two-Sample Assuming Unequal Variances

	Ratio	
Mean	0.014233357	0.441667
Variance	0.000182357	0.248669
Observations	120	120
Hypothesized Mea	0	
df	119	
t Stat	-9.386173357	
P(T<=t) one-tail	2.66318E-16	
t Critical one-tail	1.657759285	
P(T<=t) two-tail	5.32637E-16	
t Critical two-tail	1.980099876	

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.152333083
R Square	0.023205368
Adjusted R Square	0.014927448
Standard Error	0.013402812
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000503569	0.000504	2.803284697	0.096719974
Residual	118	0.021196972	0.00018		
Total	119	0.021700541			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.012411395	0.001637414	7.579876	8.62354E-12	0.009168869	0.01565392	0.009168869	0.01565392
X Variable 1	0.004125198	0.002463833	1.674301	0.096719974	-0.000753861	0.009004257	-0.000753861	0.009004257

2. **Transforming variables as a way of improving the regression model.**

Transformations are used to achieve specific goals, such as ensuring linearity, achieving normalcy, or stabilizing the variance. To get the variables in value, first, transform the footprint as inside = 1 and outside = 0. Also, generate the ratio of Total Households in Area and Households with Account that became Y/dependent value for the analysis.

3.**Distinguishing between correlation and causality in the context of the regression model.**

Correlation describes an association between variables. When one variable changes, so do the other. However, a direct or indirect causal relationship is not always to blame for this covariation. There is a cause-and-effect link between the variables, which is what the concept of causation means. There is a strong a correlation between Total Households in Area and Households with Account which does not have a signification impact on Inside/Outside Footprint to have impactful causation.

	Ratio	*Numbers*
Ratio	1	
Numbers	0.152333	1

4. **Comment on "Is the presence of a physical bank branch creating the demand for checking accounts?"**

This case's outcome demonstrates the lack of significance of the dummy variable, Inside/Outside Footprint, which denotes the existence of a real bank branch. When footprints are shifted from inside to outside, the ratio of households with accounts to all households in an area falls by 0.0041 units while the number of households with accounts rises by 1437 units. A physical bank branch's existence does not greatly impact the demand for checking accounts, according to the data set presented.