NILS BAKER'S CASE REFLECTION

Business Statistics – DAT-8565 – FMBAN1

SREEKAR BATHULA

HULT INTERNATIONAL BUSINESS SCHOOL

Nils Baker's Case Study Solutions

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

The t-test is used when comparing the average variable based on two different categories. Regression analysis helps in predicting a response variable based on an explanatory variable. Regression analysis becomes a t-test when the response variable is regressed against a binary explanatory variable. The Nils Baker data qualifies for t-test analysis since the inside/outside footprint is the explanatory variable. Three t-tests and three regressions have been conducted to investigate the differences in average household data against footprint. The models comparing total households, households with accounts, and the quotient have the same t-test value, for instance, p is 0.00 and alpha=0.05 for households with accounts/total households in the area. Hence p-value is less than the alpha value we reject the hypothesis.

t-Test: Two-Sample Assuming Unequal Variances

	Ratio	
Mean	0.014233	0.441667
Variance	0.000182	0.248669
Observations	120	120
Hypothesized Mea	0	
df	119	
t Stat	-9.38617	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.657759	
P(T<=t) two-tail	5.33E-16	
t Critical two-tail	1.9801	

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

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

The variable transformation aims at improving the model's predictive ability. The first model, households with accounts against footprint shows that the intercept is insignificant (t stat =-2.50, p=0.006) while the slope is significant for the ratio of households with accounts to total households (t stat =1.54, p=0.064) in modeling the demand for checking accounts. Transforming the demand for checking accounts into demand for checking accounts per total household makes the intercept significant (t stat =-3.73, p=0.0001. In modeling the demand and reduces the r-squared to 0.023. The variable transformation was not significant in this case.

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

Correlation measures the strength of association between selected variables. There is a weakly strong positive relationship between households with accounts and footprints in the selected area, r=0.15. Causation describes how much the changes in one variable affect the other variable. In this case, changing the footprint state from 1(Inside) to 2(outside) increases the households with accounts by 1437 units.

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

Changing footprints from inside to outside results in a 0.0041 unit decrease in the Household with account/total household in area and increases the households with accounts by 1437 units. The first model is insignificant (F =2.8033, p=0.096), while the second model is significant (F = 5.8389, p=0.0172) in modeling the response variables. It is worth concluding that the presence of a physical bank branch creates demand for checking accounts.

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.000503569	0.000504	2.803285	0.096719974
Residual	118	0.021196972	0.00018		
Total	119	0.021700541			

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	61156083.83	61156084	5.838872	0.017208262
Residual	118	1235926749	10473956		
Total	119	1297082833			