Jaden Obi — March 24, 2024

Predictive Analytics & Business Forecasting

REGRESSION #1:

Equation: UNTITLED Workfile: ANDYS BURGER HEAVEN_MIDTERM_C

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALES
Method: Least Squares
Date: 03/24/24 Time: 11:49
Sample: 1 75
Included observations: 75

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	121.9002	6.526291	18.67832	0.0000
PRICE	-7.829074	1.142865	-6.850394	0.0000

R-squared	0.391301	Mean dependent var	77.37467
Adjusted R-squared	0.382963	S.D. dependent var	6.488537
S.E. of regression	5.096858	Akaike info criterion	6.121430
Sum squared resid	1896.391	Schwarz criterion	6.183230
Log likelihood	-227.5536	Hannan-Quinn criter.	6.146106
F-statistic	46.92790	Durbin-Watson stat	2.224741
Prob(F-statistic)	0.000000		

REGRESSION #2

Equation: UNTITLED Workfile: ANDYS BURGER HEAVEN_MIDTERM_C

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALES
Method: Least Squares
Date: 03/24/24 Time: 11:53
Sample: 1 75
Included observations: 75

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	118.9136	6.351638	18.72172	0.0000
PRICE	-7.907854	1.095993	-7.215241	0.0000
ADVERT	1.862584	0.683195	2.726283	0.0080

R-squared	0.448258	Mean dependent var	77.37467
Adjusted R-squared	0.432932	S.D. dependent var	6.488537
S.E. of regression	4.886124	Akaike info criterion	6.049854
Sum squared resid	1718.943	Schwarz criterion	6.142553
Log likelihood	-223.8695	Hannan-Quinn criter.	6.086868
F-statistic	29.24786	Durbin-Watson stat	2.183037
Prob(F-statistic)	0.000000		

REGRESSION #3:

Equation: UNTITLED   Workfile: ANDYS BURGER HEAVEN_MIDTERM_C

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: SALES
Method: Least Squares
Date: 03/24/24   Time: 11:48
Sample: 1 75
Included observations: 75

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	109.7190	6.799045	16.13742	0.0000
PRICE	-7.640000	1.045939	-7.304442	0.0000
ADVERT	12.15124	3.556164	3.416950	0.0011
ADVERT_SQRD	-2.767963	0.940624	-2.942688	0.0044

R-squared	0.508235	Mean dependent var	77.37467
Adjusted R-squared	0.487456	S.D. dependent var	6.488537
S.E. of regression	4.645283	Akaike info criterion	5.961440
Sum squared resid	1532.084	Schwarz criterion	6.085039
Log likelihood	-219.5540	Hannan-Quinn criter.	6.010792
F-statistic	24.45932	Durbin-Watson stat	2.043061
Prob(F-statistic)	0.000000		

When we compare the different regressions, we can see that the adjusted R-squared value increases as we add each additional independent variable. In regression #1, it was 0.382963. In regression #2, the R-squared value increased to 0.432932. In regression #3, the R-squared value increased again to 0.487456. This increase in value tells us that the additional independent variables are justified – they improve how on how much the variance in the data can be explained.

Looking at the Akaike info criterion, we can still support the claim that the addition of independent variables is justified. This statistical measure considers both the goodness of fit of the model and the complexity, while also taking overfitting into account. When we compare the AIC values, we can see that it decreases each time we add another independent variable. This means it improves the goodness of fit without overly increasing complexity.

HYPOTHESIS TESTING:

Regression #1: *Assuming $\alpha = 0.05$*

H_0: ß$_1 = 0$
H_a: ß$_1 \neq 0$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.62554053							
R Square	0.391300955							
Adjusted R Square	0.382962612							
Standard Error	5.096857529							
Observations	75							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	1219.09103	1219.09103	46.92790295	1.97078E-09			
Residual	73	1896.390837	25.97795667					
Total	74	3115.481867						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	121.9001736	6.526290698	18.67832421	1.5876E-29	108.893295	134.9070522	108.893295	134.9070522
price	-7.829073515	1.142864644	-6.850394365	1.97078E-09	-10.10679947	-5.551347565	-10.10679947	-5.551347565

Significance F-value = 1.97078E-09 < 0.05, therefore **reject H_0**

P-value (price) = 1.97078E-09 < 0.05, therefore **reject H_0**

Regression #2: *Assuming $\alpha = 0.05$*

H_0: $ß_1 = ß_2 = 0$
H_a: At least one beta-coefficient $\neq 0$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.669520547							
R Square	0.448257762							
Adjusted R Square	0.432931589							
Standard Error	4.886123971							
Observations	75							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	1396.53893	698.2694649	29.24785948	5.04086E-10			
Residual	72	1718.942937	23.87420746					
Total	74	3115.481867						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	118.9136104	6.351637506	18.72172495	2.21429E-29	106.2518524	131.5753684	106.2518524	131.5753684
price	-7.907854327	1.095993022	-7.215241491	4.424E-10	-10.09267649	-5.723032168	-10.09267649	-5.723032168
advert	1.862584271	0.683195474	2.726283095	0.008038182	0.500658985	3.224509557	0.500658985	3.224509557

Significance F-value = 5.04086E-10 < 0.05, therefore **reject H_0**

P-value (price) = 4.424E-10 < 0.05, therefore **reject H_0**

P-value (advert) = 0.008038182 < 0.05, therefore **reject H_0**

Regression #3: *Assuming $\alpha = 0.05$*

H_0: $ß_1 = ß_2 = ß_3 = 0$
H_a: At least one beta-coefficient $\neq 0$

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.712906133							
R Square	0.508235155							
Adjusted R Square	0.487456358							
Standard Error	4.645283021							
Observations	75							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	1583.397408	527.799136	24.45931648	5.59996E-11			
Residual	71	1532.084459	21.57865435					
Total	74	3115.481867						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	109.719036	6.799045455	16.13741763	1.87037E-25	96.16212439	123.2759476	96.16212439	123.2759476
price	-7.640000035	1.045938884	-7.304442117	3.23648E-10	-9.725542935	-5.554457134	-9.725542935	-5.554457134
advert	12.15123567	3.556163941	3.416950364	0.001051598	5.060446168	19.24202518	5.060446168	19.24202518
advert_sqrd	-2.767963089	0.940624031	-2.942688043	0.004392666	-4.643514137	-0.892412041	-4.643514137	-0.892412041

Significance F-value = 5.59996E-11 < 0.05, therefore **reject H_0**

P-value (price) = 3.23648E-10 < 0.05, therefore **reject H_0**

P-value (advert) = 0.001051598 < 0.05, therefore **reject H_0**

P-value (advert_sqrd) = 0.004392666 < 0.05, therefore **reject H_0**

In regression #1, the R-squared value was 0.391301. This means that approximately 39.13% of the variation in sales can be explained by a change in price. In regression #2, the R-squared value increased to 0.448258. Using this model, approximately 44.82% of the variation in sales can be explained by a change in both price and money put into advertising. In regression #3, the R-squared value increased again to 0.508235. Using this model, we can see that approximately 50.82% of the variation in the data can be explained by a change in price, money spent for advertising, and the squared value of advertising.

Do the signs of your regression coefficients make sense for each model?

Yes, we expected the coefficient for price to be negative. When the product becomes more expensive, there would be less people willing to purchase it. We expected the coefficient for advertising to be positive. Spreading awareness of the product would result in more purchases to be made. We expected the coefficient for advertising squared to be negative due to diminishing returns.

Interpret the regression slopes for price and advertising in model #2

For every $1 that the price increases, you can expect to lose about $7.91 in sales.
For every $1,000 spent on advertising, you can expect to generate about $1,862 in new sales.

What was the rationale for including advertising squared in model #3?

In the second model, we failed to realize that after a while, advertising can only do so much to positively affect monthly sales. Advertising squared is included in the third model to account for diminishing returns. Because the rate at which sales increasing due to advertisements is decreasing, we expected the value of this slope to be negative.

CONFIDENCE INTERVALS:

Regression #1:

Equation: UNTITLED Workfile: ANDYS BURGER HEAVEN_MIDTERM_OBI::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Coefficient Confidence Intervals
Date: 03/24/24 Time: 19:10
Sample: 1 75
Included observations: 75

		90% CI		95% CI		99% CI	
Variable	Coefficient	Low	High	Low	High	Low	High
C	121.9002	111.0274	132.7729	108.8933	134.9071	104.6390	139.1614
PRICE	-7.829074	-9.733082	-5.925065	-10.10680	-5.551348	-10.85180	-4.806346

Regression #2:

Equation: UNTITLED Workfile: ANDYS BURGER HEAVEN_MIDTERM_OBI::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Coefficient Confidence Intervals
Date: 03/24/24 Time: 19:12
Sample: 1 75
Included observations: 75

		90% CI		95% CI		99% CI	
Variable	Coefficient	Low	High	Low	High	Low	High
C	118.9136	108.3299	129.4973	106.2519	131.5754	102.1081	135.7191
PRICE	-7.907854	-9.734101	-6.081608	-10.09268	-5.723032	-10.80769	-5.008019
ADVERT	1.862584	0.724180	3.000989	0.500659	3.224510	0.054950	3.670218

We are 95% confident that for every $1 increase in price, monthly sales will decrease between $10.09 and $5.72.

We are 95% confident that for every $1,000 increase in advertisement spending, sales will increase between approximately $500.65 and $3,224.51.

Using the 95% confidence interval, we reject the null hypothesis for both independent variables. Zero doesn't fall within either of the confidence intervals given.

Regression #3:

Equation: UNTITLED Workfile: ANDYS BURGER HEAVEN_MIDTERM_OBI::Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Coefficient Confidence Intervals
Date: 03/24/24 Time: 19:13
Sample: 1 75
Included observations: 75

		90% CI		95% CI		99% CI	
Variable	Coefficient	Low	High	Low	High	Low	High
C	109.7190	98.38775	121.0503	96.16212	123.2759	91.72289	127.7152
PRICE	-7.640000	-9.383161	-5.896839	-9.725543	-5.554457	-10.40846	-4.871543
ADVERT	12.15124	6.224534	18.07794	5.060446	19.24203	2.738555	21.56392
ADVERT_SQRD	-2.767963	-4.335607	-1.200319	-4.643514	-0.892412	-5.257666	-0.278260