DATA ANALYTICS

Unit 3: Regression Model for forecasting

Jyothi R.
Department of Computer Science and Engineering

CONTENTS

- Forecasting is one of the most important and frequently addressed problems in analytics.
- Inaccurate forecasting can have significant impact on both top line and bottom line of an organization.
- For example, non-availability of product in the market can result in customer dissatisfaction, whereas, too much inventory can erode the organization's profit.
- Thus, it becomes necessary to forecast the demand for a product and service as accurately as possible.
- Every organization prepares long-range and short-range planning for the organization and forecasting demand for product and service is an important input for both long-range and short-range planning

REGRESSION MODEL FOR FORECASTING

- Regression is probably more appropriate method for forecasting when the data has values of predictor (explanatory) variables in addition to the dependent variable *Yt*.
- In the data provided in Table 1, we also have information such as promotion expenses and whether the competition was on promotion or not.
- Using the values of these predictor variables is likely to give better forecast than the exponential smoothing techniques discussed in the previous sections.
- Parker and Segura (1971) claimed that regression method can predict more accurately than less scientific methods such as exponential smoothing.

REGRESSION MODEL FOR FORECASTING

- The forecasted value at time *t*, *Ft*, can be written as a regression equation

$$F_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_n X_{nt} + \epsilon_t$$

  - Here F_t is the forecasted value of *Yt*, and *X*1*t*, *X*2*t*, etc. are the predictor variables measured at time *t*.
  - The regression equation for Example 13.1 is $F_t = \beta_0 + \beta_1$
  - Here F_t is the forecasted value of Y_t, and X_{1t}, X_{2t}, etc. are the predictor variables measured at time *t*.
  - For the data in Table 1, the regression outputs using SPSS are shown in Tables 1 and 2.
  - The model is developed based on first 36 months data.

REGRESSION MODEL FOR FORECASTING

Model	*R*	*R*-Square	Adjusted *R*-Square	Std. Error of the Estimate	Durbin–Watson
1	0.928	0.862	0.853	207017.359	1.608

- TABLE 1: Model Summary
- The *R*-square for the model is 0.862 (note that we will need high *R*-square value for forecasting applications) and the Durbin–Watson statistic value is 1.608.
- Since this is a time-series data we need to check whether the errors, e_t,, are correlated (auto-correlation).
- For n = 36 (sample size) and number of predictor variables = 2, the Durbin–Watson critical values are , D_L = 1.153 and , D_U = 1.376.

REGRESSION MODEL FOR FORECASTING

- Since the model Durbin–Watson statistic
- D = 1.608 (4 – D = 2.392) lies within d_U and (4 – dU), we can conclude
- that there is no auto-correlation.
- Whenever regression model is used for forecasting, it should be checked for autocorrelation among the errors.
- Presence of auto-correlation may lead to inclusion of a non-significant variable in the model since the standard error of the regression coefficient is underestimated when autocorrelation of errors is present

REGRESSION MODEL FOR FORECASTING

- TABLE 2: Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	808471.843	278944.970		2.898	0.007
	Promotion Expenses	22432.941	1953.674	0.825	11.482	0.000
	Competition Promotion	−212646.036	77012.289	−0.198	−2.761	0.009

REGRESSION MODEL FOR FORECASTING

- The regression model (based on values in Table 13.12) is given by Eqn 1.

$$F_t = 808471.843 + 22432.941X_{1t} - 212646.036X_{2t}$$

- Where

$$X_{1t} = \text{Promotion expenses at time } t$$

$$X_{2t} = \begin{cases} 1 & \text{Competition is on promotion} \\ 0 & \text{Otherwise} \end{cases}$$

- As expected, the sales increases as the promotion expenses increase and the sales decreases whenever the competition is on promotion.
- The forecasted values for period 37 to 48 using the regression model [Eq.1] is shown in Table 3.

REGRESSION MODEL FOR FORECASTING

- TABLE 3:TABLE 3: Forecasts using regression model

Period	Y_t	X_{1t}	X_{2t}	F_t	$(Y_t - F_t)^2$	$\frac{\lvert Y_t - F_t \rvert}{Y_t}$
37	3216483	121	1	3310211.67	8785063205	0.02914
38	3453239	128	0	3679888.29	5.137E+10	0.065634
39	5431651	170	0	4622071.81	6.5542E+11	0.149048
40	4241851	160	0	4397742.4	2.4302E+10	0.036751
41	3909887	151	1	3983199.9	5374781013	0.018751
42	3216438	120	1	3287778.73	5089499329	0.02218
43	4222005	152	0	4218278.88	13884007.5	0.000883
44	3621034	125	0	3612589.47	71310120.7	0.002332
45	5162201	170	0	4622071.81	2.9174E+11	0.104632
46	4627177	160	0	4397742.4	5.264E+10	0.049584
47	4623945	168	0	4577205.93	2184540571	0.010108
48	4599368	166	0	4532340.05	4492746215	0.014573

REGRESSION MODEL FOR FORECASTING

- The RMSE and MAPE based on regression model are 302969 and 0.0419 (or 4.19%), respectively.
- The RMSE and MAPE for regression based forecasting are much smaller than the values that we obtained so far using moving average and exponential smoothing techniques.
- For Example 13.1, the moving average method resulted in an RMSE of 734725.84 and MAPE is 14.03%.
- The RMSE and MAPE for single exponential smoothing are 742339.22 and 13.94%, respectively

REGRESSION MODEL FOR FORECASTING

- The plot of actual demand and forecasted demand using regression model is shown in Figure 1.

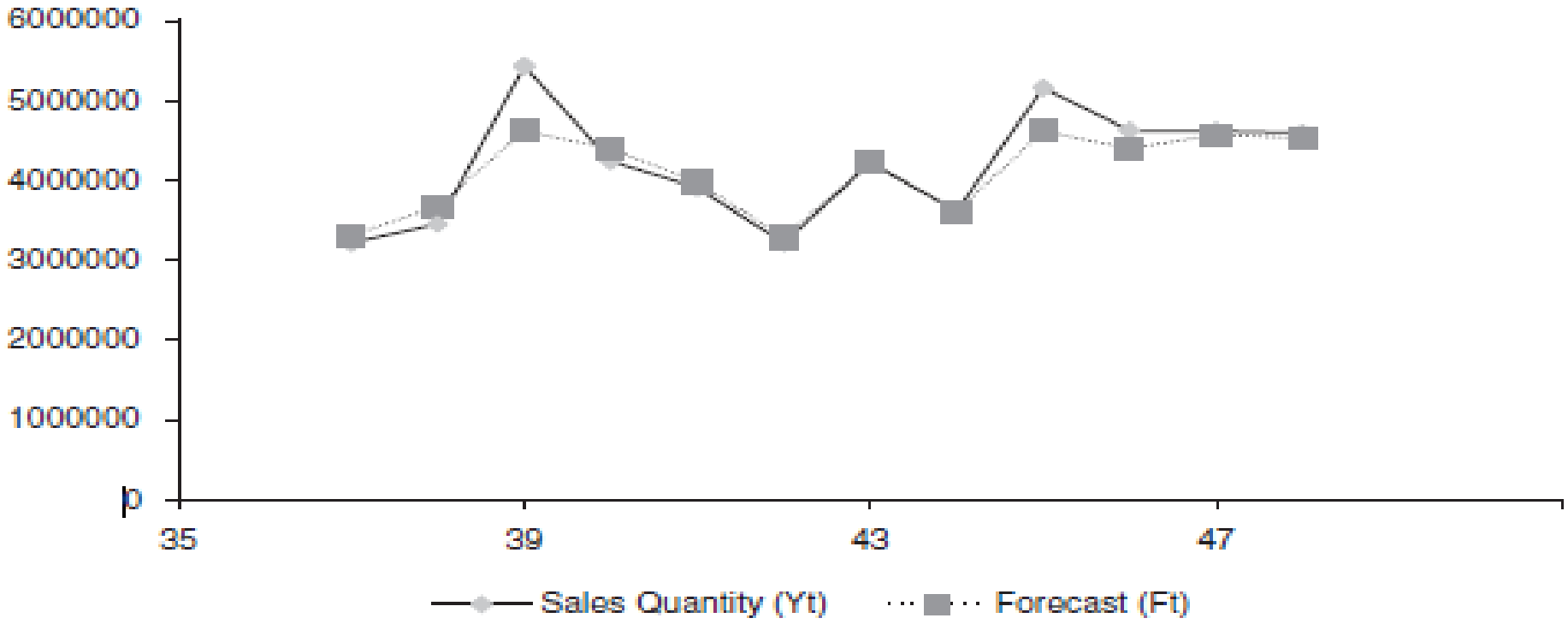

Figure 1: Actual sales quantity and forecasted sales using regression model.

DATA ANALYTICS

REGRESSION MODEL FOR FORECASTING

- STEP 1: Estimate the seasonality index (using techniques such as method of averages or ratio to moving average).
- STEP 2: De-seasonalize the data using either additive or multiplicative model. For example, in multiplicative model, the de-seasonalized data
- $Y_{d,t} = Y_t / S_t$,
- where $Y_{d,t}$ is the de-seasonalized data and S_t is the seasonality index for period *t.*
- STEP 3: Develop a forecasting model on the de-seasonalized data ($F_{d,t}$).
- STEP 4: The forecast for period $t + 1$ is $F_{t+1} = F_{d,t+1} \times S_{t+1}$.

REGRESSION MODEL FOR FORECASTING Example

- Hiccup Viking (HV) is The Vice President of Viking Cookies that specialized into chocolate chip cookies (Choco-Chip).
- Viking Cookies believes that demand for cookies is seasonal and is driven by several factors such as school holidays, festivals, etc.
- The shelf life of Choco-Chip cookies is 6 months and excess inventory and running out of stock can have financial impact.
- Hiccup would like to develop a forecasting model that they can use for forecasting the demand.
- The past monthly demand (quantity of 200 gram packets) for four years (January 2013 to December 2016) along with average price per unit during that month is shown in Table 4.
- Develop a forecasting model using regression to predict demand between months 37
- and 48, given that the data is seasonal.

DATA ANALYTICS

REGRESSION MODEL FOR FORECASTING Example

- TABLE 4: Monthly demand (quantity of 200 gram packets) along with average price per unit.

Period	Month	Demand in Units	Average Price	Period	Demand in Units	Average Price
1	January	10500472	37	25	10658309	36
2	February	10123572	34	26	8677622	38
3	March	7372141	36	27	7330354	37
4	April	7764303	38	28	8115471	37
5	May	6904463	40	29	8481936	34
6	June	10068862	34	30	8778999	37
7	July	6436190	40	31	10145039	32
8	August	9898436	34	32	8497839	38
9	September	6803825	39	33	8792138	34
10	October	8333787	36	34	8485358	36
11	November	7541964	39	35	8575904	36
12	December	8540662	37	36	9885156	32
13	January	10229437	37	37	11023467	35
14	February	8453201	38	38	7942451	40
15	March	7997459	35	39	12492798	32
16	April	8557825	35	40	9756258	32
17	May	7818397	36	41	8992741	32
18	June	8944499	37	42	7397807	40
19	July	8904086	36	43	9710611	32
20	August	8463682	39	44	8328379	39
21	September	7723957	37	45	11873063	32
22	October	7731422	39	46	10642507	32
23	November	8441834	35	47	10635075	32
24	December	7485122	40	48	10578547	32

REGRESSION MODEL FOR FORECASTING Example

- **Solution:**
- Since the demand is seasonal, the first step in forecasting is to estimate the seasonality index.
- We can use first 36 months data to estimate the seasonality index using
- method of averages explained in Section 13.7.1. Table 13.15 gives the seasonality
- index for various months. For example, the seasonality index for January is 1.2251.
- That is, in January the demand will increase by 22.51% from the trend.

DATA ANALYTICS

REGRESSION MODEL FOR FORECASTING Example

- TABLE 5 Seasonality index for various months

Month	Demand (2012)	Demand (2013)	Demand (2014)	Average	Seasonality Index
1	10500472	10229437	10658309	10462739	1.2251
2	10123572	8453201	8677622	9084798	1.0637
3	7372141	7997459	7330354	7566651	0.8860
4	7764303	8557825	8115471	8145866	0.9538
5	6904463	7818397	8481936	7734932	0.9057
6	10068862	8944499	8778999	9264120	1.0847
7	6436190	8904086	10145039	8495105	0.9947
8	9898436	8463682	8497839	8953319	1.0483
9	6803825	7723957	8792138	7773307	0.9102
10	8333787	7731422	8485358	8183522	0.9582
11	7541964	8441834	8575904	8186567	0.9585
12	8540662	7485122	9885156	8636980	1.0113
Average of monthly averages				8540659	

REGRESSION MODEL FOR FORECASTING Example

- De-seasonalized data is calculated by dividing the value of Y_t with the corresponding seasonality index. The de-seasonalized data for periods 1 to 48 is shown in Table 6:
- TABLE 6 De-seasonalized demand seasonality index is rounded to 2 decimals

Month	Demand	Seasonality Index	De-seasonalized Demand	Month	Demand	Seasonality Index	De-seasonalized Demand
1	10500472	1.23	8571459.88	25	10658309	1.23	8700301.09
2	10123572	1.06	9517214.68	26	8677622	1.06	8157870.71
3	7372141	0.89	8321110.54	27	7330354	0.89	8273944.56
4	7764303	0.95	8140603.02	28	8115471	0.95	8508790.52
5	6904463	0.91	7623682.26	29	8481936	0.91	9365476.36
6	10068862	1.08	9282556.42	30	8778999	1.08	8093422.43

REGRESSION MODEL FOR FORECASTING Example

- Regression output for the de-seasonalized demand and average price using
- Microsoft Excel are shown in Table 7:
- TABLE 7 Regression output using SPSS for data in Table 6 (based on first 36 cases)

Model		Unstandardized Coefficients		T	Sig.
		B	Std. Error		
1	(Constant)	20812014.673	717702.417	28.998	0.000
	Average Price	−335945.859	19616.915	−17.125	0.000

REGRESSION MODEL FOR FORECASTING Example

- Regression model for demand forecasting based on first 36 months of de-seasonalized data is given by
- $F_{d,t}$ = 20812014.673 – 335945.859 × Average Price
- The forecasted values are given in Table 8.
- TABLE 8 Forecasted values for the data in Table 5

REGRESSION MODEL FOR FORECASTING Example

- TABLE 8 Forecasted values for the data in Table 5

Month	Demand	Seasonality Index (S_t)	De-seasonalized Demand	$F_{d,t}$	$F_t = F_{d,t} * S_t$	$(Y_t - F_t)^2$	$\lvert Y_t - F_t \rvert / Y_t$
37	11023467	1.2251	8998377	9053910	11091497	4628131462	0.006171
38	7942451	1.0637	7466733	7374180	7844001	9692313467	0.012395
39	12492798	0.8860	14100918	10061747	8914269	1.2806×10^{13}	0.286447
40	9756258	0.9538	10229099	10061747	9596642	2.5477×10^{10}	0.01636
41	8992741	0.9057	9929491	10061747	9112521	1.4347×10^{10}	0.01332
42	7397807	1.0847	6820092	7374180	7998831	3.6123×10^{11}	0.081244
43	9710611	0.9947	9762683	10061747	10008080	8.8488×10^{10}	0.030633
44	8328379	1.0483	7944523	7710126	8082657	6.0379×10^{10}	0.029504
45	11873063	0.9102	13045128	10061747	9157730	7.373×10^{12}	0.228697
46	10642507	0.9582	11106956	10061747	9641005	1.003×10^{12}	0.094104
47	10635075	0.9585	11095071	10061747	9644592	9.8106×10^{11}	0.093134
48	10578547	1.0113	10460573	10061747	10175223	1.6267×10^{11}	0.038127

RMSE and MAPE values are 1381119.09 and 0.0775 (7.75%), respectively.

References

Text Book:

"Business Analytics, The Science of Data-Driven Making", U. Dinesh Kumar, Wiley 2017

Chapter-13 Concept of stationarity, DF and ADF test and transforming non stationary process to a stationary one 13.9.1-13.9.1 in text

Image Courtesy

https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Time+Series+Analysis:+The+Basics

THANK YOU

Jyothi R

Assistant Professor, Department of Computer Science

jyothir@pes.edu