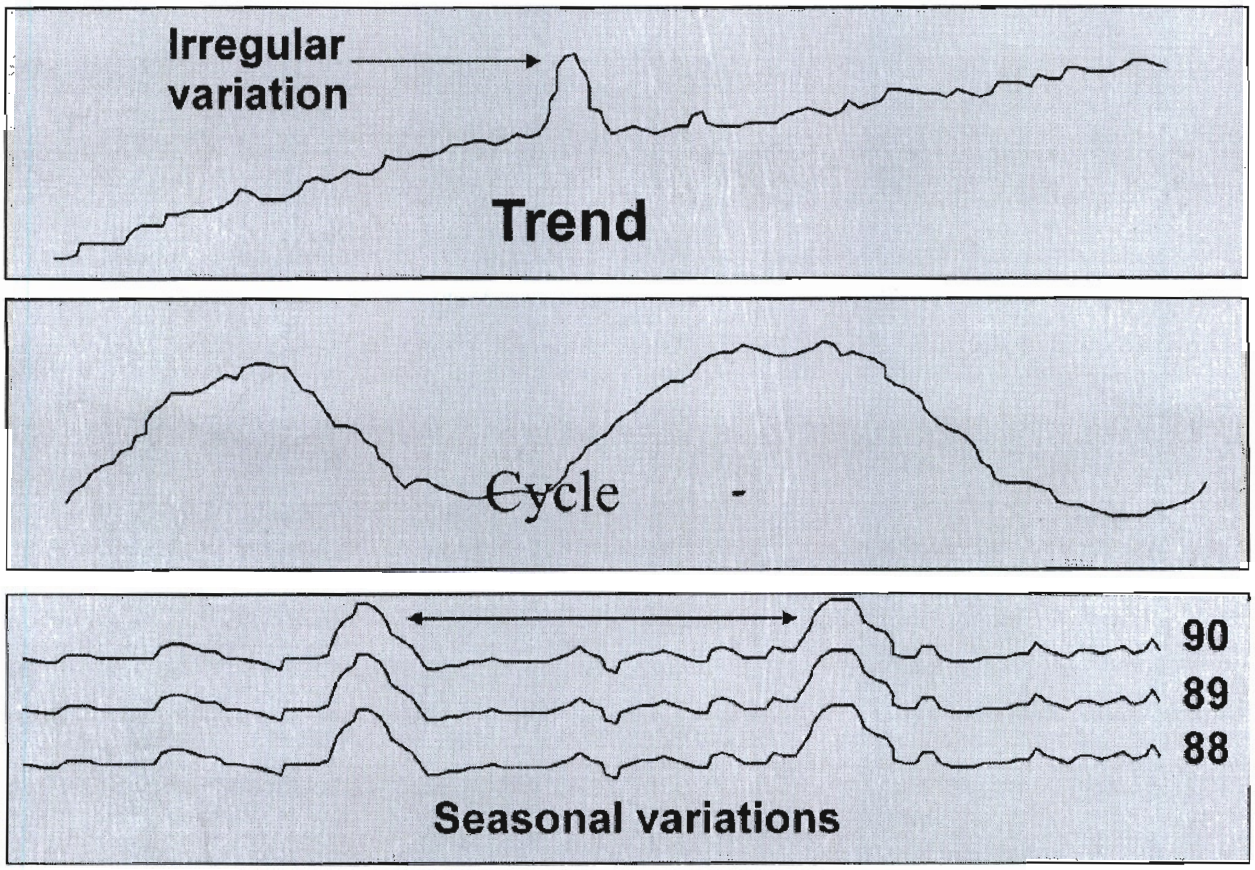
Forecasting

1. Assumes causal system: past ==> future. The future is going to resemble that of the past
2. Forecasts rarely perfect because of randomness, What that means "Forecast is always wrong, with 50% over forecast and 50% under forecast" ?
3. Forecasts more accurate for groups (product families, i.e. passenger cars) vs. individuals (i.e. Toyota Camry)
4. Forecast accuracy decreases as time horizon increases

Time Series Forecasts

1. Trend - long-term movement in data
2. Seasonality - short-term regular variations in data
3. Irregular variations - caused by unusual circumstances
4. Random variations - caused by chance



Seasonal Variations

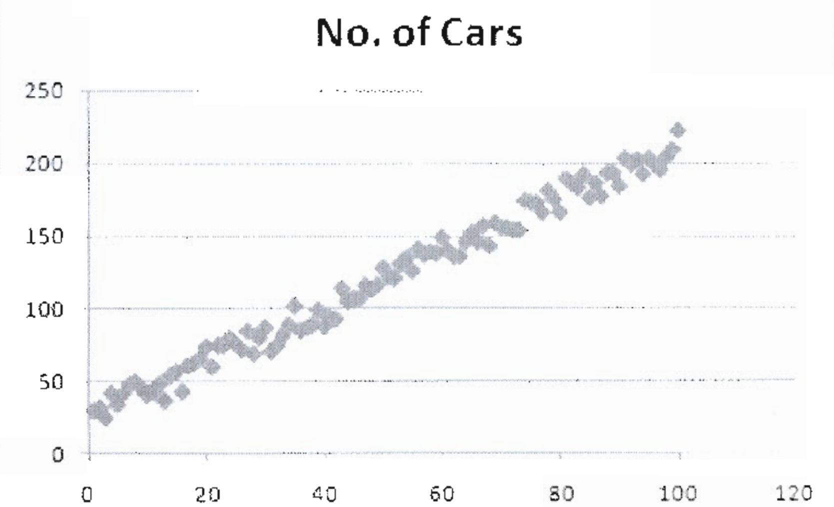
- Regular repeating movements in time series values that can be tied to recurring events
- Annual variations: weather, summer/winter sports equipment
- Vacations/holidays: airline travel, greeting cards, resort
- Daily, Weekly, Monthly: rush traffic hours, theaters and restaurants, banks, mail volume, sales of toys, beer, automobiles, turkeys, highway usage, hotel registrations, gardening, public transportations, electric power plants

Regression 1 Problem from Test 1, Fall 2008

1. Table 2 below shows the Summary Output for using regression analysis to forecast weekly number of cars sold in the last 100 weeks. Answer the following questions based on the information provided in Table 2.

Table 2. Simple Linear Regression Analysis for Car Sales

	A	B
1	SUMMARY OUTPUT	
2		
3	*Regression Statistics*	
4	Multiple R	0.9897
5	R Square	0.9795
6	Adjusted R Square	0.9793
7	Standard Error	7.8728
8	Observations	100
9		
10	ANOVA	
11		*df*
12	Regression	1
13	Residual	98
14	Total	99
15		



	A	B	C	D	E	F	G
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
17	Intercept	25.9636	1.5864	16.3660	0.0287	22.8154	29.1119
18	Week	1.8671	0.0273	68.4571	0.0147	1.8129	1.9212
19							

	A	B	C
20	*Week*	*No. of Cars*	*Predicted No. of Cars*
21	*1*	*30*	...
22	...	...	...
23	...	...	...
24	99	209	...
25	100	222	...
26	*101*		
27			

(1) (10 pts.) What is the general linear model to be used to model the linear trend for the number of cars sold overtime?

(2) Answer the following two questions:

a. (2 pts) What is the estimated value of b_0 .

b. (4 pts) How do you interpret the meaning of b_0?

(3) You are asked to predict the weekly number of cars to be sold for the week 101:

a. (6 pts.) What is the Excel@ formula of the estimated regression function in Cell C26 of Table 2?

b. (3 pts.) What is the predicted number of cars to be sold in Week 101? (show details)

(4) With Table 2, use the *p*-value approach to test the population parameter β_0, and state your conclusion. Assume a significant level of α value of 5%.

a. (3 pts.) What are the null (H_0) and alternative (H_1) hypothesis?

b. (5 points) What are the decision rules?

c. (3 pts.) What is your conclusion and what that means to your car sales forecasting?

(5) In terms of the prediction confidence interval:

a. (3 pts.) What is the margin of error for an approximated 95% prediction interval of the number of cars to be sold in Week 101?

b. (5 pts.) What is the approximated 95% prediction interval of number of cars to be sold in Week 101?

See file Regressionlecture.xls

Regression / TIF08

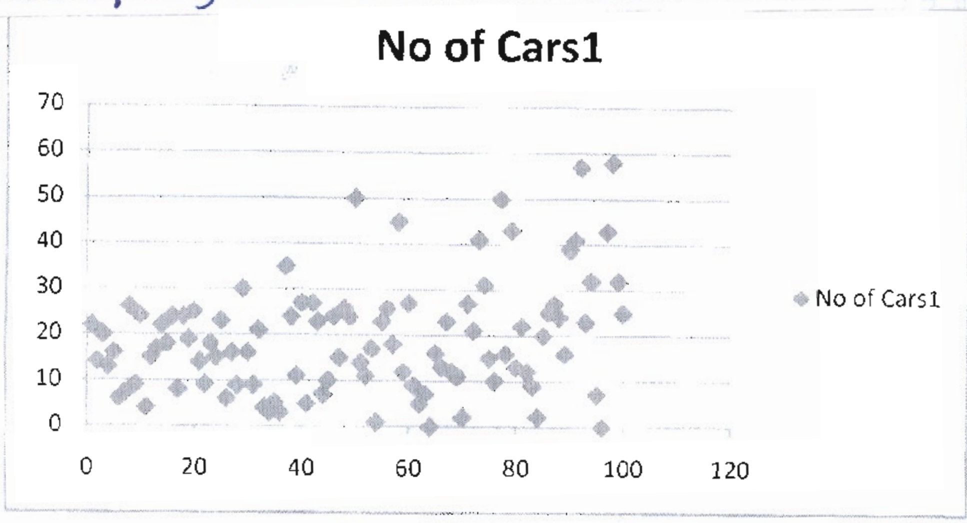
(Y) No of Cars Sold versus Time (Week) (T) and Interest Rate (R)

Dependent Variable — independent Variables

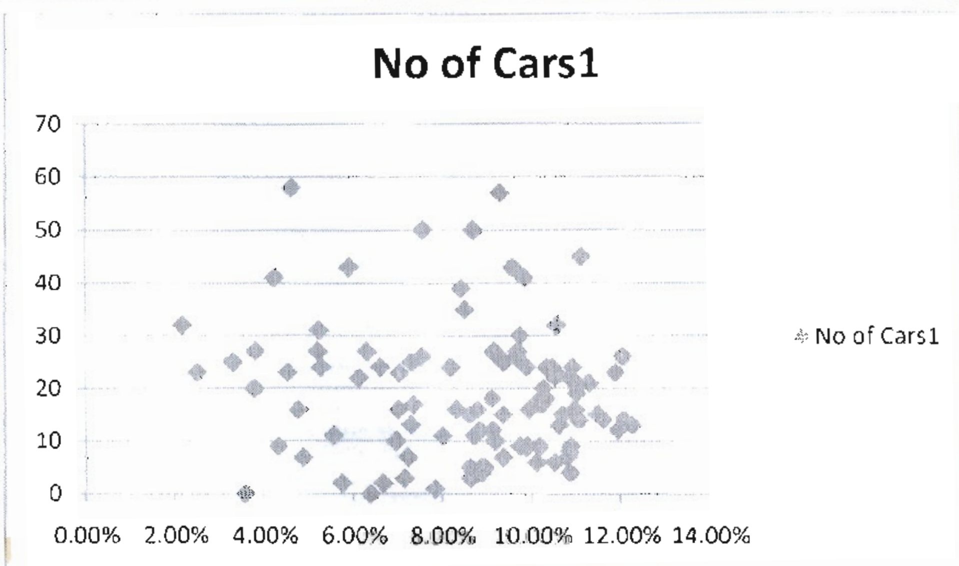
(Cars) $Y = f(T, R) = f(\text{Week}, \text{Rate})$

Scatter plot of cars vs. Week & cars vs. Rate

Cars $Y = f(\text{Week } T)$



Cars $Y = f(\text{Rate } R)$



Study the scatter plots: Y_i (cars) $= f$ (T_i week)

- stationary vs. non stationary
  - upper trend

- Fit a straight line to the data

- Identify the unique Least Sum of Squared Error (LSE) line

Values of b_0, b_1, b_2,

a. Use Excel@ Solver to MIN SSE (=SUMXMY2(,))

b. Use Excel@ Data/Data Analysis/Regression to get Summary Output

c. Use Add Trendline/Display Equation & R-Square on scatter plot

d. Use Excel@: =INTERCEPT( ), =Slope( ).

For Y_i (cars) $= f$ (T_i Week)

- any two points connect a straight line

- Y intercept (b_0) and Run and Rise (b_1)

- The LSE Least Sum of Squared Error Line

Equations for b_0 and b_1

$$b_1 = \frac{SSXY}{SSX} = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(X-\bar{X})^2} = \frac{\Sigma(XY) - \frac{(\Sigma X)(\Sigma Y)}{n}}{\Sigma(X^2) - \frac{(\Sigma X)^2}{n}}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

General linear Regression Model

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 R_i + \varepsilon_i$$

- Y_i: dependent Var: No. of cars sold for population data
- ε_i: Errorterm
- R_i: Independent Var: interest rate
- β_2: Slope for interest rate: marginal No. of cars sold when Interest rate changes ±1%.
- T_i: Independent Var: Time in Week
- β_1: Slope for Time in Week: marginal No. of cars sold when Time in week changes ±1 Week.
- β_0: Y intercept: No. of cars sold when $T_i = \phi$ and $R_i = \phi$%.

Estimated Linear Regression Equation

$$\hat{Y}_i(\text{Cars}) = b_0 + b_1 T_i(\text{week}) + b_2 R_i(\text{Rate})$$
$$= 2.44 + 0.18 T_i + 89.16 R_i$$

- $\hat{Y}_i$: dependent var. No. of Cars Sold for Sample data
- b_2: slope for Interest Rate: Each 1% change in Interest rate leads to 89.16 change in No. of Cars Sold
- b_1: slope for Time in Week: marginal change in No. of in Week leads to 0.18 change in No. of Cars Sold
- b_0: Y intercept: 2.44 Cars Sold when $T(\text{week}) = \phi$ and $R_i(\text{interest rate}) = \phi$.

Fcst in Week $T_{10} = 10$ and Interest Rate $R_{10} = 10.46\%$

$$\hat{Y}_{10}(\text{Cars}) = 2.44 + 0.18 * 10 + 89.16 * .1046 = 13.56$$

The actual $Y_{10}(\text{cars}) = 24$

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	*Regression Statistics*						
4	Multiple R	0.326					
5	R Square	0.106					
6	Adjusted R Square	0.088					
7	Standard Error	11.922					
8	Observations	100					
9							
10	ANOVA						
11		*df*	*SS*	*MS*	*F*	*Significance F*	
12	Regression	2	1640.974228	820.4871139	5.772305219	0.004279847	
13	Residual	97	13787.77577	142.1420183			
14	Total	99	15428.75				
15							
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
17	Intercept	2.439	8.063	0.302	0.763	-13.564	18.443
18	Week	0.180	0.057	3.190	0.002	0.068	0.293
19	Interest Rate	89.162	67.394	1.323	0.189	-44.597	222.922
20							
21							
22	SUMMARY OUTPUT						
23							
24	*Regression Statistics*						
25	Multiple R	0.857388378					
26	R Square	0.73511483					
27	Adjusted R Square	0.722207839					
28	Standard Error	11.86694183					
29	Observations	100					
30							
31	ANOVA						
32		*df*	*SS*	*MS*	*F*	*Significance F*	
33	Regression	2	38300.21778	19150.10889	135.9858189	7.23218E-29	
34	Residual	98	13800.78222	140.8243083			
35	Total	100	52101				
36							
37		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
38	Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
39	Week	0.195	0.030	6.392	5.56393E-09	0.134	0.255
40	Interest Rate	108.623	19.987	5.435	4.01264E-07	68 960	148.286

Learn to Understand the Regression Summary Output

$R^2 = 0.106 = 10.6\%$ or 10.6% of total variation in car sales canbe explained by the regression model.

and $1 - R^2 = 1 - .106 = .894 = 89.4\%$ of total Variation is car sales cannot be explained by the regression model.

A better Regression Model is the one with Significant large portion of total variation explained by the model or $R^2 \uparrow$

How to Select Independent Variables to be included in the final Model?

Test $\beta_0 = \phi$

Hypothesis: $H_0: \beta_0 = \phi$

$H_1: \beta_0 \neq \phi$

Decision Rule: If p-value $\geq \alpha = .05$, then conclude $H_0: \beta_0 = \phi$.

otherwise if p-value $< \alpha = .05$, then conclude $H_1: \beta_0 \neq \phi$.

in the table, p-value for $(b_0 = 0.7629) > (\alpha = .05)$, conclude $H_0: \beta_0 = \phi$.

i.e. The model should drop b_0. $\Rightarrow Y_i = \beta_1 T_i + \beta_2 R_i$

$= 0.195 T_i + 108.62 R_i$

Use Excele to get Regression Summary Output, confirm

Sequence of Hypothesis tests: start with the highest p-value in Regression Summary Output.

Forecast No. of Cars Sold with Final Model when $T = 10$, $R = 10.46\%$

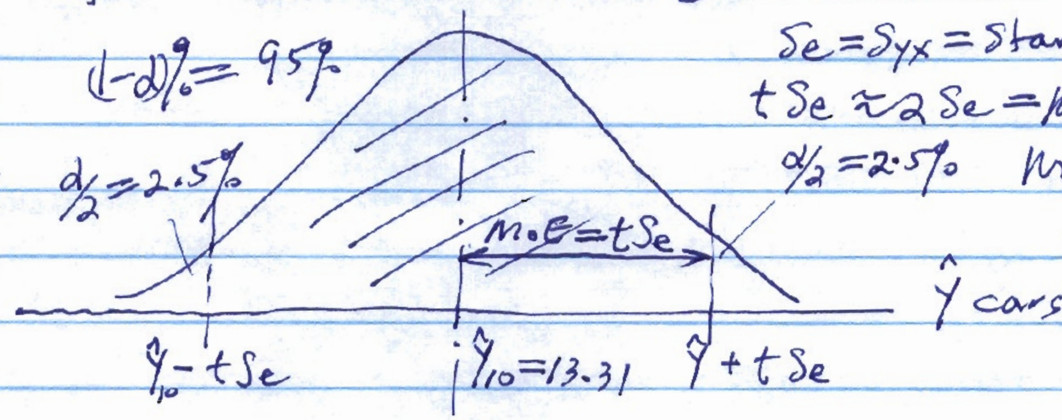
$\hat{Y}_{10}(cars) = 0.195 * 10 + 108.62 * .1046 = 13.31$ cars

Use Confidence Interval to assess the goodness of the model:

$S_e = S_{yx}$ = Standard Error

$tS_e \approx 2S_e$ = M.E or Marginal of Error

S_e = RMSE



Lower Limit is $\hat{Y}_{10} - tS_e = 13.31$ Cars $- 2 * 11.87 = \phi$

Upper Limit is $\hat{Y}_{10} + tS_e = 13.31$ Cars $+ 2 * 11.87 = 37.04$

Meaning: