

12.19). Below is the MINITAB generated output for the given problem is given.

The regression equation is

$$y = -103 + 0.605 x_1 + 8.92 x_2 + 1.44 x_3 + 0.014 x_4$$

Predictor Coef SE Coef T P

Constant -102.7 207.9 -0.49 0.636

x1 0.6054 0.3689 1.64 0.145

x2 8.924 5.301 1.68 0.136

x3 1.437 2.392 0.60 0.567

x4 0.0136 0.7338 0.02 0.986

S = 15.5793 R-Sq = 74.5% R-Sq(adj) = 59.9%

Analysis of Variance

Source DF SS MS F P

Regression 4 4957.2 1239.3 5.11 0.030

Residual Error 7 1699.0 242.7

Total 11 6656.3

12.43). The full model for the given problem is given below. Here we have used MINITAB to fit the required model.

The regression equation is

$$y = 351 - 1.27 \text{ Oil Viscosity}(x_1) - 0.154 \text{ Load}(x_2)$$

Predictor Coef SE Coef T P

Constant 350.99 74.75 4.70 0.018

Oil Viscosity(x1) -1.272 1.169 -1.09 0.356

Load(x2) -0.15390 0.08953 -1.72 0.184

S = 25.4979 R-Sq = 86.2% R-Sq(adj) = 77.0%

Analysis of Variance

Source DF SS MS F P

Regression 2 12161.6 6080.8 9.35 0.051

Residual Error 3 1950.4 650.1

Total 5 14112.0

The restricted model for the given problem is given below. In this model we have restricted oil viscosity. The MINITAB output is as follows:

The regression equation is

$$y = 400 - 0.231 \text{ Load}(x_2)$$

Predictor Coef SE Coef T P

Constant 399.97 61.03 6.55 0.003

Load(x2) -0.23067 0.05636 -4.09 0.015

S = 26.0767 R-Sq = 80.7% R-Sq(adj) = 75.9%

Analysis of Variance

Source DF SS MS F P

Regression 1 11392 11392 16.75 0.015

Residual Error 4 2720 680

Total 5 14112

Now we restrict the variable load to fit the restricted model. The MINITAB output for the fitted model is given below:

The regression equation is  
 $y = 228 - 2.86 \text{ Oil Viscosity}(x1)$   
Predictor Coef SE Coef T P  
Constant 227.83 26.01 8.76 0.001  
Oil Viscosity(x1) -2.8560 0.8780 -3.25 0.031  
S = 31.1112 R-Sq = 72.6% R-Sq(adj) = 65.7%  
Analysis of Variance  
Source DF SS MS F P  
Regression 1 10240 10240 10.58 0.031  
Residual Error 4 3872 968  
Total 5 14112

12.45). a) Now we have to fit a linear regression model including two indicator variables. Here we use MINITAB to fit the regression model to the given dataset. The output is as follows:

The regression equation is  
 $\text{MPG}(y) = -7.03 + 13.0 Z1 + 0.36 Z2 - 0.000038 \text{ Odometer} + 0.337 \text{ Octane}$   
Predictor Coef SE Coef T P  
Constant -7.025 7.088 -0.99 0.333  
Z1 12.985 1.115 11.65 0.000  
Z2 0.358 1.061 0.34 0.739  
Odometer -0.000037 0.00001 -2.37 0.028  
Octane 0.33735 0.08380 4.03 0.001  
S = 2.17777 R-Sq = 90.5% R-Sq(adj) = 88.6%  
Analysis of Variance  
Source DF SS MS F P  
Regression 4 899.69 224.92 47.43 0.000  
Residual Error 20 94.85 4.74  
Total 24 994.54

From the above output we can observe that the adjusted r-square value is 88.6 i.e. approximately 89% of the variations in the response can be explained by the fitted model or the response variable. Also the p-value is less than the 0.05 level of significance. Hence the regression explained by the above model is significant.

b) From the above output we can see that the coefficient of Z1 is 12.985, it is regression coefficient of the categorical variable sedan. Hence we can say that the sedan appears to get the best gas mileage.

c) From the above output the coefficient of the Z2 is 0.358 which is very less when compared to the Z1. Hence we can say that they are not as good as sedan in gas mileage. Also its corresponding p-value is 0.739, which is more than the 0.05 level of significance. Thus, we can say that they are not significantly different.

12.49)

### Best Subsets Regression: y versus x1, x2, x3, x4

Response is y

Vars	R-Sq	R-Sq(adj)	Mallows		x x x x			
			C-p	S	1	2	3	4
1	89.7	88.2	39.6	2.1743	X			
1	67.0	62.3	137.9	3.8903		X		
2	99.1	98.7	1.1	0.71138		X	X	
2	98.8	98.4	2.1	0.79420	X		X	
3	99.1	98.5	3.0	0.77327	X	X	X	
3	99.1	98.5	3.1	0.77845		X	X	X
4	99.1	98.2	5.0	0.86104	X	X	X	X

R-square corresponding to the model containing x2 and x3 is highest.

Hence the most suitable model for estimating the length of the infant is given below:

Regression Analysis: y versus x2, x3

The regression equation is

$$y = 2.18 + 0.958 x_2 + 3.33 x_3$$

Predictor Coef SE Coef T P

Constant 2.183 2.801 0.78 0.465

x2 0.95758 0.05927 16.16 0.000

x3 3.3253 0.2332 14.26 0.000

S = 0.711383 R-Sq = 99.1% R-Sq(adj) = 98.7%

Analysis of Variance

Source DF SS MS F P

Regression 2 318.20 159.10 314.39 0.000

Residual Error 6 3.04 0.51

Total 8 321.24

From the above output we can observe that the adjusted r-square value is 98.7%, i.e. this model can explain 99 % of the variations in the response.

12.51). a) MINITAB Model for  $\mu_{Y|x} = \beta_0 + \beta_1 x$ .

The regression equation is  
 $y = -587 + 428x$   
 Predictor Coef SE Coef T P  
 Constant -587.2 595.2 -0.99 0.343  
 x 428.43 54.41 7.87 0.000  
 S = 1051.22 R-Sq = 83.8% R-Sq(adj) = 82.4%  
 PRESS = 18811057 R-Sq(pred) = 76.99%  
 Analysis of Variance  
 Source DF SS MS F P  
 Regression 1 68505327 68505327 61.99 0.000  
 Residual Error 12 13260648 1105054  
 Total 13 81765975

From the output we noticed that the adjusted r-square value is 82.4 i.e. approximately 82% of the variations in the response can be explained by the fitted model. We can say that the fitted model is good for prediction.

**b). MINITAB Model for  $\mu_{Y|x} = \beta_0 + \beta_1x + \beta_{11}x^2$ .**  
 $y = 1180 - 192x + 35.2x^2$   
 Predictor Coef SE Coef T P VIF  
 Constant 1180.0 544.0 2.17 0.053  
 x -191.7 143.5 -1.34 0.209 17.8  
 x^2 35.209 7.915 4.45 0.001 17.8  
 S = 656.287 R-Sq = 94.2% R-Sq(adj) = 93.2%  
 PRESS = 8706974 R-Sq(pred) = 89.35%  
 Analysis of Variance  
 Source DF SS MS F P  
 Regression 2 77028141 38514071 89.42 0.000  
 Residual Error 11 4737834 430712  
 Total 13 81765975

For this quadratic model the adjusted r-square value is 93.2 i.e. approximately 93% of the variations in the response can be explained by the above regression equation. This model is also good for estimating the response values for the given values of predictors.

c) The second model i.e. the quadratic model is more preferable than the first model (simple regression model) because its residual sum squares ( $s^2$ ) and PRESS are less whereas coefficient of determination ( $R^2$ ) value is greater than the simple linear regression model.

12.55).

a) Here we use SAS technology to fit the regression model for all possible models.

output is

Model C(p) R-Square MSE Variables in Model

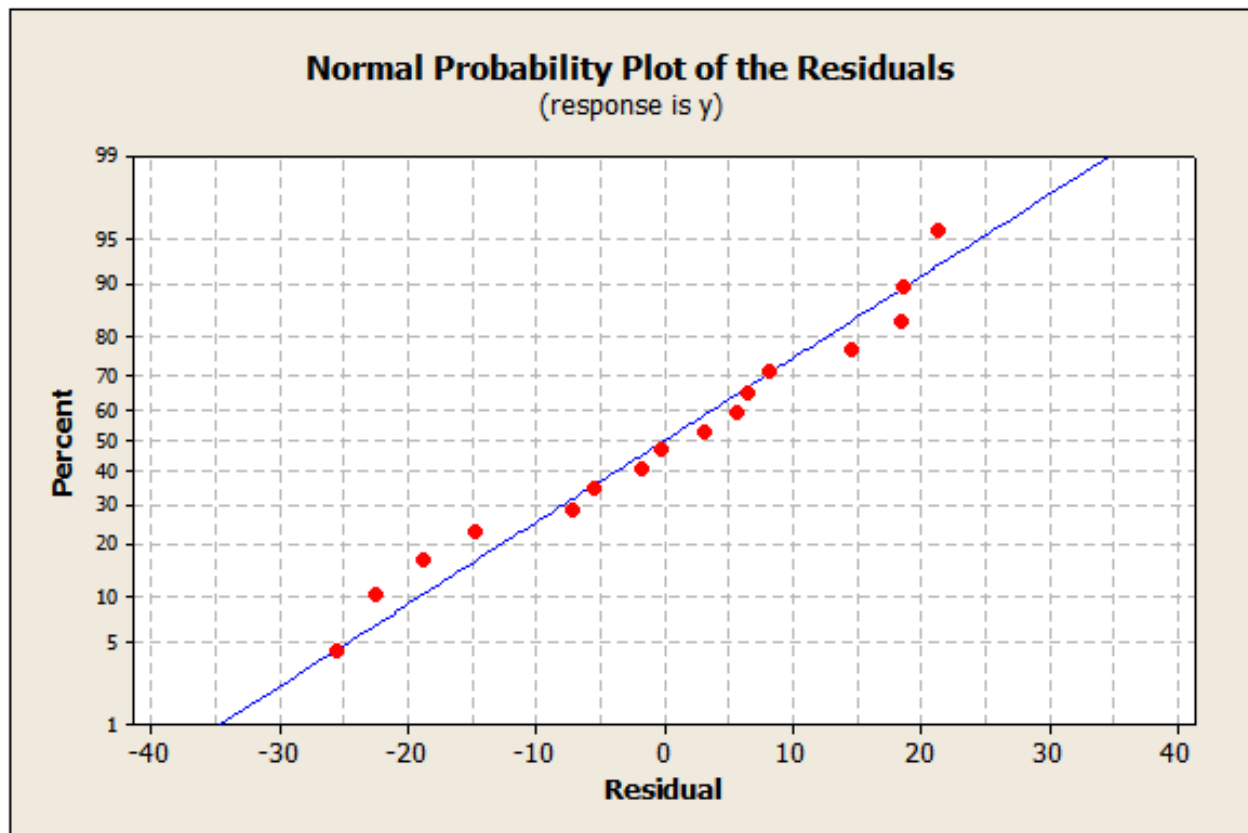
```

1 0.0932 0.0614 319.28718 x3
1 0.4642 0.0326 329.08343 x1
1 0.5112 0.0290 330.32456 x6
1 0.7635 0.0094 336.98490 x5
1 0.8224 0.0048 338.54120 x2
1 0.8401 0.0035 339.00827 x4
2 1.0255 0.1443 313.49112 x1 x3
2 1.6099 0.0990 330.10629 x5 x6
2 1.7505 0.0880 334.10455 x3 x6
2 1.7666 0.0868 334.56361 x3 x5
2 1.9717 0.0709 340.39400 x2 x3
2 2.0052 0.0683 341.34712 x1 x6
2 2.0921 0.0615 343.81763 x3 x4
3 2.2141 0.2073 314.62140 x3 x5 x6
2 2.2575 0.0487 348.52081 x2 x6
2 2.2983 0.0455 349.68025 x4 x6
2 2.3751 0.0396 351.86536 x1 x4
2 2.4245 0.0357 353.26862 x1 x5
2 2.4480 0.0339 353.93782 x1 x2
3 2.5573 0.1806 325.19301 x1 x3 x6
2 2.7259 0.0123 361.83936 x2 x5
2 2.7479 0.0106 362.46445 x4 x5
2 2.7972 0.0068 363.86473 x2 x4
3 2.8654 0.1567 334.68270 x1 x3 x5
3 2.9933 0.1468 338.62270 x1 x2 x3
3 3.0255 0.1443 339.61452 x1 x3 x4
3 3.2701 0.1253 347.14884 x2 x5 x6
...
5 6.8130 0.1608 399.68356 x1 x2 x3 x4 x5
5 6.9059 0.1536 403.11725 x1 x2 x4 x5 x6
6 7.0000 0.3015 369.63229 x1 x2 x3 x4 x5 x6

```

From the above output we can find that none of the model is good enough for prediction. In the above subset of models the highlighted models is better to the other models. But the model with x1, x2, x3, x4, x5, and x6 variables is some what good as its r-square value is 0.3015 and MSE is 332.67029.

b).



12.60).

a). Here we used SAS PROC LOGISTIC procedure to fit a logistic regression model:

The output is as follows:

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Standard Wald

Parameter	DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	-2.7620	0.1272	471.4872	<.0001
Dose 1	0.0308	0.00197	243.4111	<.0001	

Odds Ratio Estimates

Point 95% Wald

Effect Estimate Confidence Limits

Dose 1	0.031	1.027	1.035
--------	-------	-------	-------

Association of Predicted Probabilities and Observed Responses

Percent Concordant	64.3	Somers' D	0.462
--------------------	------	-----------	-------

Percent Discordant	18.2	Gamma	0.560
--------------------	------	-------	-------

Percent Tied	17.5	Tau-a	0.167
--------------	------	-------	-------

b). SAS output of  $\chi^2$ -tests revealing the significance of the regression coefficients  $\beta_0$  and  $\beta_1$ .

Standard Wald

Parameter	DF	Estimate	Error	Chi-Square	Pr >	ChiSq
Intercept	1	-2.7620	0.1272	471.4872	<.0001	
Dose 1	0.0308	0.00197	243.4111	<.0001		