U Wroclaw, Fall 2015

Applied Stats

DISCUSSION/LAB 9: MULTIPLE LINEAR REGRESSION Analysis of Covariance- dummy variables

We will use data set: FRESH15.MTW (.xlsx). Weights are in kilograms and BMI is the body mass index. Measurements of students' weight and BMI were made in September of their freshman year (1st year) and then in April of their freshmen year. There is a popular belief that freshmen students gain 15 kg of weight their freshmen year. The data is reprinted from a journal article. Do all tests of significance on the significance level of 0.05, unless specified otherwise.

1. **What is the best regression model for all the freshmen in the year of the data collection?**
2. **Is the relation between weight in September and weight in April different for Men and Women?**

- **DATA SET IS FRESH15**

Results for: FRESH15v16.MTW

Regression Analysis: WTAPR versus WTSEP, BMISP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	7469.71	3734.86	255.79	0.000
WTSEP	1	3974.64	3974.64	272.22	0.000
BMISP	1	33.68	33.68	2.31	0.134
Error	64	934.47	14.60		
Total	66	8404.18			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.82113	88.88%	88.53%	86.31%

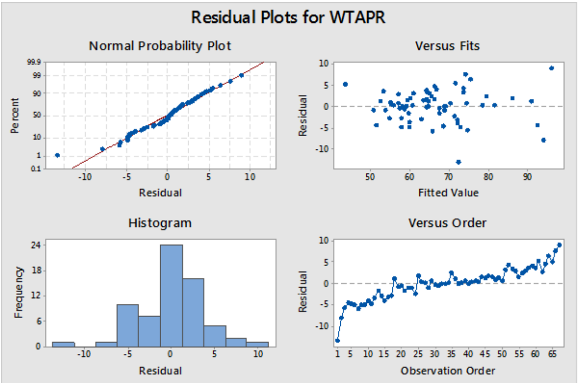
Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.27	3.30	0.69	0.494	
WTSEP	0.8887	0.0539	16.50	0.000	1.67
BMISP	0.279	0.184	1.52	0.134	1.67

Regression Equation

WTAPR = 2.27 + 0.8887 WTSEP + 0.279 BMISP

Residual Plots for WTAPR



From the point of view of statistical inference, the model is fine (see graphs). From the point of view of fit and prediction, it looks like if Sep weight is given, then sepBMI is not significant (see partial t-test or F test). We will try to reduce the model. Try model without SepBMI:

Regression Analysis: WTAPR versus WTSEP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	7436.0	7436.04	499.25	0.000
WTSEP	1	7436.0	7436.04	499.25	0.000
Error	65	968.1	14.89		
Total	66	8404.2			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.85934	88.48%	88.30%	87.14%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5.05	2.78	1.82	0.074	
WTSEP	0.9406	0.0421	22.34	0.000	1.00

Regression Equation

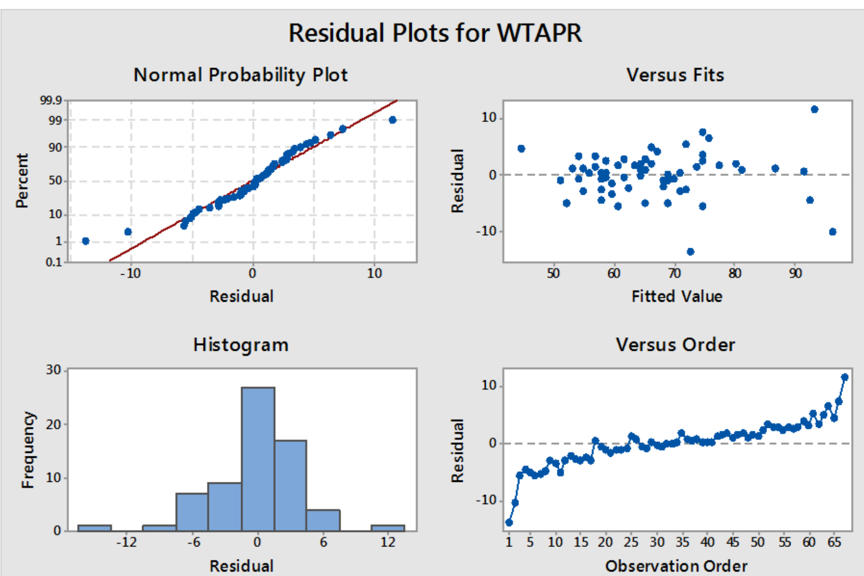
WTAPR = 5.05 + 0.9406 WTSEP

Fits and Diagnostics for Unusual Observations

Obs	WTAPR	Fit	Resid	Std Resid		
1	59.000	72.767	-13.767	-3.60	R	
2	86.000	96.280	-10.280	-2.87	R	X
4	88.000	92.518	-4.518	-1.24		X
18	92.000	91.578	0.422	0.12		X
67	105.000	93.459	11.541	3.18	R	X

R Large residual
X Unusual X

Residual Plots for WTAPR



From the point of view of statistical inference, the model is fine (see graphs). From the point of view of fit and prediction, it also looks reasonable.

Conclusion: Use `WTAPR = 5.05 + 0.9406 WTSEP to model WTAPR on WTSEP.`

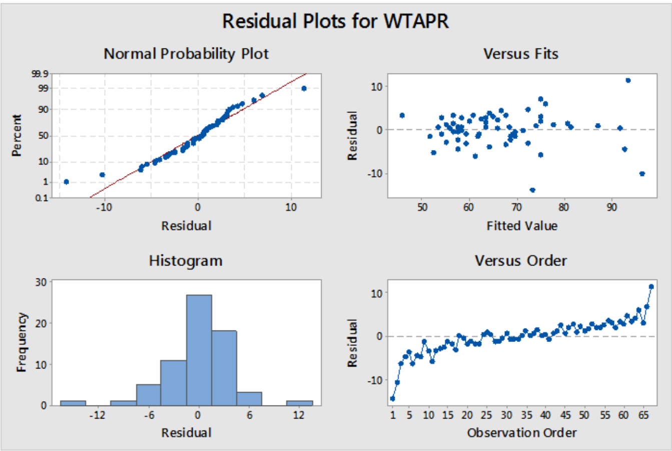
Is the relation between weight in September and weight is April different for Men and Women? Introduce a dummy variable z=1(Men) and 0(Women).

Regression Analysis: WTAPR versus WTSEP, z

Write the complex model: WTAPR = Bo + B1*WTSEP +B2*z + B3*z*WTSEP

```
Method
Categorical predictor coding  (1, 0)
Analysis of Variance
Source          DF   Adj SS   Adj MS  F-Value  P-Value
Regression       3  7464.04  2488.01   166.72    0.000
  WTSEP          1   957.00   957.00    64.13    0.000
  z              1     4.29     4.29     0.29    0.594
  WTSEP*z        1     7.99     7.99     0.54    0.467
Error           63   940.14    14.92
Total           66  8404.18

Model Summary
      S    R-sq  R-sq(adj)  R-sq(pred)
3.86302  88.81%     88.28%      86.51%

Coefficients
Term        Coef  SE Coef  T-Value  P-Value    VIF
Constant   10.84     6.08     1.78    0.080
WTSEP      0.834    0.104     8.01    0.000   6.11
z
  1        -4.17     7.77    -0.54    0.594  67.67
WTSEP*z
  1        0.090    0.123     0.73    0.467  93.47

Regression Equation
z
0  WTAPR = 10.84 + 0.834 WTSEP
1  WTAPR = 6.67 + 0.9242 WTSEP

Fits and Diagnostics for Unusual Observations

Obs   WTAPR    Fit   Resid  Std Resid
  1   59.00  73.21  -14.21      -3.74  R
  2   86.00  96.31  -10.31      -2.99  R  X
 65   49.00  45.87    3.13       0.92     X
 67  105.00  93.54   11.46       3.24  R
R  Large residual
X  Unusual X
```

Residual Plots for WTAPR



State and test all needed hypotheses.

Regression Analysis: WTAPR versus WTSEP

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	7436.0	7436.04	499.25	0.000
WTSEP	1	7436.0	7436.04	499.25	0.000
Error	65	968.1	14.89		
Total	66	8404.2			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.85934	88.48%	88.30%	87.14%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	5.05	2.78	1.82	0.074	
WTSEP	0.9406	0.0421	22.34	0.000	1.00

Regression Equation

WTAPR = 5.05 + 0.9406 WTSEP

Fits and Diagnostics for Unusual Observations

Obs	WTAPR	Fit	Resid	Std Resid		
1	59.000	72.767	-13.767	-3.60	R	
2	86.000	96.280	-10.280	-2.87	R	X
4	88.000	92.518	-4.518	-1.24		X
18	92.000	91.578	0.422	0.12		X
67	105.000	93.459	11.541	3.18	R	X

R Large residual
X Unusual X

Regression Analysis: WTAPR versus WTSEP, SEX

Method

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	7456.05	3728.02	251.65	0.000
WTSEP	1	3884.05	3884.05	262.18	0.000
SEX	1	20.01	20.01	1.35	0.249
Error	64	948.13	14.81		
Total	66	8404.18			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
3.84897	88.72%	88.37%	87.03%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	7.10	3.29	2.16	0.035	
WTSEP	0.8984	0.0555	16.19	0.000	1.75
SEX					
M	1.45	1.24	1.16	0.249	1.75

ANSWER: