

STAT 615 (Baron)

Regression

Diagnostics - I

```
PROC REG LINEPRINTER;
MODEL weight=height / R ;
PAINT RSTUDENT. >2 OR RSTUDENT.<-2 / SYMBOL='*';
PLOT weight*height RSTUDENT.*PREDICTED. ;
REWEIGHT RSTUDENT.>2 OR RSTUDENT.<-2 / WEIGHT = 0.25;
REFIT;
PRINT;
OUTPUT OUT=D RSTUDENT=R;
```

R - analysis of residuals
identify potential outliers and paint them on the plot;

```
PROC ANOVA DATA=C;
CLASS height;
MODEL weight=height;
```

To obtain SS_{PE} for Lack-of-fit test

```
PROC MODEL DATA=C;
PARMS b0 b1;
weight=b0+b1*height;
fit weight / WHITE BREUSCH=(height);
fit weight / BREUSCH=(gender smokes);
```

Testing homoscedasticity

```
PROC UNIVARIATE DATA=D NORMAL PLOT;
VAR R;
RUN;
```

Testing Normality

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	31592	31592	144.38	<.0001
Error	90	19692	218.80247		
Corrected Total	91	51284			

$SSE = 19692$

Root MSE	14.79197	R-Square	0.6160
Dependent Mean	145.15217	Adj R-Sq	0.6117
Coeff Var	10.19067		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-204.74078	29.15972	-7.02	<.0001
height	1	5.09177	0.42375	12.02	<.0001

Output Statistics

Obs	Dep Var	Predicted Value	Std Error Mean Predict	Residual	Std Error Residual	Student Residual	-2 -1 0 1 2	Cook's D
1	140.0000	131.3159	1.9246	8.6841	14.666	0.592	*	0.003
2	145.0000	161.8665	2.0768	-16.8665	14.645	-1.152	**	0.013
3	160.0000	169.5041	2.5467	-9.5041	14.571	-0.652	*	0.006
4	190.0000	166.9582	2.3815	23.0418	14.599	1.578	***	0.033
91	150.0000	136.4076	1.7053	13.5924	14.693	0.925	*	0.006
92	108.0000	109.6758	3.3309	-1.6758	14.412	-0.116		0.000

i y_i $\hat{y}_i$ $S_{\hat{y}_i}$ e_i S_{e_i} r_i

One * for each 0.5 deviation of r_i from 0.

Weighted Least Squares (reweighted outliers)

Weight: REWEIGHT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	29692	29692	177.18	<.0001
Error	90	15082	167.58030		
Corrected Total	91	44775			

Root MSE	12.94528	R-Square	0.6632
Dependent Mean	144.48315	Adj R-Sq	0.6594
Coeff Var	8.95972		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-203.66894	26.19112	-7.78	<.0001
height	1	5.06527	0.38053	13.31	<.0001

Lack-of-fit test:

The ANOVA Procedure

Class Level Information

Class	Levels	Values
height	21	61 61.75 62 62.75 63 64 65 65.5 66 67 68 69 69.5 70 71 71.5 72 73 73.5 74 75

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	35192.59814	1759.62991	7.76	<.0001
Error	71	16091.27143	226.63763		
Corrected Total	91	51283.86957			

Partition 92 obs.
into 21 groups

$$SS_{PE} = 16091$$

$$\text{Then } F_{LOF} = \frac{(19672 - 16091)/19}{16091/71}$$

The MODEL Procedure

Heteroscedasticity Test

Equation	Test	Statistic	DF	Pr > ChiSq	Variables
weight	White's Test	2.05	2	0.3597	Cross of all vars height, 1
	Breusch-Pagan	1.60	1	0.2062	
Equation	Test	Statistic	DF	Pr > ChiSq	Variables
weight	Breusch-Pagan	2.08	2	0.3530	gender, smokes, 1

Try different
models for
non-constant
variance

The UNIVARIATE Procedure
Variable: R (Studentized Residual without Current Obs)

Tests for Normality

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.984703	Pr < W 0.3585
Kolmogorov-Smirnov	D 0.072229	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.067134	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.414496	Pr > A-Sq >0.2500

No evidence
against
 $H_0: \epsilon_i \sim \text{Normal}$

No evidence
against
 $H_0: \text{Var } \epsilon_i = \text{const}$

