

# Methods Homework 10

Tim Vigers

12/2/2018

1. The relationship between plasma beta-carotene (the response) and smoking status (current smokers, former smokers, and never smokers).

a. Obtain the sample size, mean, standard deviation (SD), and standard error of the mean (SE) for plasma beta-carotene levels within each of the three smoking groups.

b. Fit a “reference cell” linear regression model (MODEL 1) regressing plasma beta-carotene levels, betaplas (the dependent variable) on smoking status (the independent variable). Make the never smokers the reference group. Write down the regression equation.

$$\hat{Y} = \beta_{\text{never}} + \beta_{\text{former}}(I_{\text{former}}) + \beta_{\text{current}}(I_{\text{current}}) = \\ 206.05096 + (-12.58139 * I_{\text{former}}) + (-84.72537 * I_{\text{current}})$$

c. Using MODEL 1, is smoking status significantly associated with plasma beta-carotene levels? Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

$$H_0 : \beta_{\text{former}} = \beta_{\text{current}} = 0$$

$$H_0 : \mu_{\text{never}} = \mu_{\text{former}} = \mu_{\text{current}}$$

$$H_1 : \beta_{\text{former}} \neq 0 \text{ or } \beta_{\text{current}} \neq 0$$

$$H_1 : \mu_{\text{never}} \neq \mu_{\text{former}} \text{ or } \mu_{\text{never}} \neq \mu_{\text{current}} \text{ or } \mu_{\text{former}} \neq \mu_{\text{current}}$$

In MODEL 1, smoking status significantly contributes to plasma beta-carotene levels ( $p = 0.0254$ ).

d. Using MODEL 1, do plasma beta-carotene levels differ between current smokers and never smokers? Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

$$H_0 : \beta_{\text{current}} = 0$$

$$H_1 : \beta_{\text{current}} \neq 0$$

$$H_0 : \mu_{\text{never}} = \mu_{\text{current}}$$

$$H_1 : \mu_{\text{never}} \neq \mu_{\text{current}}$$

According to MODEL 1, never smokers are significantly different from current smokers ( $p = 0.0070$ ).

e. Using MODEL 1, do plasma beta-carotene levels differ between former smokers and never smokers? Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

$$H_0 : \beta_{\text{former}} = 0$$

$$H_1 : \beta_{\text{former}} \neq 0$$

$$H_0 : \mu_{\text{never}} = \mu_{\text{former}}$$

$$H_1 : \mu_{\text{never}} \neq \mu_{\text{former}}$$

According to MODEL 1, never smokers are not significantly different from former smokers ( $p = 0.5725$ ).

f. Using MODEL 1, do plasma beta-carotene levels differ between current smokers and former smokers? Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion. (USE the variance-covariance matrix for the betas to answer this question).

$$H_0 : \beta_{\text{current}} - \beta_{\text{former}} = 0$$

$$H_1 : \beta_{\text{current}} - \beta_{\text{former}} \neq 0$$

$$H_0 : \mu_{\text{current}} = \mu_{\text{former}}$$

$$H_1 : \mu_{\text{current}} \neq \mu_{\text{former}}$$

$$t = \frac{\hat{\beta}_{\text{current}} - \hat{\beta}_{\text{former}}}{\sqrt{\text{Var}(\hat{\beta}_{\text{current}}) + \text{Var}(\hat{\beta}_{\text{former}}) - 2 * \text{Cov}(\hat{\beta}_{\text{current}}, \hat{\beta}_{\text{former}})}} =$$

$$\frac{-84.72537 + 12.58139}{\sqrt{975.26042464 + 495.94112724 - 2 * (209.6809913)}} = -2.224464$$

```
t <- -2.224464
p <- pt(-abs(t), df = 312)
p
```

```
## [1] 0.01341622
```

$$-2.224464 \quad t_{312}; p \approx 0.013$$

According to MODEL 1, current smokers are significantly different from former smokers ( $p < 0.05$ ).

## 2. Cell means model

a. Fit a “cell means” linear regression model (MODEL 2) predicting plasma beta-carotene levels from smoking status. Write down the regression equation.

$$\hat{Y} = \beta_{\text{never}} I_{\text{never}} + \beta_{\text{former}} I_{\text{former}} + \beta_{\text{current}} I_{\text{current}}$$

$$E[\text{plasma beta-carotene}] = 206.05096 * I_{\text{never}} + 193.46957 * I_{\text{former}} + 121.32558 * I_{\text{current}}$$

b. Use the cell means model (MODEL 2) to test if smoking status is significantly associated with plasma beta-carotene levels.

According to MODEL 2, smoking status is significantly associated with plasma beta-carotene levels ( $F = 117.49$ ,  $p < 0.0001$ ).

c. Use the cell means model (MODEL 2) to test whether plasma beta-carotene levels differ between current smokers and former smokers. Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

$$H_0 : \mu_{\text{current}} = \mu_{\text{former}} \text{ or } \beta_{\text{current}} = \beta_{\text{former}}$$

$$H_1 : \mu_{\text{current}} \neq \mu_{\text{former}} \text{ or } \beta_{\text{current}} \neq \beta_{\text{former}}$$

There is a significant difference in plasma beta-carotene levels between current and former smokers ( $p = 0.0268$ ).

d. Use the cell means model (MODEL 2) to test whether plasma beta-carotene levels differ between non-smokers (the average of never smokers and former smokers) and current smokers. Write the null and alternative hypotheses in terms of the appropriate beta coefficient(s) and also in terms of the appropriate means, test the null hypothesis, and state your conclusion.

$$H_0 : \frac{157}{272} * \beta_{\text{never}} + \frac{115}{272} * \beta_{\text{former}} - \beta_{\text{current}} = 0$$

$$H_1 : \frac{157}{272} * \beta_{\text{never}} + \frac{115}{272} * \beta_{\text{former}} - \beta_{\text{current}} \neq 0$$

$$H_0 : \frac{157}{272} * \mu_{\text{never}} + \frac{115}{272} * \mu_{\text{former}} - \mu_{\text{current}} = 0$$

$$H_1 : \frac{157}{272} * \mu_{\text{never}} + \frac{115}{272} * \mu_{\text{former}} - \mu_{\text{current}} \neq 0$$

Weighting the average of never smokers and former smokers by sample size, there is a significant difference in plasma beta-carotene levels between smokers and non-smokers ( $p = 0.0080$ ).

**3. Perform an independent samples t-test comparing plasma beta-carotene levels in current smokers versus former smokers. Compare your results to those obtained in parts (1F and 2C) and explain any differences.**

Overall the t-test is very close to the earlier results, because both found a significant difference between current and former smokers (whether you use a t-test assuming equal or unequal variances). The p-values are slightly different, but I think that's because the linear contrast uses more degrees of freedom when you get a p-value from the t distribution. The means calculated by proc ttest are the same as the parameter estimates in MODEL 2 for current and former smokers, which is also a good sign. There are likely some rounding differences between the two procedures that contribute to the small differences in t values. However, the important thing is that the conclusions are the same for the two approaches.

**The SAS System****The MEANS Procedure**

<b>Analysis Variable : betaplas</b>				
<b>smoke</b>	<b>N Obs</b>	<b>Mean</b>	<b>Std Dev</b>	<b>Std Error</b>
1	157	206.0509554	193.2085626	15.4197220
2	115	193.4695652	191.6395246	17.8704778
3	43	121.3255814	78.8116262	12.0186603

## The SAS System

The REG Procedure  
Model: MODEL1  
Dependent Variable: betaplas

Number of Observations Read	315
Number of Observations Used	315

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	244625	122312	3.72	0.0254
Error	312	10271014	32920		
Corrected Total	314	10515638			

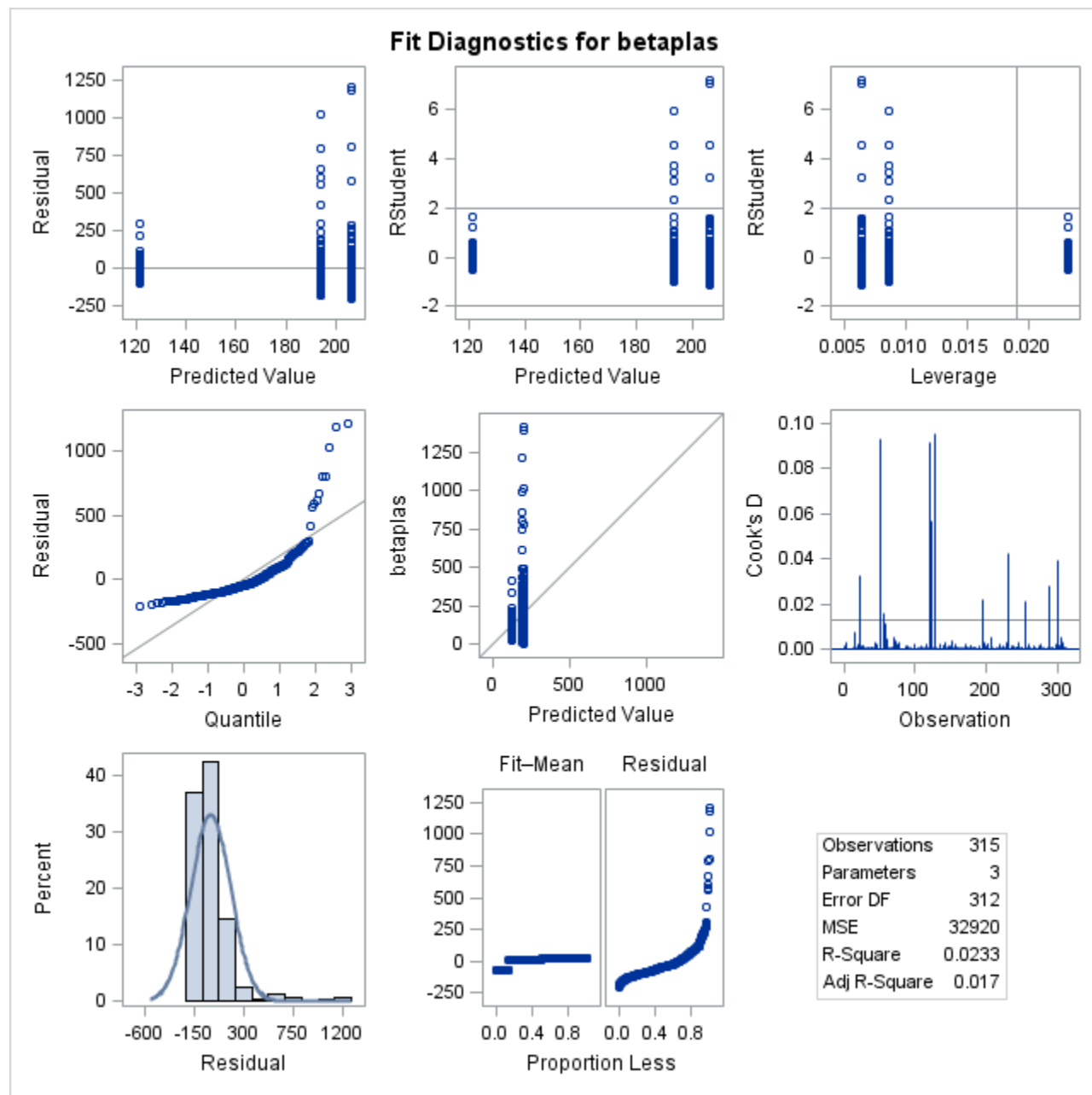
Root MSE	181.43846	R-Square	0.0233
Dependent Mean	189.89206	Adj R-Sq	0.0170
Coeff Var	95.54821		

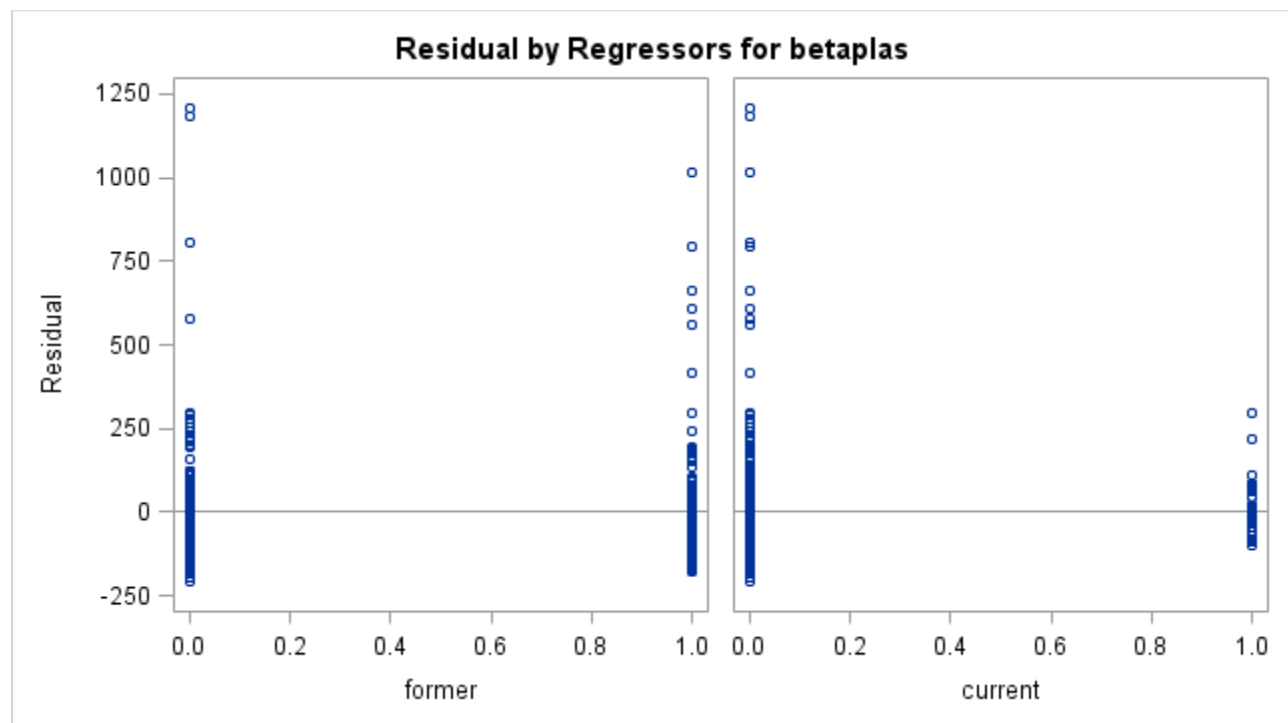
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	206.05096	14.48037	14.23	<.0001
former	1	-12.58139	22.26974	-0.56	0.5725
current	1	-84.72537	31.22916	-2.71	0.0070

Covariance of Estimates			
Variable	Intercept	former	current
Intercept	209.6809913	-209.6809913	-209.6809913
former	-209.6809913	495.94112724	209.6809913
current	-209.6809913	209.6809913	975.26042464

## The SAS System

The REG Procedure  
Model: MODEL1  
Dependent Variable: betaplas







## The SAS System

The REG Procedure  
 Model: MODEL2  
 Dependent Variable: betaplas

Number of Observations Read	315
Number of Observations Used	315

**Note:** No intercept in model. R-Square is redefined.

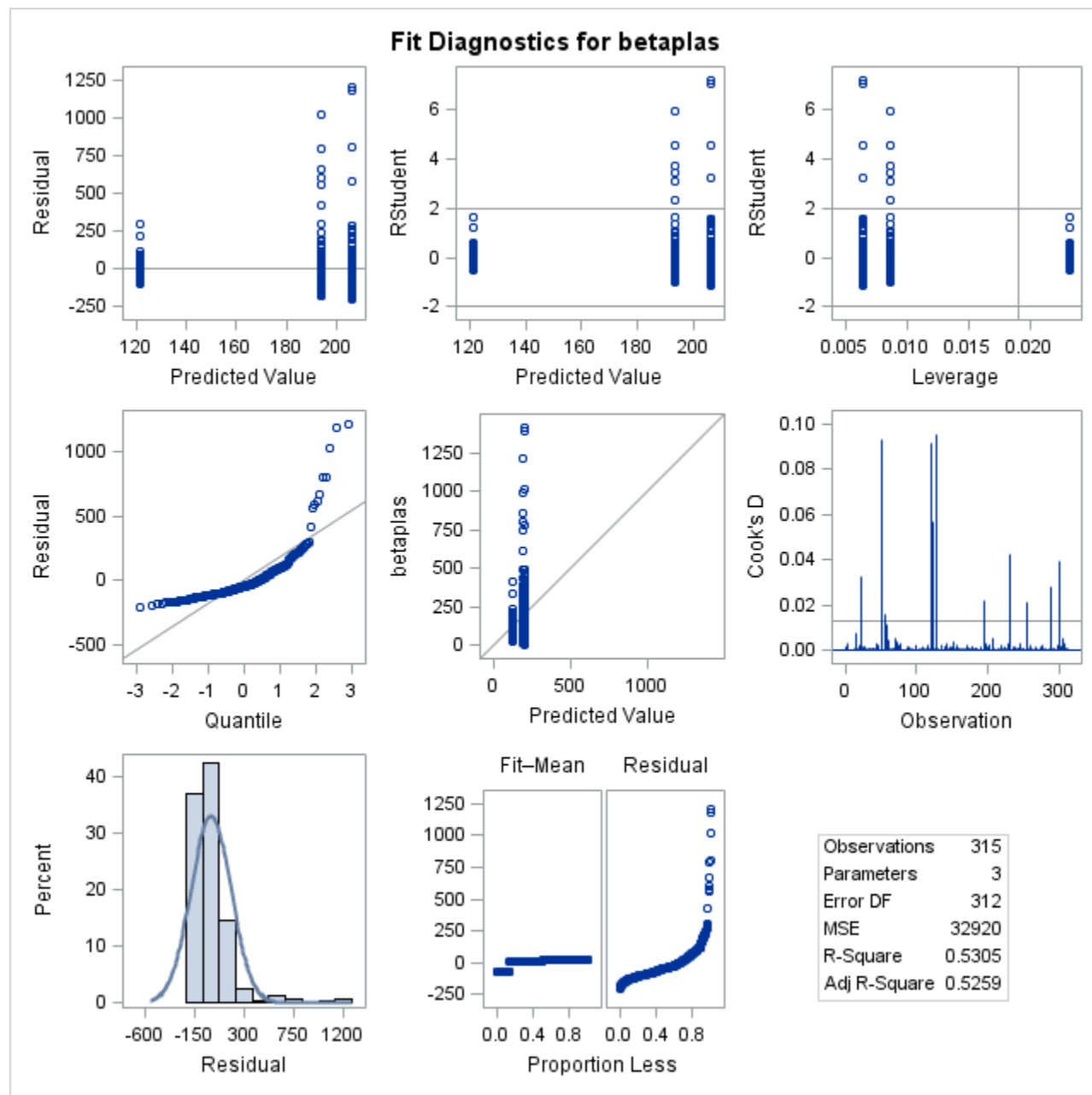
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	11603208	3867736	117.49	<.0001
Error	312	10271014	32920		
Uncorrected Total	315	21874222			

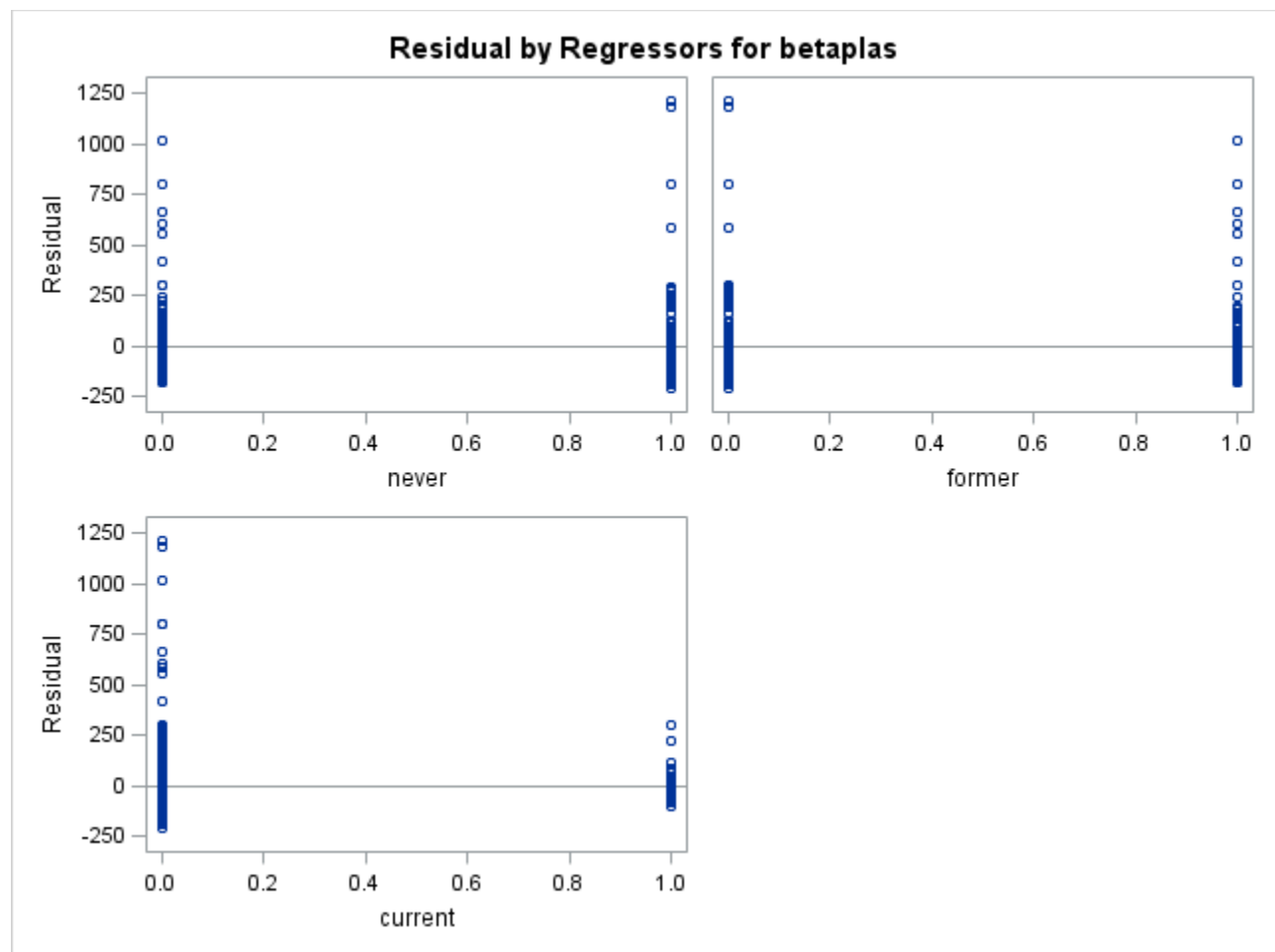
Root MSE	181.43846	R-Square	0.5305
Dependent Mean	189.89206	Adj R-Sq	0.5259
Coeff Var	95.54821		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
never	1	206.05096	14.48037	14.23	<.0001
former	1	193.46957	16.91922	11.43	<.0001
current	1	121.32558	27.66911	4.38	<.0001

## The SAS System

The REG Procedure  
Model: MODEL2  
Dependent Variable: betaplas





---

## The SAS System

### The REG Procedure Model: MODEL2

Test 1 Results for Dependent Variable betaplas				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	162896	4.95	0.0268
Denominator	312	32920		

---

## The SAS System

### The REG Procedure Model: MODEL2

Test 2 Results for Dependent Variable betaplas				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	234324	7.12	0.0080
Denominator	312	32920		

## The SAS System

### The TTEST Procedure

Variable: betaplas

smoke	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
2		115	193.5	191.6	17.8705	16.0000	1212.0
3		43	121.3	78.8116	12.0187	25.0000	418.0
Diff (1-2)	Pooled		72.1440	168.8	30.1819		
Diff (1-2)	Satterthwaite		72.1440		21.5361		

smoke	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
2		193.5	158.1	228.9	191.6	169.7	220.2
3		121.3	97.0709	145.6	78.8116	64.9834	100.2
Diff (1-2)	Pooled	72.1440	12.5261	131.8	168.8	152.0	189.9
Diff (1-2)	Satterthwaite	72.1440	29.6010	114.7			

Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	156	2.39	0.0180
Satterthwaite	Unequal	154.6	3.35	0.0010

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	114	42	5.91	<.0001

