

BIST P8130: Biostatistics Methods I

Recitation 06 Simple Linear Regression in SAS

Zilan Chai

Fall 2017

This recitation's big ideas:

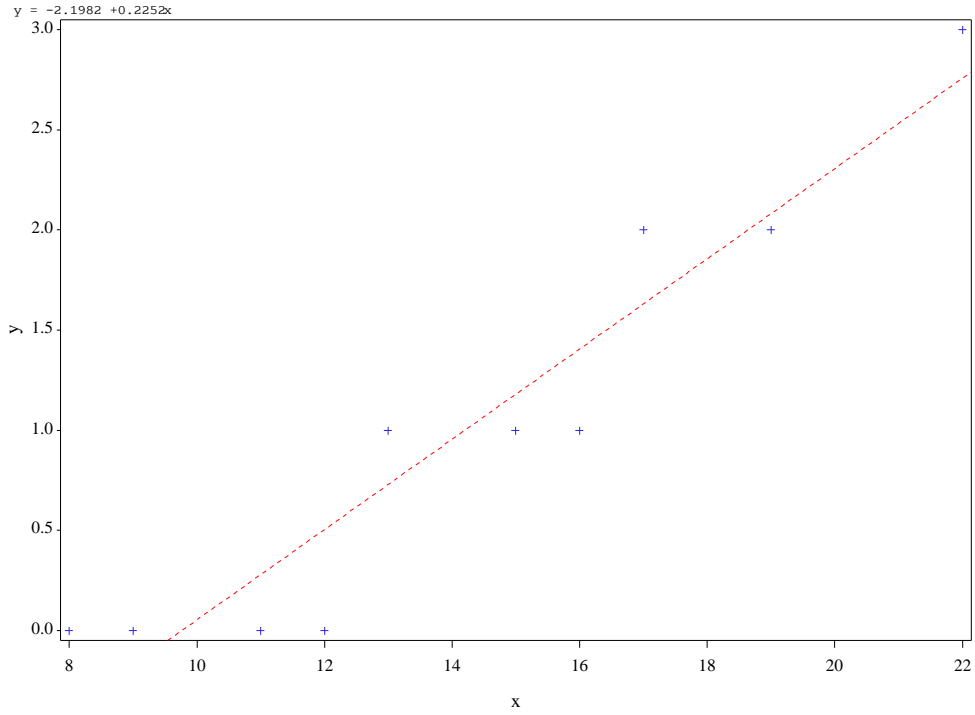
- Use PROC REG to do a simple linear regression in SAS

Simple linear regression is a technique that is used to explore the nature of the relationship between two continuous variables. Regression enables us to investigate the change in one variable, called the **response**, which corresponds to a given change in the other, known as the **explanatory variable**. One of the goals of regression is to predict or estimate the value of the response that is associated with a fixed value of the explanatory variable.

Example: Kutner 1.21

Obs	x	y
1	16	1
2	9	0
3	17	2
4	12	0
5	22	3
6	13	1
7	8	0
8	15	1
9	19	2
10	11	0

```
proc reg data = airfeight;  
model y = x;  
plot y*x / nostat cline = red;  
run;
```



Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	9.00901	9.00901	72.73	<.0001
Error	8	0.99099	0.12387		
Corrected Total	9	10.00000			

Root MSE	0.35196	R-Square	0.9009
Dependent Mean	1.00000	Adj R-Sq	0.8885
Coeff Var	35.19572		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-2.19820	0.39119	-5.62	0.0005
x	1	0.22523	0.02641	8.53	<.0001

Example:

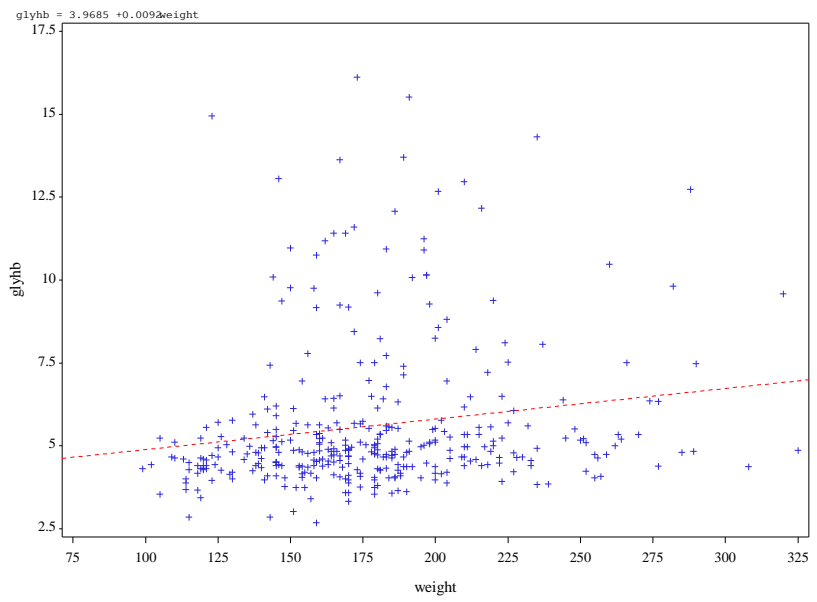
Researchers have noted that the over-weight people are more likely to develop diabetes than people are not over-weight. A number of studies have examined the relationship between weight and glycosylated hemoglobin. (High level of glycosylated hemoglobin indicates high blood glucose level for a period of time) The results of one such study are summarized below:

Obs	glyhb	weight
1	4.309999943	121
2	4.440000057	218
...
364	4.389999866	154
365	13.63000011	167
366	4.489999771	197

SAS can quickly perform a linear regression analysis of this:

SAS Syntax:

```
proc reg data = diab;  
  model glyhb = weight;  
  plot glyhb*weight / nostat cline = red;  
run;
```



Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	51.13777	51.13777	10.54	0.0013
Error	364	1765.71987	4.85088		
Corrected Total	365	1816.85764			

Root MSE	2.20247	R-Square	0.0281
Dependent Mean	5.60730	Adj R-Sq	0.0255
Coeff Var	39.27867		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	3.96850	0.51770	7.67	<.0001
weight	1	0.00920	0.00283	3.25	0.0013

Let's interpret the output:

- (1) What is the least square estimate of the population regression line?
- (2) Interpret the intercept of the fitted line.
- (3) Interpret the slope of the fitted line.

The option `CLB` produces 95% CIs for the individual β 's:

```
proc reg data=diab;
model glyhb = weight / clb;
output out=lin p=pred r=res;
run;
```

Parameter Estimates							
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	95% Confidence Limits	
Intercept	1	3.96850	0.51770	7.67	<.0001	2.95045	4.98656
weight	1	0.00920	0.00283	3.25	0.0013	0.00363	0.01477