1. Smokers in the sample have a higher floor for their FEV results but a much lower ceiling as well. THe ceiling and floor for smokers and their FEV are about equal distance from the mean but the non-smokers have a much higher ceiling than their floor. Overall the non-smoker mean and mode are lower than the smokers mean and mode but smokers have a smaller range than non-smokers.

2. The estimate $\hat{\beta} = 0.71072$ tells us that there is an increase of 0.71072 for the value of the FEV if the individual is a smoker but the p-value is < 0.001 so it is statistically insignificant.

4. Smoking Status is a factor for FEV but Age is also a factor for FEV. Older individuals are also more likely to smoke as seen in the "Distribution by Age of Smoking" boxplot but is not a direct result of it because younger individuals in the study population also were found to be smokers.

5. The value of the estimate $\hat{\beta} = 0.71072$ in part 2 while the value of $\hat{\beta} = -0.20899$ in part 5 which means that now there is a decrease in FEV if an individual smokes.

```
DATA fev;
    INFILE '/home/u58684395/tutorial5/fev.txt';
    INPUT age fev height sex smoke;
RUN;
PROC SORT Data=fev;
    BY smoke;
RUN;

PROC BOXPLOT DATA=fev;
    PLOT fev*smoke;
RUN;

PROC REG DATA=fev PLOTS=none;
    MODEL fev=smoke;
    TITLE 'Regression of FEV Against Smoking Status';
RUN;

PROC REG DATA=fev PLOTS=none;
    MODEL fev=age;
    TITLE 'Regression of FEV Against Age';
RUN;
PROC BOXPLOT DATA=fev;
    PLOT age*smoke;
RUN;

PROC REG DATA=fev PLOTS=none;
    MODEL fev=smoke age;
    OUTPUT out=regout p=pred;
    TITLE 'Regression of FEV Against Smoke and Possible Condfounder of Age';
RUN;
```

Distribution of fev by smoke
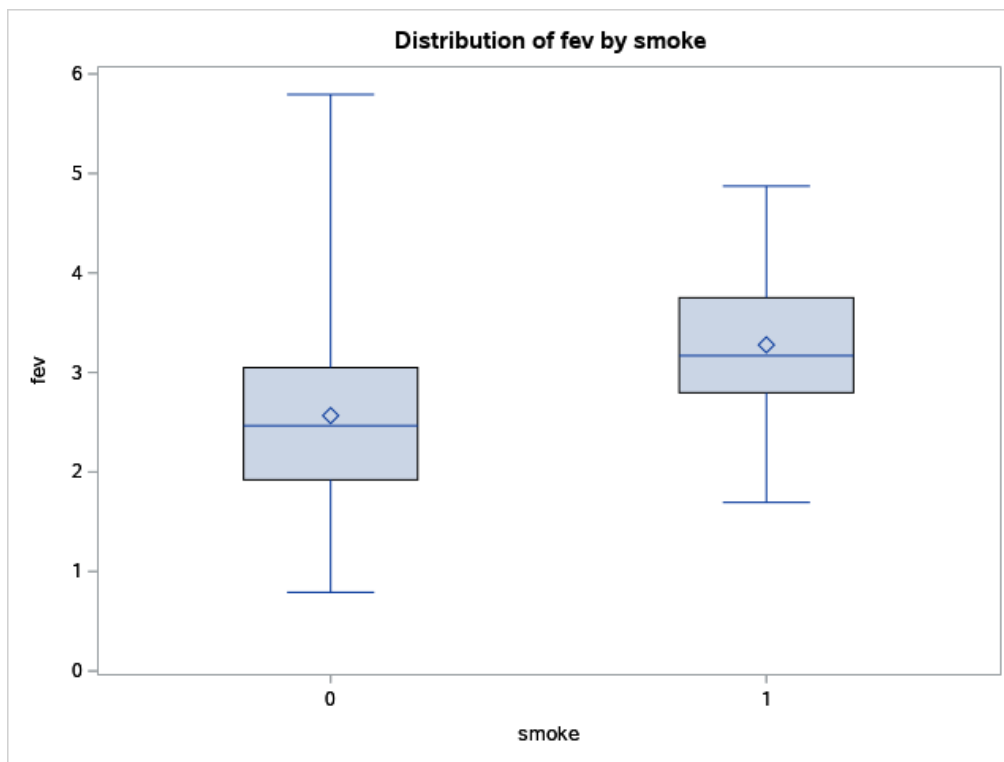
Regression of FEV Against Smoking Status

The REG Procedure
Model: MODEL1
Dependent Variable: fev

Number of Observations Read	654
Number of Observations Used	654

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	29.56968	29.56968	41.79	<.0001
Error	652	461.35015	0.70759		
Corrected Total	653	490.91984			

Root MSE	0.84119	R-Square	0.0602
Dependent Mean	2.63678	Adj R-Sq	0.0588
Coeff Var	31.90198		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	2.56614	0.03466	74.04	<.0001
smoke	1	0.71072	0.10994	6.46	<.0001

Regression of FEV Against Age

The REG Procedure
Model: MODEL1
Dependent Variable: fev

Number of Observations Read	654
Number of Observations Used	654

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	280.91916	280.91916	872.18	<.0001
Error	652	210.00068	0.32209		
Corrected Total	653	490.91984			

Root MSE	0.56753	R-Square	0.5722
Dependent Mean	2.63678	Adj R-Sq	0.5716
Coeff Var	21.52349		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.43165	0.07790	5.54	<.0001
age	1	0.22204	0.00752	29.53	<.0001

Regression of FEV Against Age
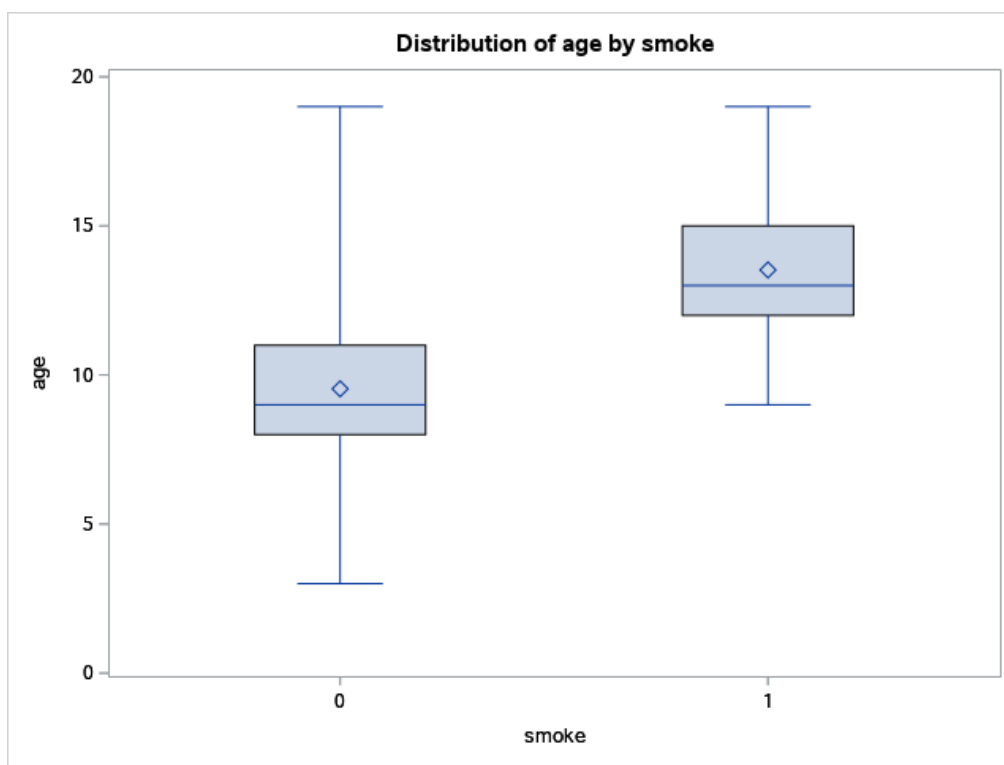
Regression of FEV Against Smoke and Possible Condfounder of Age

The REG Procedure
Model: MODEL1
Dependent Variable: fev

Number of Observations Read	654
Number of Observations Used	654

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	283.05825	141.52913	443.25	<.0001
Error	651	207.86159	0.31930		
Corrected Total	653	490.91984			

Root MSE	0.56506	R-Square	0.5766
Dependent Mean	2.63678	Adj R-Sq	0.5753
Coeff Var	21.43003		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.36737	0.08144	4.51	<.0001
smoke	1	-0.20899	0.08075	-2.59	0.0099
age	1	0.23060	0.00818	28.18	<.0001