In [34]:
```
import pandas as pd
#import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline

data=pd.read_csv('actionsSG_Total.csv')

plt.scatter(data['SGpercent'],data['Totalpercent'],color='k')
plt.title("Linear Model")


meanSGpercent=data['SGpercent'].mean()
meanTotalpercent=data['Totalpercent'].mean()  ##there is a better way...

x = np.array([data['SGpercent'].min(), data['SGpercent'].max()])

print  "Correlation Coef r= ", np.corrcoef(data['SGpercent'],data['Totalpercen
t'])
#Linear equation
a= data['SGpercent'].cov(data['Totalpercent'])/data['SGpercent'].var()
b=a*(-meanSGpercent)+meanTotalpercent #from equation y-!y=a(x-!x)
y=a*x+b
print ('Linear equation y= %f *x+( %f)' %(a,b))
plt.plot(x, y)

print ('Mean SGpercent= %f , Mean Totalpercent= %f' %(meanSGpercent,meanTotalp
ercent))
plt.grid(True,which="both")
plt.xlabel("SGpercent")
plt.ylabel("Totalpercent")
plt.show()
```

```
Correlation Coef r=  [[ 1.          0.65221047]
 [ 0.65221047  1.        ]]
Linear equation y= 0.244479 *x+( -0.404928)
Mean SGpercent= -0.477500 , Mean Totalpercent= -0.521667
```