Example of multiple linear regression, and exercises due 11-12-07.

Example 1. Data is gathered on 7 taxable properties. Here are the scores observed for independent variables and dependent variable y.

x_{i1} tax paid last year	x_{i2} (1 business, 0 if not)	y_i audit – determined tax due this year
2300	0	3000
7200	0	7800
1200	0	1321
4700	0	5100
1900	0	2200
5300	1	5800

1a. Determine the least squares coefficients of a fit of the model $y_i = \beta_0 1 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$.

```
In[55]:= xtax = {{1, 2300, 0}, {1, 7200, 0}, {1, 1200, 0}, {1, 4700, 0}, {1, 1900, 0}, {1, 5300, 1}}

Out[55]= {{1, 2300, 0}, {1, 7200, 0}, {1, 1200, 0}, {1, 4700, 0}, {1, 1900, 0}, {1, 5300, 1}}

In[56]:= ytax = {3000, 7800, 1321, 5100, 2200, 5800}

Out[56]= {3000, 7800, 1321, 5100, 2200, 5800}

In[57]:= betahattax = betahat[xtax, ytax] 1.0

Out[57]= {256.027, 1.0486, -13.6328}
```

1b. From the above, $\hat{\beta}$ = {256.027, 1.0486, -13.6328}.
Calculate the anticipated audit-determined tax for the **average property** if
μ_1 = 3900 is the population average tax for all properties in the community
μ_2 = 0.18 is the fraction of properties in the community that are businesses
Note: this is in fact the regression-based estimate of μ_y. It is computed as the dot product of {1, 3900, 0.18} with $\hat{\beta}$.

```
In[58]:= betahattax.{1, 3900, 0.18}

Out[58]= 4343.13
```

1c. So, based on a with-replacement sample of the above six properties, the regression-based estimator of μ_y is 4343.13 (dollars). In order to use this estimate we had to know the population means 3900 and 0.18 as above. If the community has 2000 properties we would estimate from this that the total audit-determined value of all 2000 properties is around 2000 times 4343.13. Sample size n = 6 is not a large enough but (just to illustrate) had n been large we'd have been entitled to margin of error

$$1.96 \sqrt{1 - R^2} \frac{s_y}{\sqrt{n}}$$

which would better the margin of error without the term $\sqrt{1 - R^2}$, applicable to just using $\overline{y}$. What is the multiple correlation R for this admittedly small sample of only six?

```
In[59]:= R[xtax, ytax] 1.0

Out[59]= 0.997381
```

From R = 0.997381 above we find $\sqrt{1 - R^2}$ = 0.0723266 indicating that this model appears to be doing a fine job of improving upon the estimate of μ_y. It is little over 7% of the width of the CI based on just $\overline{y}$.

```
In[60]:= Sqrt[1 - 0.997381^2]

Out[60]= 0.0723266
```

1d. Based on our fit to six sample properties, what does the model estimate for the audit-determined tax of a business property whose tax last year was 5900? Just feed the properties x-scores {1, 5900, 1} into the fitted model i.e. take the dot product $\hat{\beta}$.{1, 5900, 1}.

```
In[53]:= betahattax.{1, 5900, 1}

Out[53]= 6429.16
```

Had the property been non-business our fitted model would anticipate an audit-determined tax of 6442.8.

```
In[61]:= betahattax.{1, 5900, 0}

Out[61]= 6442.8
```

1e. Find the residuals $\hat{\epsilon} = y_i - x.\hat{\beta}$ of the fit.

```
In[54]:= resid[xtax, ytax] 1.0

Out[54]= {332.182, -5.98189, -193.353, -84.4699, -48.3765, 0}
```

1f. According to the fitted model what is the increase in projected audit-determined tax for every $1000 increase in tax paid last year? ans. The coefficient of tax paid last year is $\hat{\beta}_1 = 1.0486$ so the anticipated increase is 1048.60 - 1000 = 48.60 (dollars).

The following are due Monday, 11-12-07.

Exercises from chapter 11.
1. 26, 28, 30, 32,
2. Use routines from Little Software4 to determine betahat, residuals, and R from example 11.12. Compare your results with those reported on page 526.
3. For exercise 35 pg.535 verify the reported betahat using Lttle Software4 (don't submit the problem itself).
4. Exercise 38 pg. 537 and verify the repoted R^2 and fitted values $\hat{y} = x\hat{\beta}$ using Little Sortware4 Below, I've worked out R^2, $\hat{\beta}$, and $\hat{y} = x\hat{\beta}$ for observations one through seven only.

```
In[24]:= y38 = {.7, .63, .013, .049, .7, .1, .04, .0065}

Out[24]= {0.7, 0.63, 0.013, 0.049, 0.7, 0.1, 0.04, 0.0065}

In[30]:= MatrixForm[y38]

Out[30]//MatrixForm=
  ( 0.7    )
  ( 0.63   )
  ( 0.013  )
  ( 0.049  )
  ( 0.7    )
  ( 0.1    )
  ( 0.04   )
  ( 0.0065 )
```

```
In[27]:= x38 = {{1, 30, 30, 10}, {1, 30, 30, 10}, {1, 30, 30, 18.41}, {1, 40, 40, 5},
           {1, 30, 30, 10}, {1, 13.8, 30, 10}, {1, 20, 40, 5}, {1, 20, 40, 15}}

Out[27]= {{1, 30, 30, 10}, {1, 30, 30, 10}, {1, 30, 30, 18.41}, {1, 40, 40, 5},
          {1, 30, 30, 10}, {1, 13.8, 30, 10}, {1, 20, 40, 5}, {1, 20, 40, 15}}

In[29]:= MatrixForm[x38]

Out[29]//MatrixForm=
         1  30    30  10
         1  30    30  10
         1  30    30  18.41
         1  40    40  5
         1  30    30  10
         1  13.8  30  10
         1  20    40  5
         1  20    40  15

In[31]:= betahat[x38, y38]

Out[31]= {1.95247, 0.00962859, -0.0487403, -0.0273348}

In[33]:= (R[x38, y38])^2

Out[33]= 0.591298
```