Lab A
Picking Stocks

Objectives:

- To find a useful linear model
- To introduce the idea of a residual as a way of measuring how good a model is
- To reinforce Excel skills from the previous lab

Before the Lab

The purpose of this lab is to present a simple example of how models are used to pick stocks to buy on the stock market. We emphasize that we are in no way recommending an investment of any kind; this is simply an example based on one very common method of analyzing stocks.

Just in case you don't know what "buying stock in a company" means, companies issue shares of stock that represent fractional ownership of that company. For example, if XYZ Company issues 100,000 shares of stock and you buy one, you now own \( \frac{1}{100,000} \) th of XYZ company (congratulations). In most cases people buy more than one share at a time, but we'll stick to one share for most of this lab.

One theory about buying stocks says that how much you should pay to buy a stock depends on how much profit the company is generating for you. To continue the above example, if XYZ company in question had profits of $73,500 last year, and you own one share = \( \frac{1}{100,000} \) th of the company, then in theory you also own \( \frac{1}{100,000} \) th of those profits, or \( \frac{73,500}{100,000} = \$0.735 \). This number, $0.735, is called the earnings per share of XYZ company, abbreviated "EPS".

Here's the big idea: when it comes to picking stocks, the higher the earnings per share, the more you should be willing to pay for that share. If ABC Company also had 100,000 shares, and you owned one of them, and their profits last year were only $49,000, then your share of those profits would be only $0.49. All other things being equal, you would be willing to pay more to get profits of $0.735 than you would to get profits of $0.49.

But how much more?

In this lab, we present you with information on 18 real companies, all of whom are gas utilities: that is, their primary business is to sell natural gas to industries, businesses, and consumers. You will get the estimated earnings per share for the coming year, as determined by the Value Line Company, and the price of the stocks in question as of Jan. 9, 2003.
Your job is to find a good linear model, then use that to make decisions about particular stocks.

**Practice Table**

Read the rest of the lab, then come back to this part. Below is what the top of the table in the data looks like when you download it. For each of the cells labeled "?", write in the space below the appropriate Excel formula for that table.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td></td>
<td>slope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;= average absolute residuals</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>(as of 1/9/03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Company</td>
<td>Next year's EPS</td>
<td>Actual price</td>
<td>Slopes</td>
<td>Predicted price</td>
<td>Residuals</td>
<td>Absolute residuals</td>
<td>% increase</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Semco Energy</td>
<td>0.73</td>
<td>5.84</td>
<td></td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Southern Union</td>
<td>1.05</td>
<td>16.6</td>
<td>?</td>
<td>?</td>
<td></td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>14</td>
<td>Cascade Nat'l Gas</td>
<td>1.3</td>
<td>19.75</td>
<td></td>
<td>?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

D13: 
E12: 
F13: 
G14: 
H13: 

**In the Lab**

0. Open the spreadsheet for Lab A (your instructor or TA will give you the URL for the lab on the Web). On the spreadsheet is the data mentioned above. The column labeled "Next year's EPS" are the earnings per share predicted for the next year, while the "Actual price" is the price of the stocks at 4:00 pm, Jan. 9, 2003. To help you out, we have also set up and organized the rest of the spreadsheet so that you just have to fill in appropriate blanks.

1. Your first task is to *make a graph*. You should have "Next year's EPS" on the x-axis and "Actual price" on the y-axis. If you get stuck, refer back to Lab 0, and/or ask your TA. Save your work when you get done.
Since in this graph you put EPS on the x-axis and price on the y-axis, then x will always refer to EPS in this lab, and y will always refer to price.

2. When you look at the points in the graph you just made, they should look pretty close to, but not exactly on, a straight line. Most of your work in lab today will be to try to find the best linear model you can for this data. In this step, you will guess an initial linear model. You'll work on improving it later.

In class you've learned about using the median slope as a good guess for the slope of a model. You'll use Excel to figure out what the median slope is.

In cell D13, you want to put the slope of the line connecting the points for Semco Energy and Southern Union. The slope is just the difference in the y-values, located in column C, divided by the difference in x-values, located in column B. So the formula for the slope you want is \( \frac{(C13-C12)}{(B13-B12)} \). Now you can copy that formula down the rest of the column to compute all the other slopes in the table. Again, if you don't remember how to copy stuff down a column, refer to Lab 0, or ask your TA.

In cell E7, the one labeled "Slope", you want to put in the median of the slopes you just calculated. To have Excel do this, type in \=median(D13:D29). If all is well, you should get a result of 12.6.

Since you're just guessing a model, you could make your original estimate for the y-intercept a couple of ways. One way is to hold a straight edge up to the graph and see about where it hits the y-axis. Another would be to pick a particular point from the data and use it and the slope you just found in the point-slope formula. However you want to do it, estimate the y-intercept in some way, and type the number into cell E8, labeled "intercept".

Saved your work lately?

3. OK, so at this point you have a slope and an intercept for your model. Your next task is to get some idea of how good that model is: how well does it fit your data? In this step, you'll add the line corresponding to your model to your graph.

First, you need to compute the y-values predicted by the model. Let's look at the first company, Semco Energy, as our example. We'll pretend that the model you have has a slope of 14 and an intercept of 5 (your results will almost surely be different). How can we use these numbers to predict the price of the stock? The answer lies with the linear formula \( y = mx + b \). If you are trying to predict the price, that means you're calculating \( y \) (since \( y \) stands for price). So you would plug in 14 for \( m \), 0.73 for \( x \) (since that's the EPS for Semco), and 5 for \( b \), and the result would be a price \( y \) of $15.22.

Now you're ready to actually type a formula into Excel. What you want Excel to compute is what you just computed, \( mx + b \). Looking at your spreadsheet, you've got your number for the slope \( m \) in cell E7, your number for the slope \( b \) in cell E8, and the EPS for Semco
is in cell B12. So the Excel formula that reproduces $mx + b$ is $=E7*B12+E8$. (Don't type this in anywhere yet.)

So we've talked about how to compute the price your model predicts for Semco Energy, given its EPS. In the column labeled "Predicted price", you're going to want to do that for each of the companies in the table. So the basic idea is that you'll type $=E7*B12+E8$ in cell E12, then copy it down the column. However, your formula refers to two numbers outside of the table and, as we learned in Lab 0, when that happens you need to add $'$s to the formula. So the correct formula for cell E12 is $=E$7*$B12+E$8. Enter THAT and copy it down the table.

Now that you've got predicted prices from your model, add that model to the graph you've already made. Save your work.

4. How'd you do? How close is your line to the actual points? Do you think you could do better? In the last major step of the Lab, you're going to learn how to improve your initial model to make it fit the given points better.

First, look at your graph. Is your line obviously too steep or too shallow? Is it obviously too high or too low? Try different values for the slope $m$ and intercept $b$ in cells E7 and E8 until you're satisfied with the graph.

To go beyond this, we need residuals. We'll go over residuals in more detail in Chapter 2, but for now, residuals are the difference between the predicted $y$-value and the actual $y$-value. In this lab, that means the difference between the price the model predicts, and the actual price. Obviously, the better the model, the smaller the difference between predicted prices and actual prices, and so the smaller the residuals. You want the model with the smallest average residual (actually, for reasons we'll talk more about in Chapter 2, the smallest average absolute residual: that is, the average of the absolute values of the residual). First, we'll get Excel to calculate the average residuals for you, then you can fine-tune your slope and intercept to make the average as small as you can make it.

In column F you'll have Excel calculate the residuals for your current model. The residual for Semco Energy is the difference between the predicted price in cell E12 and the actual price in cell C12. So the formula you need for cell F12 is $=E12-C12$. Copy that down column F.

But you want the absolute residuals, so in column G you're going to compute the absolute value of these residuals. In Excel, the absolute value function is called abs, so type $=$abs(F12) in cell G12 and copy that down. Finally, the average of the absolute residuals goes in cell G9: type in $=$average(G12:G29).

Now adjust your slope up and down, and your intercept up and down, looking for ways to make the average residuals as small as you can. If your line is already pretty good, you may just want to change the slope by plus-or-minus 0.1, and your intercept by plus-or-minus 0.1. As a target, see if you can get the average residuals below 1.75.
5. You're basically done, but here's one more quick calculation for column H that you'll need for the "After the Lab" section. In column H, you want to compute the % increase (negative numbers would mean a decrease) in the predicted price of your model over the actual price. Column E shows the increase in the predicted price over the actual price, so just divide the number in E by the number in C to get the % increase. Type \((E12 - C12) / C12\) in cell H12. Before you copy it down, we're going to change the formatting to get Excel to write the result as a percentage. To do that, after you enter the formula in cell H12, choose Format/Cells . . . and click on the "Number" tab. Then click on the "Percentage" option and make sure you'll get at least one decimal place, then hit "OK". Now copy the resulting formula down column H.

You're done with the In the Lab portion. Save your work, print it, and quit.

After the Lab

1. Write a sentence explaining the meaning of the slope for the model you found. Do the same for the intercept.

2. (This one is supposed to be really easy, don't look for tricks.) Look at the results for the company AGL Resources, towards the middle of the list. The Actual price listed is $24.41, while, depending on exactly what model you end up with, the Predicted price is around $26. If you could buy a stock for $24.41 and you had a model that said the stock ought to be worth $26, should you buy it? In general, should you consider buying stocks with positive residuals or negative residuals?

3. (follow up to #2). Suppose you had $10,000 to invest, and decided to invest it in AGL Resources. How many shares of that stock could you buy? If the price rose to the price predicted by the model, how much money would you have? What does this have to do with the percentage calculated in column H?

4. Which stock looks best to buy: that is, which stock has the greatest difference between actual price and predicted price? Which stock looks best to sell?

5. Not a question, just a comment. Unfortunately, of course, it's never this easy. For example, the reason why the stock that looks best to buy is so cheap that is that it's considered much more risky than the other stocks in the list: exaggerating somewhat, this company is believed to have a higher chance of falling on its face than the other stocks in the list. So you have to factor a lot of other things into the price of a stock than just its earnings, and that's where the fun (?) comes into it.