STATISTICS TOPIC NOTES

Multiple Linear Regression with Two Predictor Variables

Preface. You can profit from what you already know about simple regression (one predictor) to help understand multiple regression (more than one predictor). Therefore, in this section of the Topic Notes, I mimic the earlier Topic Notes on simple regression. I especially call attention to what is new or different. A key to understanding how multiple regression differs from simple regression is the *ceteris paribus* concept, which will be explained.

Regression is the most important statistical tool for reducing uncertainty about business questions. Regression is about relationships between numerical variables. For example, we can give a better estimate of the rent of an apartment if we know its area and number of bathrooms; we can better forecast sales of a product if we know the price of the product and how much will be spent on advertising. In order to use the area of an apartment and its number of bathrooms to estimate the rent, we need to understand the relationship between area and bathrooms as inputs and rent as output; in order to use price and advertising to forecast sales, we need to understand the relationship between price and advertising as inputs and sales as output. If we have rent, area, and bathroom counts on a representative sample of apartments, regression can provide an equation to estimate rent by plugging in the value of area and the number of bathrooms; if we have data on sales at different levels of price and advertising, regression can provide an equation to forecast sales by plugging in the value of price and advertising.

> Ex: Estimated RENT = 143.67 + .3875*AREA + 89.93*BATHROOMS explains/predicts monthly apartment rents in terms of the area of the apartment in square feet and the number of bathrooms in the apartment.

The variable that we want to estimate (rent) or to forecast (sales) is variously called the **response** variable, the **dependent** variable, or the ***Y*** variable. The variables that we will plug in are called the **predictor** variables, the **independent** variables, or the ***X*** variables. The regression equation will be **linear** – that is, it will be an equation of the form $y = \alpha + \beta_1 x + \beta_2 x_2$. (Later, we shall explore the possibility of nonlinear relationships like $y = \alpha + \beta_1 x + \beta_2 x^2$ or even $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2{}^2$.)

In order to make the presentation of the regression model concrete, I will use the example of Austin apartment rents at length. The *Y* variable will be monthly rent. There will be two predictor variables, the area of the apartment and the number of bathrooms. We have a random sample of 60 Austin apartments for which rent, area, and the bathroom tally are known for each apartment. The objective is to develop a multiple linear regression equation of the form $Rent = \alpha + \beta_1 Area + \beta_2 Bathrooms$, so that the values of *Area* and *Bathrooms* may be plugged into the right-hand side of the equation to get an estimate of *Rent*. I will discuss the regression model, how to do it in Excel, and interpretation of some of the computer output.

Let us begin.

To set the stage for regression, let me pose a *pre*-regression question and recall how we can answer it using the Topic Note on Estimation and Sampling Distributions. I will then show how multiple regression can improve our answer. The next several paragraphs are repeated verbatim from the Topic Note on Simple Linear Regression.

[Begin verbatim repeat.]

Question: What is the mean rent of all Austin apartments?

The question asks for the population mean of an uncertainty distribution. Which distribution? Clearly, it is the distribution that has all Austin apartment rents as its population of outcomes (Population 1) and an unknown probability curve. Since we do not know either the outcomes or the probability curve, we cannot directly calculate the population mean. Therefore, we are uncertain. To reduce the uncertainty, I have drawn a random sample of 60 apartments from the unknown population of Austin apartments and obtained their rents:

Table 1. Rents of 60 randomly selected Austin apartments.

519	530	450	425	470	415	505	470	625	470	659	605
765	580	520	770	700	399	445	470	745	480	650	929
475	995	495	445	450	585	565	700	540	460	750	695
575	565	420	510	785	525	650	455	650	600	455	455
415	620	575	635	485	495	515	550	595	575	430	1050

Each of these 60 draws comes from the same unknown population of rents with the probability of selection governed by the same unknown probability curve. Therefore, the known distribution of these 60 rents should be representative of the unknown distribution of all Austin apartment rents. Therefore, the sample mean of the known 60 rents should be representative of the unknown population mean of all Austin apartments. Therefore, I calculate the sample mean of these 60 rents $\bar{x}$ = \$572.27 and use this value as an estimate of μ = unknown population mean rent of all Austin apartments.[1]

But the estimate, \$572.27, is almost certainly not the true value of μ. We should summarize our uncertainty about the estimate. One way to do that is to address the further question, How much can we expect \$572.27 to differ from the true mean μ? To answer that, we reason that our sample mean is a representative sample mean. So our sample mean probably differs from the true mean μ by about as much as a typical sample mean differs from μ. But how much does a typical sample mean deviate from μ? Answer: By the standard deviation of sample means – that is, by the standard deviation of Population 3, which is $\sigma / \sqrt{n}$. To calculate this, we need σ and n. The sample size n is known: n = 60. But σ, the standard deviation of Population 1, is not known. However, the sample of 60 rents taken from Population 1 are representative of Population 1. So

[1] The logic for this was explained in the Topic Note on Estimation and Sampling Distributions. If you need to review the logic, now would be a good time.

the standard deviation s of the 60 rents is a good estimate of σ. We calculate s = \$140.52. Therefore, $\sigma / \sqrt{n}$ approximately = 140.52 / $\sqrt{60}$ = \$18.14.[2]

Therefore, we estimate the mean rent of the population of all Austin apartments to be \$572.27 and we can expect the actual mean to deviate from this estimate by about ±\$18.14. Furthermore, since the sample size 60 is "sufficiently large", the Population 3 of sample means is approximately normal. Thus, approximately 68% of all sample means are within ±\$18.14 of the true mean μ. Therefore, our sample mean \$572.27 has about 68% chance of being within ±\$18.14 of μ. Wrapping up this assessment of our uncertainty, we say that our best estimate of the mean rent μ of all Austin apartments is \$572.27 and we are about 68% certain that the true mean μ lies in the interval \$572.27 ±\$18.14.

The preceding few paragraphs summarize how we develop a good estimate of the mean rent of all Austin apartments and assess the uncertainty of the estimate. If any of this is not clear to you, please review the appropriate sections of the previous Topic Notes.

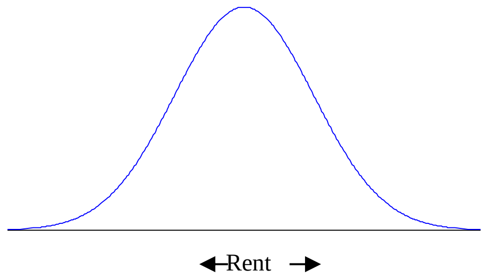
[End verbatim repeat.]

Now I will show how multiple regression can improve the answer.

It is intuitive that large apartments with two bathrooms should usually rent for more than smaller apartments with one bathroom. But the \$572.27 estimate is for the mean rent of all apartments, regardless of their size or number of bathrooms. If we want an estimate for the mean rent of all apartments that have 1,000 square feet of area and two bathrooms, then \$572.27 may not be a very good estimate. It seems intuitive that the mean rent of 1000-square foot apartments with two bathrooms should be more than the mean rent of 800-square foot apartments with one bathroom, which should be more than the mean rent of 600-square foot apartments with one bathroom. To be sure, we should not expect all 1000-square foot apartments with two bathrooms to rent for the same amount – they may still differ in terms of number of bedrooms, age, and other factors that affect rent. Some 1000-square foot apartments with two bathrooms may rent for less than some 800-square foot apartments with one bathroom. But the average 1000-square foot apartment with two bathrooms should command higher rent than the average 800-square foot apartment with one bathroom. It is important to note that even if we know that a group of apartments all have the same area and bathrooms, we are still uncertain about the rent because there are many factors other than area and bathrooms that affect rent. Since we are uncertain about the rents of 1000-square foot apartments with two bathrooms, there is a distribution of rent for 1000-square foot apartments with two bathrooms – with a set of possible outcomes (rents) and a probability curve. The following graph shows one *possible* probability curve for the rents of 1000-square foot apartments with two bathrooms.

[2] The logic of using the standard deviation of the individual values in the sample as an estimate of the standard deviation of the individual values in Population 1 was also explained in the Topic Note on Estimation and Sampling Distributions.

Figure 1

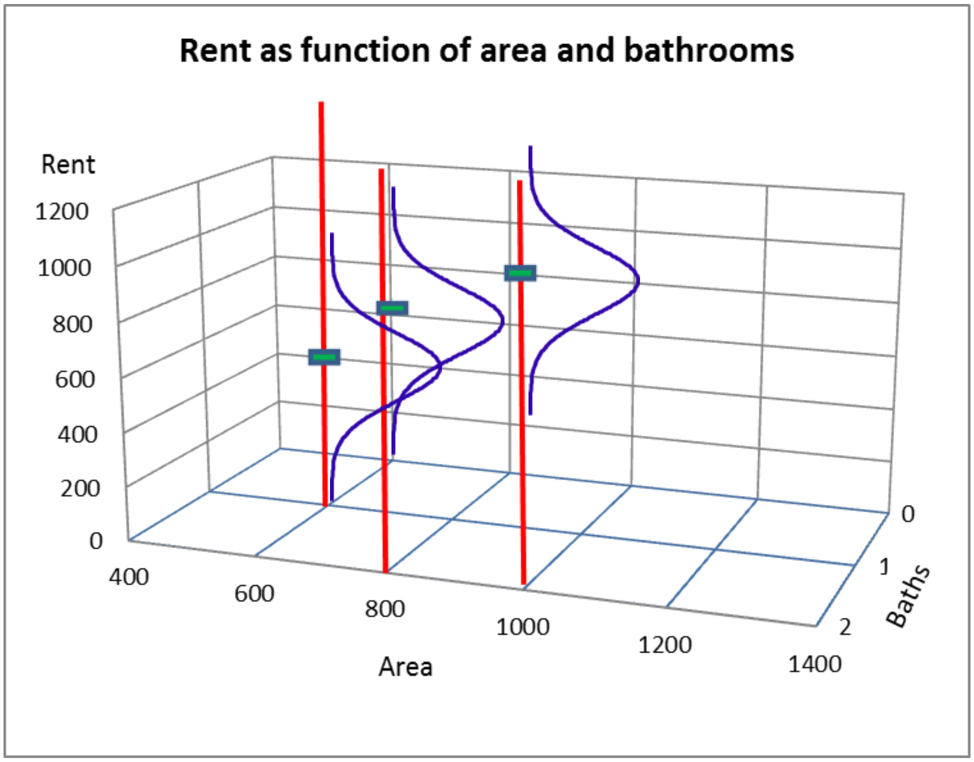


It is important to note that every apartment in this distribution has the same area and bathrooms – 1000 square feet and two bathrooms. But their rents still vary. There is a mean rent, where the positive deviations of higher-rent 1000-square foot apartments with two bathrooms exactly balance the negative deviations of lower-rent 1000-square foot apartments with two bathrooms. And there is a standard deviation, which measures the average magnitude of the deviations of 1000-square foot apartments with two bathrooms from their mean rent. We do not know either the mean or the standard deviation.

Similarly, since we are uncertain about the rents of 800-square foot apartments with two bathrooms, there is a distribution of rent for 800-square foot apartments with two bathrooms – with a set of possible outcomes (rents) and a probability curve. Likewise for rents of 600-square foot apartments with two (or one) bathrooms. Do you understand that for every value of area and number of bathrooms, there is a distribution of rent for apartments of that area and bathroom count – with a set of possible outcomes (rents) and a probability curve?

The following graph shows one *possible* set of probability curves for the rents of 600-sq ft apartments with one bath, 800-sq ft apartments with two baths, and 1000-sq ft apartments with two baths. In order to show them all on the same graph along with area, I have rotated the rent axis from horizontal to vertical, put area on the foreground horizontal axis, and bathroom count on the side axis. Each red pole marks one of the three distributions and is placed at one of the three different combinations of area and bathroom count. The vertical values on each red pole are the possible rents of apartments that have the given area and given number of bathrooms. Although I have used the same shape for all three probability curves, the true curves may have different shapes and they may occupy different vertical positions than I have drawn. Also, apartments have other possible areas and numbers of bathrooms. 600, 800, and 1000 are not the only possible areas, and one and two are not the only possible number of baths. Potentially, every possible combination of area and bathrooms has a distribution of rents that could be shown on the graph. Every distribution has a mean and a standard deviation. In the figure, I have marked the mean of each of the three distributions shown by a green dash. ***The objective of regression is to estimate these means and standard deviations.***

Figure 2



Before discussing how we estimate these means and standard deviations, let me give a few reasons why these estimates are useful.

Having good estimates of the means is useful for two reasons:

- We can see how much the typical rent should be for an apartment of a given area and given number of bathrooms. This gives a benchmark value for renters, developers, tax collectors, appraisers, insurance companies, and others. If an apartment of that area and bathrooms actually rents for more or less than the mean, the apartment should have compensating factors that justify the deviation.
- We can see how the mean rent changes as the area and bathrooms change. This tells us how much we can expect apartment rent to increase if area and/or bathrooms increase by a specified amount. So how much more rent is 100 additional square feet and one bathroom expected to be worth? This is useful information to renters, developers, tax collectors, appraisers, insurance companies, and others.

Having a good estimate of the standard deviation of the distribution is valuable because it summarizes the remaining uncertainty about the rent after taking the area and bathrooms into account. Since the apartments in a specific one of the vertical distributions all have the same area and bathrooms, something other than area and bathrooms must be responsible for their differences in rent. Their standard deviation tells us how much the rent of apartments of given size and bathrooms deviate, on average, from their expected rent. If this typical deviation is large, then factors other than area and bathrooms must be important in explaining why the rent is at the level it is. But if the typical deviation is small, then factors other than area and bathrooms have little to say about why the rent is at the level it is.

Now back to the theme:

The objective of regression is to estimate the means and standard deviations of the distributions of Y for each combination of x's.

How to do this? One approach: You might try to adapt the procedure that we followed for estimating the mean rent of all Austin apartments that was reviewed at the beginning of this set of topic notes above. First get the areas and bathroom counts of the 60 sample apartments. (I have the data and I will show them to you in a moment.) Then divide the sample of 60 apartments into groups. Put the 600-square-foot one-bathroom apartments into one group, the 800-square-foot two-bathroom apartments into a second group, and the 1000-square-foot two-bathrooms apartments into a third group. Then calculate the mean rent of the sample apartments in each group and summarize the uncertainty of the estimate for each group by the standard error of each group's mean.

Unfortunately, this straightforward application does not work well. A big problem is that there is only one apartment with 600 square feet and one bathroom in the sample, none with 800 square feet and two bathrooms, and none with 1000 square feet and two bathrooms. So you might revise the groups to include apartments that are "close" to 600, 800, and 1000 in area – like the ranges 550-650, 750-850, 950-1050. But this still leaves a small number of apartments in each group – too small for averaging to work well in balancing the large deviations in each group and too small to invoke the Central Limit Theorem in assessing the uncertainty of the estimates for each group. The problem is that there are too many vertical distributions to estimate and too few data for each distribution for the basic approach to work well unless we sample a lot more apartments.

The solution adopted by multiple linear regression is to make some strong simplifying assumptions about the distributions and their relationships with each other. These assumptions will allow us to use *all* of the data to estimate each individual distribution and to use *all* of the data in assessing the uncertainty of the estimate for each individual distribution. However, these strong assumptions may not be correct. Fortunately, there are some simple diagnostic tests that we can use to check whether the assumptions are plausible. Often they are plausible. Later in the course, we will see some ways to correct matters if the assumptions are wrong.

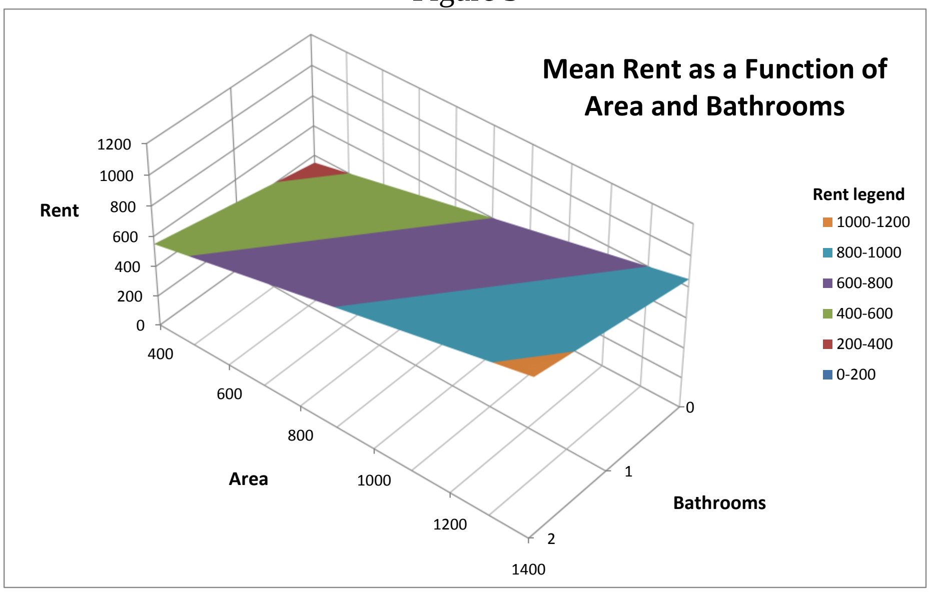
The Assumptions of the Multiple Linear Regression Model

There are four assumptions. These four assumptions are essentially the same as the four assumptions in simple linear regression – just adapted to having more than one predictor. I emphasize that it is important to check whether the assumptions are plausible each time that you use regression. If one or more assumptions are violated, then the conclusions drawn from the analysis may not be correct. Later in the course, I will present some simple graphical tests that you can employ to check the validity of the regression assumptions.

Assumption 1 (L – the Linear assumption): The means of the distributions of rent (of *Y*, in general) lie on a linear equation $\alpha + \beta_1 Area + \beta_2 Baths$ (in general, $\alpha + \beta_1 x_1 + \beta_2 x_2$).

This assumption is illustrated graphically in the figure below. The plane $\alpha + \beta_1 Area + \beta_2 Baths$ that connects the means of each distribution is called the ***true* regression equation**. We do not know what the equation is because we do not know what the true mean of each distribution is. Therefore, we are uncertain about the true regression equation. Even if Assumption 1 is correct, we do not know the values of α, β_1, and β_2. Therefore, there is an uncertainty distribution with outcomes and probabilities for each of α, β_1, and β_2. We will estimate α, β_1, and β_2 in order to estimate the true regression equation. But this ***estimated* regression equation** will deviate from the true regression equation. We will need to assess the uncertainty of the estimates.

Figure 3



There are four important points to understand about the L Assumption:

- It is a strong simplifying assumption. It replaces the difficult task of estimating the mean of rent (*Y*) for *every* combination of area (x_1) and bathrooms (x_2) by the much simpler task of estimating only three parameters – α, β_1, and β_2. We can use all of the data to estimate each of the three parameters. If the L assumption is correct, it eases our task considerably. But it is an assumption and should be tested for plausibility.
- Suppose the L assumption is wrong and we proceed anyway. Then the actual means will not lie on a plane. Nonetheless, the regression analysis will produce a plane and it may be rather far from some of the true means.
- α is the *Y*-intercept of the true regression equation. The plane crosses the vertical axis at the value α. Moreover, α is the value that you get when you plug area = 0 and bathrooms = 0 into the true regression equation. Theoretically, α is the true mean rent of all apartments that have zero area and no bathrooms. Of course, that is a nonsensical idea. There are no apartments with zero area and no baths. This is a warning that regression models may not hold if extrapolated beyond the range of the available data. In other

cases, plugging x_1 = 0 and x_2 = 0 into a regression equation may be quite valid (for example, if *Y* is stock price, and x_1 is earnings per share and x_2 is change in sales from previous year, both of which are sometimes negative). However, we could still interpret α as the value of fixed costs that all apartments must cover in their rents, regardless of their areas or bathrooms.

- β_1 represents the expected change in rent for an increase of one square foot in area – *if* the number of bathrooms remains the same. If β_1 = 0.3875, then additional area is worth 38.75 cents more rent per square foot, on average, when the number of bathrooms does not change. β_2 represents the expected change in rent for an increase of one bathroom – if the area remains the same. If β_2 = 89.929, then an additional bathroom is worth $89.93 more rent, on average, when the area does not change. β_1 and β_2 are variable cost factors. They show the effects on rent of varying the area and/or the number of bathrooms. In most regressions, β_1 and β_2 are of more interest than α. (See more about the important concept of *ceteris paribus* later in this Topic Note.)

Assumption 2 (H – the constant variance or Homoscedasticity assumption): The standard deviations of the distributions of rent (in general, of *Y*) are the same for every combination of values of area and bathrooms (in general, of all *x* predictors).

In the illustrative Figure 2 above, I drew the distributions with the same standard deviation.

The H assumption is satisfied if the uncertainty about rent is the same for every level of area and bathrooms. Since rents vary *within* the vertical distributions for reasons *other than* differences in area and/or bathrooms, this says that the other-than-rent-and-bathrooms factors that affect rent operate in the same manner at every level of area and bathrooms to spread out the rents. This is a fairly strong assumption. Fortunately, there is a simple graphical test for its plausibility. I will discuss this test later in the course.

Suppose the H assumption is wrong and we proceed anyway. Then the uncertainty about rent varies with area and/or bathrooms, sometimes more uncertainty in one combination, sometimes less in another. However, the regression model will produce a common estimate of uncertainty for all of them. This estimate will be too big for some and too small for others. This means that the uncertainty within vertical distributions will be overestimated for some and underestimated for others. So we will be underconfident about rent for some areas and/or bathrooms and overconfident about rent for other areas and/or bathrooms.

In case the H assumption is true, then all of the vertical distributions have the same standard deviation. In this case I will use the symbol σ_e to stand for this one value. (As a mnemonic, you may think of the subscript *e* as standing for "equal".) In case the H assumption is false, then the vertical distributions may have standard deviations that vary, depending upon the values of x_1 and x_2. In that case I will use the symbol σ_{x_1,x_2}, where the subscript x_1, x_2 indicates that the value of the standard deviation may change from one x_1, x_2 combination to another.

Assumption 3 (I – the Independence assumption): The distributions are independent of each other.

This assumption can be interpreted graphically in terms of the above figure. Suppose we sample an apartment with 600 square feet of area and one bathroom, and its rent is more than expected – that is, the rent lies above the true regression plane. If we then look at sampled apartments at other levels of area and bathrooms (say at 800 and 1000 square feet with two baths, for example), the fact that the 600-square foot one bath apartment has higher rent than expected should not affect whether or not the 800 and 1000-square foot two-bath apartments are above or below the regression equation. In general, the rents of two apartments are independent if the value of one rent relative to its mean (above or below the regression equation) gives no information about whether or not the rent of the other apartment is above or below its mean.

If the I assumption is not correct, there can be serious problems with regression analysis. However, if the apartments are drawn as random samples, then the I assumption is automatically satisfied.

Assumption 4 (N – the Normal assumption): The vertical distributions are all normal.

In the above Figure 2, I have shown the three illustrative distributions as normal. You will see later that the N assumption is not important for *estimation of parameters* if the sample size is sufficiently large. The N assumption will be important for *prediction of individual values* – but even there, we have a work-around. Finally, there are graphical and quantitative tests for the plausibility of the N assumption.

The four assumptions say that that the rent distributions are independent and have the same normal shape, but different locations along a linear equation (a plane). You can get any of the distributions just by sliding any one of the distributions around along the plane of the regression equation.

All four assumptions can be tested for plausibility (later).

Here are the rents, areas, and bathrooms of the 60 sample apartments:

Table 2. Rents, areas, and bathroom counts of 60 randomly selected Austin apartments.

Rent	Area	Baths	Rent	Area	Baths	Rent	Area	Baths	Rent	Area	Baths
519	725	1	425	620	1	505	672	1	470	751	1
765	995	2	770	1040	2	445	660	1	480	608	1
475	481	1	445	520	1	565	755	1	460	900	1
575	925	2	510	880	1	650	810	1	600	860	1
415	600	1	635	832	1	515	611	1	575	925	1
530	668	1	470	545	1	470	705	1	659	944	2
580	725	1	700	921	2	470	564	1	650	940	2
995	1421	2	450	577	1	700	1250	2	750	1048	2
565	672	1	785	1080	2	455	512	1	455	474	1
620	1025	1	485	710	1	550	630	1	430	700	1
450	781	1	415	605	1	625	850	1	605	921	1
520	800	1	399	680	1	745	1156	2	929	1229	2
495	870	1	585	730	1	540	932	1	695	896	2
420	700	1	525	687	1	650	755	1	455	630	1
575	800	1	495	703	1	595	1093	2	1050	1864	2

How to Run Multiple Regression in Excel

[Essentially the same as in simple linear regression]

- First, set up your data with the predictor (X) variable(s) in contiguous columns next to each other. The response (Y) variable should also be in a single column. Put labels (variable names) at the tops of the columns. Be sure that you have activated the *Analysis ToolPak* and *Analysis ToolPak VBA* add-ins, if not already activated.[3]
- Click the Data tab to open the Data ribbon in Excel.
- Click the Data Analysis option in the Analysis box at the right on the ribbon. Among the options, scroll down and select Regression and click OK.
- In the Regression dialogue box, fill in the Y range and the X range (all of the predictor columns).
- Click the "Labels" box if you have column header names.
- Select the output range to store the computer output. This is the cell that will be the upper left hand corner of the output.
- Check the "Residuals" box and perhaps the "Residual Plots" box.
- Then OK.

Missing data: Any row that has one or more missing values will be omitted from the regression analysis – the whole row will be omitted.

How to Run Multiple Regression with the StatTools Add-in[4]

[Essentially the same as in simple linear regression]

- *First, define your dataset with the StatTools* Data Set Manager. This utility allows you to select a rectangular set of data in a spreadsheet and apply a name to it so that StatTools will know where the data are that you want to analyze. Unlike Excel, the X predictors do not need to be contiguous, but each variable needs to be in a single column. Put descriptive labels at the tops of the rectangle. StatTools will use those labels, by default, as the names of your variables. Click on StatTools → Data Set Manager. StatTools will guess the extent of your data, but you can type in or select the correct range if StatTools guesses incorrectly. Then fill in the dialogue box with the range of your data and give it a name. You may have more than one StatTools data set.
- To start the regression, click on StatTools → Regression & Classification → Regression… . Then fill in the dialogue box:
- For Regression Type, select "Multiple" from the pull-down menu options.
- For Data Set, select the name of the set containing your data from the pull-down menu options – assuming you have defined your dataset per the first bullet item above. (If you omitted the Data Set Manager step, go back and do it.)
- Select response and predictor variables by clicking their names. "D" means dependent (response or Y) variable. "I" means independent (predictor or X) variable(s). You can select

[3] Click the File tab, then Options -> Manage: Excel Add-ins -> Go. In the Add-ins dialogue box that pops up, be sure the Analysis ToolPak and Analysis ToolPak – VBA boxes have checks. Then OK.

[4] StatTools is part of the Palisades DecisionTools Suite. This suite of management science add-ins for Excel is available to you for download from the McCombs site given on your syllabus.

only one “D” dependent variable, but you can select multiple “I” independent predictors for multiple regression.

- In the “Graphs” section, click the box “Residuals vs Fitted Values”. This will help you verify L and H and thus help verify the assumptions of the regression model. This is the most important option in this section of the dialogue box, but the others can also be helpful on occasion. For example, a plot of the residuals in time order can be useful if the data are a time series. If time is one of the predictors, you can get this plot by checking “Residuals vs X values”. If time is not a predictor, you can get this plot anyway by StatTools → Time Series & Forecasting → Time series graph and selecting the Residual column.
- Click OK. Your regression output will be put into a new Excel workbook.

Missing data: Any row that has one or more missing values will be omitted from the regression analysis – the whole row will be omitted.

Interpreting Regression Output

The objective of regression is to estimate the mean rent of all apartments of a given area and number of bathrooms and to assess the uncertainty of the estimate. Let us assume for now that the four regression assumptions (L,H,I N) are all satisfied. (We will return later to the issue of testing the assumptions.) Enter the rent, area, and bathrooms data shown above into Excel and run the regression. The following table shows the primary table from the StatTools output. Excel’s Data Analysis add-in provides the same output, but in a little different format.

It is easy to become overwhelmed with the amount of output that regression provides. I have labeled the output with 15 sets of numbers that will be interpreted – eventually – but only a few right now. Let us begin with the basic parts, #1, #2, and #5.

Figure 4. Multiple regression output (Y = rent, X_1 = area, X_2 = bathrooms)

Summary	Multiple R	R-Square	Adjusted R-Square	StErr of Estimate
	0.8948	0.8007	0.7937	63.83
	4	3		2

ANOVA Table	Degrees of Freedom	Sum of Squares	Mean of Squares	F-Ratio	p-Value
Explained	2	932,883	466,441	114.4998	< 0.0001
Unexplained	57	232,203	4,074		
	12	11	13	14	15

Regression Table	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95% Lower	Upper
Constant	143.6693	29.5135	4.8679	< 0.0001	84.5696	202.7689
Area	0.3875	0.0498	7.7773	< 0.0001	0.2877	0.4872
Bathrooms	89.9290	27.7507	3.2406	0.0020	34.3592	145.4989
	1	5	6	7	8	

#1. The estimated regression equation.

The numbers at **1** in Figure 4 are the regression estimates of the intercept and the two slopes. The estimate of the intercept is $\boldsymbol{a}$ **= 143.6693**. The estimate of the area slope is $\boldsymbol{b_1}$ **= 0.3875** and the estimate of the bathrooms slope is $\boldsymbol{b_2}$ = **89.9290**. These are not the true values of the intercept and slopes. That is why I use the Roman characters $\boldsymbol{a}$, $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ to refer to the estimates, in order to distinguish them from the corresponding true intercept and true slope, denoted by the Greek characters α , β_1 and β_2 .

Thus the estimated regression equation is **143.6693 + 0.3875 *Area* + 89.9290 *Bathrooms***. Since the intercept results from plugging in *Area* = 0 and *Baths* = 0, we might interpret the intercept as saying that an apartment of zero area and no bathrooms is estimated to rent for \$143.6693 – a nonsensical interpretation for these variables, [5] but perhaps not for other variables. We could more reasonably interpret the intercept as a kind of fixed cost for apartments, apart from the variable costs resulting from different areas and numbers of bathrooms.

Ceteris paribus. In multiple regression, the slope coefficients are interpreted ***ceteris paribus***[6], i.e. holding all other model predictors constant.[7] ***This is the major difference between multiple-predictor regression and one-predictor regression*****.**

The coefficient of AREA is interpreted as saying that each additional square foot of apartment space costs about \$0.3875 more, *among apartments with the same number of bathrooms*.[8] Similarly, the coefficient of BATHROOMS means that each additional bathroom is estimated to increase the rent by about \$89.93, *among apartments with the same area*. If the values of both predictors change, then the estimated effect on rent is obtained by adding the effects of both coefficients.

Let us extend the implications of the *ceteris paribus* interpretations a bit further:

- If a landlord builds a 100-sq ft addition to an apartment and does not turn the addition into a bathroom, the rent of the apartment is expected to be 100*.3875 = \$38.75 more per month.
- If a landlord converts existing space in an apartment into a bathroom, but makes no addition to the area, then the additional rent is expected to be \$89.93.
- If a landlord builds a 100-sq ft addition and turns it into a bathroom, the additional rent is expected to be \$38.75 + \$89.93 = \$128.68.
- If we compare apartment A with apartment B and both have the same number of bathrooms (it does not matter whether each has one bathroom, or each has two bathrooms, or three, etc.) but A has 100 additional square feet, then we expect A to rent for \$38.75 more than B.
- If apartments A and B have the same area, but A has one more bathroom than B, then we expect A to rent for \$89.93 more than B.
- If apartment A has 100 more square feet and one more bathroom than B, then we expect A to rent for \$38.75+\$89.93 = \$128.68 more than B. (Note also that the number of bedrooms, age, etc. are not being held constant.)

[5] There is no apartment in the dataset with area near zero. The smallest apartment has 474 square feet and one bathroom. Avoid extrapolating regression results beyond the range of available data. The model may not hold there.

[6] *Ceteris paribus* is Latin for "all other things being the same."

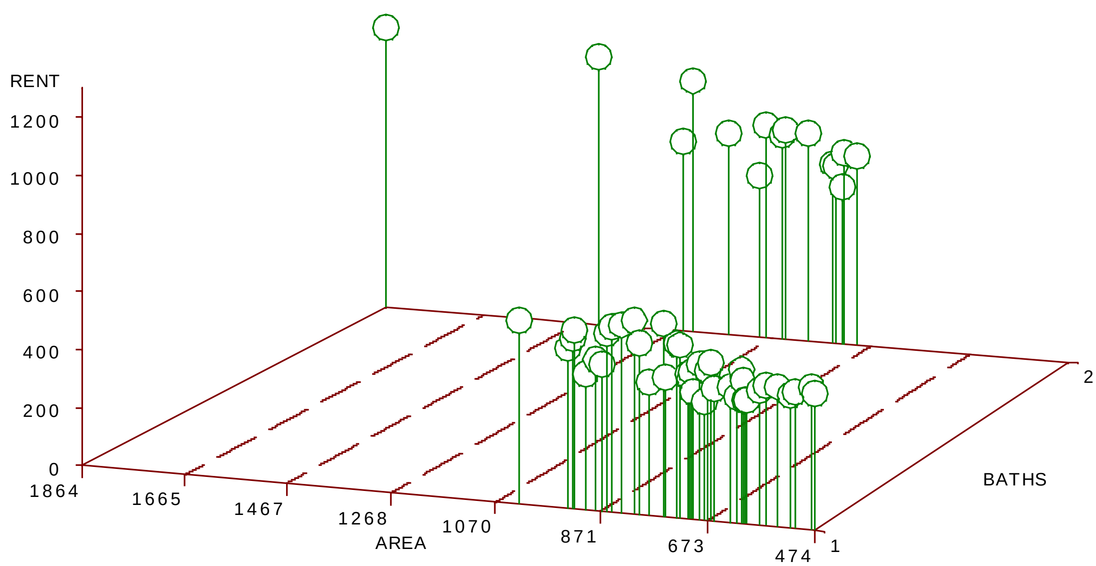
[7] Note that *ceteris paribus* holds fixed the values of all model predictors other than the one being examined, but variables not used explicitly in the model (and whose effects may be manifest in the residuals) are NOT held fixed. For example, in the regression of RENT on AREA and BATHROOMS, the value of AGE is not held constant, nor is the color of paint used in the apartment.

[8] The italicized phrase is the *ceteris paribus* qualification to the interpretation of the coefficient of AREA.

- If a landlord adds 100 extra square feet to an existing apartment, but we do not know what use is made of the extra space, then how much do we expect the additional rent to be? In this case, all other things are not being held constant -- and we do not know if the extra space will contain an extra bathroom -- so we must use the *simple* regression of RENT on AREA, from the Topic Notes on simple regression, to estimate 100*.5050 = $50.50 as the expected increase in monthly rent.[9]
- If a landlord adds a bathroom to an existing apartment, but we do not know whether the space for the bathroom came out of existing space or was added on, we should estimate the expected extra rent by running a simple regression of RENT on BATHROOMS and using the coefficient of BATHROOMS as our estimate. This regression has not been run in our examples to date.
- If we compare apartment C with apartment D and we do not know how many bathrooms each has, but C has 100 additional square feet, then we expect C to rent for $50.50 more than D. (Again, *ceteris paribus* does not apply.)
- If C has one more bathroom than D and we do not know the areas of C and D, the additional rent of C would be estimated by the coefficient of BATHROOMS in a simple regression of RENT on BATHROOMS. (Again, *ceteris paribus* does not apply.)

To visualize the regression equation, we need three-dimensional plotting capability. Here is a three-dimensional plot of the data (not done with Excel), with the rent of the apartments represented by the heights of the small bubbles above the plane, which shows the AREA and BATHROOMS:

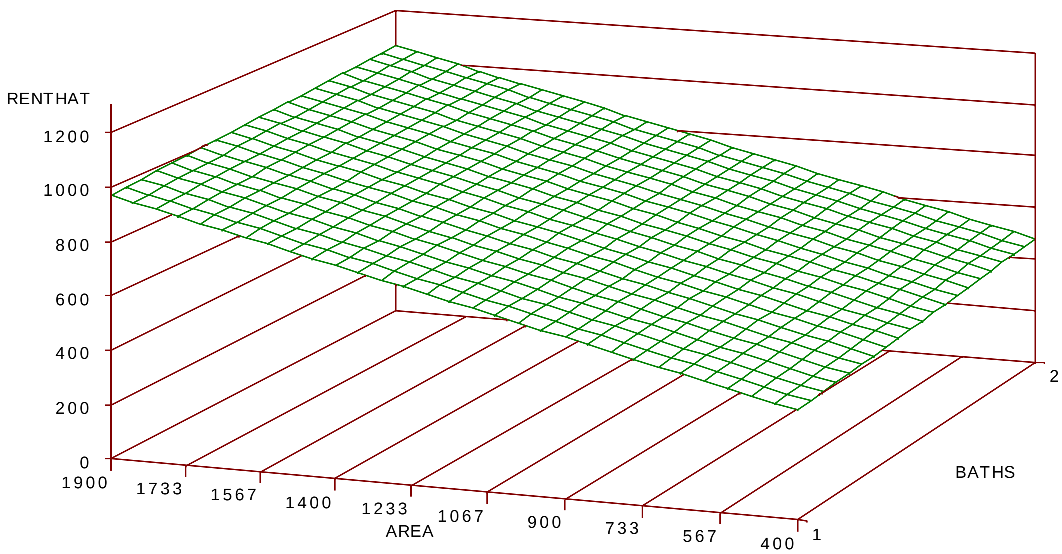
Figure 5



[9] In the simple regression, the higher coefficient of AREA reflects the fact that in some apartments with extra space, that space will be used for a high value-adding extra bathroom, and in some apartments it will not -- the higher 0.505 coefficient represents, in a sense, a blending of the effects of area and bathrooms. In the multiple regression, this blending does not occur – the effects of the two predictors are separate: The 0.3875 coefficient of AREA reflects the "pure" value of extra space that is not used for another bathroom, because bathrooms are held constant. Likewise, 89.93 is the "pure" effect of an extra bathroom when the area is not increased.

And here is a three-dimensional plot of the regression equation (plane), RENT = 143.6693 + 0.387459 * AREA + 89.9290 * BATHROOMS :

Figure 6



In Figure 6, above, the wireframe plane is composed of crossing lines that geometrically display the *ceteris paribus* effect of the predictors on rent. There are two sets of lines in the wireframe, called *warp* and *woof* in weaving. The warp lines run parallel to the AREA axis; the woof lines run parallel to the BATHROOMS axis. The slope of the warp lines is the coefficient of AREA (0.3875); the slope of the woof lines is the coefficient of BATHROOMS (89.93). These properties are shown clearly in the next two figures, which isolate the warp and woof lines.

Figure 7

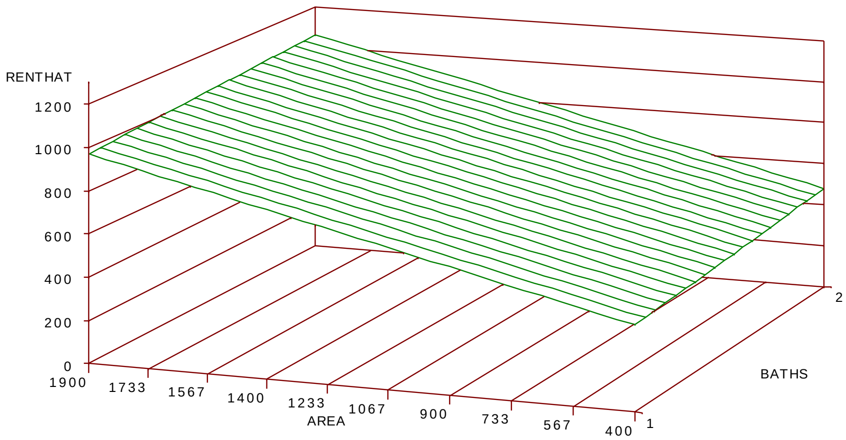


Figure 7 shows the warp lines. These lines show the effect on rent of changing area while holding bathrooms constant. Each warp line corresponds to a different value of BATHROOMS (never mind that only whole numbers are logically possible values for BATHROOMS). The number of bathrooms is constant as you move from right to left along any given warp line. All of the warp lines have the same slope, namely the coefficient of area, \$0.3875.[10]

Figure 8

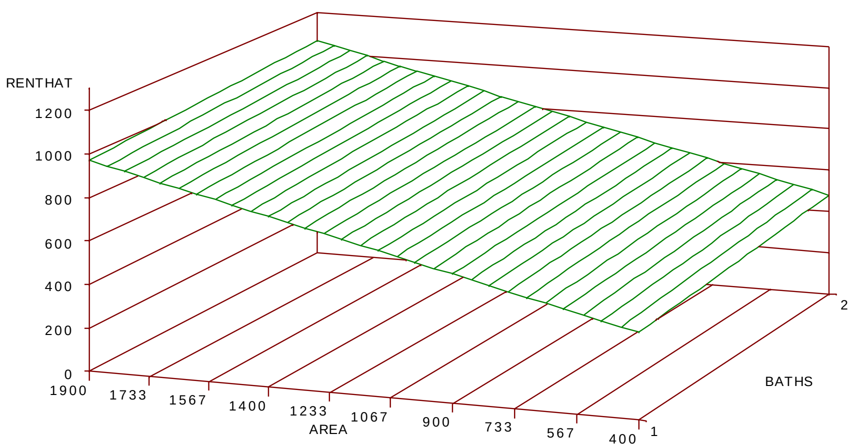


Figure 8 shows the woof lines. These lines show the effect on rent of changing bathrooms while holding area constant. Each woof line corresponds to a different value of AREA. The area is constant as you move from front to back along any given woof line. All of the woof lines have the same slope, namely the coefficient of bathrooms in the regression equation, \$89.93.[11]

---

[10] Math note: The slope is also the partial derivative of the regression plane with respect to AREA. The slope is the directional derivative of the plane in the direction of increasing AREA, perpendicular to the BATHROOMS axis.
[11] Math note: The slope is also the partial derivative of the regression plane with respect to BATHROOMS. The slope is the directional derivative of the plane in the direction of increasing BATHROOMS, perpendicular to the AREA axis.

Examples:

- To estimate the mean rent of all 600-square foot apartments having one bathroom, we plug in 600 and 1: 143.6693 + 0.3875 * 600 + 89.9290 * 1 = 466.10;
- To estimate the mean rent of all 800-square foot apartments having two baths, we plug in 800 and 2: 143.6693 + 0.3875 * 800 + 89.9290 * 2 = 633.53;
- To estimate the mean rent of all 1000-square foot apartments having two baths, we plug in 1000 and 2: 143.6693 + 0.3875 * 1000 + 89.9290 * 2 = 711.03.

See discussion in #2, next up, for assessing the uncertainty of these estimates.

#2. "Standard error of the estimate" [called just "standard error" in Excel's output]

In Figure 4, the number at **2** is the estimate of the common standard deviation σ_e referred to in the H assumption. This is not σ_e itself (which is unknown), but an estimate of σ_e. It estimates the average deviation of the individual rents from the mean rent in each of the vertical distributions depicted in Figure 2. The "standard error of the estimate" is $63.83. We can interpret this as saying that use of area and bathrooms in regression allows us to estimate an apartment's rent to within about $63.83 on average. Contrast this number with the (ordinary) standard deviation of RENT, which at $140.52 is more than twice as large. Also contrast it with the standard error of the estimate from the simple regression of RENT on AREA, which is $68.86. Thus, on average, knowing the AREA and BATHROOMS of an apartment enables us to estimate its RENT much more accurately than not knowing either its AREA or its BATHROOMS, and somewhat more accurately ($5.03 more accurately) than knowing only its AREA:

- If we do not know the area or number of bathrooms of an apartment, the sample allows us to estimate its rent to within ±$140.52 on average.
- If we know the area of an apartment, the sample allows us to estimate its rent to within ±$68.86 on average.
- If we know both the area and the number of bathrooms, the sample allows us to estimate its rent to within ±$63.83 on average.

Knowing the area and number of bathrooms of an apartment reduces the uncertainty of the rent by more than half.

Moreover, we can go further and say that about 68% of the rents in each vertical distribution are within $63.83 of the mean – as long as the N assumption is true. If the N assumption is not true, we can still say that the average apartment in a vertical distribution deviates by about $63.83 from the mean, but we cannot go further and say that 68% of the apartments deviate by less than $68.86 from the mean.

The name that StatTools assigns to this quantity ("standard error of the estimate") is unfortunate. The name that Excel assigns to the quantity ("standard error") in its output is even more unfortunate. For those names suggest that this quantity is a "standard error" in the sense of a standard deviation of a sampling distribution (Population 3).[12] It is NOT. It is a Population 1 estimate: $63.83 estimates the uncertainty of individual apartment rents due to factors other than

[12] Most other software publishers call this quantity "root mean-square error", which is both more accurate and less confusing. I am sorry that software publishers cannot settle upon one clearly descriptive name. But that is the way things are.

area and bathrooms. This is the estimated variability of the subsets of Population 1 up and down the red poles in Figure 2. It is important to note that the so-called "standard error of the estimate" does NOT assess the uncertainty of the estimates of *mean* rents of 600-sq ft apartments with one bath, of 800-sq ft apartments with two baths, and of 1000-sq ft apartments with two baths (these means are the green tabs in Figure 2).

The uncertainty of individual apartments within the red poles of Figure 2 is called *prediction error* and is a Population 1 concept; the uncertainty of the green tabs on the red poles of Figure 2 is called *estimation error* and is a Population 3 concept. These two concepts are fully explained in a Topic Note on prediction and estimation. For now, you should note that $63.83 assesses prediction error and applies to Population 1 estimates. We need a different number to assess estimation error. I will discuss this next. My discussion should recall for you the *pre*-regression development of the Three Populations concept in the Topic Note on Estimation and Sampling Distributions – and also the extension of the concept to simple regression in the Topic Note on Simple Linear Regression. If the ideas from these Topic Notes are fuzzy to you, I suggest that you review them before continuing. Since the ideas are so similar in multiple regression, I will only sketch their development. I will illustrate with the case of 1000-square foot apartments having two bathrooms:

Based upon my sample of 60 apartments, the estimate of mean rent of all 1000-square foot apartments with two bathrooms is the plug-in estimate: 143.6693 + 0.3875 * 1000 + 89.9290 * 2 = 711.03. If I had drawn a different sample of 60 apartments, I would have had a different set of rents, areas, and bathrooms. Consequently, the estimated regression equation would have been different. So, when I plug in area = 1000 and bathrooms = 2, I would get a different estimate of the mean rent of 1000-square foot apartments having two bathrooms. Imagine calculating the regression equation for every *possible* sample of 60 apartments. Each regression equation has the form $Rent = a + b_1 * Area, + b_2 * Baths$ where a, b_1 and b_2 vary from sample to sample. Then plug *Area* = 1000 and *Baths* = 2 into each equation and calculate the estimates of mean rent of 1000-square foot apartments having two bathrooms. Collect all of these estimates together in one pile. That pile is a Population 3. It is the sampling distribution of the estimates of the mean of 1000-square foot apartments having two bathrooms. This sampling distribution has three important properties:

- (Property 1 – Central Limit Theorem) The distribution is (approx.) normal if the sample size 60 is sufficiently large (it is – 30 or more is the rule of thumb).
- (Property 2) The mean of the sampling distribution equals the true mean rent $\alpha + 1000\beta_1 + 2\beta_2$ of all apartments having 1000 square feet and two bathrooms.
- (Property 3) The standard deviation of the sampling distribution is approximately equal to the "*standard error of the estimate*" $\div \sqrt{n}$.[13,14] Since $n = 60$, then the "*standard error of the estimate*" $\div \sqrt{n} = 63.83 \div \sqrt{60} = 8.24$.

We can now fully assess the uncertainty of the estimates calculated in the examples at the end of the discussion of #1 as follows:

[13] Note how this formula parallels the formula for the standard error of the mean from the pre-regression setting of the Topic Note on Estimation and Sampling Distributions. All three Properties are nearly exact parallels of the 3 Properties from that Topic Note.

[14] This is an approximation that is too small. It is exact if the *x*-values are equal to their means. But it is substantially too small if the *x*-values are far from their means. There is an exact formula (not presented here, nor included in Excel or StatTools calculations).

- To estimate the mean rent of all 600-square foot apartments having one bathroom, we plug in 600 and 143.6693 + 0.3875 * 600 + 89.9290 * 1 = 466.10 ± 8.24 on average, with 68% confidence that the true mean will lie therein;
- To estimate the mean rent of all 800-square foot apartments having two baths, we plug in 800 and 2: 143.6693 + 0.3875 * 800 + 89.9290 * 2 = 633.53 ± 8.24 on average, with 68% confidence that the true mean will lie therein;
- To estimate the mean rent of all 1000-square foot apartments having two baths, we plug in 1000 and 2: 143.6693 + 0.3875 * 1000 + 89.9290 * 2 = 711.03 ± 8.24 on average, with 68% confidence that the true mean will lie therein.

I hope that you understand the difference between the above three estimates of means and the following three estimates of individual apartments – why the estimates are the same, but the +/- uncertainties are different. The above three are *estimation* uncertainties; the following three are *prediction* uncertainties:

- To predict the rent of a 600-square foot apartment having one bathroom, we plug in 600 and 143.6693 + 0.3875 * 600 + 89.9290 * 1 = 466.10 ± 63.83 on average, with 68% confidence that the true rent will lie therein if **N** holds;
- To predict the rent of a 800-square foot apartment having two baths, we plug in 800 and 2: 143.6693 + 0.3875 * 800 + 89.9290 * 2 = 633.53 ± 63.83 on average, with 68% confidence that the true rent will lie therein if N holds;
- To predict the rent of a 1000-square foot apartment having two baths, we plug in 1000 and 2: 143.6693 + 0.3875 * 1000 + 89.9290 * 2 = 711.03 ± 63.83 on average, with 68% confidence that the true rent will lie therein if **N** holds.

#5. Standard error of the coefficients

In Figure 4, the numbers at **5** assess the uncertainty of the estimates of the intercept and two slopes. The development of this concept in regression and the associated Properties is analogous to the *pre*-regression development of the Three Populations in the Topic Note on Estimation and Sampling Distributions, to the simple regression extension in the Topic Note on Simple Linear Regression, and to the discussion just presented above in #2. If those ideas are fuzzy to you, I suggest that you review them before continuing. Since the ideas are so similar in regression, I will only sketch their development:

If I had drawn a different sample of 60 apartments, I would have had a different set of rents, areas, and bathrooms. Consequently, the estimated regression equation would have been different. That is, the equation would have a different intercept and different slopes. Imagine calculating the regression equation for every *possible* sample of size 60. Each regression equation has the form $Rent = a + b_1 * Area, + b_2 * Baths$ where a, b_1 and b_2 vary from sample to sample. Put all of these b_1 slopes together in one pile, one b_1 slope for every sample. That pile is a Population 3: the sampling distribution of the estimated area slopes. It has three important properties:

- (Property 1 – Central Limit Theorem) The distribution is (approx.) normal if the sample size 60 is sufficiently large (it is – 30 or more is the rule of thumb).
- (Property 2) The mean of the sampling distribution of b_1 equals the true slope β_1.

- (Property 3) The standard deviation of the sampling distribution is approximately equal to the value printed on the output at #5, namely 0.0498.[15]

Similarly, if you put all of the b_2 slopes together in one pile, one b_2 slope for every possible sample, that pile is a Population 3. It is the sampling distribution of the estimated bathroom slopes. It has three important properties:

- (Property 1 – Central Limit Theorem) The distribution is (approx.) normal if the sample size 60 is sufficiently large (it is – 30 or more is the rule of thumb).
- (Property 2) The mean of the sampling distribution of b_2 equals the true slope β_2.
- (Property 3) The standard deviation of the sampling distribution is approximately equal to the value printed on the output at #5, namely 27.7507.[16]

Similarly, if you put all of the intercepts *a* together in one pile, one intercept for every possible sample, that pile is a Population 3. It is the sampling distribution of the estimated intercepts. It has three important properties:

- (Property 1 – Central Limit Theorem) The distribution is (approx.) normal if the sample size 60 is sufficiently large (it is – 30 or more is the rule of thumb).
- (Property 2) The mean of the sampling distribution equals the true intercept α.
- (Property 3) The standard deviation of the sampling distribution is approximately equal to the value printed on the output at #5, namely 29.5135. [17]

Using these Properties, we can fully assess the uncertainty of the coefficient estimates printed on the output at #2.

- We estimate the true slope of area to be 0.3875 ± 0.0498, on average, with 68% confidence that the true slope will lie therein;
- We estimate the true slope of bathrooms to be 89.9290 ± 27.2707, on average, with 68% confidence that the true intercept will lie within;
- We estimate the true intercept to be 143.6693 ± 29.5135, on average, with 68% confidence that the true intercept will lie therein.

[15] There is a formula for calculating this quantity, but you are not responsible for knowing that formula.
[16] There is a formula for calculating this quantity, but you are not responsible for knowing that formula.
[17] There is a formula for calculating this quantity, but you are not responsible for knowing that formula.

SUMMARY

The purpose of multiple regression is to estimate the mean *Y* for each given combination of all values of predictors and to estimate the uncertainty (standard deviation) of *Y* for each given combination of all values of predictors. In order to avoid having too many means and standard deviations to make reasonable estimates from the data, regression imposes four strong modeling assumptions on the data:

- **L** – As the predictor values vary, the means of the distributions of *Y* fall on a linear equation: *mean of Y* = $\alpha + \beta_1 x_1 + \beta_2 x_2$ (for two predictors).
- **H** – The standard deviations of the distributions of *Y* are the same for every combination of predictor values.
- **I** – The distributions of *Y* at all combinations of predictor values are independent of each other.
- **N** – The distributions of *Y* at all combinations of predictor values are normal.

If these four assumptions hold, then only a limited number of parameters need to be estimated. If there are two predictors, we must estimate α, β_1, β_2 and the common standard deviation σ_e of the *Y*'s at each combination of predictor values – instead of having to estimate a different mean and standard deviation at every combination of predictor values. This is a great economy. The means can be estimated by plugging x_1 and x_2 into the estimated regression line $a + b_1 x_1 + b_2 x_2$, which re-uses the estimates *a*, b_1, and b_2 of intercept and slopes repeatedly every time the mean of *Y* at another combination of x_1 and x_2 is desired.

Moreover, estimates of the parameters α, β_1, β_2 and the common standard deviation σ_e of the *Y*'s at each *x*, do not need to be calculated by hand – they can be read off computer output.

I explained how to run regression with Excel and with the StatTools add-in.

Regression output has many parts. In this Topic Note, I made a start on interpreting the output. In particular,

- (#1) The coefficients in the regression table output give the estimates *a*, b_1, and b_2 of intercept and slopes α, β_1, β_2. These estimates are used to calculate the plug-in estimates of the mean of *Y* at any given x_1 and x_2: $a + b_1 x_1 + b_2 x_2$.
- (#2) The "standard error of the estimate" provides the estimate of the common standard deviation σ_e of the *Y*'s at each combination of predictor values. This has two uses:
    - to assess the *remaining* uncertainty about the *value* of *Y* after taking your knowledge of the predictor values into account; and
    - (after dividing by $\sqrt{n}$) to assess the uncertainty about the *mean* of *Y* at a given combination of predictor values.
- (#5) The standard errors of the coefficient estimates assess the uncertainty of the sampling distribution of the intercept and slope estimates. These figures say how far you can expect the intercept and slope estimates *a*, b_1, and b_2 to be from the true intercept and slopes α, β_1, β_2.

Most of the features of multiple linear regression are analogous to the corresponding features of simple linear regression. The most important difference is the **ceteris paribus** interpretation of coefficients in multiple regression.

- In the true multiple regression $\alpha + \beta_1 x_1 + \beta_2 x_2$, the value of β_1 is the effect of a unit increase in x_1 on the mean of *Y when* x_2 *remains unchanged*. Similarly, the value of β_2 is the effect of a unit increase in x_2 on the mean of *Y when* x_1 *remains unchanged*.
- Likewise, in the multiple regression estimate $a + b_1 x_1 + b_2 x_2$, the value of the estimate b_1 is the estimated effect of a unit increase in x_1 on the mean of *Y when* x_2 *remains unchanged*. Similarly, the value of the estimate b_2 is the estimated effect of a unit increase in x_2 on the mean of *Y when* x_1 *remains unchanged*.
- In simple regression, the true $\alpha + \beta x$ or the estimate $a + bx$, all other variables than *x* are free to change. In multiple regression, all other variables than x_1 and x_2 are free to change, but the predictor variables x_1 and x_2 remain unchanged in the interpretation.