Regression Analysis III ... F Tests for SLR; MLR

This session presents one additional method of inference for SLR and then begins discussion of Multiple Linear Regression.

Hypothesis Testing in SLR ... F Tests

There is a 2nd method of testing $H_0 : \beta_1 = 0$ in SLR that (only in the special case of SLR) turns out to be equivalent to two-sided t testing, but provides some additional motivation/intuition beyond that provided by t testing. This is built on the "sums of squares" used in Session 6 to define R^2. Computations for the test are usually organized in a so-called "ANOVA table" (an ANalysis

Of VAriance table). There are many such tables in applied statistics. The particular one appropriate here is as below.

ANOVA Table (for SLR)

Source	SS	df	MS	F
Regression	SSR	1	$MSR = \frac{SSR}{1}$	$F = \frac{MSR}{MSE}$
Error	SSE	$n-2$	$MSE = \frac{SSE}{n-2}$	
Total	$SSTot$	$n-1$		

Recall from Session 6 that

$$\begin{aligned} SSTot &= (n-1)\, s_y^2 = \sum (y_i - \bar{y})^2 \\ SSE &= \sum (y_i - \hat{y}_i)^2 \\ SSR &= SSTot - SSE \\ R^2 &= \frac{SSR}{SSTot} \end{aligned}$$

and note from the previous session that $s = \sqrt{SSE/(n-2)}$ so that in the present notation

$$MSE = s^2$$

In the ANOVA table, the statistic $F = MSR/MSE$ can be used to test $H_0 : \beta_1 = 0$. If the least squares line is an effective representation of an (x, y) data set, SSE should be small compared to $SSTot$, making SSR large, leading to large F.

Example In the Real Estate Example, the ANOVA table is

ANOVA Table (for SLR)

Source	SS	df	MS	F
Regression (size)	727.85	1	727.85	58.43
Error	99.65	8	12.46	
Total	827.50	9		

and just as expected, the coefficient of determination (from Session 6) is

$$R^2 = \frac{727.85}{827.50} = .88$$

and the square of the estimated standard deviation of y for any fixed x (from Session 7) is

$$s^2 = (3.53)^2 = 12.46 = MSE$$

It is intuitively reasonable that large observed F should count as evidence against $H_0 : \beta_1 = 0$ in favor of the alternative that $H_a : \beta_1 \neq 0$, i.e. that the predictor x is of some help in describing/predicting/explaining y. But what is "large"? To answer this we need to introduce a new set of distributions and state a new probability fact. This is

Under the normal SLR model, if $H_0 : \beta_1 = 0$ is true, then

$$F = \frac{MSR}{MSE}$$

has the (Snedecor) F distribution with ("numerator" and "denominator") degrees of freedom $df_{\text{num}} = 1$ and $df_{\text{denom}} = n - 2$.

F distributions are distributions over positive values and are right skewed. Several are shown below.

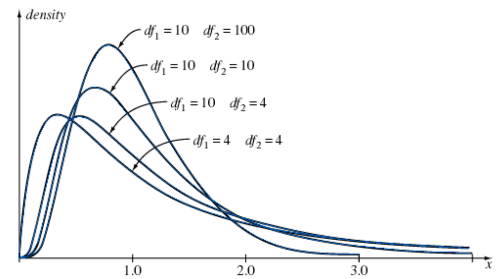

Figure 1: Several F Distributions

Tables E in MMD&S beginning on page T-12 provide upper percentage points of F distributions. One locates an appropriate numerator degrees of freedom across the top margin of the table, finds an appropriate denominator degrees of freedom and a desired (small) right tail probability down the left margin, and then reads an appropriate cut-off value out of the body of the table.

Example In the Real Estate Example, the observed value of $F = 58.43$ exceeds the upper .001 point of the F distribution for 1 and 8 degrees of freedom (25.41) shown on page T-13 of MMD&S. The p-value for testing $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$ using this F test is less than .001. There is clear evidence in the data that x is of help in predicting/explaining y.

It appears to this point that we have two different ways of testing $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$ (the t test of the previous session and this F test). But in

fact these two tests produce the same p-values. Why? As it turns out,

$$\begin{aligned}(\text{the value of the } t \text{ statistic for } H_0 : \beta_1 = 0)^2 &= \\ \text{the value of the } F \text{ statistic for } H_0 &: \beta_1 = 0\end{aligned}$$

while values in the first column of the F table (F percentage points for $df_{\text{num}} = 1$) are the squares of values in the t table (with $df = df_{\text{denom}}$). For example

$$\begin{aligned}(1.86)^2 &= (\text{upper 5\% point of the } t \text{ distribution with } df = 8)^2 \\ &= 3.46 \\ &= (\text{upper 10\% point of the } F \text{ distribution with } df = 1, 8)\end{aligned}$$

A reasonable question is "If the t test and F test are equivalent, why present both?" There are two responses to this. In the first place, there is the helpful intuition about SLR provided by the ANOVA table. And there is also the fact that the t and F tests have important and genuinely different generalizations in multiple linear regression.

Exercise For the small fake SLR data set, "by hand" make the ANOVA table for testing $H_0 : \beta_1 = 0$.

Multiple Linear Regression ... Introduction and Descriptive Statistics

Real business problems usually involve not one, but instead many predictor variables. That is, some important response variable y may well be related to one of more or k predictors

$$x_1, x_2, \ldots, x_k$$

The point of Ch 11 of MMD&S is to generalize the SLR analysis of Chapters 2 and 10 based on an approximate relationship

$$y \approx \beta_0 + \beta_1 x$$

to the more complicated MLR (multiple linear regression) context where

$$y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

It turns out that everything from Chs 2 and 10 (Sessions 6 and 7 and the first part of this one) can be generalized to this more complicated context.

Example The Real Estate Example actually had not one but two predictor variables. The second predictor besides home size, was what we'll call x_2 that was some measure of "condition" of the home, stated on a 1-10 scale, 1 being worst condition and 10 being best. The full data set for $n = 10$ homes was

Selling Price, y	Size, x_1	Condition, x_2
60.0	23	5
32.7	11	2
57.7	20	9
45.5	17	3
47.0	15	8
53.3	21	4
64.5	24	7
42.6	13	6
54.5	19	7
57.5	25	2

A multiple linear regression analysis seeks to explain/understand y as

$$y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

that is,

$$price \approx \beta_0 + \beta_1 size + \beta_2 condition$$

The first step in a multiple linear regression analysis is to fit an equation

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

to n data cases

$$(x_1, x_2, \ldots, x_k, y)$$

Just as in SLR, the principle of least squares can be used. That is, it is possible to choose coefficients

$$b_0, b_1, b_2, \ldots, b_k$$

to make as small as possible the quantity

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki}))^2$$

This is a calculus problem whose solution can conveniently be written down only using matrices. We'll not do that here. Instead, we will simply rely upon

JMP and presume that the SAS Institute (that distributes JMP) knows how to do least squares. The symbols

$$b_0, b_1, b_2, \ldots, b_k$$

will henceforth stand for "least squares coefficients" **that for our purposes in Stat 328 can only come from JMP** (or some equivalent piece of software). NOTICE that SLR formulas for least squares coefficients DO NOT WORK HERE! **You can not get** $b_0, b_1, b_2, \ldots, b_k$ **"by hand."**

The geometry of what is going on in least squares for the case of $k = 2$ predictors is illustrated in the following graphic. One attempts to minimize (by jiggling the plane and choosing the b coefficients) the sum of squared vertical distances from plotted points to the plane.

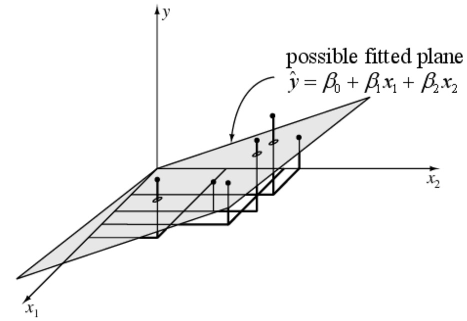

Figure 2: Six Data Points (x_1, x_2, y) and a Possible Fitted Plane

Example In the Real Estate Example (as in all MLR problems) we will use JMP to do least squares. The JMP `Fit Model` report below indicates that least square coefficients are

$$b_0 = 9.78, b_1 = 1.87, \text{ and } b_2 = 1.28$$

That is, the least squares equation is

$$\begin{aligned} \hat{y} &= 9.78 + 1.87x_1 + 1.28x_2 \\ &= 9.78 + 1.87\,size + 1.28\,condition \end{aligned}$$

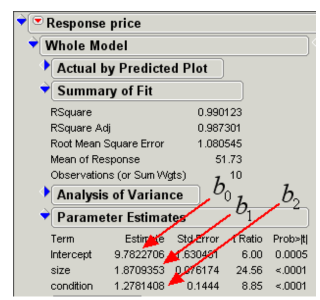

Figure 3: JMP `Fit Model` Report for the Real Estate Example

Notice that b_0 and b_1 for MLR are NOT the same as the corresponding values from SLR! The coefficients b in MLR are rates of change of y with respect to one predictor, *assuming that all other predictors are held fixed.* (They are increases in response that accompany a unit change in the predictor if all others are held fixed.) A SLR coefficient is a rate of change *ignoring any other possible predictors.*

Example In the Real Estate Example

- $b_1 = 1.87$ indicates an \$18.70/ ft^2 increase in price, holding home condition fixed

- $b_2 = 1.28$ indicates that for homes of a fixed size, increase in home condition number by 1 is accompanied by a \$1280 increase in price

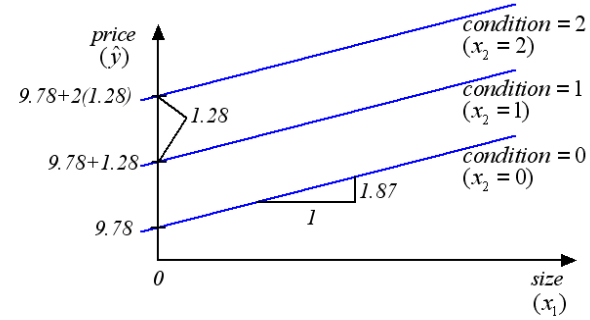

Figure 4: $\hat{y} = 9.78 + 1.87x_1 + 1.28x_2$

Again,

$$b_j = \text{an estimate of the change in mean response } (y) \text{ that accompanies a unit change in the } j\text{th predictor } (x_j) \text{ if all other predictors are held fixed}$$

To measure the "goodness of fit" for a least squares fitted equation in the context of multiple predictors, we'll use the same kind of thinking that we used in SLR. That is, we'll use a version of R^2/the coefficient of determination appropriate to MLR. The "total sum of squares"

$$SSTot = \sum (y_i - \bar{y})^2$$

is defined independent of any reference to what equation one is fitting ... it is the same whether one is considering "size" as a single predictor variable or is considering using both "size" and "condition" to describe price. Once one has

fit any equation (with a single or multiple predictors) to a data set, one has a set of fitted or predicted values $\hat{y}_i$ (coming from the fitted equation involving whatever set of predictors is under discussion). These can be used to make up an "error sum of squares"

$$SSE = \sum (y_i - \hat{y}_i)^2$$

The difference between raw variation in response and that which remains unaccounted for after fitting the equation can be called a "regression sum of squares"

$$SSR = SSTot - SSE$$

And finally, the ratio

$$R^2 = \frac{SSR}{SSTot}$$

serves as a measure of "the fraction of raw variation in y 'accounted for' by the fitted equation" or "the fraction of raw variation in y 'accounted for' by the predictor variable $x_1, x_2, \ldots, x_k$" regardless of how many of these predictors are involved.

The formulas above look exactly like those from SLR. The "news" here is simply that the interpretation of $\hat{y}_i$ has changed. Each predicted value for a given case must be obtained by plugging the values of all k predictors for that case into the fitted equation

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k$$

This and the subsequent calculation of SSE and R^2 is a job only for JMP (except for purposes of practice with a very small example to make sure that one understands what is involved in computing these values).

Example In the Real Estate Example with

$$y = price, x_1 = size, \text{ and } x_2 = condition$$

- SLR on x_1 produces $R^2 = .88$

- MLR on both x_1 and x_2 produces $R^2 = .99$

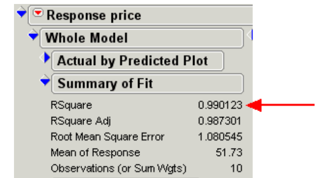

Figure 5: R^2 for MLR in the Real Estate Example

Not surprisingly, adding x_2 improves my ability to explain/predict y. (R^2 can never decrease with the addition of a predictor variable ... it can only increase ... a sensible question is whether the increase provided by an additional predictor is large enough to justify the extra complication involved with using it.)

If one does SLR on x_2 in the Real Estate Example, one gets $R^2 = .14$. There are a couple of points to ponder here. In the first place, it seems that considered one at a time, size is the more important determiner of price. Secondly, note that R^2 for the MLR is NOT simply the sum of R^2 values for the two SLR's ($.99 \neq .88 + .14$). This is typical, and reflects the fact that in the Real Estate Data set, x_1 and x_2 have some correlation, and so in some sense, x_1 is already accounting for some of the predictive power of x_2 ... thus adding x_2 to the prediction process one doesn't expect to increase R^2 by the full potential of x_2.

In Session 6, we noted that in SLR R^2 can be thought of as a squared correlation between the single predictor x and y. In the more complicated MLR context, where there are several x's, R^2 is still a squared correlation (albeit a less "natural" one)

$$R^2 = (\text{sample correlation between } y_i \text{ and } \hat{y}_i \text{ values})^2$$

Exercise For an augmented version of the small fake data set used in class for SLR (below and on the handout) it turns out that MLR of y on both x_1 and x_2 produces the fitted equation

$$\hat{y} = .8 - 1.8x_1 + 2x_2$$

Find the 5 values $\hat{y}, SSE$, and R^2 for this fitted equation.

x_1	x_2	y	$\hat{y}$	$(y - \hat{y})$	$(y - \hat{y})^2$
-2	0	4			
-1	0	3			
0	1	3			
1	1	1			
2	1	-1			

The MLR Model and Point Estimates of Parameters

The most convenient/common model supporting inference making in MLR is the *normal multiple linear regression model*. This can be described in several different ways.

In words, the (normal) multiple linear regression model is:

> Average y is linear in each of k predictor variables $x_1, x_2, \ldots, x_k$ (that is, $\mu_{y|x_1,x_2,\ldots,x_k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$) and for given $(x_1, x_2, \ldots, x_k)$, y is normally distributed with standard deviation σ (that doesn't depend upon the values of $x_1, x_2, \ldots, x_k$).

Pictorially, (for $k = 2$) the (normal) multiple linear regression model is:

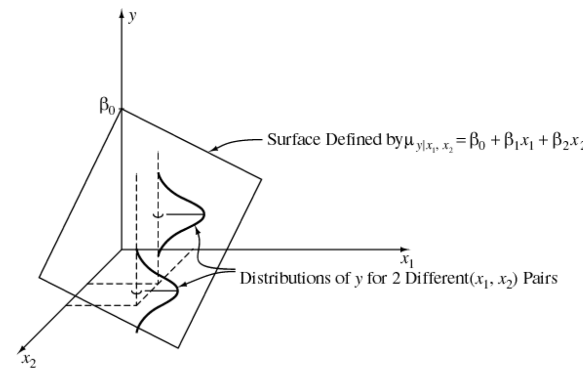

Figure 6: Pictorial Representation of the $k = 2$ Normal MLR Model

Notice on the figure that the models says that the variability in the y distribution *doesn't change with* (x_1, x_2).

In symbols, the (normal) multiple linear regression model is:

$$\begin{aligned} y &= \mu_{y|x_1,x_2,\ldots,x_k} + \epsilon \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon \end{aligned}$$

In this equation, $\mu_{y|x_1,x_2,\ldots,x_k} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ lets the average of the response y change with the predictors x, and ϵ allows the observed response to deviate from the relationship between x and average y ... it is the difference between what is observed and the average response for that $(x_1, x_2, \ldots, x_k)$ ϵ is assumed to be normally distributed with mean 0 and standard deviation σ (that doesn't depend upon $(x_1, x_2, \ldots, x_k)$).

Simple single-number estimates of the parameters of the normal MLR are exactly analogous to those for the SLR model. That is

- parameters $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ will be estimated using the least squares coefficients $b_0, b_1, b_2, \ldots, b_k$ (read from a JMP report only!)

- σ will be estimated using a value derived from the MLR error sum of squares, that is, for MLR

$$"s" = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-k-1}}$$

  JMP calls this value the "root mean square error" (just as for SLR) and it may be read directly from a JMP MLR report.

We will drop the quote marks placed around "s" above but have utilized them at this first occurrence of the formula to point out that this "sample standard deviation" is NEITHER the simple one from Session 1 nor the SLR estimate of σ from Session 7. This is one specifically crafted for the MLR context.

Example In the Real Estate Example with

$$y = price, x_1 = size, \text{ and } x_2 = condition$$

- SLR on x_1 produces $s = 3.53$ ($1000)

- MLR on both x_1 and x_2 produces $s = 1.08$ ($1000)

  The estimated standard deviation of price for any fixed combination of size (x_1) and condition (x_2) is substantially smaller than the estimated standard deviation of price for any fixed size. This should make perfect sense to the reader. Intuitively, if I put additional constraints on the population

of homes I'm talking about by constraining *both* size and condition, I make the population more homogeneous and should therefore reduce variation in selling price.

Confidence and Prediction Intervals in MLR

All that can be done in the way of making interval inferences for SLR carries over directly to MLR. In particular

- **confidence limits for** σ can be based on a χ^2 distribution with $df = n - k - 1$. That is, one may use the interval

$$\left(s\sqrt{\frac{n-k-1}{U}}, s\sqrt{\frac{n-k-1}{L}} \right)$$

for U and L percentage points of the χ^2 distribution with $df = n - k - 1$.

- **confidence limits for any** β_j (the rate of change of mean y with respect to x_j when all other predictors are held fixed, the change in mean y that accompanies a unit change in x_j when all other predictors are held fixed) are

$$b_j \pm t\text{SE}_{b_j}$$

  where t is a percentage point of the t distribution with $df = n - k - 1$ and SE_{b_j} is a standard error (estimated standard deviation) for b_j that can NOT be calculated "by hand," but **must be read from a JMP report**. DO NOT TRY TO CARRY OVER THE SLR FORMULA FOR THE STANDARD ERROR OF THE ESTIMATED SLOPE!

Example In the Real Estate Example with

$$y = price, x_1 = size, \text{ and } x_2 = condition$$

95% confidence limits for the standard deviation of selling price for homes of any fixed size and condition are

$$\left(s\sqrt{\frac{n-k-1}{U}}, s\sqrt{\frac{n-k-1}{L}} \right)$$

that is

$$\left(1.08\sqrt{\frac{10-2-1}{16.01}}, 1.08\sqrt{\frac{10-2-1}{1.69}} \right)$$

or

$$(.71, 2.19)$$

Then reading $\text{SE}_{b_1} = .076$ and $\text{SE}_{b_2} = .14$ from a JMP MLR report for these data, 95% confidence limits for β_1 are

$$b_1 \pm t\text{SE}_{b_1}$$

that is

$$1.87 \pm 2.365\,(.076) \quad \text{or} \quad 1.87 \pm .18$$

while 95% confidence limits for β_2 are

$$b_2 \pm t\text{SE}_{b_2}$$

that is

$$1.28 \pm 2.365\,(.14) \quad \text{or} \quad 1.28 \pm .34$$

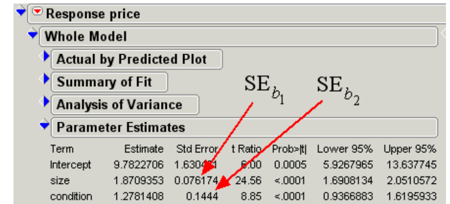

Figure 7: Standard Errors for Least Squares Coefficients from JMP Report

Exercise For the augmented version of the small fake data set, recall that MLR of y on both x_1 and x_2 produces the fitted equation

$$\hat{y} = .8 - 1.8x_1 + 2x_2$$

and that earlier you found that $SSE = .4$. As it turns out (from JMP) $\text{SE}_{b_1} = .283$ and $\text{SE}_{b_2} = .816$. Use this information and make 95% confidence limits for σ, β_1, and β_2.

Focusing now on inferences not for model parameters, but related more directly to responses, consider confidence limits for the mean response and prediction limits for a new response. Suppose that a set of values of the predictor variables

$$(x_1, x_2, \ldots, x_k)$$

is of interest, and that for these values of the predictors one has computed a fitted/predicted value

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

- **confidence limits for** $\mu_{y|x_1,x_2,\ldots,x_k} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_kx_k$ **are** then

$$\hat{y} \pm t\text{SE}_{\hat{\mu}}$$

where t is a percentage point of the t distribution for $df = n - k - 1$ and the standard error $\text{SE}_{\hat{\mu}}$ is peculiar to the particular set of values

$(x_1, x_2, \ldots, x_k)$ and (for that particular set of values) must be read from a JMP report

- **prediction limits for** y_{new} **at the values of the predictors in question are** then

$$\hat{y} \pm t\text{SE}_{\hat{y}}$$

where t is a percentage point of the t distribution for $df = n - k - 1$ and the standard error $\text{SE}_{\hat{y}}$ is peculiar to the particular set of values $(x_1, x_2, \ldots, x_k)$ and (for that particular set of values) must either be read from a JMP report or computed as

$$\text{SE}_{\hat{y}} = \sqrt{s^2 + \left(\text{SE}_{\hat{\mu}}\right)^2}$$

where $\text{SE}_{\hat{\mu}}$ is read directly from a JMP report

For cases $(x_1, x_2, \ldots, x_k)$ in the data set used to fit an equation, JMP allows one to save to the data table (and later read off)

$$\text{"Predicted Values"} = \hat{y}$$

$$\text{"Std Error of Predicted"} = \text{SE}_{\hat{\mu}}$$

and

$$\text{"Std Error of Individual"} = \text{SE}_{\hat{y}}$$

Alternatively, saving

"Mean Confidence Interval" and/or "Indiv Confidence Interval"

produces the (confidence and/or prediction) limits

$$\hat{y} \pm t\text{SE}_{\hat{\mu}} \text{ and/or } \hat{y} \pm t\text{SE}_{\hat{y}}$$

directly.

For cases $(x_1, x_2, \ldots, x_k)$ not in the data set used to fit an equation, one may add a row to the data table with the x's in it and then save

"Prediction Formula" and "StdErr Pred Formula"

and read out of the table

$$\hat{y} \text{ and } \text{SE}_{\hat{\mu}}$$

($\text{SE}_{\hat{y}}$ will have to be computed from s and $\text{SE}_{\hat{\mu}}$)

Example In the Real Estate Example with

$$y = price, x_1 = size, \text{ and } x_2 = condition$$

consider 2000 ft^2 homes of condition number 9. Note that the third home in the data set had these characteristics. After saving "predicted values," "std

error of predicted," and "std error of individual" one may read from the JMP data table the values

$$\begin{aligned} \hat{y} &= 58.704 \\ \text{SE}_{\hat{\mu}} &= .6375 \\ \text{SE}_{\hat{y}} &= 1.255 \end{aligned}$$

So 95% confidence limits for the mean price of such homes are

$$\hat{y} \pm t\text{SE}_{\hat{\mu}}$$
$$58.704 \pm 2.365\,(.6375) \quad \text{or} \quad 58.704 \pm 1.508$$

while 95% prediction limits for an additional price of a home of this size and condition are

$$\hat{y} \pm t\text{SE}_{\hat{y}}$$
$$58.704 \pm 2.365\,(1.255) \quad \text{or} \quad 58.704 \pm 2.956$$

				$\hat{y}$	$\text{SE}_{\hat{\mu}}$	$\text{SE}_{\hat{y}}$
	price	size	condition	Predicted price	StdErr Pred price	StdErr Indiv price
1	60	23	5	59.2044859	0.47136818	1.17888334
2	32.7	11	2	32.9188402	0.81997314	1.35644163
3	57.7	20	9	58.7042432	0.63745178	1.25456077
4	45.5	17	3	45.4225927	0.49185604	1.18722379
5	47	15	8	48.071426	0.60192582	1.23688829
6	55.3	21	4	54.1844746	0.42755336	1.16205845
7	64.5	24	7	63.6317028	0.57052959	1.22191734
8	42.6	13	6	41.7732739	0.57099169	1.22213316
9	54.5	19	7	54.2770264	0.42062699	1.15952793
10	57.5	25	2	59.1119342	0.76573083	1.32435707

Figure 8: Augmented Data Table for MLR Analysis of Real Estate Example

Note that, as expected,

$$1.255 = \sqrt{(1.08)^2 + (.6375)^2}$$

Exercise For the augmented version of the small fake data set, recall that MLR of y on both x_1 and x_2 produces the fitted equation

$$\hat{y} = .8 - 1.8x_1 + 2x_2$$

and that earlier you found that $s = .4472$. As it turns out for the first case in the data set (that has $x_1 = -2$ and $x_2 = 0$) from JMP, $\text{SE}_{\hat{\mu}} = .3464$. Use this information and make 95% confidence limits for mean response when $x_1 = -2$ and $x_2 = 0$ and then make 95% prediction limits for the same set of conditions.