The Simple Linear Regression Model

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

The Simple Linear Regression Model

The Simple Linear Regression Model

The model given in ALR4, page 21, states that

$$\begin{aligned}
E(Y|X = x) &= \beta_0 + \beta_1 x & (1) \\
\text{Var}(Y|X = x) &= \sigma^2 & (2)
\end{aligned}$$

Essentially, the model says that conditional mean of Y is linear in X, with an intercept of β_0 and a slope of β_1, while the conditional variance is constant.

The Simple Linear Regression Model

Because there is conditional variability in Y, the scores on Y cannot generally be perfectly predicted from those on X, and so, to account for this, we say

$$
\begin{align}
y_i &= E(Y|X = x_i) + e_i \tag{3} \\
&= \beta_0 + \beta_1 x_i + e_i \tag{4}
\end{align}
$$

As we pointed out in Psychology 310, the "statistical errors" e_i are defined tautologically as

$$
e_i = y_i - (\beta_0 + \beta_1 x_i) \tag{5}
$$

However, since we do not know the population β_1 and β_0, the statistical errors are unknown and can only be estimated.

The Simple Linear Regression Model

We make two important assumptions concerning the errors:

1. We assume that $E(e_i|x_i) = 0$, so if we could draw a scatterplot of the e_i versus the x_i, we would have a null scatterplot, with no patterns.
2. We assume the errors are all independent, meaning that the value of the error for one case gives no information about the value of the error for another case.

Under these assumptions, if the population is bivariate normal, the errors will be normally distributed.

The Simple Linear Regression Model

One way of thinking about any regression model is that it involves a *systematic* component and an *error* component.

1. If the simple regression model is correct about the systematic component, then the errors will appear to be random as a function of x.
2. However, if the simple regression model is incorrect about the systematic component, then the errors will show a systematic component and be somewhat predictable as a function of x.
3. This is shown graphically in Figure 2.2 from the third edition of ALR.
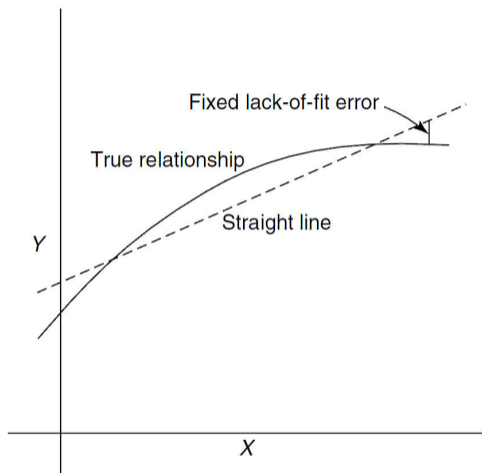
The Simple Linear Regression Model

FIG. 2.2 Approximating a curved mean function by straight line cases adds a fixed component to the errors.

Statistical Notation in ALR

ALR uses a notation in Chapter 2 that is a mixture of standard and not-quite-so-standard notation, but adjustment to it should be easy. Table 2.1 in ALR shows the standard symbols.

Table 2.1 Definitions of Symbols[a]

Quantity	Definition	Description
$\bar{x}$	$\sum x_i/n$	Sample average of x
$\bar{y}$	$\sum y_i/n$	Sample average of y
SXX	$\sum(x_i - \bar{x})^2 = \sum(x_i - \bar{x})x_i$	Sum of squares for the xs
SD_x^2	$SXX/(n-1)$	Sample variance of the xs
SD_x	$\sqrt{SXX/(n-1)}$	Sample standard deviation of the xs
SYY	$\sum(y_i - \bar{y})^2 = \sum(y_i - \bar{y})y_i$	Sum of squares for the ys
SD_y^2	$SYY/(n-1)$	Sample variance of the ys
SD_y	$\sqrt{SYY/(n-1)}$	Sample standard deviation of the ys
SXY	$\sum(x_i - \bar{x})(y_i - \bar{y}) = \sum(x_i - \bar{x})y_i$	Sum of cross-products
s_{xy}	$SXY/(n-1)$	Sample covariance
r_{xy}	$s_{xy}/(SD_x SD_y)$	Sample correlation

[a]In each equation, the symbol $\sum$ means to add over all n values or pairs of values in the data.

Parameters and Estimates

We distinguish between the regression *parameters* and their estimates from sample data. *Parameters* are unknown quantities that characterize a model. *Estimates* of parameters are computable functions of data and are therefore statistics. To keep this distinction clear, parameters are denoted by Greek letters like α, β, γ and σ , and estimates of parameters are denoted by putting a "hat" over the corresponding Greek letter.

Fitted Values

- The sample-based estimates of β_0 and β_1 are denoted $\hat{\beta}_0$ and $\hat{\beta}_1$, respectively.
- The *fitted value* for case i is given $\hat{\mathsf{E}}(Y|X = x_i)$, for which we use the shorthand notation $\hat{y}_i$,

$$\hat{y}_i = \hat{\mathsf{E}}(Y|X = x_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{6}$$

- In other words, the fitted values are obtained by applying the sample regression equation to the sample data.

Residuals

In a similar vein, we define the sample residuals: for the ith case, we have

$$\hat{e}_i = y_i - \hat{y}_i \tag{7}$$

The Least Squares Criterion

Residuals are the distances of the points from the sample-based regression line in the up-down direction, as shown in ALR4 Figure 2.2. (Figure 2.3 in ALR3.)
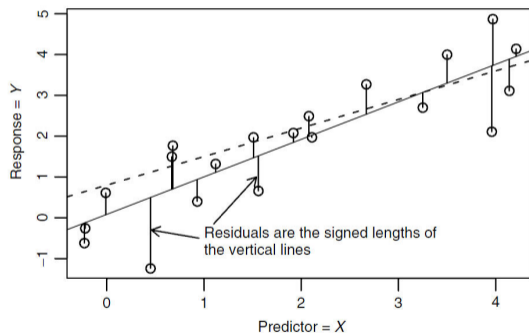

FIG. 2.3   A schematic plot for OLS fitting. Each data point is indicated by a small circle, and the solid line is a candidate OLS line given by a particular choice of slope and intercept. The solid vertical lines between the points and the solid line are the residuals. Points below the line have negative residuals, while points above the line have positive residuals.

The Least Squares Criterion

Consider *any* conceivable estimates of β_0 and β_1, and call them β_0^** and β_1^*. The residual sum of squares (RSS) *for a given* β_0^*, β_1^*, is the sum of squares of the sample residuals around the regression line defined by that particular pair of values, i.e.,

$$\text{RSS}(\beta_0^*, \beta_1^*) = \sum_{i=1}^{n} \left[y_i - (\beta_0^* + \beta_1^* x_i) \right]^2 \tag{8}$$

The OLS estimates $\hat{\beta}_0$, $\hat{\beta}_1$ are the values that minimize RSS. There are several well-known identities useful in computing RSS in OLS regression. For example:

$$
\begin{aligned}
\text{RSS}(\hat{\beta}_0, \hat{\beta}_1) &= \text{SYY} - \frac{\text{SXY}^2}{\text{SXX}} \tag{9} \\
&= \text{SYY} - \hat{\beta}_1^2 \text{SXX} \tag{10}
\end{aligned}
$$

The Least Squares Solution

A solution to the "least squares problem" is given in ALR appendix A.3. The actual solution, as you know from Psychology 310, is

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\text{SXY}}{\text{SXX}} = r_{yx}\frac{SD_y}{SD_x} \qquad &(11)\\
\hat{\beta}_0 &= \overline{y} - \hat{\beta}_1\overline{x} \qquad &(12)
\end{aligned}
$$

In computing the sample residuals, we utilize the two estimates given above, so an unbiased estimate of σ^2 has $n-2$ in its denominator, i.e.,

$$
\hat{\sigma}^2 = \frac{\text{RSS}}{n-2} \qquad (13)
$$

Analyzing the Forbes Data

We can easily fit a simple linear regression for the Forbes data. Let's predict *Lpres* from *Temp*.

The easy way to get the regression coefficients is to use the linear model function in R.

```
> LogPressure <- log(forbes$pres)
> BoilingPoint <- forbes$bp
> fit <- lm(LogPressure ~ BoilingPoint)
> summary(fit)

Call:
lm(formula = LogPressure ~ BoilingPoint)

Residuals:
       Min        1Q     Median        3Q        Max
-0.0073622 -0.0033863 -0.0015865  0.0004322  0.0313139

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.9708662  0.0769377  -12.62 2.17e-09 ***
BoilingPoint  0.0206224  0.0003789   54.42  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00873 on 15 degrees of freedom
Multiple R-squared: 0.995,        Adjusted R-squared: 0.9946
F-statistic:  2962 on 1 and 15 DF,  p-value: < 2.2e-16
```

Detailed Regression Computations in R

We can use the capabilities of R to perform the computational formulas as given in ALR.

```
> X <- BoilingPoint
> Y <- LogPressure
> SXY <- sum((X-mean(X))*(Y-mean(Y)))
> SXX <- sum((X-mean(X))^2 )
> SYY <- sum((Y-mean(Y))^2 )
> beta.hat.1 <- SXY/SXX
> beta.hat.0 <- mean(Y) - beta.hat.1 * mean(X)
> e.hat <- Y - (beta.hat.0 + beta.hat.1 * X)
> RSS <- sum(e.hat^2)
> n <- length(Y)
> sigma.hat.squared <- RSS / (n-2)
> sigma.hat <- sqrt(sigma.hat.squared)
```

Detailed Calculations

Here are the results:

```
> SYY

[1] 0.2268754

> SXX

[1] 530.7824

> SXY

[1] 10.94599

> beta.hat.1

[1] 0.02062236

> beta.hat.0

[1] -0.9708662

> RSS

[1] 0.001143315

> sigma.hat.squared

[1] 7.622099e-05

> sigma.hat

[1] 0.008730463
```

Introduction

As with any statistic, we are interested in the distributional properties of our least squares estimators. Are they unbiased? What are their standard errors?

Section 2.4 of ALR, and the associated appendix sections, A.3 and A.4, develop formulas for these properties that are given in many traditional regression textbooks.

These formulas can be confusing to someone with an intermediate level of statistical background, because:

- The notation is, in an important sense, inconsistent, or at least incomplete. (Explanation below.)
- The derivation of several classic formulas is based on an assumption that is clearly inappropriate, so the classic formulas are not correct in most applications.
- For now, we'll simply give the formulas and discuss them briefly.

Unbiasedness

The estimates are unbiased, i.e.,

$$
\begin{align}
E(\hat{\beta}_0) &= \beta_0 \tag{14} \\
E(\hat{\beta}_1) &= \beta_1 \tag{15} \\
E(\hat{\sigma}^2) &= \sigma^2 \tag{16}
\end{align}
$$

Variance of Estimators

If we assume that errors have constant variance and are uncorrelated, then

$$\text{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\text{SXX}} \right) \tag{17}$$

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\text{SXX}} \tag{18}$$

$$\text{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{n-2} \tag{19}$$

$$\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \frac{\overline{x}}{\text{SXX}} \tag{20}$$

Variance of Estimators

Why do we care about those formulas?

Because, as we shall see later, we use them for constructing confidence intervals and hypothesis tests.

Optimality Properties

Weisberg discusses optimality properties of OLS estimators on page 27 of ALR:

The Gauss-Markov theorem provides an optimality result for OLS estimates. Among all estimates that are linear combinations of the ys and unbiased, the OLS estimates have the smallest variance. If one believes the assumptions and is interested in using linear unbiased estimates, the OLS estimates are the ones to use.

When the errors are normally distributed, the ols estimates can be justified using a completely different argument, since they are then also maximum likelihood estimates, as discussed in many mathematical statistics texts, for example, Casella and Berger (1990).

Estimated Variances

To estimate the sampling variance of our estimates, we simply substitute $\hat{\sigma}^2$ for σ^2 in the preceding formulas. For example,

$$
\begin{align}
\widehat{\text{Var}}(\hat{\beta}_0) &= \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\text{SXX}} \right) \tag{21} \\
\widehat{\text{Var}}(\hat{\beta}_1) &= \frac{\hat{\sigma}^2}{\text{SXX}} \tag{22}
\end{align}
$$

Estimated Standard Errors

At this point, in Section 2.5, ALR continues the prevailing tradition in departing from its own notational standard. The "standard error" of a statistic was originally defined as a population quantity, i.e., the square root of the sampling variance. As a population quantity, a standard error also has an estimator, as we remember from Psychology 310. ALR uses the notation se() to represent the *estimator of a standard error* rather than the standard error itself. So, when Weisberg writes

$$se(\hat{\beta}_1) = \sqrt{\widehat{Var(\hat{\beta}_1)}} \tag{23}$$

he is not actually referring to the standard error, but, rather, its estimate. In the ALR notation, $\sqrt{Var(\hat{\beta})}$ is used to indicate the actual standard error (i.e., the population quantity).

Estimated Standard Errors

Frankly, I don't agree with this notational digression, although I should be clear that many authors use it. In a more consistent (if somewhat messier) notation, one should use se( ) to stand for the population quantity and $\widehat{se}$( ) to stand for the estimated standard error. I suspect the convention of dispensing with the "hat" in the standard error notation was adopted for typographical convenience in the "old days" of painstaking mathematical typing.

In any case, remember that when regression textbooks talk about "standard errors," they are actually talking about estimated standard errors. Asymptotically, it doesn't matter, but at small samples it can.

Ultimately, of course, notation is a matter of personal preference. However, in this case, a deliberate notational inconsistency has been introduced.

Interpreting *p*-values

- As you learned in Psychology 310, *p*-values are interpreted in such a way that if the *p*-value is less than α, then the null hypothesis is rejected at the α significance level.
- ALR has an extensive discussion revisiting this topic.

Power Calculation

When the null hypothesis is true, the F-statistic has a central F distribution. When it is false, *and* the assumption of fixed X holds, then the F-statistic has a non-central F distribution with 1 and $n-2$ degrees of freedom, and a noncentrality parameter λ given by

$$\lambda = \frac{\beta_1^2 \mathtt{SXX}}{\sigma^2} \tag{24}$$

The above equation for λ is not very useful in the context of regression analysis as we normally think about it.

Power Calculation

However, once λ is computed, the power of a test is determined as follows:

- First, calculate a rejection point (critical value) under the assumption that the null hypothesis is true.
- Then calculate the probability that a noncentral $F_{1,n-2,\lambda}$ exceeds the critical value.

We shall illustrate such a calculation in a homework exercise, but first we develop a more meaningful formula for λ in terms of the coefficient of determination, defined on the next slide.

The Coefficient of Determination R^2

The proportion of total variation accounted for by the regression equation is called the
coefficient of determination, and is denoted by R^2. There are a number of formulas for R^2. In
general, even in the multivariate case,

$$R^2 = \frac{SSreg}{\text{SYY}} = 1 - \frac{\text{RSS}}{\text{SYY}} \tag{25}$$

R^2, in the case of a single predictor, is simply r_{xy}^2.

Revisiting Power Calculation

It is fairly easy to show, in the case of a single predictor, that for a population correlation ρ,

$$\lambda = n\frac{\rho^2}{1-\rho^2} \tag{26}$$

Proof. Substituting some well known identities, (i.e. $\beta_1 = \rho\sigma_y/\sigma_x$, $\sigma^2 = (1-\rho^2)\sigma_y^2$, and $n\sigma_x^2 = \text{SXX}$), we get

$$\lambda = \frac{\beta_1^2\text{SXX}}{\sigma^2} \tag{27}$$

$$= \frac{\rho^2(\sigma_y^2/\sigma_x^2)n\sigma_x^2}{(1-\rho^2)\sigma_y^2} \tag{28}$$

$$= n\frac{\rho^2}{1-\rho^2} \tag{29}$$

Note in the above that the X scores are considered fixed, and so the population variance of X is $\sigma_x^2 = \text{SXX}/n$.

Revisiting Power Calculation

- The preceding expression allows one to calculate power in a linear regression in terms of the population ρ^2 value, a much more natural metric for most users than SXX and β_1^2.
- However, careful consideration of the typical application of this formula reveals once again the artificiality of the "fixed X" scores model that treats the X scores as if they were fixed and known (in a sense the entire population). In general, the X scores are random variates just like the Y scores, SXX will vary from sample to sample, the fixed scores model is not really appropriate, and the power value is an approximation.
- In the case of multiple regression, the approximation can be off by a substantial amount, but it is usually adequate.

Confidence Intervals and Tests
Introduction

In Section 2.6 of ALR, Weisberg introduces several of the classic parametric hypothesis tests and confidence intervals calculated in connection with simple linear regression.

1. *The Intercept β_0.*
2. *The Slope β_1.*
3. *Predicted Values from a New Data Point.*
4. *Fitted Values (Conditional Mean Estimates) on the Regression Line.*
5. *Residuals.*

We shall now consider each of these in turn, demonstrating calculations as we go.

Confidence Intervals and Tests
The Intercept β_0: Simple Confidence Interval

The (estimated) standard error for β_0 is $\text{se}(\beta_0) = \hat{\sigma}(1/n + \overline{x}^2/\text{SXX})^{1/2}$, and a $100(1-\alpha)\%$ confidence interval is

$$\hat{\beta}_0 \pm t^* \, \text{se}(\hat{\beta}_0) \tag{30}$$

where $t*$ is a critical value from the t distribution with $n-2$ degrees of freedom, i.e., $t^* = t_{n-2, 1-\alpha/2}$.

Confidence Intervals and Tests
The Intercept β_0: Simple Hypothesis Testing

The above $100(1 - \alpha)\%$ confidence interval may be used directly to test a two-sided hypothesis about the value of β_0. Specifically, to test the null hypothesis that $\beta_0 = \beta_0^*$, at the α significance level, simply observe whether or not the confidence interval excludes β_0^*.

If an actual *p*-value is required, one can use the test statistic

$$t_{n-2} = \frac{\hat{\beta}_0 - \beta_0^*}{\text{se}(\hat{\beta}_0)} \tag{31}$$

Confidence Intervals and Tests
The Slope β_1

The (estimated) standard error of β_1 is

$$\mathrm{se}(\hat{\beta}_1) = \frac{\hat{\sigma}}{\mathrm{SXX}} \tag{32}$$

A confidence interval for β_1 may be constructed in the standard manner, with endpoints given by

$$\hat{\beta}_1 \pm t^* \, \mathrm{se}(\hat{\beta}_1) \tag{33}$$

Confidence Intervals and Tests
Two Kinds of Intervals around a Regression Line

One often sees confidence regions plotted in connection with a regression line. There are actually two distinctly different kinds of plots:

1. A regression line has been calculated from a data set, then a new value x_* becomes available, prior to the availability of the associated y_*. What is an appropriate confidence interval for the predicted value?

2. A regression line involves an (infinite) set of "fitted values" that represent conditional means for $Y|X = x$. What is a confidence interval for such a fitted value?

Confidence Intervals and Tests
A Predicted Value from a New Data Point

The first kind of interval is calculated as follows. The estimated value of y_* is obtained by substituting x_* into the estimated regression line, i.e.,

$$\tilde{y}_* = \hat{\beta}_0 + \hat{\beta}_1 x_* \tag{34}$$

Under the assumptions of fixed predictors regression, the conditional sampling variance of $\tilde{y}_*$ given x_* is a function of x_* itself, i.e.,

$$\text{Var}(\tilde{y}_*|x_*) = \sigma^2 + \sigma^2 \left(\frac{1}{n} + \frac{(x_* - \overline{x})^2}{\text{SXX}} \right) \tag{35}$$

Recalling that $\text{SXX} = (n-1)S_x^2$, we can, after a little reduction, write a somewhat more revealing version of the formula as

$$\text{Var}(\tilde{y}_*|x_*) = \sigma^2 \left(\frac{n+1}{n} + \frac{1}{n-1} \left(\frac{x_* - \overline{x}}{S_x} \right)^2 \right) \tag{36}$$

Confidence Intervals and Tests
A Predicted Value from a New Data Point

Since the standard deviation varies as a function of x_*, a simultaneous confidence region plot will be curved. Here is a formula and notation for the estimated standard error of prediction for a given x_*:

$$\text{sepred}(\tilde{y}_*|x_*) = \hat{\sigma}\left(\frac{n+1}{n} + \frac{1}{n-1}\left(\frac{x_* - \overline{x}}{S_x}\right)^2\right)^{1/2} \tag{37}$$

Confidence Intervals and Tests
A Predicted Value from a New Data Point

- Weisberg gives an extensive example of calculation of a prediction interval in section 2.8.3 of ALR. This is done in the standard way, i.e., the estimate plus or minus the standard error times the critical value of t.
- He points out that a prediction interval for log(*Pressure*) can be converted into one for *Pressure* by tranforming the endpoints of the former, since the $\log(x)$ transformation is strictly increasing in x.

Confidence Intervals and Tests
A Fitted Value (Conditional Mean)

In some situations one may be interested in obtaining an estimate of $E(Y|X = x)$. For example, in the heights data, one might estimate the population mean height of all daughters of mothers with a particular height x_*. This quantity is estimated by the fitted value $\hat{y} = \beta_0 + \beta_1 x_*$.

Confidence Intervals and Tests
A Fitted Value (Conditional Mean)

Regardless of whether or not you consider the fitted value itself an "estimate," you can estimate *it* with the quantity $\tilde{y}_* = \hat{\beta}_0 + \hat{\beta}_1 x_*$. This estimate has an estimated standard error of

$$\text{sefit}(\tilde{y}_*|x_*) = \hat{\sigma} \left(\frac{1}{n} + \frac{1}{n-1} \left(\frac{x_* - \overline{x}}{S_x} \right)^2 \right)^{1/2} \tag{38}$$

Confidence Intervals and Tests
A Fitted Value (Conditional Mean)

Again, for a single value, a confidence interval for such an estimated conditional mean can be calculated with the standard approach, e.g.,

$$\tilde{y}_* \pm t^* \, \text{sefit}(\tilde{y}_* | x_*) \tag{39}$$

where t^* is the $1 - \alpha/2$ critical value from the t distribution with $n - 2$ degrees of freedom.

Confidence Intervals and Tests
A Fitted Value (Conditional Mean)

The above confidence interval might be considered appropriate when only one conditional mean is of interest. On page 36 of ALR, Weisberg discusses using the Scheffe correction to allow simultaneous computation of all estimated conditional means.

This can be done by substituting $(2F^*)^{1/2}$ for t^* in the above formula. F^* is the critical value from the central F distribution with 2 and $n-2$ degrees of freedom. (More generally, for multiple regression models with p' predictors including the intercept term, one substitutes $(p'F^*)^{1/2}$, where F^* is the critical value from the central F with p' and $n-p'$ degrees of freedom.

The function shown below computes the Scheffe correction by computing the uncorrected interval and then expanding it by the ratio of the two critical values. This function should work for multiple regression as well as simple bivariate regression.

Confidence Intervals and Tests
A Fitted Value (Conditional Mean)

```
> ## Create a function to compute the Scheffe corrected confidence
> ## interval for the regression line
> scheffe.rescaled.ci <- function(model,conf.level,new){
+ ## Get df and number of predictors from model object
+  df <- model$df.residual
+  p <- model$rank
+  alpha <- 1-conf.level
+ ## NOTE Scheffe value uses 1-tailed F critical value
+  scheffe.crit <- sqrt(p*qf(1-alpha,p,df))
+  ci <- predict(model,new,interval="confidence",level=conf.level)
+ ## Create multiplier to expand the width of the ci
+  multiplier <- scheffe.crit/qt(1-alpha/2,df)
+  ## Recompute the ci
+  ci[,2] <- ci[,1] -(ci[,1]-ci[,2])*multiplier
+  ci[,3] <- ci[,1] +(ci[,3]-ci[,1])*multiplier
+  return(ci)
+ }
```

Confidence Intervals and Tests
Plotting Confidence Intervals

- We can plot the prediction intervals and the confidence intervals for fitted values using R.
- Note that ALR recommends the Scheffe correction for the latter, but not for the former. One might ask, "Why?"
- Ostensibly, this is because in the former case, we are graphing what the confidence interval *would be* if we had observed a value x_*, while in the latter case, we are asking what the theoretical confidence intervals would be for the entire run of the regression line, based on the current data.

Confidence Intervals and Tests

Plotting Confidence Intervals

Here is some commented code:

```
> ##fit the simple regression model
> attach(Heights)
> m1 <- lm(dheight~mheight)
> ##Create a run of 50 points across the x-axis
> new <- data.frame(mheight=seq(55.4,70.8,length=50))
> ##create the confidence intervals
> ##  first for the prediction intervals
> pred.w.plim <- predict(m1, new, interval="prediction")
> ##  next for the fitted value (conditional mean)
> pred.w.clim <- scheffe.rescaled.ci(m1,0.95,new)
> #Then we use matplot --
> #  cbind takes all 3 columns of pred.w.clim
> #  and last two of pred.w.plim
> matplot(new$mheight,cbind(pred.w.clim, pred.w.plim[,-1]),
+                col=c("black","red","red","blue","blue"),bty="l",
+                lty=c(2,1,1,1,1), type="l", ylab="Daughter's Height",
+                xlab="Mother's Height")
> legend("bottomright", c("Prediction Interval", "Fitted Value C.I."),
+    lty = c(1, 1),col=c("blue","red"))
```

Confidence Intervals and Tests
Plotting Confidence Intervals
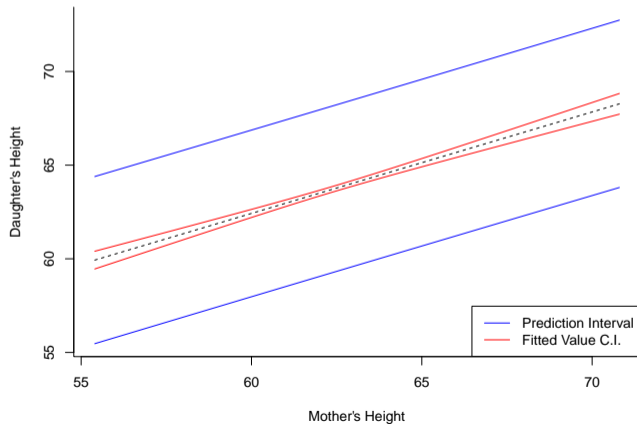
Here is the plot:

Residuals

Weisberg makes the following points about residual plots.

- Plots of residuals versus other quantities are used to find failures of assumptions.
- The most common plot, especially useful in simple regression, is the plot of residuals versus the fitted values.
- A plot with a slope of zero and even scatter indicates assumptions are realistic.
- Curvature might indicate that the fitted mean function is inappropriate.
- Residuals that seem to increase or decrease in average magnitude with the fitted values might indicate nonconstant residual variance.
- A few relatively large residuals may be indicative of outliers, cases for which the model is somehow inappropriate.

Confidence Intervals and Tests
Residual Plot for the Forbes Data

Here is the code for the Forbes data residual plot. Note the code for labeling the 12th case, which is an outlier. Weisberg discusses the effect of dropping the outlier and reanalyzing the data in section 2.8 of ALR4.

```
> m1 <- lm(LogPressure~BoilingPoint)
> plot(predict(m1),residuals(m1),
+       xlab="Fitted values", ylab="Residuals")
> text(predict(m1)[12],residuals(m1)[12],labels="12",adj=-1)
> abline(0,0)
```