Data Analysis
Spring Semester, 2023
February 28, 2023
Lecture 10

Chapter 11
Linear Regression and Correlation

The research context is that two variables have been observed for each of *n* participants. The research team then has a spreadsheet with *n* pairs of observations $(x_i, y_i),\ i = 1, \ldots, n$. One of the variables (here *y*) is the outcome variable or dependent variable. This is the variable hypothesized to be affected by the other variable in scientific research. The other variable (here *x*) is the independent variable. It may be hypothesized to predict the outcome variable or to cause a change in the outcome variable. The research task is to document the association between independent and dependent variables. An example of a research project seeking to document a causal association would be a clinical trial in which x_i was the dosage of a medicine randomly assigned to a participant (say simvastatin) and y_i was the participant's response after a specified period taking the medicine (say cholesterol reduction after 3 months). An example of a study seeking to document the value of a predictive association would be an observational study in which x_i was the score of a statistics student on the first examination in a course and y_i was the student's score on the final examination in the course.

散点图

A recommended first step is to create the scatterplot of observations, with the vertical axis representing the dependent variable and the horizontal axis representing the independent variable. The "pencil test" is to hold up a pencil to the scatterplot and examine whether that describes the data well. If so, then it is reasonable to assume that a ***linear model*** (such as $\beta_0 + \beta_1 x$) describes the data. The linear model is reasonable for many data sets in observational studies. A more object procedure is to use a "nonlinear smoother" such as LOWESS to estimate the association. If the LOWESS curve is not well approximated by a line, then the assumption of linearity is not reasonable.

Estimating the Linear Model Parameters (section 11.2)

OLS (ordinary least squares) is the most used method to estimate the parameters of the linear model. An arbitrary linear model $b_0 + b_1 x$ is used as a *fit* for the dependent variable values. The method uses the *residual* $y_i - b_0 - b_1 x_i$. The fitting model is judged by how small the set of residuals is. OLS uses each residual and focuses on the magnitude of the residuals by examining the sum of squares function $SS(b_0, b_1) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2$. The OLS method is to find the arguments $(\hat{\beta}_0, \hat{\beta}_1)$ that make $SS(b_0, b_1)$ as small as possible. This minimization is a standard calculus problem. Step 1 is to calculate the partial derivatives of $SS(b_0, b_1)$ with respect to each argument. First, the partial with respect to b_0:

$$\frac{\partial SS(b_0,b_1)}{\partial b_0} = \frac{\partial}{\partial b_0}\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^{n}\frac{\partial}{\partial b_0}(y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^{n} 2(y_i - b_0 - b_1 x_i)\frac{\partial (y_i - b_0 - b_1 x_i)}{\partial b_0}$$

$$\frac{\partial SS(b_0,b_1)}{\partial b_0} = \sum_{i=1}^{n}(-2)(y_i - b_0 - b_1 x_i).$$

1.计算残差

Second, the partial with respect to b_1:

$$\frac{\partial SS(b_0,b_1)}{\partial b_1} = \frac{\partial}{\partial b_1}\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^{n}\frac{\partial}{\partial b_1}(y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^{n} 2(y_i - b_0 - b_1 x_i)\frac{\partial (y_i - b_0 - b_1 x_i)}{\partial b_1}$$

$$\frac{\partial SS(b_0,b_1)}{\partial b_1} = \sum_{i=1}^{n}(-2x_i)(y_i - b_0 - b_1 x_i).$$

Step 2 is to find the arguments $(\hat{\beta}_0, \hat{\beta}_1)$ that make the two partial derivatives zero. The resulting equations are called the *normal equations*:

$$\sum_{i=1}^{n}(-2)(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \text{ and}$$

$$\sum_{i=1}^{n}(-2)(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0.$$

These equations have a very important interpretation. Let $r_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i,\ i = 1, \ldots, n$.

2.计算系数

The first normal equation is equivalent to $\sum_{i=1}^{n} r_i = 0$, and the second is $\sum_{i=1}^{n} r_i x_i = 0$. That is, there are two constraints on the *n* residuals. The OLS residuals must sum to zero, and the OLS residuals are orthogonal to the independent variable values. The *n* residuals then have $n-2$ degrees of freedom.

Step 3 is to solve this two linear equation system in two unknowns. Start by using the first normal equation to solve for $\hat{\beta}_0$:

$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{\beta}_0 - \sum_{i=1}^{n}\hat{\beta}_1 x_i = n\bar{y}_n - n\hat{\beta}_0 - \hat{\beta}_1(n\bar{x}_n) = 0$. Solving for $\hat{\beta}_0$ yields $\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$. Next, insert the solution for $\hat{\beta}_0$ in the second normal equation and solve for $\hat{\beta}_1$:

$$0=\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)x_i=\sum_{i=1}^{n}\{[y_i-(\bar{y}_n-\hat{\beta}_1\bar{x}_n)-\hat{\beta}_1 x_i]x_i\}=\sum_{i=1}^{n}[(y_i-\bar{y}_n)x_i]-\sum_{i=1}^{n}[\hat{\beta}_1(x_i-\bar{x}_n)x_i],$$

The solution is $\hat{\beta}_1=\dfrac{\sum_{i=1}^{n}(y_i-\bar{y}_n)x_i}{\sum_{i=1}^{n}(x_i-\bar{x}_n)x_i}$. There are several modifications of this formula that are helpful. The first results from noting that

$\sum_{i=1}^{n}(x_i-\bar{x}_n)^2=\sum_{i=1}^{n}(x_i-\bar{x}_n)x_i-\sum_{i=1}^{n}(x_i-\bar{x}_n)\bar{x}_n=\sum_{i=1}^{n}(x_i-\bar{x}_n)x_i$ and 高亮为0

$\sum_{i=1}^{n}(y_i-\bar{y}_n)(x_i-\bar{x}_n)=\sum_{i=1}^{n}(y_i-\bar{y}_n)x_i-\sum_{i=1}^{n}(y_i-\bar{y}_n)\bar{x}_n=\sum_{i=1}^{n}(y_i-\bar{y}_n)x_i$. The OLS solution is then $\hat{\beta}_1=\dfrac{\sum_{i=1}^{n}(y_i-\bar{y}_n)(x_i-\bar{x}_n)}{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}$. This is a very commonly quoted formula.

The second shows the relation of $\hat{\beta}_1$ and the Pearson product moment correlation. The Pearson product moment correlation is a dimensionless measure of association. The formula is $r(x,y)=\dfrac{\sum_{i=1}^{n}(y_i-\bar{y}_n)(x_i-\bar{x}_n)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}}$. The Cauchy-Schwartz inequality shows that $|r(x,y)|\le 1$. A correlation of +1 or -1 shows a perfect linear association. A correlation of 0 means no linear association. The numerator of $\hat{\beta}_1$ and $r(x,y)$ are the same. Starting with $\hat{\beta}_1$,

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(y_i-\bar{y}_n)(x_i-\bar{x}_n)}{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}=\frac{\sum_{i=1}^{n}(y_i-\bar{y}_n)(x_i-\bar{x}_n)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}\bullet\sqrt{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}}\bullet\frac{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}}.$$ That is,

$$\hat{\beta}_1=\frac{\sum_{i=1}^{n}(y_i-\bar{y}_n)(x_i-\bar{x}_n)}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}\bullet\sqrt{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}}\bullet\frac{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}}=r(x,y)\bullet\frac{\sqrt{(n-1)s_Y^2}}{\sqrt{(n-1)s_X^2}}=\frac{s_Y}{s_X}\bullet r(x,y).$$ The second formula is then $\hat{\beta}_1=\dfrac{s_Y}{s_X}\bullet r(x,y)$.

The next variation will be used in calculating the distributional properties of $\hat{\beta}_1$ and uses the identity that

$\sum_{i=1}^{n}(y_i-\bar{y}_n)(x_i-\bar{x}_n)=\sum_{i=1}^{n}[(x_i-\bar{x}_n)y_i]-\sum_{i=1}^{n}[(x_i-\bar{x}_n)\bar{y}_n]=\sum_{i=1}^{n}(x_i-\bar{x}_n)y_i$. Then $\hat{\beta}_1=\dfrac{\sum_{i=1}^{n}(x_i-\bar{x}_n)y_i}{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}$

Fisher's Decomposition of the Total Sum of Squares

The total sum of squares of the dependent variable is defined to be $TSS=\sum_{i=1}^{n}(y_i-\bar{y}_n)^2$ with $n-1$ degrees of freedom. The ith residual was defined above to be $r_i=y_i-\hat{\beta}_0-\hat{\beta}_1x_i,\ i=1,\ldots,n$. After substituting for $\hat{\beta}_0$, $r_i=y_i-\bar{y}_n-\hat{\beta}_1(x_i-\bar{x}_n),\ i=1,\ldots,n$.

Fisher's decomposition is a fundamental tool for the analysis of the linear model. It starts with $TSS=\sum_{i=1}^{n}(y_i-\bar{y}_n)^2=\sum_{i=1}^{n}[y_i-\bar{y}_n-\hat{\beta}_1(x_i-\bar{x}_n)+\hat{\beta}_1(x_i-\bar{x}_n)]^2=\sum_{i=1}^{n}[r_i+\hat{\beta}_1(x_i-\bar{x}_n)]^2$.
Next $TSS=\sum_{i=1}^{n}[r_i+\hat{\beta}_1(x_i-\bar{x}_n)]^2=\sum_{i=1}^{n}[r_i^2+\hat{\beta}_1^2(x_i-\bar{x}_n)^2+2\hat{\beta}_1r_i(x_i-\bar{x}_n)]$, and

$TSS=\sum_{i=1}^{n}r_i^2+\sum_{i=1}^{n}\hat{\beta}_1^2(x_i-\bar{x}_n)^2+2\hat{\beta}_1\sum_{i=1}^{n}r_i(x_i-\bar{x}_n)$. The first sum $\sum_{i=1}^{n}r_i^2=SSE$, the sum of squared errors and has $n-2$ degrees of freedom. The second sum $\sum_{i=1}^{n}\hat{\beta}_1^2(x_i-\bar{x}_n)^2$ is called the regression sum of squares and has 1 degree of freedom. It can be simplified:

$$\text{RegSS}=\sum_{i=1}^{n}\hat{\beta}_1^2(x_i-\bar{x}_n)^2=\hat{\beta}_1^2\sum_{i=1}^{n}(x_i-\bar{x}_n)^2=[r(x,y)]^2\left[\frac{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}}\right]^2\sum_{i=1}^{n}(x_i-\bar{x}_n)^2 \text{ and}$$

$$\text{RegSS}=[r(x,y)]^2\left[\frac{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y}_n)^2}}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}_n)^2}}\right]^2\sum_{i=1}^{n}(x_i-\bar{x}_n)^2=[r(x,y)]^2\sum_{i=1}^{n}(y_i-\bar{y}_n)^2=[r(x,y)]^2TSS.$$

Finally, the third sum $2\hat{\beta}_1\sum_{i=1}^{n}r_i(x_i-\bar{x}_n)=2\hat{\beta}_1(\sum_{i=1}^{n}r_ix_i-\sum_{i=1}^{n}r_i\bar{x}_n)=2\hat{\beta}_1(0-0)=0$.

This is conventionally displayed in an Analysis of Variance Table as below:

Analysis of Variance Table
One Predictor Linear Regression

Source	DF	Sum of Squares	Mean Square	F
Regression	1	$\sum_{i=1}^{n}\hat{\beta}_1^2(x_i-\bar{x}_n)^2=[r(x,y)]^2 TSS$	$[r(x,y)]^2 TSS$	$\frac{(n-2)[r(x,y)]^2}{1-[r(x,y)]^2}$
Error	$n-2$	$\sum_{i=1}^{n}r_i^2=\{1-[r(x,y)]^2\}TSS$	$\frac{\{1-[r(x,y)]^2\}TSS}{(n-2)}$	
Total	$n-1$	$TSS=(n-1)s_{DV}^2$		

11.3 Inferences

There must be a probabilistic model for the data so that researchers can make inferences and find confidence intervals. The model for one predictor linear regression is $Y_i=\beta_0+\beta_1 x_i+\sigma_{Y|x}Z_i$. The outcome or dependent (random) variables $Y_i, i=1,\ldots,n$ are each assumed to be the sum of the linear regession expected value $\beta_0+\beta_1 x_i$ and a random error term $\sigma_{Y|x}Z_i$. The random variables $Z_i, i=1,\ldots,n$ are assumed to be independent standard normal random variables. The parameter β_0 is the intercept parameter and is fixed but unknown. The parameter β_1 is the slope parameter and is also fixed but unknown. This parameter is the focus of the statistical analysis. The parameter $\sigma_{Y|x}$ is also fixed but unknown. Another description of this model is that $Y_i, i=1,\ldots,n$ are independent normally distributed random variables with Y_i having the distribution $N(\beta_0+\beta_1 x_i,\sigma_{Y|x}^2)$. That is, $E(Y_i \mid X=x_i)=\beta_0+\beta_1 x_i$, and $\mathrm{var}(Y_i \mid X=x_i)=\sigma_{Y|x}^2$. The assumption that $\mathrm{var}(Y_i \mid X=x_i)=\sigma_{Y|x}^2$ is called the *homoscedasticity* assumption.

There are four assumptions. There are two important assumptions: the outcome variables $Y_i, i=1,\ldots,n$ are independent and $E(Y_i \mid X=x_i)=\beta_0+\beta_1 x_i$ for $i=1,\ldots,n$. Homoscedasticity is less important. The assumption that $Y_i, i=1,\ldots,n$ are normally distributed random variables is least important.

Variance Calculations

The most complex variance formula in this course so far is:

$$\mathrm{var}(aX+bY)=a^2\,\mathrm{var}\,X+b^2\,\mathrm{var}\,Y+2ab\,\mathrm{cov}(X,Y).$$

More complex calculations are required for the variance-covariance matrix of the OLS estimates. The easiest way is to use the variance-covariance matrix of a random vector. Let *Y* be an $n \times 1$ vector of random variables $(Y_1, Y_2, \ldots Y_n)^T$. That is, each component of the vector is a random variable. Then the expected value of vector *Y* is the $n \times 1$ vector whose components are the respective means of the random variables; that is, $E(Y) = (EY_1, EY_2, \ldots EY_n)^T$. The variance-covariance matrix of the random vector *Y* is the $n \times n$ matrix whose diagonal entries are the respective variances of the random variables and whose off-diagonal elements are the covariances of the random variables. That is,

$$\mathrm{vcv}(Y) = \begin{bmatrix} \mathrm{var}(Y_1) & \mathrm{cov}(Y_1, Y_2) & \cdots & \mathrm{cov}(Y_1, Y_n) \\ \mathrm{cov}(Y_2, Y_1) & \mathrm{var}(Y_2) & \cdots & \mathrm{cov}(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(Y_n, Y_1) & \mathrm{cov}(Y_n, Y_2) & \cdots & \mathrm{var}(Y_n) \end{bmatrix}.$$

In terms of expectation operator calculations, $\mathrm{vcv}(Y) = E[(Y - EY)(Y - EY)^T] = \Sigma$.

Variance of a Set of Linear Combinations

Let *W* be the $m \times 1$ random vector of linear combinations of Y given by $W = MY$, where M is a matrix of constants having *m* rows and *n* columns. Then $E(W) = E(MY) = ME(Y)$, The definition of the variance-covariance matrix of *W* is $vcv(W) = E[(W - EW)(W - EW)^T] = E[(MY - MEY)(MY - MEY)^T]$, and
$vcv(W) = E[(MY - MEY)(MY - MEY)^T] = E\{M(Y - EY)[M(Y - EY)]^T\}$
From matrix algebra, when *A* is an $n \times m$ matrix and *B* is an $m \times p$ matrix, then $(AB)^T = B^T A^T$. Then $[M(Y - EY)]^T = (Y - EY)^T M^T$, and
$\mathrm{vcv}(W) = \mathrm{vcv}(MY) = E\{M(Y - EY)(Y - EY)^T M^T\} = M\{E[(Y - EY)(Y - EY)^T]\}M^T$ from the linear operator property of *E*. Since $\mathrm{vcv}(Y) = E[(Y - EY)(Y - EY)^T] = \Sigma$,
$\mathrm{vcv}(W) = \mathrm{vcv}(MY) = M \times \mathrm{vcv}(Y) \times M^T = M\Sigma M^T$

Examples

The first use of this result is to find the variance of a linear combination of values from *Y*, an $n \times 1$ vector of random variables. Let ***a*** be an $n \times 1$ vector of constants, and let $W = a^T Y$. Then $\mathrm{var}(a^T Y) = a^T \times \mathrm{vcv}(Y) \times (a^T)^T = a^T \times \mathrm{vcv(Y)} \times \mathrm{a}$. This is the completely general form of $\mathrm{var}(aX + bY) = a^2 \mathrm{var}(X) + b^2 \mathrm{var}(Y) + 2ab\,\mathrm{cov}(X, Y)$.

The second example is fundamental to this chapter. The OLS estimates of the parameters are always the same functions of the observed data: $\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$ and $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n) y_i}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}$. It is then reasonable to study the random variables

$$\bar{Y}_n = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{1}{n}Y_1 + \frac{1}{n}Y_2 + \cdots + \frac{1}{n}Y_n$$

and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)Y_i}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} = \frac{(x_1 - \bar{x}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}Y_1 + \frac{(x_2 - \bar{x}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}Y_2 + \cdots + \frac{(x_n - \bar{x}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}Y_n.$$

Let $w_i = \dfrac{(x_i - \bar{x}_n)}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2},\ i = 1, \ldots, n$.

Then $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x}_n)Y_i}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2} = w_1 Y_1 + w_2 Y_2 + \cdots + w_n Y_n$.

Let $\begin{pmatrix} \bar{Y}_n \\ \hat{\beta}_1 \end{pmatrix} = \begin{bmatrix} 1/n & 1/n & \cdots & 1/n \\ w_1 & w_2 & \cdots & w_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, which has the form *MY*, where

$$M = \begin{bmatrix} 1/n & 1/n & \cdots & 1/n \\ w_1 & w_2 & \cdots & w_n \end{bmatrix}.$$

In the model $Y_i = \beta_0 + \beta_1 x_i + \sigma_{Y|x} Z_i$, $\mathrm{vcv}(Y) = \sigma^2_{Y|x} I_{n\times n}$. Then

$$\mathrm{vcv}\begin{pmatrix} \bar{Y}_n \\ \hat{\beta}_1 \end{pmatrix} = \begin{bmatrix} 1/n & 1/n & \cdots & 1/n \\ w_1 & w_2 & \cdots & w_n \end{bmatrix} \times \sigma^2_{Y|x} I_{n\times n} \times \begin{bmatrix} 1/n & w_1 \\ 1/n & w_2 \\ \vdots & \vdots \\ 1/n & w_n \end{bmatrix} = \sigma^2_{Y|x} \begin{bmatrix} 1/n & 1/n & \cdots & 1/n \\ w_1 & w_2 & \cdots & w_n \end{bmatrix} \times \begin{bmatrix} 1/n & w_1 \\ 1/n & w_2 \\ \vdots & \vdots \\ 1/n & w_n \end{bmatrix}.$$

Then,

$$\mathrm{vcv}\begin{pmatrix}\bar{Y}_n\\ \hat{\beta}_1\end{pmatrix}=\sigma^2_{Y|x}\begin{bmatrix}1/n & 1/n & \cdots & 1/n\\ w_1 & w_2 & \cdots & w_n\end{bmatrix}\times\begin{bmatrix}1/n & w_1\\ 1/n & w_2\\ \vdots & \vdots\\ 1/n & w_n\end{bmatrix}=\sigma^2_{Y|x}\begin{bmatrix}\sum_{i=1}^n \frac{1}{n^2} & \sum_{i=1}^n \frac{w_i}{n}\\ \sum_{i=1}^n \frac{w_i}{n} & \sum_{i=1}^n w_i^2\end{bmatrix}.$$

Now $\sum_{i=1}^n \frac{w_i}{n}=\sum_{i=1}^n \frac{1}{n}\frac{(x_i-\bar{x}_n)}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}=\frac{1}{n\sum_{i=1}^n (x_i-\bar{x}_n)^2}\sum_{i=1}^n (x_i-\bar{x}_n)=0$, and

$$\sum_{i=1}^n w_i^2=\sum_{i=1}^n\left[\frac{(x_i-\bar{x}_n)}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}\right]^2=\frac{1}{[\sum_{i=1}^n (x_i-\bar{x}_n)^2]^2}\sum_{i=1}^n (x_i-\bar{x}_n)^2=\frac{1}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}.$$

The final result is that $\mathrm{vcv}\begin{pmatrix}\bar{Y}_n\\ \hat{\beta}_1\end{pmatrix}=\sigma^2_{Y|x}\begin{bmatrix}\sum_{i=1}^n \frac{1}{n^2} & \sum_{i=1}^n \frac{w_i}{n}\\ \sum_{i=1}^n \frac{w_i}{n} & \sum_{i=1}^n w_i^2\end{bmatrix}=\begin{bmatrix}\frac{\sigma^2_{Y|x}}{n} & 0\\ 0 & \frac{\sigma^2_{Y|x}}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}\end{bmatrix}$

To summarize this result, $\mathrm{var}(\bar{Y}_n)=\frac{\sigma^2_{Y|x}}{n}$, $\mathrm{var}(\hat{\beta}_1)=\frac{\sigma^2_{Y|x}}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}$, and

$\mathrm{cov}(\bar{Y}_n,\hat{\beta}_1)=0$.

Testing a null hypothesis about β_1

The last detail before deriving tests and confidence intervals for the slope of the regression function is to find $E(\hat{\beta}_1)$. Then

$E(\hat{\beta}_1)=E(\sum_{i=1}^n w_iY_i)=\sum_{i=1}^n w_iE(Y_i)=\sum_{i=1}^n w_i(\beta_0+\beta_1x_i)=\beta_0(\sum_{i=1}^n w_i)+\beta_1(\sum_{i=1}^n w_ix_i)$. Now, from above, $\sum_{i=1}^n w_i=\frac{1}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}\sum_{i=1}^n (x_i-\bar{x}_n)=0.$

The second sum is
$$\sum_{i=1}^n w_ix_i=\sum_{i=1}^n \frac{(x_i-\bar{x}_n)}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}x_i=\frac{\sum_{i=1}^n (x_i-\bar{x}_n)x_i}{\sum_{i=1}^n (x_i-\bar{x}_n)^2}=1.$$

Then $E(\hat{\beta}_1) = \beta_1$. Under the data model, the distribution of $\hat{\beta}_1$ is $N(\beta_1, \frac{\sigma^2_{Y|x}}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2})$.

The key null hypothesis is $H_0 : \beta_1 = 0$, and the alternative hypothesis is $H_1 : \beta_1 \neq 0$. The test statistic is $\hat{\beta}_1$, and the null distribution is $N(0, \frac{\sigma^2_{Y|x}}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2})$. The standard score form of the statistic is $Z = \frac{\hat{\beta}_1 - 0}{\sqrt{\frac{\sigma^2_{Y|x}}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}}$. When the level of significance is α and $\sigma^2_{Y|x}$ is known, then $H_0 : \beta_1 = 0$ is rejected when $|Z| \geq |z_{\alpha/2}|$. When $\sigma^2_{Y|x}$ is not known, it is estimated by $\hat{\sigma}^2_{Y|x} = MSE$. This requires the use of the Student's t distribution.

The studentized form of the statistic is $T_{n-2} = \frac{\hat{\beta}_1 - 0}{\sqrt{\frac{MSE}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}}$. Then, $H_0 : \beta_1 = 0$ is rejected when $|T_{n-2}| \geq |t_{\alpha/2, n-2}|$. An equivalent approach is to use $TS = \frac{MS\ REG}{MSE} = F$. Under $H_0 : \beta_1 = 0$, the null distribution of F is a central F with 1 numerator and $n-2$ denominator degrees of freedom.

Confidence interval for β_1

When $\sigma^2_{Y|x}$ is known, the $(1-\alpha)$% confidence interval for β_1 is $\hat{\beta}_1 \pm |z_{\alpha/2}| \sqrt{\frac{\sigma^2_{Y|x}}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}$. When $\sigma^2_{Y|x}$ is not known, the $(1-\alpha)$% confidence interval for β_1 is $\hat{\beta}_1 \pm |t_{\alpha/2, n-2}| \sqrt{\frac{MSE}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}}$.

Distribution of the Estimated Intercept

Since $\hat{\beta}_0 = \bar{Y}_n - \hat{\beta}_1 \bar{x}_n$, $\hat{\beta}_0$ is normally distributed with

$E(\hat{\beta}_0) = E(\bar{Y}_n - \hat{\beta}_1 \bar{x}_n) = E(\bar{Y}_n) - E(\hat{\beta}_1 \bar{x}_n) = E(\bar{Y}_n) - \bar{x}_n E(\hat{\beta}_1) = E(\bar{Y}_n) - \beta_1 \bar{x}_n$ because $E(\hat{\beta}_1) = \beta_1$.

Now $E(\bar{Y}_n) = E(\dfrac{Y_1 + Y_2 + \cdots + Y_n}{n}) = \dfrac{E(Y_1) + E(Y_2) + \cdots + E(Y_n)}{n} = \dfrac{(\beta_0 + \beta_1 x_1) + \cdots + (\beta_0 + \beta_1 x_n)}{n}$,

with $E(\bar{Y}_n) = \dfrac{(\beta_0 + \beta_1 x_1) + \cdots + (\beta_0 + \beta_1 x_n)}{n} = \dfrac{n\beta_0 + \beta_1 \sum x_i}{n} = \beta_0 + \beta_1 \bar{x}_n$. Then

$E(\hat{\beta}_0) = E(\bar{Y}_n) - \beta_1 \bar{x}_n = \beta_0 + \beta_1 \bar{x}_n - \beta_1 \bar{x}_n = \beta_0$.

Finally,

$$\operatorname{var}(\hat{\beta}_0) = \operatorname{var}(\bar{Y}_n - \hat{\beta}_1 \bar{x}_n) = \operatorname{var}(\bar{Y}_n) + (\bar{x}_n)^2 \operatorname{var}(\hat{\beta}_1) - 2\bar{x}_n \operatorname{cov}(\bar{Y}_n, \hat{\beta}_1) = \frac{\sigma_{Y|x}^2}{n} + \frac{(\bar{x}_n)^2 \sigma_{Y|x}^2}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2} - 2 \bullet 0 .$$

In summary, $\hat{\beta}_0$ is $N(\beta_0, \dfrac{\sigma_{Y|x}^2}{n} + \dfrac{(\bar{x}_n)^2 \sigma_{Y|x}^2}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2})$

Confidence Interval for $\hat{Y}(x)$

Since $\hat{Y}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$,

$E[\hat{Y}(x)] = E(\hat{\beta}_0 + \hat{\beta}_1 x) = E(\hat{\beta}_0) + xE(\hat{\beta}_1) = \beta_0 + \beta_1 x$.

For its variance,

$\operatorname{var}[\hat{Y}(x)] = \operatorname{var}[\bar{Y}_n + \hat{\beta}_1 (x - \bar{x}_n)] = \operatorname{var}(\bar{Y}_n) + (x - \bar{x}_n)^2 \operatorname{var}(\hat{\beta}_1) + 2(x - \bar{x}_n) \operatorname{cov}(\bar{Y}_n, \hat{\beta}_1)$, with

$$\operatorname{var}[\hat{Y}(x)] = \frac{\sigma_{Y|x}^2}{n} + \frac{(x - \bar{x}_n)^2 \sigma_{Y|x}^2}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2} + 2(x - \bar{x}_n) \bullet 0 = \sigma_{Y|x}^2 \left(\frac{1}{n} + \frac{(x - \bar{x}_n)^2}{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}\right) .$$

In summary, $\hat{Y}(x)$ has the normal distribution $N(\beta_0 + \beta_1 x, \sigma^2_{Y|x}(\frac{1}{n} + \frac{(x-\bar{x}_n)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}))$. When $\sigma^2_{Y|x}$ is known, the 95% confidence interval for $E[\hat{Y}(x)] = \beta_0 + \beta_1 x$ is

$$\hat{Y}(x) \pm 1.960\sqrt{\sigma^2_{Y|x}(\frac{1}{n} + \frac{(x-\bar{x}_n)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2})}.$$

When $\sigma^2_{Y|x}$ is not known, an estimate of $\sigma^2_{Y|x}$ is used, and the t-percentile is used rather than the z-percentile, here 1.960. If the four assumptions are met, then $E(MSE) = \sigma^2_{Y|x}$.

The 95% confidence interval for $E[\hat{Y}(x)] = \beta_0 + \beta_1 x$ is then

$$\hat{Y}(x) \pm t_{1.960,n-2}\sqrt{MSE(\frac{1}{n} + \frac{(x-\bar{x}_n)^2}{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2})}.$$