1. Introduction

The principal purpose of this lecture is to demonstrate how matrices can be used to simplify the development of statistical models.

A secondary purpose is to review, and extend, some material in linear models.

I will take up the following topics:

- Expressing linear models for regression, dummy regression, and analysis of variance in matrix form.
- Deriving the least-squares coefficients using matrices.
- Distribution of the least-squares coefficients.
- The least-squares coefficients as maximum-likelihood estimators.
- Statistical inference for linear models.
2. Linear Models in Matrix Form

The general linear model is

$$ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i $$

where

- $y_i$ is the value of the response variable for the $i$th of $n$ observations.
- $x_{i1}, x_{i2}, \ldots, x_{ik}$ are the values of $k$ regressors for observation $i$. In linear regression analysis, $x_{i1}, x_{i2}, \ldots, x_{ik}$ are the values of $k$ quantitative explanatory variables.
- $\beta_0, \beta_1, \ldots, \beta_k$ are $k+1$ parameters to be estimated from the data, including the constant or intercept term, $\beta_0$.
- $\varepsilon_i$ is the random error variable for the $i$th observation.

The statistical assumptions of the linear model concern the behaviour of the errors; the standard assumptions include:

- **Linearity**: The average error is zero, $E(\varepsilon_i) = 0$; equivalently, $E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$.
- **Constant error variance**: The variance of the errors is the same for all observations, $V(\varepsilon_i) = \sigma^2_\varepsilon$; equivalently, $V(y_i) = \sigma^2_\varepsilon$.
- **Normality**: The errors are normally distributed, and so $\varepsilon_i \sim N(0, \sigma^2_\varepsilon)$; equivalently, $y_i \sim N(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}, \sigma^2_\varepsilon)$.
- **Independence**: The errors are independently sampled — that is $\varepsilon_i$ and $\varepsilon_j$ are independent for $i \neq j$; equivalently, $y_i$ and $y_j$ are independent.
- Either the $x$-values are fixed (with respect to repeated sampling) or, if random, the $x$s are independent of the errors.
The linear model may be rewritten as

\[ y_i = \begin{bmatrix} 1, x_{i1}, x_{i2}, \ldots, x_{ik} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \varepsilon_i \]

\[ = x'_i \beta + \varepsilon_i \]

There is one such equation for each observation, \( i = 1, \ldots, n \).

Collecting these \( n \) equations into a single matrix equation:

\[ \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \]

\[ = X \beta + \varepsilon \]

– The \( X \) matrix in the linear model is called the model matrix (or the design matrix).
– Note the column of 1s for the constant.
Similarly, the assumptions of linearity, constant variance, normality, and independence can be recast as

$$\varepsilon \sim N_n(0, \sigma^2 I_n)$$

where $N_n(0, \sigma^2 I_n)$ denotes the multivariate-normal distribution with

- mean vector 0,
- and covariance matrix

$$\sigma^2 I_n = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

- equivalently,

$$y \sim N_n(X\beta, \sigma^2 I_n)$$

2.1 Dummy Regression Models

The matrix equation $y = X\beta + \varepsilon$ suffices not just for linear regression models, but — with suitable specification of the regressors — for linear models generally.

For example, consider the dummy-regression model

$$y_i = \alpha + \beta x_i + \gamma d_i + \delta(x_i d_i) + \varepsilon_i$$

where

- $y$ is income in dollars,
- $x$ is years of education,
- and the dummy regressor $d$ is coded 1 for men and 0 for women.
Recall that this model implies potentially different intercepts and slopes — that is, potentially different regression lines — for the two groups:

- for men,
  \[ y_i = \alpha + \beta x_i + \gamma 1 + \delta(x_i 1) + \varepsilon_i \]
  \[ = (\alpha + \gamma) + (\beta + \delta) x_i + \varepsilon_i \]

- for women
  \[ y_i = \alpha + \beta x_i + \gamma 0 + \delta(x_i 0) + \varepsilon_i \]
  \[ = \alpha + \beta x_i + \varepsilon_i \]

- and so \( \gamma \) is the difference in intercepts between men and women, and \( \delta \) is the difference in slopes.

- Because men and women can have different slopes, this model permits gender to interact with education in determining income.

Written as a matrix equation, the dummy-regression model becomes.

\[
\begin{bmatrix}
  y_1 \\
  \vdots \\
  y_{n_1} \\
  y_{n_1+1} \\
  \vdots \\
  y_n
\end{bmatrix} =
\begin{bmatrix}
  1 & x_1 & 0 & 0 \\
  \vdots & \vdots & \vdots & \vdots \\
  1 & x_{n_1} & 0 & 0 \\
  \vdots & \vdots & \vdots & \vdots \\
  1 & x_{n_1+1} & 1 & x_{n_1+1} \\
  \vdots & \vdots & \vdots & \vdots \\
  1 & x_n & 1 & x_n
\end{bmatrix}
\begin{bmatrix}
  \alpha \\
  \beta \\
  \gamma \\
  \delta
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_1 \\
  \vdots \\
  \varepsilon_{n_1} \\
  \varepsilon_{n_1+1} \\
  \vdots \\
  \varepsilon_n
\end{bmatrix}
\]

where, for clarity, the \( n_1 \) observations for women precede the \( n - n_1 \) observations for men.
Reminder: When a categorical explanatory variable has more than two (say, \( m \)) categories, it generates a set of \( m - 1 \) dummy regressors — that is, one fewer dummy variable than the number of categories.

- For example, a five-category regional classification might produce the following four dummy regressors:

<table>
<thead>
<tr>
<th>Region</th>
<th>( d_1 )</th>
<th>( d_2 )</th>
<th>( d_3 )</th>
<th>( d_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Quebec</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ontario</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Prairies</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- Here, BC is arbitrarily selected as the baseline category, to which other categories will implicitly be compared.

2.2 Analysis of Variance Models

Analysis of variance or ANOVA models are linear models in which all of the explanatory variables are factors — that is, categorical variables.

The simplest case is one-way ANOVA, where there is a single factor.

- The one-way ANOVA model is usually written with double-subscript notation as

\[
y_{ij} = \mu + \alpha_j + \varepsilon_{ij}
\]

for levels \( j = 1, \ldots, m \) of the factor, and observations \( i = 1, \ldots, n_j \) of level \( j \).
The matrix form of the one-way ANOVA model is:

\[
\begin{align*}
\begin{bmatrix}
  y_{11} \\
  \vdots \\
  y_{n1,1}
\end{bmatrix}
&= 
\begin{bmatrix}
  1 & 1 & 0 & \cdots & 0 & 0 \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  1 & 1 & 0 & \cdots & 0
\end{bmatrix}
\begin{bmatrix}
  \mu \\
  \alpha_1 \\
  \alpha_2 \\
  \vdots \\
  \alpha_{m-1} \\
  \alpha_m
\end{bmatrix}
+ 
\begin{bmatrix}
  \varepsilon_{11} \\
  \vdots \\
  \varepsilon_{n1,1}
\end{bmatrix}
\end{align*}
\]

This formulation of the model is problematic because there is a redundant column in the model matrix (which is therefore of deficient rank \(m\)):

- For example, the first column is the sum of the remaining columns.
- This will create a problem when we try to fit the model by least squares, but more fundamentally, it reflects a redundancy among the parameters of the model.

A common solution to the problem is to reduce the parameters by one. There are many ways to do this, all providing equivalent fits to the data. For example:

- Eliminating the constant, \(\mu\), produces a so-called means model,

  \[
  y_{ij} = \alpha_j + \varepsilon_{ij}
  \]

  where \(\alpha_j\) now represents the population mean of level \(j\).

- Eliminating one of the \(\alpha_j\) produces a dummy-variable solution, with the omitted coefficient corresponding to the baseline category (here category \(m\)).
Alternatively, we can place a linear constraint on the parameters, most commonly, the \textit{sigma constraint}

\[
\sum_{j=1}^{m} \alpha_j = 0
\]

- Under this constraint

\[
\alpha_m = - \sum_{j=1}^{m-1} \alpha_j
\]

need not appear explicitly, producing the model matrix
3. Least-Squares Fit

The fitted linear model is

\[ y = Xb + e \]

where

- \( b = [b_0, b_1, ..., b_k]' \) is the vector of fitted coefficients.
- \( e = [e_1, e_2, ..., e_n]' = y - Xb \) is the vector of residuals.

We want the coefficient vector \( b \) that minimizes the residual sum of squares, expressed as a function of \( b \):

\[
S(b) = \sum e_i^2 = e'e = (y - Xb)'(y - Xb)
\]

\[
= y'y - y'Xb - b'X'y + b'X'Xb
\]

\[
= y'y - (2y'X)b + b'(X'X)b
\]

- The last line of the equation is justified because \( y'X(b) \) and \( b'(X'X)b \) are both scalars, and consequently equal.
Noting that \( y'y \) is a constant (with respect to \( b \)), \((2y'X)b\) is a linear function of \( b \), and \( b'(X'X)b\) is a quadratic form in \( b \),

\[
\frac{\partial S(b)}{\partial b} = 0 - 2X'y + 2X'Xb
\]

- Setting the derivative to 0 produces the *normal equations* for the linear model

\[
-2X'y + 2X'Xb = 0
\]

\[
X'Xb = X'y
\]

a system of \( k+1 \) linear equations in \( k+1 \) unknowns (i.e., \( b_0, b_1, \ldots, b_k \)).

- We can solve the normal equations uniquely for \( b \) if as the \((k + 1) \times (k + 1)\) matrix \( X'X \) is nonsingular, which will be the case as long as

  - there are at least as many observations as coefficients — that is, \( n \geq k + 1 \).
  - no column of the model matrix \( X \) is a perfect linear function of the other columns.

- When \( X'X \) is nonsingular, the least-squares solution is

\[
b = (X'X)^{-1}X'y
\]

- Looking inside of the matrices in the normal equations,

  - the matrix \( X'X \) contains sums of squares and cross-products for the regressors (including the column of 1s).
  - \( X'y \) contains sums of products between the regressors and the response.

- The normal equations, therefore, are

\[
\begin{align*}
b_0n &+ b_1 \sum x_{i1} + \cdots + b_k \sum x_{ik} = \sum y_i \\
b_0 \sum x_{i1} &+ b_1 \sum x_{i1}^2 + \cdots + b_k \sum x_{i1}x_{ik} = \sum x_{i1}y_i \\
\vdots &\vdots \\
b_0 \sum x_{ik} &+ b_1 \sum x_{ik}x_{i1} + \cdots + b_k \sum x_{ik}^2 = \sum x_{ik}y_i
\end{align*}
\]

- An example, using Duncan’s regression of occupational prestige on the income and education levels of 45 U.S. occupations:
• Matrices of sums of squares and products:
\[
X'X = \begin{bmatrix}
45 & 1884 & 2365 \\
1884 & 105,148 & 122,197 \\
2365 & 122,197 & 163,265 \\
\end{bmatrix}
\]
\[
X'y = \begin{bmatrix}
2146 \\
118,229 \\
147,936 \\
\end{bmatrix}
\]
• The inverse of \(X'X\):
\[
(X'X)^{-1} = \begin{bmatrix}
0.1021058996 & -0.0008495732 & -0.0008432006 \\
-0.0008495732 & 0.0000801220 & -0.0000476613 \\
-0.0008432006 & -0.0000476613 & 0.0000540118 \\
\end{bmatrix}
\]
• The regression coefficients:
\[
b = (X'X)^{-1}X'y = \begin{bmatrix}
-6.06466 \\
0.59873 \\
0.54583 \\
\end{bmatrix}
\]

4. Distribution of the Least-Squares Coefficients

It is simple to show that least-squares coefficients are unbiased estimators of the population regression coefficients:
\[
b = (X'X)^{-1}X'y
\]
and so (assuming a fixed model matrix \(X\)),
\[
E(b) = (X'X)^{-1}X'E(y) = (X'X)^{-1}X'(X\beta) = \beta
\]
The covariance matrix of \(b\) follows from the covariance matrix of \(y\), which is \(\sigma^2\varepsilon I_n\):
\[
V(b) = \left[(X'X)^{-1}X'y\right]'V(y)\left[(X'X)^{-1}X'y\right]
\]
\[
= \left[(X'X)^{-1}X'y\right]'\sigma^2\varepsilon I_n\left[(X'X)^{-1}X'y\right]
\]
\[
= \sigma^2\varepsilon(X'X)^{-1}X'X(X'X)^{-1}
\]
\[
= \sigma^2\varepsilon(X'X)^{-1}
\]
• Because the error variance $\sigma^2_\varepsilon$ is an unknown parameter, the covariance matrix of $\mathbf{b}$ must be estimated:

$$\hat{V}(\mathbf{b}) = s^2_\varepsilon (\mathbf{X}'\mathbf{X})^{-1}$$

where

$$s^2_\varepsilon = \frac{\sum e_i^2}{n-k-1}$$

is the estimated error variance, and $e_i$ is the residual for observation $i$.

Because the response vector $\mathbf{y}$ is multinormally distributed, so is $\mathbf{b}$; that is

$$\mathbf{b} \sim N_{k+1} \left[ \boldsymbol{\beta}, \sigma^2_\varepsilon (\mathbf{X}'\mathbf{X})^{-1} \right]$$

Notice the strong analogy between the formulas for the slope coefficient in least-squares simple regression (i.e., with a single $x$) and for the coefficients of the linear model in matrix form:

<table>
<thead>
<tr>
<th></th>
<th>Simple Regression</th>
<th>Linear Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td>$y_i = \alpha + \beta x_i + \varepsilon_i$</td>
<td>$y = \mathbf{X}\boldsymbol{\beta} + \varepsilon$</td>
</tr>
<tr>
<td></td>
<td>$y^* = x^* \beta + \varepsilon$</td>
<td></td>
</tr>
<tr>
<td><strong>Least-Squares Estimator</strong></td>
<td>$b = \frac{\sum x^* y^<em>}{\sum x^</em> x^*}$</td>
<td>$b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$</td>
</tr>
<tr>
<td></td>
<td>= $(\sum x^* x^<em>)^{-1} \sum x^</em> y^*$</td>
<td></td>
</tr>
<tr>
<td><strong>Sampling Variance</strong></td>
<td>$V(b) = \frac{\sigma^2_\varepsilon}{\sum x^* x^*}$</td>
<td>$V(b) = \sigma^2_\varepsilon (\mathbf{X}'\mathbf{X})^{-1}$</td>
</tr>
<tr>
<td></td>
<td>= $\sigma^2_\varepsilon (\sum x^* x^*)^{-1}$</td>
<td></td>
</tr>
<tr>
<td><strong>Distribution</strong></td>
<td>$b \sim N\left[ \beta, \sigma^2_\varepsilon (\sum x^* x^*)^{-1} \right]$</td>
<td>$b \sim N_{k+1} \left[ \beta, \sigma^2_\varepsilon (\mathbf{X}'\mathbf{X})^{-1} \right]$</td>
</tr>
</tbody>
</table>
• In the scalar formulas the following short-hand notation is used:

\[
x^* = x_i - \bar{x} \\
y^* = y_i - \bar{y}
\]

5. Maximum-Likelihood Estimation of the Normal Linear Model

► The standard assumptions of the linear model provide a probability model for the data \( y \) (thinking of the model matrix \( X \) as fixed or conditioning on it):

\[
y \sim N_n(X\beta, \sigma^2 \mathbf{I}_n)
\]

• Then, from the formula for the normal distribution,

\[
p(y) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2} \right]
\]

– Note: \( \exp(a) \) in a formula means \( e^a \), for the constant \( e \approx 2.718 \).

► In maximum-likelihood estimation, recall, we find the values of the parameters that make the probability of observing the data as high as possible.
• The likelihood function is the same as the probability (or probability-density) of the data, except thought of as a function of the parameters.

• Here,
\[
L(\beta, \sigma^2_\varepsilon) = \left(2\pi\sigma^2_\varepsilon\right)^{-n/2} \exp \left[-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2_\varepsilon}\right]
\]

► As is usually the case, it is simpler to work with the log of the likelihood.

• Whatever values of the parameters maximize the log-likelihood also maximize the likelihood, since the log function is monotone (strictly increasing).

• For the linear model:
\[
\log L(\beta, \sigma^2_\varepsilon) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2_\varepsilon - \frac{1}{2\sigma^2_\varepsilon}(y - X\beta)'(y - X\beta)
\]

• To justify this result, recall that taking logs turns multiplication into addition, division into subtraction, and exponentiation into multiplication; moreover, \(\log e^a = a\).

► To maximize the log-likelihood, we need its derivatives with respect to the parameters.

• Finding the derivatives is simplified by noticing that \((y - X\beta)'(y - X\beta)\) is just the sum of squared errors.

• Differentiating,
\[
\frac{\partial \log L(\beta, \sigma^2_\varepsilon)}{\partial \beta} = -\frac{1}{2\sigma^2_\varepsilon}(2X'y\beta - 2X'y)
\]
\[
\frac{\partial \log L(\beta, \sigma^2_\varepsilon)}{\partial \sigma^2_\varepsilon} = -\frac{n}{2} \left(\frac{1}{\sigma^2_\varepsilon}\right) + \frac{1}{\sigma^2_\varepsilon}(y - X\beta)'(y - X\beta)
\]

• Setting the partial derivatives to 0 and solving for maximum-likelihood estimates of the parameters produces
\[
\hat{\beta} = (X'X)^{-1}X'y
\]
\[
\hat{\sigma}^2_\varepsilon = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n} = \frac{e'e}{n}
\]
where \(e = y - X\hat{\beta}\) is the vector of residuals.
Notice that

- The MLE \( \hat{\beta} \) is just the least-squares coefficients \( \beta \).
- The MLE of the error variance, \( \hat{\sigma}_e^2 = \sum e_i^2 / n \) is biased.
  - The usual unbiased estimator, \( s_e^2 \), divides by residual degrees of freedom \( n - k - 1 \) rather than by \( n \).
  - The MLE is consistent, however, since the bias (along with the variance of the estimator) goes to zero as \( n \) get larger.

6. Statistical Inference for Least-Squares Estimation

Statistical inference for \( \beta \) based on the least-squares coefficients \( b \) uses the estimated covariance matrix \( \hat{V}(b) = s_e^2(X'X)^{-1} \).

The simplest case is inference for an individual coefficient, \( b_j \):

- The standard error of the coefficient is the square root of the \( j \)th diagonal entry of the estimated covariance matrix (indexing the matrix from 0):
  \[
  SE(b_j) = \sqrt{s_e^2[(X'X)^{-1}]_{jj}}
  \]

- Because the error variance has been estimated, hypothesis tests and confidence intervals use the \( t \)-distribution with \( n - k - 1 \) degrees of freedom.
• For example:
  – To test
    \[ H_0: \beta_j = 0 \]
    we compute
    \[ t_0 = \frac{b_j}{\text{SE}(b_j)} \]
  – To form a 95-percent confidence interval for \( \beta_j \) we take
    \[ \beta_j = b_j \pm t_{.975,n-k-1} \text{SE}(b_j) \]
    where \( t_{.975,n-k-1} \) is the .975 quantile of the \( t \)-distribution with \( n - k - 1 \) degrees of freedom.

More generally, suppose that we want to test the linear hypothesis

\[ H_0: \begin{bmatrix} \mathbf{L} \end{bmatrix}_{(q \times k+1)(k+1 \times 1)} \begin{bmatrix} \mathbf{\beta} \end{bmatrix}_{(q \times 1)} = \begin{bmatrix} \mathbf{c} \end{bmatrix}_{(q \times 1)} \]

where the hypothesis matrix \( \mathbf{L} \) and the right-hand-side vector \( \mathbf{c} \) (usually \( 0 \)) encode the hypothesis.

• For example, in Duncan’s regression of prestige on income and education, the hypothesis matrix
  \[
  \mathbf{L} = \begin{bmatrix}
  0 & 1 & 0 \\
  0 & 0 & 1 
  \end{bmatrix}
  \]
  and right-hand-side vector
  \[
  \mathbf{c} = \begin{bmatrix}
  0 \\
  0 
  \end{bmatrix}
  \]
  specify the hypothesis
  \[ H_0: \beta_1 = 0, \beta_2 = 0 \]
• Likewise, again for Duncan’s regression, the one-row hypothesis matrix
  \[ L = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \]
  and right-hand-side \( c = [0] \) correspond to the hypothesis
  \( H_0: \beta_1 - \beta_2 = 0 \)
  that is
  \( H_0: \beta_1 = \beta_2 \)
• Under the hypothesis \( H_0 \), the statistic
  \[ F_0 = \frac{(Lb - c)' \left[ L(X'X)^{-1}L \right]^{-1} (Lb - c)}{q s_e^2} \]
  follows an \( F \)-distribution with \( q \) and \( n - k - 1 \) degrees of freedom.

► Example: For Duncan’s regression, the sum of squared residuals is
  \( e'e = 7506.699 \), and so
  \[ s_e^2 = \frac{7506.699}{45 - 2 - 1} = 178.7309 \]
• The estimated covariance matrix of the least-squares coefficients is
  \[ \hat{V}(b) = s_e^2 (X'X)^{-1} \]
  \[ = 178.7309 \begin{bmatrix} 0.1021058996 & -0.0008495732 & -0.0008432006 \\ -0.0008495732 & 0.0000476613 & 0.0000540118 \\ -0.0008432006 & -0.0000476613 & 0.0000540118 \end{bmatrix} \]
  \[ = \begin{bmatrix} 18.249387 & -0.151844 & -0.150705 \\ -0.151844 & 0.014320 & -0.008519 \\ -0.150705 & -0.008519 & 0.009653 \end{bmatrix} \]
• The estimated standard errors of the regression coefficients are, therefore,
  \[
  \text{SE}(b_0) = \sqrt{18.249387} = 4.272 \\
  \text{SE}(b_1) = \sqrt{0.014320} = 0.1197 \\
  \text{SE}(b_2) = \sqrt{0.009653} = 0.09825
  \]

• and, a 95-percent confidence interval for \( \beta_1 \) (the income coefficient) is
  \[
  \beta_1 = 0.5987 \pm 2.0181 \times 0.1197 \\
  = 0.5987 \pm 0.2416
  \]

• To test the hypothesis that both slope coefficients are 0,
  \[ H_0: \beta_1 = \beta_2 = 0 \]
  we have
  \[
  L = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
  Lb = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -6.06466 \\ 0.59873 \\ 0.54583 \end{bmatrix} = \begin{bmatrix} 0.59873 \\ 0.54583 \end{bmatrix} \text{ (i.e., the two slopes)}
  \]
\[ F_0 = \frac{(Lb)' \left[ L(X'X)^{-1}L' \right]^{-1} Lb}{qs_e^2} \]

\[
= \frac{[0.599, 0.546] \left[ \begin{array}{c} 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right] \left[ \begin{array}{ccc} 0.1021 & -0.0008 & -0.0008 \\ -0.0008 & 0.0001 & -0.0000 \\ -0.0008 & -0.0000 & 0.0001 \end{array} \right] \left[ \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right]}{2 \times 178.7309}
\]

\[ = 101.22 \text{ with 2 and 42 degrees of freedom, } p \approx 0 \]

---

- To test the hypothesis that the slopes are equal:
  \[ L = \left[ \begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right] \]

\[ Lb = \left[ \begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right] \left[ \begin{array}{c} -6.06466 \\ 0.59873 \\ 0.54583 \end{array} \right] = 0.05290 \text{ (i.e., the difference in slopes)} \]

\[ F_0 = \frac{(Lb)' \left[ L(X'X)^{-1}L' \right]^{-1} Lb}{qs_e^2} \]

\[
= \frac{0.053 \left[ \begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right] \left[ \begin{array}{ccc} 0.1021 & -0.0008 & -0.0008 \\ -0.0008 & 0.0001 & -0.0000 \\ -0.0008 & -0.0000 & 0.0001 \end{array} \right] \left[ \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right]}{1 \times 178.7309}
\]

\[ = 0.068 \text{ with 1 and 42 degrees of freedom, } p = .80 \]