STAT 210
Applied Statistics and Data Analysis
Linear Regression II:
Matrix Formulation and Tests

Joaquin Ortega

Fall 2020

Ordinary Least Squares in Matrix Notation

Ordinary Least Squares in Matrix Notation

Recall that the error sum of squares in matrix notation is given by

$$SSE = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta). \tag{1}$$

Multiplying out the terms in this expression we have

$$SSE = \mathbf{Y}'\mathbf{Y} - \beta'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta. \tag{2}$$

Since $\beta'\mathbf{X}'\mathbf{Y}$ is a scalar, it is equal to its transpose so

$$\beta'\mathbf{X}'\mathbf{Y} = (\beta'\mathbf{X}'\mathbf{Y})' = \mathbf{Y}'\mathbf{X}\beta$$

and we get

$$SSE = \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta. \tag{3}$$

The derivative of SSE is given by

$$\frac{\partial SSE}{\partial \beta} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta$$

Ordinary Least Squares in Matrix Notation

Setting this expression equal to zero and solving for β we obtain

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{4}$$

which is the matrix version of the normal equations.

Recall that for the simple linear regression model

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

and therefore

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}.$$

Ordinary Least Squares in Matrix Notation

Recall that if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

In our case

$$\det(\mathbf{X}'\mathbf{X}) = n\sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2 = n\sum_{i=1}^{n}(x_i - \bar{x})^2$$

and

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n. \end{pmatrix} \qquad (5)$$

Also,

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}.$$

Ordinary Least Squares in Matrix Notation

Therefore

$$
\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \frac{1}{n\sum_{i=1}^{n}(x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix} \\
&= \frac{1}{n\sum_{i=1}^{n}(x_i - \bar{x})^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i \\ -\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i + n\sum_{i=1}^{n} x_i y_i \end{pmatrix} \\
&= \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}
\end{aligned} \tag{6}
$$

It is possible to prove that this coincides with the expressions we obtained in a previous video.

Sampling Distribution of $\hat{\beta}$.

Sampling Distribution of $\hat{\beta}$.

We have assumed that $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ and therefore

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}).$$

Since $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, $\hat{\beta}$ also has a normal distribution.

Let us now see that $E(\hat{\beta}) = \beta$.

$$\begin{aligned} E[\hat{\beta}] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \boldsymbol{\epsilon})] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon})] \\ &= E[\mathbf{I}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\epsilon})] \\ &= \beta. \end{aligned}$$

so $\hat{\beta}$ is an unbiassed estimator of β.

Sampling Distribution of $\hat{\beta}$.

Let us calculate the variance-covariance matrix for

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Let $A = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, then $\hat{\beta} = A\mathbf{Y}$ and the covariance matrix of $\hat{\beta}$ is

$$\begin{aligned}
Var(\hat{\beta}) &= E\left[(\hat{\beta} - E[\hat{\beta}])(\hat{\beta} - E[\hat{\beta}])'\right] \\
&= E\left[(A\mathbf{Y} - E[A\mathbf{Y}])(A\mathbf{Y} - E[A\mathbf{Y}])'\right] \\
&= E\left[A(\mathbf{Y} - E[\mathbf{Y}])(A(\mathbf{Y} - E[\mathbf{Y}]))'\right] \\
&= E\left[A(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{Y} - E[\mathbf{Y}])'A'\right] \\
&= AE\left[(\mathbf{Y} - E[\mathbf{Y}])(\mathbf{Y} - E[\mathbf{Y}])'\right]A' \\
&= A\sigma^2\mathbf{I}_n A'.
\end{aligned}$$

Sampling Distribution of $\hat{\beta}$.

Thus

$$
\begin{aligned}
Var(\hat{\beta}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}
$$

Therefore

$$\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{7}$$

Sampling Distribution of $\hat{\boldsymbol{\beta}}$.

Consequently, $\hat{\boldsymbol{\beta}}$ is a linear unbiased estimator for $\boldsymbol{\beta}$.

In fact, it is the best such estimator in the sense that the variances of $\hat{\beta}_0$ and $\hat{\beta}_1$ are minimal. $\hat{\boldsymbol{\beta}}$ is known as a Best Linear Unbiased Estimator (BLUE).

The unbiased estimator for the error variance σ^2 is

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n} \hat{\epsilon}_i^2. \tag{8}$$

Sampling Distribution of $\hat{\beta}$.

Let $V = (\mathbf{X}'\mathbf{X})^{-1}$, then

$$Var(\hat{\beta}_0) = \sigma^2 v_{11}, \qquad Var(\hat{\beta}_1) = \sigma^2 v_{22} \tag{9}$$

where v_{ii} is the i-th diagonal element of V.

V can be obtained with the command

```
summary(lm.object)$cov.unscaled
```

where `lm.object` is a linear model object.

Sampling Distribution of $\hat{\beta}$.

Since

$$\hat{\beta} \sim N(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})) = N(\beta, Var(\hat{\beta})),$$

an estimate of $Var(\hat{\beta})$ is

$$\begin{aligned} s^2(\hat{\beta}) &= \begin{pmatrix} s^2(\hat{\beta}_0) & s^2(\hat{\beta}_0, \hat{\beta}_1) \\ s^2(\hat{\beta}_1, \hat{\beta}_0) & s^2(\hat{\beta}_1) \end{pmatrix} \\ &= \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= MSE \cdot (\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \tag{10}$$

The `R` function `vcov()` computes $s^2(\hat{\beta})$ when applied to a linear model object.

Back to Example 1

Let us consider again the first model we fitted (`lm1`) to the `crabs` data,

- the first command below gives matrix $V = (\mathbf{X}'\mathbf{X})^{-1}$,
- the second gives $s^2(\hat{\beta})$,
- the third recovers the estimated standard deviation for the residuals and
- the fourth verifies that $s^2(\hat{\beta}) = \hat{\sigma}^2 V$.

```
summary(lm1)$cov.unscaled
```

```
##              (Intercept)            CL
## (Intercept)  0.107204469 -3.183394e-03
## CL          -0.003183394  9.915418e-05
```

Back to Example 1

```r
vcov(lm1)
```

```
##              (Intercept)            CL
## (Intercept)  0.055114505 -0.0016366035
## CL          -0.001636603  0.0000509758
```

```r
summary(lm1)$sigma
```

```
## [1] 0.7170121
```

```r
(summary(lm1)$sigma^2)*summary(lm1)$cov.unscaled
```

```
##              (Intercept)            CL
## (Intercept)  0.055114505 -0.0016366035
## CL          -0.001636603  0.0000509758
```

Hypothesis tests

Hypothesis tests

A test statistic for $H_0 : \beta_i = \beta_{i,0}$ versus $H_1 : \beta_i \neq \beta_{i,0}$ can be obtained from the pivotal quantity

$$\frac{\hat{\beta}_i - \beta_{i,0}}{s(\hat{\beta}_i)} \tag{11}$$

for $i = 0, 1$, which has a t_{n-2} distribution under H_0.

For the test $H_0 : \beta_i = 0$ versus $H_1 : \beta_i \neq 0$, the function `summary` applied to a linear model object will provide

$$t_{obs} = \hat{\beta}_i / s(\hat{\beta})$$

and the corresponding p-value:

$$p - \text{value} = 2P(t_{n-2} \geq |t_{obs}|).$$

Back to Example 1

Let us review again the results for the initial model

```r
summary(lm1)
```

```
##
## Call:
## lm(formula = FL ~ CL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.86395 -0.51746 -0.02826  0.50456  1.77009
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.15316    0.23477   0.652    0.515
## CL           0.48060    0.00714  67.313   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.717 on 198 degrees of freedom
## Multiple R-squared:  0.9581, Adjusted R-squared:  0.9579
## F-statistic:  4531 on 1 and 198 DF,  p-value: < 2.2e-16
```

Back to Example 1

The summary results show that for the intercept, the null hypothesis of zero intercept cannot be rejected at the usual levels while for the slope the null hypothesis of zero slope would be rejected.

The p-levels are 0.515 and $< 2e - 16$, respectively.

Confidence Intervals

Confidence Intervals

Using the pivotal quantity (11) we can get confidence intervals for the parameters of the model.

A $100(1-\alpha)\%$ confidence interval for $\beta_i, i = 0, 1$ is

$$CI_{1-\alpha}(\beta_i) = \left(\hat{\beta}_i - t_{n-2,1-\alpha/2}, \hat{\beta}_i + t_{n-2,1-\alpha/2}\right)$$

Example 2: GPA and SAT

This is an example from Ugarte, Militino and Arnholt, *Probability and Statistics with R*, Chapman and Hall 2008.

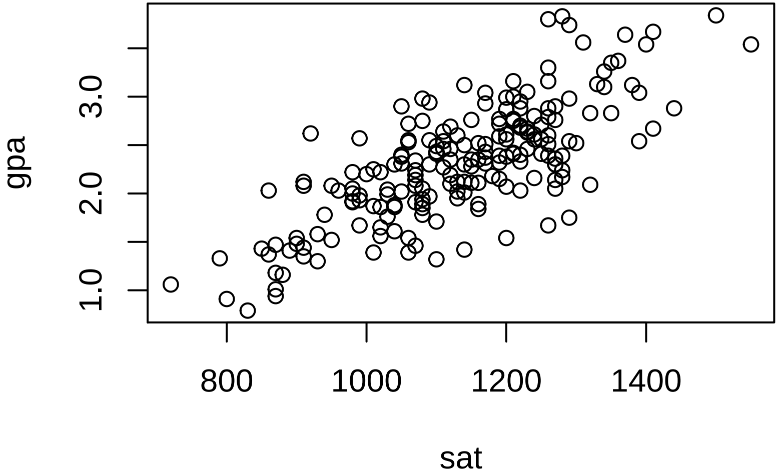
The data set `Grades` has information about the grades for 200 first-year students at a comprehensive state university in the USA.

The scores correspond to first semester college grade point average (GPA) and scholastic aptitude test (sat) scores.

We want to study the relation between these two scores.

Example 2: GPA and SAT

```r
library(PASWR); attach(Grades)
plot(sat, gpa)
```



Example 2: GPA and SAT

First, we are going to use equations

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{12}$$

and

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

to obtain the estimated parameters for the regression line:

```
y <- gpa
x <- sat
b1 <- sum((x-mean(x))*(y-mean(y)))/sum((x-mean(x))^2)
b0 <- mean(y)-b1*mean(x)
c(b0, b1)
```

```
## [1] -1.19206381  0.00309427
```

Example 2: GPA and SAT

Next, we use the normal equations in matrix form

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{13}$$

```
X <- cbind(rep(1,200), x)
Y <- matrix(y, ncol=1)
betahat <- solve(t(X)%*%X)%*%t(X)%*%Y
t(betahat)
```

```
##                           x
## [1,] -1.192064 0.00309427
```

Example 2: GPA and SAT

Finally, we use the `lm` function

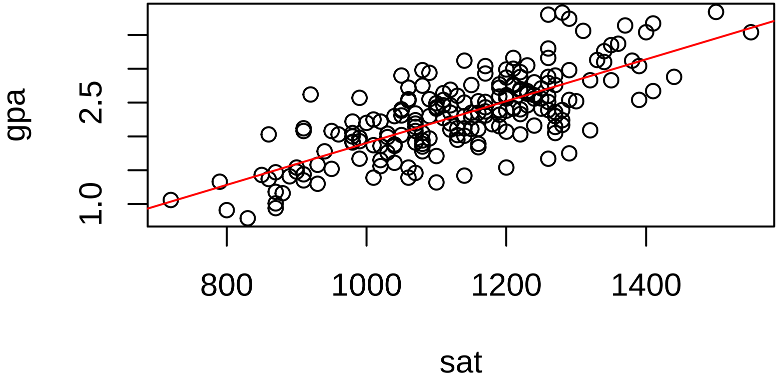
```
model <- lm(y~x)
model$coefficients
```

```
## (Intercept)           x
## -1.19206381  0.00309427
```

Example 2: GPA and SAT

The estimated regression line is $\hat{y} = -1.192 + 0.003094x$. Therefore, an increase in 100 points in the sat test results in an increase of 0.3094 points in the gpa score.

```r
plot(sat, gpa); abline(model, col='red')
```



Example 2: GPA and SAT

Let us now find the covariance matrix for $\hat{\beta}$. Recall that

$$s^2(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} = MSE \cdot (\mathbf{X}'\mathbf{X})^{-1}$$

```
XTX <- t(X)%*%X
solve(XTX)
```

```
##                              x
##    0.310137964 -2.689270e-04
## x -0.000268927  2.370131e-07
```

Example 2: GPA and SAT

Another way to obtain this matrix is

```r
(XTXb <- summary(model)$cov.unscaled)
```

```
##                (Intercept)             x
## (Intercept)  0.310137964 -2.689270e-04
## x           -0.000268927  2.370131e-07
```

The mean squared error (MSE) is

```r
sum(resid(model)^2)/(200-2)
```

```
## [1] 0.1595551
```

which can also be obtained using

```r
(mse <- summary(model)$sigma^2)
```

```
## [1] 0.1595551
```

Example 2: GPA and SAT

The covariance matrix for the estimators is given by

```
(beta.cov <- mse * XTXb)
```

```
##               (Intercept)             x
## (Intercept)  4.948408e-02 -4.290866e-05
## x           -4.290866e-05  3.781665e-08
```

which is also obtained with

```
vcov(model)
```

```
##               (Intercept)             x
## (Intercept)  4.948408e-02 -4.290866e-05
## x           -4.290866e-05  3.781665e-08
```

Example 2: GPA and SAT

We now want to test whether there is a linear relationship at the $\alpha = 0.01$ significance level. The test statistic is $\hat{\beta}_1 = 0.0030943$. Under the assumptions we have made for the model,

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2(\hat{\beta}_1)).$$

Under H_0 we have

$$\frac{\hat{\beta}_1 - \beta_1}{s(\hat{\beta}_1)} \sim t_{200-2}.$$

The quantile for the rejection region is

```r
qt(0.995,198)
```

```
## [1] 2.600887
```

Example 2: GPA and SAT

The value of the standardized observed statistic under H_0 is

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma(\hat{\beta}_1)} = \frac{0.0030943 - 0}{\sqrt{3.781665e - 08}}$$

```
model$coefficients[2]/sqrt(vcov(model)[2,2])
```

```
##        x
## 15.91171
```

The test statistic falls in the the rejection region and therefore the null hypothesis of no linear relation is rejected. The *p*-value is given by

```
2*(1-pt(15.912,198) )
```

```
## [1] 0
```

Example 2: GPA and SAT

This can also be obtained from the `summary` of the model,

```r
summary(model)$coef
```

```
##                Estimate  Std. Error  t value     Pr(>|t|)
## (Intercept) -1.19206381 0.222450180 -5.35879 2.316666e-07
## x            0.00309427 0.000194465 15.91171 2.922995e-37
```

Finally, we want confidence intervals for the regression coefficients β_0 and β_1 at the 99% level. The formulae for these intervals are

$$CI_{0.99}(\beta_0) = (\hat{\beta}_0 - t_{.995,n-p} s_{\hat{\beta}_0}, \hat{\beta}_0 + t_{.995,n-p} s_{\hat{\beta}_0})$$
$$CI_{0.99}(\beta_1) = (\hat{\beta}_1 - t_{.995,n-p} s_{\hat{\beta}_1}, \hat{\beta}_1 + t_{.995,n-p} s_{\hat{\beta}_1})$$

Example 2: GPA and SAT

These intervals can be obtained in `R` with the following commands

```r
confint(model, level=.99)
```

```
##                    0.5 %       99.5 %
## (Intercept) -1.770631656 -0.613495968
## x            0.002588489  0.003600052
```