Notes 9

Maryclare Griffin

3/23/2023

These notes are based on Chapters 2 and 6 of KNNL.

From now on, we will assume the **normal error linear regression model** for a dependent variable or response Y and independent variables, predictors, or covariates X_1, … X_{p-1} is defined as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where:

- The elements of $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$ are parameters
- The elements of the $n \times p$ matrix $\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{n,p-1} \end{pmatrix}$ are known constants
- $\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$ is a random error term elements that are ϵ_i that are independent and normally distributed with mean $E\left\{\epsilon_i\right\} = 0$ and variance $\sigma^2\left\{\epsilon_i\right\} = \sigma^2$.

Under the **normal error linear regression model**, we have shown that:

- $\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}$ is the least squares estimator of $\boldsymbol{\beta}$
- $\boldsymbol{b}$ is also the maximum likelihood estimator for $\boldsymbol{\beta}$
- $E\left\{\boldsymbol{b}\right\} = \boldsymbol{\beta}$
- $\sigma^2\left\{\boldsymbol{b}\right\} = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$
- Letting $E = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}$, $s^2 = \frac{1}{n-p}\sum_{i=1}^n e_i^2$ is unbiased for σ^2
- An estimator of $\sigma^2\left\{\boldsymbol{b}\right\} = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ is $\boldsymbol{s}^2\left\{\boldsymbol{b}\right\} = s^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$

Note: If $\boldsymbol{Z}$ is a $p \times 1$ vector of independent normal random variables with $p \times 1$ mean $E\{\boldsymbol{Z}\} = \boldsymbol{\mu}$ and $p \times p$ variance

$$\sigma^2\{\boldsymbol{Z}\} = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 \end{pmatrix}$$

and $\boldsymbol{A}$ is a fixed $p \times p$ matrix, then $\boldsymbol{V} = \boldsymbol{A}\boldsymbol{Z}$ is a **multivariate normal** random variable with mean $E\{\boldsymbol{V}\} = \boldsymbol{A}\boldsymbol{\mu}$ and variance $\sigma^2\{\boldsymbol{V}\} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}'$.

It follows that $\boldsymbol{b}$ is a multivariate normal random variable with mean $E\{\boldsymbol{b}\} = \boldsymbol{\beta}$ and variance $\sigma^2\{\boldsymbol{b}\} = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. What if we want to talk about the distribution of an element of $\boldsymbol{b}$, b_k?

Note: If $\boldsymbol{Z}$ is a $p \times 1$ multivariate normal random variable with $p \times 1$ mean $E\{\boldsymbol{Z}\} = \boldsymbol{\mu}$ and $p \times p$ variance

$$\sigma^2\{\boldsymbol{Z}\} = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_p^2 \end{pmatrix}$$

then each element Z_k is a (univariate) normal random variable with mean $E\{Z_k\} = \mu_k$ and variance $\sigma^2\{Z_k\} = \sigma_k^2$.

Accordingly, each element b_k of $\boldsymbol{b}$ is a normal random variable with mean $E\{b_k\} = \beta_k$ and variance $\sigma^2\{b_k\} = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}_{kk}$, where $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}_{kk}$ refers to the k-th diagonal element of $\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ in row k and column k.

We will refer to the normal distribution with mean $E\{b_k\} = \beta_k$ and variance $\sigma^2\{b_k\} = \sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}_{kk}$ as the **sampling distribution** of b_k.

Example 1: Consider data from a company that manufactures refrigeration equipment, called the Toluca company. They produce refrigerator parts in lots of different sizes, and the amount of time it takes to produce a lot of refrigerator parts depends on the number of parts in the lot and several other variable factors. Let X be the number of refrigerator plots in a lot, and let Y refer to the amount of time it takes to produce a size of lot X. Imagine that we magically knew that the true values $\beta_0 = 62$, $\beta_1 = 3.5$, and $\sigma^2 = 2,500$ (this would not happen in real life). We can simulate values of b_1 under this model. The **sampling distribution** describes the distribution of the simulated values.

```
load("~/Dropbox/Teaching/STAT525/Spring2023/bookdata/toluca.RData")
# Extract number of observations
n <- nrow(data)
# Construct the design matrix
X <- cbind(rep(1, n), data$X)

# Set true values
beta <- c(62, 3.5)
sigma.sq <- 2500

# Decide how many simulated datasets we want to create
nsim <- 10000
# Create a vector where we'll record corresponding b1 estimates
b1s <- numeric(nsim)
# To ensure that we obtain the same results every time we run this
# code, we need to set a seed. You can pick any number - I have picked 100.
```

```r
set.seed(100)

for (i in 1:nsim) {
  # Simulate errors
  epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma.sq))
  # Simulate response
  Y <- X%*%beta + epsilon
  # Fit model to simulated data
  XtX.inv <- solve(t(X)%*%X)
  b <- XtX.inv%*%t(X)%*%Y
  # Save coefficient
  b1s[i] <- b[2]
}

# Make a histogram of the simulated values
hist(b1s, xlab = expression(b[1]), freq = FALSE, main = "")
b1vals <- seq(2, 5, length.out = 1000)
# Add the normal density for the sampling distribution to compare
lines(b1vals, dnorm(b1vals, mean = beta[2], sd = sqrt(sigma.sq*XtX.inv[2, 2])),
      col = "blue")
```

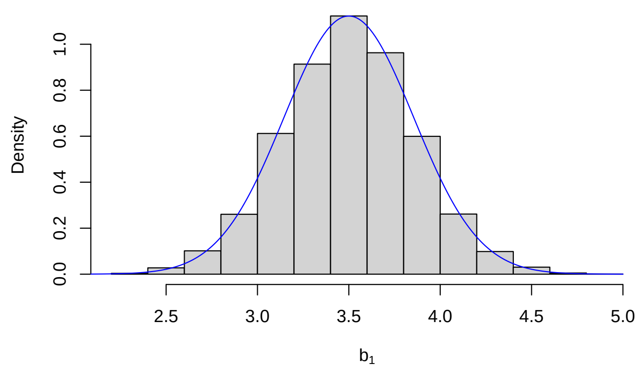

Figure 1: Example 1

In practice, we won't know the true $\boldsymbol{\beta}$. However, we might want to ask questions about what the true value of β_k might be. For example, we might want to ask if the true value of β_k is equal to some specific number, which we'll call c. A natural thing to do would be to compare $b_k - c$ to the sampling distribution when $\beta_k = c$. Conveniently if $\beta_k = c$, the **sampling distribution** of $b_k - c$ doesn't depend on c at all! It is a normal distribution with mean 0 and variance $\sigma^2\left\{b_k\right\}$.

Example 2: Consider the same data. Imagine that we magically knew the true value $\sigma^2 = 2,500$ (this would not happen in real life). What if we want to ask if $\beta_1 = 0$? It would make sense to compare $b_1 - 0 = b_1$ to the sampling distribution of b_1 when $\beta_1 = 0$.

```
# Extract response
Y <- data$Y
# Compute least squares estimate
b <- XtX.inv%*%t(X)%*%Y

# Plot the sampling distribution of b_1 if \beta_1 = 0
b1vals <- seq(-5, 5, length.out = 1000)
# Add the normal density for the sampling distribution to compare
plot(b1vals, dnorm(b1vals, mean = 0, sd = sqrt(sigma.sq*XtX.inv[2, 2])),
      col = "blue", type = "l",
     xlab = expression(b[1]), ylab = "Density")
abline(v = b[2], lty = 2, col = "red")
legend("topleft", lty = 2, col = "red",
       legend = expression(paste("Observed ", b[1], sep = "")),
       bty = "n")
```

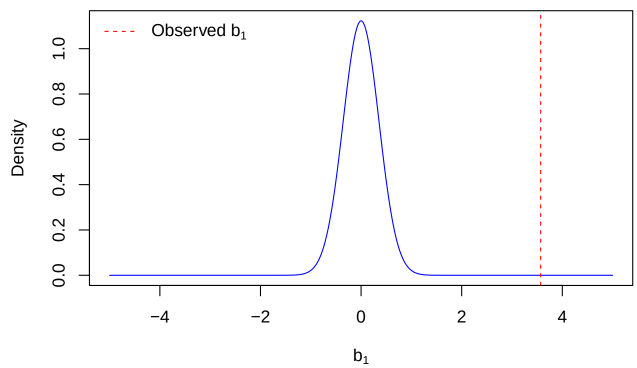

Figure 2: Example 2

We can see that the observed value of b_1 is very extreme when compared to the sampling distribution of b_1 when $\beta_1 = 0$. This suggests that the true value of β_1 is unlikely to be 0. However, we had to pretend that we knew σ^2 and accordingly, $\sigma^2\{b_1\}$ to reach this conclusion. This is not realistic in practice.

This is really useful, but we can't quite make use of it in practice because we won't know $\sigma^2\{b_k\}$. This leads us to define another quantity if we want to ask questions about b_k,

$$\frac{b_k - \beta_k}{s\{b_k\}}.$$

This quantity is an example of a **studentized statistic**. Importantly, its sampling distribution depends on neither β_k or $\sigma^2\{b_k\}$. In fact, the sampling distribution of $\frac{b_k-\beta_k}{s\{b_k\}}$ is a t distribution with $n-p$ degrees of freedom under the normal errors regression model!

To understand why $\frac{b_k-\beta_k}{s\{b_k\}}$ is distributed according to a t distribution under the normal errors regression model, we will first revisit the definition of a t random variable.

Note: Let z and v be independent standard normal (with mean 0 and variance 1) and $\chi^2(\nu)$ random variables. We define a t random variable as $\frac{z}{\sqrt{\frac{v}{\nu}}}$.

We can recognize the standard normal part of $\frac{b_k - \beta_k}{s\{b_k\}}$ by dividing the numerator and denominator by $\sigma\{b_k\}$:

$$\frac{b_k - \beta_k}{s\{b_k\}} = \frac{\frac{b_k - \beta_k}{\sigma\{b_k\}}}{\frac{s\{b_k\}}{\sigma\{b_k\}}}$$

The numerator $\frac{b_k - \beta_k}{\sigma\{b_k\}}$ is a normal random variable with mean 0 and variance 1, i.e. a standard normal random variable. What about the denominator, $\frac{s\{b_k\}}{\sigma\{b_k\}}$? Is it the square root of a $\chi^2(\nu)$ random variable that is independent of the numerator, divided by ν?

> ***Note:*** Let $z_1, \dots, z_\nu$ be independent standard normal (with mean 0 and variance 1) and $\chi^2(\nu)$ random variables. We define a $\chi^2(\nu)$ random variable as $\sum_{i=1}^{\nu} z_i^2$.

Expanding the denominator, we can rewrite it as follows and simplify:

$$\begin{aligned}\frac{s\{b_k\}}{\sigma\{b_k\}} &= \sqrt{\frac{\left(\frac{1}{n-p}\sum_{i=1}^n e_i^2\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}_{kk}}{\sigma^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}_{kk}}} \\ &= \sqrt{\frac{\sum_{i=1}^n \left(\frac{e_i}{\sigma}\right)^2}{n-p}}\end{aligned}$$

Using methods that are beyond the scope of this class, the numerator $\sum_{i=1}^n \left(\frac{e_i}{\sigma}\right)^2$ corresponds to the sum of $n-p$ independent standard normal random variables. Accordingly, $\frac{s\{b_k\}}{\sigma\{b_k\}}$ is equal to the square root of a $\chi^2(n-p)$ random variable divided by its degrees of freedom.

Now that we have decomposed $\frac{b_k - \beta_k}{\sigma\{b_k\}}$ into the ratio of a standard normal random variable $\frac{b_k - \beta_k}{\sigma\{b_k\}}$ and the square root of a $\chi^2(n-p)$ random variable $\sum_{i=1}^n \left(\frac{e_i}{\sigma}\right)^2$ divided by its degrees of freedom, one question remains. Are $\frac{b_k - \beta_k}{\sigma\{b_k\}}$ and $\sum_{i=1}^n \left(\frac{e_i}{\sigma}\right)^2$ independent?

They are, but it's not obvious! The easiest way to show this is to recognize that $\frac{b_k - \beta_k}{\sigma\{b_k\}}$ only depends on the data through $\boldsymbol{b} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and $\sum_{i=1}^n \left(\frac{e_i}{\sigma}\right)^2$ only depends on the data through $\boldsymbol{e}$.

The covariance of $\boldsymbol{b}$ and $\boldsymbol{e}$ can be computed using linear algebra.

$$\begin{aligned}\sigma\{\boldsymbol{e}, \boldsymbol{b}\} &= E\left\{\boldsymbol{e}\boldsymbol{b}'\right\} - E\{\boldsymbol{e}\}E\{\boldsymbol{b}\}' \\ &= E\left\{\left(\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)\boldsymbol{\epsilon}\boldsymbol{Y}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right\} - \boldsymbol{0}\boldsymbol{\beta}' \\ &= \left(\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)E\left\{\boldsymbol{\epsilon}\boldsymbol{Y}'\right\}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\ &= \left(\boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\right)\left(\sigma^2\boldsymbol{I}_n\right)\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\ &= \sigma^2\left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \\ &= \sigma^2\left(\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) = \boldsymbol{0}.\end{aligned}$$

This allows us to conclude that $\boldsymbol{b}$ and $\boldsymbol{e}$ are uncorrelated, and because $\boldsymbol{b}$ and $\boldsymbol{e}$ are also both normal under the normal errors linear regression model, we can conclude that they are independent as well. Accordingly, $\frac{b_k - \beta_k}{\sigma\{b_k\}}$ and $\sum_{i=1}^n \left(\frac{e_i}{\sigma}\right)^2$ are independent.

> ***Note:*** As $\nu \to \infty$, a t distribution with ν degrees of freedom becomes indistinguishable from a normal distribution with mean 0 and variance 1, often called a standard normal distribution. It is common practice to use the normal distribution in the place of the t distribution when $\nu \geq 30$.

Example 3: Consider the same data, and once again imagine that we magically knew that the true values $\beta_0 = 62$, $\beta_1 = 3.5$, and $\sigma^2 = 2,500$ (this would not happen in real life). We can simulate values of $\frac{b_1 - \beta_1}{s\{b_1\}}$ under this model. Again, the **sampling distribution** describes the distribution of the simulated values.

```
# Create a vector where we'll record corresponding
# studentized b1 estimates
sb1s <- numeric(nsim)

for (i in 1:nsim) {
  # Simulate errors
  epsilon <- rnorm(n, mean = 0, sd = sqrt(sigma.sq))
  # Simulate response
  Y <- X%*%beta + epsilon
  # Fit model to simulated data
  XtX.inv <- solve(t(X)%*%X)
  b <- XtX.inv%*%t(X)%*%Y
  s.sq <- sum((Y - X%*%b)^2)/(n - 2)
  # Save studentized statistic
  sb1s[i] <- (b[2] - beta[2])/sqrt(s.sq*XtX.inv[2, 2])
}

# Make a histogram of the simulated values
hist(sb1s, xlab = expression(paste(group("(", b[1] - beta[1], ")"),
                                   "/",
                                   s~group("{", b[1], "}"), sep = "")),
     freq = FALSE, main = "",
     ylim = c(0, 0.5))
sb1vals <- seq(-5, 5, length.out = 1000)
# Add the t density for the sampling distribution to compare
lines(sb1vals, dt(sb1vals, df = n - 2),
      col = "blue")
# Add a standard normal density for comparison
lines(sb1vals, dnorm(sb1vals, mean = 0, sd = 1),
      col = "red")
legend("topleft", col = c("blue", "red"),
       legend = c(expression(t[n-2]), "Standard Normal"),
       lty = 1, bty = "n")
```

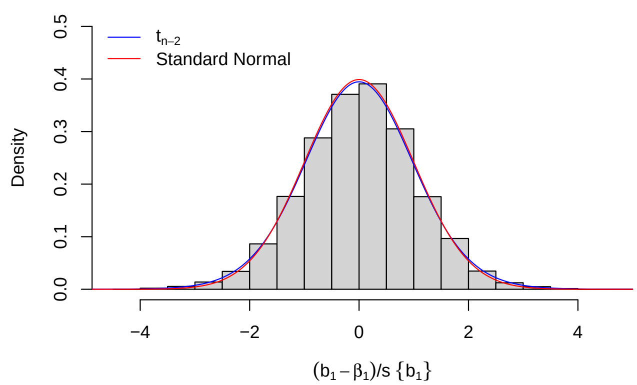

Figure 3: Example 3

Using the studentized statistic, we can to ask if the true value of β_k is equal to some specific number, which we'll call c, without needing to know σ^2 or accordingly $\sigma^2\{b_k\}$. The natural thing to do would be to compare $\frac{b_k - c}{s\{b_k\}}$ to the sampling distribution of $\frac{b_k - c}{s\{b_k\}}$ when $\beta_k = c$. Conveniently if $\beta_k = c$, the **sampling distribution** of $\frac{b_k - c}{s\{b_k\}}$ doesn't depend on c or any other unknown parameters!! It is a t distribution with $n - p$ degrees of freedom.

Example 4: Consider the same data. What if we want to ask if $\beta_1 = 0$? It would make sense to compare $\frac{b_1 - 0}{s\{b_1\}} = \frac{b_1}{s\{b_1\}}$ to the sampling distribution of $\frac{b_1}{s\{b_1\}}$ when $\beta_1 = 0$.

```
# Extract response
Y <- data$Y
# Compute least squares estimate
b <- XtX.inv%*%t(X)%*%Y
b1 <- b[2]
s.sq <- sum((Y - X%*%b)^2)/(n - 2)
s.b1 <- sqrt(s.sq*XtX.inv[2, 2])
# Plot the sampling distribution of b_1/s{b_1} if \beta_1 = 0
sb1vals <- seq(-15, 15, length.out = 1000)
# Add the normal density for the sampling distribution to compare
plot(sb1vals, dt(b1vals, n - 2),
     xlab = expression(paste(b[1],
                              "/",
                              s~group("{", b[1], "}"), sep = "")),
      col = "blue", type = "l", ylab = "Density")
abline(v = b1/s.b1, lty = 2, col = "red")
legend("topleft", lty = 2, col = "red",
       legend = expression(paste("Observed ", b[1],
                                 "/",
                                 s~group("{", b[1], "}"), sep = "")),
       bty = "n")
```

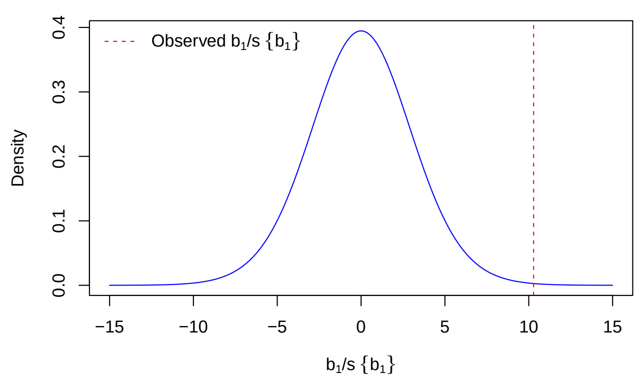

Figure 4: Example 4

We can see that the observed value of $\frac{b_1}{s\{b_1\}}$ is very extreme when compared to the sampling distribution of $\frac{b_1}{s\{b_1\}}$ when $\beta_1 = 0$. This suggests that the true value of β_1 is unlikely to be 0.