Properties of OLS Estimators

Setting: observe data $y = (y_1, \dots, y_n)$
$X \in \mathbb{R}^{n \times p}$

Fit multivariate linear regression model:

$$y = X\beta + \varepsilon \qquad (1)$$

Make the assumption $\varepsilon_1, \dots, \varepsilon_n \overset{iid}{\sim} N(0, \sigma^2)$

Then, solving the least squares problem:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \{(y - X\beta)^T (y - X\beta)\}$$

leads to the normalizing equations:

$$X^T X \beta = X^T y$$

If $(X^T X)^{-1}$ exists, then we can directly (and exactly) obtain

$$\hat{\beta} = (X^T X)^{-1} X^T y \qquad (2)$$

Linearity of expectation (and a little algebra) gives:

i) $E[\hat{\beta}] = \beta$ and ii) $\mathrm{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$

* Notes:

① (i) implies that our OLS estimators $\hat{\beta}$ are unbiased. This promotes the use of LR for interpretation/ explanation of the effects of X on y.

② $Var(\hat{\beta})$ depends on σ^2, which is an unknown parameter. that we must estimate.

③ (i) & (ii) rely only on the weaker possible assumption, mamely that $E[\varepsilon_i] = 0$, $Var(\varepsilon_i) = \sigma^2$, $Corr(\varepsilon_i, \varepsilon_j) = 0$.

* Big Question here: how can we make inference on our model? In other words, can we quantify the uncertainty of $\hat{\beta}$ and formally "test" for significance of an effect?

Answer: Yes! We now need to assume $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Then, properties of normal random variables give us a lot.

* Estimation of σ^2:

Let $\hat{y} = X\hat{\beta}$ and define the ith residual e_i by $e_i = y_i - \hat{y}_i$

Let p = # of predictors in the regression.

Then an unbiased estimator for σ^2 is

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p}$$

If we have $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ then

$$\frac{(n-p)}{\sigma^2} s^2 \sim \chi^2_{n-p} \qquad (3)$$

Question (HW): why is (3) true?

It relies upon the two mathematical facts:

Fact 1: Suppose $E[\varepsilon_i] = 0$, $var(\varepsilon_i) = \sigma^2$, and $corr(\varepsilon_i, \varepsilon_j) = 0$. Then,

$$E\left[\sum_{i=1}^{n} e_i^2\right] = (n-p)\sigma^2$$

Fact 2: Suppose $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Then,

$$\frac{\sum_{i=1}^{n} e_i^2}{\sigma^2} \sim \chi^2_{n-p}$$ and is independent of $\hat{\beta}$.

This justifies the equation for s^2. This is referred to as the mean squared error,

* ## Confidence Intervals and Tests

I. For σ^2

(a) Testing

Consider the hypothesis test

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_1 : \sigma^2 > \sigma_0^2$$

It is natural then to reject H_0 if $s^2 > c$, where c is a critical value that depends on your level of confidence $0 < \alpha < 1$.

~~Let A be a number between 0 & 1 such that~~

$$\cancel{Pr(\chi^2_n \leq A) = A}$$

So, we want $Pr(s^2 > c \mid H_0) = \alpha$;

i.e the probability of false rejection set to α.

Then,

$$Pr(s^2 > c \mid H_0) = Pr(s^2 > c \mid \sigma^2 = \sigma_0^2)$$

$$= Pr\left(\frac{(n-p)}{\sigma_0^2} s^2 > \frac{(n-p)}{\sigma_0^2} c\right)$$

$$= Pr\left(\chi^2_{n-p} > \frac{(n-p)}{\sigma_0^2} c\right) = \alpha$$

Thus, $c = \frac{\sigma_0^2}{n-p} \chi^2_{n-p,1-\alpha}$ where

$\chi^2_{n-p,1-\alpha}$ is the α critical point of the χ^2_{n-p} distribution that satisfies

$$Pr(\chi^2_{n-p} > \chi^2_{n-p,1-\alpha}) = \alpha$$

(or) $$Pr(\chi^2_{n-p} \le \chi^2_{n-p,1-\alpha}) = 1-\alpha.$$

So, decision rule is to reject H_0 if

$$s^2 > \frac{\sigma_0^2}{n-p} \chi^2_{n-p,1-\alpha} \qquad (*)$$

(b) <u>Confidence Interval</u>

* Inversion trick *

Note that by definition,

$$Pr(\chi^2_{n-p,\alpha/2} < \chi^2_{n-p} < \chi^2_{n-p,1-\alpha/2}) = 1-\alpha$$

well, (LHS) $= Pr\left(\chi^2_{n-p,\alpha/2} < \frac{(n-p)s^2}{\sigma^2} < \chi^2_{n-p,1-\alpha/2}\right)$

$$= Pr\left(\frac{(n-p)s^2}{\chi^2_{n-p,1-\alpha/2}} < \sigma^2 < \frac{(n-p)s^2}{\chi^2_{n-p,\alpha/2}}\right)$$

Thus, a $(1-\alpha)100\%$ CI for σ^2 is given by

$$\left[\frac{(n-p)s^2}{\chi^2_{n-p,1-\alpha/2}}, \frac{(n-p)s^2}{\chi^2_{n-p,\alpha/2}}\right]. \qquad (**)$$

Note : With a distribution on s^2, we were readily able to derive confidence intervals and hypothesis tests for σ^2. The same is true for $\hat{\beta}$ and β.

II. For β

We will consider linear combinations of the parameters $\beta_0, \beta_1, \dots, \beta_p$. In other words, we will look at

$$\theta = \sum_{j=0}^{p} c_j \beta_j = c^T \beta$$

where $c_0, c_1, \dots, c_p$ are specified constants.

Examples :

i) $c_j = 1$ and $c_i = 0$, $i \neq j$

Then $c^T \beta = \beta_j$

ii) setting c_j = variable x_j

Then $c^T \beta = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$

By Gauss - Markov, the BLUE for θ is

$\hat{\theta} = c^T \hat{\beta}$ where $\hat{\beta}$ = OLS estimators

In this case, $E[\hat{\theta}] = E[\theta]$ and

$$
\begin{aligned}
Var(\hat{\theta}) &= E[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T] \\
&= E[c^T(\hat{\beta}-\beta)(\hat{\beta}-\beta)c] \\
&= c^T E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)]c \\
&= c^T Var(\hat{\beta}) c \\
&= \sigma^2 c^T (X^T X)^{-1} c
\end{aligned}
$$

since we assume $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, it follows that all linear combinations of normal random variables are also normal and so since $\hat{\theta}$ is a linear function of Y (and thus ε), we have :

$$\hat{\theta} \sim N(\theta, \sigma^2 c^T (X^T X)^{-1} c)$$

$$\Rightarrow \frac{\hat{\theta} - \theta}{\sigma\sqrt{c^T (X^T X)^{-1} c}} \sim N(0, 1) \qquad (*)$$

And like we did for σ^2, $(*)$ can be used to construct hypothesis tests and confidence intervals for θ.

Of course in practice, we often don't know σ^2 and must estimate it using s^2. Doing so leads to the following

$$\frac{\hat{\theta} - \theta}{s\sqrt{c^T(X^TX)^{-1}c}} \sim t_{\cancel{n-p-1}\ n-k} \qquad (**)$$

$k =$ # non-zero elements in c

It follows that we can develop Hypothesis tests as t-tests from $(**)$. Also, a $(1-\alpha)100\%$ confidence interval for θ is given by

(HW) fill in gaps here!

$$\left[\hat{\theta} - t_{n-k,1-\frac{\alpha}{2}}\, s\sqrt{c^T(X^TX)^{-1}c}\ ,\ \hat{\theta} + t_{n-k,1-\frac{\alpha}{2}}\, s\sqrt{c^T(X^TX)^{-1}c}\right]$$

* A note in terminology:
- the standard deviation of $\hat{\beta}_j$ is given by $\sigma\sqrt{(X^TX)^{-1}_{jj}}$.
- When σ is unknown, we estimate the above w/ the <u>standard error</u>:

$$SE(\hat{\beta}_j) = s\sqrt{(X^TX)^{-1}_{jj}}, \quad \text{where } s = \sqrt{MSE}$$