STATISTICAL PROPERTIES OF LEAST SQUARES ESTIMATORS

Situation:

Assumption: $E(Y|x) = \eta_0 + \eta_1 x$ (linear mean function)

Data: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

Least squares estimator: $\hat{E}(Y|x) = \hat{\eta}_0 + \hat{\eta}_1 x$, where

$$\hat{\eta}_1 = \frac{SXY}{SXX} \qquad \hat{\eta}_0 = \bar{y} - \hat{\eta}_1 \bar{x}$$

$$SXX = \sum (x_i - \bar{x})^2 = \sum x_i (x_i - \bar{x})$$
$$SXY = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i - \bar{x}) y_i$$

Comment: If we also assume $e|x$ (equivalently, $Y|x$) is normal with constant variance, then the least squares estimates are the same as the maximum likelihood estimates of η_0 and η_1.

Properties of $\hat{\eta}_0$ and $\hat{\eta}_1$:

1) $$\hat{\eta}_1 = \frac{SXY}{SXX} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{SXX} = \sum_{i=1}^{n}\frac{(x_i - \bar{x})}{SXX}y_i = \sum_{i=1}^{n}c_i y_i$$

where $c_i = \dfrac{(x_i - \bar{x})}{SXX}$

Thus: If the x_i's are fixed (as in the blood lactic acid example), *then $\hat{\eta}_1$ is a linear combination of the y_i's.*

Note: Here we want to think of each y_i as a random variable with distribution $Y|x_i$. Thus, if each $Y|x_i$ is normal, then $\hat{\eta}_1$ is also normal. If the $Y|x_i$'s are not normal but n is large, then $\hat{\eta}_1$ is approximately normal. This will allow us to do inference on $\hat{\eta}_1$. (Details later.)

2) $\sum c_i = \sum \dfrac{(x_i - \bar{x})}{SXX} = \dfrac{1}{SXX}\sum (x_i - \bar{x}) = 0$ (as seen in establishing the alternate expression for SXX)

3) $\sum x_i c_i = \sum x_i \dfrac{(x_i - \bar{x})}{SXX} = \dfrac{1}{SXX}\sum x_i(x_i - \bar{x}) = \dfrac{SXX}{SXX} = 1.$

Remark: Recall the analogous properties for the residuals $\hat{e}_i$.

4) $\hat{\eta}_0 = \bar{y} - \hat{\eta}_1 \bar{x} = \frac{1}{n}\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} c_i y_i \bar{x} = \sum_{i=1}^{n} (\frac{1}{n} - c_i \bar{x}) y_i$, also a linear combination of the y_i's, hence …

5) The sum of the coefficients in (4) is $\sum_{i=1}^{n} (\frac{1}{n} - c_i \bar{x}) = \sum_{i=1}^{n} (\frac{1}{n}) - \bar{x}\sum_{i=1}^{n} c_i = n(\frac{1}{n}) - \bar{x}0 = 1$.

Sampling distributions of $\hat{\eta}_0$ and $\hat{\eta}_1$:

Consider $x_1, \ldots, x_n$ as fixed (i.e., condition on $x_1, \ldots, x_n$).

Model Assumptions ("The" Simple Linear Regression Model Version III):

- $E(Y|x) = \eta_0 + \eta_1 x$ (linear mean function)
- $Var(Y|x) = \sigma^2$ (Equivalently, $Var(e|x) = \sigma^2$) (constant variance)
- (*NEW*) $y_1, \ldots, y_n$ are independent observations. (independence)

The new assumption means we can consider $y_1, \ldots, y_n$ as coming from n independent random variables $Y_1, \ldots, Y_n$, where Y_i has the distribution of $Y|x_i$.

Comment: We do *not* assume that the x_i's are distinct. If, for example, $x_1 = x_2$, then we are assuming that y_1 and y_2 are independent observations from the same conditional distribution $Y|x_1$.

Since $y_1, \ldots, y_n$ are random variables, so is $\hat{\eta}_1$ -- but it depends on the choice of $x_1, \ldots, x_n$, so we can talk about the conditional distribution $\hat{\eta}_1|x_1, \ldots, x_n$.

Expected value of $\hat{\eta}_1$ (as the y's vary):

$$
\begin{aligned}
E(\hat{\eta}_1|x_1, \ldots, x_n) &= E(\sum_{i=1}^{n} c_i y_i |x_1, \ldots, x_n) \\
&= \sum c_i E(y_i|x_1, \ldots, x_n) \\
&= \sum c_i E(y_i|x_i) && \text{(since } y_i \text{ depends only on } x_i\text{)} \\
&= \sum c_i (\eta_0 + \eta_1 x_i) && \text{(model assumption)} \\
&= \eta_0 \sum c_i + \eta_1 \sum c_i x_i \\
&= \eta_0 0 + \eta_1 1 = \eta_1
\end{aligned}
$$

Thus: *$\hat{\eta}_1$ is an unbiased estimator of η_1.*

Variance of $\hat{\eta}_1$ (as the y's vary):

$$
\begin{aligned}
Var(\hat{\eta}_1|x_1, \ldots, x_n) &= Var(\sum_{i=1}^{n} c_i y_i |x_1, \ldots, x_n) \\
&= \sum c_i^2 \, Var(y_i|x_1, \ldots, x_n)
\end{aligned}
$$

$= \sum c_i^2 \text{ Var}(y_i|x_i)$ (since y_i depends only on x_i)

$= \sum c_i^2\sigma^2$

$= \sigma^2\sum c_i^2$

$= \sigma^2\sum\left(\frac{(x_i - \bar{x})}{SXX}\right)^2$ (definition of c_i)

$= \frac{\sigma^2}{(SXX)^2}\sum(x_i - \bar{x})^2$

$= \frac{\sigma^2}{SXX}$

For short: $\text{Var}(\hat{\eta}_1) = \frac{\sigma^2}{SXX}$

$\therefore$ s.d.($\hat{\eta}_1$) $= \frac{\sigma}{\sqrt{SXX}}$

Comments: This is vaguely analogous to the sampling standard deviation for a mean $\bar{y}$:

s.d. (estimator) = $\frac{\textit{population standard deviation}}{\sqrt{\textit{something}}}$

However, here the "something," namely SXX, is more complicated. However, we can still analyze this formula to see how the standard deviation varies with the conditions of sampling. For $\bar{y}$, the denominator is the square root of n, so we see that as n becomes larger, the sampling standard deviation of $\bar{y}$ gets smaller. Here, recalling that $SXX = \sum(x_i - \bar{x})^2$, we reason that:

- If the x_i's are far from $\bar{x}$, SXX is ________, so s.d.($\hat{\eta}_1$) is ________.
- If the x_i's are close to $\bar{x}$, SXX is ________, so s.d.($\hat{\eta}_1$) is ________.

Thus if you are designing an experiment, choosing the x_i's to be _________ from their mean will result in a more precise estimate of $\hat{\eta}_1$. (Assuming the linear model fits!)

Expected value and variance of $\hat{\eta}_0$:

Using the formula $\hat{\eta}_0 = \sum_{i=1}^{n}(\frac{1}{n} - c_i\bar{x})y_i$, calculations (left to the interested student) similar to those for $\hat{\eta}_1$ will show:

- $E(\hat{\eta}_0) = \eta_0$ (So $\hat{\eta}_0$ is an unbiased estimator of η_0.)
- $\text{Var}(\hat{\eta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SXX}\right)$, so

$$\text{s.d}(\hat{\eta}_0) = \sigma\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SXX}}$$

Analyzing the variance formula:

- The variance of $\hat{\eta}_0$ is _______________ than the variance of $\hat{\eta}_1$.
  → Does this agree with intuition?
- A larger sample size tends to give a ____________ variance for $\hat{\eta}_0$.
- A larger $\bar{x}$ gives a ____________ variance for $\hat{\eta}_0$.
  → Does this agree with intuition?
- The spread of the x_i's affects the variance of $\hat{\eta}_0$ in the same way it affects the variance of $\hat{\eta}_1$.

Covariance of $\hat{\eta}_0$ and $\hat{\eta}_1$: Similar calculations (left to the interested student) will show

$$\text{Cov}(\hat{\eta}_0, \hat{\eta}_1) = -\sigma^2 \frac{\bar{x}}{SXX}$$

Thus:

- $\hat{\eta}_0$ and $\hat{\eta}_1$ are not independent
  → Does this agree with intuition?
- The sign of $\text{Cov}(\hat{\eta}_0, \hat{\eta}_1)$ is opposite that of $\bar{x}$.
  → Does this agree with intuition?

Estimating σ^2: To use the variance formulas above for inference, we need to estimate σ^2 (= Var(Y|x_i), the same for all i).

First, some plausible reasoning: *If* we had lots of observations $y_{i_1}, y_{i_2}, \ldots, y_{i_m}$ from Y|x_i, then we could use the univariate standard deviation

$$\frac{1}{m-1}\sum_{j=1}^{m}(y_{i_j} - \bar{y}_i)^2$$

of these m observations to estimate σ^2. (Here $\bar{y}_i$ is the mean of $y_{i_1}, y_{i_2}, \ldots, y_{i_m}$, which would be our best estimate of E(Y| x_i) just using $y_{i_1}, y_{i_2}, \ldots, y_{i_m}$)

We don't typically have lots of y's from one x_i, so we might try (reasoning that $\hat{E}(Y \mid x_i)$) is our best estimate of E(Y|x_i))

$$\frac{1}{n-1}\sum_{i=1}^{n}[y_i - \hat{E}(Y \mid x_i)]^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\hat{e}_i^{\,2}$$

$$= \frac{1}{n-1} RSS.$$

However (just as in the univariate case, we need a denominator n-1 to get an unbiased estimator), a lengthy calculation (omitted) will show that

$$E(RSS| x_1, \dots , x_n) = (n-2)\,\sigma^2$$

(where the expected value is over all samples of the y_i's with the x_i's fixed)

Thus we use the estimate

$$\hat{\sigma}^2 = \frac{1}{n-2} RSS$$

to get an unbiased estimator for σ^2:

$$E(\hat{\sigma}^2 | x_1, \dots , x_n) = \sigma^2.$$

[If you like to think heuristically in terms of losing one degree of freedom for each calculation from data involved in the estimator, this makes sense: Both $\hat{\eta}_0$ and $\hat{\eta}_1$ need to be calculated from the data to get RSS.]

Standard Errors for $\hat{\eta}_0$ and $\hat{\eta}_1$: Using

$$\hat{\sigma} = \sqrt{\frac{RSS}{n-2}}$$

as an estimate of σ in the formulas for s.d ($\hat{\eta}_0$) and s.d($\hat{\eta}_1$), we obtain the *standard errors*

$$\text{s.e. } (\hat{\eta}_1) = \frac{\hat{\sigma}}{\sqrt{SXX}}$$

and

$$\text{s.e.}(\hat{\eta}_0) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{SXX}}$$

as estimates of s.d ($\hat{\eta}_1$) and s.d ($\hat{\eta}_0$), respectively.