LECTURE 3: SIMPLE REGRESSION II

\( \hat{\alpha} \) and \( \hat{\beta} \) are the LS estimators

\[ \hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \] are the estimated values

THE CORRELATION COEFFICIENT:

\[ R^2 = (\text{squared}) \text{ correlation between } y \text{ and } \hat{y} \]

\[ r = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}. \]

Note: \( \hat{y} \) is a linear function of \( x \).

So \( \text{corr} (y, \hat{y}) = |\text{corr} (y, x)|. \)
Correlation

**Proposition:** $-1 < r < 1$

$$r^2 = \frac{\left( \sum (x_i - \bar{x}) (y_i - \bar{y}) \right)^2}{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}.$$ 

Use Cauchy-Schwartz

$$\left( \sum x_i y_i \right)^2 \leq \sum x_i^2 \sum y_i^2$$

$$\Rightarrow r^2 \leq 1 \quad \Rightarrow \quad -1 \leq r \leq 1$$

**Proposition:** $\beta$ and $r$ have the same sign.

**Proof:**

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})y_i}{\sum (x_i - \bar{x})^2} = r \frac{\sqrt{\sum (y_i - \bar{y})^2}}{\sqrt{\sum (x_i - \bar{x})^2}}$$
Correlation (more)

\[ \sum e_i^2 = \sum (y_i - \bar{y})^2 - \hat{\beta}^2 \sum (x_i - \bar{x})^2 \]

SSR = TSS - SS explained by x

**Proposition:**

\[ r^2 = 1 - \frac{\text{SSR}}{\text{TSS}} = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} \]

**Proof**

\[ \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} = 1 - \hat{\beta}^2 \frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2} = 1 - r^2 \]

\[ \Rightarrow r^2 = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2} \]
MAXIMUM LIKELIHOOD ESTIMATORS:

Assumption: Normality

\[ p(y|x) = \mathcal{N}(\alpha + \beta x, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left( -\frac{1}{2} \left( \frac{y - \alpha - \beta x}{\sigma} \right)^2 \right) \]

Likelihood Function:

\[ L(\alpha, \beta, \sigma^2) = \prod_{i=1}^{n} p(y_i | x_i) = (2\pi\sigma^2)^{-n/2} \exp\left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \right) \]

The maximum likelihood (ML) estimators maximize L. The log likelihood function is

\[ \ell(\alpha, \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \]
Proposition: The LS estimators are also the ML estimators. What is the maximum in $\sigma^2$?

$$\sigma_{ML}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{n}$$

Why?

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

$$\Rightarrow \quad \sigma_{ML}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

is this a maximum in $\sigma$?

$$\frac{\partial^2 \ell}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum (y_i - \alpha - \beta x_i)^2 = \frac{-n}{2\sigma^4} < 0$$
Distribution of Estimators

These are linear combinations of normal random variables, hence they are normal. The means and variances have already been obtained:

_Distribution of $s$ and $\sigma$

**Fact:** $\Sigma e^2$ can be written as a sum of squares of $(n - 2)$ independent normal random variables with means zero and variances sigma-squared.

**Proposition:** $s^2$ is unbiased and $\mathcal{V}s^2 = 2\sigma^4/(n-2)$.  
**Proof:** Note that $(n-2)s^2/\sigma^2$ is distributed as $\chi^2(n-2)$
More Distributions

⇒ \( E(s^2 / \sigma^2)(n-2) = (n-2) \Rightarrow E(s^2) = \sigma^2 \)

⇒ \( V(s^2 / \sigma^2)(n-2) = 2(n-2) \) so \( V(s^2) = 2\sigma^4 / (n-2) \)

**Proposition:** \( s^2 \) has higher variance than \( \sigma^2_{ml} \)

**Proof:** Note that \( \frac{n\sigma^2_{ML}}{\sigma^2} \) is distributed as \( \chi^2(n-2) \).

⇒ \( E\sigma^2_{ML} = \frac{\sigma^2(n-2)}{n} \)

⇒ \( V\left(\frac{n\sigma^2_{ML}}{\sigma^2}\right) = 2(n-2) \Rightarrow V(\sigma^2_{ML}) = \frac{2\sigma^4(n-2)}{n^2} \)

⇒ \( \frac{V(s^2)}{V(\sigma^2_{ML})} = \frac{1 / (n-2)}{(n-2)n^2} = \frac{n^2}{(n-2)^2} > 1 \)
INFERENCE:

\[ \hat{\beta} \sim \mathcal{N}(\beta, \sigma^2) \] where \( \sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \) \Rightarrow \frac{\hat{\beta} - \beta}{\sigma} \sim \mathcal{N}(0, 1) \]

**Definition:** A 95% confidence interval for \( \hat{\beta} \) is given by \( (\hat{\beta} - z_{0.025} \cdot \sigma) \)

where \( z \) is standard normal.

**Problem:** The variance is unknown.

**Fact:** If \( z \sim \mathcal{N}(0,1) \) and \( v \sim \chi^2(k) \) and they are independent, then \( t = \frac{z}{\sqrt{v / k}} \)

is distributed as \( t(k) \).

**Proposition:**

\[ \frac{\hat{\beta} - \beta}{s / \sqrt{\sum (x_i - \bar{x})^2}} \sim t(n - 2) \]
Proof

\[
\frac{(\hat{\beta} - \beta) \sqrt{\sum (x_i - \bar{x})^2}}{\sigma} \sim n(0,1)
\]

\[
\frac{s^2}{\sigma^2} (n - 2) \sim \chi^2 (n - 2)
\]

\[
\frac{(\hat{\beta} - \beta) \sqrt{\sum (x_i - \bar{x})^2}}{s / \sigma} = \frac{(\hat{\beta} - \beta)}{s / \sqrt{\sum (x_i - \bar{x})^2}} \sim t(n - 2)
\]

Independence?

\[
E(\hat{\beta} - \beta)e_j = E[(\hat{\beta} - \beta)(e_j - \bar{e})]
\]

\[
= E[(\hat{\beta} - \beta)((\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_j + \epsilon_j - (\alpha - \hat{\alpha}) - (\beta - \hat{\beta})\bar{x} - \bar{\epsilon})]
\]

\[
= E[(\hat{\beta} - \beta)(-((\hat{\beta} - \beta)(x_j - \bar{x}) + (\epsilon_j - \bar{\epsilon}))]
\]

\[
= -(x_j - \bar{x})E[(\hat{\beta} - \beta)^2] + E((\hat{\beta} - \beta)(\epsilon_j - \bar{\epsilon})]
\]

\[
= -\frac{\sigma^2 (x_j - \bar{x})}{\sum (x_i - \bar{x})^2} + \frac{E(\epsilon_j - \bar{\epsilon}) \sum (x_i - \bar{x}) \epsilon_i}{\sum (x_i - \bar{x})^2}
\]
Continuation of independence argument

\[ E \frac{(\varepsilon_j - \bar{\varepsilon}) \sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2} = \frac{\sigma^2 (x_j - \bar{x})}{\sum (x_i - \bar{x})^2} - E \frac{\varepsilon \sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2}. \]

\[ E \frac{\varepsilon \sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2} = 0. \]

Thus

\[ E (\hat{\beta} - \beta) e_j = 0. \]
Violations of Assumptions

I. \( E(y_i) = \alpha + x_i \beta \)

II. \( V(y_i | x_i) = V(\varepsilon_i) = \sigma \)

The alternative is \((heteroskedasticity)\).

Is the LS estimator unbiased? Is it BLUE?

If the \( \sigma_i \) are known we can run the 'transformed' regression, and will get best linear unbiased estimates and correct standard errors.

\( w_i = 1/\sigma_i \), let \( w_i y_i = \alpha w_i + \beta x_i w_i + \varepsilon_i w_i \).

\( E w_i y_i = \alpha w_i + \beta x_i w_i \) and \( V(w_i y_i) = V(\varepsilon_i w_i) = 1 \)

The Gauss-Markov Theorem tells that LS is BLUE in the transformed model.
The LS estimator in the transformed model is

$$
\hat{\beta}_w = \frac{\sum (x_i w_i - \bar{x} w) w_i y_i}{\sum (x_i w_i - \bar{x} w)^2} \neq \hat{\beta}.
$$

with

$$
V(\hat{\beta}) = \frac{\sum (x_i - \bar{x})^2 \sigma_i^2}{\left(\sum (x_i - \bar{x})^2\right)^2}
$$

Note: The variance of $\beta_w$ is less than the variance of $\beta$.

“Heteroskedasticity Consistent” standard errors:

$$
V(\hat{\beta}) = E\left[\frac{\sum (x_i - \bar{x}) \varepsilon_i}{\sum (x_i - \bar{x})^2}\right]^2 = E\left[\frac{\sum (x_i - \bar{x})^2 \varepsilon_i^2}{\left(\sum (x_i - \bar{x})^2\right)^2}\right]
$$

insert $e$ for $\varepsilon$ and remove the expectation.
More on Heteroskedasticity

Essentially this works because $\sum \hat{e}_i^2 / n$ is a reasonable estimator for $\sum \sigma_i^2 / n$, although of course $\hat{e}_i^2$ is not a good estimator for $\sigma_i^2$.

*Testing for heteroskedasticity:*
Split the sample; regress $e^2$ on stuff

III. $E \varepsilon_i \varepsilon_j = 0$

The alternative is $E \varepsilon_i \varepsilon_j \neq 0$

Is the LS estimator unbiased? Is it BLUE?

*Testing for correlated errors:*
We need a hypothesis about the correlation.
More (last) on violations of assumptions

IV. Normality

\[ E(y_i|x_i) = \alpha + \beta x_i ; \quad V(y_i|x_i) = \sigma^2 \text{ but } \varepsilon_i \sim f(\varepsilon) \neq N(0,\sigma^2) \]

The usual suspect is a heavy-tailed distribution. Is the LS estimator unbiased? Is it BLUE?

Example:

\[ f(\varepsilon) = \frac{1}{2\phi} \exp(-|\varepsilon/\phi|) \]

The variance of the ML estimator is half that of the LS estimator asymptotically. The minimum absolute deviation (MAD) estimator works. It is a robust estimator.