Consider once again, population level
\[ Y = \beta_0 + \beta_1 X + \varepsilon, \quad \varepsilon \text{ independent } X, \quad \text{E}(\varepsilon) = 0, \quad \text{var}(\varepsilon) = \sigma^2 \]
so on the sample
\[ Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad \text{with } \varepsilon_i \text{ iid } \sim \varepsilon \]

**How would we produce a point-estimate of E(Y|X)** i.e. the mean of Y at a given X (within a given X-subpopulation)?

We could use the estimated regression line
\[ \hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i \quad \text{(fitted value)} \]
but also on non-sample values... in general
\[ \hat{Y}(x) = \hat{b}_0 + \hat{b}_1 x \quad \text{(estimate of E(Y|x))} \]
If the model we postulated is correct:

- \( E(\hat{Y}(x)) = E(Y/x) = \beta_0 + \beta_1 x \)

(our estimate does not contain a "systematic error" with respect to what we are trying to estimate…)

- \( \sigma^2_{\hat{Y}(x)} = \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_i (X_i - \bar{x})^2} \right) \sigma^2 \)

(the calculation takes into account the variances of \( b_0 \) and \( b_1 \), that go into \( \hat{Y}(x) = b_0 + b_1 x \)…)

Compare with

- \( \sigma^2_{b_0} = \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_i (X_i - \bar{x})^2} \right) \sigma^2 \)

clearly so, as \( b_0 = \hat{Y}(0) \)

2
As usual, we can unbiasedly estimate the variance of \( \hat{\theta}(x) \) as:

\[
S^2_{\hat{\theta}(x)} = \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) s^2
\]

(recall \( s^2 = \text{MSE} \))

and take sq. root \( S_{\hat{\theta}(x)} \) to estimate st. dev.

An interesting point:

\( \sigma_{\hat{\theta}(x)}, \sigma_{\theta}(x) \) and their estimates

\( S_{\hat{\theta}(x)}, S_{\theta}(x) \) depend on how far \( x \) is from the sample mean \( \bar{x} \). The larger is \( (x - \bar{x})^2 \), the larger is the variability of our estimate... and its estimate
If we also assume normality of the errors:

\[ E \sim N(0, \sigma^2) \] - population

\[ E_i \ iid \sim N(0, \sigma^2) \] - sample

we have

\[ \hat{\beta}(x) \sim N(\mu(x), \frac{\sigma^2}{\beta + \beta_1 x}) \]

and as usual

\[ \frac{\hat{\beta}(x) - \mu(x)}{s_\hat{\beta}(x)} \sim t_{n-2} \]

which we can use to construct confidence intervals
If our model is correct and normality of the errors holds; the interval

\[ \hat{y}(x) \pm t_{\alpha} \cdot s_{\hat{y}(x)} \]

will contain the unknown \( E(Y|x) = \beta_0 + \beta_1 x \) with prob. \( 1-\alpha \)

Since \( s_{\hat{y}(x)} \) is an increasing function of \( (x - \bar{x})^2 \); given the sample X's that go into its calculation, and the level \( 1-\alpha \) that gives us the \( t_{\alpha} \) in the formula....

**THE INTERVAL WILL BE WIDER THE FURTHER APART \( x \) IS FROM \( \bar{x} \).**

As we move away from the "center" of the sample X's \( (\bar{x}) \), we need to take wider intervals to achieve the same confidence level
We can construct a "band" about the estimated regression line.

A confidence band for $E(Y|X)$, level $1 - \alpha$

The intervals at various $x$'s are centered at $b_0 + b_1x$ ... and get wider moving away from $\bar{x}$.

Compare this with our remark that using the estimated regression line outside the sample range of $X$ might be unsafe!
Up to now we have considered estimating
- by a point $\hat{\theta}(x)$
- by an interval $\hat{\theta}(x) \pm t_\alpha S(x)$

under normality...

How would we predict the value of a new $Y$-draw from a given $X$-subpopulation?

Completely different enterprise... we are not trying to make point-estimation or inference on an unknown but fixed quantity (e.g. $\beta_1$, or $E(Y|x)$...).

We want to say something -predict- about the value of a further random observation of $Y$ in a certain $X$-subpop.

Symbol: $Y_{\text{NEW}}(x)$
Notice that if our model is correct:

- \( E(Y_{\text{NEW}}(x)) = \beta_0 + \beta_1 x = E(Y|x) \)
- \( \text{var}(Y_{\text{NEW}}(x)) = \text{var}(Y|x) = \text{var}(E) = \sigma^2 \)

and if the errors are normal:

- \( Y_{\text{NEW}}(x) \sim N\left(E(Y|x), \sigma^2\right) \)

\[ \beta_0 + \beta_1 x \]

How could we "point-predict" \( Y_{\text{NEW}}(x) \)?

Cannot do any better than taking

\[ \hat{Y}_{\text{NEW}}(x) = \hat{Y}(x) = b_0 + b_1 x \]

\[ \text{prediction of the new draw} \quad \uparrow \quad \text{point estimate of its mean} \quad \uparrow \quad \text{along the estimated regression line} \]

8
When we evaluate the variability of this "point-prediction", there are two sources to take into account:

(A) the variability of \( \hat{P}(x) \) itself

i.e. \[ \sigma_{\hat{P}(x)}^2 = \left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) \sigma^2 \]

(B) the variability intrinsic in the new draw i.e. \( \sigma^2 \)

assuming the new draw to be independent from the previous \( y_1, \ldots, y_n \) that went into the calculation of \( b_0, b \), and hence \( \hat{P}(x) \)

... we can add (A) and (B) to get

\[ \sigma^2 + \sigma_{\hat{P}(x)}^2 = \left( 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) \sigma^2 \]

\[ \uparrow \quad \text{variability in the new draw} \]

\[ \uparrow \quad \text{variability in estimating its mean} \]
As usual, this can be unbiasedly estimated as

\[ s_{\text{pred}(x)}^2 = \left( 1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) s^2 \]

An important point:

The prediction variability, and its estimate \( s_{\text{pred}(x)}^2 \), likewise the mean-estimate variability \( s_{\hat{y}(x)}^2 \), and its estimate \( s_{\hat{y}(x)}^2 \), depend on \((x - \bar{x})^2\) and GROW AS \( x \) MOVES AWAY FROM \( \bar{x} \).

Moreover: the former exceed the latter by \( \sigma^2 \) \( (s^2) \) for estimates...

...because of intrinsic variability in the new draw!
It can be shown that under normality of the errors

\[ \frac{\hat{y}(x) - \hat{y}_{\text{new}}(x)}{s_{\text{pred}}(x)} \sim t_{n-2} \]

which we can use to construct a prediction interval: If our model is correct and normality of the errors holds, the interval

\[ \hat{y}(x) \pm t_{\alpha} s_{\text{pred}}(x) \]

will contain a new \( y \)-draw from the given \( x \)-subpopulation with prob. 1-\( \alpha \). The interval will be wider the further apart \( x \) is from \( \bar{x} \), and at any \( x \) it will be wider (by \( s \), but under the sq. root...) than the corresponding confidence interval for the mean.
We can construct a prediction band for $y_{new}(x)$ about the estimated regression line, level $1-\alpha$.

The intervals at various $x$'s are centered at $\hat{b} + b_1 x$ ... and get wider moving away from $x$. At each $x$, they are wider than the confidence intervals for the mean.

(Minitab to produce bands?)
Another band (pg 67 of the book) Working-Hotelling confidence band FOR THE WHOLE REGRESSION LINE

On the sample data, construct a region of \( \mathbb{R}^2 \) that, if our model is correct and the errors are normal, covers (entirely) the fixed and unknown regression line \( \beta_0 + \beta_1 x \) with prob. \( 1 - \alpha \).

Such region is given by

\[ \hat{y}(x) \pm \hat{w}_\alpha \hat{S}(x), \ x \in \mathbb{R} \]

where \( \hat{w}_\alpha = \sqrt{2 \hat{F}_\alpha} \)

- Usual shape (wider as moving away from \( \bar{x} \))
- Slightly wider than level 1-\( \alpha \) confidence band for the mean

\[ \hat{F}_{1, n-2} \]

\[ \hat{F}_\alpha \]

\[ \hat{X} \]

\[ \hat{X} \]

\[ \beta_0 + \beta_1 x \]

\[ \hat{S}(x) \]

\( x \in \mathbb{R} \)
Important: different meaning of level

- level 1-\(\alpha\) confidence band for the mean
  \[ \hat{Y}(x) \pm t_{d, \alpha} s_{\hat{Y}(x)}, \ x \in \mathbb{R}^d \]
  at any fixed \(x\), the interval \( \hat{Y}(x) \pm t_{d, \alpha} s_{\hat{Y}(x)} \)
  has level 1-\(\alpha\), i.e. contains the unknown \(E[Y|x]\)
  with pr. 1-\(\alpha\)

- level 1-\(\alpha\) confidence band for whole regression
  line \[ \hat{Y}(x) \pm W_d s_{\hat{Y}(x)}, \ x \in \mathbb{R}^d \]
  overall, the region has level 1-\(\alpha\), i.e.
  contains the unknown \(\beta_0 + \beta_1 x\) with prob. 1-\(\alpha\)

The: data examples
Thus: start diagnostics?

14