Estimating $\sigma^2$

- We can do simple prediction of $Y$ and estimation of the mean of $Y$ at any value of $X$.

- To perform inferences about our regression line, we must estimate $\sigma^2$, the variance of the error term.

- For a random variable $Y$, the estimated variance is:

- In regression, the estimated variance of $Y$ (and also of $\varepsilon$) is:

\[
\sum (Y - \hat{Y})^2
\]

is called the error (residual) sum of squares (SSE).
- It has $n - 2$ degrees of freedom.

- The ratio $\text{MSE} = \text{SSE} / \text{df}$ is called the mean squared error.
• MSE is an unbiased estimate of the error variance $\sigma^2$.

• Also, $\sqrt{MSE}$ serves as an estimate of the error standard deviation $\sigma$.

**Partitioning Sums of Squares**

• If we did not use $X$ in our model, our estimate for the mean of $Y$ would be:

Picture:

For each data point:
• $Y - \Bar{Y} =$ difference between observed $Y$ and sample mean $Y$-value
• $Y - \hat{Y} =$ difference between observed $Y$ and predicted $Y$-value
• $\hat{Y} - \Bar{Y} =$ difference between predicted $Y$ and sample mean $Y$-value

• It can be shown:
● TSS = overall variation in the $Y$-values
● SSR = variation in $Y$ accounted for by regression line
● SSE = extra variation beyond what the regression relationship accounts for

Computational Formulas:

$$TSS = S_{YY} = \sum Y^2 - \frac{(\sum Y)^2}{n}$$

$$SSR = \frac{(S_{XY})^2}{S_{XX}} = \hat{\beta}_1 S_{XY}$$

$$SSE = S_{YY} - \frac{(S_{XY})^2}{S_{XX}} = S_{YY} - \hat{\beta}_1 S_{XY}$$

Case (1): If SSR is a large part of TSS, the regression line accounts for a lot of the variation in $Y$.

Case (2): If SSE is a large part of TSS, the regression line is leaving a great deal of variation unaccounted for.

**ANOVA test for $\beta_1$**

● If the SLR model is useless in explaining the variation in $Y$, then $\overline{Y}$ is just as good at estimating the mean of $Y$ as $\hat{Y}$ is.

=> true $\beta_1$ is zero and $X$ doesn’t belong in model

● Corresponds to case (2) above.
• But if (1) is true, and the SLR model explains a lot of the variation in Y, we would conclude $\beta_1 \neq 0$.

• How to compare SSR to SSE to determine if (1) or (2) is true?

• Divide by their degrees of freedom. For the SLR model:

• We test:

• If MSR much bigger than MSE, conclude $H_a$. Otherwise we cannot conclude $H_a$.

The ratio $F^* = \frac{MSR}{MSE}$ has an F distribution with $df = (1, n - 2)$ when $H_0$ is true.

Thus we reject $H_0$ when

where $\alpha$ is the significance level of our hypothesis test.
t-test of $H_0: \beta_1 = 0$

- Note: $\beta_1$ is a parameter (a fixed but unknown value)
- The estimate $\hat{\beta}_1$ is a random variable (a statistic calculated from sample data).
- Therefore $\hat{\beta}_1$ has a sampling distribution:

- $\hat{\beta}_1$ is an unbiased estimator of $\beta_1$.
- $\hat{\beta}_1$ estimates $\beta_1$ with greater precision when:
  - the true variance of $Y$ is small.
  - the sample size is large.
  - the $X$-values in the sample are spread out.

Standardizing, we see that:

Problem: $\sigma^2$ is typically unknown. We estimate it with MSE. Then:
To test $H_0: \beta_1 = 0$, we use the test statistic:

Advantages of t-test over F-test:
(1) Can test whether the true slope equals any specified value (not just 0).
Example: To test $H_0: \beta_1 = 10$, we use:

(2) Can also use t-test for a one-tailed test, where:
$H_a: \beta_1 < 0$ or $H_a: \beta_1 > 0$.

$H_a$ Reject $H_0$ if:
The value \( \sqrt{\frac{MSE}{S_{xx}}} \) measures the precision of \( \hat{\beta}_1 \) as an estimate.

**Confidence Interval for \( \beta_1 \):**

- The sampling distribution of \( \hat{\beta}_1 \) provides a confidence interval for the true slope \( \beta_1 \):

**Example (House price data):**

Recall: \( S_{YY} = 93232.142, S_{XY} = 1275.494, S_{XX} = 22.743 \)

Our estimate of \( \sigma^2 \) is \( MSE = SSE / (n - 2) \)

\( SSE = \)

\( MSE = \)

and recall

- To test \( H_0: \beta_1 = 0 \) vs. \( H_a: \beta_1 \neq 0 \) (at \( \alpha = 0.05 \))
Table A2: \( t_{0.025(56)} \approx 2.004. \)

- With 95\% confidence, the true slope falls in the interval

**Interpretation:**

**Inference about the Response Variable**

- We may wish to:

(1) Estimate the mean value of \( Y \) for a particular value of \( X \). Example:

(2) Predict the value of \( Y \) for a particular value of \( X \). Example:

The point estimates for (1) and (2) are the same: The value of the estimated regression function at \( X = 1.75 \). Example:
• Variability associated with estimates for (1) and (2) is quite different.

\[ \text{Var}[\hat{E}(Y \mid X)] = \]

\[ \text{Var}[\hat{Y}_{\text{pred}}] = \]

• Since \( \sigma^2 \) is unknown, we estimate \( \sigma^2 \) with MSE:

CI for \( E(Y \mid X) \) at \( x^* \):

Prediction Interval for \( Y \) value of a new observation with \( X = x^* \):

Example: 95% CI for mean selling price for houses of 1750 square feet:
Example: 95% PI for selling price of a new house of 1750 square feet:

**Correlation**

- $\hat{\beta}_1$ tells us something about whether there is a linear relationship between $Y$ and $X$.
- Its value depends on the **units of measurement** for the variables.

- The correlation coefficient $r$ and the coefficient of determination $r^2$ are **unit-free** numerical measures of the **linear association** between two variables.

- $r =$

(measures strength and direction of linear relationship)

- $r$ always between -1 and 1:
- $r > 0 \rightarrow$
• \( r < 0 \rightarrow \)

• \( r = 0 \rightarrow \)

• \( r \) near -1 or 1 →

• \( r \) near 0 →

• Correlation coefficient (1) makes no distinction between independent and dependent variables, and (2) requires variables to be numerical.

**Examples:**
House data:

Note that \( r = \hat{\beta}_1 \left( \frac{s_x}{s_y} \right) \) so \( r \) always has the same sign as the estimated slope.

• The population correlation coefficient is denoted \( \rho \).
• Test of \( H_0: \rho = 0 \) is equivalent to test of \( H_0: \beta_1 = 0 \) in SLR (p-value will be the same)

• Software will give us \( r \) and the p-value for testing \( H_0: \rho = 0 \) vs. \( H_a: \rho \neq 0 \).

• To test whether \( \rho \) is some nonzero value, need to use transformation – see p. 318.
• The square of $r$, denoted $r^2$, also measures strength of linear relationship.

• Definition: $r^2 = \frac{SSR}{TSS}$.

  **Interpretation of $r^2$:** It is the proportion of overall sample variability in $Y$ that is explained by its linear relationship with $X$.

  Note: In SLR, $F = \frac{(n-2)r^2}{1-r^2}$.

• Hence: large $r^2 \rightarrow$ large F statistic $\rightarrow$ significant linear relationship between $Y$ and $X$.

**Example** (House price data):

Interpretation:

---

**Regression Diagnostics**

• We assumed various things about the random error term. How do we check whether these assumptions are satisfied?

• The (unobservable) error term for each point is:
As “estimated” errors we use the **residuals** for each data point:

- **Residual plots** allow us to check for four types of violations of our assumptions:
  
  1. The model is misspecified (linear trend between $Y$ and $X$ incorrect)
  2. Non-constant error variance (spread of errors changes for different values of $X$)
  3. Outliers exist (data values which do not fit overall trend)
  4. Non-normal errors (error term is not (approx.) normally distributed)

- A residual plot plots the residuals $Y - \hat{Y}$ against the predicted values $\hat{Y}$.
- If this residual plot shows **random scatter**, this is good.
- If there is some notable pattern, there is a possible violation of our model assumptions.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Violation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• We can verify whether the errors are approximately normal with a Q-Q plot of the residuals.

• If Q-Q plot is roughly a straight line → the errors may be assumed to be normal.

Example (House data):

Remedies for Violations – Transforming Variables

• When the residual plot shows megaphone shape (non-constant error variance) opening to the right, we can use a variance-stabilizing transformation of \( Y \).

• Picture:

• Let \( Y^* = \log(Y) \) or \( Y^* = \sqrt{Y} \) and use \( Y^* \) as the dependent variable.
• These transformations tend to reduce the spread at high values of $\hat{Y}$.

• Transformations of $Y$ may also help when the error distribution appears non-normal.

• Transformations of $X$ and/or of $Y$ can help if the residual plot shows evidence of a nonlinear trend.

• Depending on the situation, one or more of these transformations may be useful:

  • Drawback: Interpretations, predictions, etc., are now in terms of the transformed variables. We must reverse the transformations to get a meaningful prediction.

  Example (Surgical data):