Stat 206: Linear Models

Lecture 3

October 5, 2015

ReCap: Properties of LS Estimators

- **LS estimators are linear functions of the responses Y_is.**

$$\hat{\beta}_1 = \sum_{i=1}^{n} \frac{X_i - \overline{X}}{\sum_{j=1}^{n}(X_j - \overline{X})^2} Y_i = \sum_{i=1}^{n} k_i Y_i$$

$$\hat{\beta}_0 = \sum_{i=1}^{n} (\frac{1}{n} - \bar{X} k_i) Y_i.$$

- The fitted values $\widehat{Y}_i$ and the residuals e_i are also linear functions of the responses Y_is.

Can you write down their respective coefficients?

- **LS estimators are unbiased**: For **all** values of β_0, β_1,

$$E(\hat{\beta}_0) = \beta_0, \quad E(\hat{\beta}_1) = \beta_1.$$

Notes: Use the fact $E(Y_i) = \beta_0 + \beta_1 X_i, \quad i = 1, \cdots n.$

- Variances of $\hat{\beta}_0, \hat{\beta}_1$:

$$\begin{aligned} \sigma^2\{\hat{\beta}_0\} &= \sigma^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2} \right] \\ \sigma^2\{\hat{\beta}_1\} &= \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}. \end{aligned}$$

Notes: Use the fact that Y_i*s are uncorrelated.*

Standard errors (SE) of the LS estimators.

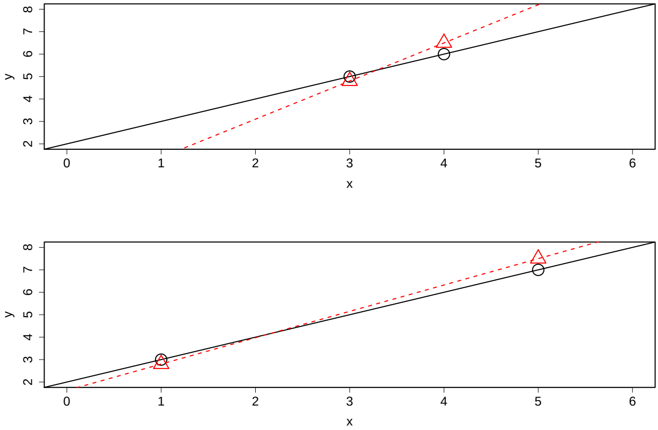
- Replace σ^2 by MSE:

$$
\begin{aligned}
s^2\{\hat{\beta}_0\} &= MSE\left[\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right], \\
s^2\{\hat{\beta}_1\} &= \frac{MSE}{\sum_{i=1}^n (X_i - \overline{X})^2}.
\end{aligned}
$$

- $s\{\hat{\beta}_0\}$ and $s\{\hat{\beta}_1\}$ are SE of $\hat{\beta}_0$ ad $\hat{\beta}_1$, respectively.
- SEs decrease with $\sum_{i=1}^n (X_i - \overline{X})^2 = (n-1)s_X^2$, which in turn increase with the sample size n and sample variance s_X^2 of X.

What are the implications?

Figure: Effects of the range of X on the variability of the fitted line



A Simulation Study

Simulate 100 data sets.

- $n = 5$ cases with the X values

$$X_1 = 1.86, \quad X_2 = 0.22, \quad X_3 = 3.55, \quad X_4 = 3.29, \quad X_5 = 1.25,$$

  fixed throughout all data sets.
- For each data set, the response variable is generated by:
  - First generate $\epsilon_1, \cdots, \epsilon_5$ i.i.d. from $N(0, 1)$.
  - Then set the response variable as:

$$Y_i = 2 + X_i + \epsilon_i, \quad i = 1, \cdots, 5.$$

- For each data set, derive the LS estimators $\hat{\beta}_0, \hat{\beta}_1$ and MSE.

- Data set 1:

case	X	Y
1	1.86	3.08
2	0.22	2.27
3	3.55	4.38
4	3.29	5.12
5	1.25	1.38

$\hat{\beta}_0 = 1.34,\ \hat{\beta}_1 = 0.94,\ MSE = 0.79.$

- Data set 2:

case	X	Y
1	1.86	2.91
2	0.22	2.13
3	3.55	5.35
4	3.29	5.76
5	1.25	2.01

$\hat{\beta}_0 = 1.19,\ \hat{\beta}_1 = 1.20,\ MSE = 0.52.$

- ..., ...
- Data set 100:

case	X	Y
1	1.86	3.36
2	0.22	2.50
3	3.55	5.93
4	3.29	5.36
5	1.25	2.67

$\hat{\beta}_0 = 1.75,\ \hat{\beta}_1 = 1.09,\ MSE = 0.24.$

Note how the X_is are kept fixed and how the LS estimators vary across these data sets.

Figure: Sampling distributions of $\hat{\beta}_0, \hat{\beta}_1, MSE$. Sample means are $1.99, 1.02, 1.04$ respectively. True parameters are $2, 1, 1$, respectively.

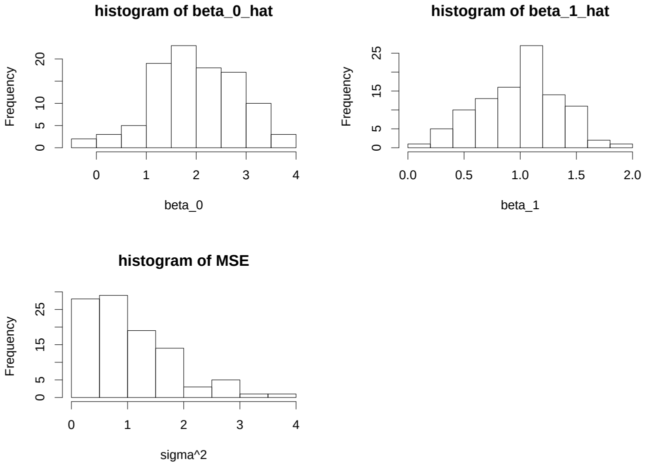
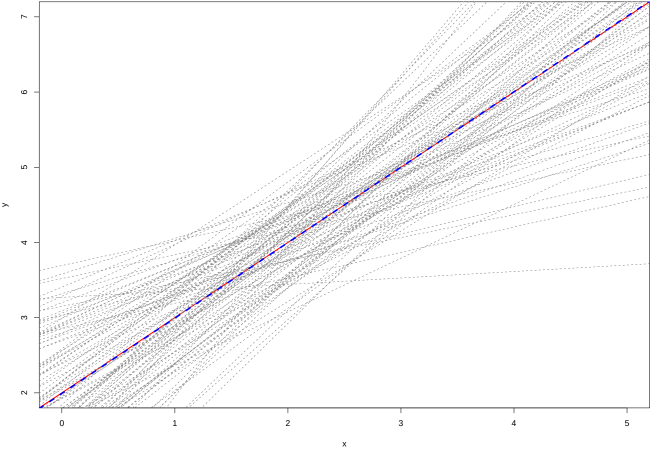


Figure: True: red solid; LS lines: grey broken; mean LS line: blue broken



We calculate the sample mean and sample standard deviation of these 100 realizations of $\hat{\beta}_0, \hat{\beta}_1$, respectively. Then compare them to the respective theoretical values.

- $\hat{\beta}_0$: Theoretical mean and standard deviation:

$$E(\hat{\beta}_0) = \beta_0 = 2, \quad \sigma\{\hat{\beta}_0\} = \sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right]} = 0.854.$$

  Sample mean and sample standard deviation: $1.99, 0.847$.

- $\hat{\beta}_1$: Theoretical mean and standard deviation:

$$E(\hat{\beta}_1) = \beta_1 = 1, \quad \sigma\{\hat{\beta}_0\} = \sqrt{\frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2}} = 0.358.$$

  Sample mean and sample standard deviation: $1.002, 0.36$.

Normal Error Model

Normal error model: Simple regression model $+$ Normality assumption.

- Model equation:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad i = 1, \ldots, n.$$

- Model assumptions: **The error terms ϵ_is are independently and identically distributed (i.i.d.) $N(0, \sigma^2)$ random variables.**

Notes: LS estimators $\hat{\beta}_0, \hat{\beta}_1$ are maximum likelihood estimators (MLE) of β_0, β_1, respectively. The MLE of σ^2 is SSE/n.

Sampling Distributions of LS Estimators

Under the Normal error model:

- $\hat{\beta}_0, \hat{\beta}_1$ are normally distributed:

$$\hat{\beta}_0 \sim N(\beta_0, \sigma^2\{\hat{\beta}_0\}), \quad \hat{\beta}_1 \sim N(\beta_1, \sigma^2\{\hat{\beta}_1\}).$$

  Notes: Use the facts (i) linear combinations of independent normal random variables are still normal random variables; (ii) $\hat{\beta}_0, \hat{\beta}_1$ *are linear combinations of the* Y_i*s.*

- SSE/σ^2 follows a χ^2 distribution with $n-2$ degrees of freedom, denoted by $\chi^2_{(n-2)}$.
- Moreover, SSE is independent with both $\hat{\beta}_0$ and $\hat{\beta}_1$.

Inference of Regression Coefficients

All inferences are under the Normal error model.

- **Studentized pivotal quantity**:

$$\frac{\hat{\beta}_1 - \beta_1}{s\{\hat{\beta}_1\}} \sim t_{(n-2)},$$

  where $t_{(n-2)}$ denotes the t-distribution with $n - 2$ degrees of freedom.

  - The numerator is the difference between the estimator and the parameter.
  - The denominator is the standard error of the estimator.
  - This quantity follows a known distribution, i.e., the t-distribution.

  Notes: Use the fact that if $Z \sim N(0,1)$, $S^2 \sim \chi^2_{(k)}$ and Z, S^2 are independent, then $\frac{Z}{\sqrt{S^2/k}} \sim t_{(k)}$.

Confidence Interval

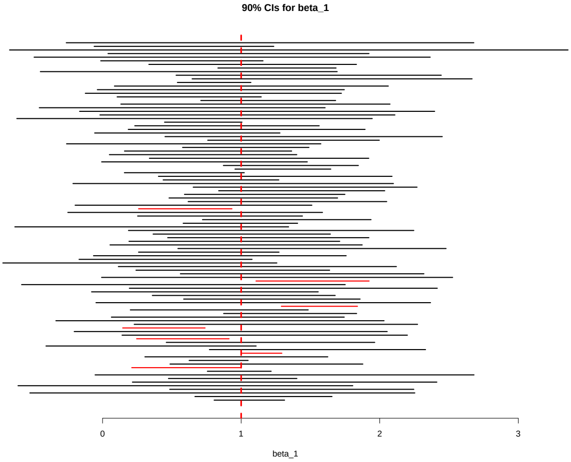
$(1 - \alpha)$-Confidence interval of β_1:

$$\hat{\beta}_1 \pm t(1 - \alpha/2; n - 2)s\{\hat{\beta}_1\},$$

where $t(1 - \alpha/2; n - 2)$ is the $(1 - \alpha/2)$th percentile of $t_{(n-2)}$.

How to construct confidence intervals for β_0?
How to interpret a confidence interval?

Figure: A Simulation Study



Heights

- Recall $n = 928$, $\overline{X} = 68.316, \sum_i X_i^2 = 4334058$, $\sum_{i=1}^n (X_i - \overline{X})^2 = \sum_{i=1}^n X_i^2 - n(\overline{X})^2 = 3038.761$. Also

$$\hat{\beta}_0 = 24.54, \ \ \hat{\beta}_1 = 0.637, \ \ MSE = 5.031.$$

  So

$$s\{\hat{\beta}_1\} = \sqrt{\frac{5.031}{3038.761}} = 0.0407.$$

- 95%-confidence interval of β_1:

$$\begin{aligned} 0.637 \pm t(0.975; 926) \times 0.0407 &= 0.637 \pm 1.963 \times 0.0407 \\ &= [0.557, 0.717]. \end{aligned}$$

  We are 95% confident that the regression slope is in between 0.557 and 0.717.

T-tests

- Null hypothesis: $H_0 : \beta_1 = \beta_1^{(0)}$, where $\beta_1^{(0)}$ is a given constant.
- T-statistic:

$$T^* = \frac{\hat{\beta}_1 - \beta_1^{(0)}}{s\{\hat{\beta}_1\}}.$$

- **Null distribution** of the T-statistic:

$$\text{Under } H_0 : \beta_1 = \beta_1^{(0)}, \quad T^* \sim t_{(n-2)}.$$

Can you derive the null distribution?

Decision rule at significance level α.

- Two-sided alternative $H_a : \beta_1 \neq \beta_1^{(0)}$: Reject H_0 if and only if $|T^*| > t(1 - \alpha/\mathbf{2}; n - 2)$, or equivalently, reject H_0 if and only if pvalue:$= P(|t_{(n-2)}| > |T^*|) < \alpha$.
- Left-sided alternative $H_a : \beta_1 < \beta_1^{(0)}$: Reject H_0 if and only if $T^* < t(\alpha; n - 2)$, or equivalently, reject H_0 if and only if pvalue:$= P(t_{(n-2)} < T^*) < \alpha$.
- Right-sided alternative $H_a : \beta_1 > \beta_1^{(0)}$: Reject H_0 if and only if $T^* > t(1 - \alpha; n - 2)$, or equivalently, reject H_0 if and only if pvalue:$= P(t_{(n-2)} > T^*) < \alpha$.

The decision rule depends on the form of the alternative hypothesis!

Why are the critical value approach and the pvalue approach equivalent?
How to conduct hypothesis testing with regard to β_0?

Heights

Test whether there is a linear association between parent's height and child's height. Use significance level $\alpha = 0.01$.

- The hypotheses: $H_0 : \beta_1 = 0$ *vs.* $H_a : \beta_1 \neq 0$.
- T statistic: $T^* = \frac{\hat{\beta}_1 - 0}{s\{\hat{\beta}_1\}} = \frac{0.637}{0.0407} = 15.7$.
- Critical value: $t(1 - 0.01/2; 928 - 2) = 2.58$. Since the observed $T^* = |15.7| > 2.58$, reject the null hypothesis at level 0.01.
- Or the pvalue $= P(|t_{(926)}| > |15.7|) \approx 0$. Since $pvalue < \alpha = 0.01$, reject the null hypothesis at level 0.01.
- Conclude that there is a significant association between parent's height and child's height.

Estimation of Mean Response

Given $X = X_h$, the mean response is $E(Y_h) = \beta_0 + \beta_1 X_h$.

- An unbiased point estimator for $E(Y_h)$ is :

$$\widehat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = \overline{Y} + \hat{\beta}_1 (X_h - \overline{X}).$$

$$E(\widehat{Y}_h) = \beta_0 + \beta_1 X_h = E(Y_h).$$

- Variance of $\widehat{Y}_h$ is:

$$\sigma^2\{\widehat{Y}_h\} = \sigma^2 \left[\frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2} \right].$$

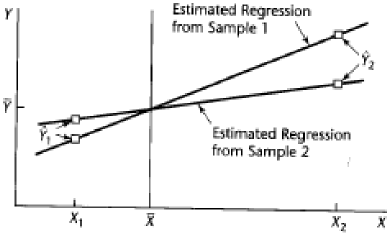
Notes: Use the fact that $\overline{Y}$ and $\hat{\beta}_1$ are uncorrelated.

- The larger the sample size and/or the larger the range of X_is, the smaller
  the variance of $\widehat{Y}_h$.

Figure: Effects of the distance of X_h from $\overline{X}$ on variability of $\widehat{Y}_h$.

Chapter 2 *Inferences in Re*

FIGURE 2.3 Effect on $\hat{Y}_h$ of Variation in b_1 from Sample to Sample in Two Samples with Same Means $\bar{Y}$ and $\bar{X}$.



The further is X_h from $\overline{X}$, the larger
is the variance of $\widehat{Y}_h$: The variability in the estimated slope $\hat{\beta}_1$ has a larger
effect on $\widehat{Y}_h$ when X_h is further away from the sample mean $\overline{X}$.

- Standard error of $\widehat{Y}_h$:

$$s^2\{\widehat{Y}_h\} = MSE\left[\frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right]. \tag{1}$$

- Under the Normal error model, $\widehat{Y}_h$ is normally distributed.
    - Studentized quantity:

$$\frac{\widehat{Y}_h - E(Y_h)}{s(\widehat{Y}_h)} \sim t_{(n-2)}.$$

    - $(1-\alpha)$- C.I.

$$\widehat{Y}_h \pm t(1-\alpha/2; n-2)s(\widehat{Y}_h).$$

Heights

Estimate the average height of children of 70*in* tall parents.

- Recall: $n = 928$, $\overline{X} = 68.316$, $\sum_{i=1}^{n}(X_i - \overline{X})^2 = 3038.761$, $\widehat{E(Y)} = 24.54 + 0.637X$ and $MSE = 5.031$.
- $\widehat{Y}_h = 24.54 + 0.637 \times 70 = 69.2$.
- Standard error:

$$s\{\widehat{Y}_h\} = \sqrt{5.031 \times \left\{\frac{1}{928} + \frac{(70 - 68.316)^2}{3038.761}\right\}} = 0.1.$$

- 95%-confidence interval of $E(Y_h)$:

$$69.2 \pm 1.8831 \times 0.1 = [68.96, 69.35], \quad t(0.975; 926) = 1.8831.$$

- We are 95% confident that the average height of children of 70*in* parents is between [68.96*in*, 69.35*in*].

Prediction of New Observation

Predict a **future observation** $Y_{h(new)}$ of the response variable corresponding to a given level of the predictor variable $X = X_h$.

- $Y_{h(new)} = \beta_0 + \beta_1 X_h + \epsilon_h$.
  - This is a new observation, so ϵ_h is assumed to be uncorrelated with ϵ_is.
  - Consequently, $Y_{h(new)}$ is uncorrelated with the observed Y_is.
- The **predicted value** for $Y_{h(new)}$ is simply the estimated mean response when $X = X_h$:

$$\widehat{Y}_h = \hat{\beta}_0 + \hat{\beta}_1 X_h = \overline{Y} + \hat{\beta}_1 (X_h - \overline{X}).$$

How would we perform prediction if the X value is unknown?

Distinction between prediction and mean estimation.

- $Y_{h(new)}$ is a "moving target" as it is a random variable. On the contrary, $E(Y_h)$ is a fixed non-random quantity.
- There are two sources of variations in the prediction process: Variability from the predicted value $\widehat{Y}_h$ and variability from the target $Y_{h(new)}$.

$$\sigma^2(pred) := Var(\widehat{Y}_h - Y_{h(new)}) = \sigma^2(\widehat{Y}_h) + \sigma^2(Y_{h(new)}) = \sigma^2(\widehat{Y}_h) + \sigma^2.$$

$$s^2\{pred\} = s^2(\widehat{Y}_h) + MSE = MSE\left[1 + \frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right]. \quad (2)$$

Note the difference between $s^2\{\widehat{Y}_h\}$ and $s^2\{pred\}$.

Prediction Intervals

- Studentized quantity:

$$\frac{\widehat{Y}_h - Y_{h(new)}}{s(pred)}.$$

- Under the Normal error model, it follows a $t_{(n-2)}$ distribution.
- $(1-\alpha)-$ prediction interval of $Y_{h(new)}$:

$$\widehat{Y}_h \pm t(1-\alpha/2; n-2)s(pred).$$

- Prediction interval is wider
  than the corresponding confidence interval of the mean response.
- With sample size becomes very large, the width of the confidence interval tend to vanish,
  but this would not happen for the prediction interval.

Heights

What would be the predicted height of the child of a 70*in* tall couple?

- Predicted height: $\widehat{Y}_h = 24.54 + 0.637 \times 70 = 69.2$. Standard error:

$$s\{pred\} = \sqrt{5.031 \times \left\{\mathbf{1} + \frac{1}{928} + \frac{(70 - 68.316)^2}{3038.761}\right\}} = 2.25.$$

- 95% prediction interval:

$$69.2 \pm 1.8831 \times 2.25 = [64.75, 73.56], \quad t(0.975; 926) = 1.8831.$$

- We are 95% confident that the child's height will be in between [64.75*in*, 73.56*in*].

Extrapolation

Extrapolation occurs when predicting the response variable for values of the predictor variable lying outside of the range of the observed data.

- Every model has a range of validity. Particularly, a model may be inappropriate when it is extended outside of the range of the observations upon which it was built.
- Extrapolations are often much less reliable than interpolation and need to be handled with caution, even though they can be of more interests to us (e.g. fortune telling).
- In the Heights example: Extrapolation would happen if we use the fitted regression line to predict heights of children of very tall or very short parents.

Analysis of Variance Approach

The basic idea of ANOVA is to attributing variation in the data to different sources.

- In regression, the variation in the observations Y_i is attributed to:
    -
    -
- ANOVA is performed through:
    - Partitioning sums of squares;
    - Partitioning degrees of freedoms;

Analysis of Variance Approach

The basic idea of ANOVA is to attributing variation in the data to different sources.

- In regression, the variation in the observations Y_i is attributed to:
  - variation of the error terms
  - variation of the values of the predictor variable(s)
- ANOVA is performed through:
  - partitioning sums of squares
  - partitioning degrees of freedoms

Partition of Total Deviations

- **Total deviations:** Difference between Y_i and the sample mean $\overline{Y}$:

$$Y_i - \overline{Y}, \quad i = 1, \cdots, n.$$

- Total deviations can be decomposed into the sum of two terms:

i.e., the *deviation of observed value around the fitted regression line (residual)* and the *deviation of fitted value from the mean*.

Partition of Total Deviations

- **Total deviations:** Difference between Y_i and the sample mean $\overline{Y}$:

$$Y_i - \overline{Y}, \quad i = 1, \cdots, n.$$

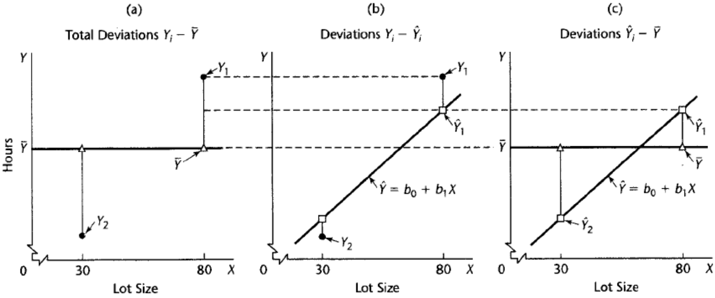
- Total deviations can be decomposed into the sum of two terms:

$$Y_i - \overline{Y} = (Y_i - \widehat{Y}_i) + (\widehat{Y}_i - \overline{Y}), \qquad i = 1, \ldots, n,$$

i.e., the *deviation of observed value around the fitted regression line (residual)* and the *deviation of fitted value from the mean*.

Figure: Partition of total deviation.

FIGURE 2.7 **Illustration of Partitioning of Total Deviations $Y_i - \bar{Y}$—Toluca Company Example (not drawn to scale; only observations Y_1 and Y_2 are shown).**



Decomposition of Total Variation

Decomposition of Total Variation

- Taking sum of squares of the total deviations and noting that the sum of the cross product terms vanishes:

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2.$$

- Decomposition of total variation:

$$SSTO = SSE + SSR$$

and

$$d.f.(SSTO) = d.f.(SSE) + d.f.(SSR).$$

Sum of Squares

- **Total sum of squares (SSTO)**:

  This is the variation of the observed Y_is around their sample mean.
- **Error sum of squares (SSE)**:

  This is the variation of the observed Y_is around the fitted regression line.

Sum of Squares

- **Total sum of squares (SSTO)**:

$$SSTO := \sum_{i=1}^{n} (Y_i - \overline{Y})^2, \quad d.f.(SSTO) = n - 1.$$

  This is the variation of the observed Y_is around their sample mean.

- **Error sum of squares (SSE)**:

$$SSE = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2, \quad d.f.(SSE) = n - 2.$$

  This is the variation of the observed Y_is around the fitted regression line.

- **Regression sum of squares (SSR)**:

  This is the variation of the fitted values around the sample mean. The the regression slope and the spread out the X_is, the larger is SSR.
- $SSR = SSTO - SSE$ is the effect of X in the variation in Y through linear regression.
- In other words, SSR is the in predicting Y by utilizing the predictor X through a linear regression model.

What is $\frac{1}{n}\sum_{i=1}^{n}\widehat{Y}_i$?

- **Regression sum of squares (SSR)**:

$$SSR = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{n}(X_i - \overline{X})^2, \quad d.f.(SSR) = 1.$$

  This is the variation of the fitted values around the sample mean: the larger the regression slope and the more spread out the X_is, the larger is SSR.

- $SSR = SSTO - SSE$ is the effect of X in reducing the variation in Y through linear regression. In other words, SSR is the reduction in uncertainty in predicting Y by utilizing the predictor X through a linear regression model.

What is $\frac{1}{n}\sum_{i=1}^{n}\widehat{Y}_i$?

Expected Values of SS

- Expected values of SS:

What is $E(SSTO)$?

- **Mean squares (MS): = SS / df(SS)**

$$MSE = \frac{SSE}{\text{d.f.}(SSE)} = \frac{SSE}{n-2}, \qquad MSR = \frac{SSR}{\text{d.f.}(SSR)} = \frac{SSR}{1}.$$

- Expected values of MS:

$$E(MSE) = \sigma^2, \qquad E(MSR) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Expected Values of SS

- Expected values of SS:

$$E(SSE) = (n-2)\sigma^2, \quad E(SSR) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

What is $E(SSTO)$?

- **Mean squares (MS): = SS / df(SS)**

$$MSE = \frac{SSE}{\text{d.f.}(SSE)} = \frac{SSE}{n-2}, \qquad MSR = \frac{SSR}{\text{d.f.}(SSR)} = \frac{SSR}{1}.$$

- Expected values of MS:

$$E(MSE) = \sigma^2, \qquad E(MSR) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Under Normal error model.

- Sampling distributions of SS : (scaled) χ^2 distributions

$$SSE \sim \sigma^2 \chi^2_{(n-2)}, \qquad SSR \sim (\sigma^2 + \beta_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2)\, \chi^2_{(1)}.$$

- *SSE* and *SSR* are independent.
  Notes: Recall SSE and $\hat{\beta}_1$ *are independent.*

What is the distribution of SSTO when $\beta_1 = 0$*?*

F Test

- $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$.
- F ratio:

$$F^* = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)}.$$

  - F^* fluctuates around $1 + \frac{\beta_1^2 \sum_{i=1}^n (X_i - \overline{X})^2}{\sigma^2}$.
  - A large value of F^* means evidence against H_0.
- Null distribution of F^*:

$$F^* \underset{H_0:\beta_1=0}{\sim} F_{1,n-2}.$$

Notes: Use the fact that if $Z_1 \sim \chi^2_{(df_1)}$, $Z_2 \sim \chi^2_{(df_2)}$ *and* Z_1, Z_2 *independent, then* $\frac{Z_1/df_1}{Z_2/df_2} \sim F_{df_1,df_2}$.

- Decision rule at level α:

$$\text{reject } H_0 \text{ if } \quad F^* > F(1-\alpha; 1, n-2),$$

where $F(1-\alpha; 1, n-2)$ is the $(1-\alpha)$-percentile of the $F_{1,n-2}$ distribution.

- **In simple linear regression, the F-test is equivalent to the t-test for testing $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$.**
Check the following.
    - $F^* = (T^*)^2$ where $T^* = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}$ is the T-statistic.
    - $F(1-\alpha; 1, n-2) = t^2(1-\alpha/2; n-2)$.

ANOVA Table

ANOVA table for simple linear regression.

Source of Variation	SS	d.f.	MS=SS/d.f.	F^*
Regression	$SSR = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$	d.f.$(SSR) = 1$	$MSR = SSR/1$	$F^* = MSR/MSE$
Error	$SSE = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$	d.f.$(SSE) = n - 2$	$MSE = SSE/(n-2)$	
Total	$SSTO = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$	d.f.$(SSTO) = n - 1$	$MSTO = SSTO/(n-1)$	

Heights

$n = 928$, $\overline{X} = 68.31578$, $\overline{Y} = 68.08227$, $\sum_i X_i^2 = 4334058$, $\sum_i Y_i^2 = 4307355$, $\sum_i X_i Y_i = 4318152$, $\hat{\beta}_1 = 0.637$, $\hat{\beta}_0 = 24.54$.

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n(\overline{Y})^2 \\
&= 4307355 - 928 \times 68.08227^2 = 5893. \\
SSR &= \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{n}(X_i - \overline{X})^2 \\
&= 0.637^2 \times \left[4334058 - 928 \times 68.31578^2\right] = 1234. \\
SSE &= SSTO - SSR = 4659.
\end{aligned}
$$

Heights (Cont'd)

Source of Variation	SS	d.f.	MS=SS/d.f.	F^*
Regression	$SSR = 1234$	d.f.$(SSR) = 1$	$MSR = 1234$	$F^* = MSR/MSE = 245$
Error	$SSE = 4659$	d.f.$(SSE) = 926$	$MSE = 5.03$	
Total	$SSTO = 5893$	d.f.$(SSTO) = 927$	$MSTO = 6.36$	

- Test whether there is a linear association between parent's height and child's height. Use significance level $\alpha = 0.01$.
- $F(0.99; 1, 926) = 6.66 < F^* = 245$, so reject $H_0 : \beta_1 = 0$ and conclude that there is a significant linear association between parent's height and child's height.
- Recall $T^* = 15.66$, $t(0.995; 926) = 2.58$ and check:

$$15.66^2 = 245, \quad 2.58^2 = 6.66.$$