Consider as usual, population level
\[ Y = \beta_0 + \beta_1 X + \varepsilon \] with \( \varepsilon \) indep.of \( X \), \( E(\varepsilon) = 0, \text{var}(\varepsilon) = \sigma^2 \)

In terms of sample data
\[ Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \] with \( \varepsilon_i \) iid \( \sim \varepsilon \)

LS estimates of the parameters of the regression
\[ b_1 = \left( \sum (x_i - \bar{x})y_i \right) \div \left( \sum (x_i - \bar{x})^2 \right) ; \quad b_0 = \bar{y} - b_1 \bar{x} \]

both linear functions of the sample \( Y \)'s.

Gauss-Markov theorem
\[ b_1 \text{ and } b_0 \text{ are unbiased, and have minimum variance among unbiased estimates linear in the sample } Y \text{'s.} \]
Estimates of variabilities:

- $HST = \frac{SST}{m-1} = \frac{1}{m-1} \sum_{i} (Y_i - \bar{Y})^2$
  is an unbiased estimate of the (overall) variance of $Y$, say $\sigma_Y^2$.

- $S^2 = HSE = \frac{SSE}{m-2} = \frac{1}{m-2} \sum_{i} (Y_i - (b_0 + b_1 x_i))^2 = \frac{1}{m-2} \sum_{i} e_i^2$
  is an unbiased estimate of the error variance $\sigma^2$ (i.e., of the common variance of $Y$ within each $X$-subpopulation: $\text{var}(Y|X) = \sigma^2$, any $X$).

- $HSR = \frac{SSR}{1} = \sum_{i} ((b_0 + b_1 x_i) - \bar{Y})^2 = \sum_{i} (\hat{Y}_i - \bar{Y})^2$
  which expresses the variability explained by the regression line - 1 d.o.f. (Its mean is $E(HSR) = \sigma^2 + \beta_1^2 \sum (x_i - \bar{x})^2$)

2
Two important reminders:

1. All our estimates are statistics to be calculated on the sample data — hence they are random variables with their own means, variances ... and more generally distributions.

2. All our discussion takes the sample y's as random ... but the sample x's as fixed/controlled (not random ...)

Recall also that the book does not use "conditional" notation (except for a section dedicated to regression "when X is random").
Variances of the LS estimates
based on their linear expressions in the sample Y's, these variances (minimal in the clan of unbiased linear estim's) are:

\[ \text{var}(b_i) = \frac{1}{n-1} \frac{s_y^2}{\sum (x_i - \bar{x})^2} \]

\[ \text{var}(b_0) = \left( \frac{1}{n} - \frac{\bar{x}}{\sum (x_i - \bar{x})^2} \right) s_y^2 \]

They can be (unbiasedly) estimated using the unbiased estimate of \( s_y^2 \):

\[ s_{b_i}^2 = \frac{1}{\sum (x_i - \bar{x})^2} s_y^2 \]

\[ s_{b_0}^2 = \left( \frac{1}{n} - \frac{\bar{x}}{\sum (x_i - \bar{x})^2} \right) s_y^2 \]

St. deviations are estimated taking square roots: \( s_{b_i} \) and \( s_{b_0} \)
Up to now, we have not assumed nor used
NORMALITY OF ERRORS. If we pass to
\[ y = \beta_0 + \beta_1 x + \varepsilon, \quad \text{indep } x, \quad \varepsilon \sim N(0, \sigma^2) \quad \text{(Population)} \]
\[ y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad \varepsilon_i \sim \text{iid } N(0, \sigma^2) \quad \text{(Sample)} \]

We know that the LS estimates \( b_1, b_0 \)

coincide with the ML estimates of \( \beta_1, \beta_0 \)

and that the ML estimate of \( \sigma^2 \) is the

"uncorrected" \( \frac{1}{n} \sum \varepsilon_i^2 = \left( \frac{n-2}{n} \right) \hat{\sigma}^2 \)

But we can say more: we can actually
give the distributions for our estimates

fixing the error distribution allows us to
derive the distribution of the estimator

this is what will allow us to construct
confidence intervals and perform testing

(... "inference", as meant in the book)
Again using their linear expressions in the sample $y$'s
\[ b_i \sim N(\beta_i, \text{var}(b_i)) ; \quad b_0 \sim N(\beta_0, \text{var}(b_2)) \]

Estimates of variability are distributed as "multiples" of chi-squares. In particular
\[ \frac{m-2}{\sigma^2} \cdot s^2 = \frac{m-2}{\sigma^2} \cdot \text{HSE} = \frac{\text{SSR}}{\sigma^2} \sim \chi^2_{m-2} \left(\text{the d.o.f. of SSR}\right) \]

If $\beta = 0$, \[ \frac{1}{\sigma^2} \cdot \text{HSE} = \frac{\text{SSR}}{\sigma^2} \sim \chi^2_{1} \left(\text{the d.o.f. of SSR}\right) \]

Moreover, under normality fitted values and residuals are independent.

From all this we get:
\[ \frac{b_i - \beta_i}{s_{b_i}} \sim t_{m-2} ; \quad \frac{b_0 - \beta_0}{s_{b_0}} \sim t_{m-2} \]

If $\beta = 0$, \[ \frac{\text{HSE}}{\text{MSE}} \sim F_{1, m-2} \quad \text{d.o.f. of SSE} \rightarrow \text{d.o.f. of SSE} \]

F distribution with 1 d.o.f. for the numerator, and $m-2$ d.o.f. for the denominator

F distrib. is only on $\mathbb{R}^+$
Confidence interval for the slope

\[ b_1 \pm t_\alpha S_{b_1} \]

contains the unknown \( \beta_1 \) with probability \( 1 - \alpha \)

Testing \( H_0: \beta_1 = 0 \) vs \( H_A: \beta_1 \neq 0 \)

test statistic \( \frac{b_1}{S_{b_1}} \sim t_{n-2} \)

level- \( \alpha \) rejection region: \( (-t_\alpha, +t_\alpha) \)

(Prob. of rejecting \( H_0 \) when true = \( \alpha \))

P-value associated to the observed value of \( \frac{b_1}{S_{b_1}} \), when testing the two-sided Hyp. system is twice the area on the right of it.

Recall \( t_\alpha \) just a symbol to indicate its dependence on \( \alpha \).
Confidence interval for the intercept

\[ b_0 \pm t_{\alpha/2} \, s_b_0 \]

contains the unknown \( \beta_0 \) with probability \( 1-\alpha \)

Testing \( \beta_0 = 0 \) vs \( \beta_0 \neq 0 \)

test statistic \( b_0/s_{b_0} \sim t_{n-2} \)

level-\( \alpha \) rejection region: \( <-t_{\alpha/2}, +t_{\alpha/2}> \)

(Pr. of rejecting \( H_0 \) when true = \( \alpha \))

p-value associated to the observed value of \( b_0/s_{b_0} \) when testing the two-sided hyp. system

is twice the area on the right of it
Testing the (overall) explanatory power of the regression

\[ H_0: \text{the regression has no explanatory power} \]

\[ H_1: \text{it does have explanatory power} \]

idea: compare the MSR \(\approx\) variability per d.a.f. explained by the reg.

to the MSE \(\approx\) variability per d.a.f. which is left unexplained, and attributed to the error

if MSR is sufficiently large compared to MSE, (NOTICE ONE-SIDED APPROACH) there is evidence against \( H_0 \).

Test statistic: \[ \frac{\text{MSR}}{\text{MSE}} \sim F_{1, n-2} \]

level-\( \alpha \) rejection region: \( > F_{\alpha} \)

(\( \alpha \) = \( \alpha \) probability of rejecting \( H_0 \) when true = \( \alpha \))

\( p \)-value associated to the observed value of \( \frac{\text{MSR}}{\text{MSE}} \) when testing the one-sided Hyp. system is the area on the right of it.
IMPORTANT:

when we restrict ourselves to a linear relationship in one explanatory variable

\[ H_0: \beta_1 = 0 \quad \text{equivalent} \quad H_0: \text{the regression has no explanatory power} \]

\[ H_1: \beta_1 \neq 0 \quad \text{equivalent} \quad H_1: \text{it does have expl. power} \]

test stat:

\[ t = \frac{b_1}{s_{b_1}} \overset{\sim}{\sim} t_{n-2} \]

\[ \frac{SSE}{HSE} = \left( \frac{b_1}{s_{b_1}} \right)^2 \]

Two-sided

\[ \text{ONE-SIDED TRANSLATION} \quad \text{TAKING THE SQUARE} \]

When we pass to polynomial (lin. in parameters...) regressions in one \( x \), or to regressions on many \( x \)'s, things become more complicated:

A t-test for each component (\( x, x^2, x^3 \)...) or variable (\( x_1, x_2, x_3 \)...) still one "overall" F-test for the regression and a "single" F-test for each component or variable given the others
Standard Regression Output

<table>
<thead>
<tr>
<th>Estimate</th>
<th>Std. Dev.</th>
<th>T Stat</th>
<th>P-Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const</td>
<td>$b_0$</td>
<td>$s_{b_0}$</td>
<td>$t_0$</td>
</tr>
<tr>
<td>X</td>
<td>$b_1$</td>
<td>$s_{b_1}$</td>
<td>$t_1$</td>
</tr>
</tbody>
</table>

↑ LS estim's of the params in the reg. line
↑ estim's of their std. dev.

THE "TEST" PART

*(*) Makes sense under normality of errors value of the test statistic, e.g., $T_i = \frac{b_i}{s_{b_i}}$ (for $\beta_i = 0$ as $H_0$)

p-value associated to it, e.g., $p_i$ (for two-sided hyp. sys., i.e., $\beta_i \neq 0$ as $H_A$)

The smaller the p-value, the more unlikely it is that $\beta_i = 0$ might hold, i.e., the stronger the evidence for $\beta_i \neq 0$ (for X playing a significant role in explaining Y along the linear model we have selected)
### Overall ANOVA Table for the Regression

<table>
<thead>
<tr>
<th>&quot;Source of variation&quot;</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Stat.</th>
<th>P-Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regr.</td>
<td>1</td>
<td>SSR</td>
<td>MSR</td>
<td></td>
<td>p</td>
</tr>
<tr>
<td>Error</td>
<td>n-2</td>
<td>SSE</td>
<td>MSE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>n-1</td>
<td>SST</td>
<td>MST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Partitioning d.o.f.'s and overall variability among sources. What is explained by the reg. and what is not -> error

- Makes sense under normality of the errors.

- Value of the test statistic, $F = \frac{MSR}{MSE}$

- P-value associated to it, p

- The smaller the p-value, the higher the explanatory power of the regression

\[ F_{n-2, n-1} \]

- Observed F
Evaluating a regression model "numerically":

- the T's of its various components \( (x_1, x_2, x^2, \ldots) \) or variables \( (x_1, x_2, x_3, \ldots) \)
- the coefficient of determ. 
  \[ R^2 = \frac{\text{SSR}}{\text{SST}} \in [0,1] \] (how close to 1)
- the adjusted \( R^2 \)
- the F in the overall ANOVA table
  \[ F = \frac{\text{MSR}}{\text{MSE}} \]

Regression line in one \( X \)
- only \( T_1 \) is of interest
- (for \( \beta_1 = 0 \) vs \( \beta_1 > 0 \))

will see later relevant for model with many comp's or variables

\[ F = T_1^2 \] (this equivalence vanishes for models with many comp's or variables)

... and "graphically":

Residual plot, more "sophisticated" residual plots,
plot for normality of residuals ... AND MANY OTHERS
will see when we get into "DIAGNOSTICS"