Interpreting output of the R function lm

(*) If H_0 is true, then F statistic will have (Snedecor's) F distribution with df = (2, 503).

3.6.3 Multiple Linear Regression

In order to fit a multiple linear regression model using least squares, we again use the `lm()` function. The syntax `lm(y∼x1+x2+x3)` is used to fit a model with three predictors, `x1`, `x2`, and `x3`. The `summary()` function now outputs the regression coefficients for all the predictors.

```
> lm.fit=lm(medv∼lstat+age,data=Boston)
> summary(lm.fit)

Call:
lm(formula = medv ∼ lstat + age, data = Boston)

Residuals:
    Min      1Q  Median      3Q     Max
 -15.98   -3.98   -1.28    1.97   23.16

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  33.2228     0.7308   45.46   <2e-16 ***
lstat        -1.0321     0.0482  -21.42   <2e-16 ***
age           0.0345     0.0122    2.83   0.0049 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 6.17 on 503 degrees of freedom
Multiple R-squared: 0.551,      Adjusted R-squared: 0.549
F-statistic:  309 on 2 and 503 DF,  p-value: <2e-16
```

$\hat{\beta} = \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$

lstat X_1

age X_2

$\widehat{SE}(\hat{\beta}_1)$

$\dfrac{\hat{\beta}_2 - 0}{\widehat{SE}(\hat{\beta}_2)}$

If $H_0: \beta_2 = 0$ is true, then $\sim$ Student's t distr. with df = n-3 = 503

p-value for the test $H_0: \beta_2 = 0$ (2-sided)

$H_A: \beta_2 \neq 0$

: $\hat{\sigma}$ - an estimate of σ

R^2: coeff. of determination

(*) $H_0: \beta_1 = 0$ and $\beta_2 = 0$ (no associations w. any of the two covariates)

H_A: at least one of β_1 or $\beta_2 \neq 0$. (*)

CC balance vs ethnicity example:

3 ~~sub~~groups : AA, Asian, Caucasian.

Model 1: The mean is the same for every group, i.e., $E(Y_i) = \mu \ \forall i$.

Model 2: ~~sub~~group-specific means, i.e.,

$$E(Y_i) = \mu_j \text{ if the } i\text{th person belongs to group } j.$$

Representation 1:

$y_i = \mu_1 + \varepsilon_i$ if ith person is AA

$y_i = \mu_2 + \varepsilon_i$ if — " — Asian

$y_i = \mu_3 + \varepsilon_i$ if — " — Caucasian

Representation 2:

$y_i = \beta_0 + \varepsilon_i$

$y_i = \beta_0 + \beta_1 + \varepsilon_i$

$y_3 = \beta_0 + \beta_2 + \varepsilon_i$

Example: 2 people from each group. Write out the design matrices for MLR.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix} \ \underset{\text{equiv.}}{\equiv} \ \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}.$$