MSc 1 DS&AI

Statistical inference part 2

Practice session 2: Simple and multiple linear regression

Exercises on simple linear regression

6.2 **(a)** Show that $E(\hat{\beta}_1) = \beta_1$ as in (6.7).

(b) Show that $E(\hat{\beta}_0) = \beta_0$ as in (6.8).

with $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

and $E[y_i] = \beta_0 + \beta_1 x_i$

6.9 **(a)** Obtain a test for $H_0 : \beta_0 = a$ versus $H_1 : \beta_0 \neq a$.

(b) Obtain a confidence interval for β_0.

Use $\hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \overbrace{\sigma^2\left[1/n + \bar{x}^2/\sum_{i=1}^{n}(x_i - \bar{x})^2\right]}^{c}\right)$, $(n-2)\dfrac{s^2}{\sigma^2} \sim \chi^2_{(n-2)}$

and $\hat{\beta}_0$ and s^2 independent. Thus deduce the distribution of $\dfrac{\hat{\beta}_0 - \beta_0}{\sigma\sqrt{c}}$ ...

6.14 Table 6.1 (Weisberg 1985, p. 231) gives the data on daytime eruptions of Old Faithful Geyser in Yellowstone National Park during August 1–4, 1978. The variables are $x =$ duration of an eruption and $y =$ interval to the next eruption. Can x be used to successfully predict y using a simple linear model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$? see file geyser.csv

(a) Find $\hat{\beta}_0$ and $\hat{\beta}_1$.

(b) Test $H_0 : \beta_1 = 0$ using (6.14).

(c) Find a confidence interval for β_1.

(d) Find r^2 using (6.16).

Exercises on multiple linear regression

Proove (i) of theorem 7.6b

Theorem 7.6b. Suppose that $\mathbf{y}$ is $N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, where $\mathbf{X}$ is $n \times (k+1)$ of rank $k+1 < n$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$. Then the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ given in Theorem 7.6a have the following distributional properties:

(i) $\hat{\boldsymbol{\beta}}$ is $N_{k+1}[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$.

use that $\hat{\beta} = (X'X)^{-1} X y$ | Thus give $\mathrm{cov}(\hat{\beta})$

7.2 Show that (7.10) follows from (7.9). Why is $\mathbf{X}'\mathbf{X}$ positive definite, as noted below (7.10)?

7.29 (**a**) Show that R^2 in (7.55) can be written in the form $R^2 = 1 - \mathrm{SSE}/\sum_i (y_i - \bar{y})^2$.

(**b**) Replace SSE and $\sum_i (y_i - \bar{y})^2$ in part (a) by variance estimators $\mathrm{SSE}/(n-k-1)$ and $\sum_i (y_i - \bar{y})^2/(n-1)$ and show that the result is the same as R_a^2 in (7.56).

7.53 When gasoline is pumped into the tank of a car, vapors are vented into the atmosphere. An experiment was conducted to determine whether y, the amount of vapor, can be predicted using the following four variables based on initial conditions of the tank and the dispensed gasoline:

data gas.csv

$$x_1 = \text{tank temperature } (^\circ\text{F})$$
$$x_2 = \text{gasoline temperature } (^\circ\text{F})$$
$$x_3 = \text{vapor pressure in tank (psi)}$$
$$x_4 = \text{vapor pressure of gasoline (psi)}$$

The data are given in Table 7.3 (Weisberg 1985, p. 138).

(**a**) Find $\hat{\boldsymbol{\beta}}$ and s^2.

(**b**) Find an estimate of $\mathrm{cov}(\hat{\boldsymbol{\beta}})$.

~~(**c**) Find $\hat{\boldsymbol{\beta}}_1$ and $\hat{\beta}_0$ using $\mathbf{S}_{xx}$ and $\mathbf{s}_{yx}$ as in (7.46) and (7.47).~~

(**d**) Find R^2 and R_a^2.