VE401, Probabilistic Methods in Eng.

Recitation Class - Week 9

Zhanpeng Zhou

UMJI-SJTU Joint Institute

April 20, 2021

Mail: zzp1012@sjtu.edu.cn / WeChat: zzp01012

Table of contents

Estimating Correlation

Estimator for correlation. The unbiased estimators for variance and covariance are given by

$$\widehat{\mathrm{Var}[X]} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

$$\widehat{\mathrm{Var}[Y]} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2,$$

$$\widehat{\mathrm{Cov}[X,Y]} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}),$$

giving

$$R := \widehat{\rho} = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2}\sqrt{\sum(Y_i - \overline{Y})^2}}.$$

Hypothesis Tests for the Correlation Coefficient

Distribution. Suppose (X, Y) follows a bivariate normal distribution with relation coefficient $\rho \in (-1, 1)$. For large sample size n, the Fisher transformation of R

$$\frac{1}{2} \ln \left(\frac{1+R}{1-R} \right) = \text{Artanh}(R)$$

is approximately normal with

$$\mu = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) = \text{Artanh}(\rho), \qquad \sigma^2 = \frac{1}{n-3}.$$

Hypothesis Tests for the Correlation Coefficient

Confidence interval. A $100(1-\alpha)\%$ confidence interval for ρ is given by

$$\left[\frac{1+R-(1-R)e^{2z_{\alpha/2}/\sqrt{n-3}}}{1+R+(1-R)e^{2z_{\alpha/2}/\sqrt{n-3}}}, \frac{1+R-(1-R)e^{-2z_{\alpha/2}/\sqrt{n-3}}}{1+R+(1-R)e^{-2z_{\alpha/2}/\sqrt{n-3}}}\right]$$

or

$$\tanh\left(\operatorname{Artanh}(R) \pm \frac{z_{\alpha/2}}{\sqrt{n-3}}\right).$$

Hypothesis Tests for the Correlation Coefficient

Test for correlation coefficient. Suppose $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a sample of size n from a bivariate normal population (X, Y) with correlation coefficient $\rho \in (-1, 1)$. The test statistic is given by

$$\begin{aligned} Z &= \frac{\sqrt{n-3}}{2}\left(\ln\left(\frac{1+R}{1-R}\right) - \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)\right) \\ &= \sqrt{n-3}(\text{Artanh}(R) - \text{Artanh}(\rho_0)). \end{aligned}$$

We reject at significance level α

- $H_0 : \rho = \rho_0$ if $|Z| > z_{\alpha/2}$,
- $H_0 : \rho \leq \rho_0$ if $Z > z_\alpha$,
- $H_0 : \rho \geq \rho_0$ if $Z < -z_\alpha$.

Simple Linear Regression Model

Model. We assume that

$$Y|x = \beta_0 + \beta_1 x + E,$$

where $\mathrm{E}[E] = 0$. We want to find estimators

$$B_0 := \widehat{\beta_0} = \text{estimator for } \beta_0, \qquad b_0 = \text{estimate for } \beta_0,$$
$$B_1 := \widehat{\beta_1} = \text{estimator for } \beta_1, \qquad b_1 = \text{estimate for } \beta_1.$$

Assumptions.

- For each value of x, the random variable follows a normal distribution with variance σ^2 and mean $\mu_{Y|x} = \beta_0 + \beta_1 x$.
- The random variables $Y|x_1$ and $Y|x_2$ are independent if $x_1 \neq x_2$.

Least Squares Estimation

Least squares estimation. We have the ***error sum of squares***

$$\mathrm{SS_E} := \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2.$$

To minimize it, we take

$$\frac{\partial \mathrm{SS_E}}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) = 0,$$
$$\frac{\partial \mathrm{SS_E}}{\partial b_1} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) x_i = 0.$$

which gives

$$b_1 = \frac{S_{xy}}{S_{xx}}, \qquad b_0 = \overline{y} - b_1 \overline{x},$$

Useful Properties

Properties.

$$S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2 = \sum_{i=1}^{n}(x_i - \overline{x})x_i = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = \sum_{i=1}^{n} x_i^2 - n\overline{x}^2,$$

$$S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(y_i - \overline{y})y_i = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2,$$

$$S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n}(x_i - \overline{x})y_i = \sum_{i=1}^{n}(y_i - \overline{y})x_i = \sum_{i=1}^{n} x_i y_i - n\overline{x}\cdot\overline{y}$$

$$= \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right).$$

$$b_1 = \frac{S_{xy}}{S_{xx}}, \qquad b_0 = \overline{y} - b_1\overline{x}, \qquad \mathrm{SS_E} = S_{yy} - b_1 S_{xy}.$$

Useful Properties

Properties. The last property follows from

$$
\begin{aligned}
\mathrm{SS_E} &= \sum_{i=1}^{n}(y_i - \overline{y} + \overline{y} - (b_0 + b_1 x_i))^2 \\
&= \sum_{i=1}^{n}(y_i - \overline{y} + \overline{y} - (\overline{y} - b_1\overline{x} + b_1 x_i))^2 \\
&= \sum_{i=1}^{n}(y_i - \overline{y})^2 + \sum_{i=1}^{n} b_1^2(\overline{x} - x_i)^2 - 2\sum_{i=1}^{n} b_1(y_i - \overline{y})(x_i - \overline{x}) \\
&= S_{yy} + b_1 \cdot \frac{S_{xy}}{S_{xx}} \cdot S_{xx} - 2b_1 S_{xy} \\
&= S_{yy} - b_1 S_{xy}.
\end{aligned}
$$

Distribution of Estimator for Variance

LSE for variance. An unbiased estimator for variance σ^2 is given by

$$S^2 = \frac{\mathrm{SS_E}}{n-2} = \frac{1}{n-2}\sum_{i=1}^{n}(Y_i - \widehat{\mu}_{Y|x_i})^2.$$

Distribution of estimator for variance. The statistic

$$\chi^2_{n-2} = \frac{(n-2)S^2}{\sigma^2} = \frac{\mathrm{SS_E}}{\sigma^2}$$

follows a chi-squared distribution with $n-2$ degrees of freedom.

Distribution of B_1

Theorem. The least squares estimator B_1 for β_1 follows a normal distribution with

$$\mathrm{E}[B_1] = \beta_1, \qquad \mathrm{Var}[B_1] = \frac{\sigma^2}{\sum(x_i - \overline{x})^2} = \frac{\sigma^2}{S_{xx}}.$$

Proof.

$$\begin{aligned} B_1 &= \frac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y}) = \frac{1}{S_{xx}} \sum (x_i - \overline{x}) Y_i \\ &= \frac{1}{S_{xx}} \sum (x_i - \overline{x})(\beta_0 + \beta_1 x_i + E_i) \\ &= \frac{1}{S_{xx}} \cdot \beta_1 S_{xx} + \frac{1}{S_{xx}} \sum (x_i - \overline{x}) E_i \\ &= \beta_1 + \frac{\sum (x_i - \overline{x}) E_i}{S_{xx}}. \end{aligned}$$

Distribution of B_1 with Estimated Variance

Distribution. The statistics

$$T_{n-2} = \frac{B_1 - \beta_1}{S/\sqrt{S_{xx}}}$$

follows T-distributions with $n-2$ degrees of freedom.

Confidence interval. The $100(1-\alpha)\%$ confidence interval of β_1 is given by

$$B_1 \pm t_{\alpha/2,n-2}\frac{S}{\sqrt{S_{xx}}}.$$

Test for Significance

Test for significance of regression. Let $(x_i, Y|x_i), i = 1, \ldots, n$ be a random sample from $Y|x$. We reject

$$H_0 : \beta_1 = 0$$

at significance level α if the test statistic

$$T_{n-2} = \frac{B_1}{S/\sqrt{S_{xx}}}$$

satisfies $|T_{n-2}| > t_{\alpha/2,n-2}$.

Distribution of B_0

Theorem. The least squares estimator B_0 for β_0 follows a normal distribution with

$$\mathrm{E}[B_0] = \beta_0, \qquad \mathrm{Var}[B_0] = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \overline{x})^2}.$$

Proof. Since

$$\begin{aligned} B_0 &= \overline{Y} - B_1 \overline{x} = \beta_0 + \beta_1 \overline{x} + \overline{E} - \left(\beta_1 + \frac{\sum (x_i - \overline{x}) E_i}{S_{xx}} \right) \overline{x} \\ &= \beta_0 + \overline{E} - \frac{\overline{x} \sum (x_i - \overline{x}) E_i}{S_{xx}}, \end{aligned}$$

we can see that

$$\begin{aligned} \mathrm{E}[B_0] = \beta_0, \qquad \mathrm{Var}[B_0] &= \frac{\sigma^2}{n} + \overline{x}^2 \frac{\sigma^2}{S_{xx}} - 2 \sum \mathrm{Cov} \left[\frac{E_i}{n}, \frac{\overline{x} \sum (x_i - \overline{x}) E_i}{S_{xx}} \right] \\ &= \sigma^2 \cdot \frac{S_{xx} + n\overline{x}^2}{n S_{xx}} = \frac{\sigma^2 \sum x_i^2}{n S_{xx}}. \end{aligned}$$

Distribution of B_0 with Estimated Variance

Distribution. The statistics

$$T_{n-2} = \frac{B_0 - \beta_0}{S\sqrt{\sum x_i^2}/\sqrt{nS_{xx}}}$$

follows T-distributions with $n-2$ degrees of freedom.

Confidence interval. The $100(1-\alpha)\%$ confidence interval of β_0 is given by

$$B_0 \pm t_{\alpha/2,n-2}\frac{S\sqrt{\sum x_i^2}}{\sqrt{nS_{xx}}}.$$

Distribution of Estimated Mean

Distribution. The estimated mean $\widehat{\mu}_{Y|x}$ follows a normal distribution with mean and variance

$$\mathrm{E}[\widehat{\mu}_{Y|x}] = \mu_{Y|x}, \qquad \mathrm{Var}[\widehat{\mu}_{Y|x}] = \left(\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}\right)\sigma^2.$$

Therefore, the statistic

$$T_{n-2} = \frac{\widehat{\mu}_{Y|x} - \mu_{Y|x}}{S\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}}$$

follows a T-distribution with $n-2$ degrees of freedom. A $100(1-\alpha)\%$ confidence interval for $\mu_{Y|x}$ is given by

$$\widehat{\mu}_{Y_x} \pm t_{\alpha/2,n-2} S\sqrt{\frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}.$$

Distribution and CI for Predictor

Predictor. The statistic $Y|x - \widehat{Y|x}$ follows a normal distribution with mean and variance

$$\mathrm{E}[Y|x - \widehat{Y|x}] = 0, \qquad \mathrm{Var}[Y|x - \widehat{Y|x}] = \left(1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}\right)\sigma^2.$$

Therefore, the statistic

$$T_{n-2} = \frac{Y|x - \widehat{Y|x}}{S\sqrt{1 + \frac{1}{n} + \frac{(x-\overline{x})^2}{S_{xx}}}}$$

follows a T-distribution with $n - 2$ degrees of freedom. A $100(1 - \alpha)\%$ confidence interval for $Y|x$ is given by

$$\widehat{Y|x} \pm t_{\alpha/2,n-2} S\sqrt{1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}}.$$

Model Analysis

Crucial quantities.

- ***Total sum of squares***:

$$\mathrm{SS_T} = S_{yy} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2.$$

- ***Error sum of squared***:

$$\mathrm{SS_E} = \sum_{i=1}^{n} (Y_i - (B_0 + B_1 x_i))^2 = S_{yy} - B_1 S_{xy} = S_{yy} - \frac{S_{xy}^2}{S_{xx}}.$$

- ***Coefficient of determination***: the proportion of the total variation in Y that is explained by the linear model.

$$R^2 = \frac{\mathrm{SS_T} - \mathrm{SS_E}}{\mathrm{SS_T}} = \frac{S_{xy}^2}{S_{xx} S_{yy}}.$$

Test for Significance with R^2

Test for significance of regression. Let $(x_i, Y|x_i), i = 1, \ldots, n$ be a random sample from $Y|x$. We reject

$$H_0 : \beta_1 = 0$$

at significance level α if the test statistic

$$T_{n-2} = \frac{B_1}{S/\sqrt{S_{xx}}} = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

satisfies $|T_{n-2}| > t_{\alpha/2,n-2}$.

Test for Correlation with R^2

Test for correlation. Let (X, Y) follow a bivariate normal distribution with correlation coefficient $\rho \in (-1, 1)$. Let R be the estimator for ρ. Then we reject

$$H_0 : \rho = 0$$

at significance level α if the test statistic

$$T_{n-2} = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

satisfies $|T_{n-2}| > t_{\alpha/2,n-2}$.

Lack-of-Fit and Pure Error

Source of $\mathrm{SS_E}$. $\mathrm{SS_E}$ is the variance of Y explained by the model.

- ***Error sum of squares due to pure error***:

$$\mathrm{SS_{E,pe}} := \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y}_i)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^2 - \sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i} Y_{ij}\right)^2.$$

  The statistic $\mathrm{SS_{E,pe}}/\sigma^2$ follows a chi-squared distribution with $\sum_{i=1}^{k}(n_i-1) = n-k$ degrees of freedom. (Recall that with sample variance S^2 of some sample $X_1,\ldots,X_n$, the statistic $(n-1)S^2/\sigma^2$ follows chi-squared distribution with $n-1$ d.o.f.)

- ***Error sum of squares due to lack of fit***:

$$\mathrm{SS_{E,lf}} := \mathrm{SS_E} - \mathrm{SS_{E,pe}}.$$

  The statistic $\mathrm{SS_{E,lf}}/\sigma^2$ follows a chi-squared distribution with $k-2$ degrees of freedom.

Testing for Lack of Fit

Test for lack of fit. Let $x_1, \ldots, x_k$ be regressors and $Y_{i1}, \ldots, Y_{in_i}$, $i = 1, \ldots, k$ the measured responses at each of the regressors. Let $\mathrm{SS}_{\mathrm{E,pe}}$ and $\mathrm{SS}_{\mathrm{E,lf}}$ be the pure error and lack-of-fit sums of squares for a linear regression model. Then we reject at significance level α

$$H_0 : \text{the linear regression model is appropriate}$$

if the test statistic

$$F_{k-2,n-k} = \frac{\mathrm{SS}_{\mathrm{E,lf}}/(k-2)}{\mathrm{SS}_{\mathrm{E,pe}}/(n-k)}$$

satisfies $F_{k-2,n-k} > f_{\alpha,k-2,n-k}$.

Calculations for Simple Linear Regression

1. Find $\sum x_i, \sum y_i, \sum x_i^2, \sum y_i^2, \sum x_i y_i$ and calculate

$$S_{xx} = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2, \quad S_{yy} = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2,$$
$$S_{xy} = \sum x_i y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum y_i\right).$$

2. Obtain b_1 and b_0 by

$$b_1 = \frac{S_{xy}}{S_{xx}}, \qquad b_0 = \overline{y} - b_1\overline{x}.$$

3. Calculate other quantities as required, e.g.,

$$\mathrm{SS_E} = S_{yy} - \frac{S_{xy}}{S_{xx}}, \qquad R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}.$$

Gradient of Matrix

Gradient. Suppose $a, x \in \mathbb{R}^n$, $A \in \text{Mat}(n \times n; \mathbb{R})$, then we have the following properties.

- $\nabla_x(a^T x) = a$, since

$$a^T x = \sum_{i=1}^{n} a_i x_i \quad \Rightarrow \quad \nabla_x(x^T x) = \begin{pmatrix} \frac{\partial a^T x}{x_1} \\ \vdots \\ \frac{\partial a^T x}{x_n} \end{pmatrix} = a.$$

- $\nabla_x(x^T A x) = 2Ax$ if $A^T = A$, since

$$x^T A x = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j = \sum_{i=1}^{n} a_{ii} x_i^2 + 2\sum_{i<j} x_i a_{ij} x_j$$

$$\Rightarrow \quad \frac{\partial x^T A x}{\partial x_i} = 2\sum_{j=1}^{n} a_{ij} x_j \quad \Rightarrow \quad \nabla_x(x^T x) = 2Ax.$$

Idempotent Matrix

Idempotent matrix. A $n \times n$ matrix P satisfying the property that $P^2 = P$ is called idempotent. Then

$$(1_n - P)^2 = 1_n - P - P + P^2 = 1_n - P$$

is also idempotent. Furthermore, its eigenvalues may only be 0 or 1, since

$$\lambda v = Pv = P^2 v = P(\lambda v) = \lambda^2 v,$$

where v is an eigenvector. This gives $\lambda^2 = \lambda$. In lecture slides, P is an ***orthogonal projection*** if

$$P^2 = P, \qquad P^T = P.$$

The Spectral Theorem of Linear Algebra

Spectral theorem. Let $A \in \mathrm{Mat}(n \times n; \mathbb{R})$ be a self-adjoint matrix, which means $A = A^* = A^T$. Then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

Corollary of spectral theorem. Let $A \in \mathrm{Mat}(n \times n; \mathbb{R})$ be a self-adjoint matrix. Then if $(v_1, \ldots, v_n)$ is an orthonormal basis of eigenvectors of A and $U = (v_1, \ldots, v_n)$, then $U^{-1} = U^T$, and

$$D = U^{-1}AU = U^T AU$$

is a diagonal matrix containing eigenvectors of A. This can be seen from

$$De_k = U^{-1}AUe_k = U^{-1}Av_k = U^{-1}\lambda_k v_k = \lambda_k e_k.$$

Important Results from Linear Algebra for Regression

Results used multiple linear regression.

- If $A \in \mathrm{Mat}(n \times n; \mathbb{R})$ is idempotent, then $1_n - A$ is idempotent, and A has eigenvalues only 0 or 1.
- If $A \in \mathrm{Mat}(n \times n; \mathbb{R})$ is symmetric, then there exists a matrix $U = (v_1, \ldots, v_n)$ of eigenvectors of A and $U^{-1} = U^T$ such that

$$D = U^T A U \quad \Rightarrow \quad A = U D U^T.$$

- In discussions of multiple linear regression, the two properties above hold for matrices

$$P = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \quad \text{and} \quad H = X(X^T X)^{-1} X^T,$$

and thus $1_n - P$ and $1_n - H$.

Polynomial Regression Model

Model. For a polynomial model, we assume that

$$Y|x = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + E \quad \Leftrightarrow \quad Y = X\beta + E,$$

where

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 & \cdots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^p \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \quad E = \begin{pmatrix} E_1 \\ \vdots \\ E_n \end{pmatrix}.$$

Assumptions.

- For each value of x, the random variable follows a normal distribution with variance σ^2 and mean $\mu_{Y|x} = \beta_0 + \beta_1 x + \cdots + \beta_p x^p$.
- The random variables $Y|x_1$ and $Y|x_2$ are independent if $x_1 \neq x_2$.

The Multilinear Model

Model. For a multilinear model, we assume that Y depends on several factors,

$$Y|x = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + E \quad \Leftrightarrow \quad Y = X\beta + E,$$

where

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}, \quad E = \begin{pmatrix} E_1 \\ \vdots \\ E_n \end{pmatrix}.$$

Assumptions.

- For each value of x, the random variable follows a normal distribution with variance σ^2 and mean $\mu_{Y|x} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$.
- The random variables $Y|x_1$ and $Y|x_2$ are independent if $x_1 \neq x_2$.

Least Squares Estimation

Least squares estimation. For both cases, we have the error sum of squares

$$\mathrm{SS_E} = \langle Y - Xb, Y - Xb \rangle = (Y - Xb)^T (Y - Xb).$$

To minimize it, we take

$$\begin{aligned}
\nabla_b \mathrm{SS_E} &= \nabla_b (Y - Xb)^T (Y - Xb) \\
&= \nabla_b \left(Y^T Y - Y^T Xb - b^T X^T Y + b^T X^T Xb \right) \\
&= -2X^T Y + 2X^T Xb = 0 \quad \Rightarrow \quad b = (X^T X)^{-1} X^T Y,
\end{aligned}$$

where we have used since both $Y^T Xb$ and $b^T X^T Y$ are scalars,

$$b^T X^T Y = (b^T X^T Y)^T = Y^T Xb.$$

Error Analysis

Crucial quantities.

- ***Total variation***: given orthogonal projection P,

$$P := \frac{1}{n}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \quad \Rightarrow \quad (1_n - P)^2 = 1_n - P,$$

giving

$$\mathrm{SS_T} = \langle (1_n - P)Y, (1_n - P)Y \rangle = \langle Y, (1_n - P)Y \rangle.$$

- ***Sum of squares error***: given orthogonal projection H,

$$\begin{aligned} H := X(X^TX)^{-1}X^T \quad \Rightarrow \quad \mathrm{SS_E} &= \langle Y - Xb, Y - Xb \rangle \\ &= \langle (1_n - H)Y, (1_n - H)Y \rangle \\ &= \langle Y, (1_n - H)Y \rangle = \langle E, (1_n - H)E \rangle. \end{aligned}$$

- ***Coefficient of multiple determination***:

$$R^2 = \frac{\mathrm{SS_R}}{\mathrm{SS_T}}, \quad \mathrm{SS_R} = \mathrm{SS_T} - \mathrm{SS_E} = \langle Y, (H - P)Y \rangle = \langle (H - P)Y, (H - P)Y \rangle.$$

Distribution of SS_E

Distribution of sum of squares error. The statistic given by the SS_E and variance σ^2

$$\begin{aligned}\frac{SS_E}{\sigma^2} &= \left\langle \frac{E}{\sigma}, (1_n - H)\frac{E}{\sigma} \right\rangle = \langle Z, (1_n - H)Z \rangle \\ &= \langle Z, UD_{n-p-1}U^T Z \rangle = \langle U^T Z, D_{n-p-1}U^T Z \rangle \\ &= \sum_{i=1}^{n-p-1} (U^T Z)_i^2, \quad \left(U^T Z \sim \mathrm{N}(0, \sigma^2 UU^T) = \mathrm{N}(0, \sigma^2 1_n) \right)\end{aligned}$$

follows a chi-squared distribution with $n - p - 1$ degrees of freedom, where the matrix U contains columns of eigenvectors of $(1_n - H)$ such that

$$U^T(1_n - H)U = D_{n-p-1} \quad \Rightarrow \quad 1_n - H = UD_{n-p-1}U^T.$$

Distribution of SS_E

- SS_E/σ^2 follows a chi-squared distribution with $n-p-1$ degrees of freedom.
- If $\beta = (\beta_0, 0, \ldots, 0)$, then SS_R/σ^2 follows a chi-squared distri-bution with p degrees of freedom.
- SS_R and SS_E are independent random variables. (Fisher-Cochran theorem.)
- An unbiased estimator for σ^2 is given by

$$\widehat{\sigma}^2 = S^2 = \frac{SS_E}{n-p-1}.$$

- The regression sum of squares can be expressed as

$$\begin{aligned} SS_R &= \langle Xb, Y\rangle - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2 \\ &= b_0 \sum_{i=1}^{n} Y_i + \sum_{j=1}^{p} b_j \sum_{i=1}^{n} x_{ji} Y_i - \frac{1}{2}\left(\sum_{i=1}^{n} Y_i\right)^2, \end{aligned}$$

for multilinear model, and substitute x_{ji} with x_i^j for polynomial model.

F-Test for Significance of Regression

F-test for significance of regression. Let $x_1, \ldots, x_p$ be the predictor variables in a multilinear model for Y. Then we reject at significance level α

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

if the test statistic

$$F_{p,n-p-1} = \frac{\mathrm{SS_R}/p}{\mathrm{SS_E}/(n-p-1)} = \frac{\mathrm{SS_R}/p}{S^2} = \frac{n-p-1}{p}\frac{R^2}{1-R^2}$$

satisfies $F_{p,n-p-1} > f_{\alpha,p,n-p-1}$.