Lecture 8

September 18, 2018

Normal Regression Theory

Normal Regression Theory

- Everything that we did so far was only under the assumption that the errors $e_1, \ldots, e_n$ were uncorrelated, had mean zero and variance σ^2.

Normal Regression Theory

- Everything that we did so far was only under the assumption that the errors $e_1, \dots, e_n$ were uncorrelated, had mean zero and variance σ^2.
- But if we want to test hypotheses about or if we want confidence intervals for linear combinations of β, we need distributional assumptions on the errors.

Normal Regression Theory

- Everything that we did so far was only under the assumption that the errors $e_1, \dots, e_n$ were uncorrelated, had mean zero and variance σ^2.
- But if we want to test hypotheses about or if we want confidence intervals for linear combinations of β, we need distributional assumptions on the errors.
- For example, consider the problem of testing the null hypothesis $H_0 : \beta_1 = 0$ against the alternative hypothesis $H_1 : \beta_1 \neq 0$.

Normal Regression Theory

- Everything that we did so far was only under the assumption that the errors $e_1, \dots, e_n$ were uncorrelated, had mean zero and variance σ^2.
- But if we want to test hypotheses about or if we want confidence intervals for linear combinations of β, we need distributional assumptions on the errors.
- For example, consider the problem of testing the null hypothesis $H_0 : \beta_1 = 0$ against the alternative hypothesis $H_1 : \beta_1 \neq 0$.
- If H_0 were true, this would mean that the first explanatory variable has no role (in the presence of the other explanatory variables) in determining the expected value of the response.

Normal Regression Theory

- Everything that we did so far was only under the assumption that the errors $e_1, \ldots, e_n$ were uncorrelated, had mean zero and variance σ^2.
- But if we want to test hypotheses about or if we want confidence intervals for linear combinations of β, we need distributional assumptions on the errors.
- For example, consider the problem of testing the null hypothesis $H_0 : \beta_1 = 0$ against the alternative hypothesis $H_1 : \beta_1 \neq 0$.
- If H_0 were true, this would mean that the first explanatory variable has no role (in the presence of the other explanatory variables) in determining the expected value of the response.
- An obvious way to test this hypothesis is to look at the value of $|\hat{\beta}_1|$ and then to reject H_0 if $|\hat{\beta}_1| > 0$ is large.

Normal Regression Theory

- Everything that we did so far was only under the assumption that the errors $e_1, \dots, e_n$ were uncorrelated, had mean zero and variance σ^2.
- But if we want to test hypotheses about or if we want confidence intervals for linear combinations of β, we need distributional assumptions on the errors.
- For example, consider the problem of testing the null hypothesis $H_0 : \beta_1 = 0$ against the alternative hypothesis $H_1 : \beta_1 \neq 0$.
- If H_0 were true, this would mean that the first explanatory variable has no role (in the presence of the other explanatory variables) in determining the expected value of the response.
- An obvious way to test this hypothesis is to look at the value of $|\hat{\beta}_1|$ and then to reject H_0 if $|\hat{\beta}_1| > 0$ is large.
- But how large?

Normal Regression Theory

- To answer this question, we need to understand how $\hat{\beta}_1$ is distributed under the null hypothesis H_0.

Normal Regression Theory

- To answer this question, we need to understand how $\hat{\beta}_1$ is distributed under the null hypothesis H_0.
- Such a study requires some distributional assumptions on the errors $e_1, \ldots, e_n$.

Normal Regression Theory

- To answer this question, we need to understand how $\hat{\beta}_1$ is distributed under the null hypothesis H_0.
- Such a study requires some distributional assumptions on the errors $e_1, \ldots, e_n$.
- The most standard assumption on the errors is that $e_1, \ldots, e_n$ are independently distributed according to the normal distribution with mean zero and variance σ^2.

Normal Regression Theory

- To answer this question, we need to understand how $\hat{\beta}_1$ is distributed under the null hypothesis H_0.
- Such a study requires some distributional assumptions on the errors $e_1, \dots, e_n$.
- The most standard assumption on the errors is that $e_1, \dots, e_n$ are independently distributed according to the normal distribution with mean zero and variance σ^2.
- This is written in multivariate normal notation as $e \sim \mathrm{N}(0, \sigma^2 I_n)$.

The Multivariate Normal Distribution

- A random vector $U = (U1, \ldots, U_p)^T$ is said to have the multivariate normal distribution with parameters $\mu($ $p \times 1$ vector) and Σ ($p \times p$ matrix) if the joint density of $U_1, \ldots, U_p$ is given by

$$(2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(-\frac{(u-\mu)^T\Sigma^{-1}(u-\mu)}{2}\right), \quad \forall u \in \mathbb{R}^d.$$

The Multivariate Normal Distribution

- A random vector $U = (U1, \dots, U_p)^T$ is said to have the multivariate normal distribution with parameters $\mu($ $p \times 1$ vector) and Σ ($p \times p$ matrix) if the joint density of $U_1, \dots, U_p$ is given by

$$(2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(-\frac{(u-\mu)^T\Sigma^{-1}(u-\mu)}{2}\right), \quad \forall u \in \mathbb{R}^d.$$

- Here $|\Sigma|$ denotes the determinant of Σ.

The Multivariate Normal Distribution

- A random vector $U = (U1, \ldots, U_p)^T$ is said to have the multivariate normal distribution with parameters $\mu($ $p \times 1$ vector) and Σ ($p \times p$ matrix) if the joint density of $U_1, \ldots, U_p$ is given by

$$(2\pi)^{-p/2}|\Sigma|^{-1/2} \exp\left(-\frac{(u-\mu)^T\Sigma^{-1}(u-\mu)}{2}\right), \quad \forall u \in \mathbb{R}^d.$$

- Here $|\Sigma|$ denotes the determinant of Σ.
- We use the notation $U \sim N_p(\mu, \Sigma)$ to express that U is multivariate normal with parameters μ and Σ.

The Multivariate Normal Distribution

- An important example of the multivariate normal distribution occurs when $U_1, \ldots, U_p$ are independently distributed according to the normal distribution with mean 0 and variance σ^2. In this case, it is easy to show $U = (U_1, \ldots, U_p)^T \sim N_p(0, \sigma^2 I_p)$.

The Multivariate Normal Distribution

The most important properties of the multivariate normal distribution are summarized below:

The Multivariate Normal Distribution

The most important properties of the multivariate normal distribution are summarized below:

- When $p = 1$, this is just the usual normal distribution.

The Multivariate Normal Distribution

The most important properties of the multivariate normal distribution are summarized below:

- When $p = 1$, this is just the usual normal distribution.
- Mean and Variance-Covariance Matrix: $\mathbb{E}U = \mu$ and $Cov(U) = \Sigma$.

The Multivariate Normal Distribution

The most important properties of the multivariate normal distribution are summarized below:

- When $p = 1$, this is just the usual normal distribution.
- Mean and Variance-Covariance Matrix: $\mathbb{E}U = \mu$ and $Cov(U) = \Sigma$.
- Independence of linear functions can be checked by multiplying matrices: Two linear functions AU and BU are independent if and only if $A\Sigma B^T = 0$.

The Multivariate Normal Distribution

The most important properties of the multivariate normal distribution are summarized below:

- When $p = 1$, this is just the usual normal distribution.
- Mean and Variance-Covariance Matrix: $\mathbb{E}U = \mu$ and $Cov(U) = \Sigma$.
- Independence of linear functions can be checked by multiplying matrices: Two linear functions AU and BU are independent if and only if $A\Sigma B^T = 0$. In particular, this means that U_i and U_j are independent if and only if the $(i, j)th$ entry of Σ equals 0.

The Multivariate Normal Distribution

The most important properties of the multivariate normal distribution are summarized below:

- When $p = 1$, this is just the usual normal distribution.
- Mean and Variance-Covariance Matrix: $\mathbb{E}U = \mu$ and $Cov(U) = \Sigma$.
- Independence of linear functions can be checked by multiplying matrices: Two linear functions AU and BU are independent if and only if $A\Sigma B^T = 0$. In particular, this means that U_i and U_j are independent if and only if the $(i, j)th$ entry of Σ equals 0.
- Every linear function is also multivariate normal: $a + AU \sim N(a + A\mu, A\Sigma A^T)$.

The Multivariate Normal Distribution

The most important properties of the multivariate normal distribution are summarized below:

- When $p = 1$, this is just the usual normal distribution.
- Mean and Variance-Covariance Matrix: $\mathbb{E}U = \mu$ and $Cov(U) = \Sigma$.
- Independence of linear functions can be checked by multiplying matrices: Two linear functions AU and BU are independent if and only if $A\Sigma B^T = 0$. In particular, this means that U_i and U_j are independent if and only if the $(i, j)th$ entry of Σ equals 0.
- Every linear function is also multivariate normal: $a + AU \sim N(a + A\mu, A\Sigma A^T)$.
- Suppose $U \sim N_p(\mu, I)$ and A is a $p \times p$ symmetric and idempotent (symmetric means $A^T = A$ and idempotent means $A^2 = A$) matrix. Then $(U - \mu)^T A(U - \mu)$ has the chi-squared distribution with degrees of freedom equal to the rank of A.

Distributions of regression quantities under normality

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \ldots, e_n$ are independent normals with mean 0 and variance σ^2.

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \ldots, e_n$ are independent normals with mean 0 and variance σ^2.
- Under this assumption, we can calculate the distributions of many of the quantities studied so far.

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \dots, e_n$ are independent normals with mean 0 and variance σ^2.
- Under this assumption, we can calculate the distributions of many of the quantities studied so far.
- **Distribution of Y:** Since $Y = X\beta + e$, we have $Y \sim N_n(X\beta, \sigma^2 I_n)$.

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \dots, e_n$ are independent normals with mean 0 and variance σ^2.
- Under this assumption, we can calculate the distributions of many of the quantities studied so far.
- **Distribution of Y:** Since $Y = X\beta + e$, we have $Y \sim N_n(X\beta, \sigma^2 I_n)$.
- **Distribution of $\hat{\beta}$:**

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \ldots, e_n$ are independent normals with mean 0 and variance σ^2.
- Under this assumption, we can calculate the distributions of many of the quantities studied so far.
- **Distribution of Y:** Since $Y = X\beta + e$, we have $Y \sim N_n(X\beta, \sigma^2 I_n)$.
- **Distribution of** $\hat{\beta}$: Because $\hat{\beta} = (X^T X)^{-1} X^T Y$ is a linear function of Y, it has a multivariate normal distribution.

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \ldots, e_n$ are independent normals with mean 0 and variance σ^2.
- Under this assumption, we can calculate the distributions of many of the quantities studied so far.
- **Distribution of Y:** Since $Y = X\beta + e$, we have $Y \sim N_n(X\beta, \sigma^2 I_n)$.
- **Distribution of $\hat{\beta}$:** Because $\hat{\beta} = (X^T X)^{-1} X^T Y$ is a linear function of Y, it has a multivariate normal distribution.
- We already saw that $E\hat{\beta} = \beta$ and $Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$.

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \ldots, e_n$ are independent normals with mean 0 and variance σ^2.
- Under this assumption, we can calculate the distributions of many of the quantities studied so far.
- **Distribution of Y:** Since $Y = X\beta + e$, we have $Y \sim N_n(X\beta, \sigma^2 I_n)$.
- **Distribution of** $\hat{\beta}$: Because $\hat{\beta} = (X^T X)^{-1} X^T Y$ is a linear function of Y, it has a multivariate normal distribution.
- We already saw that $E\hat{\beta} = \beta$ and $Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. Thus $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2 (X^T X)^{-1})$.

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \ldots, e_n$ are independent normals with mean 0 and variance σ^2.
- Under this assumption, we can calculate the distributions of many of the quantities studied so far.
- **Distribution of Y:** Since $Y = X\beta + e$, we have $Y \sim N_n(X\beta, \sigma^2 I_n)$.
- **Distribution of $\hat{\beta}$:** Because $\hat{\beta} = (X^T X)^{-1} X^T Y$ is a linear function of Y, it has a multivariate normal distribution.
- We already saw that $E\hat{\beta} = \beta$ and $Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. Thus $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2 (X^T X)^{-1})$.
- **Distribution of Fitted Values:** $Y = HY$.

Distributions of regression quantities under normality

- We assume that $e \sim N_n(0, \sigma^2 I_n)$.
- Equivalently, $e_1, \ldots, e_n$ are independent normals with mean 0 and variance σ^2.
- Under this assumption, we can calculate the distributions of many of the quantities studied so far.
- **Distribution of Y:** Since $Y = X\beta + e$, we have $Y \sim N_n(X\beta, \sigma^2 I_n)$.
- **Distribution of** $\hat{\beta}$: Because $\hat{\beta} = (X^T X)^{-1} X^T Y$ is a linear function of Y, it has a multivariate normal distribution.
- We already saw that $E\hat{\beta} = \beta$ and $Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. Thus $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2 (X^T X)^{-1})$.
- **Distribution of Fitted Values:** $Y = HY$. Thus $\mathbb{E}\hat{Y} = H\mathbb{E}(Y) = HX\beta = X\beta$. Also $Cov(\hat{Y}) = Cov(HY) = \sigma^2 H$. Therefore $\hat{Y} \sim N_n(X, \sigma^2 H)$.

Distributions of regression quantities under normality

Distributions of regression quantities under normality

- **Distribution of Residuals:** $\hat{e} = (I - H)Y$.

Distributions of regression quantities under normality

- **Distribution of Residuals:** $\hat{e} = (I - H)Y$. We saw that $\mathbb{E}\hat{e} = 0$ and $Cov(\hat{e}) = \sigma^2(I - H)$.

Distributions of regression quantities under normality

- **Distribution of Residuals:** $\hat{e} = (I - H)Y$. We saw that $\mathbb{E}\hat{e} = 0$ and $Cov(\hat{e}) = \sigma^2(I - H)$. Therefore $\hat{e} \sim N_n(0; \sigma^2(I - H))$.

Independence of residuals and $\hat{\beta}$.

Independence of residuals and $\hat{\beta}$.

- Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T = 0$.

Independence of residuals and $\hat{\beta}$.

- Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T = 0$.
- This can be used to verify that $\hat{\beta} = (X^TX)^{-1}X^TY$ and $\hat{e} = (I - H)Y$ are independent.

Independence of residuals and $\hat{\beta}$.

- Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T = 0$.
- This can be used to verify that $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{e} = (I - H)Y$ are independent.
- To see this, observe that both are linear functions of $Y \sim N_n(X\beta; \sigma^2 I_n)$.

Independence of residuals and $\hat{\beta}$.

- Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T = 0$.
- This can be used to verify that $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{e} = (I - H)Y$ are independent.
- To see this, observe that both are linear functions of $Y \sim N_n(X\beta; \sigma^2 I_n)$. Thus if $A = (XTX)^{-1} X^T$, $B = (I - H)$ and $\Sigma = \sigma^2 I_n$, then

Independence of residuals and $\hat{\beta}$.

- Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T = 0$.
- This can be used to verify that $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{e} = (I - H)Y$ are independent.
- To see this, observe that both are linear functions of $Y \sim N_n(X\beta; \sigma^2 I_n)$. Thus if $A = (XTX)^{-1}X^T$, $B = (I - H)$ and $\Sigma = \sigma^2 I_n$, then

$$A\Sigma B^T \ = \ (XTX)^{-1}X^T(\sigma^2 I_n)(I{-}H) \ =$$

Independence of residuals and $\hat{\beta}$.

- Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T = 0$.
- This can be used to verify that $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{e} = (I - H)Y$ are independent.
- To see this, observe that both are linear functions of $Y \sim N_n(X\beta; \sigma^2 I_n)$. Thus if $A = (XTX)^{-1}X^T$, $B = (I - H)$ and $\Sigma = \sigma^2 I_n$, then

$$A\Sigma B^T = (XTX)^{-1}X^T(\sigma^2 I_n)(I-H) = \sigma^2(X^T X)^{-1}(X^T - X^T H) = 0$$

Independence of residuals and $\hat{\beta}$.

- Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T = 0$.
- This can be used to verify that $\hat{\beta} = (X^T X)^{-1} X^T Y$ and $\hat{e} = (I - H)Y$ are independent.
- To see this, observe that both are linear functions of $Y \sim N_n(X\beta; \sigma^2 I_n)$. Thus if $A = (XTX)^{-1} X^T$, $B = (I - H)$ and $\Sigma = \sigma^2 I_n$, then

  $$A\Sigma B^T = (XTX)^{-1} X^T (\sigma^2 I_n)(I-H) = \sigma^2 (X^T X)^{-1} (X^T - X^T H) = 0$$

- Because $X^T H = (HX)^T = X^T$, we conclude that $\hat{\beta}$ and $\hat{e}$ are independent.

Independence of residuals and $\hat{\beta}$.

- Recall that if $U \sim N_p(\mu, \Sigma)$, then AU and BU are independent if and only if $A\Sigma B^T = 0$.
- This can be used to verify that $\hat{\beta} = (X^TX)^{-1}X^TY$ and $\hat{e} = (I - H)Y$ are independent.
- To see this, observe that both are linear functions of

$Y \sim N_n(X\beta; \sigma^2 I_n)$. Thus if $A = (XTX)^{-1}X^T$, $B = (I - H)$ and $\Sigma = \sigma^2 I_n$, then

$$A\Sigma B^T = (XTX)^{-1}X^T(\sigma^2 I_n)(I-H) = \sigma^2(X^TX)^{-1}(X^T - X^TH) = 0.$$

- Because $X^TH = (HX)^T = X^T$, we conclude that $\hat{\beta}$ and $\hat{e}$ are independent.
- Also check that $\hat{Y}$ and $\hat{e}$ are independent.

Distribution of RSS

Distribution of RSS

- Recall

$$RSS = \hat{e}^T \hat{e} = Y^T (I - H) Y = e^T (I - H) e.$$

Distribution of RSS

- Recall

$$RSS = \hat{e}^T\hat{e} = Y^T(I - H)Y = e^T(I - H)e.$$

- So

$$\frac{RSS}{\sigma^2} = \left(\frac{e}{\sigma}\right)^T (I - H) \left(\frac{e}{\sigma}\right).$$

Distribution of RSS

- Recall
$$RSS = \hat{e}^T\hat{e} = Y^T(I - H)Y = e^T(I - H)e.$$
- So
$$\frac{RSS}{\sigma^2} = \left(\frac{e}{\sigma}\right)^T (I - H)\left(\frac{e}{\sigma}\right).$$
- Because $e/\sigma \sim N_n(0, I)$ and $I - H$ is symmetric and idempotent with rank $n - p - 1$, we have
$$\frac{RSS}{\sigma^2} \sim \chi^2_{n-p-1}.$$

Hypothesis Testing: How to test $H_0 : \beta_j = 0$.

Hypothesis Testing: How to test $H_0 : \beta_j = 0$.

The are two equivalent ways of testing this hypothesis. One is the t-test which we shall study today. The other is the F-test which we shall look at in the next class.

t-test

t-test

- It is natural to base the test on the value of $\hat{\beta}_j$ i.e., reject if $|\hat{\beta}_j|$ is large. How large?

t-test

- It is natural to base the test on the value of $\hat{\beta}_j$ i.e., reject if $|\hat{\beta}_j|$ is large. How large?
- To answer this, we need to look at the distribution of $\hat{\beta}_j$ under H_0 (called the null distribution).

t-test

- It is natural to base the test on the value of $\hat{\beta}_j$ i.e., reject if $|\hat{\beta}_j|$ is large. How large?
- To answer this, we need to look at the distribution of $\hat{\beta}_j$ under H_0 (called the null distribution).
- Under normality of the errors, we have seen that $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2 (X^T X)^{-1})$.

t-test

- It is natural to base the test on the value of $\hat{\beta}_j$ i.e., reject if $|\hat{\beta}_j|$ is large. How large?
- To answer this, we need to look at the distribution of $\hat{\beta}_j$ under H_0 (called the null distribution).
- Under normality of the errors, we have seen that $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^T X)^{-1})$. In other words,

$$\hat{\beta}_j \sim N(\beta_j, v_j \sigma^2)$$

  where v_j is the j-th entry of the diagonal of the matrix $(X^T X)^{-1}$.

t-test

- It is natural to base the test on the value of $\hat{\beta}_j$ i.e., reject if $|\hat{\beta}_j|$ is large. How large?
- To answer this, we need to look at the distribution of $\hat{\beta}_j$ under H_0 (called the null distribution).
- Under normality of the errors, we have seen that $\hat{\beta} \sim N_{p+1}(\beta, \sigma^2(X^T X)^{-1})$. In other words,

$$\hat{\beta}_j \sim N(\beta_j, v_j \sigma^2)$$

  where v_j is the j-th entry of the diagonal of the matrix $(X^T X)^{-1}$.
- Under the null hypothesis, when $\beta_j = 0$, we thus have

$$\frac{\hat{\beta}_j}{\sigma \sqrt{v_j}} \sim N(0, 1).$$

t-test

- This can be used to construct a test but the problem is that σ is unknown.

t-test

- This can be used to construct a test but the problem is that σ is unknown. One therefore replaces it by the estimate $\hat{\sigma}$ to construct the test statistic:

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma} v_j} = \frac{\hat{\beta}_j/(\sigma\sqrt{v_j})}{\hat{\sigma}/\sigma},$$

t-test

- This can be used to construct a test but the problem is that σ is unknown. One therefore replaces it by the estimate $\hat{\sigma}$ to construct the test statistic:

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma} v_j} = \frac{\hat{\beta}_j/(\sigma\sqrt{v_j})}{\hat{\sigma}/\sigma},$$

- Now the numerator here is $N(0, 1)$.

t-test

- This can be used to construct a test but the problem is that σ is unknown. One therefore replaces it by the estimate $\hat{\sigma}$ to construct the test statistic:

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma} v_j} = \frac{\hat{\beta}_j/(\sigma\sqrt{v_j})}{\hat{\sigma}/\sigma},$$

- Now the numerator here is $N(0,1)$.
- The denominator is

$$\hat{\sigma}/\sigma = \sqrt{\frac{RSS}{\sigma(n-p-1)}} = \sqrt{\frac{\chi^2_{n-p-1}}{n-p-1}}$$

.

t-test

- This can be used to construct a test but the problem is that σ is unknown. One therefore replaces it by the estimate $\hat{\sigma}$ to construct the test statistic:

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\hat{\sigma}\sqrt{v_j}} = \frac{\hat{\beta}_j/(\sigma\sqrt{v_j})}{\hat{\sigma}/\sigma},$$

- Now the numerator here is $N(0,1)$.
- The denominator is

$$\hat{\sigma}/\sigma = \sqrt{\frac{RSS}{\sigma(n-p-1)}} = \sqrt{\frac{\chi^2_{n-p-1}}{n-p-1}}$$

.
Moreover, the numerator and the denominator are independent.

t-test

- Therefore we get

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} \sim t_{n-p-1}.$$

where t_{n-p-1} denotes the t-distribution with $n - p - 1$ degrees of freedom.

t-test

- Therefore we get

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} \sim t_{n-p-1}.$$

  where t_{n-p-1} denotes the t-distribution with $n-p-1$ degrees of freedom.

- p-value for testing $H_0 : \beta_j = 0$ can be got by

$$\mathbb{P}\left(|t_{n-p-1}| > \frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)}\right).$$

t-test

- Therefore we get

$$\frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} \sim t_{n-p-1}.$$

  where t_{n-p-1} denotes the t-distribution with $n-p-1$ degrees of freedom.
- p-value for testing $H_0 : \beta_j = 0$ can be got by

$$\mathbb{P}\left(|t_{n-p-1}| > \frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)}\right).$$

- Note that when $n-p-1$ is large, the t-distribution is almost the same as a standard normal distribution.