Exercises for Linear Statistical Models

1. Suppose we have data $x_1, \ldots, x_n$ and $y_i, \ldots, y_n$, which we model as

$$Y_i = b_0 + b_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2).$$

   (a) Write down the formula for the sum of squares criterion.

   (b) Take the partial derivatives of the sum of squares criterion to derive the normal equations.

   (c) Solve the normal equations to show that the least squares estimates are

$$\widehat{b}_1 = \frac{\sum_{i=1}^n (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^n (x_i - \overline{x})^2}$$
$$\widehat{b}_0 = \overline{y} - \widehat{b}_1 \overline{x}$$

2. In terms of the estimates $\widehat{b}_0$ and $\widehat{b}_1$, write down the formulas for the fitted values, the residuals, and $\widehat{\sigma}^2$ (the variance estimate).

3. Explain the difference between an estimate and an estimator.

   Answer:

   An estimate is a function of the data, for example $\widehat{b} = \overline{y}$, whereas an estimator is a random variable, a function of the random variables used to model the data, for example $\widehat{B} = \overline{Y}$.

4. Write down how you would explain all of the assumptions of the simple linear model for $y_1, \ldots, y_n$ given $x_1, \ldots, x_n$ to your grandmother (who was a chemical engineer).

5. Suppose we have data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. The quantity sxy is given to the left of the equals sign and is always equal to **exactly one** of the three expressions to the right of the equals sign. Circle the correct expression and show why sxy is equal to it.

$$\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) \quad = \quad \sum_{i=1}^n (x_i - \overline{x}) y_i \qquad \sum_{i=1}^n (x_i - \overline{x})\overline{y} \qquad \sum_{i=1}^n \overline{x}(y_i - x_i)$$

   Answer:

   The first one is the correct answer. To see why, we subtract zero in a clever way

and regroup the terms:

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\overline{x})y_i &= \sum_{i=1}^{n}(x_i-\overline{x})y_i - \overline{y}\sum_{i=1}^{n}(x_i-\overline{x}) \\ &= \sum_{i=1}^{n}(x_i-\overline{x})y_i - \sum_{i=1}^{n}(x_i-\overline{x})\overline{y} \\ &= \sum_{i=1}^{n}\left[(x_i-\overline{x})y_i - (x_i-\overline{x})\overline{y}\right] \\ &= \sum_{i=1}^{n}\left[(x_i-\overline{x})(y_i-\overline{y})\right]\end{aligned}$$

6. Derive a formula for $\mathrm{Cov}(\widehat{B}_0, \widehat{B}_1)$

   Answer:

$$\begin{aligned}\mathrm{Cov}(\widehat{B}_0, \widehat{B}_1) &= \mathrm{Cov}(\overline{Y} - \widehat{B}_1\overline{x}, \widehat{B}_1) \\ &= \mathrm{Cov}(\overline{Y}, \widehat{B}_1) - \overline{x}\mathrm{Cov}(\widehat{B}_1, \widehat{B}_1) \\ &= \mathrm{Cov}(\overline{Y}, \widehat{B}_1) - \overline{x}\mathrm{Var}(\widehat{B}_1) \\ &= \mathrm{Cov}(\overline{Y}, \widehat{B}_1) - \overline{x}\sigma^2/sxx\end{aligned}$$

   The last step uses the formula for $\mathrm{Var}(\widehat{B}_1)$. Now we must consider the first term.

$$\begin{aligned}\mathrm{Cov}(\overline{Y}, \widehat{B}_1) &= \mathrm{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}Y_i, \frac{1}{sxx}\sum_{i=1}^{n}(x_i-\overline{x})Y_i\right) \\ &= \frac{1}{n}\frac{1}{sxx}\sum_{i=1}^{n}(x_i-\overline{x})\mathrm{Var}(Y_i) \\ &= \frac{1}{n}\frac{1}{sxx}\sum_{i=1}^{n}(x_i-\overline{x})\sigma^2 \\ &= 0\end{aligned}$$

   The second-to-last step uses the fact that the Y_i's are indepdent. So the answer is $-\overline{x}\sigma^2/sxx$.

7. Use the formulas for the variances and covariances of $\widehat{B}_0$ and $\widehat{B}_1$ to derive a simplified expression for the prediction variance $\mathrm{Var}(\widehat{B}_0 + \widehat{B}_1 x_i)$.

8. Suppose we have data $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. Which of the following equations

qualify as statistical models for $y_1, \ldots, y_n$?

$$
\begin{aligned}
&Y_i = b_0 + b_1 x_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2) \\
&Y_i = b_0 + b_1 x_i \\
&Y_i = e^{\cos(x_i) + \varepsilon_i}, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2) \\
&0 \leq Y_i \leq 10 \\
&Y_i \overset{ind}{\sim} \text{Uniform}(0, 10)
\end{aligned}
$$

Answer:

Yes to the first one. It is the simple linear model that fully expresses Y_i as a random variable and explains the assumptions underlying the random term ε_i.

No to the second because because the right side of the equation is not random.

Yes to the third one. Even though this is a nonlinear model, it still specifies Y_i as a random variable and delineates all of the assumptions. This happens to be a linear model for $\log(Y_i)$, a transformation of the response.

No to the fourth one because it merely says that Y_i is between 0 and 10, without specifying the distribution on that interval.

Yes to the fifth one, because it specifies the distribution–uniform–on the interval 0 to 10.

9. Under the simple linear model, there is a formula for the least squares estimators of the regression coefficients. Which property of the estimators allows us to conclude that the estimators are normally distributed?

10. Select all that are correct: A confidence interval for the slope b_1 . . .

    (a) contains the slope values that we rejected using a hypothesis test

    (b) contains the slope values that we failed to reject using a hypothesis test

    (c) is a set of plausible values of the slope given the data

    (d) has endpoints that are random variables

    Answer:

    (a) is wrong. It contains values that we fail to reject.

    (b) is right

    (c) is right, this is good interpretation of the definition of a confidence interval

    (d) is wrong. We *model* the endpoints as relizations of random variables. The endpoints themselves are numbers that we can write down and therefore are non-random.

11. Name two desirable properties of estimators, and explain why they are desirable.

12. Let

$$\boldsymbol{Y} = X\boldsymbol{b} + \boldsymbol{\varepsilon}, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2 I)$$

where $\boldsymbol{Y}$ is $n \times 1$ and X is $n \times p + 1$.

What are the dimensions of $\boldsymbol{b}$ and $\boldsymbol{\varepsilon}$?

What is the expectation of $\boldsymbol{Y}$?

What is the expectation of $X^T\boldsymbol{Y}$?

What is the expectation of $\widehat{\boldsymbol{B}} = (X^TX)^{-1}X^T\boldsymbol{Y}$?

What is the covariance matrix for ε?

What is the covariance matrix for $\widehat{\boldsymbol{B}}$?

Answer:

$\boldsymbol{b}$ must be $p + 1 \times 1$ in order that $X\boldsymbol{b}$ is $n \times 1$

$\boldsymbol{\varepsilon}$ must be $n \times 1$ to match the dimensions of $X\boldsymbol{b}$

$$E(\boldsymbol{Y}) = E(X\boldsymbol{b} + \boldsymbol{\varepsilon}) = \boldsymbol{X}\boldsymbol{b} + E(\boldsymbol{\varepsilon}) = \boldsymbol{X}\boldsymbol{b}$$

$$E(X^T\boldsymbol{Y}) = X^TE(\boldsymbol{Y}) = X^TX\boldsymbol{b}$$

$$E((X^TX)^{-1}X^T\boldsymbol{Y}) = (X^TX)^{-1}E(X^T\boldsymbol{Y}) = (X^TX)^{-1}X^TX\boldsymbol{b} = \boldsymbol{b}$$

$\varepsilon_1, \ldots, \varepsilon_n$ independent implies that $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$ if $i \neq j$. When $i = j$, $\text{Cov}(\varepsilon_i, \varepsilon_i) = \text{Var}(\varepsilon_i)$, which means that $\text{Cov}(\boldsymbol{\varepsilon}_i) = \sigma^2 I_n$

$$\begin{aligned}
\text{Cov}(\widehat{\boldsymbol{B}}) &= \text{Cov}((X^TX)^{-1}X^T\boldsymbol{Y}) \\
&= (X^TX)^{-1}X^T\text{Cov}(\boldsymbol{Y})X(X^TX)^{-1} \\
&= (X^TX)^{-1}X^T(\sigma^2 I)X(X^TX)^{-1} \\
&= \sigma^2(X^TX)^{-1}X^TX(X^TX)^{-1} = \sigma^2(X^TX)^{-1}
\end{aligned}$$

13. Recall the simple linear model

$$Y_i = b_0 + b_1x_i + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2).$$

Write down the design matrix X.

Calculate the entries of X^TX

Calculate the entries of $(X^TX)^{-1}$

Calculate $X^T\boldsymbol{y}$.

Show that $\widehat{\boldsymbol{b}} = (X^TX)^{-1}X^T\boldsymbol{y}$ produces the same least squares estimates that we originally derived. You'll have to use the formula for the inverse of a 2×2 matrix.

Answer:

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n 1 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\overline{x} \\ n\overline{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

$$(X^T X)^{-1} = \begin{bmatrix} n & n\overline{x} \\ n\overline{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} = \frac{1}{n\sum_{i=1}^n x_i^2 - n^2(\overline{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\overline{x} \\ -n\overline{x} & n \end{bmatrix}$$

$$X^T \boldsymbol{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix} = \begin{bmatrix} n\overline{y} \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

$$\begin{aligned} (X^T X)^{-1} X^T \boldsymbol{y} &= \frac{1}{n\sum_{i=1}^n x_i^2 - n^2(\overline{x})^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\overline{x} \\ -n\overline{x} & n \end{bmatrix} \begin{bmatrix} n\overline{y} \\ \sum_{i=1}^n x_i y_i \end{bmatrix} \\ &= \frac{1}{n\sum_{i=1}^n x_i^2 - n^2(\overline{x})^2} \begin{bmatrix} n\overline{y}\sum_{i=1}^n x_i^2 - n\overline{x}\sum_{i=1}^n x_i y_i \\ -n^2\overline{x}\,\overline{y} + n\sum_{i=1}^n x_i y_i \end{bmatrix} \end{aligned}$$

Let's look at the second entry:

$$\begin{aligned} &= \frac{n\sum_{i=1}^n x_i y_i - n^2\overline{x}\,\overline{y}}{n\sum_{i=1}^n x_i^2 - n^2(\overline{x})^2} \\ &= \frac{\sum_{i=1}^n x_i y_i - n\overline{x}\,\overline{y}}{\sum_{i=1}^n x_i^2 - n(\overline{x})^2} \\ &= \frac{sxy}{sxx} \end{aligned}$$

where the last line applies after applying "the trick" a few times. This is our usual formula for $\widehat{b}_1$

The first entry is a different formula for $\widehat{b}_0$ than you have seen, because we have written $\widehat{b}_0$ in terms of $\widehat{b}_1$. But if you expand $\widehat{b}_0 = \overline{y} - \overline{x}\widehat{b}_1$, you will get the first entry.

14. The residual sum of squares criterion for the multiple linear model is

$$rss(\boldsymbol{b}^*) = \sum_{i=1}^{n}(y_i - \sum_{j=0}^{p} b_j^* x_{ij})^2$$

Derive the kth normal equation by differentiating the residual sum of squares.

15. Show that the set of $p+1$ normal equations can be written as $X^T\boldsymbol{y} = X^T X\widehat{\boldsymbol{b}}$

16. Let

$$X = \begin{bmatrix}\boldsymbol{x}_0 & \cdots & \boldsymbol{x}_p\end{bmatrix}$$

And suppose for this problem that $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_p$ are orthonormal, which means that $\boldsymbol{x}_j^T\boldsymbol{x}_k = 0$ if $j \neq k$ and 1 if $j = k$.

Show that the least squares estimates of $\boldsymbol{b}$ are $\widehat{b}_j = \boldsymbol{x}_j^T\boldsymbol{y}$.

Answer:

First, let's write down $X^T X$

$$X^T X = \begin{bmatrix}\boldsymbol{x}_0^T \\ \vdots \\ \boldsymbol{x}_p^T\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_0 & \cdots & \boldsymbol{x}_p\end{bmatrix} = \begin{bmatrix}\boldsymbol{x}_0^T\boldsymbol{x}_0 & \boldsymbol{x}_0^T\boldsymbol{x}_1 & \boldsymbol{x}_0^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_0^T\boldsymbol{x}_p \\ \boldsymbol{x}_1^T\boldsymbol{x}_0 & \boldsymbol{x}_1^T\boldsymbol{x}_1 & \boldsymbol{x}_1^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_1^T\boldsymbol{x}^p \\ \boldsymbol{x}_2^T\boldsymbol{x}_0 & \boldsymbol{x}_2^T\boldsymbol{x}_1 & \boldsymbol{x}_2^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_2^T\boldsymbol{x}^p \\ \vdots & & & & \\ \boldsymbol{x}_p^T\boldsymbol{x}_0 & \boldsymbol{x}_p^T\boldsymbol{x}_1 & \boldsymbol{x}_p^T\boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p^T\boldsymbol{x}_p\end{bmatrix}$$

$$= \begin{bmatrix}1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \cdots & 1\end{bmatrix} = I$$

This means that

$$(X^T X)^{-1}X^T\boldsymbol{y} = X^T\boldsymbol{y} = \begin{bmatrix}\boldsymbol{x}_0^T \\ \vdots \\ \boldsymbol{x}_p^T\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}\boldsymbol{x}_0^T\boldsymbol{y} \\ \vdots \\ \boldsymbol{x}_p^T\boldsymbol{y}\end{bmatrix}$$

In general, the columns of X will not be orthonormal, or even orthogonal. This result applies only in the case when the columns of X are orthonormal.

17. What are the two defining properties of projection matrices?

Answer:

Symmetric ($P = P^T$) and idempotent ($PP = P$)

18. Show that $X(X^T X)^{-1}X^T$ is a projection matrix.

19. Suppose that

$$U = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 0 \end{bmatrix} \text{ and } M = UU^T$$

Calculate the entries of M and show that M is a projection matrix.

20. Consider

$$P = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Show that P is a projection matrix.

Describe the space that P projects onto.

21. Show that the fitted values are in the column space of X. In other words, show that the fitted values can be written as a linear combination of the columns of X, $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_p$.

Answer:

The fitted values are $\widehat{\boldsymbol{y}} = X(X^TX)^{-1}X^T\boldsymbol{y}$. We can rewrite this as $\widehat{\boldsymbol{y}} = X\widehat{\boldsymbol{b}}$, which is equal to $\widehat{b}_0\boldsymbol{x}_0 + \cdots + \widehat{b}_p\boldsymbol{x}_p$, which is a linear combination of the columns of X. In fact, it's not just any linear combination of the columns of X; it's the particular linear combination whose coefficients are the regression coefficients. Since we say that the coefficients of the linear combination are the coordinates of the vector with respect to the basis, the regression coefficient estimates are the coordinates of the fitted values with respect to the basis defined by the columns of X.

22. Let $\boldsymbol{b} = (b_0, b_1, b_2)$ and $\widehat{\boldsymbol{B}}$ be the least squares estimator for $\boldsymbol{b}$. Further, let $M = \sigma^2(X^TX)^{-1}$, and M_{ij} be the (i, j) entry of M.

What is the expected value of $\begin{bmatrix} 0 & 1 & -1 \end{bmatrix}\widehat{\boldsymbol{B}}$?

What is the variance of $\begin{bmatrix} 0 & 1 & -1 \end{bmatrix}\widehat{\boldsymbol{B}}$?

23. Let

$$\boldsymbol{Y} = X\boldsymbol{b} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 I)$$

Recall that if $\boldsymbol{Z} \sim N(\boldsymbol{\mu}, \Sigma)$, then $\boldsymbol{a} + M\boldsymbol{Z} \sim N(\boldsymbol{a} + M\boldsymbol{\mu}, M\Sigma M^T)$.

What is the distribution of $\boldsymbol{Y}$?

What is the distribution of $X^T\boldsymbol{Y}$?

What is the distribution of $\widehat{\boldsymbol{B}} = (X^TX)^{-1}X^T\boldsymbol{Y}$?

Does $\boldsymbol{Y}^T\boldsymbol{Y}$ follow a normal distribution? Why or why not?

Answer:

Because $\boldsymbol{Y}$ is a linear transformation of $\boldsymbol{\varepsilon}$, and $\boldsymbol{\varepsilon}$ is MVN,

$$\boldsymbol{Y} \sim N(X\boldsymbol{b}, \sigma^2 I)$$

Because $X^T\boldsymbol{Y}$ is a linear transformation of $\boldsymbol{Y}$, and $\boldsymbol{Y}$ is MVN,

$$X^T\boldsymbol{Y} \sim N(X^TX\boldsymbol{b}, \sigma^2 X^TX)$$

Because $\widehat{\boldsymbol{B}}$ is a linear transformatino of $\boldsymbol{Y}$, and $\boldsymbol{Y}$ is MVN,

$$(X^TX)^{-1}X^T\boldsymbol{Y} \sim N((X^TX)^{-1}X^TX\boldsymbol{b}, \sigma^2(X^TX)^{-1}(X^TX)(X^TX)^{-1}) = N(\boldsymbol{b}, \sigma^2(X^TX)^{-1})$$

24. The following is from a longitudinal study measuring heights, weights, etc. of a group of girls. Height2 = height at age 2, Height9 = height at age 9, Height18 = height at age 18, and LegCirc9 = leg circumference at age 9.

Below is output from a regression of Height18 on Height2, Height9, and LegCirc9

```
Call: lm(formula = Height18 ~ Height2 + Height9 + LegCirc9 )

                Estimate Std. Error t value Pr(>|t|)
(Intercept)      35.6880    11.2669   3.167  0.00233
Height2           0.2945     0.1827   1.612  0.11163
Height9           0.9016     0.1195   7.547 1.71e-10
LegCirc9         -0.5986     0.2089  -2.865  0.00559

Residual standard error: 3.404 on 66 degrees of freedom
Multiple R-squared:  0.6997,Adjusted R-squared:  0.6861
F-statistic: 51.27 on 3 and 66 DF,  p-value: < 2.2e-16
```

(a) How many individual girls were in this dataset?

Answer:

The degrees of freedom is 66. Since there are 3 covariates in the model + 1 intercept, the degrees of freedom was calculated as $n - 4 = 66$, which means that $n = 70$.

(b) The model fit to the height data was

$$Y_i = b_0 + b_1x_{i1} + b_2x_{i2} + b_3x_{i3} + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2),$$

where the covariates are listed in the same order as in the output above. What is the estimate of σ?

Answer:

In the output, $\widehat{\sigma}$ is "Residual standard error", which is listed as 3.404.

(c) Give an interpretation for the result of the t-test for the Height2 regression coefficient. Why does this make sense for height data?

Answer:

We do not have sufficient evidence to reject the hypothesis that b_1 is zero. In other words, 0 is a plausible value for b_1 given the data we have observed. It is not surprising that we don't see evidence for height at age 2 as being important after accounting for height at age 9, since information about height at age 9 likely makes information about height at age 2 irrelevant.

(d) Can we conclude that height at age 2 is not important for predicting height at age 18?

Answer:

No, from this model, we can only conclude that we do not have sufficient evidence to say that height at age 2 is important, after accounting for height at age 9 and leg circumference at age 9.

(e) What is the residual sum of squares for this model?

Answer:

We know that $\widehat{\sigma}^2 = rss/df$, so $rss = \widehat{\sigma}^2 df = 3.404^2 * 66 = 764.7563$.

(f) What is syy for this dataset, and what is the standard deviation of the response?

Answer:

We know that $R^2 = 1 - rss/syy$, so $syy = rss/(1 - R^2) = 764.7563/(1 - .6997) = 2546.641$

25. Show that $X^T\widehat{\boldsymbol{e}} = \mathbf{0}$, that is, the observed residual vector is orthogonal to every column of the design matrix.

26. Let $T = Z/\sqrt{W/m}$. T has a t distribution with m degrees of freedom if what 3 things are true?

Answer:

1. $Z \sim N(0,1)$, 2. $W \sim \chi^2_m$, and 3. Z independent of W.

27. Suppose we have a hypothesis for the multiple linear model that can be written as

$$H_0 : \boldsymbol{c}^T\boldsymbol{b} = a$$

where $\boldsymbol{c}$ is a $p+1$ by 1 vector, a is a known (hypothesized) scalar, and, as usual, $\boldsymbol{b}$ is the vector of regression coefficients. For example, the hypothesis $H_0 : b_2 - b_1 = 0$ can be written in this form.

(a) Write down the t statistic for this test in terms of $\widehat{\boldsymbol{b}}$, X, and $\widehat{\sigma}^2$.

Answer:

$$t = (\boldsymbol{c}^T\widehat{\boldsymbol{b}} - a)/\sqrt{\widehat{\sigma}^2\boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}}$$

(b) Show that the random version of the t statistic has a T distribution.

Answer:

We can write the random version of the statistic as

$$T = \frac{\boldsymbol{c}^T\widehat{\boldsymbol{B}} - a}{\sqrt{\sigma^2\boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}}} \frac{1}{\sqrt{\frac{\widehat{\sigma}^2(n-p-1)}{\sigma^2(n-p-1)}}}$$

The first term is $N(0,1)$ because under the null hypothesis, it is a mean-zero normal divided by its standard deviation. The second term is a Chi-squared random variable with $n-p-1$ degrees of freedom divided by $n-p-1$. And the first and second terms are independent because $\widehat{\boldsymbol{B}}$ is independent of $\widehat{\sigma}^2$, due to the fact that $\widehat{\sigma}^2$ is a function of the (random version) of the residuals, and the residuals are independent $\widehat{\boldsymbol{B}}$.