STAC67: Regression Analysis

Lecture 13

Thibault Randrianarisoa

October 21, 2026

University of Toronto Scarborough

Review of Lecture 12

Tests and intervals with several predictors

With normal errors, each coefficient is tested and estimated with the t distribution:

$$\frac{\hat{\beta}_k - \beta_k}{s\{\hat{\beta}_k\}} \sim t(n-p), \qquad s^2\{\hat{\beta}_k\} = \text{entry } k \text{ of the diagonal of } MSE\,(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

The test of $\beta_k = 0$ is about X_k given the other predictors.

At a new point $\boldsymbol{X}_h$, with $s^2\{\hat{Y}_h\} = MSE\,\boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h$ and $s^2\{\text{pred}\} = MSE + s^2\{\hat{Y}_h\}$:

- mean response $E(Y_h)$: $\hat{Y}_h \pm t(1-\alpha/2;\, n-p)\, s\{\hat{Y}_h\}$;
- new observation $Y_{h(\text{new})}$: $\hat{Y}_h \pm t(1-\alpha/2;\, n-p)\, s\{\text{pred}\}$.

Beware of hidden extrapolation: a large $\boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h$ signals a point unlike the data, even when each predictor is within its range.

The ANOVA table and the overall F test

$SSR = \|\hat{y} - \bar{y}1\|_2^2$ / $SSE = \|y - \hat{y}\|_2^2$ / $SST = \|y - \bar{y}1\|_2^2$

Source	SS	df	MS	F^*
Regression	SSR	$p-1$	$MSR = SSR/(p-1)$	MSR/MSE
Error	SSE	$n-p$	$MSE = SSE/(n-p)$	
Total	SST	$n-1$		

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{p-1} = 0$ against H_a: at least one $\beta_k \neq 0$.

With normal errors, under H_0, $F^* = MSR/MSE \sim F(p-1,\ n-p)$. Reject H_0 at level α when $F^* > F(1-\alpha;\ p-1,\ n-p)$.

⚠ Rejecting says that the predictors, *taken together*, explain some variation. It does not say which ones, nor that the model is good.

The acceleration data

Model $\texttt{acc} = \beta_0 + \beta_1 \texttt{wt} + \beta_2 \texttt{hp} + \epsilon$, $n = 390$, $p = 3$:

Source	SS	df	MS	F^*
Regression	1794.7	2	897.3	295.7
Error	1174.3	387	3.034	
Total	2969.0	389		

$F^* = 295.7 > F(0.95;\ 2,\ 387) = 3.02$: we reject H_0. Weight and horsepower together are related to acceleration.

In R: `anova(lm(acc ~ 1, data = autompg), fit2)` compares the model with the model that has no predictor.

R^2, adjusted R^2, and F

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}, \qquad R^2_a = 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{n-1}{n-p}\frac{SSE}{SST}.$$

R^2 never decreases when a predictor is added, even pure noise; R^2_a charges for each parameter and can decrease. Acceleration data, then with a column of noise added:

$$R^2\text{: } 0.60448 \to 0.60452, \qquad R^2_a\text{: } 0.6024 \to 0.6014.$$

Dividing the numerator and the denominator of F^* by SST:

$$F^* = \frac{SSR/(p-1)}{SSE/(n-p)} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)}: \quad \text{here } \frac{0.6045/2}{(1-0.6045)/387} = 295.7.$$

With one predictor ($p = 2$), $F^* = (t^*)^2 = (n-2)\,R^2/(1-R^2)$.

Reading `summary(fit2)`

$Pr(T > t\text{-value})$ with $t(n-p)$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 18.4695924  0.3265144   56.57   <2e-16 ***
wt           0.0022919  0.0002064   11.10   <2e-16 ***
hp          -0.0932587  0.0045541  -20.48   <2e-16 ***
Residual standard error: 1.742 on 387 degrees of freedom
Multiple R-squared:  0.6045,    Adjusted R-squared:  0.6024
F-statistic: 295.7 on 2 and 387 DF,  p-value: < 2.2e-16
```

- `Estimate`, `Std. Error`: $\hat{\beta}_k$ and $s\{\hat{\beta}_k\}$, the square root of entry k of the diagonal of $MSE\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
- `t value` $= \hat{\beta}_k / s\{\hat{\beta}_k\}$, with its p-value from $t(n-p)$.
- `Residual standard error` $= \sqrt{MSE}$, on $n-p$ degrees of freedom.
- The last line is the overall F test, on $p-1$ and $n-p$ degrees of freedom.

Why the F test works

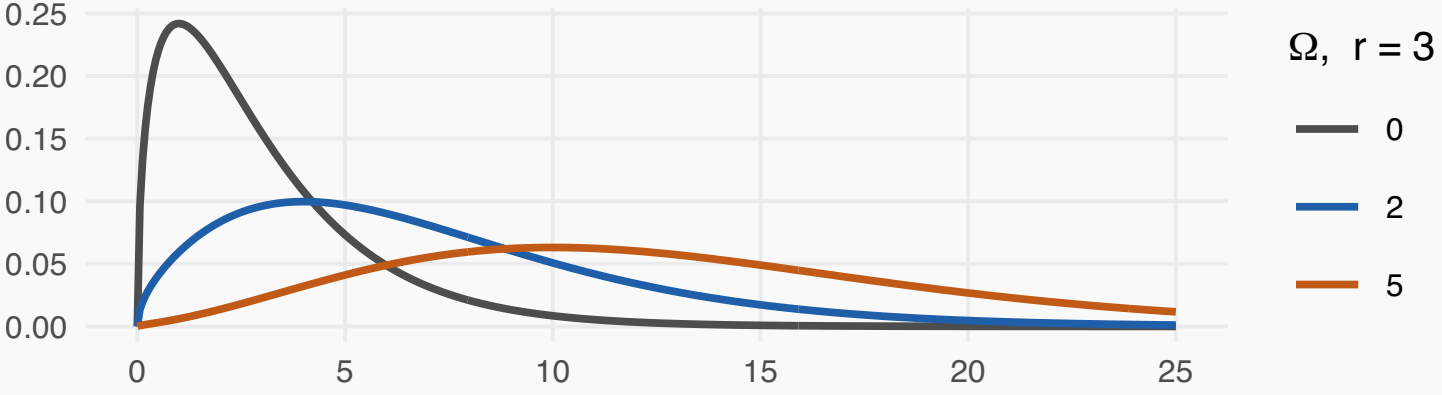
The noncentral χ^2 distribution

Recall: if $Z_1, \ldots, Z_r$ are independent $N(0,1)$, then $\sum_{i=1}^r Z_i^2 \sim \chi^2_r$.

Definition

If $Z_1, \ldots, Z_r$ are independent, with $Z_i \sim N(\mu_i, 1)$, then $W = \sum_{i=1}^r Z_i^2$ has a noncentral χ^2 distribution with r degrees of freedom and noncentrality parameter $\Omega = \frac{1}{2}\sum_{i=1}^r \mu_i^2$.

- If $\Omega = 0$ (all $\mu_i = 0$), it is the usual χ^2_r.
- $E(W) = \sum_i \left(1 + \mu_i^2\right) = r + 2\Omega$: the larger Ω, the further to the right.



Some books, and R's `dchisq(x, df, ncp)`, use $\lambda = \sum_i \mu_i^2 = 2\Omega$.

The distribution of quadratic forms

On Friday, the F distribution of MSR/MSE rested on a result we took for granted. Here it is, stated without proof.

Theorem

Let $\boldsymbol{Z} \sim N(\boldsymbol{\mu},\ \sigma^2\boldsymbol{V})$, with $\boldsymbol{V}$ nonsingular, and let $\boldsymbol{A}$ and $\boldsymbol{B}$ be constant matrices, $\boldsymbol{A}$ symmetric.

1. If $\boldsymbol{AV}$ is idempotent, then $\boldsymbol{Z}'\boldsymbol{AZ}/\sigma^2$ has a noncentral χ^2 distribution (previous slide), with $r(\boldsymbol{A})$ degrees of freedom (the rank of $\boldsymbol{A}$) and noncentrality parameter $\Omega = \boldsymbol{\mu}'\boldsymbol{A}\boldsymbol{\mu}/(2\sigma^2)$. If $\Omega = 0$, it is the usual $\chi^2_{r(\boldsymbol{A})}$.
2. $\boldsymbol{Z}'\boldsymbol{AZ}$ and $\boldsymbol{Z}'\boldsymbol{BZ}$ ($\boldsymbol{B}$ symmetric) are independent if $\boldsymbol{AVB} = \boldsymbol{0}$.
3. $\boldsymbol{Z}'\boldsymbol{AZ}$ and $\boldsymbol{BZ}$ are independent if $\boldsymbol{BVA} = \boldsymbol{0}$.

For the response, $\boldsymbol{Z} = \boldsymbol{Y}$ and $\boldsymbol{V} = \boldsymbol{I}$: the conditions become $\boldsymbol{AA} = \boldsymbol{A}$ and $\boldsymbol{AB} = \boldsymbol{0}$.

$y \sim N(X^T\beta, \sigma^2 I)$

Applying it to SSE and SSR

$\boldsymbol{Y} \sim N(\boldsymbol{\mu},\, \sigma^2\boldsymbol{I})$ with $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, $SSE = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ and $SSR = \boldsymbol{Y}'(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J})\boldsymbol{Y}$ (Lecture 12).

Applying it to SSE and SSR

$\boldsymbol{Y} \sim N(\boldsymbol{\mu},\, \sigma^2 \boldsymbol{I})$ with $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, $SSE = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ and $SSR = \boldsymbol{Y}'(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J})\boldsymbol{Y}$ (Lecture 12).

Both matrices are idempotent (Lecture 12), their ranks are their traces, $n - p$ and $p - 1$, and their product is $\mathbf{0}$, since $\boldsymbol{H}\boldsymbol{H} = \boldsymbol{H}$ and $\boldsymbol{J}\boldsymbol{H} = \boldsymbol{J}$:

$$\left(\boldsymbol{H} - \tfrac{1}{n}\boldsymbol{J}\right)(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{H} - \boldsymbol{H}\boldsymbol{H} - \tfrac{1}{n}\boldsymbol{J} + \tfrac{1}{n}\boldsymbol{J}\boldsymbol{H} = \mathbf{0}.$$

Applying it to SSE and SSR

$\boldsymbol{Y} \sim N(\boldsymbol{\mu},\, \sigma^2 \boldsymbol{I})$ with $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$, $SSE = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$ and $SSR = \boldsymbol{Y}'(\boldsymbol{H} - \frac{1}{n}\boldsymbol{J})\boldsymbol{Y}$ (Lecture 12).

Both matrices are idempotent (Lecture 12), their ranks are their traces, $n - p$ and $p - 1$, and their product is $\mathbf{0}$, since $\boldsymbol{H}\boldsymbol{H} = \boldsymbol{H}$ and $\boldsymbol{J}\boldsymbol{H} = \boldsymbol{J}$:

$$\left(\boldsymbol{H} - \tfrac{1}{n}\boldsymbol{J}\right)(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{H} - \boldsymbol{H}\boldsymbol{H} - \tfrac{1}{n}\boldsymbol{J} + \tfrac{1}{n}\boldsymbol{J}\boldsymbol{H} = \mathbf{0}.$$

- $SSE/\sigma^2 \sim \chi^2_{n-p}$: its Ω is 0, because $(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X} = \mathbf{0}$.
- SSR/σ^2 is noncentral χ^2 with $p - 1$ degrees of freedom and $\Omega = \sum_i (\mu_i - \bar{\mu})^2/(2\sigma^2)$.
- SSR and SSE are independent (part 2 of the theorem).

The overall F test, justified

Under $H_0 : \beta_1 = \cdots = \beta_{p-1} = 0$, every $\mu_i = \beta_0$, so $\Omega = 0$ and $SSR/\sigma^2 \sim \chi^2_{p-1}$, independent of $SSE/\sigma^2 \sim \chi^2_{n-p}$. Hence

$$F^* = \frac{MSR}{MSE} = \frac{\big(SSR/\sigma^2\big)/(p-1)}{\big(SSE/\sigma^2\big)/(n-p)} \sim F(p-1,\ n-p) \quad \text{under } H_0 :$$

a ratio of two independent χ^2 variables, each divided by its degrees of freedom. σ^2 cancels, so F^* can be computed.

The overall F test, justified

Under $H_0 : \beta_1 = \cdots = \beta_{p-1} = 0$, every $\mu_i = \beta_0$, so $\Omega = 0$ and $SSR/\sigma^2 \sim \chi^2_{p-1}$, independent of $SSE/\sigma^2 \sim \chi^2_{n-p}$. Hence

$$F^* = \frac{MSR}{MSE} = \frac{\big(SSR/\sigma^2\big)/(p-1)}{\big(SSE/\sigma^2\big)/(n-p)} \sim F(p-1,\ n-p) \quad \text{under } H_0 :$$

a ratio of two independent χ^2 variables, each divided by its degrees of freedom. σ^2 cancels, so F^* can be computed.

Under H_a, the μ_i differ and $\Omega > 0$: SSR/σ^2 is pushed to the right, and F^* with it. This is why the test rejects only for large values of F^*, as $E(MSR) \geq \sigma^2$ suggested on Friday.

One linear combination of the coefficients

Estimating $c'\beta$

Many questions are about a linear combination
$\theta = \boldsymbol{c}'\boldsymbol{\beta} = c_0\beta_0 + c_1\beta_1 + \cdots + c_{p-1}\beta_{p-1}$, for a constant vector $\boldsymbol{c}$:

- $\boldsymbol{c} = (0, \ldots, 1, \ldots, 0)'$: a single coefficient β_k;
- $\boldsymbol{c} = \boldsymbol{X}_h$: the mean response at $\boldsymbol{X}_h$ (Lecture 11);
- $\boldsymbol{c} = \boldsymbol{X}_{h_1} - \boldsymbol{X}_{h_2}$: the difference between two mean responses.

Estimating $c'\beta$

Many questions are about a linear combination
$\theta = \boldsymbol{c}'\boldsymbol{\beta} = c_0\beta_0 + c_1\beta_1 + \cdots + c_{p-1}\beta_{p-1}$, for a constant vector $\boldsymbol{c}$:

- $\boldsymbol{c} = (0, \ldots, 1, \ldots, 0)'$: a single coefficient β_k;
- $\boldsymbol{c} = \boldsymbol{X}_h$: the mean response at $\boldsymbol{X}_h$ (Lecture 11);
- $\boldsymbol{c} = \boldsymbol{X}_{h_1} - \boldsymbol{X}_{h_2}$: the difference between two mean responses.

It is estimated by $\hat{\theta} = \boldsymbol{c}'\hat{\boldsymbol{\beta}}$, a linear function of $\hat{\boldsymbol{\beta}}$:

$$E(\hat{\theta}) = \boldsymbol{c}'\boldsymbol{\beta}, \qquad Var(\hat{\theta}) = \boldsymbol{c}'\,\boldsymbol{\sigma^2}\{\hat{\boldsymbol{\beta}}\}\,\boldsymbol{c} = \sigma^2\,\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}.$$

Estimating $c'\beta$

Many questions are about a linear combination
$\theta = \boldsymbol{c}'\boldsymbol{\beta} = c_0\beta_0 + c_1\beta_1 + \cdots + c_{p-1}\beta_{p-1}$, for a constant vector $\boldsymbol{c}$:

- $\boldsymbol{c} = (0, \ldots, 1, \ldots, 0)'$: a single coefficient β_k;
- $\boldsymbol{c} = \boldsymbol{X}_h$: the mean response at $\boldsymbol{X}_h$ (Lecture 11);
- $\boldsymbol{c} = \boldsymbol{X}_{h_1} - \boldsymbol{X}_{h_2}$: the difference between two mean responses.

It is estimated by $\hat{\theta} = \boldsymbol{c}'\hat{\boldsymbol{\beta}}$, a linear function of $\hat{\boldsymbol{\beta}}$:

$$E(\hat{\theta}) = \boldsymbol{c}'\boldsymbol{\beta}, \qquad Var(\hat{\theta}) = \boldsymbol{c}'\,\boldsymbol{\sigma^2}\{\hat{\boldsymbol{\beta}}\}\,\boldsymbol{c} = \sigma^2\,\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}.$$

With normal errors, and $s^2\{\hat{\theta}\} = MSE\,\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}$: $\dfrac{\hat{\theta} - \theta}{s\{\hat{\theta}\}} \sim t(n-p)$.

Hence the interval $\hat{\theta} \pm t(1-\alpha/2;\ n-p)\,s\{\hat{\theta}\}$, and the test of $H_0 : \theta = \theta_0$ with $t^* = (\hat{\theta} - \theta_0)/s\{\hat{\theta}\}$, as before.

Example: two cars

Model $\texttt{acc} = \beta_0 + \beta_1\,\texttt{wt} + \beta_2\,\texttt{hp} + \epsilon$. How much faster, on average, is a car of 3500 lb and 130 hp than one of 3000 lb and 100 hp?

$$\theta = E(Y \mid 3500, 130) - E(Y \mid 3000, 100) = 500\,\beta_1 + 30\,\beta_2, \qquad \boldsymbol{c} = (0, 500, 30)'.$$

Recall that `vcov(fit2)` is $MSE\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

```
fit2 <- lm(acc ~ wt + hp, data = autompg)
cc <- c(0, 500, 30)
theta_hat <- sum(cc * coef(fit2))
s_theta <- sqrt(drop(t(cc) %*% vcov(fit2) %*% cc))
c(theta_hat = theta_hat, s_theta = s_theta)
```

```
##   theta_hat     s_theta
## -1.65182813  0.07030124
```

Example: two cars

Model $\texttt{acc} = \beta_0 + \beta_1\,\texttt{wt} + \beta_2\,\texttt{hp} + \epsilon$. How much faster, on average, is a car of 3500 lb and 130 hp than one of 3000 lb and 100 hp?

$$\theta = E(Y \mid 3500, 130) - E(Y \mid 3000, 100) = 500\,\beta_1 + 30\,\beta_2, \qquad \boldsymbol{c} = (0, 500, 30)'.$$

Recall that `vcov(fit2)` is $MSE\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

```
fit2 <- lm(acc ~ wt + hp, data = autompg)
cc <- c(0, 500, 30)
theta_hat <- sum(cc * coef(fit2))
s_theta <- sqrt(drop(t(cc) %*% vcov(fit2) %*% cc))
c(theta_hat = theta_hat, s_theta = s_theta)
```

```
##   theta_hat     s_theta
## -1.65182813  0.07030124
```

95% interval: $-1.65 \pm 1.966 \times 0.070 = (-1.79,\ -1.51)$ s. The heavier but more powerful car reaches 60 mph 1.5 to 1.8 s sooner on average.

The general linear hypothesis

Several restrictions at once

A general linear hypothesis imposes q linear restrictions on $\boldsymbol{\beta}$:

$$H_0 : \boldsymbol{K}'\boldsymbol{\beta} = \boldsymbol{m} \qquad \text{against} \qquad H_a : \boldsymbol{K}'\boldsymbol{\beta} \neq \boldsymbol{m},$$

where $\boldsymbol{K}'$ is a constant $q \times p$ matrix of rank q (one row per restriction) and $\boldsymbol{m}$ is a constant $q \times 1$ vector.

Several restrictions at once

A general linear hypothesis imposes q linear restrictions on $\boldsymbol{\beta}$:

$$H_0 : \boldsymbol{K}'\boldsymbol{\beta} = \boldsymbol{m} \qquad \text{against} \qquad H_a : \boldsymbol{K}'\boldsymbol{\beta} \neq \boldsymbol{m},$$

where $\boldsymbol{K}'$ is a constant $q \times p$ matrix of rank q (one row per restriction) and $\boldsymbol{m}$ is a constant $q \times 1$ vector.

Examples, with $p = 5$ parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$:

- $H_0 : \beta_3 = \beta_4 = 0$: $\boldsymbol{K}' = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, $\boldsymbol{m} = \boldsymbol{0}$, $q = 2$.
- $H_0 : \beta_1 = \beta_2$: $\boldsymbol{K}' = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \end{bmatrix}$, $\boldsymbol{m} = 0$, $q = 1$.
- The overall test: $\boldsymbol{K}' = [\, \boldsymbol{0} \mid \boldsymbol{I}_4 \,]$, $\boldsymbol{m} = \boldsymbol{0}$, $q = p - 1 = 4$.

The F statistic

$\boldsymbol{K}'\hat{\boldsymbol{\beta}}$ is a linear function of $\hat{\boldsymbol{\beta}}$. Under H_0, with normal errors,

$$\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m} \sim N\big(\boldsymbol{0},\ \sigma^2\, \boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\big).$$

The F statistic

$\boldsymbol{K}'\hat{\boldsymbol{\beta}}$ is a linear function of $\hat{\boldsymbol{\beta}}$. Under H_0, with normal errors,

$$\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m} \sim N\big(\mathbf{0},\ \sigma^2\, \boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\big).$$

The theorem on quadratic forms, with $\boldsymbol{A}$ the inverse of $\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}$ (so $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{I}$), gives a χ^2_q variable below, independent of SSE since $\hat{\boldsymbol{\beta}}$ is (part 3: $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{H}) = \mathbf{0}$). Hence

$$F^* = \frac{(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'\big[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\big]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})\,/\,q}{MSE} \sim F(q,\, n{-}p) \text{ under } H_0.$$

The F statistic

$\boldsymbol{K}'\hat{\boldsymbol{\beta}}$ is a linear function of $\hat{\boldsymbol{\beta}}$. Under H_0, with normal errors,

$$\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m} \sim N\big(\mathbf{0},\ \sigma^2\, \boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\big).$$

The theorem on quadratic forms, with $\boldsymbol{A}$ the inverse of $\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}$ (so $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{I}$), gives a χ^2_q variable below, independent of SSE since $\hat{\boldsymbol{\beta}}$ is (part 3: $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{H}) = \mathbf{0}$). Hence

$$F^* = \frac{(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'\big[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\big]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})\,/\,q}{MSE} \sim F(q,\, n{-}p) \text{ under } H_0.$$

Reject H_0 at level α when $F^* > F(1-\alpha;\ q,\ n-p)$.

The F statistic

$\boldsymbol{K}'\hat{\boldsymbol{\beta}}$ is a linear function of $\hat{\boldsymbol{\beta}}$. Under H_0, with normal errors,

$$\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m} \sim N\big(\boldsymbol{0},\ \sigma^2\, \boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\big).$$

The theorem on quadratic forms, with $\boldsymbol{A}$ the inverse of $\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}$ (so $\boldsymbol{A}\boldsymbol{V} = \boldsymbol{I}$), gives a χ^2_q variable below, independent of SSE since $\hat{\boldsymbol{\beta}}$ is (part 3: $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{H}) = \boldsymbol{0}$). Hence

$$F^* = \frac{(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})'\big[\boldsymbol{K}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{K}\big]^{-1}(\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m})\,/\,q}{MSE} \sim F(q,\, n{-}p) \text{ under } H_0.$$

Reject H_0 at level α when $F^* > F(1-\alpha;\ q,\ n-p)$.

With one restriction ($q = 1$), F^* is the square of the t^* of the previous section.

In R: do displacement and cylinders matter?

Full model: `acc ~ wt + hp + disp + cyl`. $H_0 : \beta_{\text{disp}} = \beta_{\text{cyl}} = 0$, so $q = 2$.

Since `vcov(full)` is $MSE\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$, the numerator and MSE combine:

```
full <- lm(acc ~ wt + hp + disp + cyl, data = autompg)
Kt <- rbind(c(0, 0, 0, 1, 0),
            c(0, 0, 0, 0, 1))
d <- Kt %*% coef(full)              # here m = 0
V <- Kt %*% vcov(full) %*% t(Kt)
F_star <- drop(t(d) %*% solve(V) %*% d) / 2
c(F_star, 1 - pf(F_star, 2, 385))
```

```
## [1] 7.6362147397 0.0005594001
```

In R: do displacement and cylinders matter?

Full model: `acc ~ wt + hp + disp + cyl`. $H_0 : \beta_{\text{disp}} = \beta_{\text{cyl}} = 0$, so $q = 2$.

Since `vcov(full)` is $MSE\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$, the numerator and MSE combine:

```
full <- lm(acc ~ wt + hp + disp + cyl, data = autompg)
Kt <- rbind(c(0, 0, 0, 1, 0),
            c(0, 0, 0, 0, 1))
d <- Kt %*% coef(full)              # here m = 0
V <- Kt %*% vcov(full) %*% t(Kt)
F_star <- drop(t(d) %*% solve(V) %*% d) / 2
c(F_star, 1 - pf(F_star, 2, 385))
```

```
## [1] 7.6362147397 0.0005594001
```

$F^* = 7.64 > F(0.95;\ 2,\ 385) = 3.02$: reject H_0. Once weight and horsepower are in the model, displacement and cylinders together still explain some variation in acceleration.

Summary: testing coefficients

H_0	Statistic	Under H_0
$\boldsymbol{c}'\boldsymbol{\beta} = \theta_0$ (one combination)	$t^* = (\boldsymbol{c}'\hat{\boldsymbol{\beta}} - \theta_0)/s\{\boldsymbol{c}'\hat{\boldsymbol{\beta}}\}$	$t(n-p)$
$\boldsymbol{K}'\boldsymbol{\beta} = \boldsymbol{m}$ (q restrictions)	F^* from $\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m}$	$F(q,\, n-p)$

Summary: testing coefficients

H_0	Statistic	Under H_0
$\boldsymbol{c}'\boldsymbol{\beta} = \theta_0$ (one combination)	$t^* = (\boldsymbol{c}'\hat{\boldsymbol{\beta}} - \theta_0)/s\{\boldsymbol{c}'\hat{\boldsymbol{\beta}}\}$	$t(n-p)$
$\boldsymbol{K}'\boldsymbol{\beta} = \boldsymbol{m}$ (q restrictions)	F^* from $\boldsymbol{K}'\hat{\boldsymbol{\beta}} - \boldsymbol{m}$	$F(q,\ n-p)$

Next: the same test from two fits (extra sums of squares), then multicollinearity: why a predictor can matter jointly but not on its own.