Intermediate Econometrics

Vincenzo Gioia

21st October 2026

Statistical Model: Inferential Results

Likelihood Inference

- Let's assume to be interested on the inference on a parameter θ (unknown)
- Given a random sample $y = (y_1, \ldots, y_n)$ of size n from Y which is distributed according to the model $f(y; \theta)$, with $\theta \in \Theta \subset R^d, d \geq 1$, we define the likelihood function for θ the function as

$$\mathcal{L}(\theta) = f(y_1, \ldots, y_n; \theta)$$

where $f(y_1, \ldots, y_n; \theta)$ is the joint density function

Statistical Model: Inferential Results

Likelihood Inference

- Under a random sampling, due to the independence condition of the random variables (and for the condition of identical distribution), we have

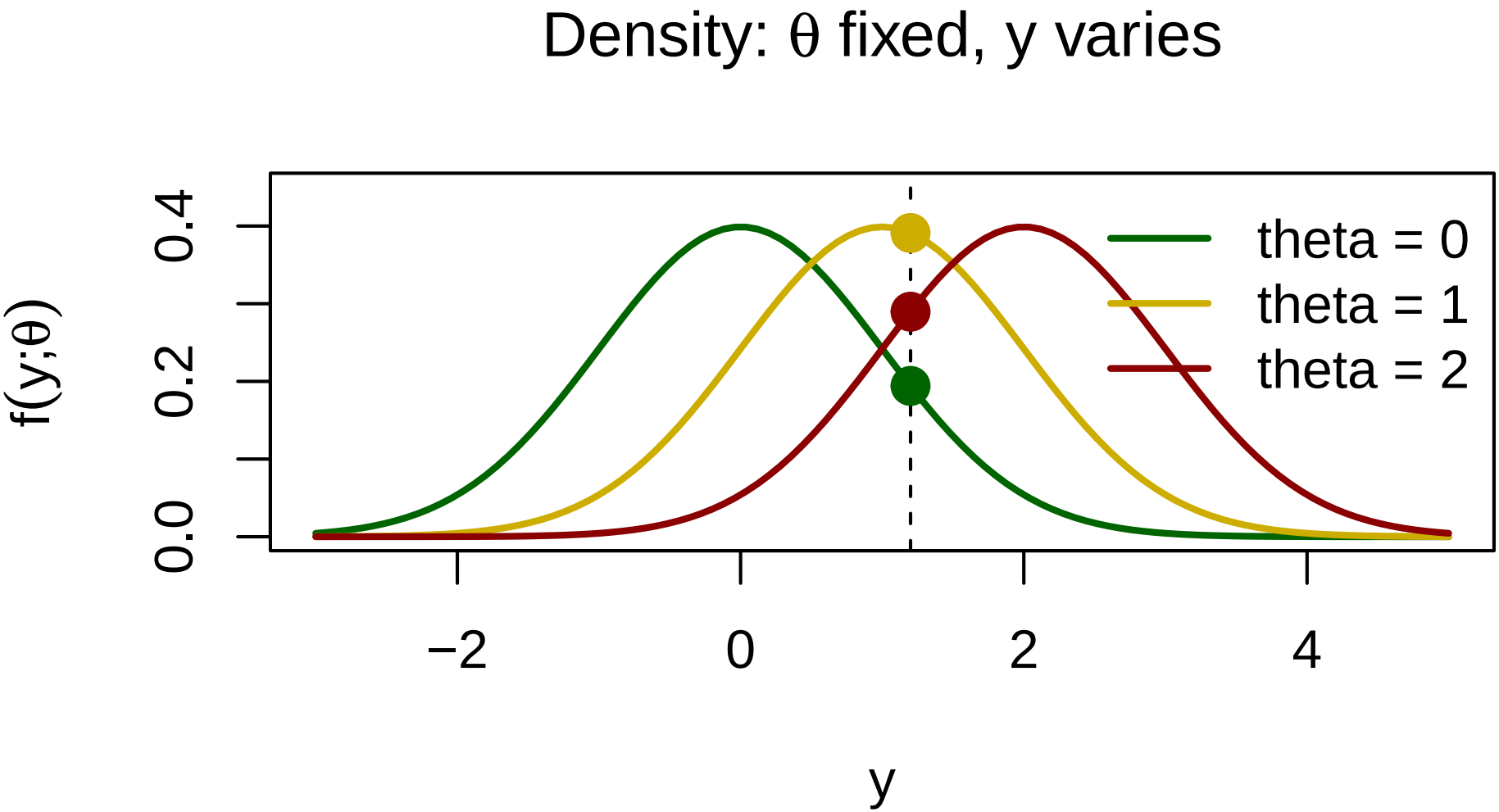
$$\mathcal{L}(\theta) = f(y_1, \ldots, y_n; \theta) = \prod_{i=1}^{n} f(y_i; \theta)$$

- For convenience, it is better to work with the log-likelihood function

$$\ell(\theta) = \log \mathcal{L}(\theta) = \sum_{i=1}^{n} \log f(y_i; \theta)$$

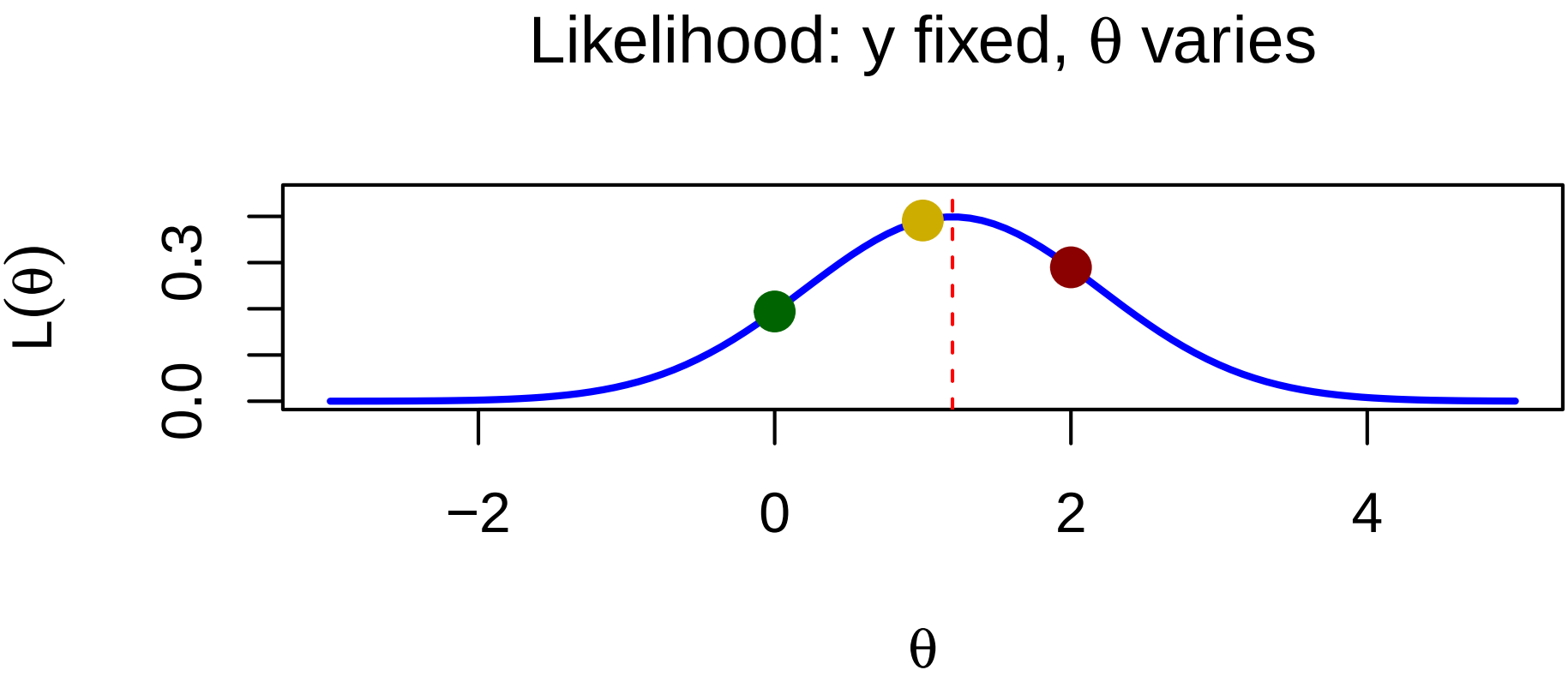
Likelihood: One Function, Two Readings

- Let $Y \sim \mathcal{N}(\theta, 1)$ and suppose we observe $y_{obs} = 1.2$
- **Reading 1** (θ fixed, y varies): $f(y; \theta)$ is a **density** in y



Likelihood: One Function, Two Readings

- **Reading 2** ($y = y_{obs}$ fixed, θ varies): $\mathcal{L}(\theta) = f(y_{obs}; \theta)$ is the **likelihood**
- It's not a probability function: It's a function of θ which is defined on Θ getting values in $\mathbb{R}$



Statistical Model: Inferential Results

Likelihood Is Not a Probability Function

- Random sample of size $n = 5$ from $\mathcal{N}(\theta, 1)$
- Under independence, $\mathcal{L}(\theta) = \prod_{i=1}^{n} f(y_i; \theta)$
- The area under $f(y; \theta)$ over y is 1 (for any fixed θ), while the area under $\mathcal{L}(\theta)$ over θ is in general **not** 1

```
set.seed(7)
y <- rnorm(5, mean = 1, sd = 1)
lik <- function(theta)
  sapply(theta,
         function(t) prod(dnorm(y, mean = t, sd = 1)))
```

Statistical Model: Inferential Results

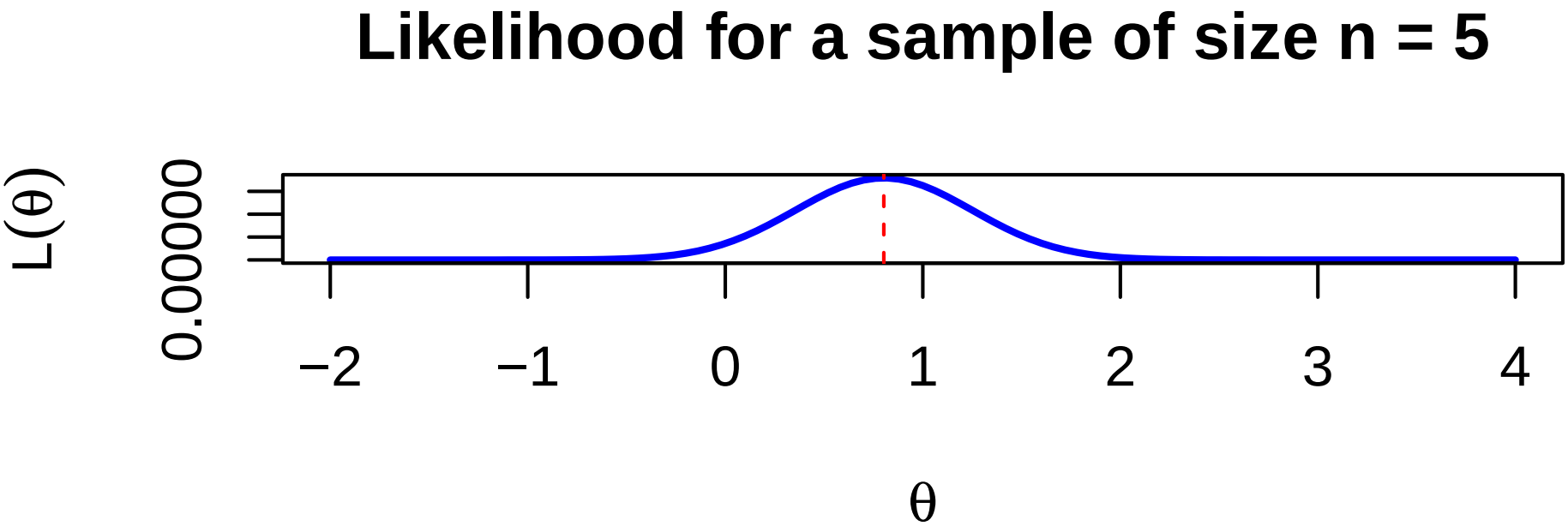
Likelihood Is Not a Probability Function

```
# Area under the density in y (theta fixed)
integrate(dnorm, -100, 100, mean = 1, sd = 1)$value
```

```
## [1] 1
```

```
# Area under the likelihood in theta (y fixed)
integrate(lik, -100, 100)$value
```

```
## [1] 0.0002015088
```



Statistical Model: Inferential Results

Maximum Likelihood

- A value $\hat{\theta} \in \Theta$ such that $\ell(\hat{\theta}) \geq \ell(\theta)$ for all $\theta \in \Theta$ is said maximum likelihood estimate (the random variable $\hat{\theta}(Y)$ is said maximum likelihood estimator)

Score

- Under regularity conditions (what we consider) the maximum likelihood estimate is obtained by solving

$$\ell'(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = 0$$

where the vector $\ell'(\theta)$ is the score vector

Asymptotic distribution of the MLE

- Finally, $\hat{\theta}$ is consistent and has asymptotic distribution $\hat{\theta} \dot{\sim} \mathcal{N}(\theta, j(\hat{\theta})^{-1})$ where

$$j(\theta) = -\ell''(\theta) = -\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta^\top}$$

The Normal model: Likelihood Inference

Likelihood Function

$$\mathcal{L}(\theta) = \mathcal{L}(\mu, \sigma^2) = \prod_{i=1}^{n} f(y_i; \mu, \sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

Log-Likelihood Function

$$\ell(\mu, \sigma^2) = \log \mathcal{L}(\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$$

The Normal model: Likelihood Inference

Score Vector: $\ell'(\theta) = (s_1, s_2)$

$$s_1 = \frac{\partial \ell(\theta)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \mu)$$

$$s_2 = \frac{\partial \ell(\theta)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \mu)^2$$

Observed information $J(\theta) = -\nabla^2 \ell(\theta) = -\begin{pmatrix} \frac{\partial^2 \ell}{\partial \mu^2} & \frac{\partial^2 \ell}{\partial \mu \, \partial \sigma^2} \\ \frac{\partial^2 \ell}{\partial \sigma^2 \, \partial \mu} & \frac{\partial^2 \ell}{\partial (\sigma^2)^2} \end{pmatrix}$.

$$\boxed{J(\mu, \sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^4} \sum_{i=1}^{n} (y_i - \mu) \\ \frac{1}{\sigma^4} \sum_{i=1}^{n} (y_i - \mu) & -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} \sum_{i=1}^{n} (y_i - \mu)^2 \end{pmatrix}}$$

The Normal model: Likelihood Inference

Maximum Likelihood Estimate

- From $s_1 = \ell_1'(\theta) = 0$ we obtain

$$\hat{\mu} = \bar{y}$$

- From $s_2 = \ell_2'(\theta) = 0$ (and substituting μ with $\hat{\mu}$), we obtain

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

- The latter is biased but it can be expressed in terms of the unbiased variance estimator using the relationship $\hat{\sigma}^2 = \frac{S^2(n-1)}{n}$

The Normal model: Likelihood Inference

Observed information $J(\hat{\theta})$

MLE:

$$\hat{\mu} = \bar{y}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2.$$

Observed information evaluated in the MLE $J(\hat{\theta})$

$$\boxed{J(\hat{\mu}, \hat{\sigma}^2) = \begin{pmatrix} \frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{pmatrix}}$$

Maximum Likelihood: Maximum and Score

Example: $Y_i \sim \mathcal{N}(\mu, \sigma^2 = 1)$, sample of size $n = 5$

- Let's simplify the previous case by assuming the σ^2 is known
- So, we have
  - $\ell(\mu) = \sum_{i=1}^{n} \log f(y_i; \mu) = c(y) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2$
  - $s1 = \frac{\partial \ell(\theta)}{\partial \mu} = \sum_{i=1}^{n}(y_i - \mu) = n \times (\bar{y} - \mu) \quad (s_1 = 0 \implies \hat{\mu} = \bar{y})$
  - $J(\mu) = -\frac{\partial^2 \ell}{\partial \mu^2} = n$
- We generate the data (assuming a value for μ) and we can build our functions

```
set.seed(7)
y <- rnorm(5, mean = 1, sd = 1)
n <- length(y)
mu_hat <- mean(y)

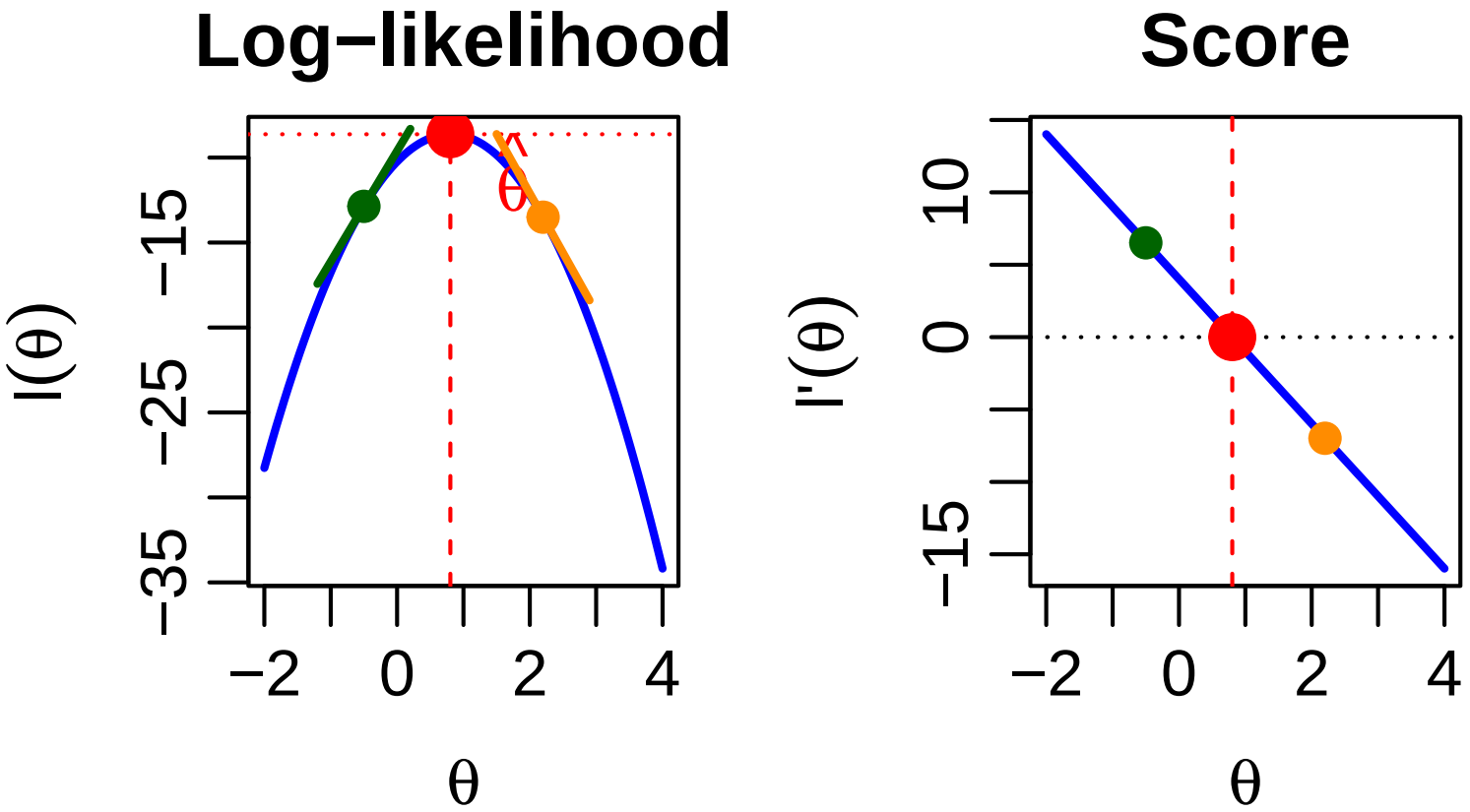
loglik <- function(mu, y)
  sapply(mu,
         function(t) sum(dnorm(y, mean = t, sd = 1, log = TRUE)))
score  <- function(mu, y) length(y) * (mean(y) - mu)
```

Maximum Likelihood: Maximum and Score

Example: $Y_i \sim \mathcal{N}(\mu, 1)$, sample of size $n = 5$

- $\hat{\theta}$ is the value where $\ell(\hat{\theta}) \geq \ell(\theta)$ for all θ: the **peak** of the log-likelihood

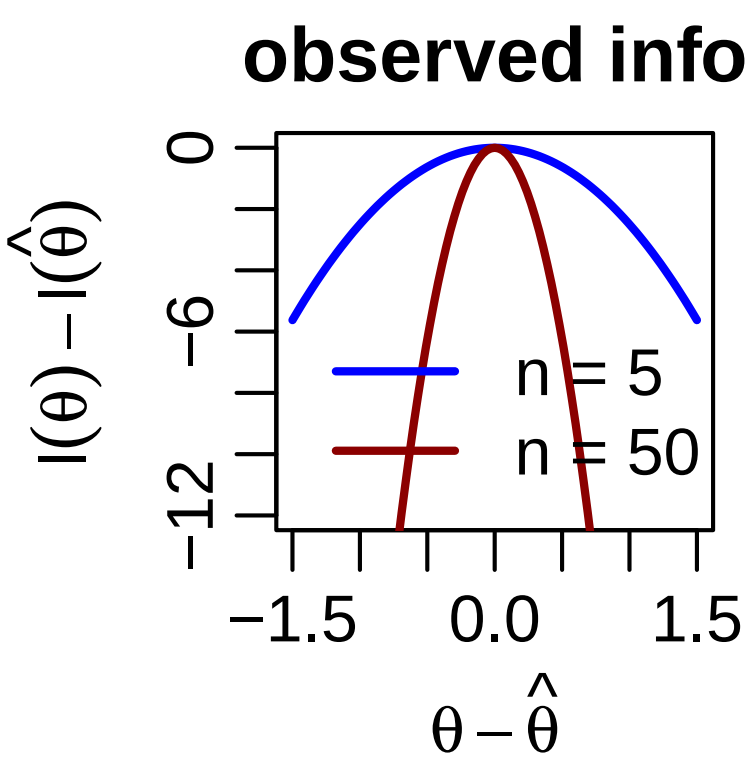
- The score $\ell'(\theta)$ is the **slope** of the log-likelihood: positive on the left of $\hat{\theta}$, negative on the right, and **zero at** $\hat{\theta}$

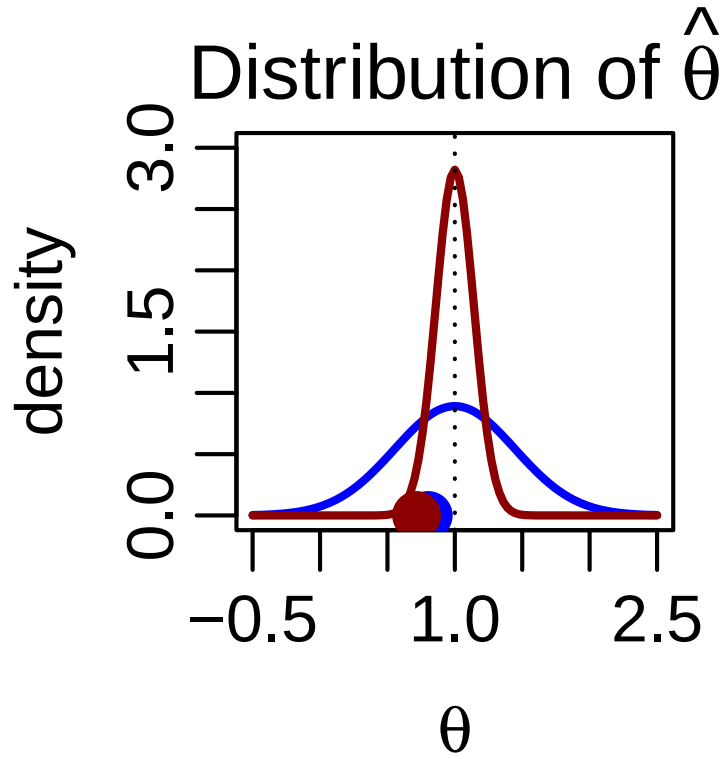


Maximum Likelihood: Information and precision

Example: $Y_i \sim \mathcal{N}(\theta, 1)$, sample of size $n = 5$

- $j(\hat{\theta}) = -\ell''(\hat{\theta})$: **curvature** of the log-likelihood at its peak
- More data give a more curved log-likelihood: larger $j(\hat{\theta})$ and smaller variance $j(\hat{\theta})^{-1}$ for $\hat{\theta}$
- $j(\hat{\theta}) = n$, so $\hat{\theta} \sim \mathcal{N}(\theta, 1/n)$ (here the approximation is exact)





The Normal model: Likelihood Inference

- In this case, we have an explicit solution for the maximum likelihood estimates. In other cases, we must obtain the estimates numerically (for instance considering the Newton-Raphson algorithm)

Optimization

- In R, there are several routines for optimization purposes (usually minimize a function): `optim()`, `nlminb()`, ...

- We analyze one of them in order to obtain the maximum likelihood estimates, comparing it with the analytic solution

Statistical Model: The Normal model

Optimization Example

- First, we need to write down the (negative) likelihood function
- Let's generate a sample of size $n = 50$ from a $\mathcal{N}(\mu = 10, \sigma^2 = 4)$
- Optimize the likelihood function using `nlminb()`, which requires:
  - the starting values for the algorithm (argument `start`)
  - the function to minimize (argument `objective`)
  - the data (argument `data`)
  - the limits of the parametric space (arguments `lower` and `upper`), if there are some constraints on the parametric space
- After running `nlminb()`, explore its structure and extract the MLE
- Then, compare the MLE obtained via `nlminb()` with those obtained via analytic solution

Statistical Model: The Normal model

Optimization Example

- First, we need to write down the (negative) likelihood function

```
nllik <- function(par, data) {
  mu <- par[1]
  sigma2 <- par[2]
  n <- length(data)
  nll <-  0.5 * n * log(sigma2) + sum((data - mu)^2) / (2 * sigma2)
  return(nll)
}
```

- Let's generate a sample of size $n = 50$ from a $\mathcal{N}(\mu = 10, \sigma^2 = 4)$

```
set.seed(13)
n <- 50
y <- rnorm(n, mean = 10, sd = 2)
```

Statistical Model: The Normal model

Optimization Example

- Optimize the likelihood function using nlminb(), which requires:
  - the starting values for the algorithm (argument start)
  - the function to minimize (argument objective)
  - the data (argument data)
  - the limits of the parametric space (arguments lower and upper), if there are some constraints on the parametric space

```
start_vals <- c(mu = 0, sigma = 1)

fit <- nlminb(start = start_vals,
              objective = nllik,
              data = y,
              lower = c(-Inf, 1e-6))
```

Statistical Model: The Normal model

Optimization Example

- After running nlminb, explore its structure and extract the MLE
- Then, compare the MLE obtained via nlminb with those obtained via analytic solution

```
str(fit)
```

```
## List of 6
##  $ par        : Named num [1:2] 9.95 3.73
##   ..- attr(*, "names")= chr [1:2] "mu" "sigma"
##  $ objective  : num 57.9
##  $ convergence: int 0
##  $ iterations : int 18
##  $ evaluations: Named int [1:2] 25 46
##   ..- attr(*, "names")= chr [1:2] "function" "gradient"
##  $ message    : chr "relative convergence (4)"
```

```
fit$par
```

```
##       mu    sigma
## 9.947162 3.726903
```

```
mean(y)
```

```
## [1] 9.947163
```

```
var(y)*(n-1)/n
```

```
## [1] 3.726902
```

The Normal Model: Linear Regression

Linear Regression Model

- The settings above can be straightforwardly extended to linear regression models
- Let's consider a simple linear regression $Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ independently
- Note that the difference is that we do not have the identically distribution because $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ where $\mu_i = \beta_1 + \beta_2 x_i$

The Normal Model: Linear Regression

Linear Regression Model

Exercise

- Adapt the negative log-likelihood function built above to the case of a simple linear regression model

- Let's generate some data:
  - set $n = 100$
  - generate x from a uniform distribution in (0,10)
  - fix $\beta_0 = 2$, $\beta_1 = 0.8$, and $\sigma = 1.5$

- Fit the model using `nlminb()`

- Then, compare the results with those obtained via `lm()` function

The Normal Model: Linear Regression

Linear Regression Model

Exercise

- Adapt the negative log-likelihood function built above to the case of a simple linear regression model

```
nllik_lm <- function(par, x, y) {
  beta0 <- par[1]
  beta1 <- par[2]
  sigma2 <- par[3]
  mu <- beta0 + beta1 * x
  n <- length(y)
  nll <-  0.5 * n * log(sigma2) + sum((y - mu)^2) / (2 * sigma2)
  return(nll)
}
```

The Normal Model: Linear Regression

Linear Regression Model

Exercise

- Let's generate some data:
  - set $n = 100$
  - generate x from a uniform distribution in (0,10)
  - fix $\beta_0 = 2$, $\beta_1 = 0.8$, and $\sigma = 1.5$

```
n <- 100
x <- runif(n, 0, 10)
beta0 <- 2; beta1 <- 0.8; sigma <- 1.5
y <- beta0 + beta1 * x + rnorm(n, 0, sigma)
```

The Normal Model: Linear Regression

Linear Regression Model

Exercise

- Fit the model using `nlminb()`

```
start_vals <- c(beta0 = 0, beta1 = 0, sigma2 = 1)

fit <- nlminb(
  start = start_vals,
  objective = nllik_lm,
  x = x,
  y = y,
  lower = c(-Inf, -Inf, 1e-6)
)
```

The Normal Model: Linear Regression

Linear Regression Model

Exercise

- Then, compare the results with those obtained via `lm()` function

```
fit$par
```

```
##     beta0     beta1    sigma2
## 2.1476594 0.7798585 2.6494339
```

```
fitLM <- lm(y ~ x)
fitLM$coefficients
```

```
## (Intercept)           x
##   2.1476602   0.7798585
```

```
sum(fitLM$residuals^2)/n
```

```
## [1] 2.649434
```

The Normal Model: Linear Regression

Linear Regression Model

- As we noticed the β estimates obtained via maximum likelihood are the same of those obtained via OLS

- Indeed, from

$$\mathcal{L}(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y_i - \beta_1 x_{i1} - \ldots - \beta_p x_{ip})^2}$$

$$\propto (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_1 x_{i1} - \ldots - \beta_p x_{ip})^2}$$

$$\propto (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)}$$

$$\ell(\beta, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)$$

- Note that the term that does not depend on the parameter has been removed and we introduce the proportionality symbol $\propto$

The Normal Model: Linear Regression

β Estimate

- For fixed σ^2, $\ell(\beta, \sigma^2)$ is maximum when the residual sum of square $RSS(\beta) = (y - X^\top \beta)^\top (y - X\beta)$ is minimum, that is

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \operatorname{RSS}(\beta) = \underset{\beta}{\operatorname{argmax}} \ell(\beta, \sigma^2)$$

- Then, the maximum likelihood estimate is equal to the OLS estimate

$$\hat{\beta}_{ML} = \hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top y$$

The Normal model: Linear Regression

σ^2 Estimate

- By substituing $\hat{\beta}$ in $\ell(\beta, \sigma^2)$ we have

$$\ell(\hat{\beta}, \sigma^2) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X^\top\hat{\beta})^\top(y - X\hat{\beta})$$

$$= -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}RSS(\hat{\beta}) = -\frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} e_i^2$$

- By getting the derivative with respect to σ^2 and equating to zero, we obtain

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}e^\top e$$

The Normal model: Linear Regression

Distribution of $\hat{\beta}$

- Since $Y \sim \mathcal{N}_n(\mu, \sigma^2 I)$, the estimator $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ is a linear transformation of a multivariate normal. So

$$\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2 (X^\top X)^{-1})$$

- Due to the results on the multivariate normal distribution, we know that also the marginals are normally distributed and in particular

$$\hat{\beta}_r \sim \mathcal{N}(\beta_r, \sigma^2 (X^\top X)^{-1}_{rr}), \quad r = 1, \ldots, p$$

having denoted with $(X^\top X)^{-1}_{rr}$ the r-th diagonal element of $(X^\top X)^{-1}$.

The Normal model: Linear Regression

Distribution of $\hat{\sigma}^2$

- The estimator σ^2 is biased, while the unbiased estimator is

$$S^2 = \frac{n}{n-p}\hat{\sigma}^2$$

for which

$$\frac{(n-p)S^2}{\sigma^2} \sim \chi^2_{n-p}$$

Joint Distribution of $(\hat{\beta}, \hat{\sigma}^2)$

- $\hat{\beta}$ is independent from $\hat{\sigma}^2$, and so $\hat{\beta}$ is independent from S^2

Statistical Model: Inferential Results

Hypothesis Test

- To verify the hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ the test statistic is based on the test of the log-likelihood ratio, that is

$$W(\theta_0) = 2(\ell(\hat{\theta}) - \ell(\theta_0))$$

whose asymptotic distribution is χ^2_d

Confidence interval/region

- The confidence region is:

$$\{\theta : W(\theta) \leq \chi^2_{d;1-\alpha}\}$$

Note

- In the normal linear model, by assuming the normality of errors the function $W(\theta)$ is a monotone increasing function of a statistic F (Fisher) or t (Students' t) and the results are reported in that terms

Normal Linear Regression: Inference on a Single Coefficient

Confidence interval

- From the results just outlined, we can obtain pivotal quantities to make inference on a single coefficient

$$t_r = \frac{\hat{\beta}_r - \beta_r}{\sqrt{S^2 (X^\top X)^{-1}_{rr}}} \sim t_{n-p}$$

- Confidence interval of level $1 - \alpha$ for β_r

$$IC^{1-\alpha}_{\beta_r} = (\hat{\beta}_r - t_{n-p;1-\alpha/2} \sqrt{S^2((X^\top X)^{-1}_{rr})}, \hat{\beta}_r + t_{n-p;1-\alpha/2} \sqrt{S^2((X^\top X)^{-1}_{rr})})$$

Normal Linear Reg.: Inference on a Single Coefficient

Hypothesis test (nullity of the single coefficient):

$$\begin{cases} H_0 : \beta_r = 0 \\ H_1 : \beta_r \neq 0 \end{cases}$$

$$t_r = \hat{\beta}_r / \sqrt{S^2 (X^\top X)^{-1}_{rr}} \overset{H_0}{\sim} t_{n-p}$$

$$\mathcal{R}_\alpha = \{ |t_r^{obs}| > t_{n-p;1-\alpha/2} \}$$

$$p - value = P(|t_{n-p}| > |t_r^{obs}|) = 2P(t_{n-p} \leq -|t_r^{obs}|)$$

Normal Linear Regression: Inference

- Consider the example and the model fitted during the first lecture

```
library(ISLR)
```

```
## Warning: il pacchetto 'ISLR' è stato creato con R versio
```

```
data("Wage")
lmFit <- lm(wage ~ age + year + education, data = Wage)
summary(lmFit)$coefficients
```

```
##                                    Estimate   Std. Error
## (Intercept)                   -2058.0770290 649.34597951 -
## age                               0.5621005   0.05713794
## year                              1.0562995   0.32377790
## education2. HS Grad              11.3976404   2.47630123
## education3. Some College         24.2344792   2.60564592
## education4. College Grad         39.7408035   2.58616318 1
## education5. Advanced Degree      64.8536598   2.80416339 2
```

- Some useful quantities

Normal Linear Regression: Inference

Confidence intervals

- By using the confint() function and by hand

```
confint(lmFit)
```

```
##                                    2.5 %       97.5 %
## (Intercept)                -3331.2866429 -784.8674152
## age                            0.4500669    0.6741341
## year                           0.4214498    1.6911493
## education2. HS Grad            6.5422157   16.2530651
## education3. Some College      19.1254409   29.3435174
## education4. College Grad      34.6699661   44.8116408
## education5. Advanced Degree   59.3553771   70.3519425
```

```
cbind(lmFit$coefficients - sqrt(hat_s2 * diag(solve(t(X) %*% X)))* qt(0.975, df = n-p),
      lmFit$coefficients + sqrt(hat_s2 * diag(solve(t(X) %*% X)))* qt(0.975, df = n-p))
```

```
##                                    [,1]         [,2]
## (Intercept)                -3331.2866429 -784.8674152
## age                            0.4500669    0.6741341
## year                           0.4214498    1.6911493
## education2. HS Grad            6.5422157   16.2530651
## education3. Some College      19.1254409   29.3435174
## education4. College Grad      34.6699661   44.8116408
## education5. Advanced Degree   59.3553771   70.3519425
```

Normal Linear Regression: Inference

Hypothesis test

```
summary(lmFit)$coefficients[,3:4]
```

```
##                                t value      Pr(>|t|)
## (Intercept)                 -3.169461  1.542629e-03
## age                          9.837604  1.694287e-22
## year                         3.262420  1.117059e-03
## education2. HS Grad          4.602687  4.344327e-06
## education3. Some College     9.300757  2.611570e-20
## education4. College Grad    15.366704  2.377262e-51
## education5. Advanced Degree 23.127632 5.135751e-109
```

```
tval <- lmFit$coefficients/(sqrt(hat_s2 * diag(solve(t(X) %*% X))))

cbind(tval, 2*pt(-abs(tval), n - p))
```

```
##                                 tval
## (Intercept)                 -3.169461  1.542629e-03
## age                          9.837604  1.694287e-22
## year                         3.262420  1.117059e-03
## education2. HS Grad          4.602687  4.344327e-06
## education3. Some College     9.300757  2.611570e-20
## education4. College Grad    15.366704  2.377262e-51
## education5. Advanced Degree 23.127632 5.135751e-109
```

Normal Linear Reg.: Inference on a Pool of Coefficients

A Special Case

- Let's suppose we want assess the model overall, that is no one of the predictors are explaining the variability of Y. The hypothesis system is

$$\begin{cases} H_0 : \beta_2 = \ldots = \beta_p = 0 \\ H_1 : \text{at least one of} \quad \beta_r \neq 0 \quad \text{for } r = 1, \ldots, p \end{cases}$$

- In this case we distinguish between

$$\text{Full model}(\mathcal{M}_\infty) : Y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i$$

$$\text{Reduced model}(\mathcal{M}_\prime) : Y_i = \beta_1 + \varepsilon_i$$

Normal Linear Reg.: Inference on a Pool of Coefficients

A Special Case

- Under $\mathcal{M}_\infty$ the estimate of the variance of the error term component is

$$\hat{\sigma}^2 = \frac{1}{n} e^\top e$$

- Under $\mathcal{M}_\prime$ is simply

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

```
hat_sigma2 <- sum(residuals(lmFit)^2)/n
tilde_sigma2 <- var(Wage$wage)*(n-1)/n
```

Normal Linear Reg.: Inference on a Pool of Coefficients

A Special Case

- The test statistic is

$$F = \frac{\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{p-1}}{\frac{\hat{\sigma}^2}{n-p}} \overset{H_0}{\sim} F_{p-1,n-p}$$

- The p-value is given by $p - value = Pr(F_{p-1;n-p} > f^{obs})$

```
fval <- ((tilde_sigma2 - hat_sigma2)/(p-1))/((hat_sigma2)/(n-p))
fval
```

```
## [1] 177.0247
```

```
pf(fval, p-1, n-p, lower = FALSE)
```

```
## [1] 3.159188e-193
```

Normal Linear Reg.: Inference on a Pool of Coefficients

A Special Case

```
summary(lmFit)
```

```
##
## Call:
## lm(formula = wage ~ age + year + education, data = Wage)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -113.323  -19.521   -3.964   14.438  219.172
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -2.058e+03  6.493e+02  -3.169  0.00154 **
## age                          5.621e-01  5.714e-02   9.838  < 2e-16 ***
## year                         1.056e+00  3.238e-01   3.262  0.00112 **
## education2. HS Grad          1.140e+01  2.476e+00   4.603 4.34e-06 ***
## education3. Some College     2.423e+01  2.606e+00   9.301  < 2e-16 ***
## education4. College Grad     3.974e+01  2.586e+00  15.367  < 2e-16 ***
## education5. Advanced Degree  6.485e+01  2.804e+00  23.128  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.89 on 2993 degrees of freedom
## Multiple R-squared:  0.2619, Adjusted R-squared:  0.2604
## F-statistic:   177 on 6 and 2993 DF,  p-value: < 2.2e-16
```

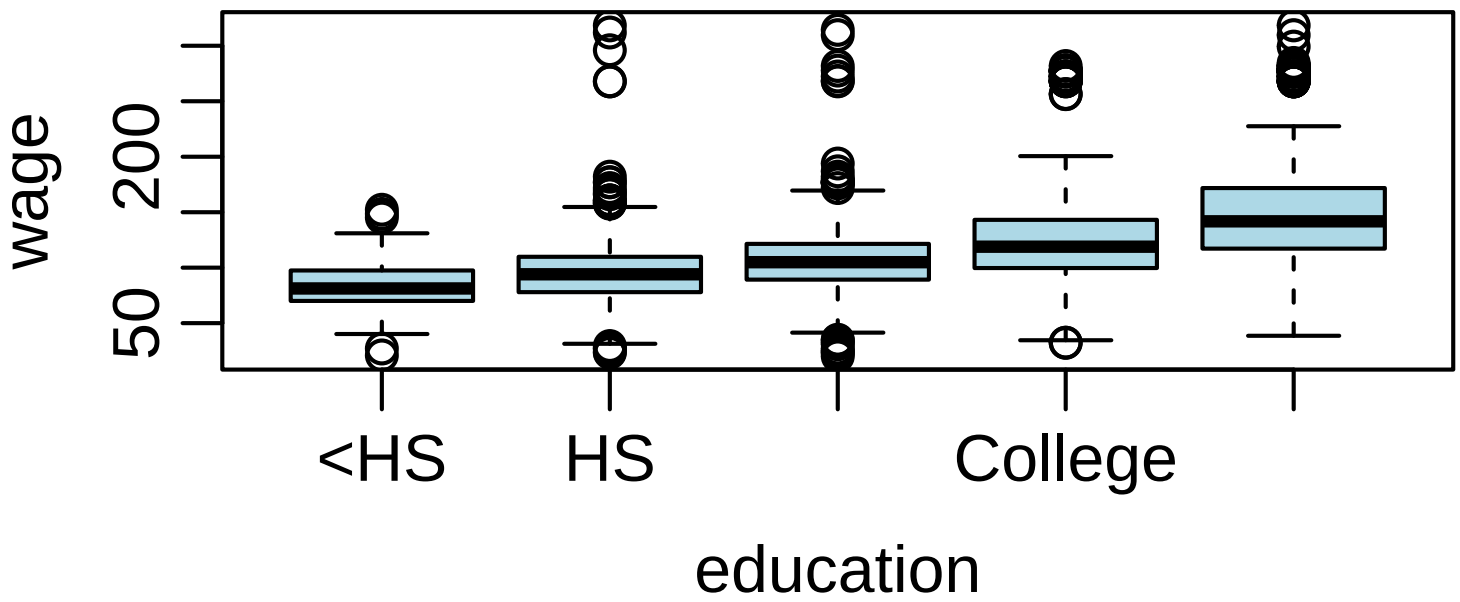
```
summary(lmFit)$fstatistic
```

```
##     value     numdf     dendf
##  177.0247    6.0000 2993.0000
```

Normal Linear Reg.: Inference on a Pool of Coefficients

Motivating question

- Does **education** matter for **wage**, once we account for age and year?
- `education` is a factor with 5 levels: it enters the model through $q = 4$ dummy variables, so the question concerns **four coefficients at once**
- The individual t-tests compare each level with the baseline, one at a time: they do not answer whether education matters as a whole



Normal Linear Reg.: Inference on a Pool of Coefficients

General case: Hypotheses on a subset of q coefficients

Partition $\beta = (\beta_A, \beta_B)$, where β_B contains the q coefficients under test:

$$\begin{cases} H_0 : \beta_B = 0 \\ H_1 : \text{at least one coefficient in } \beta_B \text{ is different from 0} \end{cases}$$

Full and reduced model

$$\text{Full model } (\mathcal{M}_\infty): \quad Y = X\beta + \varepsilon \quad (p \text{ parameters})$$

$$\text{Reduced model } (\mathcal{M}_0): \quad Y = X_A\beta_A + \varepsilon \quad (p - q \text{ parameters})$$

- $\mathcal{M}_0$ is obtained from $\mathcal{M}_\infty$ by dropping the q columns of X related to β_B (nested models)
- $\mathrm{RSS}_0 \geq \mathrm{RSS}_1$: a model with fewer variables cannot fit better

Normal Linear Reg.: Inference on a Pool of Coefficients

F test

Let $\hat{\sigma}^2 = \mathrm{RSS}_1/n$ under $\mathcal{M}_\infty$ and $\tilde{\sigma}^2 = \mathrm{RSS}_0/n$ under $\mathcal{M}_0$

$$F = \frac{(\mathrm{RSS}_0 - \mathrm{RSS}_1)/q}{\mathrm{RSS}_1/(n-p)} = \frac{(\tilde{\sigma}^2 - \hat{\sigma}^2)/q}{\hat{\sigma}^2/(n-p)} \overset{H_0}{\sim} F_{q,\, n-p}$$

- p-value: $\Pr(F_{q;\, n-p} > f^{obs})$
- If the dropped variables are useless, $\mathrm{RSS}_0 \approx \mathrm{RSS}_1$ and F is small; a large F means that removing them worsens the fit too much
- $q = 1$: a single coefficient, and $F = t^2$

Normal Linear Reg.: Inference on a Pool of Coefficients

```
fitFull <- lm(wage ~ age + year + education, data = Wage)
fitRed  <- lm(wage ~ age + year, data = Wage)    # without education

round(summary(fitFull)$coefficients, 3)
```

```
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  -2058.077    649.346  -3.169    0.002
## age                              0.562      0.057   9.838    0.000
## year                             1.056      0.324   3.262    0.001
## education2. HS Grad             11.398      2.476   4.603    0.000
## education3. Some College        24.234      2.606   9.301    0.000
## education4. College Grad        39.741      2.586  15.367    0.000
## education5. Advanced Degree     64.854      2.804  23.128    0.000
```

Normal Linear Reg.: Inference on a Pool of Coefficients

```r
n_obs  <- nrow(Wage)
p_full <- length(coef(fitFull))
q      <- p_full - length(coef(fitRed))
c(n = n_obs, p = p_full, q = q)
```

```
##    n    p    q
## 3000    7    4
```

```r
hat_sigma2   <- sum(residuals(fitFull)^2)/n_obs
tilde_sigma2 <- sum(residuals(fitRed)^2)/n_obs

Fobs <- ((tilde_sigma2 - hat_sigma2)/q)/(hat_sigma2/(n_obs - p_full))
Fobs
```

```
## [1] 223.3169
```

```r
pf(Fobs, q, n_obs - p_full, lower.tail = FALSE)
```

```
## [1] 6.24877e-168
```

Normal Linear Reg.: Inference on a Pool of Coefficients

anova(reduced model, full model)

- `Res.Df`: residual degrees of freedom, $n - (p - q)$ and $n - p$
- `Df`: number of coefficients tested, q
- `Sum of Sq`: $\mathrm{RSS}_0 - \mathrm{RSS}_1$

```
anova(fitRed, fitFull)
```

```
## Analysis of Variance Table
##
## Model 1: wage ~ age + year
## Model 2: wage ~ age + year + education
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1   2997 5004607
## 2   2993 3854286  4   1150320 223.32 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The p-value is essentially zero: we reject H_0 and conclude that education is associated with wage, given age and job class