Estimativa de σ^2 e Teorema de Gauss-Markov

Renato

ESTIMATIVA de σ^2

(a)

$$\sigma^2 = Var(\varepsilon_i) = Var(y_i \mid \underset{\sim}{x}_i) = \mathbb{E}\left(\left(y_i - \mathbb{E}(y_i \mid \underset{\sim}{x}_i)\right)^2\right)$$

~~Se~~ Se o modelo de regressão linear está correto, devemos ter $\mathbb{E}(y_i \mid \underset{\sim}{x}_i) = \underset{\sim}{x}_i' \cdot \underset{\sim}{\beta} \approx \underset{\sim}{x}_i' \cdot \hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}$

Isto é, podemos olhar para o resíduo

$$r_i = y_i - \hat{y}_i = y_i - \underset{\sim}{x}_i' \cdot \hat{\beta}$$

Como uma ~~aproximaç~~ para o erro

$$\varepsilon_i = y_i - \mathbb{E}(y_i \mid \underset{\sim}{x}_i)$$

Para obter um valor aproximado para

$$\sigma^2 = \mathbb{E}\left(y_i - \mathbb{E}(y_i \mid \underset{\sim}{x_i})\right)^2 .$$

podemos usar

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 =$$ Média dos resíduos (ao quadrado)

~~Na~~ $\hat{\sigma}^2$ é um valor aleatório (uma variável aleatória)

Portanto, $\hat{\sigma}^2$ possui uma distribuição de probabilidade, tem $\mathbb{E}(\hat{\sigma}^2)$ e também $\mathrm{Var}(\hat{\sigma}^2)$.

Além disso, $\boxed{\sigma^2 \neq \hat{\sigma}^2}$

por saídas

(c)

Na verdade, usamos uma ligeira modificação:

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

Se $p << n$, teremos $\frac{1}{n-(p+1)} \approx \frac{1}{n}$

Por exemplo, no caso do Cement dataset, $n = 1030$ e $p = 8$

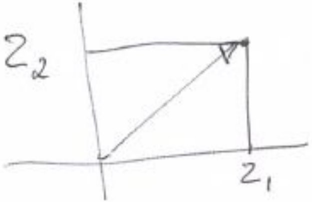
$$\Rightarrow 0.0009794319 = \frac{1}{1030-(8+1)} \approx \frac{1}{1030} = 0.0009708738$$

~~Para obter~~ A distribuição de $\hat{\sigma}^2$:

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = \|\underset{\sim}{Y}-\hat{\underset{\sim}{Y}}\|^2 \sim \chi^2_{n-(p+1)}$$ graus de liberdade

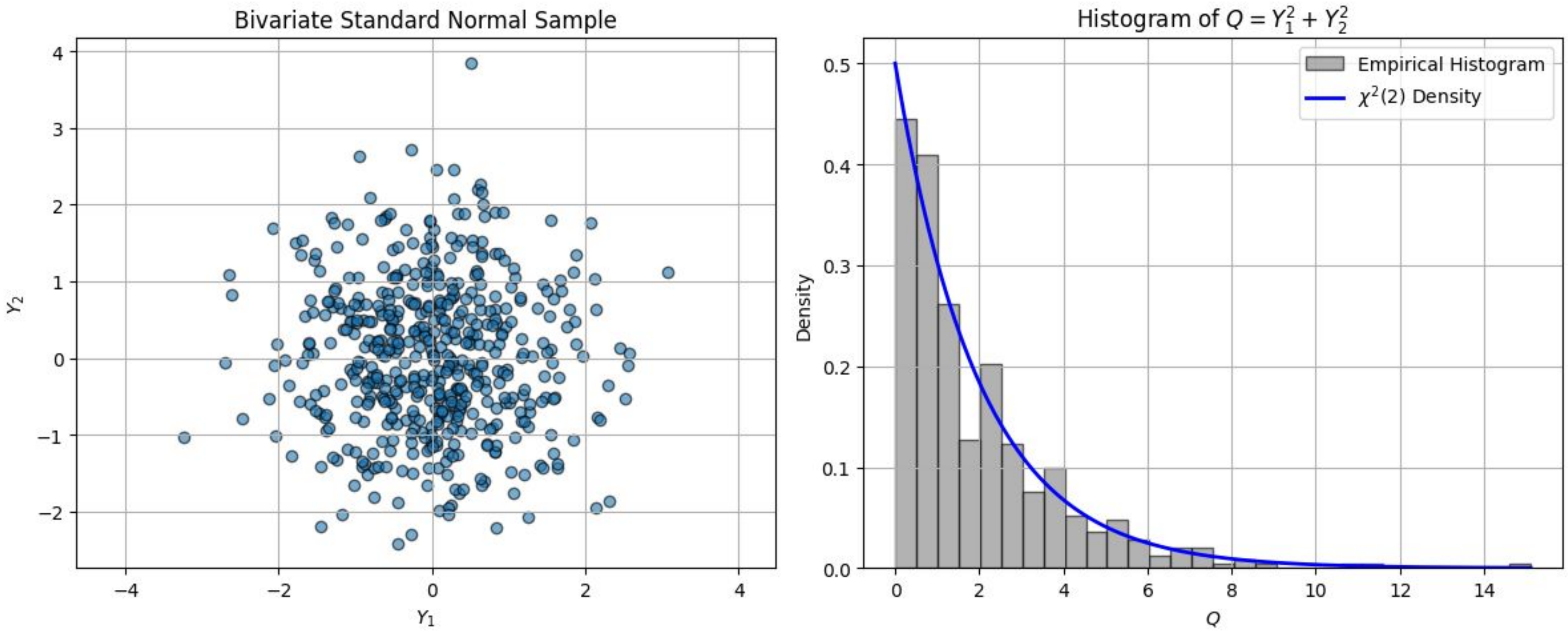
O que é uma v.a. χ^2 ??

Por definição, χ^2_k = ~~comprimento~~ soma de k $N(0,1)^2$ indep.

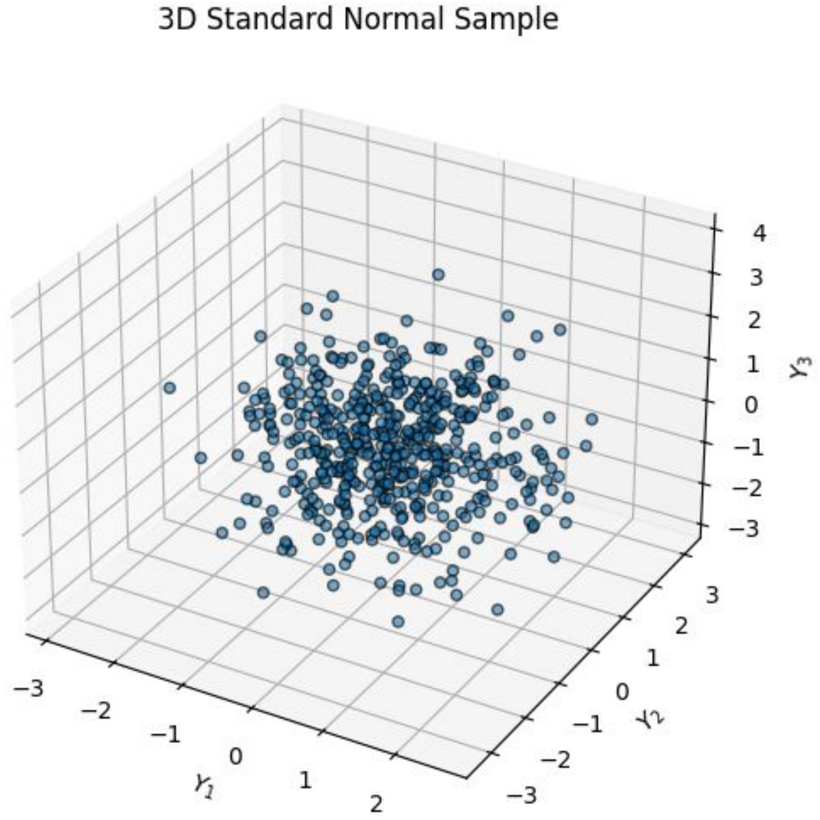


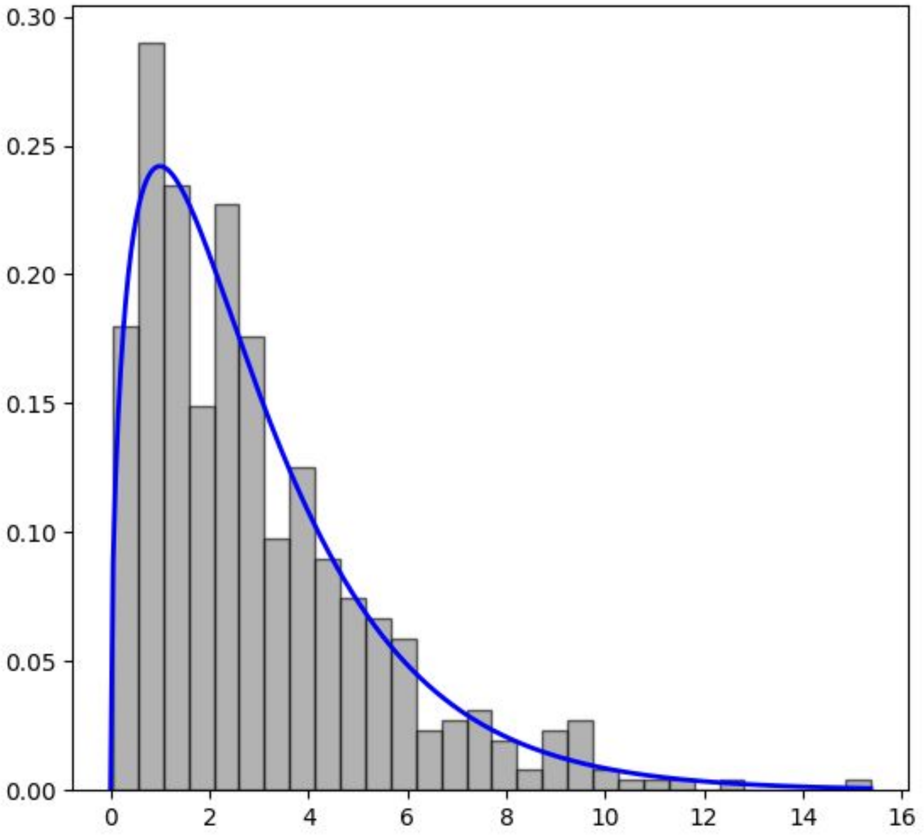
$Z_1 \sim N(0,1)$
$Z_2 \sim N(0,1)$ > indep.

$$Z_1^2 + Z_2^2 \sim \chi^2_2$$



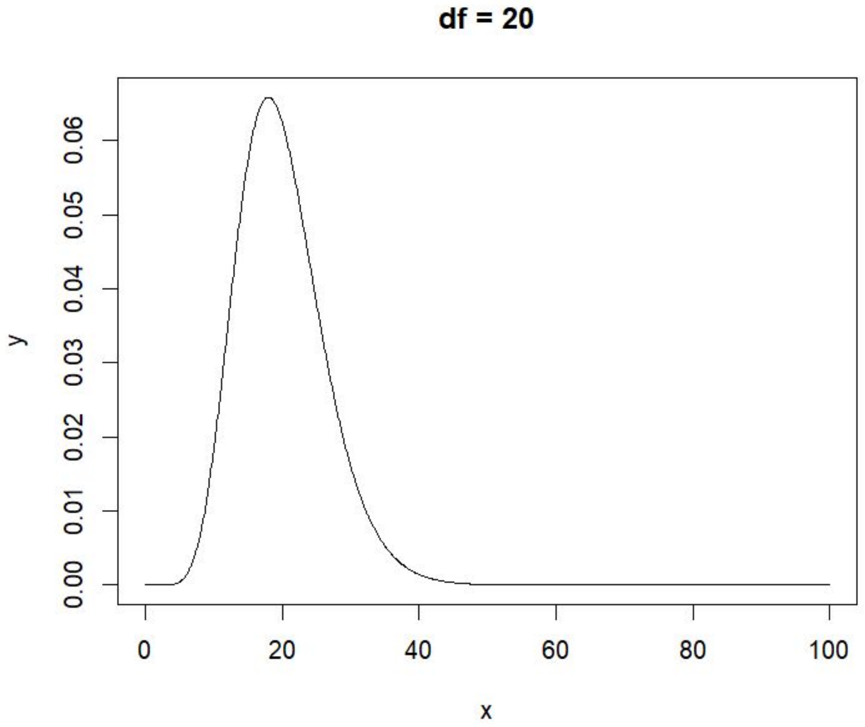
Bivariate Standard Normal Sample
Y_1
Y_2
Histogram of $Q = Y_1^2 + Y_2^2$
Empirical Histogram
$\chi^2(2)$ Density
Density
Q

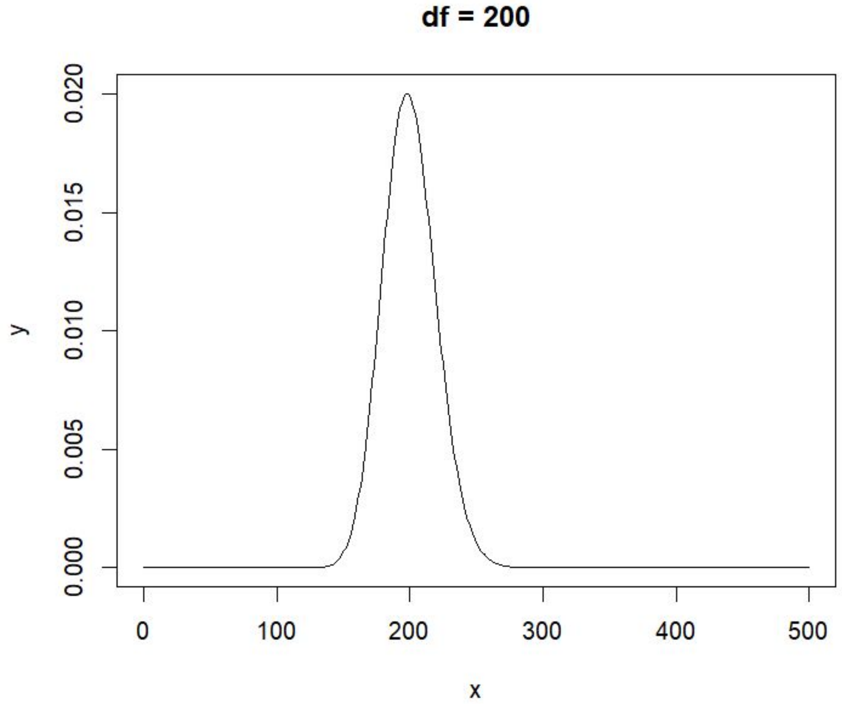

3D Standard Normal Sample
Y_1
Y_2
Y_3

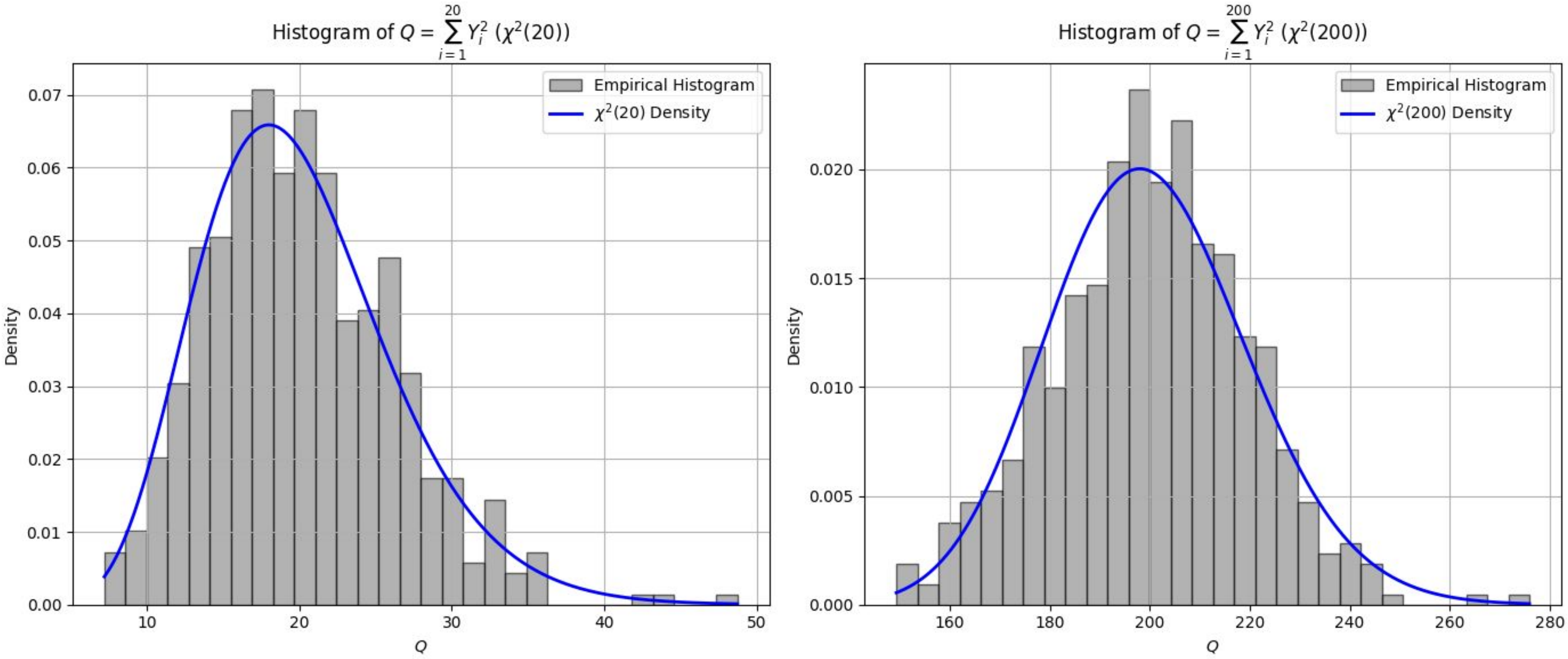


```python
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from scipy.stats import chi2

# Set seed for reproducibility
np.random.seed(42)

# Step 1: Generate 500 samples from 3D standard normal
n_samples = 500
mean = [0, 0, 0]
cov = np.identity(3)
Y = np.random.multivariate_normal(mean, cov, size=n_samples)
Y1, Y2, Y3 = Y[:, 0], Y[:, 1], Y[:, 2]

# Step 2: Compute Q = Y1^2 + Y2^2 + Y3^2
Q = Y1**2 + Y2**2 + Y3**2

# Step 3: Create subplots
fig = plt.figure(figsize=(14, 6))

# Left-hand side: 3D scatter plot
ax1 = fig.add_subplot(1, 2, 1, projection='3d')
ax1.scatter(Y1, Y2, Y3, alpha=0.6, edgecolor='k')
ax1.set_title("3D Standard Normal Sample")
ax1.set_xlabel("$Y_1$")
ax1.set_ylabel("$Y_2$")
ax1.set_zlabel("$Y_3$")

# Right-hand side: Histogram of Q with chi-squared(3) density
ax2 = fig.add_subplot(1, 2, 2)
x_vals = np.linspace(0, max(Q), 300)
ax2.hist(Q, bins=30, density=True, alpha=0.6, color='gray', edgecolor='black', label="Empirical Histogram")
ax2.plot(x_vals, chi2.pdf(x_vals, df=3), 'b-', linewidth=2, label=r'$\chi^2(3)$ Density')
ax2
```

Qui-quadrado: densidade

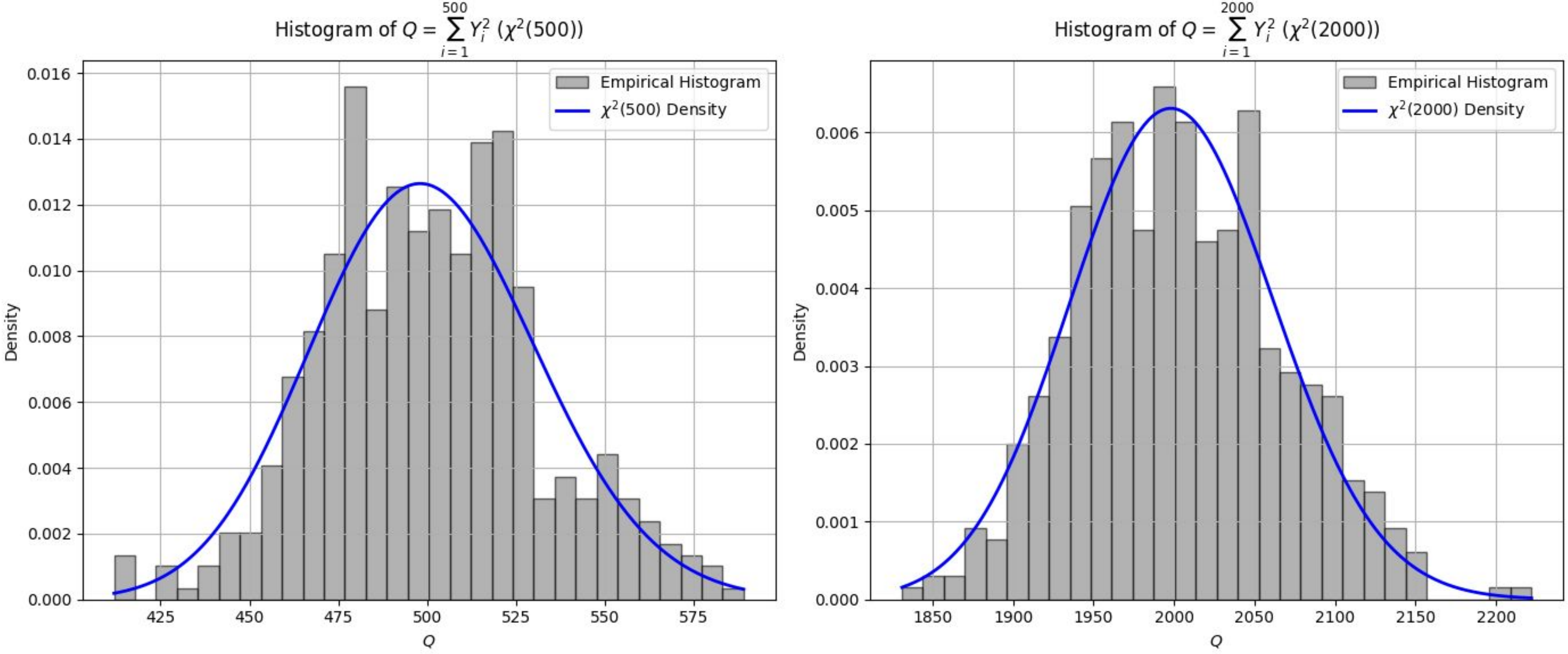






Histogram of $Q = \sum_{i=1}^{20} Y_i^2$ ($\chi^2(20)$)
Empirical Histogram
$\chi^2(20)$ Density
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
10
20
30
40
50
Q
Histogram of $Q = \sum_{i=1}^{200} Y_i^2$ ($\chi^2(200)$)
Empirical Histogram
$\chi^2(200)$ Density
Density
0.000
0.005
0.010
0.015
0.020
160
180
200
220
240
260
280
Q

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import chi2

# Set seed for reproducibility
np.random.seed(42)

# Define simulation parameters
n_samples = 500
dfs = [20, 200]  # Degrees of freedom

# Prepare figure
fig, axes = plt.subplots(1, 2, figsize=(14, 6))

for i, df in enumerate(dfs):
    # Step 1: Generate samples from multivariate standard normal
    Y = np.random.normal(0, 1, size=(n_samples, df))

    # Step 2: Compute Q = sum of squares (chi-squared variable)
    Q = np.sum(Y**2, axis=1)

    # Step 3: Plot histogram with chi-squared density
    ax = axes[i]
    ax.hist(Q, bins=30, density=True, alpha=0.6, color='gray', edgecolor='black', label='Empirical Histogram')

    x_vals = np.linspace(min(Q), max(Q), 300)
    ax.plot(x_vals, chi2.pdf(x_vals, df), 'b-', linewidth=2, label=fr'$\chi^2({df})$ Density')

    ax.set_title(fr"Histogram of $Q = \sum_{{i=1}}^{{{df}}} Y_i^2$ ($\chi^2({df})$)")
    ax.set_xlabel("$Q$")
    ax.set_ylabel("Density")
    ax.legend()
    ax.grid(True)

plt.tight_layout()
plt.show()
```



Histogram of $Q = \sum_{i=1}^{500} Y_i^2$ ($\chi^2(500)$)
Empirical Histogram
$\chi^2(500)$ Density
Density
Q
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0.014
0.016
425
450
475
500
525
550
575
Histogram of $Q = \sum_{i=1}^{2000} Y_i^2$ ($\chi^2(2000)$)
Empirical Histogram
$\chi^2(2000)$ Density
Density
Q
0.000
0.001
0.002
0.003
0.004
0.005
0.006
1850
1900
1950
2000
2050
2100
2150
2200

Chi-squared distribution

- If Q ~ Chi-squared with k degrees of freedom, then

1. E(Q) = k
2. Var(Q) = 2k → DP(Q) = sqrt(2k)

When k increases, the density becomes more concentrated around E(Q):

DP(Q)/E(Q) = sqrt(2k) / k = sqrt(2/ k) → 0 when k increases

$\hat{\sigma}^2$ é baseado em $\|\underset{\sim}{r}\|^2 = \|\underset{\sim}{Y} - \hat{\underset{\sim}{Y}}\|^2$ (F)

Contraste: $\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$

Versus $\underset{\sim}{Y} = X\hat{\underset{\sim}{\beta}} + \underset{\sim}{r} \quad \left(= \hat{\underset{\sim}{Y}} + (\underset{\sim}{Y} - \hat{\underset{\sim}{Y}})\right)$

$\underset{\sim}{\varepsilon} \neq \underset{\sim}{r}$

não observado

vetor aleatório conhecido

Quais as diferenças entre $\underset{\sim}{\varepsilon}$ e $\underset{\sim}{r}$??

$\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$

$\underset{\sim}{r} \sim$??

$$\underset{\sim}{r} = \underset{\sim}{Y} - \hat{\underset{\sim}{Y}} = \underset{\sim}{Y} - H\underset{\sim}{Y} = (I-H)\underset{\sim}{Y}$$

$$\underset{\sim}{Y} \sim N_n(X\underset{\sim}{\beta}, \sigma^2 I_n) \Rightarrow (I-H)\underset{\sim}{Y} \sim N_n\Big(\underbrace{(I-H)X\underset{\sim}{\beta}}_{(*)}, \underbrace{\sigma^2 (I-H) I_n (I-H)'}_{(**)}\Big)$$

$$(*)\ (I-H)X\underset{\sim}{\beta} = (X - HX)\cdot\underset{\sim}{\beta} = \left(X - X(X'X)^{-1}X'X\right)\underset{\sim}{\beta} = (X-X)\cdot\underset{\sim}{\beta} = \underset{\sim}{0}$$

$$(**) = \sigma^2 (I-H)(I-H)' \underset{\text{simétrica}}{=} \sigma^2 (I-H)(I-H) \underset{\text{idempotente: } (I-H)^2 = (I-H)}{=} \sigma^2 (I-H)$$

Assim, $\begin{cases} \underset{\sim}{r} \sim N_n(\underset{\sim}{0}, \sigma^2 \underbrace{(I_n - H)}_{\text{matriz "cheia"}}) \\ \underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n) \end{cases}$ enquanto

(0)

$$\begin{cases} \underset{\sim}{r} \sim N_n\left(\underset{\sim}{0},\ \sigma^2(I-H)\right) \\ \underset{\sim}{\varepsilon} \sim N_n\left(\underset{\sim}{0},\ \sigma^2 I\right) \end{cases}$$

Então $\frac{1}{\sigma^2}\|\underset{\sim}{\varepsilon}\|^2 = \frac{1}{\sigma^2}\underset{\sim}{\varepsilon}'\underset{\sim}{\varepsilon} = \underbrace{\sum_{i=1}^{n}\left(\frac{\varepsilon_i}{\sigma}\right)^2}_{\text{soma de } (N(0,1))^2 \text{ independentes}}$

$$\Longrightarrow \frac{1}{\sigma^2}\|\underset{\sim}{\varepsilon}\|^2 \sim \chi^2_n$$

$\boxed{\text{não-observado}} \Longrightarrow \boxed{\|\underset{\sim}{\varepsilon}\|^2 \sim \sigma^2 \chi^2_n}$

$\underset{\sim}{r} \sim N\left(0, \sigma^2 (I_n - H)\right)$

~~[illegible]~~ $\|\underset{\sim}{r}\|^2 = \underset{\sim}{r}' \underset{\sim}{r} = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$

<u>Teorema</u> No modelo de regressão linear,

$\frac{1}{\sigma^2} \|\underset{\sim}{r}\|^2 \sim \chi^2$ com degrees of freedom igual à dimensão do espaço vetorial $\underbrace{\mathbb{R}^n \ominus \mathcal{M}(X)}$

(espaço $\perp$ a $\mathcal{M}(X)$)

observado

Isto é, $\boxed{\frac{1}{\sigma^2} \|\underset{\sim}{r}\|^2 \sim \chi^2_{n-(p+1)}}$

(2)

Como $\mathbb{E}\left(\chi^2_k\right) = k$, temos que

$$\mathbb{E}\left(\frac{1}{\sigma^2}\|\underset{\sim}{r}\|^2\right) = \mathbb{E}\left(\chi^2_{n-(p+1)}\right) = \underbrace{n-(p+1)}$$

n.º de observ. − (n.º de features + 1)

n.º de obs. − n.º de colunas em X

Então

$$\mathbb{E}\left(\frac{\|\underset{\sim}{r}\|^2}{n-p-1}\right) = \sigma^2$$

$\Longrightarrow$ Usamos que $\dfrac{\|\underset{\sim}{r}\|^2}{n-p-1}$ deve ser $\approx \sigma^2$

③

Isto é,

$$\sigma^2 \approx \hat{\sigma}^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} \underbrace{(y_i - \hat{y}_i)^2}_{r_i^2}$$

$\boxed{\text{esperamos isto}}$

Veja que somamos $\underline{\underline{n}}$ termos r_i^2 mas não dividimos por $\underline{\underline{n}}$, e sim por $n-(p+1)$

```python
#generate OLS regression results for all features
import statsmodels.api as sm

X_sm = sm.add_constant(X)
model = sm.OLS(y,X_sm)
print(model.fit().summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 csMPa   R-squared:                       0.616
Model:                           OLS   Adj. R-squared:                  0.613
Method:                Least Squares   F-statistic:                     204.3
Date:               Fri, 15 Oct 2021   Prob (F-statistic):          6.29e-206
Time:                       16:43:15   Log-Likelihood:                -3869.0
No. Observations:               1030   AIC:                             7756.
Df Residuals:                   1021   BIC:                             7800.
Df Model:                          8
Covariance Type:           nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const              -23.3312     26.586     -0.878      0.380     -75.500      28.837
cement               0.1198      0.008     14.113      0.000       0.103       0.136
slag                 0.1039      0.010     10.247      0.000       0.084       0.124
flyash               0.0879      0.013      6.988      0.000       0.063       0.113
water               -0.1499      0.040     -3.731      0.000      -0.229      -0.071
superplasticizer     0.2922      0.093      3.128      0.002       0.109       0.476
coarseaggregate      0.0181      0.009      1.926      0.054      -0.000       0.037
fineaggregate        0.0202      0.011      1.887      0.059      -0.001       0.041
age                  0.1142      0.005     21.046      0.000       0.104       0.125
==============================================================================
```

A soma dos quadrados dos resíduos não aparece neste output

```python
#generate OLS regression results for all features
import statsmodels.api as sm

X_sm = sm.add_constant(X)
model = sm.OLS(y,X_sm)
print(model.fit().summary())
```

$$DP(\hat{\beta}_j) = \sqrt{\mathbb{V}(\hat{\beta}_j)} = \hat{\sigma}\sqrt{(\mathbf{X}'\mathbf{X})^{-1}[jj]}$$

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 csMPa   R-squared:                       0.616
Model:                           OLS   Adj. R-squared:                  0.613
Method:                Least Squares   F-statistic:                     204.3
Date:               Fri, 15 Oct 2021   Prob (F-statistic):          6.29e-206
Time:                       16:43:15   Log-Likelihood:                -3869.0
No. Observations:               1030   AIC:                             7756.
Df Residuals:                   1021   BIC:                             7800.
Df Model:                          8
Covariance Type:           nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const              -23.3312     26.586     -0.878      0.380     -75.500      28.837
cement               0.1198      0.008     14.113      0.000       0.103       0.136
slag                 0.1039      0.010     10.247      0.000       0.084       0.124
flyash               0.0879      0.013      6.988      0.000       0.063       0.113
water               -0.1499      0.040     -3.731      0.000      -0.229      -0.071
superplasticizer     0.2922      0.093      3.128      0.002       0.109       0.476
coarseaggregate      0.0181      0.009      1.926      0.054      -0.000       0.037
fineaggregate        0.0202      0.011      1.887      0.059      -0.001       0.041
age                  0.1142      0.005     21.046      0.000       0.104       0.125
```

Teorema de Gauss-Markov

https://pt.wikipedia.org/wiki/Andrei_Markov

Recapitulando:

Predizer $\underset{\sim}{Y}$: problema de interesse

~~...~~ **Abordagem**: usar preditor linear

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Visto como problema numérico:

solução é $\hat{\underset{\sim}{\beta}} = \underbrace{(X'X)^{-1}X'\underset{\sim}{Y}}_{A\underset{\sim}{Y}} = \arg\min_{\beta} \|\underset{\sim}{Y} - X\underset{\sim}{\beta}\|^2$

Visto como problema de álgebra linear:

$\hat{\underset{\sim}{Y}} = X\hat{\underset{\sim}{\beta}} =$ projeção $\perp$ de $\underset{\sim}{Y}$ em $\mathcal{M}(X) =$ sub-espaço vetorial das colunas de X

$\hat{Y} = H\underset{\sim}{Y}$ $\quad H =$ hat matrix

Adotamos uma visão estocástica: suponha que exista um β verdadeiro fixo e que os ~~modelo~~ dados sejam realmente gerados pelo modelo.

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \quad \text{com } \underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$$

Então: $$\underset{(p+1)\times 1}{\hat{\underset{\sim}{\beta}}} = \underbrace{(X'X)^{-1}X'}_{A}\underset{\sim}{Y} \sim N_{p+1}\left(\underset{\sim}{\beta}, \sigma^2 (X'X)^{-1}\right)$$

Consequência:

$$\begin{cases} E(\underbrace{\hat{\underset{\sim}{\beta}}}_{\text{aleatório}}) = \overbrace{\underset{\sim}{\beta}}^{\text{verdadeiro, desconhecido}} \rightsquigarrow \hat{\underset{\sim}{\beta}} \text{ é não-viciado para estimar } \underset{\sim}{\beta} \\[2ex] Var(\hat{\underset{\sim}{\beta}}) = \sigma^2 (X'X)^{-1} \end{cases}$$

Usamos $$\hat{\sigma}^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$ para estimar σ^2

Teorema de Gauss-Markov

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{Y} = A\underset{\sim}{Y}$$

é a "melhor maneira" de estimar o vetor $\underset{\sim}{\beta}$ desconhecido

Uma "melhor maneira": abordagem numérica

$$\hat{\underset{\sim}{\beta}} = \arg\min_{\underset{\sim}{\beta}} \|\underset{\sim}{Y} - X\underset{\sim}{\beta}\|^2$$

Outra "melhor maneira": se existe um $\underset{\sim}{\beta}$ verdadeiro e fixo, então $\hat{\underset{\sim}{\beta}}$ é o que menos erra ao estimar este $\underset{\sim}{\beta}$ desconhecido.

Os próximos passos/slides

- Entender mais precisamente o que significa “menor erro” para estimar beta
- Como medir isso?
- Beta é um VETOR
- E é um vetor ALEATÓRIO

- Como minimizar o erro de estimação?
- O Teorema de Gauss-Markov e sua demonstração

A solução de mínimos quadrados (LS) (Least Squares) é ①

$$\underset{(p+1)\times 1}{\hat{\underset{\sim}{\beta}}_{LS}} = \underbrace{(X'X)^{-1}X'}_{(p+1)\times n} \underset{n\times 1}{\underset{\sim}{Y}} = A \cdot \underset{\sim}{Y}$$

$$\text{e } \hat{\underset{\sim}{\beta}}_{LS} \sim N_{p+1}\left(\underset{\sim}{\beta}, \ \sigma^2 (X'X)^{-1}\right)$$

verdadeiro
e desconhecido

Seja $\underset{(p+1)\times 1}{\hat{\underset{\sim}{\beta}}_{*}} = \underset{(p+1)\times n}{C} \ \underset{n\times 1}{\underset{\sim}{Y}}$ um outro candidato para estimar o verdadeiro $\underset{\sim}{\beta}$ que seja <u>melhor</u> que

$\hat{\underset{\sim}{\beta}}_{LS} = A \cdot \underset{\sim}{Y}$ onde $\underset{(p+1)\times n}{A} = (X'X)^{-1}X'$

Vamos mostrar que **outro** candidato melhor que ②
$\hat{\beta}_{LS}$ **não existe**.

Como a matriz C de $\hat{\beta}_* = C\underset{\sim}{Y}$ deve ser diferente da matriz A de $\hat{\beta}_{LS} = \underbrace{(X'X)^{-1}X'}_{A}\underset{\sim}{Y}$

Vamos escrever $C = A + D = (X'X)^{-1}X' + D$

Uma propriedade de $\hat{\underset{\sim}{\beta}}_{LS} = A\underset{\sim}{Y}$ é ser

não-viciado: $\mathbb{E}(\hat{\underset{\sim}{\beta}}^{LS}) = \underset{\sim}{\beta}$ para todo $\underset{\sim}{\beta}$ possível

Esta é uma propriedade desejável de um estimador e vamos querer que o candidato $\hat{\underset{\sim}{\beta}}_{*} = C\underset{\sim}{Y}$ também seja não-viciado.

Isto implica que a matriz $C = A + D = (X'X)^{-1}X' + D$ deve satisfazer $DX = \underset{(p+1)\times(p+1)}{0}$

Para ver que isto é verdade, vamos calcular:

$$\underset{\sim}{\beta} = \mathbb{E}(\hat{\underset{\sim}{\beta}}_*) = \mathbb{E}(C\underset{\sim}{Y}) = C\,\mathbb{E}(\underset{\sim}{Y}) = CX\underset{\sim}{\beta} \quad (4)$$

$$= \left((X'X)^{-1}X' + D\right)X\underset{\sim}{\beta} = \left((X'X)^{-1}X'\cancel{X} + DX\right)\cdot\underset{\sim}{\beta}$$

$$= (I + DX)\underset{\sim}{\beta} = \underset{\sim}{\beta} + DX\underset{\sim}{\beta}$$

Isto é, devemos ter $\cancel{\underset{\sim}{\beta}} = \cancel{\underset{\sim}{\beta}} + DX\underset{\sim}{\beta}$ para todo $\underset{\sim}{\beta}$

$$\Rightarrow DX\underset{\sim}{\beta} = \underset{(p+1)\times 1}{0} \quad \text{para todo } \underset{\sim}{\beta}$$

$$\Rightarrow DX = \underset{(p+1)\times(p+1)}{0}$$

Portanto, estamos buscando $\hat{\underset{\sim}{\beta}}_* = \left((X'X)^{-1}X' + D\right)\underset{\sim}{Y}$ ⑤

$= (A + D)\cdot\underset{\sim}{Y}$

que seja "melhor" que $\hat{\underset{\sim}{\beta}}_{LS} = (X'X)^{-1}X'Y = A\underset{\sim}{Y}$

⊕ O que significa "ser melhor"? Significa ter um erro de estimação esperado menor. Significa um estimador que erra menos. Isto é, o erro de estimação dos coeficientes

$$\hat{\underset{\sim}{\beta}}_* - \underset{\sim}{\beta} = \begin{bmatrix} \hat{\beta}_0^* - \beta_0 \\ \vdots \\ \hat{\beta}_p^* - \beta_p \end{bmatrix}$$ deve ser menor que o

erro de estimação cometido por $\hat{\underset{\sim}{\beta}}_{LS}$ que é

$$\hat{\beta}_{LS} - \underset{\sim}{\beta} = \begin{bmatrix} \hat{\beta}_0^{LS} - \beta_0 \\ \vdots \\ \hat{\beta}_p^{LS} - \beta_p \end{bmatrix}$$

Mas como avaliar isto? (6)

β é desconhecido e arbitrário: não podemos calcular os erros diretamente.

Além disso, os erros de estimação são <u>vetores</u> aleatórios.

Vamos considerar uma <u>coordenada genérica</u> de β e comparar os <u>tamanhos típicos</u> dos erros de estimação $(\hat{\beta}_j^* - \beta_j)$ e $(\hat{\beta}_j^{LS} - \beta_j)$

Qual o tamanho típico desses erros? (7)

Ele pode ser medido olhando-se o **valor esperado** dos erros (ao quadrado):

$\mathbb{E}(\hat{\beta}_j^{LS} - \beta_j)^2$. Mas como $\hat{\beta}^{LS}$ é não-viciado para estimar β, temos $\mathbb{E}(\hat{\beta}_j^{LS}) = \beta_j$

Portanto,

$$\mathbb{E}(\hat{\beta}_j^{LS} - \beta_j)^2 = \mathbb{E}\left(\hat{\beta}_j^{LS} - \mathbb{E}(\hat{\beta}_j^{LS})\right)^2 = \mathbb{V}ar(\hat{\beta}_j^{LS})$$

Mas já sabemos que $\hat{\beta}^{LS} \sim N_{p+1}\left(\beta, \sigma^2 (X'X)^{-1}\right)$

Assim, $\mathbb{V}ar(\hat{\beta}_j^{LS}) = \sigma^2 (X'X)^{-1}[jj] \longrightarrow$ elemento $[jj]$ na diagonal de $(X'X)^{-1}$

Olhar a diagonal de $\sigma^2(X'X)^{-1}$ fornece os tamanhos típicos dos erros de $\hat{\beta}^{LS}$ em cada coordenada. (8)

Como $\hat{\beta}^*$ também é não-viesado para estimar β (o que implica $DX=0$) então

$$\mathbb{E}\left(\hat{\beta}_j^* - \beta_j\right)^2 = \mathbb{E}\left(\hat{\beta}_j^* - \mathbb{E}(\hat{\beta}_j^*)\right)^2 = \mathbb{V}ar(\hat{\beta}_j^*)$$

O que é a matriz de variância-covariância de $\hat{\beta}^*$?

$$\underset{(p+1)\times(p+1)}{\mathbb{V}\text{ar}\left(\hat{\beta}^*\right)} = \mathbb{V}\text{ar}\left(C\underset{\sim}{Y}\right)$$

$$= C\,\mathbb{V}\text{ar}\left(\underset{\sim}{Y}\right)C' \qquad \text{(ver exerc. 2 de Lista}$$

$$= C\left(\sigma^2 I_n\right)C' = \sigma^2\, C C'$$

$$= \sigma^2\left((X'X)^{-1}X' + D\right)\left((X'X)^{-1}X' + D\right)'$$

$$= \sigma^2\left((X'X)^{-1}X' + D\right)\left(X(X'X)^{-1} + D'\right)$$

$$= \sigma^2\left((X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'D' + DX(X'X)^{-1} + DD'\right)$$

$$= \sigma^2\left((X'X)^{-1} + (X'X)^{-1}\underbrace{(DX)'}_{=0} + \underbrace{(DX)}_{=0}(X'X)^{-1} + DD'\right)$$

$$= \sigma^2(X'X)^{-1} + \sigma^2 DD'$$

$$= \underbrace{\mathbb{V}\text{ar}\left(\hat{\underset{\sim}{\beta}}^{LS}\right)}_{(p+1)\times(p+1)} + \sigma^2\underbrace{DD'}_{(p+1)\times(p+1)}$$

O elemento [jj] na diagonal de $\mathrm{Var}(\hat{\beta}^*)$ é positivo (pois é uma variância) e é igual ao elemento [jj] na diagonal de $\mathrm{Var}(\hat{\beta}^{LS})$ (também positivo) somado ao elemento [jj] da matriz DD'.

Basta mostrar agora que este elemento [jj] de DD' é positivo

(10)

Mas $[DD'][jj]$ = produto interno da linha j de D pela coluna j de D'

Mas coluna j de D' = linha j de D

$\Rightarrow [DD'][jj]$ = produto interno da linha j de D pela mesma linha j de D

= $(\text{comprimento})^2$ da linha de D

e isto é sempre $\underline{\geq 0}$

Concluímos que, em toda coordenada,

$$\underbrace{\mathbb{V}\mathrm{ar}\left(\hat{\beta}_j^{*}\right)}_{\substack{\text{(tamanho)}^2 \\ \text{típico do} \\ \text{erro de} \\ \text{estimação} \\ \text{de } \hat{\beta}^{*} \text{ na} \\ \text{coordenada } j.}} \geq \underbrace{\mathbb{V}\mathrm{ar}\left(\hat{\beta}_j^{LS}\right)}_{\substack{\text{(tamanho)}^2 \text{ típico} \\ \text{do erro de estimação} \\ \text{de } \hat{\beta}^{LS} \text{ na coordenada } j}}$$

<u>Ou seja</u>, se o modelo de regressão realmente gera os dados com um β verdadeiro, então o estimador linear da forma $\hat{\beta} = A\,\underset{\sim}{Y}$ com $A = (X'X)^{-1}X'$ é o que terá os <u>menores</u> $(\text{erros})^2$ de estimação <u>em média em todas as coordenadas</u>