COMPUTATIONAL FINANCE & RISK MANAGEMENT
UNIVERSITY *of* WASHINGTON
Department of Applied Mathematics

CRFM 541: Investment Science

6. The Normal Linear Regression Model

Kjell Konis
Acting Assistant Professor, Applied Mathematics
University of Washington

Outline

Outline

Why a Statistical Model?

- Last time, simulated data from the model

$$\mathbf{y} = \begin{bmatrix} 1 & x_1 & \sin(x_1) & x_1^2 & e^{x_1} \\ 1 & x_2 & \sin(x_2) & x_2^2 & e^{x_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \sin(x_n) & x_n^2 & e^{x_n} \end{bmatrix} \begin{bmatrix} 0 \\ 0.5 \\ 2 \\ 0 \\ 0 \end{bmatrix} + \boldsymbol{\varepsilon} \qquad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0,1)$$

- R code: simulating the data

```
> x <- runif(100, -4, 4)
> y <- 0.5*x + 2*sin(x) + rnorm(100)
```

- R code: fitting the model

```
> m1 <- lm(y ~ x + I(sin(x)) + I(x^2) + I(exp(x)))
> coef(m1)
(Intercept)           x   I(sin(x))      I(x^2)   I(exp(x))
    0.06147     0.50714     2.01958    -0.04822     0.02011
```

Why a Statistical Model?

```
> summary(m1)
Call:
lm(formula = y ~ x + I(sin(x)) + I(x^2) + I(exp(x)))

Residuals:
   Min     1Q Median     3Q    Max
-2.213 -0.712  0.124  0.691  2.151

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0615     0.1555    0.40   0.6935
x             0.5071     0.1744    2.91   0.0045
I(sin(x))     2.0196     0.2684    7.52    3e-11
I(x^2)       -0.0482     0.0587   -0.82   0.4134
I(exp(x))     0.0201     0.0376    0.53   0.5943

Residual standard error: 0.96 on 95 degrees of freedom
Multiple R-squared:  0.83, Adjusted R-squared:  0.823
F-statistic:  116 on 4 and 95 DF,  p-value: <2e-16
```

Why a Statistical Model?

- How many standard errors away from 0 is significant?
- Depends on the distribution of the statistic
- What is the probability of a *Six Sigma* (or worse) event?

$T \sim \mathcal{N}$
```
> 2 * pnorm(-6)
[1] 1.973e-09
```

$T \sim t_1$
```
> 2 * pt(-6, df = 1)
[1] 0.1051
```

- Ratio

```
> pt(-6, df = 1) / pnorm(-6)
[1] 53283109
```

Hunting for α

```
> summary(lm(Asset.returns ~ SP500.returns))

Call:
lm(formula = Asset.returns ~ SP500.returns)

Residuals:
     Min       1Q   Median       3Q      Max
-0.09001 -0.00891 -0.00030  0.00878  0.05954

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.002633   0.000984    2.67    0.008
SP500.returns 2.023243   0.123288   16.41   <2e-16

Residual standard error: 0.0155 on 248 degrees of freedom
Multiple R-squared:  0.521, Adjusted R-squared:  0.519
F-statistic:  269 on 1 and 248 DF,  p-value: <2e-16
```

Outline

Multivariate Normal Distribution

- A random vector $\mathbf{V} = \begin{bmatrix} V_1 & V_2 & \dots & V_m \end{bmatrix}^\intercal$ is a vector of random variables
- Let $\mathcal{F} = F(v_1, v_2, \dots, v_n)$ be the joint distribution of the random variables comprising the random vector $\mathbf{V}$
- Then we say that $\mathbf{V} \sim \mathcal{F}$
- A random vector $\boldsymbol{\varepsilon}$ in $\mathbb{R}^m$ has the *Multivariate Normal* distribution if and only if $\boldsymbol{\gamma}^\intercal \boldsymbol{\varepsilon}$ has the univariate normal distribution $\forall \boldsymbol{\gamma} \in \mathbb{R}^m$
- How to use this definition to determine the properties of the Multivariate Normal distribution?
- The moment generating function (MGF) of a random vector $\mathbf{V}$ is

  $$M_W(\theta) = \mathbb{E}\left[e^{\boldsymbol{\theta}^\intercal \mathbf{W}}\right] \qquad \boldsymbol{\theta} \in \mathbb{R}^m$$

  (when the expectation exists)
- The following properties can be proved using the Multivariate Normal MGF

Properties of Multivariate Normal Random Vectors

1. Moment generating function of $\mathbf{V} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$M_{\mathbf{V}}(\mathbf{u}) = \exp\left[\boldsymbol{\mu}^\intercal \mathbf{u} + \frac{1}{2}\mathbf{u}^\intercal \boldsymbol{\Sigma}\mathbf{u}\right]$$

2. Let $\mathbf{V} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, given an $n \times m$ matrix $\mathbf{B}$ and an $n \times 1$ vector $\mathbf{b}$

$$\mathbf{BV} + \mathbf{b} \sim \mathcal{N}_n(\mathbf{b} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\intercal)$$

3. $\mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ density (assuming $\boldsymbol{\Sigma}$ nonsingular)

$$f_{\mathbf{V}}(\mathbf{v}) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{v} - \boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{v} - \boldsymbol{\mu})\right]$$

4. Constant density isosurfaces are ellipsoidal

Properties of Multivariate Normal Random Vectors

5. Marginals of Gaussian are Gaussian (converse NOT true)
6. $\mathbf{\Sigma}$ diagonal $\iff$ independent
7. $\mathbf{V} \sim \mathcal{N}_m(\boldsymbol{\mu}, \mathbf{\Sigma})$

$$\mathbf{AV} \text{ independent of } \mathbf{BV} \iff \mathbf{A\Sigma B}^\intercal = \mathbf{0}$$

The χ^2 and F Distributions

- Let $\mathbf{Z} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I})$, then $\|\mathbf{Z}\|^2$ has the χ^2 *distribution with m degrees of freedom*

  [i.e. χ^2_m is the distribution of the sum of the squares of m independent standard normal random variables]

- Let $U \sim \chi^2_p$ and $W \sim \chi^2_q$ be independent random variables, then

  $$V \sim \frac{(U/p)}{(W/q)}$$

  has the F *distribution* with p and q degrees of freedom

Gaussian Quadratic Forms

1. If $\mathbf{Z} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I})$ and $\mathbf{H}$ is a projection of rank $r \leq m$, then

$$\mathbf{Z}^\top \mathbf{H} \mathbf{Z} \sim \chi^2_r$$

2. If $\mathbf{V} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ nonsingular, then

$$(\mathbf{V} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{V} - \boldsymbol{\mu}) \sim \chi^2_m$$

Outline

The Normal Linear Regression Model

The *Simple Linear Regression Model*

$$Y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i$$

The *Multiple Linear Regression Model*

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i$$

Notation:

- Y_i is called the *response* variable; a random quantity
- x_{ij} is called a *predictor* variable; a non-random quantity
- $\beta_0, \ldots, \beta_p$ are the *parameters* of the model
  Alternatively: $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \ldots & \beta_p \end{bmatrix}^\intercal$ is the *parameter* of the model
- $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ is called the *noise*, *disturbance*, or *error* term

The Normal Linear Regression Model

4 **Key Assumptions**

1. Linearity of the conditional expectation

$$E(Y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$$

2. Independent noise: $\varepsilon_1, \ldots, \varepsilon_n$ are jointly independent

3. Constant variance

$$\mathrm{Var}(Y_i) = \mathrm{Var}(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i) = \mathrm{Var}(\varepsilon_i) = \sigma^2$$

4. Normal noise: $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ for all i

Normal Linear Regression Model: Matrix Form

- Normal Linear Regression Model: Matrix Form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & \dots & x_{mp} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

- Assumptions 2, 3, and 4

$$\boldsymbol{\varepsilon} \sim \mathcal{N}_m(\mathbf{0}, \sigma^2 \mathbf{I})$$

- Location-scale property

$$\mathbf{Y} \sim \mathcal{N}_m(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

Outline

Method of Maximum Likelihood

- The joint density of a random sample from a parametric family with parameter θ is

$$f_{X_1,\ldots,X_m}(x_1,\ldots,x_m|\theta) = \prod_{i=1}^{m} f_X(x_i|\theta)$$

- Let $x_1,\ldots,x_m$ (the data) be the realization of a random sample $X_1,\ldots,X_m \overset{iid}{\sim} f_X(x|\theta)$
- The *likelihood* of the parameter θ is the function of θ defined by

$$L(\theta|x_1,\ldots,x_m) = f_{X_1,\ldots,X_m}(x_1,\ldots,x_m|\theta) = \prod_{i=1}^{m} f_X(x_i|\theta)$$

- The *maximum likelihood estimator* $\hat{\theta}_{ML}$ of a parameter θ is the value of θ corresponding to the largest likelihood possible, i.e.,

$$L(\hat{\theta}_{ML}|x_1,\ldots,x_m) \geq L(\theta|x_1,\ldots,x_m)$$

for all possible values of θ

Calculation of $\hat{\theta}_{ML}$

- Often, maximizing the log of the likelihood is easier than maximizing the likelihood itself

- General algorithm for finding $\hat{\theta}_{ML}$
  1. Write down the likelihood function $L(\theta|x_1, \ldots, x_m)$
  2. $\ell(\theta|x_1, \ldots, x_m) = \log L(\theta|x_1, \ldots, x_m)$ (the log likelihood of θ)
  3. Find $\hat{\theta}_{ML}$ such that

$$\frac{d}{d\theta}\ell(\theta|x_1, \ldots, x_m)\Big|_{\hat{\theta}_{ML}} = 0$$

  4. Verify that $\hat{\theta}_{ML}$ is a maximum

Example: Find $\hat{\mu}$ for a Normal Population

- Let $x_1, \ldots, x_m$ be the realization of a random sample $X_1, \ldots, X_m \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$
- Find the maximum likelihood estimator of μ

- Likelihood function

$$
\begin{aligned}
L(\mu, \sigma^2 | x_1, \ldots, x_m) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_m | \mu, \sigma^2) \\
&= \prod_{i=1}^{m} f_X(x_i | \mu, \sigma^2) \\
&= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right]
\end{aligned}
$$

Example: Find $\hat{\mu}$ for a Normal Population

- Compute the log likelihood function

$$\begin{aligned}
\ell(\mu, \sigma^2 | x_1, \ldots, x_m) &= \log L(\mu, \sigma^2 | x_1, \ldots, x_m) \\
&= \log \left[\prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right] \\
&= \sum_{i=1}^{m} \log \frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{m} \left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right] \\
&= -m \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^{m} (x_i - \mu)^2 \\
&\propto -\frac{1}{2\sigma^2} \sum_{i=1}^{m} (x_i - \mu)^2
\end{aligned}$$

Example: Find $\hat{\mu}$ for a Normal Population

- Take the derivative of the log likelihood

$$\begin{aligned}\frac{d}{d\mu}\ell(\mu, \sigma^2 | x_1, \ldots, x_m) &= \frac{d}{d\mu}\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(x_i - \mu)^2\right] \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{m}(x_i - \mu)\end{aligned}$$

- Set the derivative equal to 0 and solve for $\hat{\mu}_{ML}$

$$\frac{1}{\sigma^2}\sum_{i=1}^{m}(x_i - \hat{\mu}_{ML}) \overset{\text{set}}{=} 0$$

$$\sum_{i=1}^{m} x_i = m\hat{\mu}_{ML} \quad \Rightarrow \quad \hat{\mu}_{ML} = \frac{1}{m}\sum_{i=1}^{m} x_i = \bar{x}$$

Example: Find $\hat{\mu}$ for a Normal Population

- Last, verify that $\hat{\mu}_{ML}$ is a maximum

$$\frac{d^2}{d\mu^2}\ell(\mu, \sigma^2 | x_1, \ldots, x_m) = \frac{d}{d\mu}\left[\frac{1}{\sigma^2}\sum_{i=1}^{m}(x_i - \mu)\right]$$
$$= -\frac{m}{\sigma^2}$$

- Since

$$\left.\frac{d^2\ell}{d\mu^2}\right|_{\mu=\hat{\mu}_{ML}} = -\frac{m}{\sigma^2} < 0$$

$\hat{\mu}_{ML}$ is a local maximum of the log likelihood function

Outline

Normal Linear Regression Model: Likelihood

- Model: $\mathbf{Y} \sim \mathcal{N}_m(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$
- By property 3 (of multivariate normal random vectors)

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right]$$

- Likelihood

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left[-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right]$$

- Loglikelihood

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2}\left[m\log(2\pi) + m\log(\sigma^2) + \frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right]$$

Maximum Likelihood Estimation: $\boldsymbol{\beta}$

- For any value of $\sigma^2 > 0$, the loglikelihood is maximum when $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ is minimum

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}}\left\{-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\} = \arg\min_{\boldsymbol{\beta}}\left\{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}$$

- Calculus solution

$$\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\beta}}\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\intercal(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]\Bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} &\overset{\text{set}}{=} 0 \\
-2\mathbf{X}^\intercal(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= 0 \\
\mathbf{X}^\intercal\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}^\intercal\mathbf{y} \\
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}
\end{aligned}$$

Maximum Likelihood Estimation: σ^2

- The MLE for σ^2 can be obtained by profile likelihood

$$\begin{aligned}\hat{\sigma}^2 &= \arg\max_{\sigma^2} \left\{ \arg\max_{\beta} \ell(\beta, \sigma^2) \right\} \\ &= \arg\max_{\sigma^2} \ell(\hat{\beta}, \sigma^2) \\ &= \arg\max_{\sigma^2} \left[-\frac{1}{2} \left\{ m \log(\sigma^2) + \frac{1}{\sigma^2} (y - X\hat{\beta})^\intercal (y - X\hat{\beta}) \right\} \right]\end{aligned}$$

- Differentiating and setting equal to zero

$$\hat{\sigma}^2 = \frac{1}{m} (y - X\hat{\beta})^\intercal (y - X\hat{\beta})$$

- Unbiased estimator

$$S^2 = \frac{1}{m - p} (y - X\hat{\beta})^\intercal (y - X\hat{\beta})$$

Outline

Least Squares Estimators

- Normal linear regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}_m(\mathbf{0}, \sigma^2\mathbf{I})$$

- Least squares estimators
  - $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{y}$
  - $\hat{\sigma}^2 = \frac{1}{m}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\intercal(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \frac{1}{m}\|\mathbf{y} - \hat{\mathbf{y}}\|^2$
  - $S^2 = \frac{1}{m-p}\|\mathbf{y} - \hat{\mathbf{y}}\|^2$

- Want to study the distributions of these estimators to
  - understand their precision
  - make confidence intervals/regions
  - test hypotheses
  - make predictions

Joint Distribution of Least Squares Estimators

Theorem Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ where

- $\boldsymbol{\varepsilon} \sim \mathcal{N}_m(\mathbf{0}, \sigma^2 \mathbf{I})$
- $\mathbf{X}$ is an $m \times p$ matrix ($m \geq p$) of rank p

Then

1. $\hat{\boldsymbol{\beta}} \sim \mathcal{N}_p\big(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^\intercal \mathbf{X})^{-1}\big)$
2. The random variables $\hat{\boldsymbol{\beta}}$ and S^2 are independent
3. $\dfrac{m-p}{\sigma^2} S^2 \sim \chi^2_{m-p}$

Joint Distribution of Least Squares Estimators

- Claim # 1

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}_p\big(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}\big)$$

- Property 2: if $\mathbf{V} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $\mathbf{B}\mathbf{V} + \mathbf{b} \sim \mathcal{N}_n(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\intercal)$
- Let $\mathbf{B} = (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal$, $\mathbf{b} = \mathbf{0}$, and $\mathbf{V} = \mathbf{Y} \sim \mathcal{N}_p(\mathbf{X}\beta, \sigma^2\mathbf{I})$, then

$$\begin{aligned}
(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{Y} &\sim \mathcal{N}_p((\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{X}\beta, (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\sigma^2\mathbf{I}\,[(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal]^\intercal) \\
\hat{\boldsymbol{\beta}} &\sim \mathcal{N}_p(\boldsymbol{\beta},\ (\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\sigma^2 I\,[(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal]^\intercal) \\
\hat{\boldsymbol{\beta}} &\sim \mathcal{N}_p(\boldsymbol{\beta},\ \sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{X}^\intercal\mathbf{X}(\mathbf{X}^\intercal\mathbf{X})^{-1}) \\
\hat{\boldsymbol{\beta}} &\sim \mathcal{N}_p(\boldsymbol{\beta},\ \sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1})
\end{aligned}$$

Joint Distribution of Least Squares Estimators

- Claim #2: the estimators $\hat{\boldsymbol{\beta}}$ and S^2 are independent
- If $\mathbf{e}$ is independent of $\hat{\boldsymbol{\beta}}$, then $S^2 = \mathbf{e}^\intercal \mathbf{e}/(m-p)$ is independent of $\hat{\boldsymbol{\beta}}$

$$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y} \qquad \hat{\boldsymbol{\beta}} = (\mathbf{X}^\intercal \mathbf{X})^{-1}\mathbf{X}^\intercal \mathbf{Y} \qquad \mathbf{Y} \sim \mathcal{N}_m(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$$

- Property 7: If $\mathbf{Y} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then

$$\mathbf{AY} \text{ independent of } \mathbf{BY} \iff \mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\intercal = \mathbf{0}$$

- Let $\mathbf{A} = (\mathbf{I} - \mathbf{H})$, $\mathbf{B} = (\mathbf{X}^\intercal \mathbf{X})^{-1}\mathbf{X}^\intercal$, and $\mathbf{Y} \sim \mathcal{N}_m(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$

$$\begin{aligned}
\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}^\intercal &= (\mathbf{I} - \mathbf{H})(\sigma^2 \mathbf{I})\left[(\mathbf{X}^\intercal \mathbf{X})^{-1}\mathbf{X}^\intercal\right]^\intercal \\
&= \sigma^2 \left[\mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1} - \mathbf{H}\mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1}\right] \\
&= \sigma^2 \left[\mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1} - \mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1}\mathbf{X}^\intercal \mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1}\right] \\
&= \sigma^2 \left[\mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1} - \mathbf{X}(\mathbf{X}^\intercal \mathbf{X})^{-1}\right] = \mathbf{0}
\end{aligned}$$

Joint Distribution of Least Squares Estimators

- Claim #3: $\dfrac{m-p}{\sigma^2}\, S^2 \sim \chi^2_{m-p}$
- Observe that $\mathbf{e} = (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon}$
- It follows that

$$
\begin{aligned}
(m-p)\, S^2 &= (m-p)\frac{\mathbf{e}^\intercal \mathbf{e}}{m-p} = \boldsymbol{\varepsilon}^\intercal (\mathbf{I} - \mathbf{H})^\intercal (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^\intercal (\mathbf{I} - \mathbf{H})\boldsymbol{\varepsilon} \\
\frac{(m-p)}{\sigma^2}\, S^2 &= \frac{\boldsymbol{\varepsilon}^\intercal}{\sigma}(\mathbf{I} - \mathbf{H})\frac{\boldsymbol{\varepsilon}}{\sigma} \\
&= \mathbf{Z}^\intercal (\mathbf{I} - \mathbf{H})\mathbf{Z} \qquad \text{where } \mathbf{Z} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I}) \\
&\sim \chi^2_{m-p}
\end{aligned}
$$

Distribution of a Single Parameter

- Recall claim #1

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}_p\big(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\intercal\mathbf{X})^{-1}\big)$$

- Let $\mathbf{c}_j$ be a vector where the j^{th} component is one and the rest are zero

- Property 2

$$\hat{\beta}_j = \mathbf{c}_j^\intercal\hat{\boldsymbol{\beta}} \sim \mathcal{N}_1(\mathbf{c}_j^\intercal\boldsymbol{\beta}, \sigma^2\mathbf{c}_j^\intercal(\mathbf{X}^\intercal\mathbf{X})^{-1}\mathbf{c}_j) = \mathcal{N}(\beta_j, \sigma^2\delta_{jj})$$

  where δ_{jj} is the j^{th} diagonal element of $(\mathbf{X}^\intercal\mathbf{X})^{-1}$

- The expected value and standard error (se) of $\hat{\beta}_j$

$$\mathbb{E}(\hat{\beta}_j) = \beta_j \qquad \text{and} \qquad \text{se}(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)} = \sqrt{\sigma^2\delta_{jj}}$$

- The estimated standard error is $\widehat{\text{se}}(\hat{\beta}_j) = \sqrt{S^2\delta_{jj}}$

Hunting for α

```
> summary(lm(Asset.returns ~ SP500.returns))

Call:
lm(formula = Asset.returns ~ SP500.returns)

Residuals:
     Min       1Q   Median       3Q      Max
-0.09001 -0.00891 -0.00030  0.00878  0.05954

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.002633   0.000984    2.67    0.008
SP500.returns 2.023243   0.123288   16.41   <2e-16

Residual standard error: 0.0155 on 248 degrees of freedom
Multiple R-squared:  0.521, Adjusted R-squared:  0.519
F-statistic:  269 on 1 and 248 DF,  p-value: <2e-16
```

Interpretation of `summary` Output

- The first column is the coefficient estimate $\hat{\beta}_j$
- The second column is the standard error estimate $\widehat{\text{se}}(\hat{\beta}_j)$
- The third column is the observed t value

$$t = \frac{\hat{\beta}_j}{\widehat{\text{se}}(\widehat{\beta}_j)}$$

- Under the *null hypothesis* $H_0 : \beta_j = 0$

$$T = \frac{\hat{\beta}_j}{\widehat{\text{se}}(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{S^2 \delta_{jj}}} = \frac{\frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 \delta_{jj}}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{U}{m-p}}} \sim t_{m-p}$$

where $Z \sim \mathcal{N}(0,1)$ and $U \sim \chi^2_{m-p}$ are independent

- The fourth column is the p-value $\text{P}(|T| > |t|)$

Hunting for α

```
> summary(lm(Asset.returns ~ SP500.returns))

Call:
lm(formula = Asset.returns ~ SP500.returns)

Residuals:
     Min        1Q    Median        3Q       Max
-0.09001 -0.00891 -0.00030  0.00878  0.05954

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.002633   0.000984    2.67    0.008
SP500.returns 2.023243   0.123288   16.41   <2e-16

Residual standard error: 0.0155 on 248 degrees of freedom
Multiple R-squared:  0.521, Adjusted R-squared:  0.519
F-statistic:  269 on 1 and 248 DF,  p-value: <2e-16
```

Outline

Confidence Intervals

- Want a $(1-\alpha)\times 100\%$ confidence interval (CI) for a linear combination of the parameters
  - $\mathbf{c}^\intercal \hat{\boldsymbol{\beta}} \sim \mathcal{N}_1(\mathbf{c}^\intercal \boldsymbol{\beta}, \sigma^2 \mathbf{c}^\intercal (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{c}) = \mathcal{N}_1(\mathbf{c}^\intercal \boldsymbol{\beta}, \sigma^2 \delta)$
  - $Z = (\mathbf{c}^\intercal \hat{\boldsymbol{\beta}} - \mathbf{c}^\intercal \beta)/(\sigma\sqrt{\delta}) \sim \mathcal{N}_1(0,1) \implies Z^2 \sim \chi^2_1$
  - Z^2 and S^2 are independent (since $\hat{\beta}$ and S^2 are independent)
  - $\frac{m-p}{\sigma^2} S^2 \sim \chi^2_{m-p}$
- It follows that

$$\frac{\frac{Z^2}{1}}{\frac{\frac{m-p}{\sigma^2} S^2}{m-p}} \sim F_{1,m-p} \Rightarrow \frac{\frac{(c^\intercal \hat{\beta} - c^\intercal \beta)^2}{\sigma^2 \delta}}{\frac{S^2}{\sigma^2}} = \left(\frac{c^\intercal \hat{\beta} - c^\intercal \beta}{\sqrt{S^2 c^\intercal (X^\intercal X)^{-1} c}} \right)^2 \sim F_{1,m-p}$$

Confidence Intervals

- But $W^2 \sim F_{1,m-p} \iff W \sim t_{m-p}$, so base the CI on

$$\frac{c^\intercal \hat{\beta} - c^\intercal \beta}{\sqrt{S^2 c^\intercal (X^\intercal X)^{-1} c}} \sim t_{m-p}$$

- Symmetric $(1-\alpha) \times 100\%$ confidence interval

$$\mathbf{c}^\intercal \hat{\boldsymbol{\beta}} \pm t_{m-p}(1-\alpha/2)\sqrt{S^2 \mathbf{c}^\intercal (\mathbf{X}^\intercal \mathbf{X})^{-1} \mathbf{c}}$$

- Use c_j to obtain a confidence interval for β_j

$$\hat{\beta}_j \pm t_{m-p}(1-\alpha/2)\sqrt{S^2 \delta_{jj}}$$

Prediction Intervals

- Want to predict y_{new} given $\mathbf{x}_{\text{new}} \in \mathbb{R}^n$
- The model predicts y_{new} by $\mathbf{x}_{\text{new}}^\top \hat{\boldsymbol{\beta}}$ but $y_{\text{new}} = \mathbf{x}_{\text{new}}^\top \boldsymbol{\beta} + \varepsilon_{\text{new}}$
- A prediction interval is different than the interval for a linear combination $\mathbf{c}^\top \boldsymbol{\beta}$ because there is additional uncertainty due to ε_{new}
- $\mathbb{E}(\mathbf{x}_{\text{new}}^\top \hat{\boldsymbol{\beta}} + \varepsilon_{\text{new}}) = \mathbf{x}_{\text{new}}^\top \boldsymbol{\beta}$
- $\text{Var}(\mathbf{x}_{\text{new}}^\top \hat{\boldsymbol{\beta}} + \varepsilon_{\text{new}}) = \text{Var}(\mathbf{x}_{\text{new}}^\top \hat{\boldsymbol{\beta}}) + \text{Var}(\varepsilon_{\text{new}}) = \sigma^2 \left[\mathbf{x}_{\text{new}}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{\text{new}} + 1\right]$
- Base the prediction interval on

$$\frac{\mathbf{x}_{\text{new}}^\top \hat{\beta} - y_+}{\sqrt{S^2 \left[1 + \mathbf{x}_{\text{new}}^\top (X^\top X)^{-1} \mathbf{x}_{\text{new}}\right]}} \sim t_{m-p}$$

- Obtain the symmetric, $(1-\alpha) \times 100\%$ prediction interval

$$\mathbf{x}_{\text{new}}^\top \hat{\boldsymbol{\beta}} \pm t_{m-p}(1-\alpha/2) \sqrt{S^2 \left[1 + \mathbf{x}_{\text{new}}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_{\text{new}}\right]}$$

COMPUTATIONAL FINANCE & RISK MANAGEMENT

UNIVERSITY *of* WASHINGTON

Department of Applied Mathematics

http://computational-finance.uw.edu