Random Explanatory variables[1]

STA 431 Spring 2023

[1]See last slide for copyright information.

Overview

1. Preparation
2. Random Explanatory Variables

Change of Variables

A big theorem

$$
\begin{aligned}
E(g(X)) &= \sum_x g(x)\, p_X(x) \\
E(g(\mathbf{x})) &= \sum_{x_1} \cdots \sum_{x_p} g(x_1, \ldots, x_p)\, p_{\mathbf{x}}(x_1, \ldots, x_p) \\
E(g(X)) &= \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \\
E(g(\mathbf{x})) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_p)\, f_{\mathbf{x}}(x_1, \ldots, x_p)\, dx_1 \ldots dx_p
\end{aligned}
$$

Indicator functions

Conditional expectation and the Law of Total Probability

$I_A(x)$ is the *indicator function* for the set A. It is defined by

$$I_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \notin A \end{cases}$$

Also sometimes written $I(x \in A)$

$$\begin{aligned} E(g(X)) &= E(I_A(X)) \\ &= \sum_x I_A(x)p(x), \text{ or} \\ & \quad \int_{-\infty}^{\infty} I_A(x)f(x)\,dx \\ &= P\{X \in A\} \end{aligned}$$

So the expected value of an indicator is a probability.

Applies to conditional probabilities too

Y given X, like regression

$$
\begin{aligned}
E(I_A(Y)|X=x) &= \sum_y I_A(y)p(y|x), \text{ or} \\
&\quad \int_{-\infty}^{\infty} I_A(y)f(y|x)\,dy \\
&= Pr\{Y \in A|X=x\}
\end{aligned}
$$

So the conditional expected value of an indicator is a *conditional* probability.

Double expectation

$$\begin{aligned} E(Y) &= E\{E(Y|X)\} \\ &= E_x\{E_y(Y|X)\} \\ &= E_x\{g(X)\} \end{aligned}$$

Showing $E(Y) = E\{E(Y|X)\}$

Again note $E\{E(Y|X)\}$ is an example of $E(g(X))$

$$
\begin{aligned}
E\{E(Y|X)\} &= \int E[Y|X=x] f_x(x)\,dx \\
&= \int \left(\int y f_{y|x}(y|x)\,dy \right) f_x(x)\,dx \\
&= \int \left(\int y \frac{f_{x,y}(x,y)}{f_x(x)}\,dy \right) f_x(x)\,dx \\
&= \int \int y\, f_{x,y}(x,y)\,dy\,dx \\
&= E(Y)
\end{aligned}
$$

Double expectation: $E\left(Y\right) = E\left\{E(Y|X)\right\}$

$$\begin{aligned} Pr\{Y \in A\} &= E\left(E[I_A(Y)|X]\right) \\ &= E\left(Pr\{Y \in A|X\}\right) \\ &= \int_{-\infty}^{\infty} Pr\{X \in A|X = x\} f_x(x)\, dx, \text{ or} \\ & \sum_x Pr\{Y \in A|X = x\} p_x(x) \end{aligned}$$

This is known as the *Law of Total Probability*

Random Explanatory Variables in Regression

Example: Multivariate Regression

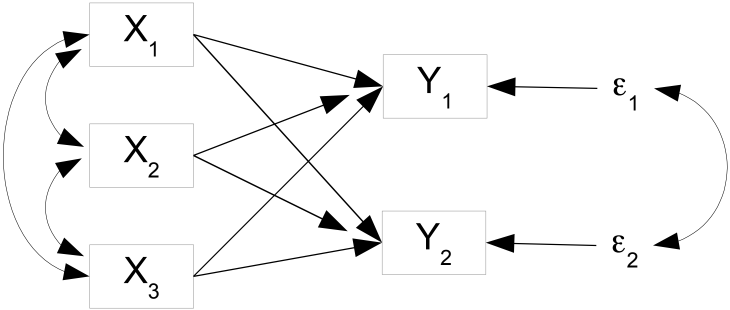
These are all vectors and matrices.

Independently for $i = 1, \ldots, n$,

$$\mathbf{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{x}_i + \boldsymbol{\epsilon}_i, \text{ where}$$

- $\mathbf{y}_i$ is an $q \times 1$ random vector of observable response variables, so the regression is multivariate; there are q response variables.
- $\mathbf{x}_i$ is a $p \times 1$ observable random vector; there are p explanatory variables. $E(\mathbf{x}_i) = \boldsymbol{\mu}_x$ and $cov(\mathbf{x}_i) = \boldsymbol{\Phi}_{p \times p}$. The vector $\boldsymbol{\mu}_x$ and the matrix $\boldsymbol{\Phi}$ are unknown parameters.
- $\boldsymbol{\beta}_0$ is a $q \times 1$ vector of unknown constants.
- $\boldsymbol{\beta}_1$ is a $q \times p$ matrix of unknown constants. These are the regression coefficients, with one row for each response variable and one column for each explanatory variable.
- $\boldsymbol{\epsilon}_i$ is a $q \times 1$ unobservable random vector with expected value zero and unknown variance-covariance matrix $cov(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi}_{q \times q}$.
- $\boldsymbol{\epsilon}_i$ is independent of $\mathbf{x}_i$.

Three explanatory variables and two response variables



Regression Equations

In scalar form,

$$\begin{aligned} y_{i,1} &= \beta_{1,0} + \beta_{1,1}x_{i,1} + \beta_{1,2}x_{i,2} + \beta_{1,3}x_{i,3} + \epsilon_{i,1} \\ y_{i,2} &= \beta_{2,0} + \beta_{2,1}x_{i,1} + \beta_{2,2}x_{i,2} + \beta_{2,3}x_{i,3} + \epsilon_{i,2} \end{aligned}$$

In matrix form,

$$\begin{array}{ccccccc} \mathbf{y}_i & = & \boldsymbol{\beta}_0 & + & \boldsymbol{\beta}_1 & \mathbf{x}_i & + & \boldsymbol{\epsilon}_i \\ \begin{pmatrix} y_{i,1} \\ y_{i,2} \end{pmatrix} & = & \begin{pmatrix} \beta_{1,0} \\ \beta_{2,0} \end{pmatrix} & + & \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \end{pmatrix} & \begin{pmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{pmatrix} & + & \begin{pmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \end{pmatrix} \end{array}$$

Simulate Multivariate Regression

```
> # Set parameter values
>     # Regression coefficients
>         beta10 = 1; beta11 = 1; beta12 = 0; beta13 = 3
>         beta20 = 2; beta21 = 0; beta22 = 4; beta23 = 0
>     # Expected values of x variables
>         mux = c(10,20,10)
>     # Variance-covariance matrix of x variables
>         Phi = rbind(c(25,  25, 15),
+                     c(25, 100, 35),
+                     c(15,  35, 25))
>     # Variance-covariance matrix of error terms
>         Psi = rbind(c(500,  750),
+                     c(750, 2000))

> source("https://www.utstat.toronto.edu/~brunner/openSEM/fun/rmvn.txt")
```

First an experiment

```
> x = rmvn(nn=100000, mu=mux, sigma=Phi)
> dim(x)
[1] 100000      3

> head(x)
          [,1]      [,2]      [,3]
[1,]  8.956959  7.537267  8.256174
[2,] 21.678814 21.764190 20.103837
[3,] 10.340543 28.986937 17.104511
[4,]  3.760735 10.528940  7.981938
[5,]  9.916082 24.939210 10.681681
[6,]  7.001012 21.927595 16.729394
```

Estimation should be very good with $n = 100,000$

```
> apply(x,MARGIN=2,FUN=mean) # Column sample means
[1] 10.00862 19.96847 10.00400

> mux # Population means, for comparison
[1] 10 20 10

> var(x) # Sample variance-covariance matrix with n-1
         [,1]      [,2]     [,3]
[1,] 24.95666  25.07949 14.95961
[2,] 25.07949 100.16989 35.05957
[3,] 14.95961  35.05957 24.95647

> Phi # Population variance-covariance matrix, for comparison
     [,1] [,2] [,3]
[1,]   25   25   15
[2,]   25  100   35
[3,]   15   35   25
```

Simulate data from the model

```
> n = 500
> x = rmvn(nn=n, mu=mux, sigma=Phi)
> epsilon = rmvn(nn=n, mu=c(0,0), sigma=Psi)
> # Extract variables (for clarity)
> x1 = x[,1]; x2 = x[,2]; x3 = x[,3]
> epsilon1 = epsilon[,1]; epsilon2 = epsilon[,2]
> # Generate y
> y1 = beta10 + beta11*x1 + beta12*x2 + beta13*x3 + epsilon1
> y2 = beta20 + beta21*x1 + beta22*x2 + beta23*x3 + epsilon2
> length(y1)
[1] 500
```

Calculate MOM estimate of $\boldsymbol{\beta}_1$ (the slopes)

$$\widehat{\boldsymbol{\beta}}_1 = \widehat{\boldsymbol{\Sigma}}_{yx}\widehat{\boldsymbol{\Sigma}}_x^{-1}$$

```
> # Calculate MOM estimate of beta1 (the slopes)
> y = cbind(y1,y2)
> Sigmahat_x  = var(x) * (n-1)/n
> Sigmahat_xy = var(x,y) * (n-1)/n
> beta1hat = t(Sigmahat_xy) %*% solve(Sigmahat_x)
> round(beta1hat,3)
     [,1]  [,2]  [,3]
y1  0.707 0.130 2.967
y2 -0.300 4.194 0.228
```

True $\boldsymbol{\beta}_1$ for Comparison

```
> # True beta1
> beta1 = rbind(c(beta11, beta12, beta13),
+               c(beta21, beta22, beta23))
> beta1

     [,1] [,2] [,3]
[1,]    1    0    3
[2,]    0    4    0

> # Estimated beta1
> round(beta1hat,3)
     [,1]  [,2]  [,3]
y1  0.707 0.130 2.967
y2 -0.300 4.194 0.228
```

MOM = Least Squares

```
> # MOM estimate of slopes
> round(beta1hat,3)
     [,1]  [,2]  [,3]
y1  0.707 0.130 2.967
y2 -0.300 4.194 0.228

> # Least squares estimate
> LSbetahat = lsfit(x,y)$coefficients                                    #
> t(round( LSbetahat ,3))
   Intercept     X1    X2    X3
Y1     0.658  0.707 0.130 2.967
Y2    -2.440 -0.300 4.194 0.228
```

But this is not how the standard theory goes

Don't you think its strange?

- In the general linear regression model, the $\mathbf{X}$ matrix is supposed to be full of fixed constants.
- This is convenient mathematically. Think of $E(\widehat{\boldsymbol{\beta}})$.
- But in any non-experimental study, ...
- View the usual model as *conditional* on $\mathcal{X} = \mathbf{X}$.
- All the probabilities and expected values in the typical regression course are *conditional* probabilities and *conditional* expected values.
- Does this make sense?

$\widehat{\boldsymbol{\beta}}$ is (conditionally) unbiased

$$E(\widehat{\boldsymbol{\beta}}|\mathcal{X} = \mathbf{X}) = \boldsymbol{\beta} \text{ for } \textit{any} \text{ fixed } \mathbf{X}$$

It's *unconditionally* unbiased too.

$$E\{\widehat{\boldsymbol{\beta}}\} = E\{E\{\widehat{\boldsymbol{\beta}}|\mathcal{X}\}\} = E\{\boldsymbol{\beta}\} = \boldsymbol{\beta}$$

Conditional size α test, Critical value f_α

$$Pr\{F > f_\alpha | \mathcal{X} = \mathbf{X}\} = \alpha$$

$$\begin{aligned} Pr\{F > f_\alpha\} &= \int \cdots \int Pr\{F > f_\alpha | \mathcal{X} = \mathbf{X}\} f(\mathbf{X})\, d\mathbf{X} \\ &= \int \cdots \int \alpha f(\mathbf{X})\, d\mathbf{X} \\ &= \alpha \int \cdots \int f(\mathbf{X})\, d\mathbf{X} \\ &= \alpha \end{aligned}$$

The moral of the story

- Don't worry.
- Even though the explanatory variables are often random, we can apply the usual fixed $\mathbf{X}$ model without fear.
- Estimators are still unbiased.
- Tests have the right Type I error probability.
- Similar arguments apply to confidence intervals and prediction intervals.
- And it's all distribution-free with respect to $\mathcal{X}$.

Copyright Information