Descriptive Statistics with Scilab

ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่

Outline

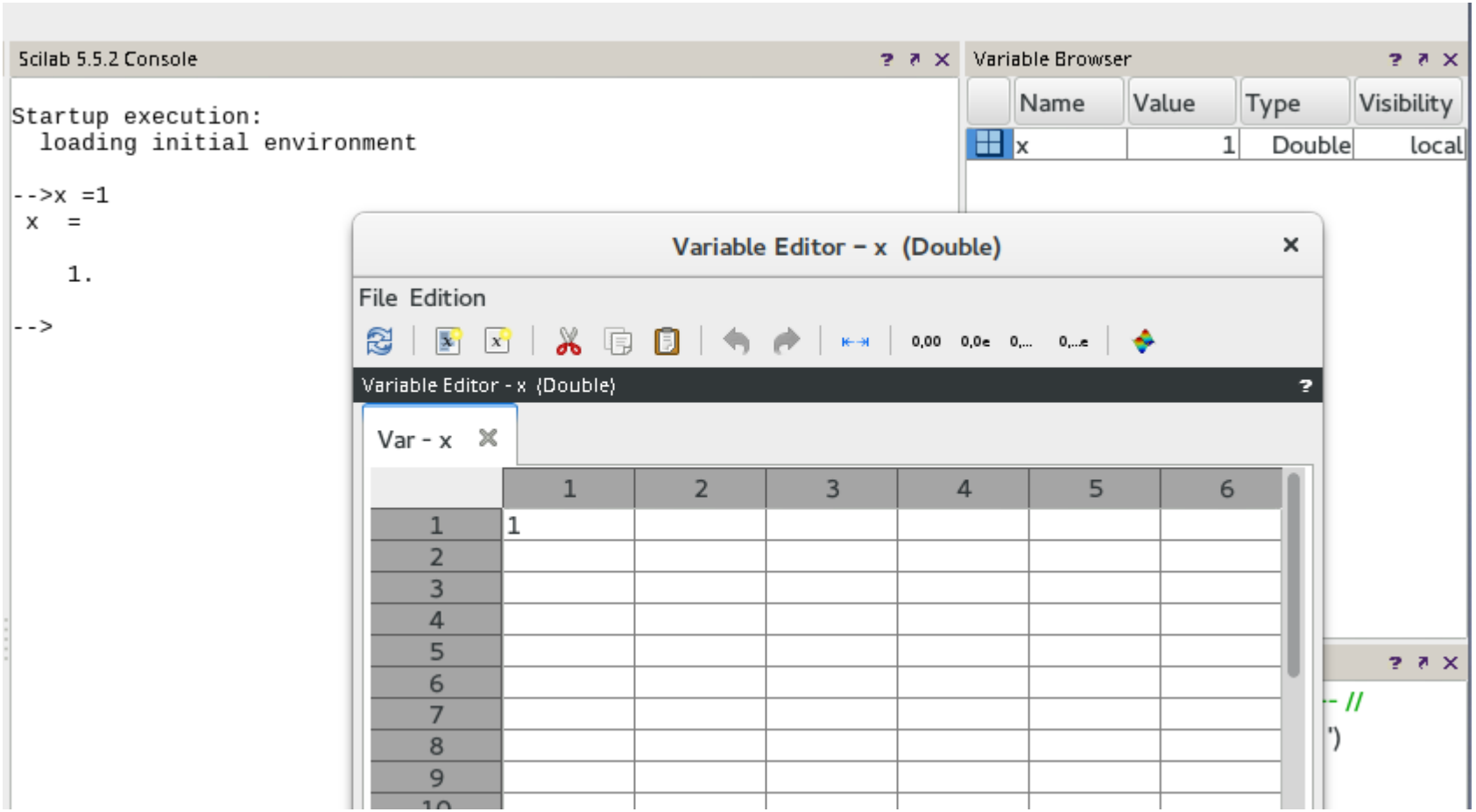
- การคำนวณเชิงสถิติโดยใช้ Scilab
  - การนำเสนอข้อมูลในรูป histogram
  - ค่าเฉลี่ย mean, มัธยฐาน median
  - ส่วนเบี่ยงเบนมาตรฐาน standard deviation
  - variance
  - Correlation of variation

การคำนวณเชิงสถิติโดยใช้ Scilab

- ตัวแปรใน scilab จะอยู่ในรูปของ matrix
  - ตัวเลขเดี่ยวๆ ก็คือ 1x1 matrix
- การสร้างตัวแปรสามารถทำได้โดย กำหนดชื่อตัวแปรและ assign ค่าให้ตัวแปรนั้น

  เช่น X = 1;
- หลังจากมีการกำหนดค่าตัวแปรแล้ว เราสามารถเข้าไปแก้ไขค่าของตัวแปรนั้นได้ โดยการ double click ชื่อตัวแปรในช่อง variable browser
- จากนั้นสามารถแก้ไขค่าตัวแปรได้ (เพิ่มมิติของตัวแปรได้ด้วย)

ตัวอย่างการแก้ไขค่าตัวแปร



การหาค่าเฉลี่ย

- สามารถทำได้โดยใช้ฟังก์ชั่น mean(X, dim)
- X คือตัวแปรที่เก็บข้อมูลที่เราต้องการจะหาค่าเฉลี่ย
- dim คือการระบุว่าจะให้หาค่าเฉลี่ยในมิติไหน
  - หากไม่ใส่คือ หาค่าเฉลี่ยของตัวเลขทุกตัวใน เมทริกซ์ X
  - หาก dim='c' คือหาค่าเฉลี่ยของตัวเลขทุกหลัก ที่อยู่ในแถวเดียวกัน
  - หาก dim='r' คือหาค่าเฉลี่ยของตัวเลขทุกแถว ที่อยู่ในหลักเดียวกัน

ตัวอย่าง

Scilab 5.5.2 Console

```
-->x = [1,1,1;2,2,2;3,3,3]
 x  =

    1.    1.    1.
    2.    2.    2.
    3.    3.    3.

-->mean(x)
 ans  =

    2.

-->mean(x,'r')
 ans  =

    2.    2.    2.

-->mean(x,'c')
 ans  =

    1.
    2.
    3.

-->
```

การหาค่ามัธยฐาน

- สามารถทำได้โดยใช้ฟังก์ชั่น median(X, dim)
- X คือตัวแปรที่เก็บข้อมูลที่เราต้องการจะหาค่าเฉลี่ย
- dim คือการระบุว่าจะให้หาค่าเฉลี่ยในมิติไหน
  - หากไม่ใส่คือ หาค่าเฉลี่ยของตัวเลขทุกตัวใน เมทริกซ์ X
  - หาก dim='c' คือหาค่าเฉลี่ยของตัวเลขทุกหลัก ที่อยู่ในแถวเดียวกัน
  - หาก dim='r' คือหาค่าเฉลี่ยของตัวเลขทุกแถว ที่อยู่ในหลักเดียวกัน

ตัวอย่าง

Scilab 5.5.2 Console

```
-->x = [1,1,1;2,2,2;3,3,3]
 x  =

    1.    1.    1.
    2.    2.    2.
    3.    3.    3.

-->median(x)
 ans  =

    2.

-->median(x,'r')
 ans  =

    2.    2.    2.

-->median(x,'c')
 ans  =

    1.
    2.
    3.

-->
```

การสร้าง Histogram

- Histogram มีประโยชน์ในการดูภาพรวมของชุดข้อมูล
- histplot(n,x,[,normalization])
  - n คือจำนวนกล่องที่ต้องการ เช่น 10
  - x คือข้อมูลที่ต้องการนำไปพลอท
  - Normalization คือ option เสริมที่บอกว่าจะ normalize ความถี่ของข้อมูลให้อยู่ระหว่าง 0 กับ 1 หรือไม่ เช่น
    - normalization=%t ให้ทำการ normalize
    - normalization=%f ไม่ต้องทำการ normalize

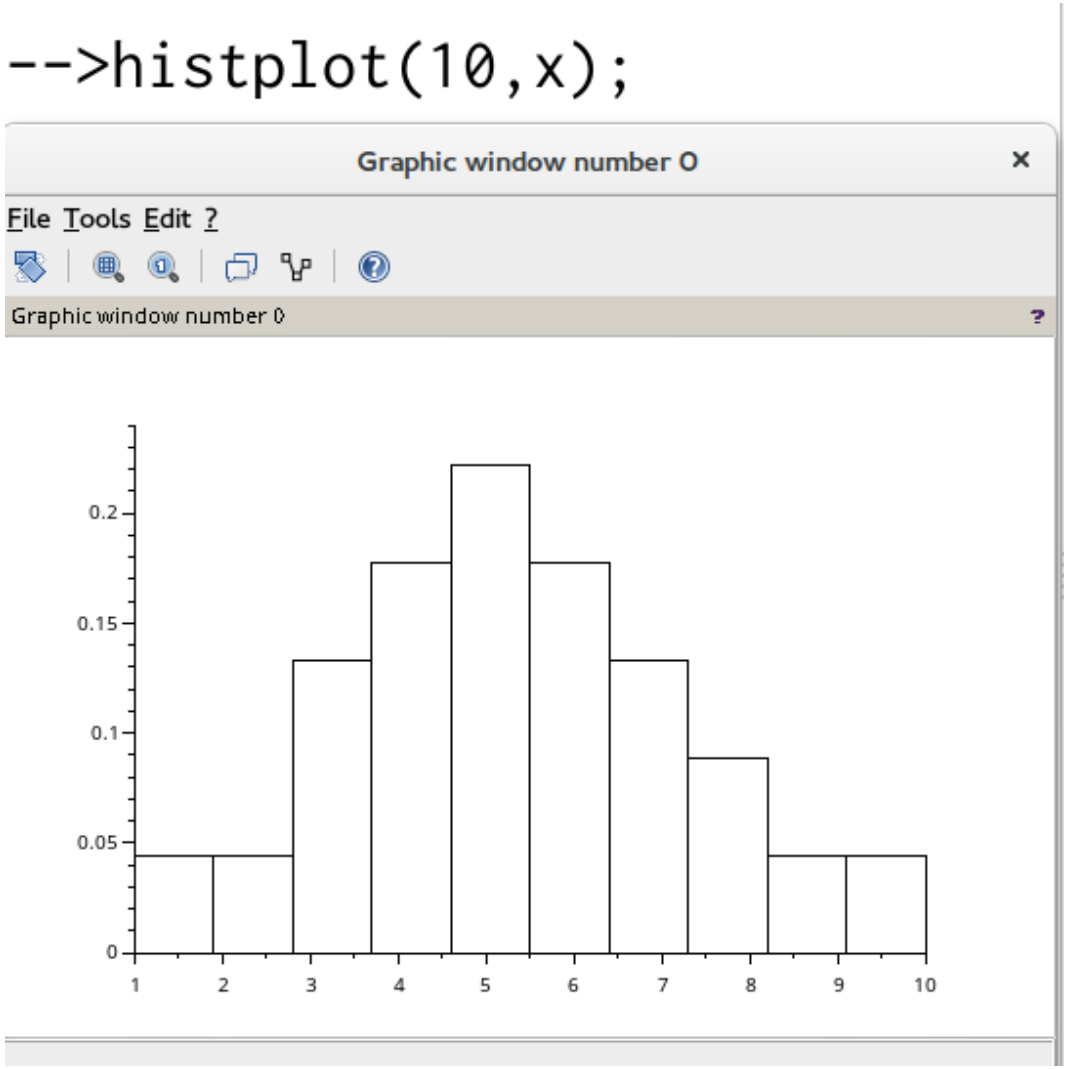
ตัวอย่าง

ข้อมูล

1
2,2
3,3,3,
4,4,4,4,
5,5,5,5,5
6,6,6,6
7,7,7
8,8
9,
10

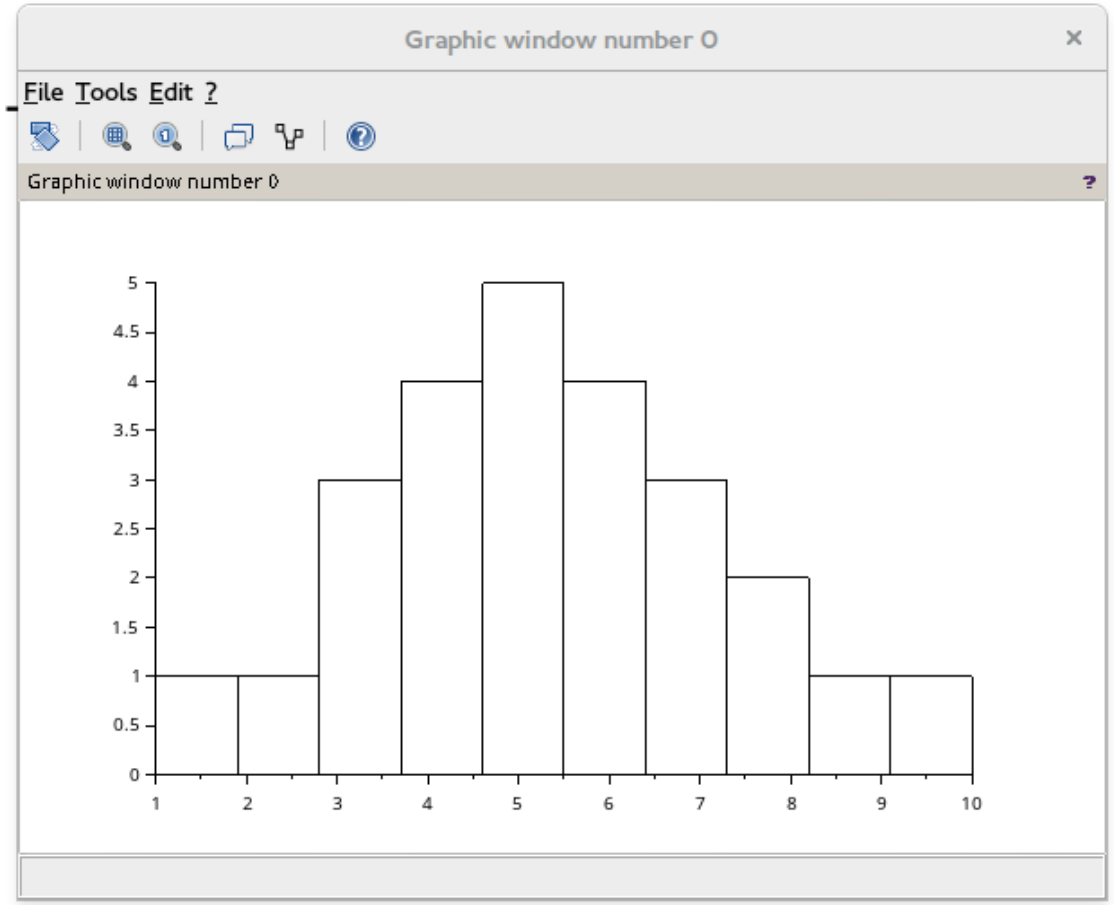
ตัวอย่าง

```
-->histplot(10,x);
```



ตัวอย่างแบบไม่ normalize

```
-->histplot(10,x,normalization=%f);
```



ส่วนเบี่ยงเบนมาตรฐาน

- ส่วนเบี่ยงเบนมาตรฐาน เป็นการวัดการกระจายของข้อมูลโดยพิจารณาสามารถคำนวณได้จาก

$$SD = \frac{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}{n-1}$$

- ใน Scilab สามารถคำนวณได้โดยใช้ฟังก์ชัน

  STDEV(X,dim)

ตัวอย่าง

- X = [1,2,3,4,5,6,7,8,9,10]

```
-->stdev(x)
 ans  =

    3.0276504
```

Variance

- คือค่าของ SD ยกกำลังสอง
- สามารถคำนวณโดยใช้ฟังก์ชัน stdev แล้วนำผลที่ได้มายกกำลังสอง หรือ
- ใช้ฟังก์ชัน variance()

ตัวอย่าง

```
-->stdev(x)^2
 ans  =

    9.1666667

-->variance(x)
 ans  =

    9.1666667

-->
```

Coefficient of variation (CV)

- เป็นการเปรียบเทียบการกระจาย ใช้เมื่อต้องการเปรียบเทียบข้อมูลสองชุดใด ๆว่ามีการกระจายมากกว่า ข้อมูลอีกชุดหนึ่ง
- สัมประสิทธิ์การแปรผัน สามารถหาได้โดย

$$CV = \frac{SD}{\bar{x}}$$