



《统计学导论》

李勇

北京师范大学统计学院

个人主页: <http://stat.bnu.edu.cn/zwj1/103803.html>

精品课: <http://course.bnu.edu.cn/course/statistics/>

精品共享课: http://www.icourses.cn/coursetatic/course_3827.html

公开课: <http://www.icourses.cn/viewVCourse.action?courseCode=10027V015>

使用说明: 鼠标单击红色字体后会跳转到链接目标位置; 同时按Alt和左箭头键, 返回上一演示位置.

课程概论

统计学是通过收集数据和分析数据来认识未知现象的一门科学，在各行各业中有广泛的应用。



课程概论



统计学是通过收集数据和分析数据来认识未知现象的一门科学，在各行各业中有广泛的应用。在统计学思想指导下处理生活、学习和工作中与（大）数据相关问题的能力，成为数字化社会公民必备的基本素质。



课程概论



统计学是通过收集数据和分析数据来认识未知现象的一门科学，在各行各业中有广泛的应用。在统计学思想指导下处理生活、学习和工作中与（大）数据相关问题的能力，成为数字化社会公民必备的基本素质。

本课程介绍统计学的思维方式和常用统计方法的原理，使学习者了解常用初等统计方法的来龙去脉，对于统计学有一个整体了解，为进一步学习《抽样调查》、《数理统计》和《线性回归模型》等后继统计学课程奠定基础。

课程概论

统计学是通过收集数据和分析数据来认识未知现象的一门科学，在各行各业中有广泛的应用。在统计学思想指导下处理生活、学习和工作中与（大）数据相关问题的能力，成为数字化社会公民必备的基本素质。

本课程介绍统计学的思维方式和常用统计方法的原理，使学习者了解常用初等统计方法的来龙去脉，对于统计学有一个整体了解，为进一步学习《抽样调查》、《数理统计》和《线性回归模型》等后继统计学课程奠定基础。

本课程将R语言融入教学过程，帮助学习者初步掌握R软件的使用方法，提高学习和研究效率：利用R语言的统计计算功能减轻学习者的数学推导和计算负担，以把主要精力集中于基本概念和方法原理的理解；利用R语言

的随机模拟功能帮助学习者体会随机现象和统计分析结果的内涵，理解统计方法的原理。

本课程第一部分介绍统计学的内涵、研究流程和研究追求目标，使学习者正确理解统计学模型的本质，以及统计软件的利与弊。



第二部分介绍概率论的基本概念，再现概率空间的形成思路，熏陶学习者的创新思维能力.



第二部分介绍概率论的基本概念，再现概率空间的形成思路，熏陶学习者的创新思维能力。

第三部分介绍随机变量的产生动因，随机变量的分布函数和基本数字特征，以及大数定律和中心极限定理，为解释统计方法原理奠定理论基础。

第二部分介绍概率论的基本概念，再现概率空间的形成思路，熏陶学习者的创新思维能力。

第三部分介绍随机变量的产生动因，随机变量的分布函数和基本数字特征，以及大数定律和中心极限定理，为解释统计方法原理奠定理论基础。

第四部分介绍数据收集的原理，包括有限总体的抽样原理和实验数据的抽取原理，使学习者理解数据收集方法的重要性。

第二部分介绍概率论的基本概念，再现概率空间的形成思路，熏陶学习者的创新思维能力。

第三部分介绍随机变量的产生动因，随机变量的分布函数和基本数字特征，以及大数定律和中心极限定理，为解释统计方法原理奠定理论基础。

第四部分介绍数据收集的原理，包括有限总体的抽样原理和实验数据的抽取原理，使学习者理解数据收集方法的重要性。

第五部分介绍用图形展示总体密度信息的基本原理，使学习者初步具备利用数据图展示和判断总体特征的能力。

第二部分介绍概率论的基本概念，再现概率空间的形成思路，熏陶学习者的创新思维能力。

第三部分介绍随机变量的产生动因，随机变量的分布函数和基本数字特征，以及大数定律和中心极限定理，为解释统计方法原理奠定理论基础。

第四部分介绍数据收集的原理，包括有限总体的抽样原理和实验数据的抽取原理，使学习者理解数据收集方法的重要性。

第五部分介绍用图形展示总体密度信息的基本原理，使学习者初步具备利用数据图展示和判断总体特征的能力。

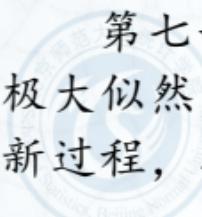
第六部分介绍总体数字特征的估计原理，以及应用这些样本特征探测离群数据的基本原理，展示离群数据研究的重要性。

第七部分参数估计评价方法的原理、矩估计的原理和
极大似然估计的原理，使学习者从这些原理中体会统计创
新过程，培养他们的创新能力。



第七部分参数估计评价方法的原理、矩估计的原理和极大似然估计的原理，使学习者从这些原理中体会统计创新过程，培养他们的创新能力。

第八部分区间估计和假设检验的原理，使学习者正确理解假设检验方法中的各个概念，以及假设检验的结论。



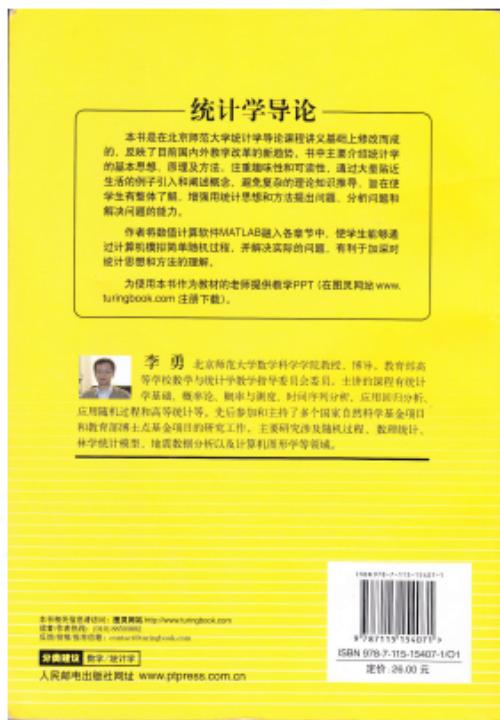
第七部分参数估计评价方法的原理、矩估计的原理和极大似然估计的原理，使学习者从这些原理中体会统计创新过程，培养他们的创新能力.

第八部分区间估计和假设检验的原理，使学习者正确理解假设检验方法中的各个概念，以及假设检验的结论.

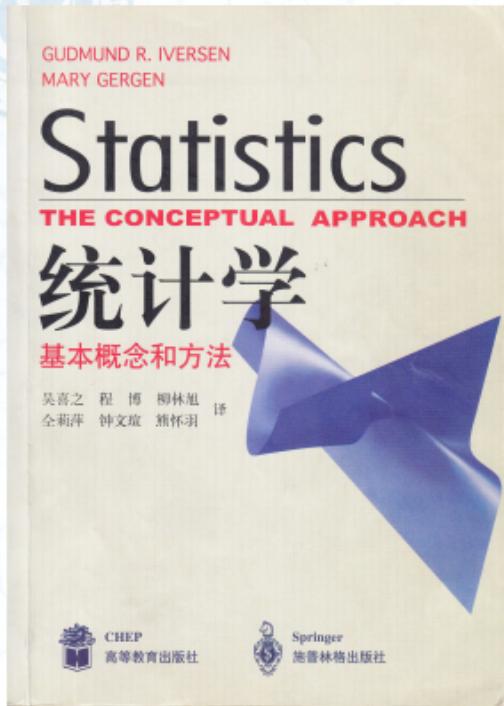
第九部分回归模型的基本原理，使学习者对于此模型有一个客观的初步理解.



2008年北京高等教育精品教材：统计学导论, 2007年版.



参考书目1：统计学—基本概念和方法，2000年版.



从数据入手：

- 如何评估统计的可靠性？
- 变量是如何定义的？
- 用了什么样的统计方法？
- 什么是“统计显著”的结果？

研究设计问题通常是我们进入统计学的中心。本书将引导学生思考并回答这些问题，教会学生如何看统计图表、调查、选举投票、广告宣传，以及其它的隐藏在数据中的统计结果。

邀请您的想象力：

本书配备了纯数学式的推导，从基础的逻辑、清晰直白的解答到复杂的案例都述了统计学的基本概念和技巧。在健康促进、婚姻、教育、休闲、科学、生活质量等方面的数据将起到特色，相信学生会发现学习统计的乐趣！不强调公式，技术或较难理解的内容（注释末）。使用计算机和统计软件帮助学生处理数据的统计中解题时，让他们把精力集中在对统计过程和概念的理解上，每一道习题提供了能让你着手可能的主题。书中指出了许多练习题答案和各章的统计表。

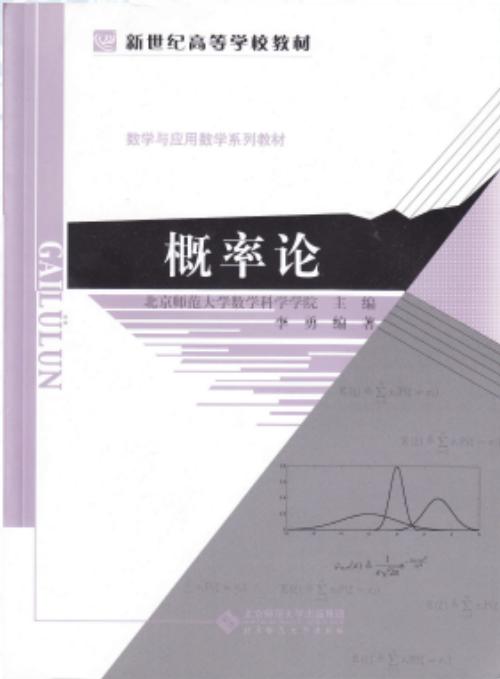
本精炼而又丰富的教材经过了千锤百炼，深受欢迎。它不仅可作为理工科、文科、金融管理类专业学生的统计教材，也可作为各行各业实际工作者和各指导员的统计“启蒙”或“扫盲”之用。

ISBN 7-04-007891-0

9 787040 078916
定价：52.00 元

ISBN 7-04-007891-0

参考书目2：概率论，2013年版.



参考书目3：概率论基础及其应用, 2007年版.



第1章 绪论

1.1 未知现象的认识过程与统计学

案例 1.1 未来的天气.

1. 神化：古时人们不能解释天气的变化，把它归结为神的支配.



第1章 绪论

1.1 未知现象的认识过程与统计学

案例 1.1 未来的天气.

1. 神化：古时人们不能解释天气的变化，把它归结为神的支配.
2. 谚语：关于气压、湿度、云、雨、冰雹等气象要素定性观测的经验总结.

第1章 绪论

1.1 未知现象的认识过程与统计学

案例 1.1 未来的天气.

1. 神化：古时人们不能解释天气的变化，把它归结为神的支配.
2. 谚语：关于气压、湿度、云、雨、冰雹等气象要素定性观测的经验总结.
3. 当地定量观测数据预报：用当地天气测量数据预报未来天气.

第1章 绪论

1.1 未知现象的认识过程与统计学

案例 1.1 未来的天气.

1. 神化：古时人们不能解释天气的变化，把它归结为神的支配.
2. 谚语：关于气压、湿度、云、雨、冰雹等气象要素定性观测的经验总结.
3. 当地定量观测数据预报：用当地天气测量数据预报未来天气.
4. 台网定量观测预报：用全球气象观测网数据预报未来天气。

1.1.1 与案例相关的概念

不确定现象（未知现象）：在特定条件下，不能预知结果的现象.



1.1.1 与案例相关的概念

不确定现象（未知现象）：在特定条件下，不能预知结果的现象.

必然现象：在特定条件下，能预知结果的现象.

1.1.1 与案例相关的概念

不确定现象（未知现象）：在特定条件下，不能预知结果的现象.

必然现象：在特定条件下，能预知结果的现象.

统计学：通过收集数据和分析数据来认识未知现象的一门科学.



1.1.1 与案例相关的概念

不确定现象（未知现象）：在特定条件下，不能预知结果的现象.

必然现象：在特定条件下，能预知结果的现象.

统计学：通过收集数据和分析数据来认识未知现象的一门科学.

大英百科全书：“统计学是一种收集数据、分析数据，并根据数据进行推断的艺术和科学”.

1.1.2 统计学研究流程

未来天气

图 1.1 未来天气变化的研究流程

1.1.2 统计学研究流程

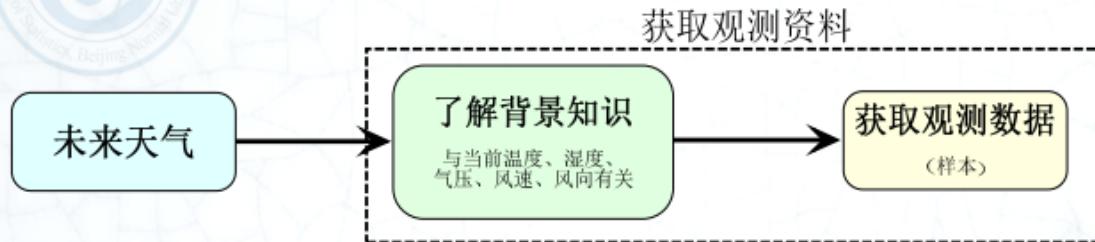


图 1.2 未来天气变化的研究流程

1.1.2 统计学研究流程

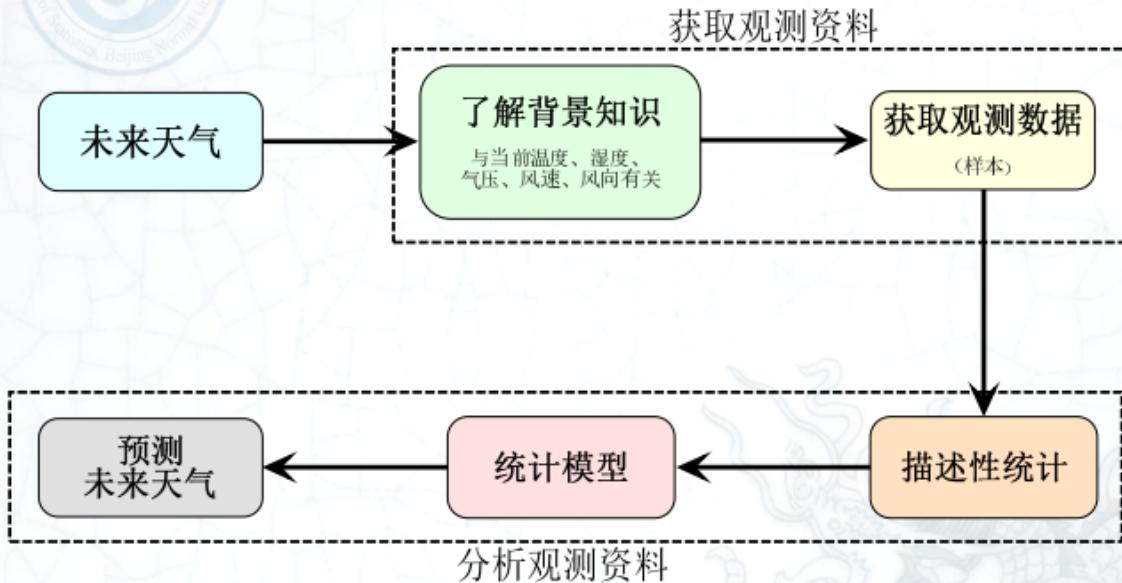


图 1.3 未来天气变化的研究流程

1.1.2 统计学研究流程

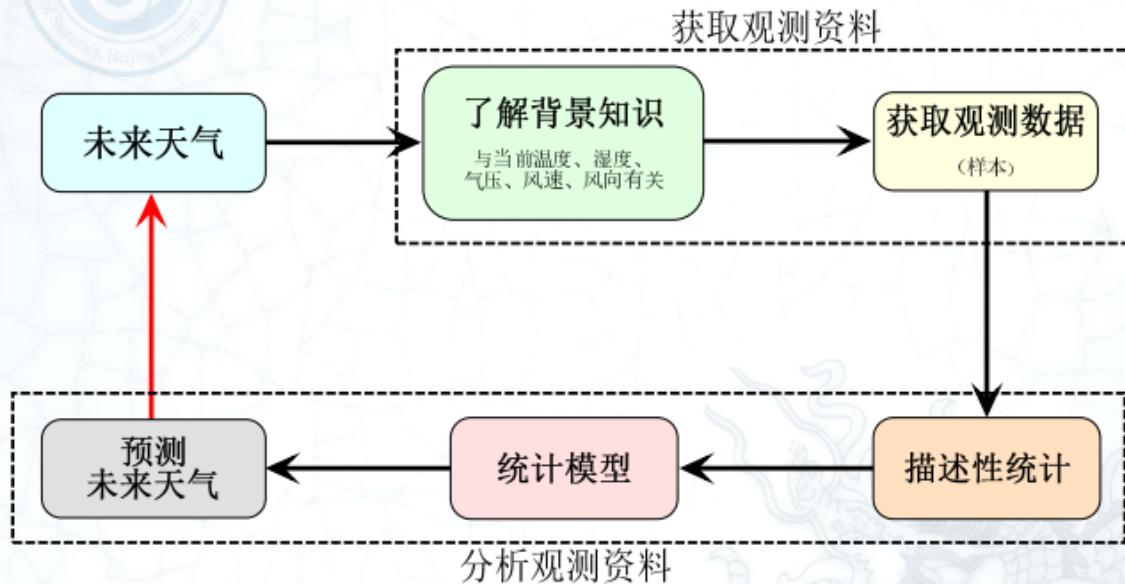


图 1.4 未来天气变化的研究流程



一般地，统计学思维模式如下：

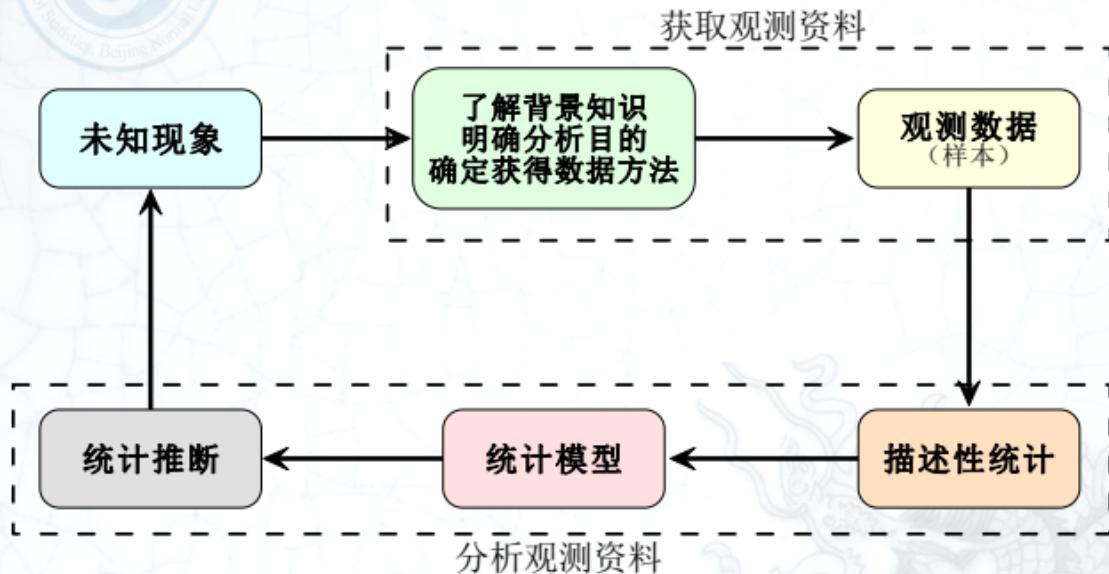


图 1.5 统计学研究流程图

例 1.1.1 了解地震的孕育过程，是人类梦寐以求的愿望。
为达此愿，应该收集哪些数据，如何收集这些数据？

解：



例 1.1.1 了解地震的孕育过程，是人类梦寐以求的愿望。
为达此愿，应该收集哪些数据，如何收集这些数据？

解：要了解未来地震发生的时间、地点和震级变化规律，当然应该收集有与这 3 个量有关的数据：地震目录、形变、和地下水位等。

例 1.1.1 了解地震的孕育过程，是人类梦寐以求的愿望。
为达此愿，应该收集哪些数据，如何收集这些数据？

解：要了解未来地震发生的时间、地点和震级变化规律，当然应该收集有与这 3 个量有关的数据：地震目录、形变、和地下水位等。由于地震是一种自然现象，只能通过观测的方式来收集有关的数据。 □

例 1.1.2 一个人在市场上买了一篮表面上看起来十分新鲜的水果. 回到家详细检查后才发现篮中有许多烂水果, 这里的分析结果为什么和实际情况不相符合?

解:



例 1.1.2 一个人在市场上买了一篮表面上看起来十分新鲜的水果. 回到家详细检查后才发现篮中有许多烂水果, 这里的分析结果为什么和实际情况不相符合?

解: 因为获取观测资料的方法有问题, 这些观测资料不能很好地代表篮中的所有水果;

例 1.1.2 一个人在市场上买了一篮表面上看起来十分新鲜的水果. 回到家详细检查后才发现篮中有许多烂水果, 这里的分析结果为什么和实际情况不相符合?

解: 因为获取观测资料的方法有问题, 这些观测资料不能很好地代表篮中的所有水果; 而在分析观测数据的过程中没有注意到这个问题, 使用的分析方法不得当, 使得分析结果与实际不符. □

统计学常识

例 1.1.2 一个人在市场上买了一篮表面上看起来十分新鲜的水果. 回到家详细检查后才发现篮中有许多烂水果, 这里的分析结果为什么和实际情况不相符合?

解: 因为获取观测资料的方法有问题, 这些观测资料不能很好地代表篮中的所有水果; 而在分析观测数据的过程中没有注意到这个问题, 使用的分析方法不得当, 使得分析结果与实际不符. □

统计学常识

1. 要用科学的方法收集数据.
2. 需要根据数据的背景特点(获取过程、质量、分布特点)调整分析方法, 以取得更好的分析结果.

例 1.1.3 摆奖机是否能公平是亿万彩民关心的问题. 请设计一个检验摇奖机是否公平摇奖的方案（流程）.

解：

例 1.1.3 摆奖机是否能公平是亿万彩民关心的问题. 请设计一个检验摇奖机是否公平摇奖的方案 (流程) .

解: 这里“公平”是指摇出每个号码的概率相等. 根据频率稳定于概率的思想, 可以按如下的方案检验摇奖机的公平性:

例 1.1.3 摆奖机是否能公平是亿万彩民关心的问题. 请设计一个检验搖奖机是否公平搖奖的方案(流程).

解: 这里“公平”是指搖出每个号码的概率相等. 根据频率稳定于概率的思想, 可以按如下的方案检验搖奖机的公平性:

1. 将带号码的球装入搖奖机后, 开动摇奖机搖出一个球, 记录下该球的号码.

例 1.1.3 摆奖机是否能公平是亿万彩民关心的问题. 请设计一个检验搖奖机是否公平搖奖的方案(流程).

解: 这里“公平”是指搖出每个号码的概率相等. 根据频率稳定于概率的思想, 可以按如下的方案检验搖奖机的公平性:

1. 将带号码的球装入搖奖机后, 开动摇奖机搖出一个球, 记录下该球的号码.
2. 重复第1步, 考察搖出的各个号码的频率的稳定性.
当所有号码的频率都比较稳定之后, 再进行下一步.

例 1.1.3 摆奖机是否能公平是亿万彩民关心的问题. 请设计一个检验搖奖机是否公平搖奖的方案(流程).

解: 这里“公平”是指搖出每个号码的概率相等. 根据频率稳定于概率的思想, 可以按如下的方案检验搖奖机的公平性:

1. 将带号码的球装入搖奖机后, 开动摇奖机搖出一个球, 记录下该球的号码.
2. 重复第1步, 考察搖出的各个号码的频率的稳定性.
当所有号码的频率都比较稳定之后, 再进行下一步.
3. 考察各个号码的频率的接近程度, 以此来判断这台搖奖机的公平性.

1.1.3 不确定现象与随机现象

案例 1.2 投掷一枚质地均匀的硬币，结果不是正面向上，就是反面向上.

1.1.3 不确定现象与随机现象

案例 1.2 投掷一枚质地均匀的硬币，结果不是正面向上，就是反面向上。随着投掷次数的增加，却呈现出如下规律性（见表1.1）：正面向上的比例接近于**0.5**。

表 1.1 投掷硬币的实验结果

实验者	抛硬币次数	出现正面次数	出现正面频率
蒲丰	4040	2048	0.5069
德莫根	4092	2048	0.5005
费勒	10000	4979	0.4979
皮尔逊	12000	6019	0.5016
皮尔逊	24000	12012	0.5005
罗曼诺夫斯基	80640	39699	0.4932

在特定条件下不能事先预知结果，且各个结果都具有频率稳定性的现象称为**随机现象**.

例 1.1.4 假定我们关心北京市一月份的平均气温，试讨论它与随机现象间的关系.

解：

在特定条件下不能事先预知结果，且各个结果都具有频率稳定性的现象称为**随机现象**.

例 1.1.4 假定我们关心北京市一月份的平均气温，试讨论它与随机现象间的关系.

解：“北京市”和“一月份”的特定条件下，不能预知这个平均气温，它是一个不确定现象.

在特定条件下不能事先预知结果，且各个结果都具有频率稳定性的现象称为**随机现象**.

例 1.1.4 假定我们关心北京市一月份的平均气温，试讨论它与随机现象间的关系.

解：“北京市”和“一月份”的特定条件下，不能预知这个平均气温，它是一个不确定现象.

考虑到温室效应气体的作用，这不是一个随机现象.

在特定条件下不能事先预知结果，且各个结果都具有频率稳定性的现象称为**随机现象**.

例 1.1.4 假定我们关心北京市一月份的平均气温，试讨论它与随机现象间的关系.

解：“北京市”和“一月份”的特定条件下，不能预知这个平均气温，它是一个不确定现象.

考虑到温室效应气体的作用，这不是一个随机现象.

大气中温室效应气体含量的变化及其缓慢，在短的时间范围内平均气温的观测数据近似地有频率稳定性，可看成随机现象。



1.2 描述未知现象的理想模型

确定性因素：能够事先确定的量.

随机性因素：不能事先确定，但具有频率稳定性的量.

1.2 描述未知现象的理想模型

确定性因素：能够事先确定的量.

随机性因素：不能事先确定，但具有频率稳定性的量.

在投掷硬币案例1.2中，“硬币质地均匀”是确定性因素.

1.2 描述未知现象的理想模型

确定性因素：能够事先确定的量.

随机性因素：不能事先确定，但具有频率稳定性的量.

在投掷硬币案例1.2中，“硬币质地均匀”是确定性因素.

在北京平均气温（例1.1.4）中，“北京”和“一月份”是确定因素.

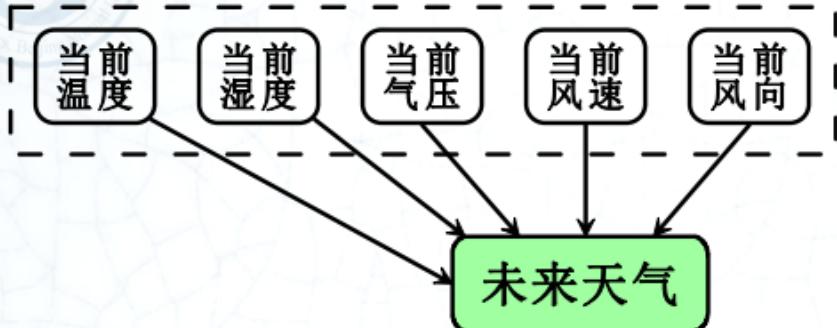


未来天气





确定因素





确定因素

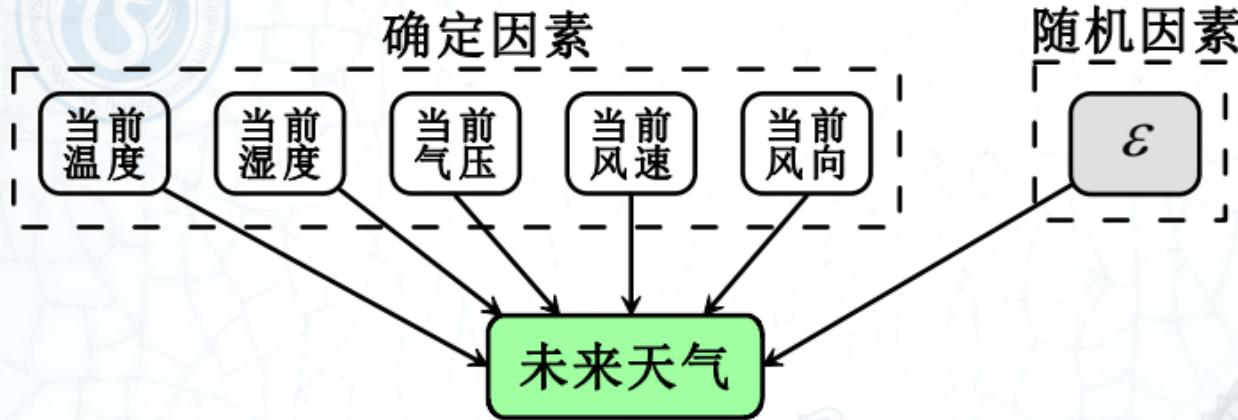


图 1.6 未来天气变化规律

$$y = f(x_1, x_2, x_3, x_4, x_5) + \varepsilon.$$

1.2.1 理想模型、现实模型与模型

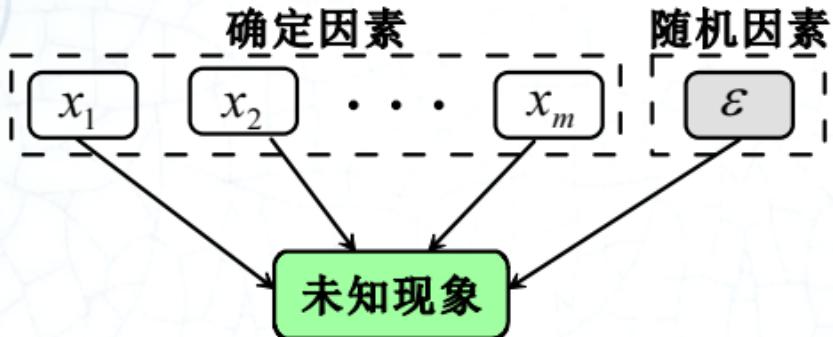
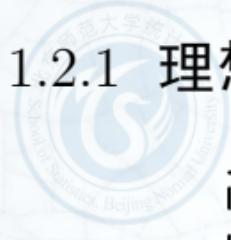


图 1.7 理想模型



1.2.1 理想模型、现实模型与模型

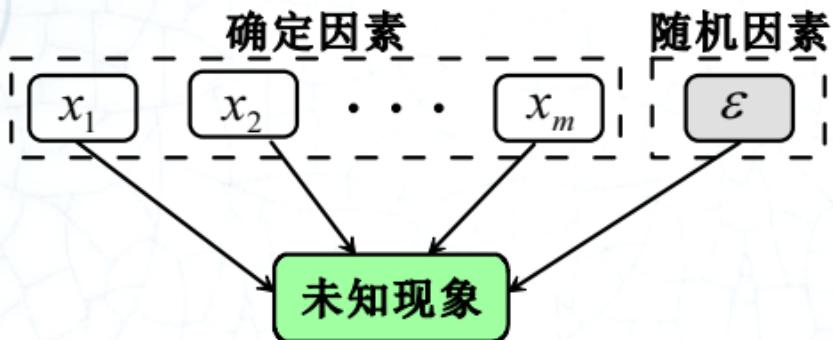


图 1.7 理想模型

$$y = f(x_1, x_2, \dots, x_m) + \varepsilon, \quad (1.1)$$

称(1.1)为描述未知现象的理想模型. 理想模型中的
 m 、 n 、 (x_1, x_2, \dots, x_m) 、 ε 和 f
都是我们感兴趣的研究对象.



在研究过程中只能得到理想模型的近似，即

$$y = \hat{f}(x_1, x_2, \dots, x_k) + \eta,$$



在研究过程中只能得到理想模型的近似，即

$$y = \hat{f}(x_1, x_2, \dots, x_k) + \eta, \quad (1.2)$$

称模型 (1.2) 为现实模型，或统计模型，简称为模型.

在研究过程中只能得到理想模型的近似，即

$$y = \hat{f}(x_1, x_2, \dots, x_k) + \eta, \quad (1.2)$$

称模型 (1.2) 为现实模型，或统计模型，简称为模型.

由于现实模型和理想模型之间的差异，意味着同一未知现象在研究的不同阶段会有不同的表征，研究的初始阶段可能表征为不具备频率稳定性的不确定现象，而在之后的研究中其表征越来越接近于随机现象或确定现象.



在研究过程中只能得到理想模型的近似，即

$$y = \hat{f}(x_1, x_2, \dots, x_k) + \eta, \quad (1.2)$$

称模型 (1.2) 为现实模型，或统计模型，简称为模型。

由于现实模型和理想模型之间的差异，意味着同一未知现象在研究的不同阶段会有不同的表征，研究的初始阶段可能表征为不具备频率稳定性的不确定现象，而在之后的研究中其表征越来越接近于随机现象或确定现象。

问题：案例1.1中的“神话”研究阶段是迷信吗？



在研究过程中只能得到理想模型的近似，即

$$y = \hat{f}(x_1, x_2, \dots, x_k) + \eta, \quad (1.2)$$

称模型 (1.2) 为现实模型，或统计模型，简称为模型。

由于现实模型和理想模型之间的差异，意味着同一未知现象在研究的不同阶段会有不同的表征，研究的初始阶段可能表征为不具备频率稳定性的不确定现象，而在之后的研究中其表征越来越接近于随机现象或确定现象。

问题：案例1.1中的“神话”研究阶段是迷信吗？

在未知现象研究过程中，认为现实模型无懈可击也是迷信，因为它仅是理想模型的近似。只有破除迷信，才能逐步认识未知现象，才是统计创新的动力。

例 1.2.5 考察模型 $y = \sin x + e$ 所产生数据的特点.



例 1.2.5 考察模型 $y = \sin x + e$ 所产生数据的特点.

解:

1. 在已知 x 和 e 的条件下, y 是必然现象.

例 1.2.5 考察模型 $y = \sin x + e$ 所产生数据的特点.

解:

- 在已知 x 和 e 的条件下, y 是必然现象. 如当 $x = 0, e = 0.5$ 时, 可以用如下的R程序代码模拟 y 的观测值

```
x=0;e=0.5;y=sin(x)+e;
```

模拟的结果永远是0.5.

例 1.2.5 考察模型 $y = \sin x + e$ 所产生数据的特点.

解:

- 在已知 x 和 e 的条件下, y 是必然现象. 如当 $x = 0, e = 0.5$ 时, 可以用如下的R程序代码模拟 y 的观测值

```
x=0;e=0.5;y=sin(x)+e;
```

模拟的结果永远是0.5.

- 在仅知 x 的条件下 y 的不能全确定, 所观测到的 y 是不确定现象. 下在已知 $x = 0$, 并且不知道 e 的情况下, 讨论 y 的观测值变化情况。



- 当真实的 e 是以0.5的可能性取1，以0.5的可能性取0的时候，所观测到的 y 值是一个随机现象。可以用如下的R程序代码模拟 y 的16次观测结果：

```
n=16;x=0;  
e=sample(0:1,n,replace = T);  
y=sin(x)+e;  
y
```

上述代码运行后，模拟的 y 在 RStudio 或 R 软件的控制台窗口显示如下：

```
[1] 1 1 0 1 0 1 1 1 1 0 0 1 0 1 1 0
```

注意，将如上代码再运行一次，得到的结果多半不同。



- 当真实的 e 是一个可以某种未知方式确定的量时，如

$$e = (-1)^{\lfloor \log_2 n \rfloor}, \quad n = 1, 2, \dots$$

时，所观测到的 y 值也具有不确定性，但此时 y 的观测不具备频率的稳定性。可以用如下的R程序代码模拟 y 的前16次观测值：

```
n=16;x=pi/2;  
e=(-1)^floor(log2(1:n));  
y=sin(x)+e;  
y
```

运行上述代码后，模拟的 y 在 RStudio 或 R 软件的控制台窗口显示如下：

```
[1] 2 0 0 2 2 2 2 0 0 0 0 0 0 0 0 2
```

1.2.2 模型的解读与追求目标

在现实研究过程中，无法知道与未知现象对应的理想模型。因此，

- 对于未知现象，在研究过程中没有正确的模型，只有近似的模型，不能迷信。

1.2.2 模型的解读与追求目标

在现实研究过程中，无法知道与未知现象对应的理想模型. 因此，

- 对于未知现象，在研究过程中没有正确的模型，只有近似的模型，不能迷信.
- 统计学追求的阶段目标：寻求更好的模型来刻画未知现象.

1.2.2 模型的解读与追求目标

在现实研究过程中，无法知道与未知现象对应的理想模型。因此，

- 对于未知现象，在研究过程中没有正确的模型，只有近似的模型，不能迷信。
- 统计学追求的阶段目标：寻求更好的模型来刻画未知现象。
- 统计学追求的理想目标：通过阶段目标逐渐认识未知现象，其途径是未知现象研究流程图。

记住这三条，这是统计学研究的基本观点与创新动力。

1.3 统计学的应用领域

在客观世界中，需要认识的现象无穷无尽，例如：



1.3 统计学的应用领域

在客观世界中，需要认识的现象无穷无尽，例如：

- 国家想要知道经济运行是否正常；

1.3 统计学的应用领域

在客观世界中，需要认识的现象无穷无尽，例如：

- 国家想要知道经济运行是否正常；
- 客运计划者希望预知春运高峰的客流量的分布；

1.3 统计学的应用领域

在客观世界中，需要认识的现象无穷无尽，例如：

- 国家想要知道经济运行是否正常；
- 客运计划者希望预知春运高峰的客流量的分布；
- 人们希望读懂遗传天书——基因序列（由A、C、G和T构成）；

1.3 统计学的应用领域

在客观世界中，需要认识的现象无穷无尽，例如：

- 国家想要知道经济运行是否正常；
- 客运计划者希望预知春运高峰的客流量的分布；
- 人们希望读懂遗传天书——基因序列（由A、C、G和T构成）；
- 人们关心大气碳浓度与气温之间的关系；

1.3 统计学的应用领域

在客观世界中，需要认识的现象无穷无尽，例如：

- 国家想要知道经济运行是否正常；
- 客运计划者希望预知春运高峰的客流量的分布；
- 人们希望读懂遗传天书——基因序列（由A、C、G和T构成）；
- 人们关心大气碳浓度与气温之间的关系；
- 在座者希望知道自己的学习能力如何；

1.3 统计学的应用领域

在客观世界中，需要认识的现象无穷无尽，例如：

- 国家想要知道经济运行是否正常；
- 客运计划者希望预知春运高峰的客流量的分布；
- 人们希望读懂遗传天书——基因序列（由A、C、G和T构成）；
- 人们关心大气碳浓度与气温之间的关系；
- 在座者希望知道自己的学习能力如何；
- 保险公司想要知道各种灾害的分布情况；

1.3 统计学的应用领域

在客观世界中，需要认识的现象无穷无尽，例如：

- 国家想要知道经济运行是否正常；
- 客运计划者希望预知春运高峰的客流量的分布；
- 人们希望读懂遗传天书——基因序列（由A、C、G和T构成）；
- 人们关心大气碳浓度与气温之间的关系；
- 在座者希望知道自己的学习能力如何；
- 保险公司想要知道各种灾害的分布情况；



- 生产管理者想要知道生产线是否在正常工作；



- 生产管理者想要知道生产线是否在正常工作；
- 药厂想要知道新研制的药品是否更有效；

- 生产管理者想要知道生产线是否在正常工作；
- 药厂想要知道新研制的药品是否更有效；
- 什么样的饮食习惯更有利健康；

- 生产管理者想要知道生产线是否在正常工作；
- 药厂想要知道新研制的药品是否更有效；
- 什么样的饮食习惯更有利健康；
- 粮食产量是否能够满足需求；

- 生产管理者想要知道生产线是否在正常工作；
- 药厂想要知道新研制的药品是否更有效；
- 什么样的饮食习惯更有利健康；
- 粮食产量是否能够满足需求；
- 某减肥产品是否象其广告声称的那样有效率；

- 生产管理者想要知道生产线是否在正常工作；
- 药厂想要知道新研制的药品是否更有效；
- 什么样的饮食习惯更有利健康；
- 粮食产量是否能够满足需求；
- 某减肥产品是否象其广告声称的那样有效率；
- 明天是否下雨；

- 生产管理者想要知道生产线是否在正常工作；
- 药厂想要知道新研制的药品是否更有效；
- 什么样的饮食习惯更有利健康；
- 粮食产量是否能够满足需求；
- 某减肥产品是否象其广告声称的那样有效率；
- 明天是否下雨；
-

这些未知现象的研究需要统计学。

在互联网的信息时代，“数据”获取成本越来越低，不断形成有待分析的行业“大数据”。统计知识成为公民的必备素质，各行业中都需要大量的掌握统计学知识的人才：

1. 政府部门：人口状况、就业状况、经济状况、发展规划……。

在互联网的信息时代，“数据”获取成本越来越低，不断形成有待分析的行业“大数据”。统计知识成为公民的必备素质，各行业中都需要大量的掌握统计学知识的人才：

1. 政府部门：人口状况、就业状况、经济状况、发展规划……。
2. 自然科学：论文是否能够发表在主要杂志上，很大程度上依赖于能否正确使用统计方法。

在互联网的信息时代，“数据”获取成本越来越低，不断形成有待分析的行业“大数据”。统计知识成为公民的必备素质，各行业中都需要大量的掌握统计学知识的人才：

1. 政府部门：人口状况、就业状况、经济状况、发展规划……
2. 自然科学：论文是否能够发表在主要杂志上，很大程度上依赖于能否正确使用统计方法。
3. 社会科学：经济、教育、心理、文字研究……
4. 工业：研究新产品、质量控制、市场开发……

在互联网的信息时代，“数据”获取成本越来越低，不断形成有待分析的行业“大数据”。统计知识成为公民的必备素质，各行业中都需要大量的掌握统计学知识的人才：

1. 政府部门：人口状况、就业状况、经济状况、发展规划……
2. 自然科学：论文是否能够发表在主要杂志上，很大程度上依赖于能否正确使用统计方法。
3. 社会科学：经济、教育、心理、文字研究……
4. 工业：研究新产品、质量控制、市场开发……
5. 农业：植物生长、产量预测、病虫害研究……

6. 日常生活（报纸杂志上常刊登以统计为基础的文章、民意调查、公正性问题、商品广告等）



6. 日常生活（报纸杂志上常刊登以统计为基础的文章、民意调查、公正性问题、商品广告等）

纽约时报专栏文章(Steve Lohr,2009年8月5日)

For Today's Graduate, Just One Word: Statistics

6. 日常生活（报纸杂志上常刊登以统计为基础的文章、民意调查、公正性问题、商品广告等）

纽约时报专栏文章(Steve Lohr,2009年8月5日)

For Today's Graduate, Just One Word: Statistics

随着研究的进展，未知现象的背景涉及越来越多的领域专业知识，这成为统计学家对于该现象的研究阻力；而大多数领域专家虽然熟知专业知识，但不掌握统计学理论的最新理论研究成果，这成为领域专家进一步研究的阻力。

6. 日常生活（报纸杂志上常刊登以统计为基础的文章、民意调查、公正性问题、商品广告等）

纽约时报专栏文章(Steve Lohr,2009年8月5日)

For Today's Graduate, Just One Word: Statistics

随着研究的进展，未知现象的背景涉及越来越多的领域专业知识，这成为统计学家对于该现象的研究阻力；而大多数领域专家虽然熟知专业知识，但不掌握统计学理论的最新理论研究成果，这成为领域专家进一步研究的阻力。

为提高复杂未知现象的研究效率，我们应该组建由领域专家和统计学家构成的跨学科交叉研究团队，在研究过程中取长补短，高效获取创新性成果，加速未知现象研究流程。

1.4 统计学、数学、概率及统计软件

在总体研究思路上，

- 统计学：建立在观测数据基础之上的归纳推理，由局部归纳总体；

1.4 统计学、数学、概率及统计软件

在总体研究思路上，

- 统计学：建立在观测数据基础之上的归纳推理，由局部归纳总体；
- 数学：建立在公理体系基础之上的演绎推理，由公理演绎理论体系；

1.4 统计学、数学、概率及统计软件

在总体研究思路上，

- 统计学：建立在观测数据基础之上的归纳推理，由局部归纳总体；
- 数学：建立在公理体系基础之上的演绎推理，由公理演绎理论体系；
- 概率论：数学的一个研究随机现象的分支，其公理体系由概率的非负性、规范性和可列可加性构成；

1.4 统计学、数学、概率及统计软件

在总体研究思路上，

- 统计学：建立在观测数据基础之上的归纳推理，由局部归纳总体；
- 数学：建立在公理体系基础之上的演绎推理，由公理演绎理论体系；
- 概率论：数学的一个研究随机现象的分支，其公理体系由概率的非负性、规范性和可列可加性构成；
- 统计学软件是已有统计原理与知识的集成.

在研究的过程中，统计学归纳是建立在统计学原理基础上的演绎；



在研究的过程中，统计学归纳是建立在统计学原理基础上的演绎；而归纳又是数学发现新知识的一条重要的途径。



在研究的过程中，统计学归纳是建立在统计学原理基础上的演绎；而归纳又是数学发现新知识的一条重要的途径。

数学研究中基于已有知识归纳出的结论需要严谨的演绎论证；

在研究的过程中，统计学归纳是建立在统计学原理基础上的演绎；而归纳又是数学发现新知识的一条重要的途径。

数学研究中基于已有知识归纳出的结论需要严谨的演绎论证；统计学研究中归纳出的结论需要实践验证，以发现新的研究问题.

在研究的过程中，统计学归纳是建立在统计学原理基础上的演绎；而归纳又是数学发现新知识的一条重要的途径。

数学研究中基于已有知识归纳出的结论需要严谨的演绎论证；统计学研究中归纳出的结论需要实践验证，以发现新的研究问题。因此

1. 研究统计模型和方法的性质需要坚实的数学和概率论基础；

在研究的过程中，统计学归纳是建立在统计学原理基础上的演绎；而归纳又是数学发现新知识的一条重要的途径。

数学研究中上基于已有知识归纳出的结论需要严谨的演绎论证；统计学研究中归纳出的结论需要实践验证，以发现新的研究问题. 因此

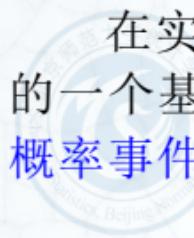
1. 研究统计模型和方法的性质需要坚实的数学和概率论基础；
2. 在统计思想指导下使用统计软件能解决大量应用问题，盲目地使用统计软件会降低结论的可信性.

在研究的过程中，统计学归纳是建立在统计学原理基础上的演绎；而归纳又是数学发现新知识的一条重要的途径。

数学研究中上基于已有知识归纳出的结论需要严谨的演绎论证；统计学研究中归纳出的结论需要实践验证，以发现新的研究问题. 因此

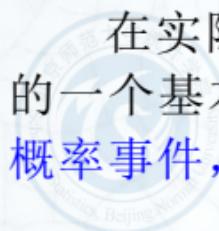
1. 研究统计模型和方法的性质需要坚实的数学和概率论基础；
2. 在统计思想指导下使用统计软件能解决大量应用问题，**盲目地使用统计软件会降低结论的可信性.**

掌握成熟的统计思想和方法，不需要高深的数学基础知识，仅需牢记几个基本的统计学原理.



在实际问题中，经常面临几个结论选择问题，其推断的一个基本原理：所得结论应该保证所观测到的事件是大概率事件，而不是与该结论相矛盾的小概率事件。

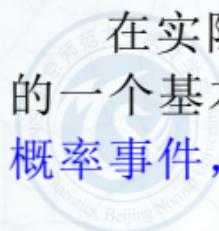




在实际问题中，经常面临几个结论选择问题，其推断的一个基本原理：**所得结论应该保证所观测到的事件是大概率事件，而不是与该结论相矛盾的小概率事件。**

例 1.4.6 如果连续投掷一枚骰子10次，结果都是1点。你能够认为这枚骰子的质地均匀吗？为什么？





在实际问题中，经常面临几个结论选择问题，其推断的一个基本原理：**所得结论应该保证所观测到的事件是大概率事件，而不是与该结论相矛盾的小概率事件。**

例 1.4.6 如果连续投掷一枚骰子**10**次，结果都是**1**点。你能够认为这枚骰子的质地均匀吗？为什么？

解：如果它是均匀的，各个面出现的可能性都是**1/6**.

在实际问题中，经常面临几个结论选择问题，其推断的一个基本原理：**所得结论应该保证所观测到的事件是大概率事件，而不是与该结论相矛盾的小概率事件。**

例 1.4.6 如果连续投掷一枚骰子**10**次，结果都是**1**点。你能够认为这枚骰子的质地均匀吗？为什么？

解：如果它是均匀的，各个面出现的可能性都是**1/6**. 从而出现连续**10**次**1**点的概率为

$$\left(\frac{1}{6}\right)^{10} \approx 0.000000016538,$$

这是几乎不会发生的与骰子质地均匀相矛盾的小概率事件。

在实际问题中，经常面临几个结论选择问题，其推断的一个基本原理：所得结论应该保证所观测到的事件是大概率事件，而不是与该结论相矛盾的小概率事件。

例 1.4.6 如果连续投掷一枚骰子10次，结果都是1点。你能够认为这枚骰子的质地均匀吗？为什么？

解：如果它是均匀的，各个面出现的可能性都是 $1/6$ 。从而出现连续10次1点的概率为

$$\left(\frac{1}{6}\right)^{10} \approx 0.000000016538,$$

这是几乎不会发生的与骰子质地均匀相矛盾的小概率事件。因此认为这枚骰子的质地不均匀。

小结

1. 统计学的定义
2. 研究未知现象流程图
3. 描述未知现象的理想模型
4. 理想模型、现实模型和模型
5. 迷信与创新
6. 模型的解读与追求目标
7. 复杂未知现象研究团队的效率
8. 统计学、数学、概率及统计软件



课程目的：本门课程概要介绍统计学的思维方式、基本的统计思想与方法，使同学对于统计学有一个初步的整体了解，体会统计学的理论与应用价值，提升统计学创新研究能力。





课程目的：本门课程概要介绍统计学的思维方式、基本的统计思想与方法，使同学对于统计学有一个初步的整体了解，体会统计学的理论与应用价值，提升统计学创新研究能力。

在终极的分析中，一切知识都是历史；

在抽象的意义下，一切科学都是数学；

在理性的基础上，所有的判断都是统计学。

C. R. Rao

精品共享课网址查看方式：登录爱课程社区 sns.icourses.cn



图 1.8 精品共享课登陆界面



学习社区 首页 学友圈 开放课堂 消息

Teacher Li 设置 登录 退出

站内信

我的发言 提到我的 我的评论

我的开放课堂 我的学习笔记 我的答疑解惑 我的学习资源 公开课学习记录

我的收藏 我的工作室

精彩专题栏目

去猜 朋友 图片 视频

还能输入 140 字

发布

全部发言

Teacher Li 2015年初，已经完成《统计学导论》课堂实况录像，需要时间加工处理后才能上传。从2009年到2015年，课程的内容已经发生了很大变化。

10月12日 18:28 转帖 | 点评 | 编辑 | 收藏

Teacher Li 大家好！我是《统计学导论》的主讲教师之一，欢迎大家一起来探讨统计学的基本思想与原理。

10月12日 18:21 转帖 | 点评 | 编辑 | 收藏

Teacher Li

0 0 0

未关注的 关注我的 学友

推荐阅读

lawg240 iwyEtc... 逻辑思维 加关注

天者地明 易易 吴可为 加关注

华平 张萌 我的第一加关注

图 1.9 精品共享课登陆界面

第二章

1.5 常用统计软件与MATLAB简介

1.5.1 常用统计软件

SPSS:

SAS:

S-PLUS:

R软件:

MATLAB:



1.5.2 MATLAB简介

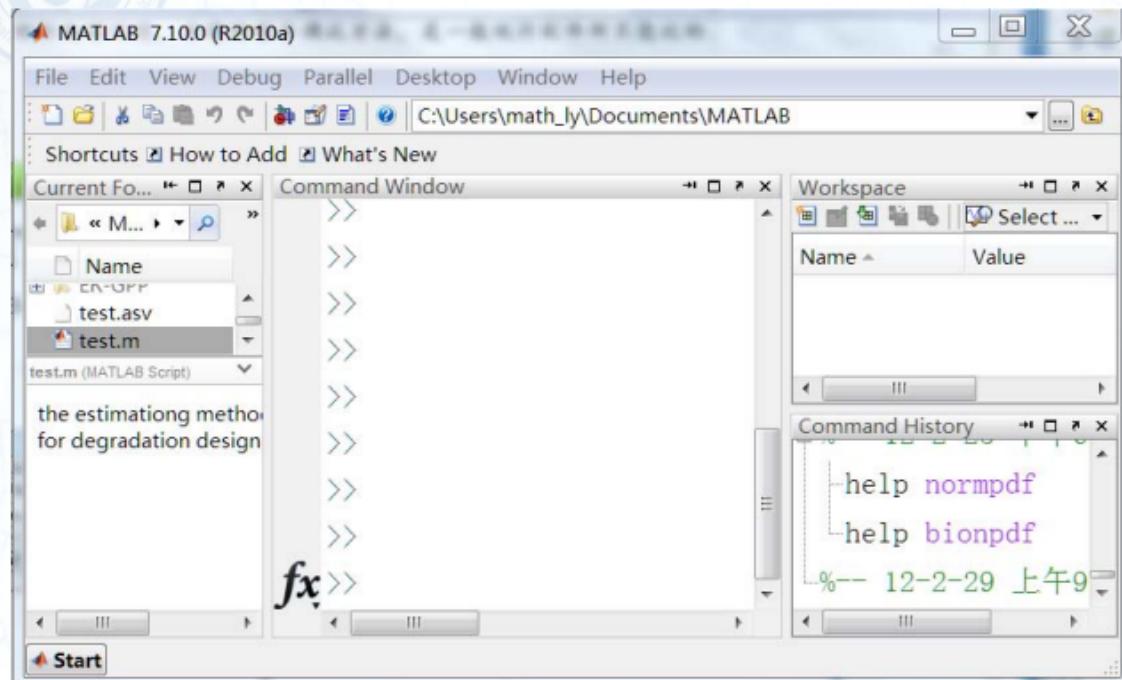


图 1.10 MATLAB软件窗口



>>

表示命令行提示符.





>>

表示命令行提示符. “回车键” 表示让MATLAB执行刚刚键入的指令或程序代码.

help:

```
>>help help
```





>>

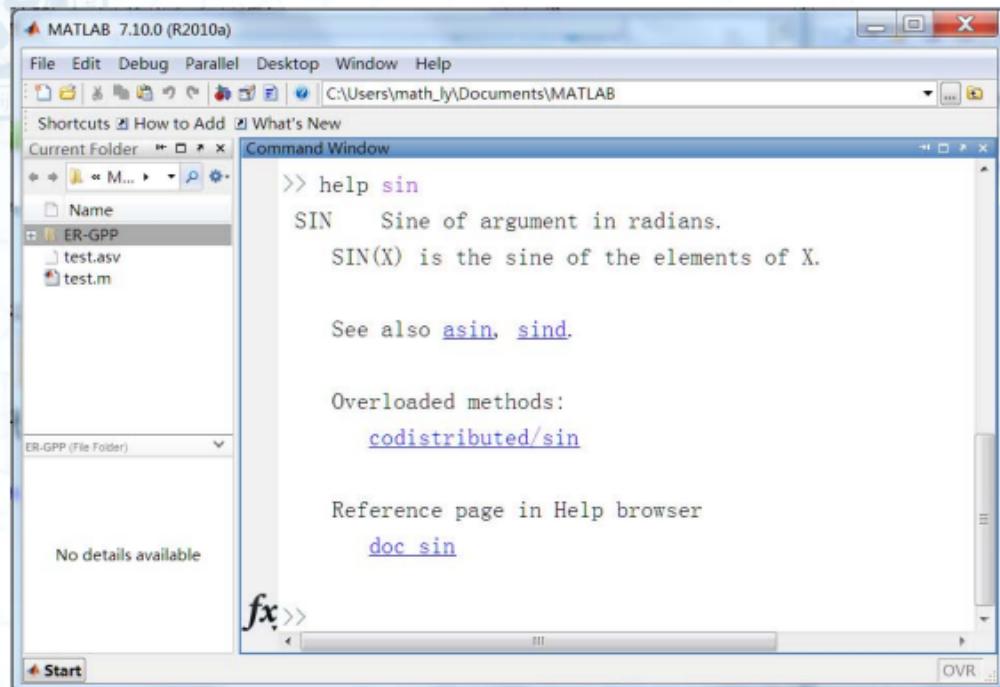
表示命令行提示符. “回车键” 表示让MATLAB执行刚刚键入的指令或程序代码.

help:

>>help help

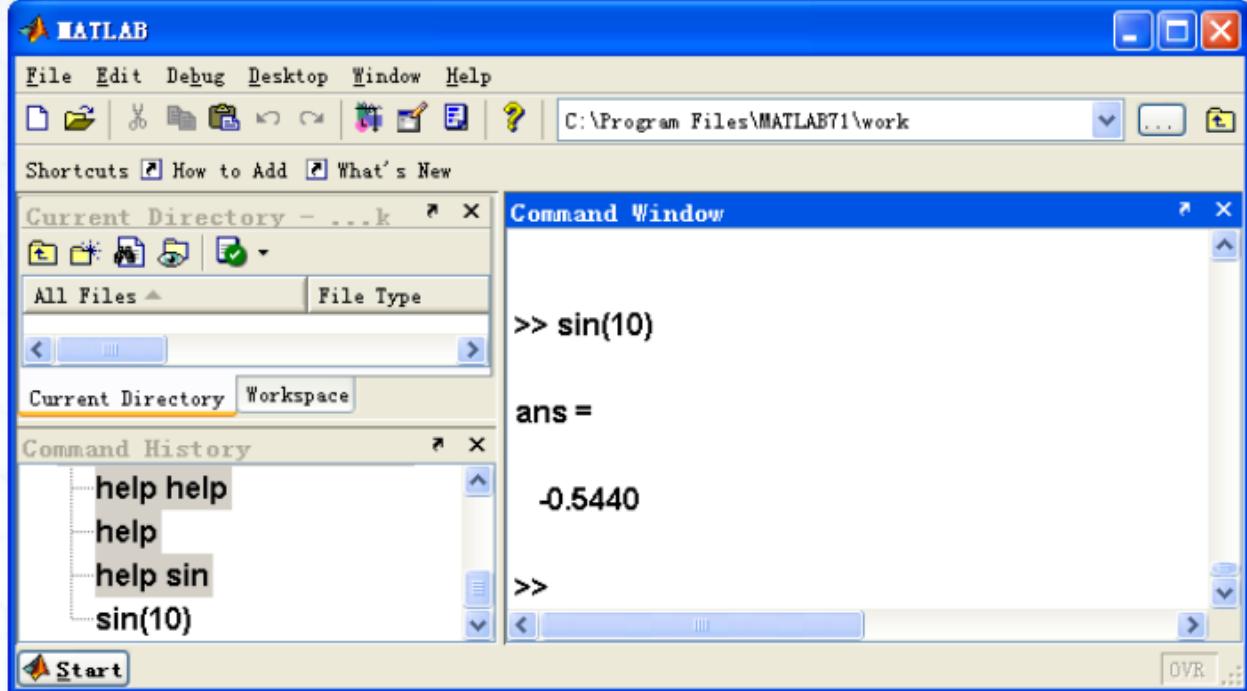
demo: demo

help sin

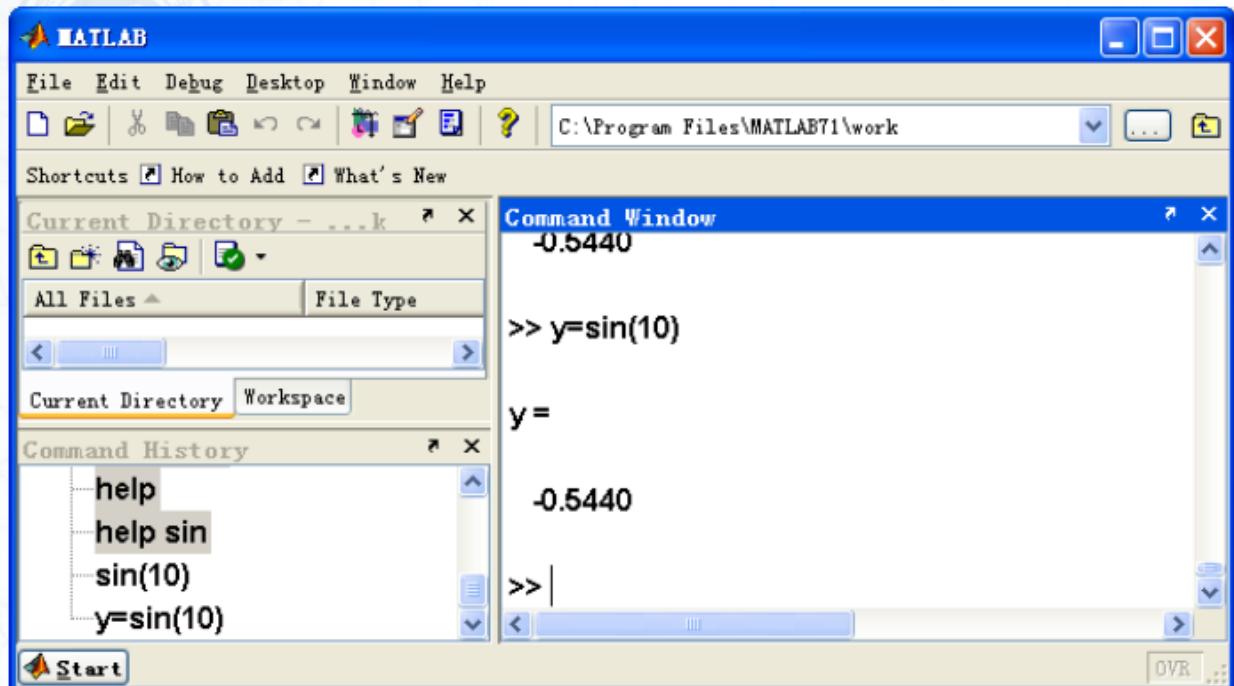


sin函数的帮助信息

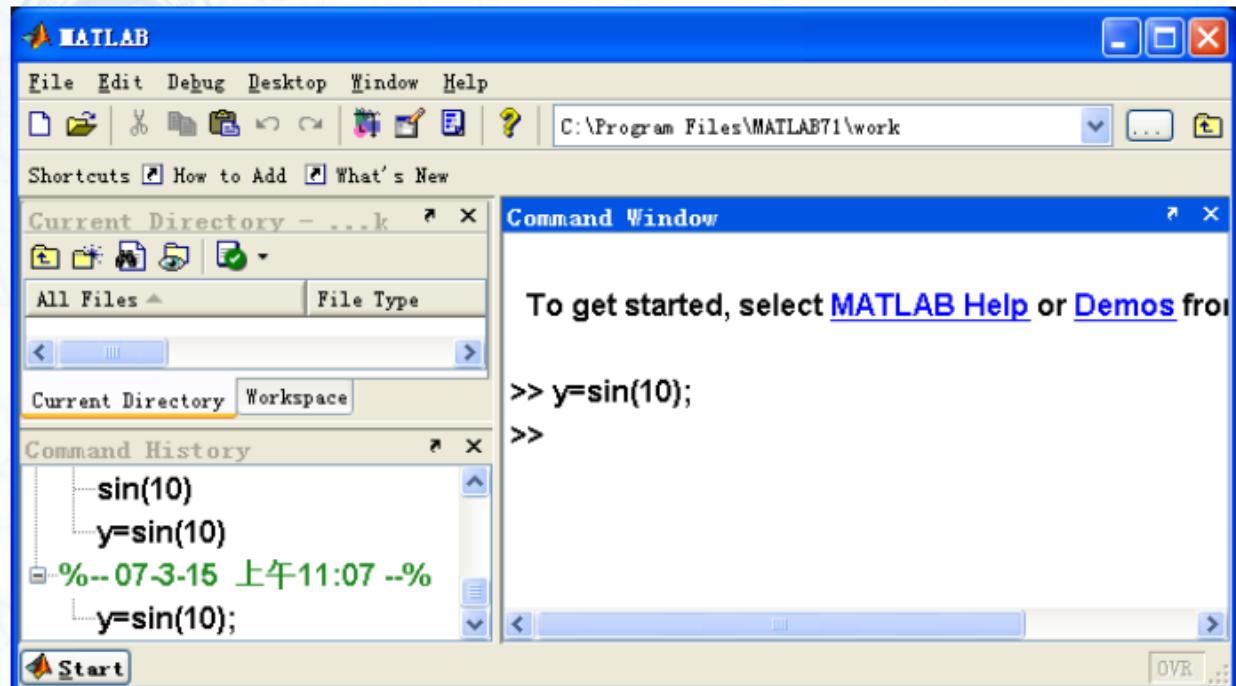
$\sin(10)$: ans,



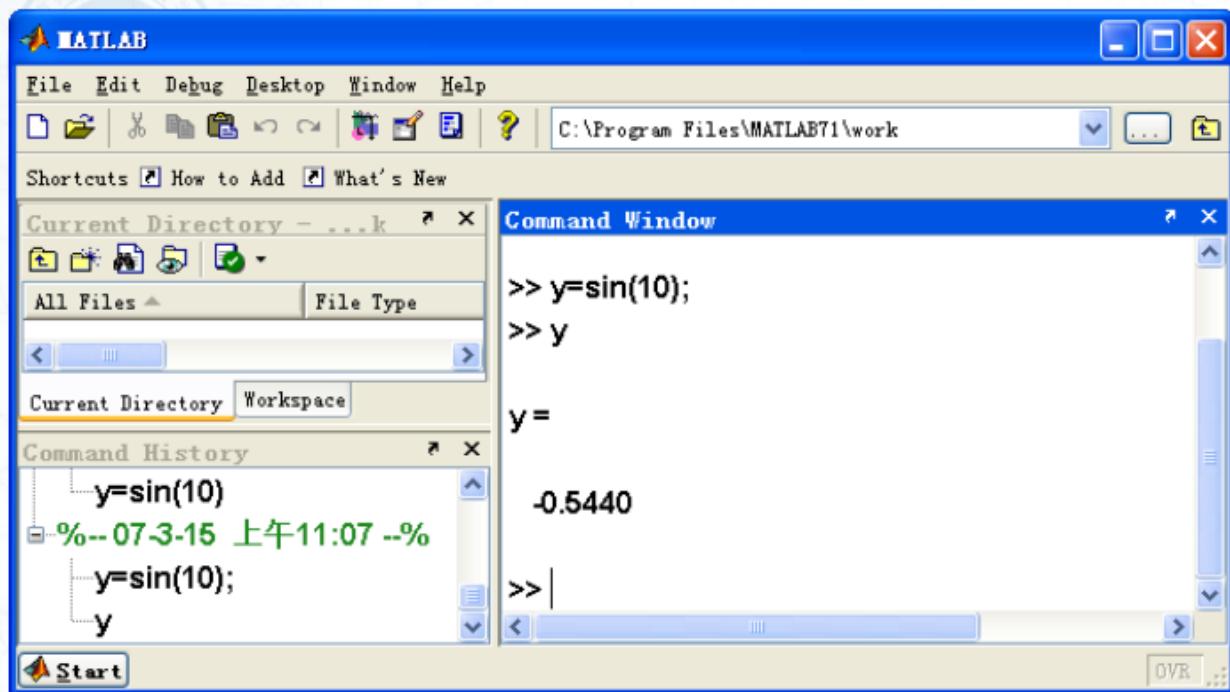
$y=\sin(10):$



```
y=sin(10);
```



输入变量名后回车：



矩阵（向量）用方括号表示. [1 2 3] [1,2,3]

(1, 2, 3)



矩阵（向量）用方括号表示. [1 2 3] [1,2,3]

(1, 2, 3)

[1;
2;
3]

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$



$$[1 \ 2 \ 1; 2 \ 4 \ 1] \quad \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 1 \end{bmatrix}$$

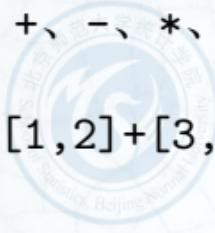
$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 1 \end{pmatrix}$$

表达式 $x=[1 \ 2 \ 1; 2 \ 4]$ 是否有问题？



+、 -、 *、 / =

[1,2]+[3,4] [4,6];



+、 -、 *、 / =

[1,2]+[3,4] [4,6];

[1,2,3]+[3,4]

[1,2]+3[4,5];



+、 -、 *、 / =

[1,2]+[3,4] [4,6];

[1,2,3]+[3,4]

[1,2]+3[4,5];

[1,2]*[1;2]5;

+、 -、 *、 / =

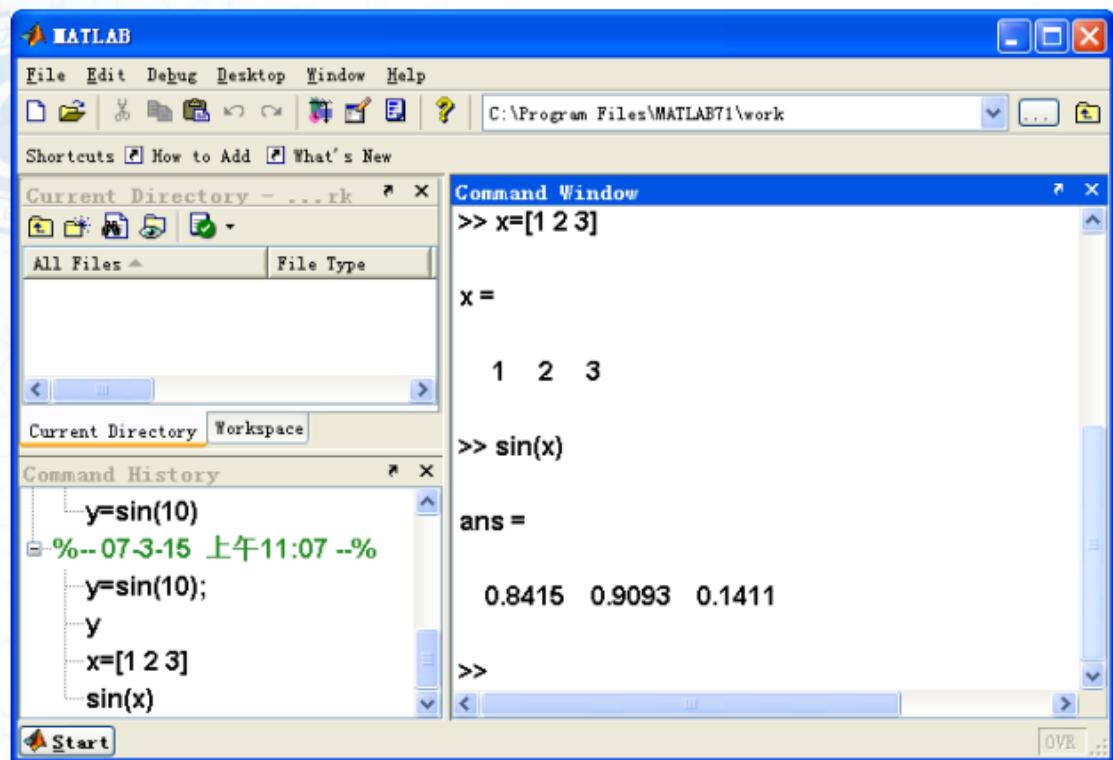
[1,2]+[3,4] [4,6];

[1,2,3]+[3,4]

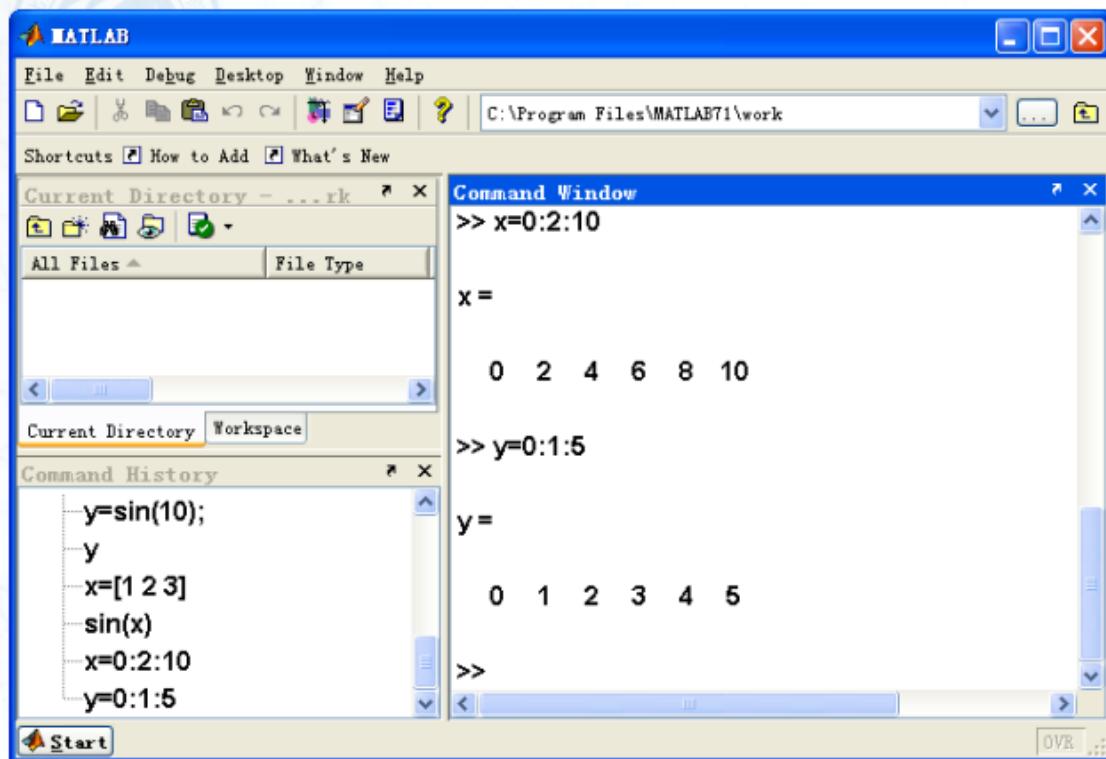
[1,2]+3[4,5];

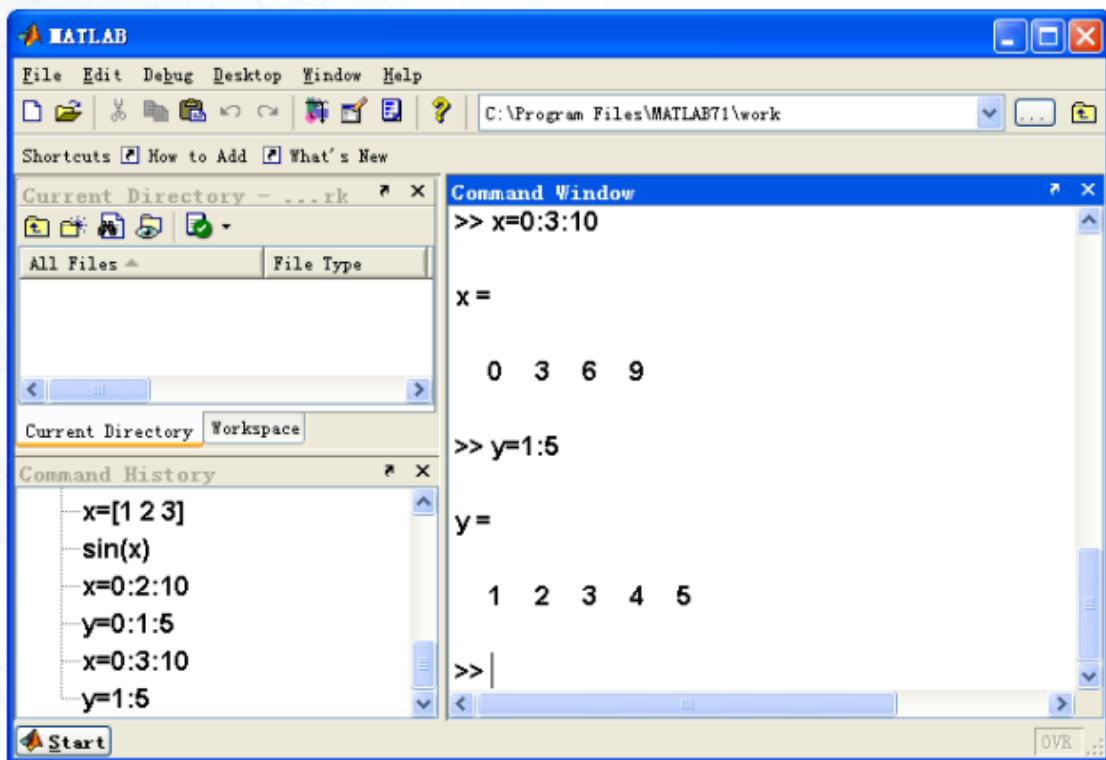
[1,2]*[1;2]5;

[1,2]*[1,2,3]



分量为等差数列的行向量表示： $0:2:10$ $0:1:5$

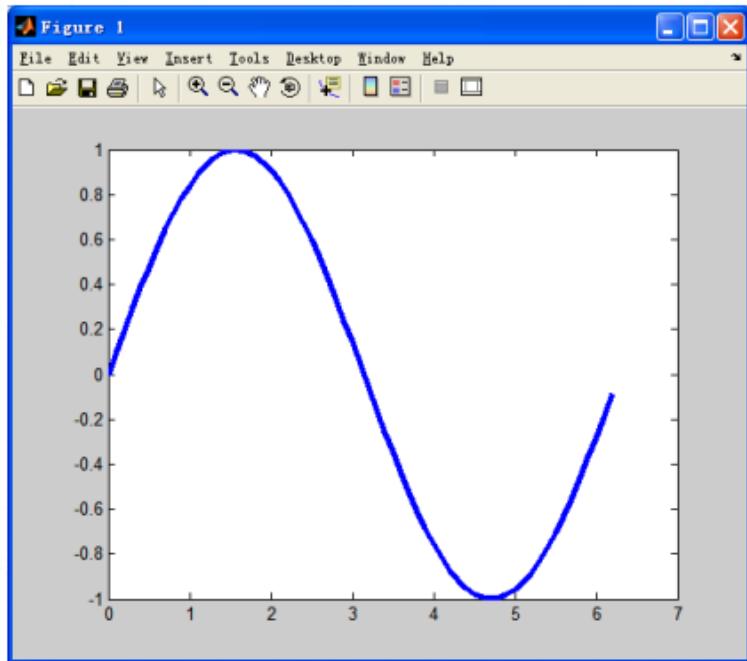




0:3:10的最后一个分量是9； 1:5是1:1:5的简写

MATLAB提供了方便的制图功能，可以绘制任何函数的图像。

```
>> x=0:0.1:2*pi;  
>> y=sin(x);  
>> plot(x,y);  
>>
```



利用MATLAB软件，可以模拟随机现象，如模拟抛硬币、掷骰子等。





利用MATLAB软件，可以模拟随机现象，如模拟抛硬币、掷骰子等。

模拟随机现象，主要是通过产生与随机现象的统计规律相一致的随机数来实现。

用 X 表示摸出球上的号码，事件 $\{X = k\}$ 发生的概率为

$$\mathbb{P}(X = k) = \frac{1}{n}, \quad 1 \leq k \leq n.$$



利用MATLAB软件，可以模拟随机现象，如模拟抛硬币、掷骰子等。

模拟随机现象，主要是通过产生与随机现象的统计规律相一致的随机数来实现。

用 X 表示摸出球上的号码，事件 $\{X = k\}$ 发生的概率为

$$\mathbb{P}(X = k) = \frac{1}{n}, \quad 1 \leq k \leq n. \quad (1.3)$$

称 X 服从离散均匀分布（参数为 n ），称 X 的观测值为离散均匀分布随机数（参数为 n ）。

unidrnd:

任何软件的随机数函数都是通过特定的迭代算法产生的，具有周期性，称为伪随机数.

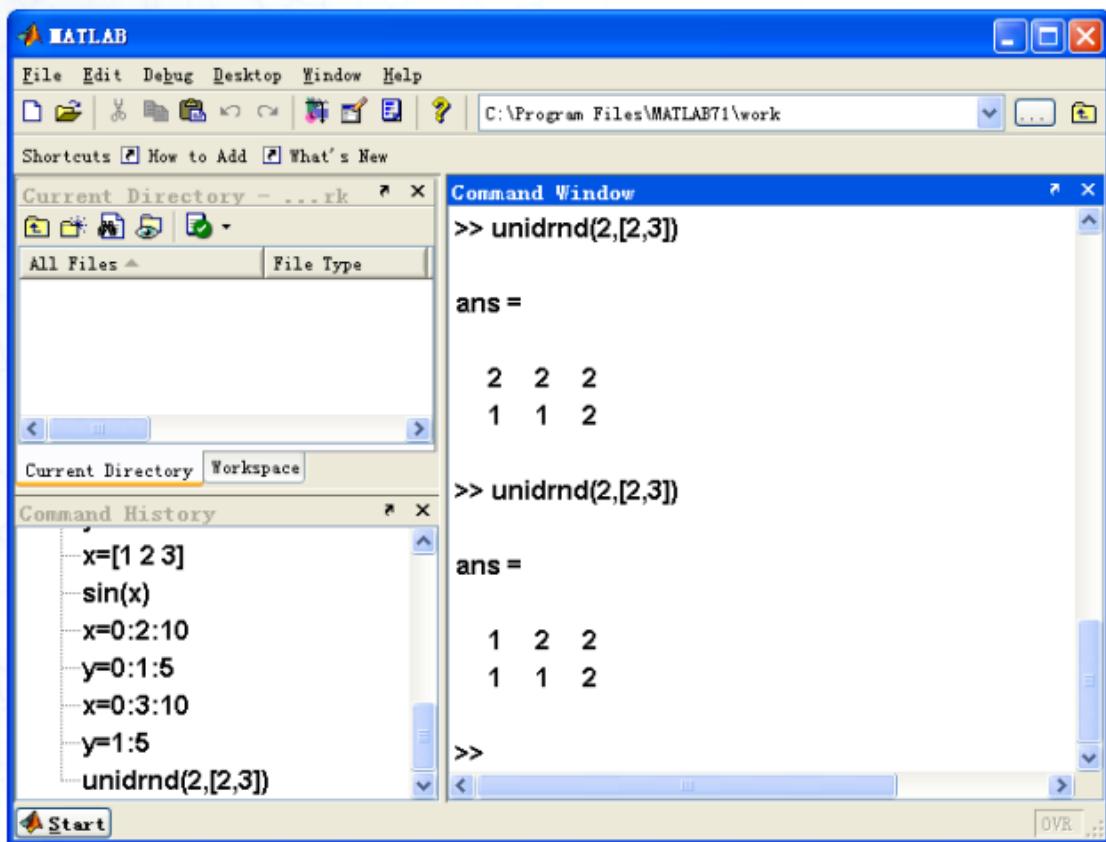
unidrnd:

任何软件的随机数函数都是通过特定的迭代算法产生的，具有周期性，称为伪随机数。

MATLAB函数的命名规则： unidrnd uniform discrete random

unidrnd(n): unidrnd(2)

unidrnd(n, [m, l]) unidrnd(n, m, l): unidrnd(2, [2,



unidrnd(2, [2,3])的运行结果

例 1.5.7 用计算机模拟100次投掷一枚均匀骰子的实验结果，并写出相应的MATLAB命令代码.

解：



例 1.5.7 用计算机模拟100次投掷一枚均匀骰子的实验结果，并写出相应的MATLAB命令代码.

解：

```
unidrnd(6,1,100)
```

第*i*个分量是第*i*次掷骰子实验结果的模拟.

例 1.5.7 用计算机模拟100次投掷一枚均匀骰子的实验结果，并写出相应的MATLAB命令代码.

解：

```
unidrnd(6,1,100)
```

第 i 个分量是第 i 次掷骰子实验结果的模拟.
逻辑运算的结果只有“真”和“假”，分别用“1”和
“0”表示.

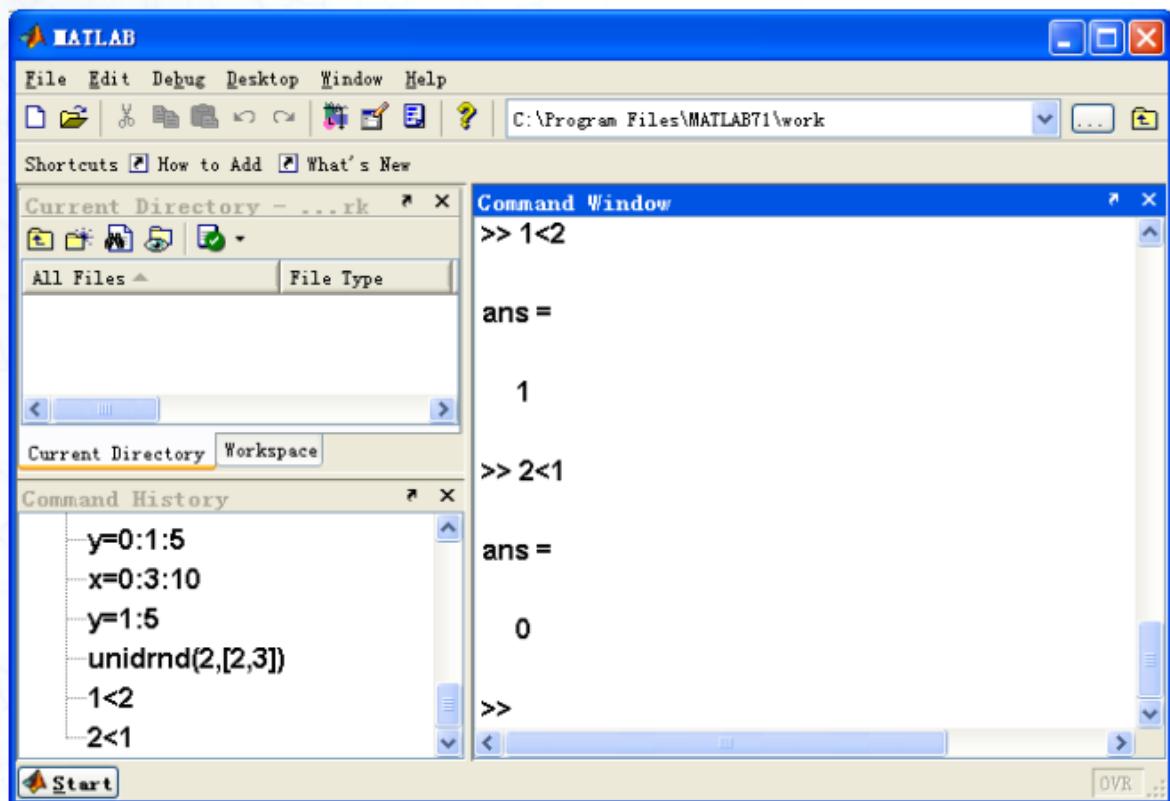
例 1.5.7 用计算机模拟100次投掷一枚均匀骰子的实验结果，并写出相应的MATLAB命令代码.

解：

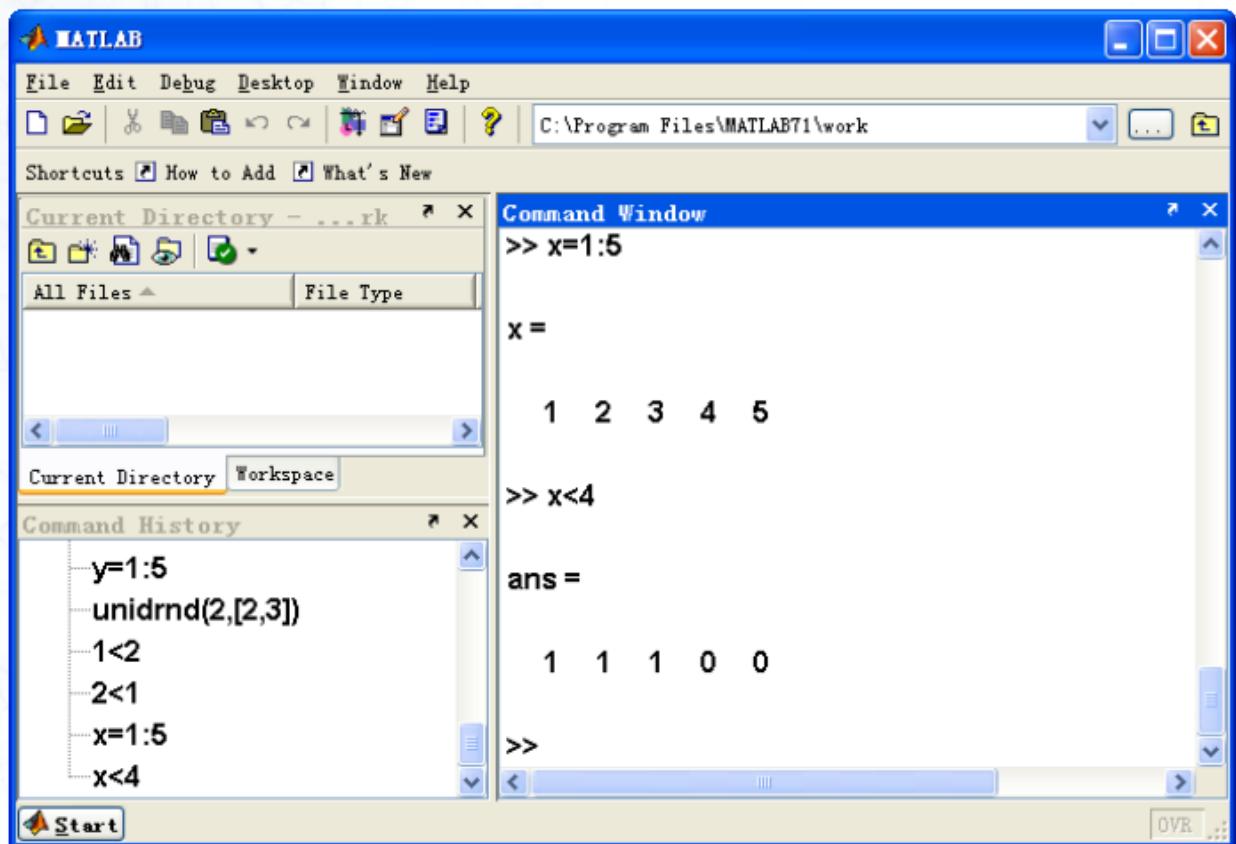
```
unidrnd(6,1,100)
```

第 i 个分量是第 i 次掷骰子实验结果的模拟.
逻辑运算的结果只有“真”和“假”，分别用“1”和
“0”表示.

>、<、>=、<=、== ^=



数之间的逻辑运算

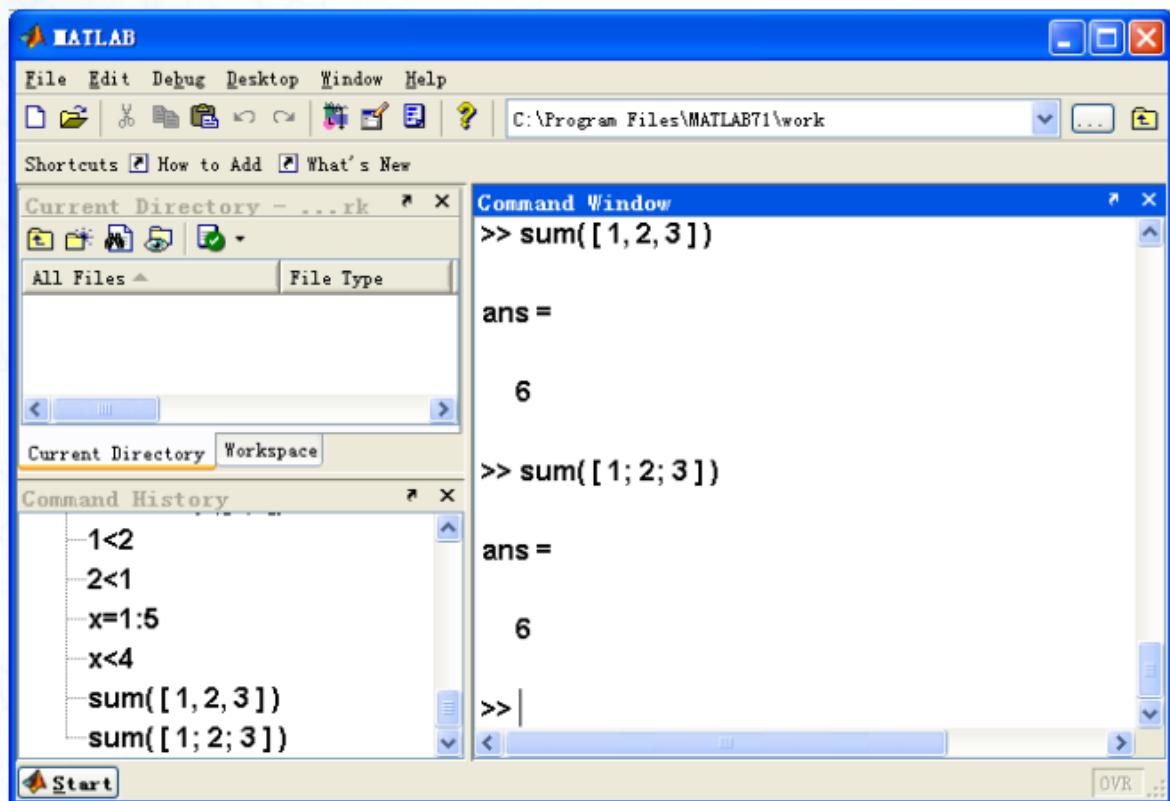


矩阵(向量)与数之间的逻辑运算

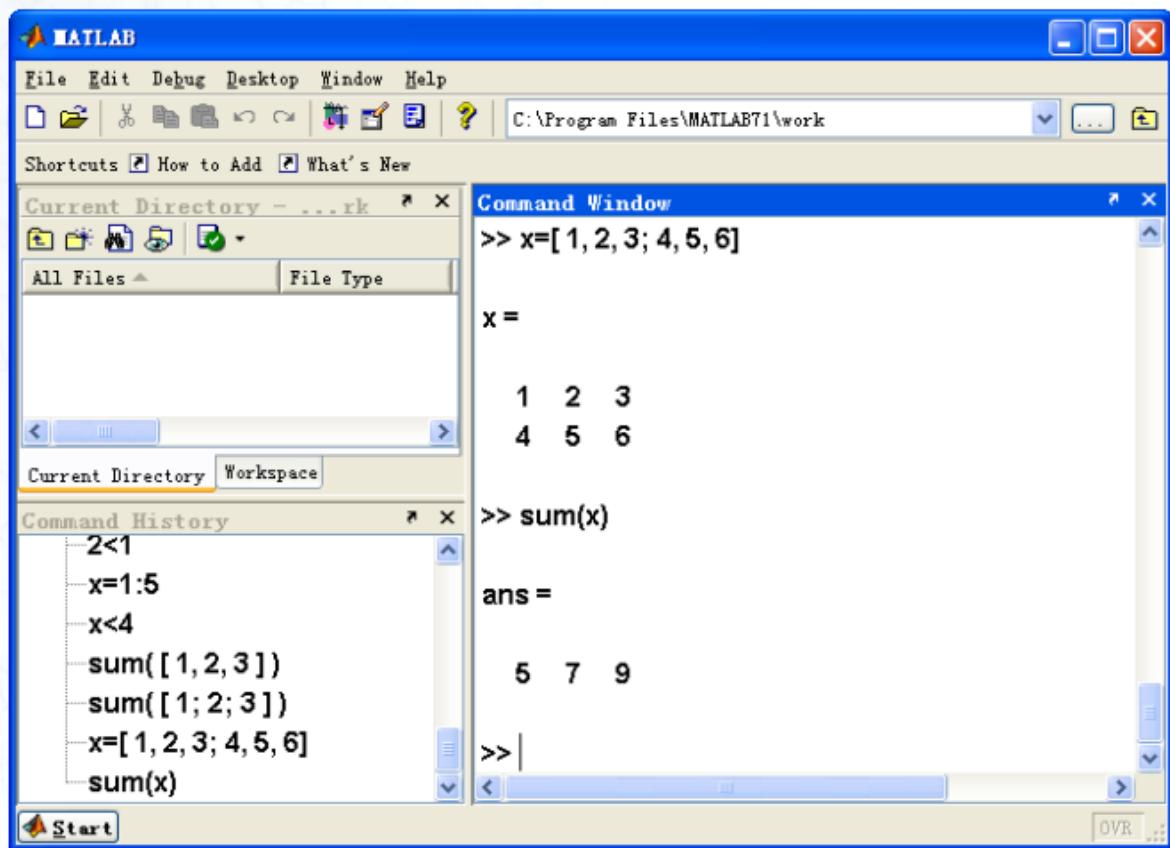
sum(X,1):

sum(X,2):

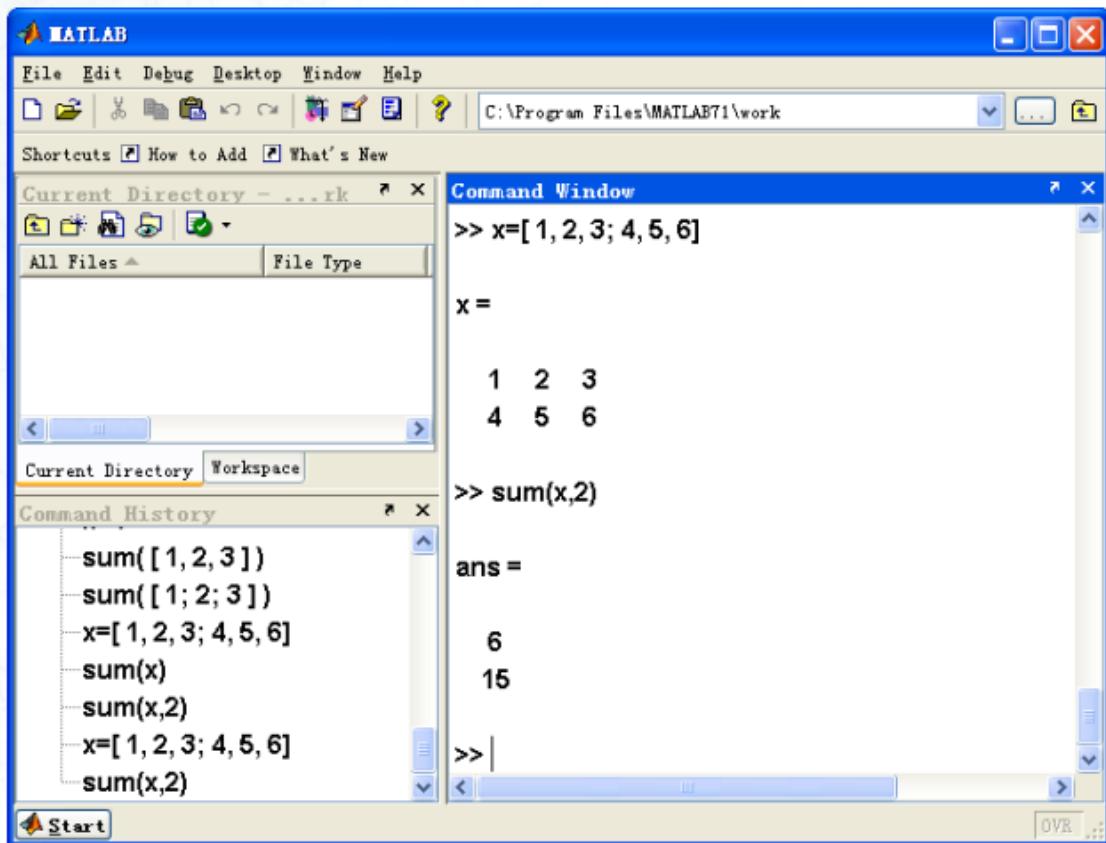
sum(X):



$\text{sum}([1,2,3])$ 与 $\text{sum}([1;2;3])$ 的运行结果



用sum求X的各个列的元素之和



用sum求X的各个行的元素之和

例 1.5.8 试模拟1000次投掷一枚均匀骰子的实验，并对于 $i = 1, 2, \dots, 10, 50, 100$ ，分别计算出前 $i \times 10$ 次实验中各个结果出现的频率。请写出完成上述任务的MATLAB命令代码。

解：

例 1.5.8 试模拟1000次投掷一枚均匀骰子的实验，并对于 $i = 1, 2, \dots, 10, 50, 100$ ，分别计算出前 $i \times 10$ 次实验中各个结果出现的频率。请写出完成上述任务的MATLAB命令代码。

解：

```
x=unidrnd(6,1000,1);
```

例 1.5.8 试模拟1000次投掷一枚均匀骰子的实验，并对于 $i = 1, 2, \dots, 10, 50, 100$ ，分别计算出前 $i \times 10$ 次实验中各个结果出现的频率。请写出完成上述任务的MATLAB命令代码。

解：

```
x=unidrnd(6,1000,1);  
y=x(1:10);
```

例 1.5.8 试模拟1000次投掷一枚均匀骰子的实验，并对于 $i = 1, 2, \dots, 10, 50, 100$ ，分别计算出前 $i \times 10$ 次实验中各个结果出现的频率。请写出完成上述任务的MATLAB命令代码。

解：

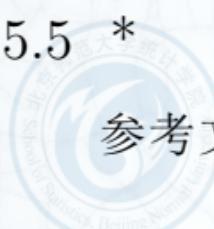
```
x=unidrnd(6,1000,1);  
y=x(1:10);  
f1=sum([y==1,y==2,y==3,y==4,...  
y==5,y==6])/10;
```

```
y=x(1:20);  
f2=sum([y==1,y==2,y==3,y==4,...  
y==5,y==6])/20  
:  
:
```



小结

1. 统计学的定义
2. 研究未知现象流程图
3. 描述未知现象的理想模型
4. 现象的理想分类
5. 关于统计学模型与追求目标
6. 统计学、数学、概率及统计软件



参考文献

- [1] 李勇, 张淑梅. 统计学导论[M]. 北京: 人民邮电出版社, 2007.
- [2] 杨振明. 概率论[M]. 北京: 科学出版社, 2008.
- [3] 唐守正, 李勇, 符利勇. 生物数学模型的统计学基础[M]. 北京: 高等教育出版社, 2015.