

◆ 课程内容

第 二 次课 (3 学时)									
<b>本次主题：</b> 第二章 监督学习与非监督学习									
<b>教学目的：</b> 1.了解监督学习和非监督学习的基本概念和区别；2.掌握监督学习的方法和应用，如线性回归、逻辑回归、决策树等；3.了解非监督学习的方法和应用，如聚类分析、主成分分析等；4.能够根据不同的问题选择学习方法。									
<b>教学重点：</b> 1.监督学习的方法和应用；2.非监督学习的方法和应用；3.学习方法的选择。									
<b>教学难点：</b> 1.监督学习和非监督学习的概念和区别；2.不同的监督学习和非监督学习方法的应用场景和优缺点									
<b>教学手段及教具：</b> 1.采用讲授和案例分析相结合的方式；2.使用幻灯片展示；									
思政要素	思政目标								
<p>监督学习与非监督学习的区分，可以类比于社会治理中的规范性和自发性。监督学习强调规则 and 标准，类似于法律法规对社会行为的指导作用；非监督学习则强调自我组织和内在规律的发现，类似于社会自发形成的秩序和道德规范。</p>	<p>通过对比监督学习和非监督学习，引导学生理解社会治理中规范性和自发性的的重要性，培养学生的法治意识和社会责任感。同时，教育学生在遵守规则的同时，也要具备创新思维和自我管理能力。</p>								
<b>讲授内容及时间分配：</b>									
<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="padding-left: 20px;">2.1 监督学习和非监督学习的基本概念</td> <td style="text-align: right;">35 分钟</td> </tr> <tr> <td style="padding-left: 20px;">2.2 监督学习的方法和应用</td> <td style="text-align: right;">35 分钟</td> </tr> <tr> <td style="padding-left: 20px;">2.3 非监督学习的方法和应用</td> <td style="text-align: right;">30 分钟</td> </tr> <tr> <td style="padding-left: 20px;">2.4 学习方法的选择</td> <td style="text-align: right;">35 分钟</td> </tr> </table>		2.1 监督学习和非监督学习的基本概念	35 分钟	2.2 监督学习的方法和应用	35 分钟	2.3 非监督学习的方法和应用	30 分钟	2.4 学习方法的选择	35 分钟
2.1 监督学习和非监督学习的基本概念	35 分钟								
2.2 监督学习的方法和应用	35 分钟								
2.3 非监督学习的方法和应用	30 分钟								
2.4 学习方法的选择	35 分钟								
<b>课后作业</b>	<ol style="list-style-type: none"> <li>1. 阅读相关论文或书籍，了解当前监督学习和非监督学习的最新研究进展</li> <li>2. 思考并总结监督学习和非监督学习的区别和联系</li> </ol>								
<b>课后总结</b>	<ol style="list-style-type: none"> <li>1. 监督学习是根据已知的输入和输出来训练模型，适用于分类和回归问题。</li> <li>2. 非监督学习是在没有明确的标签或目标的情况下，通过发现数据集的结构和模式来训练模型。</li> <li>3. 根据问题的特性和数据的类型，选择合适的学习方法和算法，是实现高效学习和准确预测的关键。</li> </ol>								

教学步骤设计			
步骤	主要任务	教师活动	学生活动
引入 (5分钟)	引入新授内容	<p>解释监督学习和非监督学习的定义和区别。</p> <p>引导学生理解监督学习的特点，即从标记的样本中学习预测模型。</p> <p>介绍非监督学习的特点，即从无标记的样本中学习数据的结构和模式。</p>	<p>理解监督学习和非监督学习的基本概念和区别。</p> <p>感受监督学习的特点，认识其在分类、回归等任务中的应用。</p> <p>理解非监督学习的特点，了解其在聚类、降维等任务中的应用。</p>
监督学习和非监督学习的基本概念 (35分钟)	介绍监督学习和非监督学习的基本概念，帮助学生理解它们的定义、特点和应用领域。	<p>解释监督学习和非监督学习的定义，明确它们在机器学习中的作用。</p> <p>引导学生思考监督学习和非监督学习的区别，例如数据标签的有无、目标任务的不同等。</p> <p>举例说明监督学习和非监督学习的应用领域，如图像分类、推荐系统等。</p>	<p>理解监督学习和非监督学习的基本概念，明确它们的定义和目标。</p> <p>比较监督学习和非监督学习的区别，思考它们在解决不同类型问题时的优势和局限性。</p> <p>探索监督学习和非监督学习的应用领域，思考它们如何应用于实际问题中。</p>
监督学习的方法和其应用 (35分钟)	介绍监督学习的常见方法和其在实际应用中的重要性。	<p>介绍监督学习的常见方法，如决策树、朴素贝叶斯、支持向量机等。</p> <p>阐述每种方法的基本原理和应用领域，如决策树在分类问题中的应用。</p> <p>提供实际案例，展示监督学习方法在解决实际问题中的成功案例。</p>	<p>理解监督学习的常见方法，掌握它们的基本原理和应用场景。</p> <p>学习各种方法的优势和限制，了解它们在不同问题上的适用性。</p> <p>分析实际案例，思考监督学习方法如何应用于解决实际问题。</p>

<p>非监督学习的方法和应用 (30分钟)</p>	<p>介绍非监督学习的常见方法和其在实际应用中的重要性</p>	<p>介绍非监督学习的常见方法,如聚类、关联规则挖掘、降维等。          阐述每种方法的基本原理和应用领域,如聚类在市场细分中的应用。          提供实际案例,展示非监督学习方法在解决实际问题中的成功案例。</p>	<p>学习非监督学习的常见方法,理解它们的基本原理和应用场景。          掌握聚类、关联规则挖掘、降维等方法的特点和优势。           分析实际案例,思考非监督学习方法如何应用于解决实际问题。</p>
<p>学习方法的选择 (35分钟)</p>	<p>探讨如何选择适合的学习方法,考虑数据特点、问题要求和模型性能等因素。</p>	<p>介绍学习方法选择的重要性和挑战,引导学生思考选择的因素。          解释数据特点对学习方法选择的影响,如高维数据、不平衡数据等。          引导学生讨论问题要求和模型性能对学习方法选择的影响。</p>	<p>理解学习方法选择的重要性,认识不同因素对选择的影响。          分析数据特点,思考如何选择适合的学习方法来处理不同类型的数据。          讨论问题要求和模型性能,考虑如何根据需求选择最合适的学习方法。</p>
<p>总结和讨论 (10分钟)</p>	<p>对本章内容进行总结,并与学生一起讨论学习的收获和潜在问题。</p>	<p>总结本章内容的关键要点,强调监督学习和非监督学习的区别和应用。          鼓励学生分享他们对监督学习和非监督学习的理解和思考。          解答学生可能存在的疑惑和困惑,澄清相关概念或问题。</p>	<p>回顾本章学习的内容,概括监督学习和非监督学习的基本概念和应用。          分享个人对监督学习和非监督学习的理解和思考,与同学们进行交流和讨论。          提出问题或困惑,寻求教师的解答和澄清。</p>

## 1. 监督学习和非监督学习的基本概念和区别

本小节主要介绍监督学习和非监督学习的概念和区别。监督学习是指给定一些有标签的数据作为训练集，通过学习训练集中输入和输出之间的关系来构建一个预测模型。常见的监督学习算法包括回归算法和分类算法。非监督学习是指在没有标签的数据上寻找数据内在的结构和规律，通过对数据进行聚类、降维等处理来挖掘数据中的隐藏信息。

监督学习和非监督学习的主要区别在于是否有标签数据。监督学习需要有标签的数据来指导模型的训练，而非监督学习则是在没有标签数据的情况下进行学习。

本小节的教学重点是让学生理解监督学习和非监督学习的基本概念和区别，帮助学生明确两者的学习目标和方法，以便后续的学习。

## 2. 监督学习的方法和应用

监督学习是指从标记样本数据中学习预测模型，进而对未知数据进行预测的过程。在监督学习中，我们需要给出输入数据和对应的输出数据，让算法能够通过这些数据来学习预测模型。监督学习的方法和应用非常广泛，包括线性回归、逻辑回归、决策树、朴素贝叶斯、支持向量机等。

**线性回归：**线性回归是一种广泛应用于机器学习中的监督学习方法，用于预测连续的输出变量。它建立在输入变量和输出变量之间的线性关系上，通过拟合一个或多个直线来解决问题。线性回归在统计学中已经有很长的历史，是最简单的回归分析之一。线性回归有很多变种，其中最常用的是普通最小二乘法（OLS）。OLS通过最小化平方误差来确定最优拟合线，即在所有可能的直线中，找到使误差最小的直线。

**逻辑回归：**逻辑回归是一种广泛应用于机器学习中的监督学习方法，用于预测二元输出变量。它通过将线性回归的输出映射到概率分布上，从而得到二元输出的概率预测结果。逻辑回归也是一种广义线性模型，逻辑回归模型通过最大似然估计来求解模型参数，从而得到最优的预测结果。

**决策树：**决策树是一种广泛应用于机器学习中的监督学习方法，它通过将输入数据分成不同的类别来预测输出变量。决策树模型由节点和边组成，节点表示属性，边表示属性值，模型的预测过程就是通过决策树从根节点一步步向下推导，最终到达一个叶子节点，叶子节点即为输出结果。

备注：  
让学生意识到数据有无标签的差异

介绍线性回归、逻辑回归、决策树等算法，并举例说明

决策树的优点在于易于理解、可解释性强、处理非线性数据能力较强等，但它也容易过拟合，需要通过剪枝等方法来优化模型。

除了上述三种常见的模型，还有许多其他的方法和模型，如支持向量机、神经网络等。这些方法在不同的问题和应用场景中表现出色，可以帮助解决许多现实世界中的难题。

支持向量机是一种二分类模型，可以处理高维数据，并且具有强大的泛化能力。其基本思想是在数据空间中找到一个最优的超平面，将数据分为两个类别，从而实现分类。支持向量机在文本分类、图像识别、生物信息学等领域得到广泛应用。

神经网络是一种基于人工神经元的计算模型，具有强大的拟合能力和泛化能力。在监督学习中，常用的神经网络包括多层感知机、卷积神经网络和循环神经网络等。这些神经网络在语音识别、图像分类、自然语言处理等领域都取得了很好的成果。

监督学习的应用非常广泛，例如在金融领域中，用于信用评估和预测股票价格；在医学领域中，用于疾病预测和药物发现等。

### 3. 非监督学习的方法和应用

在机器学习领域，非监督学习是指从数据本身中学习结构和模式的一种方法，与监督学习相比，非监督学习没有目标变量或标签，而是寻找数据中的内在结构和模式。非监督学习的方法和应用非常广泛，其中聚类和主成分分析是最常见的两种方法。

聚类是指将数据点分成相似的组或簇的过程，这些组内部的数据点具有较高的相似度或距离度量。聚类算法可以应用于许多领域，例如市场营销、生物学、图像处理等，用于发现数据的相似性或者将数据划分为不同的类别。聚类方法可以分为层次聚类和划分聚类两种方法。层次聚类是通过不断合并或分裂簇来构建层次结构，直到所有数据点都被划分为单个簇。划分聚类是通过指定簇数来将数据划分为固定数量的簇。

主成分分析是一种常用的降维技术，它可以将高维数据映射到低维空间中，同时尽可能地保留原始数据的信息。主成分分析的目标是找到数据中的主要方差，然后使用这些方差来确定数据的主成分。主成分分析可以应用于许多领域，例如图像处理、信号处理、金融和医学等领域，用于降低维度、去除冗余信息、发现数据的潜在结构等。

非监督学习的一个重要挑战是如何评估算法的效果，因为非监督学习没有目标变量或标签，所以通常使用内部评估和外部评估两种方法来评估算法的效果。内部评估方法使用数据本身的性质来评估算法的效果，例如聚类算法中使用轮廓系数和 DB 指数等指标。外

介绍聚类分析、主成分分析等算法，并简单解释使用场景

部评估方法使用先验知识或已知类别的标签来评估算法的效果，例如聚类算法中使用准确率和召回率等指标。

总之，非监督学习是机器学习领域的一个重要分支，聚类和主成分分析是其最常用的方法之一。通过对非监督学习算法的研究和应用，可以更好地理解数据的结构和模式，从而为实际问题的解决提供帮助。

#### 4. 学习方法的选择

在机器学习领域，存在大量的学习方法，不同的学习方法针对不同的问题具有不同的优缺点。因此，在实际应用中，我们需要根据问题的性质和要求，选择合适的学习方法来解决问题。

在选择学习方法时，需要根据问题的性质和要求来确定选择的基本原则。下面列举了一些基本原则：

- (1) 选择具有强大预测能力的模型。在实际应用中，我们往往关心模型的预测能力。因此，我们需要选择具有强大预测能力的模型。
- (2) 选择具有良好可解释性的模型。在一些应用场景中，需要对模型的结果进行解释，因此需要选择具有良好可解释性的模型。
- (3) 考虑计算效率。在实际应用中，需要选择计算效率高的模型，尤其是对于大规模数据集。
- (4) 考虑模型的鲁棒性。在实际应用中，需要选择具有较好鲁棒性的模型，能够处理异常数据。
- (5) 考虑模型的可扩展性。在实际应用中，需要选择具有良好可扩展性的模型，能够处理不同规模的数据集。

学习方法选择的具体步骤：

- (1) 确定问题的性质和要求。在选择学习方法前，需要先确定问题的性质和要求。例如，需要解决的是分类问题还是回归问题，数据集是否存在噪声等。
- (2) 选择候选学习方法。根据问题的性质和要求，可以选择一些候选学习方法。例如，对于分类问题，可以选择支持向量机、决策树等方法；对于回归问题，可以选择线性回归、岭回归等方法。
- (3) 评估候选学习方法。对于选择的候选学习方法，需要进行评估。评估方法包括交叉验证、留出法等。
- (4) 选择最终学习方法。根据评估结果，可以选择最终的学习方法。需要注意的是，在实际应用中，最终选择的学习方法不一定是最优的，需要综合考虑各种因素来确定最终的学习方法。

帮助学生了解如何选择适合自己的学习方法，如实践、理论学习等

## 教学案例—房价预测

### 案例描述：

在这个案例中，我们将使用监督学习方法来预测房屋的售价。我们将基于一些特征，如房间数量、卧室数量、面积等，来建立一个回归模型，以预测房屋的售价。

### 要求：

- (1) 数据集：我们将使用一个包含房屋特征和对应售价的数据集。数据集应包括多个样本，每个样本包含特征和对应的售价。
- (2) 特征工程：在建立回归模型之前，需要对特征进行预处理和工程化。这可能包括特征选择、缺失值处理、数据标准化等步骤。
- (3) 模型训练：选择合适的回归模型，并使用数据集对模型进行训练。
- (4) 模型评估：使用评估指标（如均方误差）对模型进行评估，了解其预测性能。
- (5) 预测应用：使用训练好的模型进行房价预测，并展示预测结果。

### 代码示例：

```
# 导入所需的库
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# 读取数据集
data = pd.read_csv('house_data.csv')

# 数据预处理和特征工程
# ...

# 划分训练集和测试集
X = data.drop('Price', axis=1)
y = data['Price']
X_train, X_test, y_train, y_test = train_test_split(X, y,
                                                    test_size=0.2, random_state=42)

# 建立回归模型
model = LinearRegression()
model.fit(X_train, y_train)
```

```
# 模型评估
y_pred = model.predict(X_test)
mse = mean_squared_error(y_test, y_pred)
print('均方误差: ', mse)

# 进行房价预测
new_data = pd.DataFrame({'Rooms': [3], 'Bedrooms': [2], 'Area':
[150]})
predicted_price = model.predict(new_data)
print('预测的房价: ', predicted_price)
```

说明:

- (1) 需要根据实际情况准备一个包含特征和对应售价的数据集，并将其命名为 'house\_data.csv'，并根据数据集的特点进行数据预处理和特征工程的步骤。
- (2) 在建立回归模型时，我们使用了线性回归模型，你可以根据实际情况选择其他合适的回归模型。
- (3) 模型评估阶段，我们使用了均方误差 (Mean Squared Error) 作为评估指标，你也可以选择其他适当的指标进行评估。

## 第 1 讲 第 1、2 节课(2 学时)

**引入:** 一句话简单介绍自己, 提出课程要求, 考勤与作业情况, 平时分组成, 参考书目《统计学(第七版)》学习指导书, 贾俊平, 中国人民大学出版社。

简单介绍本门课程: 统计学是一门既有趣又十分有用的学科, 但计算机时代到来前, 统计的应用受到限制。计算机和互联网普及的今天, 尤其是统计软件的使用, 既促进了统计科学的发展, 也使得统计教学和学习发生了革命性变化。如 SAS, SPSS, 但价格不菲。R 是基于 R 语言的一种优秀的免费统计软件, 功能十分强大, 更新速度快, 使用灵活, 绘图功能强大。

**教学内容:** 第 1 章 导论

**教学目的:** 使学生理解统计学的含义、理解描述统计和推断统计、掌握概率抽样方法、了解统计学的应用领域、了解数据的类型、理解统计中的几个基本概念。

**育人目标:** 1.通过统计学的基本理论与研究方法的学习, 更好地利用统计学的基本方法去了解和认识我国经济建设与发展的成就。2.改革开放以来, 我国经济取得了举世瞩目的成就, 无论经济发展规模还是人民生活水平都有长足的进步, 这些说法是否有依据? 这些依据是什么? 它们从何而来? 通过统计学的学习将能很好地回答这些问题。

**重点:** 统计学相关基本概念, 统计数据类型

**难点:** 统计学数据类型

**教法说明:** 重点讲解数据与统计学的内容。

**教学手段:** 多媒体教学。

**讲授内容及时间分配:**

自我介绍与新课导言, 10 分钟

1.1 统计及其应用领域, 20 分钟

1.2 统计数据类型, 25 分钟

1.3 统计中的几个基本概念, 30 分钟

思考与练习, 15 分钟

**课后作业**      学习指导书

**课后总结**      统计学的用途?  
结合自身专业如何设计统计学学习?

教学过程	备注
<p><b>自我介绍与新课导言</b></p> <p>统计学是一门既有趣又十分有用的学科，但计算机时代到来前，统计的应用受到限制。计算机和互联网普及的今天，尤其是统计软件的使用，既促进了统计科学的发展，也使得统计教学和学习发生了革命性变化。</p> <p><b>第 1 章 导论</b></p> <p><b>1.1 统计及其应用领域</b></p> <p><b>【思政元素】</b>中华人民共和国《统计法》明确规定：统计的基本任务是对经济社会发展情况进行统计调查、统计分析，提供统计资料和统计咨询意见，实行统计监督。国家建立集中统一的统计系统，实行统一领导、分级负责的统计管理体制。国务院和地方各级人民政府、各有关部门应当加强对统计工作的组织领导，为统计工作提供必要的保障。</p> <p><b>什么是统计学？</b> (statistics)</p> <ol style="list-style-type: none"> <li>1. 统计学是收集、分析、表述和解释数据的科学(不列颠百科全书)</li> <li>2. 统计是一门收集、分析、解释和提供数据的科学(韦伯斯特国际辞典)</li> <li>3. 统计指的是一组方法，用来设计实验、获得数据，然后在这些数据的基础上组织、概括、演示、分析、解释和得出结论(Mario F.Triola,《初级统计学》)</li> </ol> <ol style="list-style-type: none"> <li>1. 收集数据：取得数据</li> <li>2. 处理数据：整理与图表展示</li> <li>3. 分析数据：利用统计方法分析数据</li> <li>4. 数据解释：结果的说明</li> <li>5. 得到结论：从数据分析中得出客观结论</li> </ol> <p><b>描述统计(descriptive statistics)</b></p> <p>研究数据收集、处理、汇总、图表描述、概括与分析等统计方法 内容：搜集数据、整理数据、展示数据、描述性分析 目的：描述数据特征、找出数据的基本规律</p> <p><b>推断统计(inferential statistics)</b></p> <p>研究如何利用样本数据来推断总体特征的统计方法 内容：参数估计、假设检验 目的：对总体特征作出推断</p> <p><b>1.2 统计数据的类型</b></p> <p><b>【思政元素】</b>2021 年中国 GDP 规模达 114.4 万亿元（17.7 万亿美元），名义增长 12.8%，稳居全球第二大经济体；中国、美国占全球 GDP 比重分别超 18% 和 23%。2021 年中国实际 GDP 增长 8.1%，两年平均增长 5.1%；同期，美国 GDP 同比 5.7%、德国 2.8%、英国 7.5%、日本 1.7%。IMF 预测 2021 年全球经济实际增速 5.9%，中国经济发展处于全球领先地位。中美经济规模差距正在快速缩小。美国 2021 年 GDP 为 23.0 万亿美元，中国 GDP 相当于美国的约 77.0%。2000-2021 年，中美经济规模差距从 9.1 万亿美元快速缩减至 5 万亿美元。中国有望在十年左右跃升为第一大经济体。如果未来几年中国经济保持年均 5.5% 左右的增长，则将在 2030 年前后超越美国成为全球第一大经济体，再回世界之巅，全球政经格局和治理版图进入新的篇章（摘自任泽平，中美经济规模快速缩小，中国有望十年之内成为第一大经济体。</p> <p>统计数据的分类——按计量尺度分</p>	<p>引入学生团队介绍</p> <p>内容讲解：思考，统计学能做什么？</p> <p>翻转课堂：学习主讲</p> <p>思考分类依据</p> <p>专业类别中，都是那些数据？如何搜集？</p>

<p>分类数据(categorical data)          顺序数据(rank data)          数值型数据(metric data)</p> <p style="text-align: center;">统计数据的分类——按收集方法分</p> <p>观测的数据(observational data)          实验的数据(experimental data)</p> <p style="text-align: center;">统计数据的分类——按时间状况分</p> <p>截面数据(cross-sectional data)          时间序列数据(time series data)          在不同时间上收集到的数据          描述现象随时间变化的情况          比如, 2000 年至 2005 年国内生产总值数据</p>	继续翻转
<p style="text-align: center;"><b>1.3 统计中的几个基本概念</b></p> <p><b>【思政元素】</b>数据显示, 2021 年, 我国国内生产总值达 1143670 亿元, 比上年增长 8.1%, 超额完成了“国内生产总值增长 6% 以上”的目标。此外, 2021 年, 我国粮食总产量为 13657 亿斤, 比上年增加 267 亿斤, 增长 2%, 粮食总产量连续 7 年保持在 1.3 万亿斤以上, 超额完成目标。数据还显示, 2021 年全国居民消费价格指数(CPI)同比上涨 0.9%, 这一终值远低于《政府工作报告》设定的“居民消费价格涨幅 3%”左右的目标。在当前全球通胀压力严峻的大背景下, 超额完成物价调控目标实属不易(选自 2022 年 3 月 5 日第十三届全国人民代表大会第五次会议上李克强总理《政府工作报告》)。</p>	案例引入: 分清总体和样本
<p style="text-align: center;"><b>总体和样本</b></p> <p><b>总体(population)</b></p> <ol style="list-style-type: none"> <li>1. 所研究的全部个体(数据)的集合, 其中的每一个个体也称为元素</li> <li>2. 分为有限总体和无限总体             <ol style="list-style-type: none"> <li>(1) 有限总体的范围能够明确确定, 且元素的数目是有限的</li> <li>(2) 无限总体所包括的元素是无限的, 不可数的</li> </ol> </li> </ol> <p>样本: 从总体中抽取的一部分元素的集合          构成样本的元素的数目称为样本容量或样本量(sample size)</p>	理论定义
<p style="text-align: center;"><b>参数和统计量</b></p> <p><b>参数(parameter)</b></p> <ol style="list-style-type: none"> <li>1. 描述总体特征的概括性数字度量, 是研究者想要了解的总体某种特征值</li> <li>2. 所关心的参数主要有总体均值(<math>\mu</math>)、标准差(<math>\sigma</math>)、总体比例(<math>\pi</math>)等</li> <li>3. 总体参数通常用希腊字母表示</li> </ol>	举例
<p><b>统计量(statistic)</b></p> <ol style="list-style-type: none"> <li>1. 用来描述样本特征的概括性数字度量, 它是根据样本数据计算出来的一些量, 是样本的函数; 关心的样本统计量有样本均值(<math>\bar{x}</math>)、样本标准差(s)、样本比例(p)等; 样本统计量通常用小写英文字母来表示</li> </ol> <p style="text-align: center;"><b>变量(variable)</b></p> <p>说明现象某种特征的概念          变量可以分为:          分类变量(categorical variable): 说明事物类别的名称          顺序变量(rank variable): 说明事物有序类别的名称          数值型变量(metric variable): 说明事物数字特征的名称</p>	专业案例

### 变量(其他分类)

随机变量和非随机变量；经验变量和理论变量

经验变量所描述的是我们周围可以观察到的事物

理论变量则是由统计学家用数学方法所构造出来的一些变量，比如， $z$  统计量、 $t$  统计量、 $\chi^2$  统计量、 $F$  统计量等

**【思政园地】统计需要实事求是——人口统计学家戴世光**



戴世光（1908.12.1-1999.8.1）湖北江夏人，1931年毕业于清华大学经济系，1936年获美国密歇根大学数理统计硕士。曾任西南联合大学、清华大学、中国人民大学教授。人口学家，更是我国 20 世纪杰出的统计学家。

戴教授 1940 年至 1992 年 50 余年一直从事教学科研工作，实事求是，勤于耕耘，成果丰硕，是 20 世纪我国杰出的统计学家，为统计学界及其从业人员实事求是地从事开展统计教学、科研、统计实务工作树立了很好的示范作用。

#### 讨论：

1.从以上示例中，您学到了什么？

2.统计学是在统计实践的基础上产生发展起来的方法论科学，其中最为关键与重要的问题是什么？

#### 本章小节

1. 统计及其应用领域
2. 统计数据类型
3. 统计中的几个基本概念

#### 总结与作业布置

总结本次课内容：本次课学习了统计学的基本概念、抽样方法。

作业：学习指导书

思政园地，  
引发讨论

总结思考，  
布置线上  
作业