

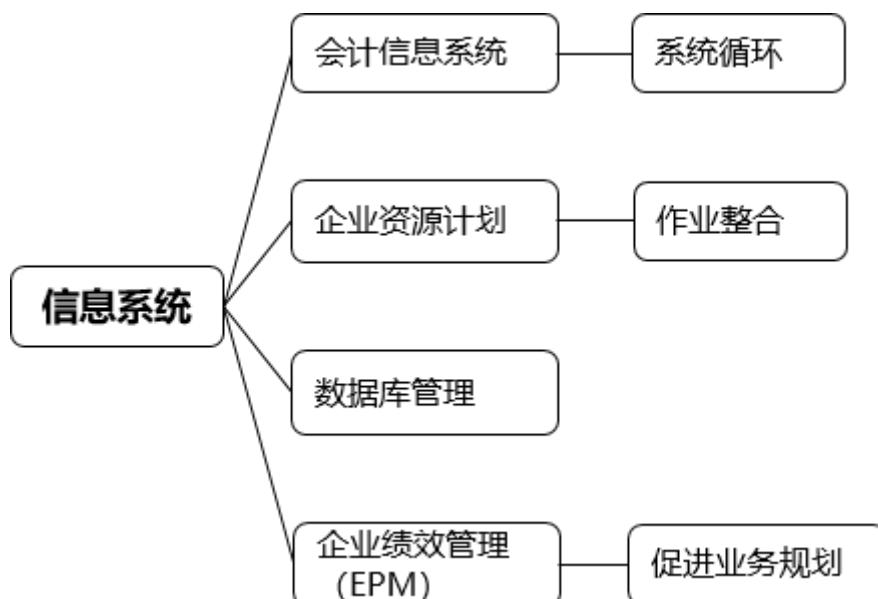
第 69 讲-信息系统 (1)

第六章 / F 科技与分析

本章考点框架

- 第 F.1 节 - 信息系统
- 第 F.2 节 - 数据管控
- 第 F.3 节 - 技术支持的财务转型
- 第 F.4 节 - 数据分析

第 F.1 节 信息系统 (1)



会计信息系统

会计信息系统 (AIS) 是基于计算机的、将会计数据转换为信息的系统。会计信息系统是利用信息技术对会计信息进行采集、存储和处理，完成会计核算任务，并能提供为进行会计管理、分析、决策用的辅助信息的系统。

会计信息系统 (AIS) 是一个收集、记录、储存和处理数据并为决策者提供信息的系统。 一个设计合理、执行有效的会计信息系统的最终目的是为决策制定者提供相关和可靠的信息。

从本质上来说，AIS 最基本的模型是**“输入-处理-输出-报告”**。

目前的会计信息系统大都属于会计电子数据处理系统 EDP (Electronic Data Processing) 和管理信息系统 MIS (Management Information System) 一类。由于技术水平的限制，会计决策支持系统 DSS (Decision Support Systems) 尚处于探索阶段。

会计信息系统可以是一个独立的系统，也可以是企业管理信息系统 MIS 的一个子系统。

会计信息系统的基本功能

- 会计数据的采集
- 会计数据的存储
- 会计数据的加工
- 会计信息的传输
- 会计信息的输出

会计信息系统在价值链中的角色

为主要和辅助活动提供了财务和非财务的信息，以此来在价值链上对业务进行决策和控制。

辅助：会计、金融、法律和行政管理，人力资源，IT 部门等。**(理解+记忆)**

会计信息系统里的数据需要以对决策制定者相对有用的形式**总结归纳，最终呈现**。

会计信息系统（AIS）的关键因素是数据的质量和可靠性，以及信息的及时性。

价值链会计管理的核心是如何实现快速和准确地反映整个价值链的会计核算与管理。包括如何产生、采集、记录、核算、反映、控制和分析价值链上的会计和财务信息，如何执行价值链上各成本中心、利润中心、物料中心等的会计管理和监控作用等。为达上述目标，价值链会计管理信息化是最重要的技术基础和赖以生存的环境。

只有在价值链企业群中实现了整体信息化后，在现代信息技术（包括：网络技术、计算机硬件技术、软件技术、分布式技术、数据库技术、数据仓库技术、数据挖掘技术和 ERP 应用系统等）的支持下，才能在整个价值链上实现信息的实时、准确地采集、记录、核算、集成、共享、跟踪、反馈，并有效地实现各中心的会计管理和监控作用，进而提高整个价值链上会计管理工作的含金量。

会计信息系统循环

业务流程与会计信息系统

- **业务流程**是指企业的采购、生产及销售等一系列活动。
- **交易循环**是由一些在某个特定的环节才出现的相关事件构成。
- **事件**是发生在某个特定时点的系列活动，是信息系统处理的基本对象。

区分管理信息系统与会计信息系统的角度：事件

对于信息系统，事件是其信息处理的基本单元，信息系统围绕着事件收集输入数据，再进行处理，最终输出有用的信息。

财务事件

- 一个能够影响到组织的资产或权益，以货币方式予以计量并反映在会计科目账户中的经济事件。

非财务事件

- 指不满足财务事件定义的另一些事件。
- 财务事件与非财务事件联系紧密，并且经常被同一信息系统进行处理。

美国的萨班斯—奥克斯利法案（SOX）要求管理层针对整个财务报告处理与生成过程设计与实施内部控制。

业务流程与会计信息系统

- **获取循环：**购买货物或服务并支付价款的过程。
- **生产循环：**将采购的资源转化成货物或服务的过程。
- **收入循环：**把货物或服务提供给客户的过程。

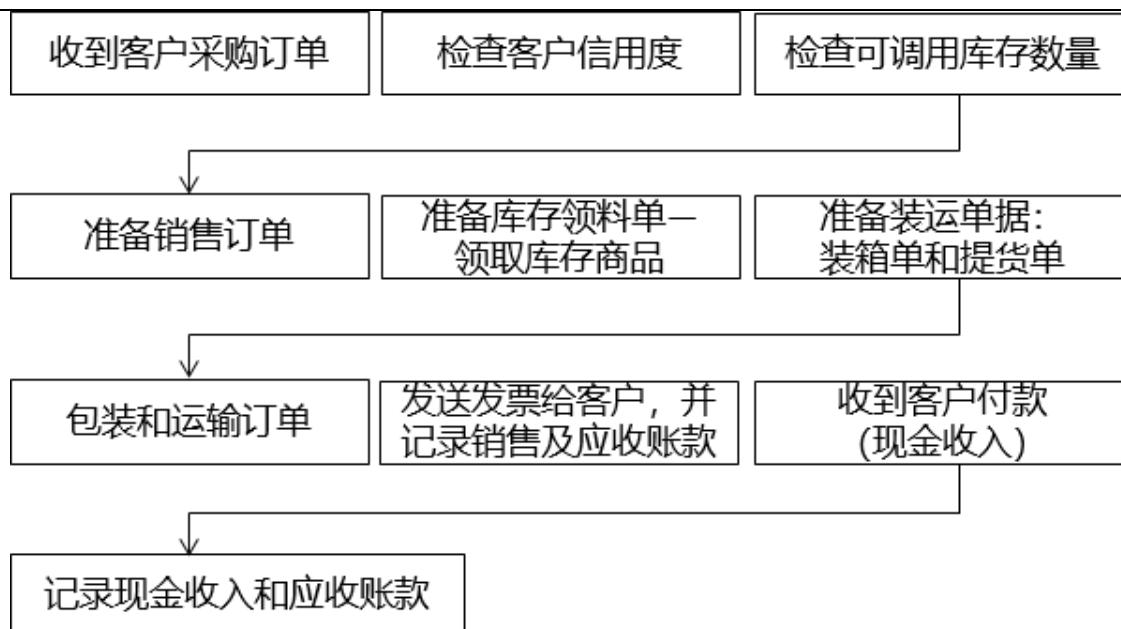
收入循环

收入循环或销售循环是致力于处理公司销售的，是与向客户提供商品和服务以及为这些销售收取现金有关的一系列业务活动和相关的信息处理操作，主要的外部信息交换是与客户进行的。

收入循环的主要目标是在正确的时间、正确的地点、以合适的价格提供正确的产品。

为了实现该目标，管理层必须做出以下六个关键决定：

1. 每个产品或服务的最佳价格是多少？
2. 是否应将信贷扩展到客户？如果是，应提供哪些信用条件？应向个人客户提供多少信用？
3. 产品在多大程度上可以并且应该根据个人客户的需求和愿望进行定制？
4. 应储存多少库存，应将库存放在何处？
5. 应如何将商品交付给客户？公司应该自行执行运输功能还是将其外包给专门从事物流的第三方？
6. 如何处理客户付款以使现金流量最大化？



支出循环

支出循环致力于为公司的经营购买商品（主要是存货），不管是简单地以批发价格购买再以零售价格转售，还是用于产品制造，它是与商品和服务的购买和付款相关的一组周期性的业务活动和相关的信息处理操作，其他支出包括日常商业经营活动，比如临时工、咨询费、云服务、差旅费等。

支出循环中，主要的外部信息交换是与供应商进行的。在组织内部，是有关购买商品和材料的需求信息从收入和生产循环、库存控制以及各个部门流入支出循环。货物和物料到达后，其收货通知会从支出循环流回到供应商处，方便备份。支出数据也从支出循环流向总账和报告程序，以包含在财务报表和各种管理报告中。

支出循环的主要目标是最大限度地减少获取和维护库存、供应品以及组织运行所需要的各种服务的总成本。

为了实现此目标，管理层必须做出以下关键决策：

1. 最佳库存和供应量是多少？
2. 组织如何合并各个部门的采购以获得最佳价格？
3. 如何使用信息技术来提高入库物流功能的效率和准确性？
4. 组织如何保持足够的现金来利用供应商提供的折扣？
5. 如何管理对供应商的付款以使现金流最大化？（延长付款浮差，缩短收款浮差）

这些问题的答案指导组织如何执行支出循环活动。

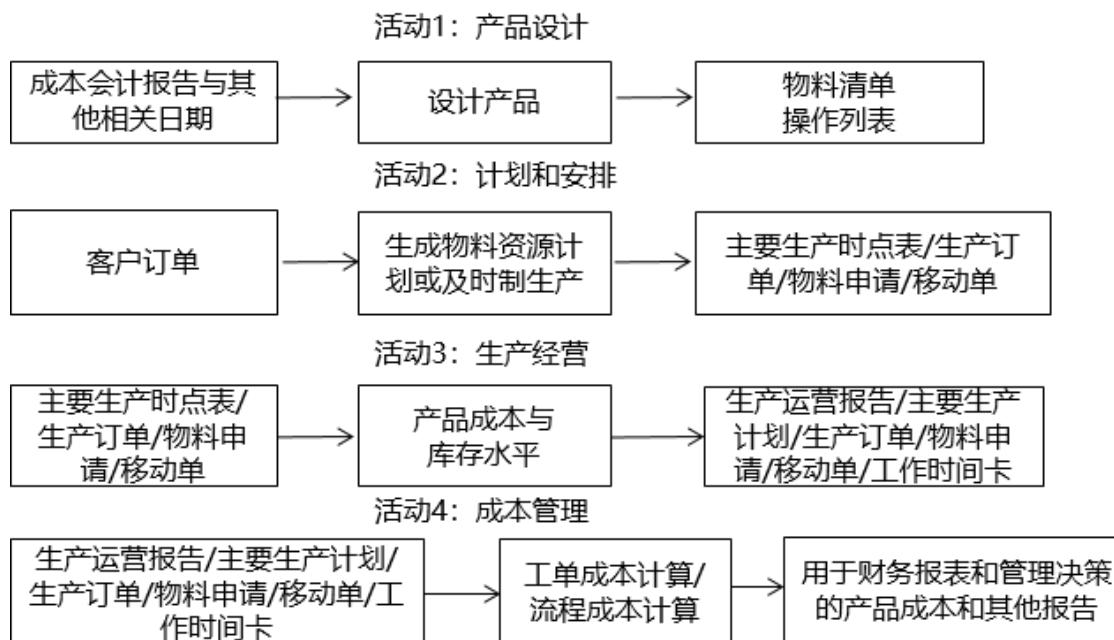


经营（生产）循环

生产循环 (Production) 是与产品制造相关的一组周期性的业务活动和相关信息的处理操作，经营 (或生产) 循环因公司和产品的不同而不同，然而，就产品循环而言，存在一种可以使用的会计数据输入、处理、输出的模式。

通常，在经营循环有四个主要的活动：

- 产品设计
- 计划和安排
- 生产经营
- 成本管理



活动 1 – 产品设计：

成本信息有助于指导产品的设计，以有效地反映不同设计的成本。输出包括物料清单（产品原料）和操作列表（产品配方）。

活动 2 – 计划和安排：

计划和安排把客户订单作为输入，以决定是否使用物料需求计划 (MRP) 或及时制生产 (JIT)。具体而言，MRP 用客户订单预测来决定何时生产、生产多少。及时制生产计划和计划产量是建立在实际的客户订单的基础上的，这就意味着库存是已被订购的，且进行及时生产以供应客户订单。

该活动的输出包括主生产计划、生产订单（授权实际开始制造产品）、材料申请从库存中提取必要的物料，并转移票据（授权将产品从一个阶段转移到下一个阶段）。

活动 3 – 生产经营：

生产经营使用来自活动 2 制造产品过程的输出文件。此项活动最主要的会计职能是查明生产成本、追踪相关的存货水平。

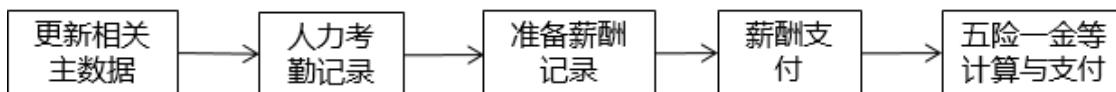
该活动的输出是经过修订和更新的输入文件（生产运营报告、主要生产计划等），以及用于计算生产过程中所用的劳动力成本的工作时间卡（为计算成本做准备）。

活动 4 – 成本管理：

成本管理是指企业生产经营过程中各项成本核算、成本分析、成本决策和成本控制等一系列科学管理行为的总称。成本管理是由成本规划、成本计算、成本控制和业绩评价四项内容组成。

人力资源管理/薪酬循环

1. 薪酬设定；
2. 构建全面的薪酬体系（岗位评价与薪酬等级、薪酬调查、薪酬计划、薪酬结构、薪酬制度的制定、薪酬制度的调整、人工成本核算）；
3. 福利和其它薪酬问题（福利保险管理、企业福利项目的设计、企业补充养老保险和补充医疗保险的设计）；
4. 评估绩效和提供反馈。

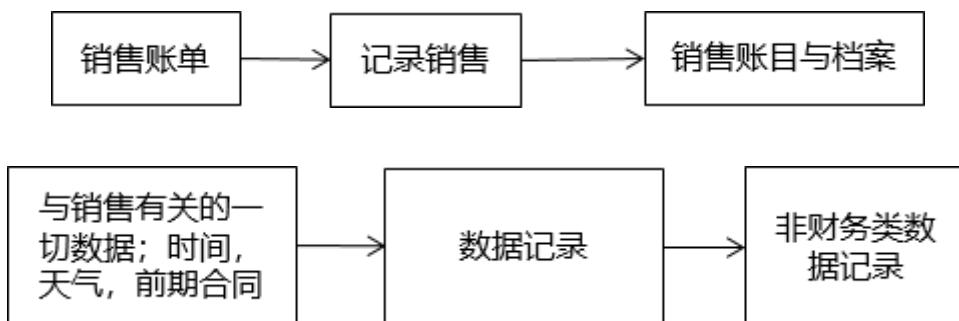
**固定资产 (PPE) 循环**

固定资产循环是指从固定资产投入使用，经过计提折旧，到固定资产报废、清理、重新购买新固定资产的业务活动过程。

总账和报告循环

总账与报告生成的循环主要包括：

1. 更新总账数据
2. 上传调整数据
3. 准备财务报告
4. 生成管理报告

财务与非财务数据循环**【例题 • 单选题】**

下列哪一项不是生产循环中的主要活动（ ）。

- A. 产品设计
- B. 计划和安排
- C. 生产经营
- D. 推广销售

【答案】D

【解析】生产循环的主要活动：产品设计计划和安排生产经营成本管理

【例题 • 单选题】

会计信息系统 (AIS) 是一个收集、记录、储存和处理数据并为决策者提供信息的系统。其在价值链中的作用是（ ）。

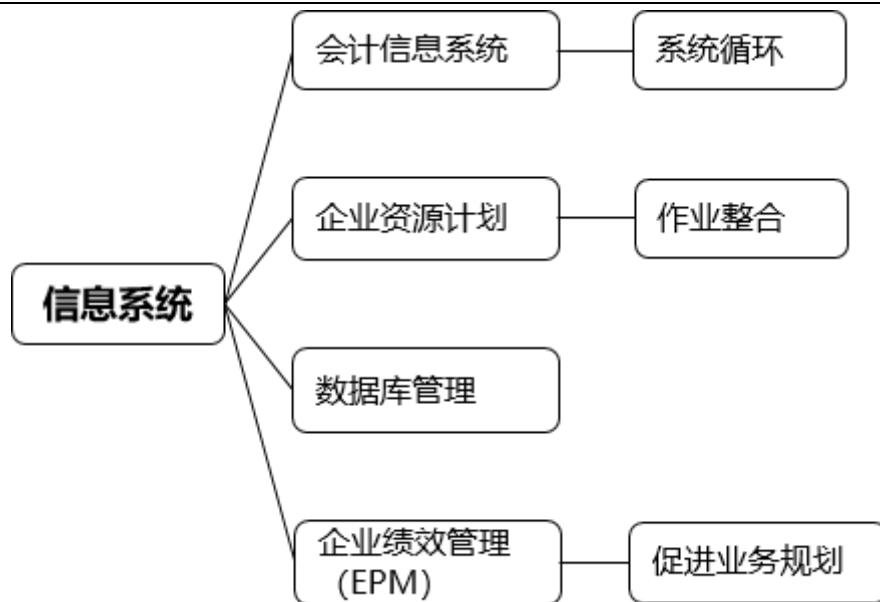
- A. 为主要和辅助活动提供财务和非财务信息，以此来在价值链上对业务进行决策和控制
- B. 为主要和辅助活动提供财务信息，以此来在价值链上对业务进行决策和控制
- C. 为主要和辅助活动提供财务和非财务信息，以此来在价值链上对业务进行控制
- D. 为主要和辅助活动提供财务信息，以此来在价值链上对业务进行控制

【答案】A

【解析】会计信息系统 (AIS) 是一个收集、记录、储存和处理数据并为决策者提供信息的系统。其在价值链中的作用是为主要和辅助活动提供财务和非财务信息，以此来在价值链上对业务进行决策和控制。

第 70 讲-信息系统 (2)

第 F.1 节 信息系统 (2)



企业资源计划

其核心思想是实现对整个供应链的有效管理，主要体现在以下三个方面：

体现对整个供应链资源进行管理的思想。

体现精益生产、同步工程和敏捷制造的思想。

体现事先计划与事中控制的思想。

ERP 系统涵盖了以下常见功能领域。在许多 ERP 系统中，这些功能领域被命名并组合在一起作为 **ERP 模块：**

生产管理：工程、材料清单 (bill of materials)、调度、产能、**工作流程管理、质量控制、成本管理、生产过程、生产工程、生产流程、生产配置、产品生命周期管理。**

进销存货管理：库存、订单输入、采购、供应商调度、货物检查、付款请求处理、佣金计算。

财务管理及会计项目：总账、现金管理、**应付账款管理、应收账款管理**、票据资金管理、固定资产管理。

成本管理：账单、时间和支出、活动管理。

人力资源管理：人力资源、薪金名册、培训管理、员工班别和出勤管理、津贴、劳健保、绩效考核。

供应链管理：和客户、供应商、员工之间的各种服务接口，采购，存货，处理索赔，仓储（收货、上架、拣货和包装）。

项目管理：项目规划，资源规划，项目花费，工作分解结构，发票，时间和费用，业绩单元，活动管理。

客户关系管理 (CRM)：销售和市场营销，佣金，服务，客户联络，呼叫中心支持。CRM 系统并不总是被认为 ERP 系统的一部分，而是业务支撑系统 (BSS) 的一部分。

数据服务：为客户、供应商及/或雇员的各种“自我服务”的界面。

企业资源计划 - 优点

1. 整合性：ERP 最大特色便是整个企业信息系统的整合，比传统单一的系统更具功能性。

2. 弹性：采用模块化的设计方式，使系统本身可因企业新增模块来支持并整合，提升企业的应变能力。

3. 集中的数据储存：将原先分散企业各角落的数据整合起来，使数据保持一致性，并提升其精确性。

4. 便利性：在整合的环境下，企业内容所产生的信息透过系统将可在企业任一地方取得与应用。

5. 提升管理绩效：ERP 系统将使部分间横向的联系有效且紧密，使得管理绩效提升。

6. 增进组织间的互动关系：透过 ERP 系统配合因特网及供应管理模式，使企业与原料供货商之间紧密结合，增加其市场变通的能力。而客户关系管理模式则使企业充分掌握市场需要取向的动脉，两者皆有助于促进企业与上下游的互动发展关系。

企业资源计划 - 缺点

1. 昂贵的投资成本：ERP 投资需要花费的软、硬件及顾问公司收取的顾问费都是庞大的开支，往往只有大型企业才有能力导入。

2. 安全性问题：为合乎电子商务的需要，整合企业功能部门在单一系统所构建出的坚实安全性，在提供网络模块后，受到强烈的挑战。

3. 不够充分的信息：系统内部的信息无法与其它企业或部门的系统整合，且也未将企业外部的情报整合其中。

4. 高潜在成本：有下列高成本的花费

- (1) 训练成本：企业内部人员的教育训练课程成本。
- (2) 数据转换成本：数据由旧 ERP 系统转置到新 ERP 系统中的成本。

【例题 · 单选题】

ERP 是将公司所有的数据存储到一个中央数据库中，是一个跨多个行业的自动化、跟踪和支持一系列管理和操作业务流程的应用程序，以下哪项不是 ERP 系统目前的局限性（ ）。

- A. 实施过程耗时、成本高昂
- B. 失败概率较高
- C. 标准化的 ERP 系统可能很难与所有现有的业务流程相适应
- D. 来自管理层的反对

【答案】D

【解析】ERP 的主要缺点有：实施过程耗时、成本高昂失败概率较高标准化的 ERP 系统可能很难与所有现有的业务流程相适应基层员工的抵触

数据库管理

数据库是“按照数据结构来组织、存储和管理数据的仓库”。是一个长期存储在计算机内的、有组织的、可共享的、统一管理的大量数据的集合。

数据库是以一定方式储存在一起、能与多个用户共享、具有尽可能小的冗余度、与应用程序彼此独立的数据集合，可视为电子化的文件柜——存储电子文件的处所，用户可以对文件中的数据进行新增、查询、更新、删除等操作。

数据库管理系统（Database Management System, DBMS）是一种操纵和管理数据库的大型软件，是用于建立、使用和维护数据库。它对数据库进行统一的管理和控制，以保证数据库的安全性和完整性。

用户通过 DBMS 访问数据库中的数据，数据库管理员也通过 DBMS 进行数据库的维护工作。它提供多种功能，可使多个应用程序和用户用不同的方法在同时或不同时刻去建立，修改和询问数据库。它使用户能方便地定义和操纵数据，维护数据的安全性和完整性，以及进行多用户下的并发控制和恢复数据库。

数据库管理系统的层次结构

- (1) 应用层
- (2) 语言翻译处理层
- (3) 数据存取层
- (4) 数据存储层
- (5) 操作系统

企业绩效管理（EPM）

上世纪 90 年代初，术语“业务绩效管理”开始出现，并与平衡计分卡方法密切相关。IT 行业在 2003 年左右才接受这个概念，这最终演变成所谓的“企业绩效管理”（EPM），根据调查机构 Gartner 公司提出的“以提高企业绩效的目标监控整个业务绩效的过程”的描述，该术语通常与公司绩效管理（CPM），业务绩效管理（BPM）和财务绩效管理（FPM）是同义词。

一个 EPM 套件或平台通常包括预置的应用程序和工具，以解决在绩效管理循环中的关键步骤：

- 建模：定义高层次的目的和商业模式。
- 规划：制定年度财务和经营计划、预算、定期预测。
- 财务关闭和整合：月度、季度和年度关闭决算。
- 报告：向内部、外部利益相关者提交报告，进行审查。
- 分析：与原来的预算或计划比较实际结果，与前期、跨部门或产品线等结果进行比较，获得一个基准各不相同的，理解“为什么”的结果，并决定行动方针，纠正偏差。

企业应用 EPM 主要的收益表现在业务绩效的提升、运营成本的下降、客户关系的改善等方面。同时还有：

- 技术相关的节约
- 生产力收益，由于时间耗费少使得效益提高
- 企业流程加强，在分析应用软件的支持下，企业流程进一步转变，由此创造了收益
- 改善决策：EPM（BPM）使企业可以直接根据业务变化制订决策
- 改进企业绩效：引进企业分析可对绩效产生积极的影响——提高边际利润，减少成本，可以达到更好的存货管理

- 改进业务可见度：改进业务可见度可以使企业将更多的注意力投入到业务本身
- 持续改进：EPM (BPM) 可持续改进业务流程

【例题 • 单选题】

企业应用 EPM 主要的收益表现在业务绩效的提升、运营成本的下降、客户关系的改善等方面。同时还有以下优。点，除了（ ）。

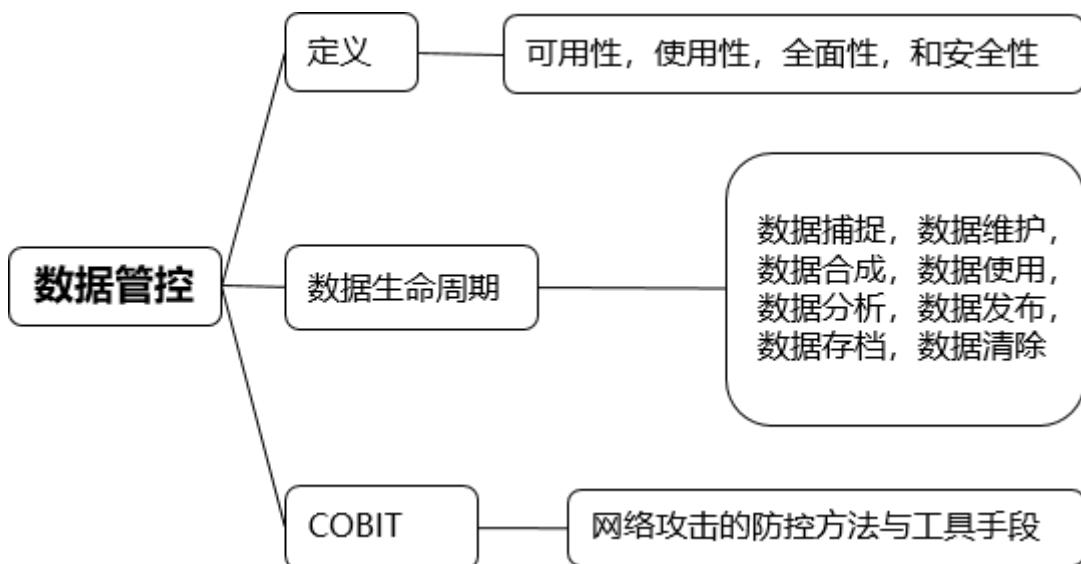
- 激励员工，发挥个体效能
- 技术相关的节约
- 生产力收益，由于时间耗费少使得效益提高
- 企业流程加强，在分析应用软件的支持下，企业流程进一步转变，由此创造了收益

【答案】A

【解析】企业应用 EPM 主要的收益表现在业务绩效的提升、运营成本的下降、客户关系的改善等方面。同时还有：技术相关的节约生产力收益，由于时间耗费少使得效益提高企业流程加强，在分析应用软件的支持下，企业流程进一步转变，由此创造了收益

第 71 讲-数据管控

第 2 节 数据管控



数据管控定义

数据管控

数据管控，又称数据治理（Data Governance），是组织中涉及数据使用的一整套管理行为。由企业数据治理部门发起并推行，关于如何制定和实施针对整个企业内部数据的商业应用和技术管理的一系列政策和流程。

国际数据管理协会（DAMA）给出的定义：数据治理是对数据资产管理行使权力和控制的活动集合。

国际数据治理研究所（DGI）给出的定义：数据治理是一个通过一系列信息相关的过程来实现决策权和职责分工的系统，这些过程按照达成共识的模型来执行，该模型描述了谁（Who）能根据什么信息，在什么时间（When）和情境（Where）下，用什么方法（How），采取什么行动（What）。

数据治理的最终目标是提升数据的价值，数据治理非常必要，是企业实现数字战略的基础，它是一个管理体系，包括组织、制度、流程、工具。

数据生命周期

生命周期由多个阶段组成，每个阶段都有自己的特征。虽然对数据生命周期所包含的阶段的实际数量存在争议，但是这八个阶段代表了关于数据生命周期的一般观点：数据捕捉、数据维护、数据合成、数据使用、数据分析、数据发布、数据存档、数据清除。

(一) 数据捕捉

为了进行分析，首先必须记录或捕捉数据，这是创建组织内尚不存在的信息的行为。被捕捉的数据可能来自公司的外部或内部。今天的物联网产生了大量可能被捕捉到的数据。

(二) 数据维护

捕捉数据后，生命周期的第二阶段可以定义为向发生数据合成和数据使用的点提供数据。

(三) 数据合成

数据合成涉及统计方法的使用，这些方法结合了许多数据来源或测试，以获得更好的总体估计或被问到的有关数据问题的答案。一些人把这个称为数据建模或使用归纳推理以转换数据。另一些人将数据合成视为数据维护的子集。

(四) 数据使用

数据使用就是如何使用数据来支持业务的任务，此阶段可以定义为将数据应用到企业需要自行运行和管理的任务。如战略规划、客户关系管理（CRM）、处理发票、向供应商发送订购单等。

(五) 数据分析

数据分析是以创建新的信息和产生业务洞察力为目的检查原始数据的科学性。它包含技能、科学、迭代探索和对过去业务绩效的调查，以获得洞察力并推动未来的业务规划。在它最基本的层次，它意味着用数据分析的方法来回答问题。一些人将数据分析视为数据使用的子集。

(六) 数据发布

数据发布，指向组织外部发送数据的，所发布的数据供外部利益相关者广泛使用，通常将数据发送到业务部门。例如向客户发送结算单。

(七) 数据存档

单个数据值可能会使用相当长的一段时间，但其生命周期最终都会到尽头。那时，组织应该将数据值存档。数据存档是从有效使用的数据中删除数据，并将其存储以备潜在的未来使用的过程，数据存档是存储数据的位置，但是它没有定期维护，且使用率可能很小。

(八) 数据清除

数据清除涉及删除不再有用或不需要的数据，在数据生命周期结束时，将从企业中删除数据项的每个副本，这通常是从归档中完成的。管理者应创建数据保留政策以实施正确的数据清除实践。

COBIT 信息及相关技术控制目标

COBIT (Control Objectives for Information and related Technology) 是目前国际上通用的信息系统审计的标准，由信息系统审计与控制协会在 1996 年公布。这是一个在国际上公认的、权威的安全与信息技术管理和控制的标准。COBIT 有过五个主要版本，目前通用的版本也叫 COBIT 5。

COBIT 5 提供一种全面的框架，以支持企业实现其企业 IT 治理和管理的目标。COBIT 将 IT 过程、IT 资源与企业的策略与目标（准则）联系起来，形成一个三维的体系结构。

(1) IT 准则维。集中反映了企业的战略目标，主要从质量、成本、时间、资源利用率、系统效率、保密性、完整性、可用性等方面来保证信息的安全性、可靠性、有效性；

(2) IT 资源维。主要包括以人、应用系统、技术、设施及数据在内的信息相关的资源，这是 IT 治理过程的主要对象；

(3) IT 过程维。则是在 IT 准则的指导下，对信息及相关资源进行规划与处理，从信息技术 I 规划与组织、采集与实施、交付与支持、监控等四个方面确定了 34 个信息技术处理过程，每个处理过程还包括更加详细的控制目标和审计方针对 IT 处理过程进行评估。

技术资源

数据 (Data) ——指最广义（例如，表面的和内在的）的对象，包括结构化和非结构化、图表、声音等。

应用系统 (Application Systems) ——人工程序和电脑程序的总和。

技术 (Technology) ——包括硬件、操作系统、数据库管理系统、网络、多媒体等。

设备 (Facilities) ——用来存放和支持信息系统的一切资源。

人员 (People) ——包括用来计划、组织、获取、传送、支持和监控信息系统和服务所需要的人员技能、意识和生产力。

COBIT 信息技术的控制目标

(1) 有效性 (Effectiveness) ——是指信息与商业过程相关，并以及时、准确、一致和可行的方式传送。

(2) 高效性 (Efficiency) ——关于如何最佳（最高产和最经济）利用资源来提供信息。

- (3) 机密性 (Confidentiality) ——涉及对敏感信息的保护，以防止未经授权的披露。
- (4) 完整性 (Integrity) ——涉及信息的精确性和完整性，以及与商业评价和期望相一致。
- (5) 可用性 (Availability) ——指在商业处理需求中，信息是可用的。还指对必要的资源和相关性能的维护。
- (6) 符合性 (Compliance) ——遵守商业运作过程中必须遵守的法律、法规和契约条款，如外部强制商业标准。
- (7) 信息可靠性 (Reliability of Information) ——为管理者的日常经营管理以及履行财务报告责任提供适当的信息。

数据安全网络攻击检测和预防

网络攻击检测和预防

网络攻击 (Cyber Attacks, 也称赛博攻击) 是指针对计算机信息系统、基础设施、计算机网络或个人计算机设备的，任何类型的进攻动作。对于计算机和计算机网络来说，破坏、揭露、修改、使软件或服务失去功能、在没有得到授权的情况下偷取或访问任何一计算机的数据，都会被视为计算机和计算机网络中的攻击。

检测和预防 方式主要有：

渗透测试，是指一项由内部审计团队或外部安全咨询企业执行的授权活动，该活动试图侵入组织的信息系统。渗透测试旨在确定系统的哪些部分最需要增加额外的保护，这些保护措施可以大幅增加侵入系统所需要的时间和精力。

生物识别控制，是指一种根据行为特征来自动验证个人身份的方法。

防火墙，是指一种特殊硬件设备或在通用电脑上运行的软件，用于控制在防火墙内部的系统和其他网络之间的内部和外部通讯。

【例题 • 单选题】

虽然对数据生命周期所包含的阶段的实际数量存在争议，但是通常有八个阶段，下列各阶段的顺序正确的是（ ）。

- A. 数据捕捉、数据合成、数据维护、数据使用
- B. 数据捕捉、数据存档、数据维护、数据合成
- C. 数据合成、数据使用、数据分析、数据发布
- D. 数据捕捉、数据合成、数据使用、数据维护

【答案】C

【解析】这八个阶段代表了关于数据生命周期的一般观点：数据捕捉、数据维护、数据合成、数据使用、数据分析、数据发布、数据存档、数据清除。

【例题 • 单选题】

以下哪一项是 COBIT 的最佳定义（ ）。

- A. 协助对 IT 流程实施适当控制的一套指南
- B. 由发起组织委员会发布的控制框架
- C. 旨在帮助股东理解管理层使用的 IT 系统的一套指南
- D. 一系列技术挑战的风险和对策

【答案】A

【解析】COBIT (Control Objectives for Information and related Technology) 是目前国际上通用的信息系统审计的标准，由信息系统审计与控制协会在 1996 年公布。

【例题 • 单选题】

某系统程序员被解雇同时其访问密码被禁用以后，该公司的服务器崩溃并丢失了所有的文件和软件。经调查，被解雇的员工在操作系统中安装了病毒并通过“后门”远程操作了此次事件。以下哪一项措施最有可能阻止此类事件的发生（ ）。

- A. 制定规章，要求对服务器上所有数据和文件进行加密
- B. 每日备份服务器上的所有数据
- C. 应用监控网络流量日志的入侵检测系统
- D. 启用探测入侵的高级防火墙

【答案】D

【解析】防火墙，是指一种特殊硬件设备或在通用电脑上运行的软件，用于控制在防火墙内部的系统和其他网络之间的内部和外部通讯。

【例题 • 单选题】

BAC 公司，由于今年连续发生企业内部文件失窃事件，决定使用生物识别控制来加强企业内部控制。下列哪项不是企业通常使用的生物识别技术（ ）。

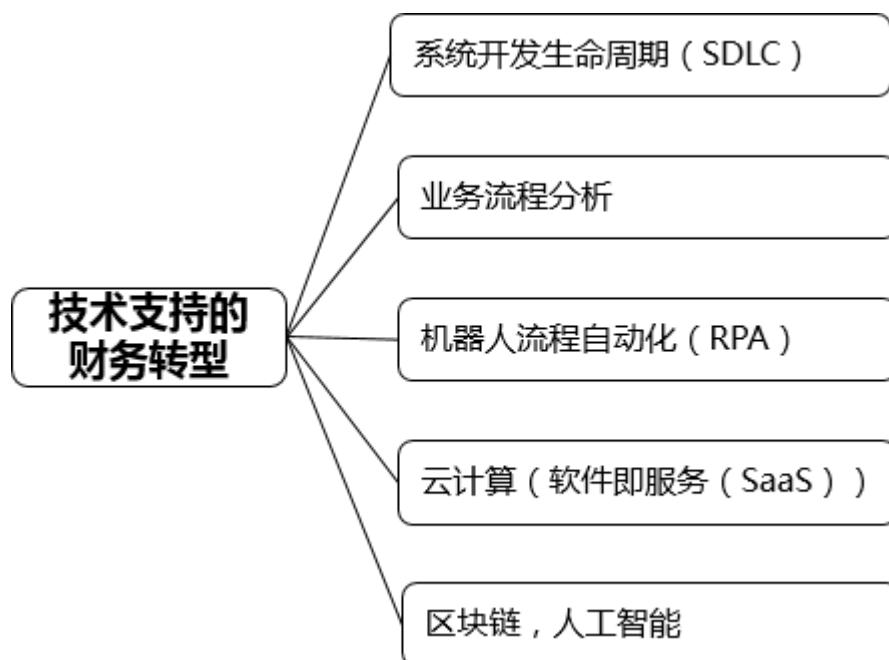
- A. 岗位绩效数据
- B. 身份证
- C. 指纹
- D. 行为特征

【答案】A

【解析】所谓生物识别技术就是通过计算机与光学、声学、生物传感器和生物统计学原理等高科技手段密切结合，利用人体固有的生理特性，（如指纹、指静脉、人脸、虹膜等）和行为特征（如笔迹、声音、步态等）来进行个人身份的鉴定。主要包括，你知道的东西，如密码；你拥有的东西，如身份证；你唯一的东西（生物识别），如你的指纹。

第 72 讲-技术支持的财务转型

第 3 节 技术支持的财务转型



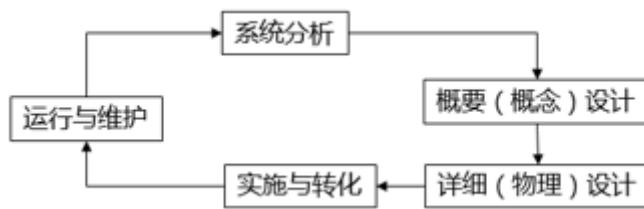
系统开发生命周期

系统开发生命周期 (SDLC, 也称软件生命周期) 是软件从产生直到报废的生命周期，周期内有“问题定义、可行性分析、总体描述、系统设计、编码、调试和测试、验收与运行、维护升级到废弃”等阶段。

这种按时间划分的思想方法是软件工程中的一种思想原则，即按部就班、逐步推进，每个阶段都要有定义、工作、审查、形成文档以供交流或备查，以提高软件的质量。

一个软件从定义、开发、运行维护、直到报废要经历一个时期的过程，这个时期称为软件的生命周期。

系统软件生命周期一般分为分析、设计、实现和测试与维护这几个阶段。



系统分析

软件开发首先需要进行需求调研和分析，完善的需求信息是软件开发项目成败的重要因素，提炼所收集的用户需求，建立完整的分析模型，把他编写成软件开发过程中的需求，规格说明和初步的用户手册。

设计阶段

软件设计可以分为两个阶段：**概要(概念)设计** 和 **详细(物理)设计**，实际上软件设计的主要任务就是把软件分解成模块机制，模块机制是指实现某个功能的数据和程序的说明。

概要设计就是结构设计，其主要目标就是给出软件模块结构，用软件结构图表示，详细设计的首要任务是设计模块的程序流程，算法和数据结构，设计人员依据软件需求规格说明文档，确定软件的体系结构，进而确定每个模块实现算法，数据结构和接口等，编写设计说明书，组织设计评审。

实施与转化

软件编码是指把软件设计转换成计算机可以接受的程序，即写成以某一段程序设计语言表示的“源程序清单”。充分了解软件开发语言、工具的特性和编程风格，有助于开发工具的选择，保证开发产品的开发质量。

在设计测试的基础上，测试软件的各个组成模块，然后再把各个模块集成起来，测试整个产品的功能和性能是否能够满足已有的规格说明。

运行与维护

维护是指已经完成对软件的研制工作并交付使用后，对软件产品所产生的错误进行改正，适应环境变化和增强功能等软件工程修订，做好软件维护工作，不仅能排除障碍，使软件能正常工作，而且还可以扩展软件功能，提高性能，为用户带来明显的经济效益。

业务流程分析

我们在之前章节已经详细了解过什么是业务流程分析。

在这里，我们需要关注的是，“业务流程分析”如何协助**提高企业系统绩效**的四个步骤：

1. 明确系统目标，相关责任人，明确项目起始点，以及目前定位。
2. 进行系统的穿行测试，并全面清晰的记录穿行结果。
3. 对现有流程进行评测，定义现有流程中的强项，以及需要提高的方面（如流程瓶颈，流程摩擦点，流程薄弱环节等）。
4. 基于上诉分析，起草改进方案。

机器人流程自动化

机器人流程自动化（Robotic Process Automation，简称 RPA）是一种根据预先设定的程序，通过模拟并增强人类与计算机的交互过程，执行基于一定规则的大批量、可重复性任务，实现工作流程自动化的软件或平台。

RPA 机器人能够模仿几乎所有人类用户的行为。它可以登录应用程序，移动文件和文件夹，复制和粘贴数据，填写表单，从文档中提取结构化和半结构化数据，抓取浏览器等。

快速实施，快速实现投资回报率

来自欧洲的人力资源服务提供商每月处理 2 500 份病假证明，每件物品的平均处理时间为 4 分钟。在三周内，他们实施了 RPA 解决方案，实现了 90% 的流程自动化。RPA 机器人从 SAP 中的事务中提取数据，将信息插入客户的系统并进行打印。人力资源服务提供商在六个月内实现了投资回报，错误率降至 0，手动工作量降至 5%，处理时间缩短 80%。

减少后台的工作量

一家全球零售商正在使用其商店盘点报告来验证数百家商店中每个商店的盘点信息。该商店的员工之前使用手动过程来提取这些报告，这一过程繁琐且不高效。通过自动化流程，商店使员工现在能够更多专注于以客户为中心的活动。RPA 机器人现在将结束报告移动到一个服务器，然后读取并合并商店盘点报告所需的信息。

改善前台的客户服务

一家在全球拥有 50 000 多家客户的贸易信用保险公司实现了信用额度请求承保流程。承销商之前曾手动收集，从内部（风险与政策）到外部（客户网站，谷歌新闻）来源的信息。通过 RPA，他们每月节省了 2 440 小时的人力工作。员工现在利用这段时间直接与客户合作。

工业生产的自动化

一家在中国的汽车发动机制造商，发动机的电子控制器在生产过程中需要通过软件写入初始内容。写入操作通过生产工人操作软件完成。6 台写入设备需要相应的工人值守并手动操作完成。通过部署 RPA+摄像头自动识别系统，能够实现全自动的机器识别和录入。不需要工人值守。每月节省了近 1 000 小时的人力。大大降低了生产成本。

【例题 • 单选题】

莱尔斯集团，目前为了应对外部发展与全球化 2.0 的趋势，准备全面开始进行 RPA 机器人流程自动化，管理层目前需要一份 RPA 方面的报告。报告中对于 RPA 机器人流程自动化进行了相关分析，其中理解不正确的是()。

- A. RPA 可以快速实施，快速实现投资回报率
- B. RPA 机器人流程自动化可以减少后台的工作量
- C. RPA 将会加强学习曲线效应，使企业有更多收益
- D. RPA 可以重视客户，改善前台的客户服务

【答案】C

【解析】当实施 RPA 机器人流程自动化时，企业成本中直接人工占比将下降，学习曲线效应不再明显。

【例题 • 单选题】

下列哪一项，不是人工智能在财务方面技术的直接应用 ()。

- A. 光学字符识别系统
- B. SOP 标准化流程手册
- C. 机器人流程自动化
- D. 电子财务记账系统

【答案】B

【解析】SOP 标准化流程手册，是在系统电子化之前，传统财务部就会执行的流程规范方法。

云计算

云计算，是指一种能够在任何地方，以按需的方式便捷地接入配置性计算资源（例如网络、服务器、存储、应用和服务）共享池的模式。云计算能够以最少的管理投入或以服务供应商互动的方式快速地提供和释放这些资源。

- (1) 按需自助式服务
- (2) 广泛的网络接入
- (3) 资源汇聚与共享
- (4) 极速弹性
- (5) 可度量的服务

三种基于云计算的服务类型：

(1) 基础设施即服务 (IaaS)，是指一种信息系统的策略，在策略下，企业外包其用于数据处理活动的设备，包括服务器、存储设备和网络组件。

(2) 平台即服务 (PaaS)，是指一种为用户提供计算平台的服务方式，包括操作系统、编程语言的执行环境、数据库服务和网页服务器。

(3) 软件即服务 (SaaS)，是指一种让业务活动能够订阅网页传递应用软件的服务。

软件即服务

SaaS 即 **Software-as-a-Service (软件即服务)** 是随着互联网技术的发展和应用软件的成熟，在 21 世纪开始兴起的一种完全创新的软件应用模式。传统模式下，厂商通过 License 将软件产品部署到企业内部多个客户终端实现交付。SaaS 定义了一种新的交付方式，也使得软件进一步回归服务本质。

企业部署信息化软件的本质是为了自身的运营管理服务，软件的表象是一种业务流程的信息化，本质还是第一种服务模式，SaaS 改变了传统软件服务的提供方式，减少本地部署所需的大量前期投入，进一步突出信息化软件的服务属性，或成为未来信息化软件市场的主流交付模式。

特性：

- (1) 互联网特性；

- (2) 多重租赁 (Multi-tenancy) 特性;
- (3) 服务 (Service) 特性;
- (4) 可扩展 (Scalable) 特性;

【例题 · 单选题】

软件即服务的最佳定义为（ ）。

- A. 本地安装和维护的预置 ERP 系统
- B. 云计算的一种，其应用程序由第三方托管并按月收费
- C. 数据库管理系统的一种，由第三方提供服务并收取年费
- D. 基础架构平台，取代企业内复杂的个人计算机

【答案】B

【解析】 云计算的一种，其应用程序由第三方托管并按月收费。三种基于云计算的服务类型：**(1) 基础设施即服务 (IaaS)** **(2) 平台即服务 (PaaS)** **(3) 软件即服务 (SaaS)**，是指一种让业务活动能够订阅网页传递应用软件的服务。

【例题 · 单选题】

目前，云计算可以简单理解为计算与存储服务的集中化，有三种主要服务方式，除了以下哪一种（ ）。

- A. 基础设施即服务 (IaaS)
- B. 平台即服务 (PaaS)
- C. 客户即服务 (CaaS)
- D. 软件即服务 (SaaS)

【答案】C

【解析】 服务方式基础设施即服务 (IaaS)：一种信息系统的策略。在该策略下，企业外包其用于数据处理活动的设备。平台即服务 (PaaS)：一种为用户提供计算平台的服务方式。软件即服务 (SaaS)：一种让业务活动能够订阅网页传递应用软件的服务。

区块链

• **区块链** 是一个信息技术领域的术语。

- 从本质上讲，它是一个**共享数据库**，存储于其中的数据或信息，具有“不可伪造”“全程留痕”“可以追溯”“公开透明”“集体维护”等特征。
- 基于这些特征，区块链技术奠定了坚实的**“信任”** 基础，创造了可靠的“合作”机制，具有广阔的运用前景。

区块链对于财务工作的改变

- 区块链 (Blockchain) 是指通过**去中心化** 和 **去信任** 的方式集体维护一个可靠数据库的技术方案。通俗一点说，区块链技术就指一种全民参与记账的方式。
- 所有的系统背后都有一个数据库，你可以把数据库看成是一个大账本。那么谁来记这个账本就变得很重要。

分布式记账

分布式记账 技术是分布在多个节点或计算设备上的数据库，每个节点都可以复制并保存一个分类账，且每个节点都可以进行独立更新。它的特征是分类账**不由任何中央机构维护**，分类账的更新是由每个节点独立构建和记录的。

节点可以对这些更新进行投票，以确保其符合大多数人的意见。这种投票又被称为**共识**，共识会通过算法自动达成。共识一旦达成，分布式分类账就会自行更新，分类账的最新的商定版本将分别保存在每个节点上。

从传统记账到分布式记账

从公司角度来看，不同规模和财务结构的公司，记账模式不一样，记账的可信度也不一样。

一方记账（给个人记账），数据随意篡改。

二方记账（出现会计和出纳的职能分开），相对安全。

三方记账（公司内部的出纳、会计，加上外部的第三方审计），相对得以保障。

审计的问题： 第三方记账的“独立性”的问题，比如说第三方记账的“利益冲突”的问题——既要从上市公司获取报酬，又要保证不能按上市公司的“意志”办事。

这是这种三方记账模式的核心矛盾所在，也是现在上市公司造假案例频出的原因所在。只要第三方记账模式下的利益冲突不解决，造假问题基本无解。

多方记账，继续把三方记账扩展至四方记账、五方记账、更多方记账，参与记账的人越多，参与记账的人越不相关，则系统就越健壮，账目越不容易篡改，造假的可能性就越低；

企业记账，记得的是财务账；而区块链记账，记得的是数据账。但本质都是一样的，都是特定格式下用数据记录交易、资金流向。

分布式记账与单纯的多方记账还不一样，区别在于它还增加了一个 **Token** 的激励元素。

这使得它不再需要从公司内部获取收益，记账本身变成了一件有收益的事，而这也从根本上解决了第三方记账的独立性问题与道德风险。

智能合约

智能合约（Smart contract）是一种旨在以信息化方式传播、验证或执行合同的计算机协议。智能合约允许在没有第三方的情况下进行可信交易，这些交易可追踪且不可逆转。智能合约概念于 1995 年由 Nick Szabo 首次提出。

智能合约的目的是提供优于传统合约的安全方法，并减少与合约相关的其他交易成本。

人工智能对财务的影响

人工智能的概述

人工智能（Artificial Intelligence，英文缩写为 AI）。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

人工智能是计算机科学的一个分支，它企图了解**智能的实质**，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。

人工智能对财务的影响

“除非我们认识到并接受变革，否则会计可能面临灭顶之灾”

《Challenges and Opportunities of Big Data and Artificial Intelligence for the Accounting and Auditing Profession》（大数据和人工智能对会计和审计职业带来的挑战和机遇）的主题报告。

目前，被广泛认为**正在改变财务工作**的十大信息技术包括：

大数据、电子发票、云计算、数据挖掘、移动支付、机器学习、移动互联、图像识别、区块链和数据安全技术。

技术的直接应用：

1. 光学字符识别系统

识别字迹等。

2. 机器人流程自动化

可以把工作流程分为各项技术，给每项技术加上区块链，计算机物联网系统自动追踪货物的流转，并对预期产生的各项数据，进行收集整理及加以分析。

3. 电子财务记账系统

企业更多的财务运营依靠电子记账来实行。

财务机器人与财务 / 审计人员的工作相比，优势在于：

(1) 审计人员由于在固定且有限的时间、人员不多的情况下进行审计，对于各项财务往来、货物流转等事件的真实性不能确保一定准确，只能提供合理保证，而且成本较高。

(2) 财务机器人可以通过各项技术，把全年的业务进行核对，准确的同时，大大提高工作效率。

传统财务人员的劣势：

(1) 财务流程中有高度重复的手工操作，耗费大量的人力和时间。

(2) 跨岗位的实务操作需要协同处理，沟通成本高且效率低下。

(3) 手工处理存在较高的容错率，且获取的数据准确性低。

(4) 人工处理财务相关的事务，无法快速响应业务变化和拓展。

(5) 受困于时间人力，某些合规和审计工作通过抽样的方式进行，无法到达 100% 全覆盖。

人工智能的优势：

(1) 规范记账、精准报税、合法节税、智能风控。

(2) 系统可以实时监控整个企业的运转，杜绝员工徇私舞弊等。

(3) 智能风控。侦测财务造假，经营数据的监控可以有效防止内外财务数据造假。

(4) 财务危机预警，以现金流、利润率、周转率等体现企业有没有财务危机的数据进行分析预测。

未来展望

1. 虽然这两年的财务机器人在会计行业引起了非常大的轰动，但是财务机器人十分昂贵，应用不成熟，所以只有极少数的大公司将他投入使用。

2. “人工智能+财资管理”已成趋势，企业借助人工智能的关键技术和核心能力，将使得企业财务和金融形成包括结算、融资、票据、投资、风控以及财务公司运营、供应链金融等有机结合的整体智能化平台。可以为企业提供多方面、多层次的财资管理分析和经营决策。

未来的岗位需求更在于管理会计，而非财务会计。但目前大部分公司职责混淆。

【例题 • 单选题】

下列哪一项，不是人工智能的在财务方面技术的直接应用（ ）。

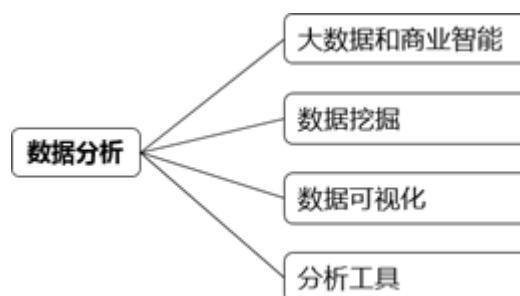
- A. 光学字符识别系统
- B. SOP 标准化流程手册
- C. 机器人流程自动化
- D. 电子财务记账系统

【答案】B

【解析】SOP 标准化流程手册，是在系统电子化之前，传统财务部就会执行的流程规范方法。

第 73 讲-数据分析 (1)

第 F.4 节 数据分析 (1)



大数据

对于“**大数据**”(Big data)研究机构Gartner给出了这样的定义。“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产。

麦肯锡全球研究所给出的定义是：一种在获取、存储、管理、分析方面**大大超出了传统数据库**软件工具能力范围的数据集合，具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征。

大数据特征：

2001年麦塔集团(META Group)分析师莱尼在一份报告中对大数据提出了“**3D 数据管理**”的观点，即认为大数据将往高速、多样、海量3个方向发展，提出了3个特性：高速性(Velocity)、多样化(Variety)、规模化(Volume)，统称3V。

大数据从“3V”到“4V”，“8V”

大数据特征：

容量(Volume)：数据的大小决定数据的价值和潜在的信息。

种类(Variety)：数据类型的多样性。

速度(Velocity)：指获得数据的速度。

价值(Value)：合理运用大数据，以低成本创造高价值。

精准性(Veracity)：从数据采集，应用场景入手到算法设计。

动态性(Vitality)：大数据的实时变化。

可视化(Visualization)：大数据的易于表达。

可变性(Variability)：妨碍了处理和有效地管理数据的过程。

大数据

大数据特征 (4V) :

一是数据体量巨大。百度资料表明，其新首页导航每天需要提供的数据超过 1.5PB（1PB=1024TB），这些数据如果打印出来将超过 5 千亿张 A4 纸。有资料证实，到目前为止，人类生产的所有印刷材料的数据量仅为 200PB。

二是数据类型多样。现在的数据类型不仅是文本形式，更多的是图片、视频、音频、地理位置信息等多类型的数据，个性化数据占绝对多数。

三是处理速度快。数据处理遵循“1秒定律”，可从各种类型的数据中快速获得高价值的信息。

四是价值密度低。以视频为例，一小时的视频，在不间断的监控过程中，可能有用的数据仅仅只有一两秒。

大数据结构化：

大数据包括结构化、半结构化和非结构化数据，非结构化数据越来越成为数据的主要部分。

据 IDC 的调查报告显示：企业中 80% 的数据都是非结构化数据，这些数据每年都按指数增长 60%。

有用的数据一般来自于结构性数据（比如交易数据），它们可以较好地符合电子表格或数据库的要求。

半结构性数据有个人网页浏览记录和传感数据。

非结构性数据涉及文本、音频和视频数据、社交网络的评论等。

第 74 讲-数据分析 (2)

第 F.4 节 数据分析 (2)

数据分析

企业利用这一过程将数据中的信息提取出来，进行处理、识别、加工、呈现，最后成为指导企业经营管理的知识和智慧。因而，企业利用这一过程的成熟程度，决定了企业使用数据的有效程度。

**大数据的机遇：**

- (1) 大数据的挖掘和应用成为核心，为企业探寻新的战略机遇带来了契机。
- (2) 对大数据的处理和分析成为新一代信息技术应用的支撑点。
- (3) 大数据的商业价值和市场需求成为推动信息产业持续增长的新引擎。
- (4) 大数据安全更加重要，为信息安全带来发展契机。

大数据的挑战：

- (1) 外部业务需求的数据转换。
- (2) 大数据技术运用仍存在困难。
- (3) 用户隐私与便利性的冲突。
- (4) 数据安全风险更加凸显。
- (5) 数据分析与管理人才紧缺。

大数据的发展：

- (1) 大数据将创造新的细分市场。
- (2) 大数据应用促使商业模式向以“数据租售”为直接盈利的模式转变。
- (3) 大数据由网络数据处理走向企业级应用。
- (4) 大数据成为智力资产和资源，信息部门从成本中心转向利润中心。
- (5) 大数据从商业行为上升到国家发展战略。

- (6) 大数据科学越来越大众化。
- (7) 从大数据技术到大数据科学的发展趋势。

商业智能 (BI)

商业智能 (Business Intelligence, 简称 BI)，又称商业智慧或商务智能，指用现代数据仓库技术、线上分析处理技术、数据挖掘和数据展现技术进行数据分析以实现商业价值。

其中，数据展现技术，数据可视化的**指示板功能也会为企业提供数据的模式和趋势。**

商业智能作为一个工具，是用来处理企业中现有数据，并将其转换成知识、分析和结论，辅助业务或者决策者做出正确且明智的决定。是帮助企业更好地利用数据提高决策质量的技术，包含了从数据仓库到分析型系统等。

【例题·单选题】

某公司希望利用商业智能 (BI) 来协助评估关键指标。IT 经理建议在 BI 工具中加入指示板功能。以下哪一项是管理层应该实施指示板功能的主要原因（ ）。

- A. 它允许管理层展示尽可能多的各类图表
- B. 它为整个公司提供数据的模式和趋势
- C. 它旨在专注于尚未达到的指标
- D. 它能够为生产活动自动续订重要原材料

【答案】B

【解析】商业智能 (Business Intelligence, 简称 BI)，又称商业智慧或商务智能，指用现代数据仓库技术、线上分析处理技术、数据挖掘和数据展现技术进行数据分析以实现商业价值。

其中，数据展现技术，数据可视化的指示板功能也会为企业提供数据的模式和趋势。

数据挖掘

数据挖掘又称数据库中的知识发现 (Knowledge Discover in Database, 简称 KDD)，是目前人工智能和数据库领域研究的热点问题，所谓数据挖掘是指从数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的非平凡过程。

数据挖掘是一种决策支持过程，它主要基于人工智能、机器学习、模式识别、统计学、数据库、可视化技术等，高度自动化地分析企业的数据，做出归纳性的推理，从中挖掘出潜在的模式，帮助决策者调整市场策略，减少风险，做出正确的决策。

技术上的定义及含义

数据挖掘 (Data Mining) 就是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程。

这个定义包括好几层含义：

- 数据源必须是真实的、大量的、含噪声的。
- 发现的是用户感兴趣的知识。
- 发现的知识要可接受、可理解、可运用。
- 并不要求发现放之四海皆准的知识，仅支持特定的发现问题。

商业角度的定义

数据挖掘是一种新的商业信息处理技术，其主要特点是对商业数据库中的大量业务数据进行抽取、转换、分析和其他模型化处理，从中提取辅助商业决策的关键性数据。简而言之，数据挖掘其实是一类深层次的数据分析方法。

SQL 结构化查询语言 (Structured Query Language, 简称 SQL) 是一种特殊目的的编程语言，是一种数据库查询和程序设计语言，用于存取数据以及查询、更新和管理关系数据库系统。

SQL 从功能上可以分为 3 部分：数据定义、数据操纵和数据控制。

【例题·单选题】

以下哪一项是数据挖掘的定义（ ）。

- A. 运用统计技术从大型数据库中提取和分析数据以识别模式和趋势的过程
- B. 用于开发公司绩效指标的系统
- C. 使用算法来促进公司内高效交流的过程
- D. 用于组织和解释复杂数据的系统，用以确保数据被准确记录在数据库中

【答案】A

【解析】运用统计技术从大型数据库中提取和分析数据以识别模式和趋势的过程，是大数据挖掘的重要作用。

数据可视化

可视化的终极目标是洞悉蕴含在数据中的现象和规律，这里面有多重含义：发现、决策、解释、分析、探索和学习。

简明定义是：通过可视表达增强人们完成某些任务的效率。比如，相同统计特征（方差、均值等）的几组数据可视化出来的结果是完全不同的。

近几年来，随着人工智能的兴起，人们逐渐发现其实一些机器能比人做得更好的事情，同时也发现了一些事情需要借助人类3亿年的进化本领。所以将可视化与分析进行结合，产生了一个新的学科，可视分析学。

可视分析学被定义为由可视交互界面为基础的分析推理科学，将图形学、数据挖掘、人机交互等技术融合在一起，形成人脑智能和机器智能优势互补和相互提升。



传统的可视化可以大致分为探索性可视化和解释性可视化，按照应用来分，可视化有多个目标：

1. 有效呈现重要特征
2. 揭示客观规律
3. 辅助理解事物概念和过程
4. 对模拟和测量进行质量监控
5. 提高科研开发效率
6. 促进沟通交流和合作

数据可视分析和数据挖掘的目标都是从数据中获取信息与知识，但是手段不同。

数据可视分析是将数据呈现给用户以易于感知的图形符号，让用户交互地理解数据。数据挖掘是通过计算机自动或者半自动地获取数据隐藏的知识，并将获取的知识直接给予用户。

也就是说，数据可视化可以看到交互界面，更适合于探索性地分析数据。而数据挖掘面对的是一堆活生生但黑不溜秋的数据，需要像挖矿一样从中发现金子。

数据可视化基础应用



1. 柱状图

柱状图是使用频率最高的图表。



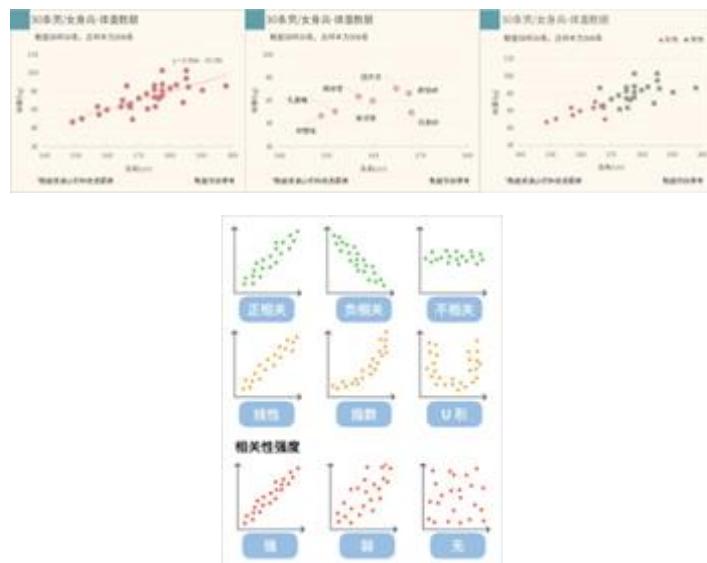
2. 条形图

与柱状图对应的就是条形图，如果优先观察数据的分布、类别名称比较长等，优先使用条形图，展示效果会比柱状图要好很多。



3. 散点图

散点图的适合场景（XY 双变量分析）：XY 两个变量之间的关联与联系，例如：身高/体重、广告投放/收入等等，如果需要分析变量之间的关系，则使用散点图



4. 气泡图

在散点图的基础上，如果数据的维度增加多一维，那么就可以使用气泡图，气泡图的性质与散点图接近，但是不适合展示大量数据，一般也可以用来观察少量数据的分布情况。



5. 雷达图

如果数据的维度暴增，变成了 4~10 维的话，那么要对比他们的差异，使用雷达图是最佳的，但是雷达图有一个限制，就是数据量不能很大，并且尽量让数据归一化进行对比（同数据量对比）否则会丧失意义。



6. 折线图

如果数据量比较大，并且数据是沿着某个方向进行有规律变化的，那么就可以使用折线图，折线图无论是展示大量数据，还是少量数据，都非常适合。



数据可视化

数据可视化工具为技术、管理人员和其他知识工作者提供了新方法，可以显著的提高他们掌握隐藏在数据中的信息的能力。

数据可视化的作用：

- (1) 加强商业信息传递效率；
- (2) 快速访问相关业务见解；
- (3) 更好地理解运营和业务活动；
- (4) 快速识别最新趋势；
- (5) 准确的客户情感分析；
- (6) 与数据直接交互；
- (7) 预测销售分析；
- (8) 深入销售分析；
- (9) 轻松理解数据；
- (10) 定制数据可视化。

虽然数据可视化的应用作用很大，也越来越受到企业用户的重视，但是从目前的技术发展水平来看，数据可视化依然存在一些局限性，这些局限性也对企业发展产生一定的影响。

数据可视化的局限性：

- (1) 基本成本（设备投入与时间投入）；
- (2) 依赖数据专家；
- (3) 报表级别不高（二维的数据延展性不强）；
- (4) 只有技术人员才能发现价值。

【例题·单选题】

医院注意到一种疾病的病例数增加，于是要求其分析师搜集过去三年的病例数据。该分析师发现，这种疾病出现在三年前的第二季度。从那以后，每年的第三和第四季度的病例数显著高于前两个季度。以下哪一种方法最能呈现分析师的发现（ ）。

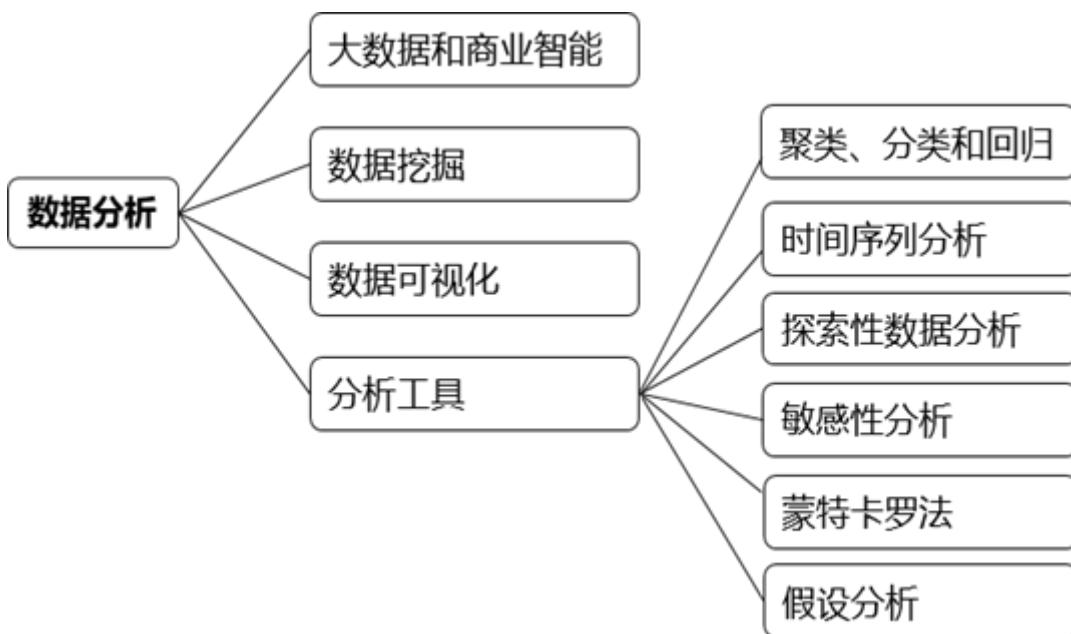
- A. 表格，显示过去三年每个月的病例数
- B. 饼状图，显示过去三年每个季度的病例数
- C. 散点图，显示过去三年每个季度的病例数的变化
- D. 条形图，显示过去三年每个季度的病例数

【答案】D

【解析】柱状图也可以采用，如果优先观察数据的分布、类别名称比较长等，优先使用条形图，展示效果会比柱状图要好很多。

第 75 讲-数据分析 (3)

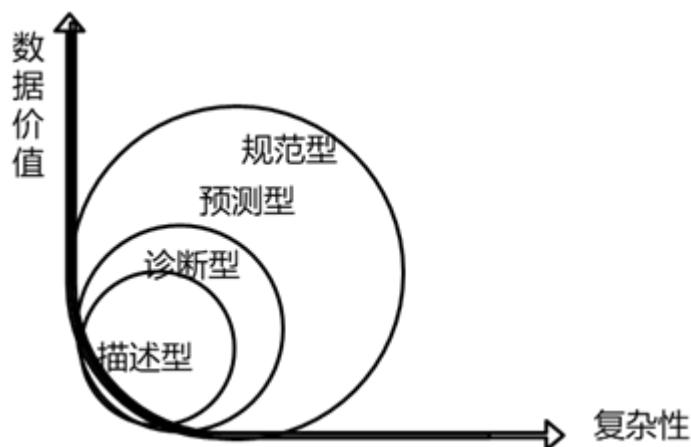
第 4 节 数据分析 (3)



分析工具

数据分析四种类型

在数据科学领域所遇到的四种类型的数据分析模式：描述型、诊断型、预测型和规范型（指导型）。不同类型的分析能提供不同的商业价值，每一种分析都有它自己的用处。



描述型：发生了什么？

这是最常见的一种。在业务中，它向分析师们提供业务的重要衡量标准的概览。

例如，每月的利润和损失账单。类似地，分析师可以获得大批客户的数据。了解客户（如，30%的客户是自雇型）的地理信息也可认为是“描述型分析”。充分利用可视化工具能增强描述型分析所带来的信息。

诊断型：为什么会发生？

这是描述型分析的下一步难题。通过评估描述型数据，诊断分析工具使得分析师们能够深入分析问题的核心原因。

设计良好的商业信息 dashboard 整合了时间序列数据（譬如，在多个联系时间点上的数据）的读入、特征的过滤和钻入功能，能够用于这类分析。

预测型：可能发生什么？

预测型分析主要是进行预测。某事件在将来发生的可能性，预测一个可量化的值，或者是估计事情可能发生的某个时间点，这些都可以通过预测模型完成。

预测模型通常运用各种可变数据来作出预测。数据成员的多样化与可能预测的目标是相关联的（如，人的年龄越大，越可能发生心脏病，我们可以说年龄与心脏病风险是线性相关的）。随后，这些数据被放在一起，产生分数或预测。

在一个充满不确定性因素的世界里，能够预测允许人们作出更好的决定。预测模型在很多领域都被用到。

规范型（指导型）：我需要做什么？

在价值和复杂度上，下一步就是指导性模型。指导性模型基于发生了什么、为什么会发生以及一系列“可能发生什么”的分析，帮助用户确定要采取的最好的措施。很显然，指导性分析不是一个单独的行为，实际上它是其他很多行为的主导。

交通应用是一个很好的例子，它帮助人们选择最好的回家路线，考虑到了每条路线的距离、在每条路上的速度、以及很关键的目前的交通限制。另一个例子是生成考试时间表，不让任何学生的时间表发生冲突。

【例题 • 单选题】

最有可能对组织产生最大影响同时也是最复杂的数据分析类型是（ ）。

- A. 诊断型分析
- B. 预测型分析
- C. 描述型分析
- D. 规范型分析

【答案】D

【解析】在价值和复杂度上，最后一步就是（规范型分析）指导性模型。指导性模型基于发生了什么、为什么会发生以及一系列“可能发生什么”的分析，帮助用户确定要采取的最好的措施。很显然，指导性分析不是一个单独的行为，实际上它是其他很多行为的主导。

聚类、分类和回归

“聚类、分类和回归” 主要被用在机器学习的领域。

给定一个样本特征，我们希望预测其对应的属性值，如果是离散的，那么这就是一个分类问题，反之，如果是连续的实数，这就是一个回归问题。

给定一组样本特征，我们没有对应的属性值，而是想发掘这组样本在多维空间的分布，比如分析哪些样本靠的更近，哪些样本之间离得很远，这就是属于聚类问题。

回归方法

回归方法是一种对数值型连续随机变量进行预测和建模的监督学习算法。使用案例一般包括房价预测、股票走势或测试成绩等连续变化的案例。

回归任务的特点是标注的数据集具有数值型的目标变量。也就是说，每一个观察样本都有一个数值型的标注真值以监督算法。

回归方法中的具体方法

- 回归树；
- 深度学习；
- 最近邻算法。

线性回归的适用条件

线性回归的理解与解释都十分直观，但是，需要注意的是，在变量是非线性关系的时候表现很弱。

分类方法

分类方法是一种对离散型随机变量建模或预测的监督学习算法。使用案例包括邮件过滤、金融欺诈和预测雇员异动等输出为类别的任务。

许多回归算法都有与其相对应的分类算法，分类算法通常适用于预测一个类别（或类别的概率）而不是连续的数值。

无论是分类还是回归，都是想建立一个预测模型，给定一个输入，可以得到一个输出；不同的只是在分类问题中，预测结果是离散的。

而在回归问题中预测结果是连续的。所以总体来说，两种问题的学习算法都很类似。所以在这个图谱上，我们看到在分类问题中用到的学习算法，在回归问题中也能使用。分类问题最常用的学习算法包括 SVM（支持

向量机），SGD（随机梯度下降算法），Bayes（贝叶斯估计），Ensemble（集成），KNN等。而回归问题也能使用SVR，SGD，Ensemble等算法，以及其它线性回归算法。

聚类方法

聚类是一种无监督学习任务，该算法基于数据的内部结构寻找观察样本的自然族群（即集群）。使用案例包括细分客户、新闻聚类、文章推荐等。

因为聚类是一种无监督学习（即数据没有标注），并且通常使用数据可视化评价结果。如果存在「正确的回答」（即在训练集中存在预标注的集群），那么分类算法可能更加合适。

聚类方法中的具体方法

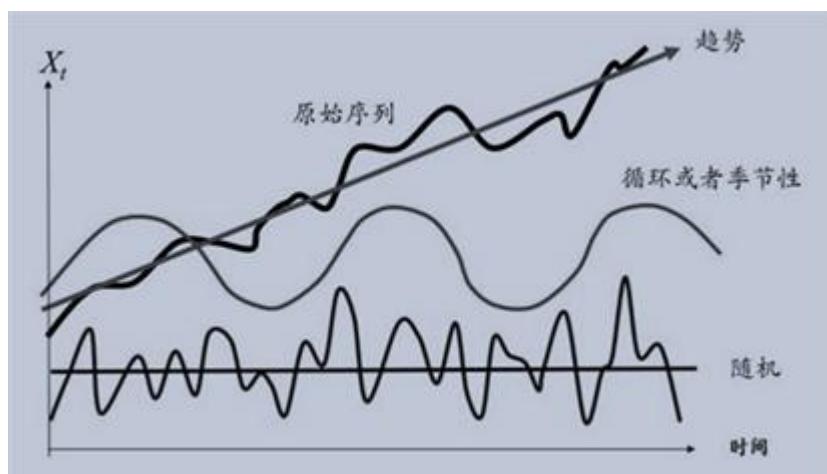
K-Means /聚类将数据划分为预定的类数 K，原理简单便于处理大数据。

- K-中心聚类；
- 系统聚类（多层次聚类）。

时间序列分析

时间序列分析 (Time-Series Analysis) 是指将原来的数据分解为四部分来看，分别是趋势、周期、季节性和随机性因素，然后综合这些因素，提出相关预测。

强调的是通过对一个区域进行一定时间段内的连续遥感观测，提取图像有关特征，并分析其变化过程与发展规模。



长期趋势变化：受某种基本因素的影响，数据依时间变化时表现为一种确定倾向，它按某种规则稳步地增长或下降。

使用的分析方法有：移动平均法、指数平滑法、模型拟合法等。

季节性周期变化：受季节更替等因素影响，序列以一固定周期规则性的变化，又称商业循环。

采用的方法：季节指数。

循环（周期性循环）变化：周期不固定的波动变化。

随机性变化：由许多不确定因素引起的序列变化。

时间序列分析优缺点讨论：

构成时间序列的要素有两个：其一是时间，其二是与时间相对应的变量水平。实际数据的时间序列能够展示研究对象在一定时期内的发展变化趋势与规律，因而可以从时间序列中找出变量变化的特征、趋势以及发展规律，从而对变量的未来变化进行有效地预测。

假定事物的过去趋势会延伸到未来；预测所依据的数据具有不规则性；撇开了市场发展之间的因果关系。

【例题 • 单选题】

时间序列分析 (Time-Series Analysis) 是指将原来的数据分解为多维度来看，分别是（ ）。

- A. 趋势、周期、季节性和随机性因素
- B. 趋势、和未来发展可能性
- C. 风险与时间结合
- D. 季节的周期性

【答案】A

【解析】时间序列分析 (Time-Series Analysis) 的衡量因素分别是：趋势、周期、季节性和随机性因素，然后综合这些因素，提出相关预测。强调的是通过对一个区域进行一定时间段内的连续遥感观测，提取图像有关特征，并分析其变化过程与发展规模。

探索性数据分析

探索性数据分析 (Exploratory Data Analysis, 简称 EDA) 主要的工作是对数据进行清洗，对数据进行描述（描述统计量，图表），查看数据的分布，比较数据之间的关系，培养对数据的直觉，对数据进行总结等。

探索性数据分析 (EDA) 与传统统计分析的区别

传统的统计分析方法通常是先假设样本服从某种分布，然后把数据套入假设模型再做分析。但由于多数数据并不能满足假设的分布，因此，传统统计分析结果常常不能让人满意。

探索性数据分析方法注重数据的真实分布，强调数据的可视化，使分析者能一目了然看出数据中隐含的规律，从而得到启发，以此帮助分析者找到适合数据的模型。“探索性”是指分析者对待解问题的理解会随着研究的深入不断变化。

应用传统统计分析方法的数据分析步骤：

提出问题 Problem => 准备数据 Data => 建模 Model => 分析 Analysis => 得出结论 Conclusions

应用探索性数据分析方法的数据分析步骤：

提出问题 Problem => 准备数据 Data => 分析 Analysis => 建模 Model => 得出结论 Conclusions

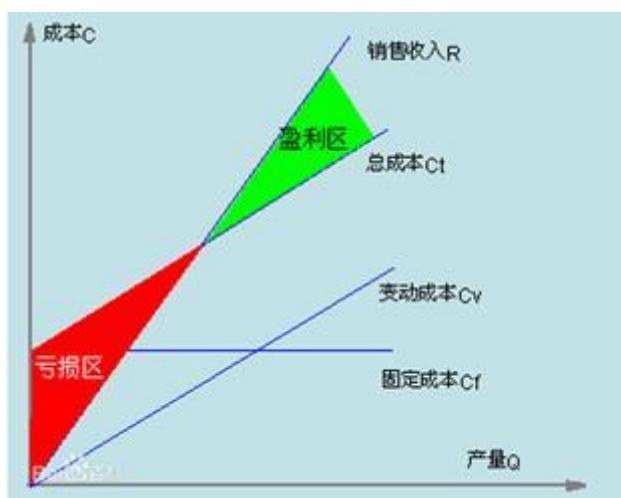
敏感性分析

敏感性分析是投资项目的经济评估中常用的分析不确定性的方法之一。

从多个不确定性因素中逐一找出对投资项目经济效益指标有重要影响的敏感性因素，并分析、测算其对项目经济效益指标的影响程度和敏感性程度，进而判断项目承受风险的能力。

敏感性分析是指从定量分析的角度研究有关因素发生某种变化对某一个或一组关键指标影响程度的一种不确定分析技术。其实质是通过逐一改变相关变量数值的方法来解释关键指标受这些因素变动影响大小的规律。

敏感性因素一般可选择主要参数（如销售收入、经营成本、生产能力、初始投资、寿命期、建设期、达产期等）进行分析。若某参数的小幅度变化能导致经济效果指标的较大变化，则称此参数为敏感性因素，反之则称其为非敏感性因素。



蒙特卡罗法

蒙特卡罗法也称统计模拟法、统计试验法。是把概率现象作为研究对象的数值模拟方法。是按抽样调查法求取统计值来推定未知量的计算方法。

蒙特卡罗是摩纳哥的著名赌城，该法为表明其随机抽样的本质而命名。故适用于对离散系统进行计算仿真试验。在仿真计算中，通过构造一个和系统性能相近似的概率模型，并在数字计算机上进行随机试验，可以模拟系统的随机特性。

蒙特卡罗法的基本思想：为了求解问题，首先建立一个概率模型或随机过程，使它的参数或数字特征等于问题的解，然后通过对模型或过程的观察或抽样试验来计算这些参数或数字特征，最后给出所求解的近似值。

(1) 根据实际问题的特点，构造简单而又便于实现的概率统计模型，使所求的解恰好是所求问题的概率分布或数学期望；

(2) 给出模型中各种不同分布随机变量的抽样方法；

(3) 统计处理模拟结果，给出问题解的统计估计值和精度估计值。

蒙特卡罗法的最大优点:

- (1) 方法的误差与问题的维数无关。
- (2) 对于具有统计性质的问题可以直接进行解决。
- (3) 对于连续性的问题不必进行离散化处理。

蒙特卡罗法的缺点:

- (1) 对于确定性问题需要转化成随机性问题。
- (2) 误差是概率误差。
- (3) 通常需要较多的计算步数。

假设(或目标寻求)分析

假设分析，又名“what if”分析，是结构化头脑风暴分析方法的一种。这种方法首先对当前的决策提出一系列假设性的问题，然后通过分析和回答这些假设性的问题来评估一个决策的风险程度；这里所说的假设性问题通常是一些质疑性的问题。

【例题 • 单选题】

某公司利用十年的数据开发了以下多元回归方程，并用它来估算其产品的成本。

$$\text{总成本} = F + AL + BM$$

其中：F=固定成本

L=人工成本（每小时）

M=原料成本（每磅）

以下哪一项改变最有可能对该模型失效产生最大的影响（ ）。

- A. 改用低成本国家的原料
- B. 大幅削减工厂间接费用
- C. 劳动生产率的重大改变
- D. 工会重新谈判的更高工资

【答案】C

【解析】注意前面章节相关的问题。不同系数的理解。