chapter 1

1. 针对收集到的 2019-nCoV 疫情数据，我们采用什么样方式处理可以保证了解比较准确的情况?

大数据采集过程中通常有一个或多个数据源，这些数据源包括同构或异构的数据库、文件系统、服务接口等，易受到噪声数据、数据值缺失、数据冲突等影响，因此需首先对收集到的大数据集合进行预处理，以保证大数据分析与预测结果的准确性与价值性。

多维的数据源、结合领域知识等

2. 公交车 GPS 数据可能创造哪些价值?应该用什么样的商业模式实现其价值?

通过 GPS 获取车辆的位路信息和里程数据，这些数据将改善车险定价技术与核保政策，提升精准定价能力。

(1)对出行信息进行实时的采集,并在驾驶 员选择路线前,将各区域拥挤状况提供给驾驶员。车辆通过 GPS 信号确定其自身位置。这些位置信息被发送给车队管理者,管理者通过地图便可观 测到各个车辆的具体位置。路线选择软件为货车指定额外的工作,并以电子指令的形式传回给驾驶员。 (2)具体的位置信息也可以保存在在途系统中,以便于日后的分析。这种在途系统公告板系统还可以监视 车辆的运行状况并在检测到特殊事件发生时,向车场提供信息。

出行流量分析、交通状况了解、判断道路路径优化等。

3. 根据你的理解，解释什么是数据分析、商务智能、商务分析、商务数据分析。

（1）数据分析：用适当的统计分析方法对收集到的大量数据进行分析，提取有用信息和知识 并加以利用的过程

（2）商务智能：利用信息技术手段，对相关数据进行加工处理，将其转化为组织所需要的知识，并支持决策的手段或工具

（3）商务分析：从组织的业务和战略出发，对组织的业务及战略进行分析、建模、模拟和预测，以支持组织相关决策的过程

（4）商务数据分析：调查分析问题出现的原因并预测未来。应对 商务环境下的计划与 解决方案;业务及战 略决策建议

雨课堂：
1-1: ACD
1-2: ABC

chapter 2

1.根据你自己的理解，怎样看待商务数据分析人员是一种复合人才的观点?

熟悉工作业务、掌握数据分析的基本方法、能够熟练使用数据分析的工具、具备项目的管理能力、拥有一定的设计技巧。商务数据分析人员需要具备管理能力、方法和工具的使用技能以及设计技巧等多个方面的能力。

2.解决问题的一般过程有哪几个步骤?主要内容是什么?

发现问题、搜集信息、分析假设、检验结果、应用结论

3.商务数据分析一般需要几个步骤?每个步骤的主要目的是什么?

（1）问题识别、确认与分解（将问题分解为可管理、可解决状态）；（2）数据收集;（3）数据分析模型选择；（4）数据整理（为了将数据预处理）；（5）数据分析（为了发现模型）；（6）结果解释与分析结果应用（为了发现知识）

4.从示例中如何看待心智模式对管理者及其行为的影响?如何完善自己的心智模式?

（1）心智模式可以影响管理者如何看待事物和管理者的认知方法。良好的心智模式可以帮助管理者发掘公司的潜能，使管理着决策更有成效。

（2）改善自己心智模式的方法主要有两种方式，一是反思自己的心智模式，通过反思与学习改善自己的心智模式；二是探询他人的心智模式，从自己与别人的心智模式的比较中完善自己的心智模式。

5.如何将大问题转换成为可以解决的小问题?

（1）自顶向下结构化分解（层次分析）（2）系统化问题分解

受限抓住加以抽象概括，形成高层次的概念，然后逐步考虑细节问题，把整个若干小问题，然后分别解决。

6.商务数据分析过程中三类分析方法的特点分别是什么?

（1） Porter 的竞争势力模型：**以竞争为导向、以现存产业为研究对象、关注产业盈利潜力。分析焦点集中在 企业外部利益相关者身上**，如现实及潜在竞争对手、供应商、顾客等，而对企业 内部诸如股东、员工等利益相关者考虑较少。

（2）SWOT 分析：从某种意义上来说隶属于企业**内部分析**方法，即根据企业自身的条件在既定内进行分析。主要理论基础也强调从结构分析入手对企业的外部环境和内部资源进行分析。

（3）波士顿矩阵分析：波士顿矩阵法的应用不但提高了管理人员的分析和战略决策能力，**同时还帮助他们以前瞻性的眼光看问题，更深刻地理解企业各项业务活动之间的联系**，加强了业务单位和企业管理人员之间的沟通，及时调整企业的业务投资组合，收获或放弃萎缩业务，加大在更有发展前景的业务中的投资，紧缩那些在没有发展前景的业务中的投资。但同时也应该看到这种方法的局限性，该方法也难以同时顾及两项或多项业务的平衡。

（4）产品生命周期法：在采用高价格的同时，**只用很少的促销努力。高价格的目的在于能够及时收回投资，获取利润；低促销的方法可以减少销售成本**。

7.是否理解课堂上讨论的常用商务分析方法?它们各自的优缺点是什么?

（1） Porter 的竞争势力模型

-优点：1.企业多元经营的决策工具，尤其当企业准备进入与其 核心能力关系不大的行业时，依据波特五力分析模型分析行业特点，可帮助企业 了解和控制投资风险。；2.企业市场定位的分析工具，在选定低成本、差异化或 集中化等战略的同时，企业作为领导者、竞争者、追随者的角色也在很大程度上 被确定下来。；3.企业利润分析的辅助工具，通过对五种因素的分析， 可明确维持企业利润来源的关键因素，有针对性地制定提高利润率的措施。

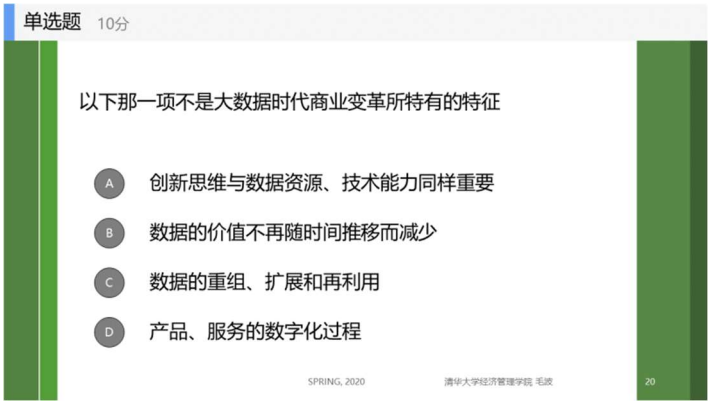
-缺点：该模型更多是一种理论思考工具，而非可以实际操作的战略工具。该模型的理论是建立在以下三个假定基础之上的：1、制定战略者需要了解整个行业的信息，显然现实中是难于做到的。2、同行业之间只有竞争关系，没有合作关系。但现实中企业之间存在多种合作关系，不一定是你死我

活的竞争关系。3、行业的规模是固定的，因此，只有通过夺取对手的份额来占有更大的资源和市场。但现实中企业之间往往不是通过吃掉对手而是与对手共同做大行业的蛋糕来获取更大的资源和市场。同时，市场可以通过不断的开发和创新来增大容量。

（2）SWOT 分析：
-优点：1.SWOT 分析的最大优点之一是它没有相关的成本。这是一个分析, 任何人在业务上可以合理地完成, 因此, 不需要任何专家或顾问的参与。**它是分析公司内任何职能或行业中的项目和建议的有效方法。**2.SWOT 分析的前提是在分析的概念中**识别优势、劣势、机遇和威胁**。理想的结果对于公司来说是非常重要的, 就是最大化优势, 尽量减少弱点, 使公司能够利用上面列出的外部机会来克服已确定的威胁。3.SWOT 分析的另一个好处是它能帮助为企业创造新的想法。通过研究 SWOT 分析中列和行中出现的问题, 作为一个社会, 它不仅提高了对潜在优势 (或劣势) 和威胁的认识, 而且还可以帮助我们在将来更有效地作出反应, 形成计划。
-缺点：1.由于典型的 SWOT 分析在前提上很简单, 所以通常不会受到批评。如果公司只专注于记录的准备工作, 那么它可能没有充分关注实现其目标的手段。2.需要进行更多的研究: 为了使 SWOT 分析真正成功, 它必须超越一个简单的优势、劣势、机会和威胁列表。3. 为了对公司业绩产生影响的分析, 业务决策必须基于可靠的、相关的和可比较的数据。然而, SWOT 数据的收集和分析可以是一个主观的过程, 反映了个人的偏见, 进行分析。

（3）波士顿矩阵分析：
-优点：波士顿矩阵法的应用不但**提高了管理人员的分析和战略决策能力**，同时还帮助他们以前瞻性的眼光看问题，更深刻地理解企业各项业务活动之度间的联系，加强了业务单位和企业管理人员之间的沟通，及时调整企业的业务投资组合，收获或放弃萎知缩业务，加大在更有发展前景的业务中的投道资，紧缩那些在没有发展前景的业务中的投资。
-缺点：由于评分等级过于宽泛，可能会造成两项或多项不同的业务位于一个象限中回;由于评分等级带有折衷性，使很多业务位于短阵的中答间区域，难以确定使用何种战赂。同时，这种方法也难以同时顾及两项或多项业务的平衡。

（4）产品生命周期法：
-优点：产品生命周期（PLC）提供了一套适用的营销规划观点。它将产品分成不同的策略时期，营销人员可针对各个阶段不同的特点而采取不同的营销组合策略。此外，产品生命周期只考虑销售和时间两个变数，简单易懂。
-缺点：1、产品生命周期各阶段的起止点划分标准不易确认。2、并非所有的产品生命周期曲线都是标准的 S 型，还有很多特殊的产品生命周期曲线。3、无法确定产品生命周期曲线到底适合单一产品项目层次还是一个产品集合层次。4、该曲线只考虑销售和时间的关系，未涉及成本及价格等其它影响销售的变数。5、易造成“营销近视症”，认为产品已到衰退期而过早将仍有市场价值的好产品剔除出了产品线。6、产品衰退并不表示无法再生。如通过合适的改进策略，公司可能再创产品新的生命周期。



答案：D

chapter 3

1.**结构化数据**的特点是什么?如何体现在商务数据分析中?

可以使用关系型数据库表示和存储，表现为二维形式的数据。 数据以行为单位，一行数据表示一个实体的信息，每一行数据的属性是相同的。

2.针对数据来源看，哪些可能属于结构化数据?

经营业务数据、网络及外部收集的数据、设备采集数据等
层次模型的数据组织 、网状模型的数据组织 、关系模型的数据组织

3.根据你的理解，数据抽样的目的是什么?如何避免缺失样本类别?

目的是利用样本数据推断出总体参数的特征。

现实世界中的数据异常杂乱，属性值缺失的情况经常发全甚至是不可避免的。造成数据缺失的原因是多方面的：信息暂时无法获取、信息被遗漏、有些对象的某个或某些属性是不可用的、有些信息（被认为）是不重要的、获取这些信息的代价太大、系统实时性能要求较高
应该尽量避免出现缺失值，数据采集时应慎重考虑其缺失值处理方法的利弊。

4.问卷调查、焦点小组访谈、德尔菲方法之间的联系和区别是什么?

问卷调查法是通过调查问卷提出问题的方式收集资料的研究方法，

焦点小组访谈是由主持人组织以一种无结构的自然形式与一组被调查对象交流 进行数据采集的方法，可以在问卷调查的基础上使用，完善和补充问卷调查反映的内容

德尔菲方法是通过专家意见汇总出整体结论的方法，通常是问卷调查法和焦小组访谈的补充。

5.从个人喜好的角度，分析一下不同内容类型问题布局方法的优劣

半封闭半开放的问题就中和了两种类型的优点和缺点，半封闭半开放的问题就是说，你在这样的问题选项上先设计出几个既定的答案让人们选择，而为了避免人们找不到自己心里所想的答案时，你就可以在最后的一个选项中写上其他，然后让人们写下自己想要的答案。

6.根据你的理解，给出一个重复数据的定义

单项数据值重复不可直接认定为重复数据，对样本各项进行统一检查才能确定重复数据的存在，因分析模型要求，非重复值也有可能是重复数据。重复数据可理解为需要剔除的数据。

7.数据预处理包括哪些内容?为什么要进行数据预处理

重复数据、异常数据、缺失数据
数据标准化，数据离散化。在真实世界中，数据通常是不完整的、不一致的、极易受到错误值或异常值的侵扰的。 因为数据库太大，而且数据集经常来自多个异种数据源，低质量的数据将导致低质量的挖掘结果

8.根据课堂上讲授的内容，哪些数据获取与清洗的步骤可能导致商务数据分析结果的偏差?

样本抽样的方法、调查方法的选择、调查问卷的设计、重复数据的处理、缺失值的处理、异常值的处理等可能导致商务数据分析结果的偏差。

答案：B

chapter 4

1.根据你的判断，观察研究法是否科学?为什么要采用观察研究法?观察研究法的特点是什么?

观察研究法是指研究者根据一定的研究目的、研究提纲或观察表，用自己的感官和辅助工具去直接观察被研究对象，从而获得资料的一种方法。科学的观察具有目的性和计划性、系统性和可重复性。

观察研究法能够启发思维，导致新的发现，扩大感性认知。

特点：通过观察直接获得资料，不需其他中间环节，观察具有及时性的优点，观察能搜集到一些无法言表的材料。然而观察研究法充满混杂因素。

2. 什么是混杂因素，如何区分研究结果中的混杂因素?实验中设立控制组就一定能排除混杂因素的影响吗?

混杂因素就是研究对象之间的个体差异，它们不是比较研究的对象，但会导致分析结果的敏感度变差 。

并不能完全排除。

3. 请阐述优化问题的一般结构。优化问题一定能得到最优解吗?为什么?

（1）目标函数：希望最大化或最小化的对象就是目标，目标函 数表现出目标与决策变量之间的关系

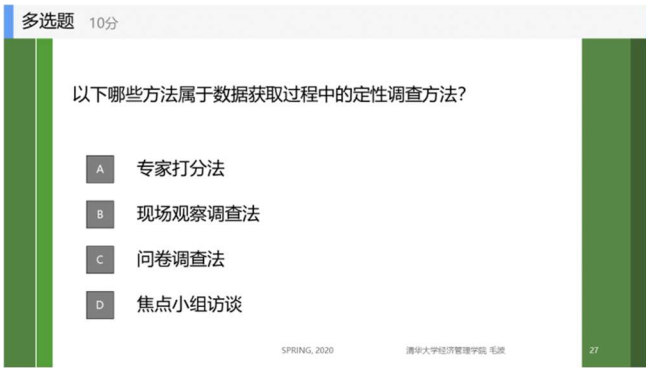
（2）决策变量：在最优化过程中你可以控制的因素就是决策变 量，决策变量的取值决定了目标函数的结果

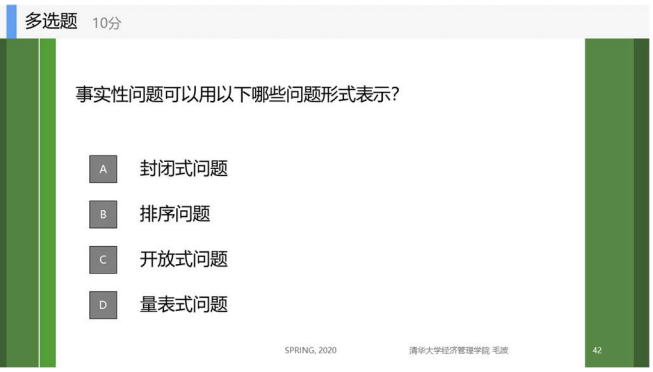
（3）约束条件：不可控制因素构成最优化问题的约束条件，它们决定了问题的参数，限定了决策变量的取值

不一定。非线性规划的单纯性法，根本不能保证最终一定能够找到最优解。

4. 如何理解“All models are wrong, but some are useful.”?

所有模型都是错的，但是有一些是有用的。因为不存在能够完美描述市场的模型，因此没有绝对的真理或者公式可以解释，只有现实才是真理。但可以用统计学去解释，预测某一些宏观现





象，并为之提供依据。

答案：ACD

答案：ABCD

chapter 5

1.在“数据-信息-知识-智慧”框架中，数据图形化的作用主要体现在哪些层次?用什么方法可以达到目的?

主要体现在信息层次，数据图形化的作用就是通过更直观地展示数据，进行数据的汇总、比较和观察。

2.多元图形的概念是什么?它们如何发挥作用?

多元图形是能对三个以上变量进行比较， 最有可能促成最有效的比较的数据图形化方法。

3.为什么说“模型与数据吻合也可能有其他情况”?

构建的模型通常都会与数据吻合，但也不可避免存在其他的可能性 。描述数据模型时，需要论述可相互 换用的多种(至少两种)因果模型或图解，以保证思考比较完善。

4.根据你的观点，什么是判断假设是否正确的证据?主观的?客观的? 数字的?文字的?

客观的和数字的。

5.怎样理解证伪法是剔除不合理假设的合适方法?

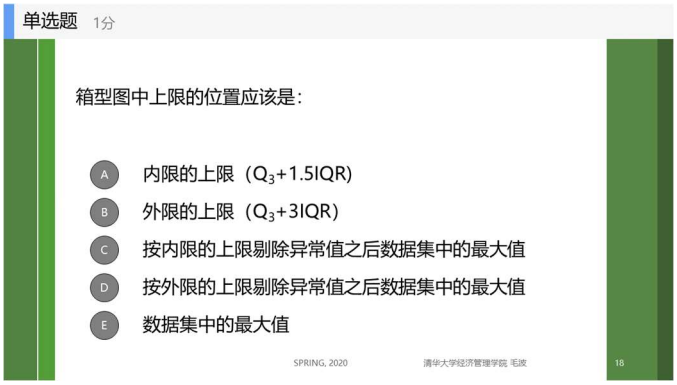
从有效的数据信息中提取出各种假设，并一一验证。假设检验的核心是证伪。请勿试图选出最合理的假设，只需剔除无法证实的假设——这就是假设检验的基础：证伪。进行假设检验时，要使用证伪法，而回避满意法（选出看上去最可信的一个假设的方法）。但并非所有假设都一定能被证伪，比如，某些证据可能会对假设不利，却无法推翻假设。优秀分析师的理想是找到复杂问题的非自觉答案。

6.不具备诊断性的证据与要检验的假设之间是否存在关联?

诊断性：是证据所具有的一种功能，能够帮助你评估所考虑的假设的相对似然，如果证据具有诊断性，就能版主对假设进行排序。进行假设检验时，重点是要识别和找出诊断证据，非诊断证据不会带来任何进展。将证据与待鉴别的假设列表确定诊断强度。以便选出强调最大的假设。

7.贝叶斯公式应用是否有条件?

需要有已知概率才可使用



答案：C

Chapter 6: 商务数据分析的过程——数据分析的方法（三）

1. 根据你自己的判断，主观概率是否合理？其主观性是其可能有偏差的根源吗？如何改善其偏差问题？

- 合理：很难说明主观概率是否合理，或者一个人的主观概率判断是否正确，但有经验的人比没有经验的人更能准确地判断形势。或者说，经验和信念有助于提高主观概率的合理性或准确性。虽然主观性较强，但又是根据经验、各方而后知识，对客观情况的了解进行分析、推理、综合判断而设定的，与主观臆测不同。因此，主观概率还是偏向于合理的。
- 其主观性是其可能有偏差的根源，因为它是一个表示自己对某事的确认程度的指标，而每个人的信念和确认程度有所不同。可以通过更多的观察和经验改变一个人的信念，将信念达成中立，改善其偏差问题。

2. 怎样形成主观概率？完全拍脑袋的方法是否可以？

- 直觉和经验，或者事件过去的相对频率的形式根据丰富的经验进行的推测等，对于有效的证据的确认，信任度能形成主观概率。完全拍脑袋的方法可以形成主观概率。

3. 启发法的核心是什么？如何看待启发法？哪些情况下启发法可能导致系统偏差？

- 启发法的核心：启发法可以用来解决复合问题和信息不安全的情况，在“直觉”和“最优化”之间建立一个桥梁
- 如何看待启发法：可能会得出正确答案，但不保证得出最优化答案
- 使用可得性启发法时，人们最容易想到的通常是过去经常发生的事件或近期发生的不寻常事件，但这些信息也可能对判断是不重要的或不够的，自然也会导致判断上的偏差。
- 人们往往趋向于在很少的数据基础上很快地得出结论。使用代表性启发法时，代表性会导致忽略样本大小。在分析事件特征或规律时，人们往往不能正确理解统计样本大小的意义，对总体进行统计的结果才是真正的结果，样本的数量愈接近真实的数量，统计的结果也就愈可信，样本愈小,与真实数量相差愈大,统计的结果愈不能反映真实的结果情况。代表性会忽略判断的难易程度，即使面对的是一个复杂的难以判断的问题，也简单地去作出判断，或经常根据不规范的和与判断无关的描述轻易地作出判断，或经常会忽略掉不熟悉或是看不懂的信息，只凭自己能够理解和熟悉的信息去作出判断，这些忽略掉的信息可能对判断是关键的。
- https://baike.baidu.com/item/%E5%90%AF%E5%8F%91%E5%BC%8F%E8%AE%A4%E7%9F%A5%E5%81%8F%E5%B7%AE/12750604?fr=aladdin

4. 根据直方图的展示是否可以直接判断分析模型的类型？

- 否。由于直方图主要用于表示资料变化情况和解析出资料的规则性，虽然直方图可以为分析决策提供思路和依据，也比较直观地显示指标的质量特性的分布状态，但不能直接判断分析模型的类型。

Chapter 6 雨课堂习题

1 根据你的理解，使用主观概率方法体现了出以下哪个观念？

① 经验是可以用来建模分析的
② 将定性数据转化为定量数据进行分析
③ 经验和判断是有价值的
④ 图形化可以分析出相关趋势

2 以下哪些对于启发法的描述是正确的？

① 启发法将定性问题转化为定量问题
② 启发法可以利用容易获得的指标置换难以获得的指标
③ 启发法可以在一定程度上简化推理过程
④ 启发法可能存在系统偏差

Chapter 7: 商务数据分析的过程——数据分析的方法（四）

1. 根据你的理解，回归是否可以预测未来？为什么？
- 可以。
2. 模型误差的含义是什么？能否用误差大小判断模型的优劣？
- 模型误差的含义：模型误差可以让预测和信念更全面。通过指出误差，可以知道切实的预期，了解更多信息，做出更好的决策。
- 能用误差大小判断模型的优劣。若误差越小，可判断模型的准确度越大，更易于查出最优的结果。
3. 根据你的理解，根据原始数据之间的关联构造的数据模型是否就可以构成分析用数据模型？为什么？
- 可以。为了分析目标，需要发现原始数据之间的关联，且采用符合分析数据之间的关联的模型。因此，可以构成分析用数据模型。
4. 为什么在数据模型分析中也要排除混杂因素？
- 通过排除混杂因素可以减少因果效应的混杂偏倚，能得出更全面的、客观的、符合分析目标的数据模型。否则，使得在判断因果关系时将对于因变量作用全部或部分归于无关因素的一种现象。

Chapter 7 雨课堂习题

1 基于已有的加薪数据和模型，你认为可以从哪些方面进一步完善能得到更准确的效果？

Chapter 8: 商务数据分析的过程——商务智能方法

1. 如何理解商务智能是一个综合性的概念？
- 商务智能将企业中现有的数据转化为知识，以各种系统和软件工具帮助企业做出明智的业务经营决策，提高效率和生产力，构造更强的客户关系，优化生成收入的战略，增加收入并使收益最大化。它不仅支持数据分析和商业决策制定的辅助，还涉及人力资源和企业利益方面的优化问题，因此可以把商务智能看作为一个综合性的概念。
2. 根据你的理解，数据仓库的核心是什么？数据库系统？技术？数据？应用？或者其他？
- 上述的都能成为核心。数据仓库能集中解决实际的业务问题、争取数据仓库的拥护者确保项目的实施、支持详尽的数据和可靠的历史数据、适应于业务的技术，这些都是数据仓库项目的关键成功因素。
3. 根据你的理解，OLAP 与 OLTP 的本质区别是什么？能否利用业务系统直接进行 OLAP？

- OLTP 是传统的关系型数据库的主要应用，主要是基本的、日常的事务处理，例如银行交易。OLAP 是数据仓库系统的主要应用，支持复杂的分析操作，侧重决策支持，并且提供直观易懂的查询结果。
- 能利用业务系统直接进行 OLAP。虽然基于数据仓库的 OLAP 的运行效率略低，但它的灵活性强、易于扩展，多维分析的数据模型等同于业务模型，因此可以直接进行 OLAP。
- https://yq.aliyun.com/articles/350209

4. 构造一个以分析清华大学学生情况的多维模型，探讨如何利用该模型的切块、切片、旋转和钻取等操作进行分析。能否使用普通数据库管理软件实现多维分析？

- Ex. 清华大学学生情况的组合维：年级、学院、地区等
- 切块：选定多维数据模型的一个三维子集进行观察
- 切片：用切片的方法从不同的角度观察
- 旋转：按照不同的顺序组合维，对数据进行考察
- 钻取：考察一个特定的维
- 否，普通数据库管理软件只能实现一或二维分析，难以实现多角度的分析。

5. 根据你的理解，在进行商务数据分析时，选择合适的数据挖掘（知识发现）方法的依据有哪些？

- 判断数据之间是否存在关系和规则（ex. 因果或关联、时间序列、分类等）

6. 分别勾画可利用预测、聚类、分类和关联方法去解决知识发现问题的现实商务问题场景，探讨其中具体方法的选择依据及模型评价准则

- 预测：一个企业收入数据的趋势，预算下一季的投资和收入政策等
  依据及评价准则：预测模型对现有数据分析建模，对未来发展走势作出判断
- 聚类：如把付费用户按照几个特定维度，如利润贡献，用户年龄，续费次数等聚类分析后得到不同特征的群体，在运营活动中为这些细分群体采取精细化，个性化的运营和服务，最终提升运营的效率和商业效果
  依据及评价准则：聚类分析的重要用途就是针对目标群体进行多指标的群体划分
  https://www.sohu.com/a/286412111_165070
- 分类：基于运营商数据的个人征信评估，国家电网客户用电异常行为分析等
  依据及评价准则：获得一个分类函数或分类模型，把相应的数据项映射到给定的类别中，易于提取分类规则
  https://blog.csdn.net/liulingyuan6/article/details/53637129/
- 关联：穿衣搭配推荐，购物篮推荐，电子商务搭配购买推荐
  依据及评价准则： 反映事件之间关联（依赖）的知识成为关联型知识（或依赖关系）。关联能找出所有的频繁项集，以及由频繁项集产生的强关联规则，这些规则的扩展性较强，适合用于推荐系统
  https://blog.csdn.net/liulingyuan6/article/details/53637846

7. 社交网络分析的数据如何收集？这些数据是否可以进行普遍意义上的数据挖掘？为什么？

- 由于社交网络分析不局限于学科或者领域，在不同场景可以通过网络爬虫、实验、问卷调查等的方式收集数据。这些数据可以进行普遍意义上的数据挖掘，因为从社交网络分析的数据中能提取人们（或者样本）感兴趣的知识和信息，实现数据挖掘的意义。

8. 怎样理解文本分析？它是一类独特的方法吗？

- 文本分析是一种非结构化数据，分析时采用信息检索和文本挖掘的方式，能与传统的数据挖掘方法结合。文本分析的数据挖掘与其他分析有些不同，需要先建立语料库，创建词频-文档矩阵，从词频-文档矩阵中提取知识。从利用词频-文档矩阵的特点来看，它是一类独特的方法。

9. 根据你的理解，文本的情感分析结果能否作为其他数据挖掘方法的输入数据？为什么？

- 否，文本的情感分析结果的形式基于文本的类型，需要特殊的挖掘方法。其他数据挖掘方法不需要经过创建文档矩阵和主观分析的过程，所以将情感分析结果作为其他数据挖掘方法的输入数据的话难以做出准确的、客观的分析。

Chapter 8 雨课堂习题

1 以下哪些内容属于日常事务处理时所需数据的特点？

① 数据有不同程度的综合
② 只包括当前的数据
③ 数据是分散的
④ 数据处理时存取操作频率高
⑤ 用户需要系统在短时间内进行反馈
⑥ 数据关系无法满足规范性要求

2 你理解云平台吗？以你的理解，你认为中台与各种云平台是一种什么样的关系？

3 以下哪些变动属于时间序列数据的要素？
① 循环变动
② 关联变动
③ 不规则变动
④ 周期变动（季节变动）
⑤ 长期变动

4 以下哪些指标不属于社交网络分析时常用的指标？
① 邻近中心性
② 节点的权重
③ 节点的入度
④ 节点的关联性

5 请思考一下如何利用优酷网站的公开信息构造优酷用户之间的社交网络联系图。

Chapter 9

1. 数据可视化是一个数据分析工具？一种分析解决问题的思路？ 或是一种思维方式？

数据可视化首先是一个数据分析可视化方法，是指通过对比来反映问题，用图形的方式来展现数据，从而更加清晰有效地传递信息，主要方法包括图表类型的选择和图表设计的准则。数据可视化有很多分析工具，如 Excel，Python 等等

数据可视化的应用，不是帮人解决问题而应该是让人能准确快速地从中获取有价值的信息，从而去解决问题。

2. 如何降低数据可视化中的 data-ink ratio？

我们可以通过减少网格线的数量，减轻网格线的数量来降低数据墨水比。

3. 为什么在数据可视化过程中还要使用表格？

图形比表格展示更多的信息，并且易读但某些场合下，表格比单纯图形更合适：① 需要保留具体的数值信息 ② 需要进行不同数据间的精确比较 ③ 数据的计量单位不同或量级不一样

4. 根据你的理解，用不同的可视化方法对同一数据集进行展示， 是否可以得到不同的结论？

用不同的可视化方法对同一数据集进行展示，可以得到不同的结论。因为您可以使用四种基本的表示类型来表示数据：

- 比较方式
- 组成
- 分布
- 关系

上面的每种方法都用不同的图形展示方法，如 比较方法（Bar chart，Line Chart 等等），组成（Pie，Stacked Chart 等等），分配（Histogram 等等），关系（Bubble，Scatter chart）。

在实际的数据可视化中，往往不是孤立地用一个基本图形，把多个图形组合、邻接，能交叉对比出更多的信息。比如在柱状图上叠加折线图，在地图上叠加散点图，把多个柱状图放在一起对比等等。

记住数据可视化展现信息是第一位的，好看倒是其次。

5. 如何判断商务数据分析时所采用数据可视化方法是否合适？

你要问自己要用数据讲什么样的故事以及要传达给听众什么消息，您就可以为项目或计划选择合适的数据可视化类型。还有，要问自己“谁是我的听众？”最后问自己，您是否要分析特定趋势？您想证明数据的**组成**吗？您是否要**比较**两组或更多组值？您想如何显示您的KPI？等等问题。

6. 如何判断一个企业是否是可视化组织？

一个企业能否被可视化，在于这个能否找出组织的战略地图(Strategic Map)，业务流程图(Business Processes Flow Chart)与组织架构图(Organization Chart)，这三者可以将企业的无形思考与运转变为有形。“可视化”本身就是将企业的战略意图、业务流程和组织架构变为一种比较直观和能够统筹安排的视图使员工进行理解，并且进一步加以分析，及时对系统的问题做出反应和调整。

https://www.0734zpw.com/n8240.html

7. 清华大学是否需要成为可视化组织？如何帮助学校建设成为可 视化组织？

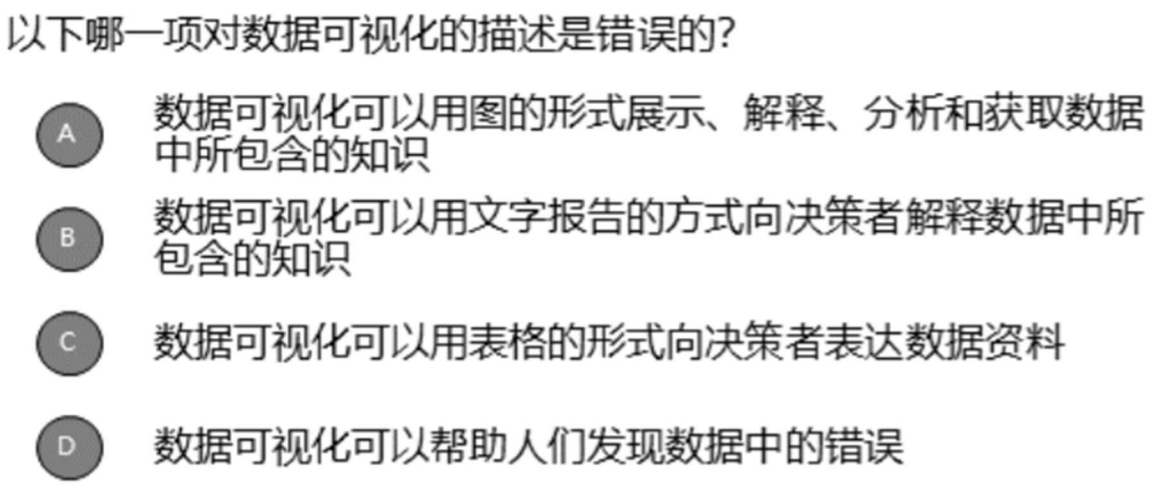
需要，尽管清华大学是一所大学，但是已经出现过许多不同层级诉求与政策相冲突的情况。这反映了系统的故障不能得到及时的反馈，系统的问题也不能得到快速的发现。

如何建设成可视化组织：首先，根据学校战略层面的思考，如愿景、使命、价值观、运作基本模式以及短期的发展建设规划等，确定组织的需要；根据需要，来匹配组织的业务流程（学校的业务可能包括很多方面，但是最主要的业务是以学生为核心的业务）；根据业务流程，匹配学校已有的组织架构。使用数据可视化的方法，来确定组织的业务和组织中的人员的效用。最终通过数据集反映的现状对可视化后的结果进行调整，得到目前形成的可视化组织架构、业务流程。

然而，这一定不是一个一成不变的行动。组织的可视化应当是持续进行的，随着学校根据外部环境对自身的战略做出调整，数据可视化就要保持动态的常态跟进。进而根据数据反映的事实和战略需要，调整业务流程和组织架构。

雨课堂 ch9

1．答案【B】



2.

(思考与小憩)

如果我们希望用刚刚讲述过的图形方式反映截止到今天疫情期间清华大学师生员工的健康状况，你认为哪一种图形工具比较合适？为什么？

【答案】我认为用折线图的方式，绘制师生员工的肺炎疫情患病比例随时间的变化趋势图。也可以用饼状图绘制感染新冠肺炎的各个群体的人数占比。用柱状图也可能绘制不同人群的不同阶段的患病人数。

3.答案【A，D】

以下哪些项对可视化图形工具的描述是正确的?

A 条形图或柱状图通常用来描述属性类信息

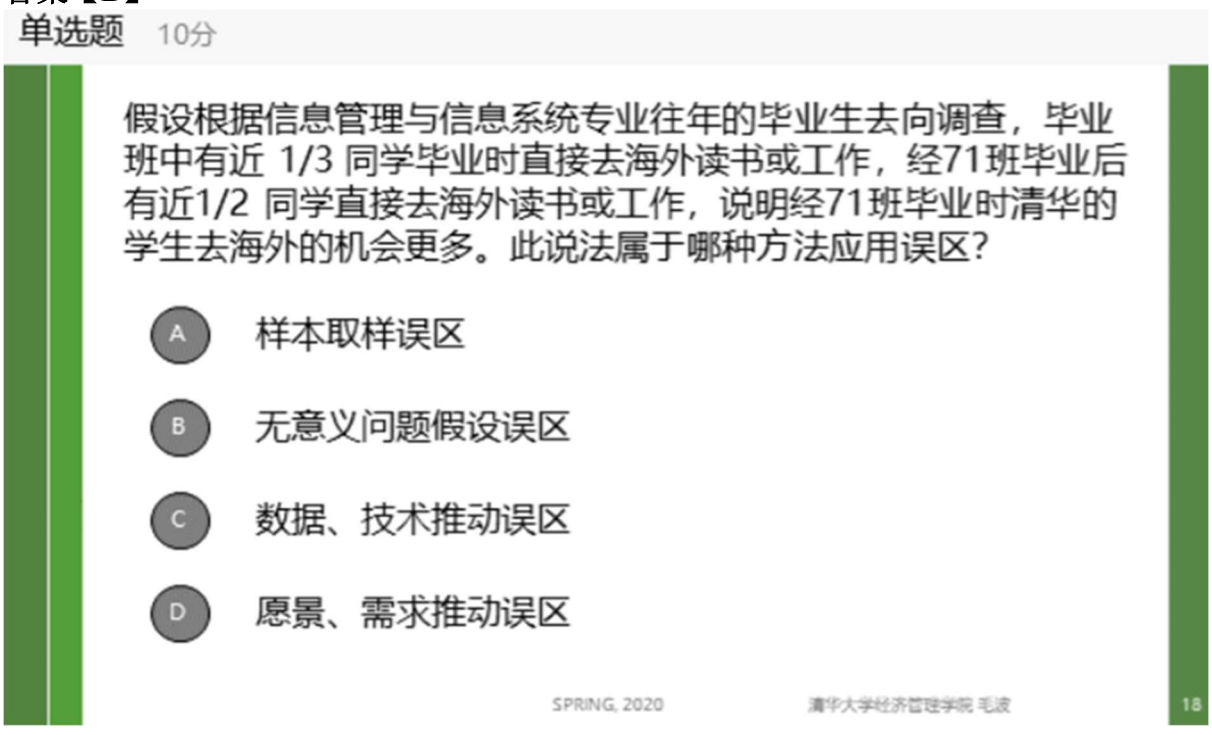
B 圆饼图通常用来表示数据的排序

C 折线图通常用来表示数据之间的比较信息

D 地理信息图通常用来表示数据在某些地理区域上的分布

雨课堂题 ch10

1. 答案【B】



2.

(思考与小憩)

如何看待近期专家们普遍降低了对那些利用零假设检验方法验证的模型的关注度。

Chapter10

1. 如何理解商务数据分析项目的 **BASP** 框架？

2. 根据你的理解，如何解释不同类型的数据导致不同类型的智能？

3. 根据你的理解，伦理道德问题是否构成了商务数据分析领域发 展的障碍？如何来解决？

4. 怎样理解“数据、技术是中性的，使用者不是中性的”这句话？

5. 从个人角度出发，如何平衡隐私保护和享受方便服务这两个方 面？

6. 根据你的理解，“幸存者偏差”和小样本事件是不是同一类事 件？

7. 如何避免陷入数据分析中的各种误区和陷阱？