北京市商务数据分析初赛理论样题

注　意　事　项

1、考试时间：90 分钟。
2、请首先按要求在试卷的标封处填写您的姓名、身份证号。
3、请仔细阅读各种题目的回答要求，在规定的位置填写您的答案。
4、不要在试卷上乱写乱画，不要在标封区填写无关的内容。

	一	二	总　分
得　分			

得　分	
评分人	

姓　名

身份证号

一、单项选择(第 1 题～第 80 题．选择一个正确的答案，将相应的字母填入题内的括号中．每题 1 分，满分 80 分．)

1. 当前大数据技术的基础是由(　　)首先提出的。

A、微软

B、百度

C、谷歌

D、亚马逊

2. 大数据的核心是(　　)。

A、分析事情的原因

B、预测事情发生的可能性

C、评估事情的现状

D、评估事情的相关性

3. 大数据最明显的特点就是(　　)。

A、数据体量大

B、数据类型繁多

C、价值密度低

D、处理速度快

4. (　　)是机器学习的成果之一。

A、可视化分析

B、语义引擎

C、预测性分析能力

D、数据质量管理

5. (　　)反映数据的精细化程度，越细化的数据价值越高。

A、规模

B、活性

C、关联度

D、颗粒度

6. 数据清洗的方法不包括(　　)。

A、缺失值处理

B、噪声数据清除

C、一致性检查

D、重复数据记录处理

7. 下列关于数据重组的说法中，错误的是(　　)。

A、数据重组是数据的重新生产和重新采集

B、数据重组能够使数据焕发新的光芒

C、数据重组有利于实现新颖的数据模式创新

D、数据重组实现的关键在于多源数据融合和数据集成

8. 大数据时代，数据使用的关键是(　　)。

A、数据收集

B、数据存储

C、数据分析

D、数据再利用

9. 大数据的本质是(　　)。

A、挖掘　　B、洞察　　C、联系　　D、搜集

10. 以下关于大数据的分析理念的说法中，错误的是(　　)。

A、在数据基础上倾向于全体数据而不是抽样数据

B、在分析方法上更注重相关分析而不是因果分析

C、在分析效果上更追究效率而不是绝对精确

D、在数据规模上强调相对数据而不是绝对数据

11. (　　)指对客观事件进行记录并可以鉴别的符号，是对客观事物的性质、状态以及相互关系等进行记载的物理符号或这些物理符号的组合。

A、数据　　B、数字　　C、文字　　D、信息

12. 大数据环境下的隐私担忧，主要表现为(　　)。

A、个人信息的被识别与暴露

B、用户画像的生成

C、恶意广告的推送

D、病毒入侵

13. 大数据与小数据的根本区别在于大数据采用(　　)方式，小数据强调抽样。

A、定向思维

B、相关思维

C、全样思维

D、实验思维

14. 下列关于数据交易市场的说法中，错误的是(　　)。

A、数据交易市场是大数据产业发展到一定程度的产物

B、商业化的数据交易活动催生了多方参与的第一方数据交易市场

C、数据交易市场通过生产数据、研发和分析数据，为数据交易提供帮助

D、数据交易市场是大数据资源化的必然产物

15. 大数据的起源是(　　)。

A、金融

B、互联网

C、公共管理

D、电信

16. 根据不同的业务需求来建立数据模型，抽取最有意义的向量，决定选取哪种方法的数据分析角色人员是(　　)。

A、数据分析员

B、研究科学家

C、数据管理人员

D、软件开发工程师

17. 下列关于舍恩伯格对大数据特点的说法中，错误的是(　　)。

A、数据规模大

B、数据类型多样

C、数据处理速度快

D、数据价值密度高

18. 当前社会中，最为突出的大数据环境是(　　)。

A、互联网　　B、物联网　　C、综合国力　　D、自然资源

19. 在数据生命周期管理实践中，(　　)是执行方法。

A、数据存储和备份规范

B、数据管理和维护

C、数据价值发觉和利用

D、数据应用开发和管理

20. 下列关于网络用户行为的说法中，错误的是(　　)。

A、网络公司能够捕捉到用户在其网站上的所有行为

B、用户离散的交互痕迹能够为企业提升服务质量提供参考

C、数字轨迹用完即自动删除

D、用户的隐私安全很难得以规范保护

21. (　　)统计分析报告是对统计报表进行说明的统计分析报告，亦称为“文字说明”，也就是们通常所说的报表说明。

A、说明型　　B、总结型　　C、调查型　　D、分析型

22. 数据采集人员通过设计具有针对性的问题，对用户需求习惯、喜好、产品使用反馈等数据进行采集，其采用的采集方式是(　　)。

A、系统日志数据采集

B、数据库采集

C、报表采集

D、调查问卷采集

23. 通过(　　)渠道，可以采集宏观经济数据、居民消费价格指数。

A、电子商务平台

B、国家统计局网站

C、百度指数

D、生意参谋

24. 电子商务常见的数据来源主要有内部数据和外部数据，以下属于外部数据的是(　　)。

A、订单数量　　B、加购数量　　C、360 趋势　　D、店铺订单信息表

25. 下列采集行为属于违法行为的是(　　)。

A、使用生意参谋工具导出自己店铺的运营数据

B、使用百度指数工具获取关键词搜索指数及用户画像数据

C、通过技术手段进入竞争对手网站数据库获取网站流量及销售数据

D、使用数据采集工具采集其他网站公开数据信息用于数据分析

26. 产品搜索指数是根据(　　)等因素综合计算得出的数值。

A、客单价

B、搜索频次

C、交易数量

D、支付金额

27. 数据统计分析的基本步骤是(　　)。

A、确定问题、收集数据、处理数据、制作数据图表、依据数据及图表分析得出结论

B、收集数据、处理数据、制作数据图表

C、确定问题、收集数据、处理数据、制作数据图表

D、确定问题、处理数据、制作数据图表、依据数据及图表分析得出结论

28. 制作报表的一般工作流程顺序应该是(　　)。

①搭建报表框架

②构思报表的大纲

③明确数据汇报的需求

姓　名

身份证号

④进行报表数据指标的选择

⑤报表填制与美化

A、①②③④⑤

B、①④②③⑤

C、③④②①⑤

D、③②④①⑤

29 以下属于使用 PowerPoint 制作数据分析报告优势的是(　　)。

A、结果可实时更新

B、可含有动态图表

C、适合演示汇报

D、适合日常数据通报

30. 以下不属于数据分析报告的标题需具备的要求的是(　　)。

A、直接

B、简洁

C、确切

D、数字化

31. 以下不属于数据分析报告的前言部分的是(　　)。

A、分析背景

B、专业术语解析

C、分析目的

D、分析思路

32. 撰写数据分析报告时应注意(　　)。

A、要基于可靠的数据源

B、根据数据的特点选择图表的配色方案，风格越多样越好

C、结论要明确、详实并有层次性，不可对未来进行预测

D、行文保持自己的语言习惯和风格即可，展现自己的独特性

33. 某网店运营人员准备分析该网店客户的产品购买款式和规格偏好，需要采集的数据是（　　）。

A、店铺商品的 SKU 销售数据

B、店铺商品采购数据

C、店铺商品销量信息

D、店铺订单的配送信息

34. 供应链是指围绕核心企业，从配套零件开始，制成中间产品以及最终产品，最后由销售网络把产品送到消费者手中的生产、交易全链条，供应链数据包括（　　）。

A、采购数据、

B、库存数据

C、物流数据

D、以上都是

35. 以下属于供应链数据指标的是（　　）。

A、直通车转化率

B、展现量

C、点击率

D、库存周转率

36. 针对国内和跨境电子商务提供数据采集和分析的工具是（　　）。

A、淘数据

B、生意参谋

C、店侦探

D、京东商智

37. 物流服务的优劣，关系到用户对于品牌、产品、卖家的印象，对于物流数据进行分析，以下不属于物流数据的是（　　）。

A、物流时效

B、物流异常量

姓　名

身份证号

批过超准不题答生考

C、库存周转率

D、物流服务满意度

38. 数据分析报告是对整个数据分析过程的总结与呈现。那么，针对数据分析报告的撰写，下列说法错误的是(　　)。

A、数据分析报告需图文并茂，让数据更加生动活泼

B、数据分析报告需要结构清晰、主次分明，能使读者正确理解报告内容

C、数据分析报告需要注重科学性和严谨性

D、|数据展示内容一般在结论部分进行

39. 下列说法错误的是(　　)。

A、数据的表现形式可以是符号、文字、数字、语音、图像、视频等

B、电子商务数据是企业进行电子商务营销活动时产生的行为数据和客户数据

C、数据分析指通过建立分析模型，对数据进行核对、检查、复算、判断等操作

D、电子商务数据分析指运用有效方法和工具收集、处理数据并获取信息的过程

40. 数据分类与处理的方法不包括(　　)。

A、数据采集

B、数据清洗

C、数据计算

D、数据排序

41. 下列属于反映比例关系的可视化图表的是(　　)。

A、旭日图　　B、散点图　　C、热力图　　D、气泡图

42. 散点图是对成组的(　　)数值进行比较，气泡图是对(　　)数值进行比较。

A、两个；两个　　B、两个，三个　　C、三个；两个　　D、四个；三个

43. 散点图矩阵通过(　　)坐标系中的一组点来展示变量之间的关系。

A、三维　　B、一维　　C、多维　　D、二维

44. 由于数据分析的重要性，将可视化与数据分析结合可形成一个新的学科是(　　)。

A、可视分析学

B、信息分析学

C、科学可视化学

D、数据可视化学

45. 什么图表最适合表现与时间有关的趋势或是两个变量之间的潜在关系?

A、流程图　　B、表面图　　C、线图　　D、网络图

46. 图表类型的选择有依赖于所要处理和展现的数据类型，例如连续数据的取值可以是任何范围，最适合用什么图表展示?

A、气泡图 B、饼图 C、曲线图 D、柱状图

47. 图表类型的选择有依赖于所要处理和展现的数据类型，例如离散数据的数值可清晰计数，最适合用什么图表展示?

A、饼图　　B、柱状图　　C、气泡图　　D、曲线图

48. 对于关系数据，我们要做的是尝试着探索事物的?

A、因果关系

B、线性关系

C、逻辑关系

D、相关关系

49. 我们用散点图可以推断出变量间的?

A、分布性　　B、相关性　　C、因果性　　D、结构性

50. 饼图不太适合表示精确数据的原因是(　　)。

A、人眼对面积的大小不敏感

姓　名

身份证号

批 过 超 准 不 题 答 生 考

B、饼图的角度不能精确展示数据

C、饼图上只能标记比例，不能标记数据值

D、对圆的设置会影响饼图对数据的展示

51. 几何体表达法对数据的不确定性可视化的缺点是(　　)。

A、需要精心选择视觉元素才能有效表达不确定性

B、易污染原有的确定性数据的可视化结果

C、容易产生视觉混淆问题

D、理解曲线较长，易引起疲劳

52. 下列属于基本图表的是(　　)。

A、瀑布图　　B、滑珠图　　C、漏斗图　　D、折线图

53. 可视化的输入是(　　)。

A、数据　　B、代码　　C、视觉形式　　D、语言

54. 可视化的目标是(　　)。

A、阻止数据爆炸

B、美观酷炫

C、清洗噪声

D、理解数据

55. 以下哪个选项不是可视化的作用？

A、传播交流　　B、信息记录　　C、数据采集　　D、数据分析

56. 使用以下哪种可视化工具不需要编程基础？

A、Tbleu　　B、3js　　C、Veg　　D、Processing

57. 下列选项中，不是地理信息数据可视化分析的应用是(　　)。

A、通过交互式发现拥堵的路口

B、通过地图分析微博数据的传播情况

C、自动计算异常的轨迹

D、通过图表了解区域之间的收入差异

58. 堆叠柱形图除了可以展示离散型时间数据，还可以展示(　　)。

A、比例的变化情况

B、数据随时间变化的趋势

C、多个部分到整体的关系

D、一个部分到整体的关系

59. 对于折线图来说，合理设置横轴长度的原因是?

A、使得折线剧烈变化，方便观察

B、正确展示折线变化趋势

C、正确设置横轴刻度

D、规范图的大小

60. 环形图采用(　　)表示各类别的占比。

A、角度　　B、弧度　　C、颜色　　D、宽度

61. 文本分析中去除无意义词语的步骤是(　　)。

A、分词　　B、停用词处理　　C、词频统计　　D、数据预处理

62. 某超市研究销售纪录数据后发现，买啤酒的人很大概率也会购买尿布，这种属于数据挖掘的(　　)问题。

A、关联规则发现　　B、分类　　C、回归　　D、聚类

63. 在假设检验中，不拒绝原假设意味着(　　)。

A、原假设肯定是正确的

B、原假设肯定是错误的

C、没有证据证明原假设是正确的

D、没有证据证明原假设是错误的

64. 线性回归能完成的任务是(　　)。

A、预测离散值　　B、预测连续值　　C、分类　　D、聚类

姓　名

身份证号

此 过 超 准 不 题 答 生 考

65. 随机森林方法属于(　　)。

A、梯度下降优化　　B、Bagging 方法　　C、Boosting 方法　　D、线性分类

66. (　　)是一个观测值，它与其他观测值的差别如此之大，以至于怀疑它是由不同的机制产生的。

A、边界点　　B、质心　　C、离群点　　D、核心点

67. 为了减小多重共线性的影响，可以使用(　　)模型。

A、岭回归　　B、逻辑回归　　C、线性回归　　D、多项式回归

68. 当不知道数据所带标签时，可以使用(　　)技术促使带同类标签的数据与带其他标签的数据相分离。

A、关联规则发现　　B、回归　　C、聚类　　D、分类

69. 现象之间线性依存关系的程度越低，则相关系数(　　)。

A、越接近于-1　　B、越接近于 1　　C、越接近于 0　　D、在 0.5-0.8 之间

70. Apriori 算法可用来解决(　　)问题。

A、分类　　B、预测　　C、聚类　　D、关联

71. 一元线性回归模型和多元线性回归模型的区别在于(　　)。

A、因变量的个数不同

B、自变量的个数不同

C、相关系数的大小不同

D、判定系数的大小不同

72. 进行假设检验时，在样本量一定的情况下，犯第一类错误的概率减小，犯第二类错误的概率就会(　　)。

A、不变　　B、减小　　C、增大　　D、不确定

73. 当某一列的数值差距比较大时，一般需要进行(　　)操作来减少预测误差。

A、相关性分析

B、特征离散

C、主成分分析

D、特征组合

74. 分类变量使用以下(　　)统计量进行缺失值填补较合适。

A、均值

B、最大值

C、中位数

D、众数

75. 决策树中不包含以下(　　)节点。

A、叶节点　　B、根节点　　C、内部节点　　D、外部节点

76. 关于 CART 算法，错误的是(　　)。

A、CART 算法既可以处理分类问题，也可以处理回归问题

B、可以处理样本不平衡问题

C、CART 算法采用信息增益率的大小来度量特征的各个划分点

D、CART 分类树采用基尼系数的大小来度量特征的各个划分点

77. 为了评估模型拟合的好坏，通常用(　　)来度量拟合的程度。

A、步长　　B、特征　　C、假设函数　　D、损失函数

78. 欠拟合的产生原因有(　　)。

A、学习到数据的特征过少

B、学习到数据的特征过多

C、学习到错误数据

D、机器运算错误

79. 主成分分析中的协方差矩阵是(　　)。

A、零矩阵　　B、对角矩阵　　C、上三角矩阵　　D、下三角矩阵

80. 在下列几个数值中，检验的 p 值为(　　)值时拒绝原假设的理由最充分。

A、95%　　B、50%　　C、5%　　D、2%

得　分	
评分人	

二、判断题(第 81 题～第 100 题．将判断结果填入括号中．正确的填“√”，错误的填“×”．每题 1 分，满分 20 分．)

81. (　　) 模型的 r2 系数为 0.36，我们认为模型对数据的拟合效果较好。

82. (　　) 支持度确定 Y 在包含 X 的事务中出现的频繁程度。

83. (　　) 得到线性回归的模型后，只要知道了自变量，就能知道因变量的准确值。

84. (　　) 检验时间序列的平稳性要用到单位根检验。

85. (　　) 支持向量机的学习策略便是间隔最大化，最终可转化为一个凸二次规划问题的求解。

86. (　　) 饼图一般用于显示数据系列中各项的大小与各项总和的比例，饼图中的数据点显示为整个饼图的百分比。

87. (　　) 当条形图数据系列的列数过多时，可以插入图表分界线将图表分为两部分，使图表表达数据更加清晰。

88. (　　) 在绘制图表的过程中，可以对不易理解的图表添加辅助图表，以便于用户理解和传达信息。

姓　名

身份证号

考生答题不准超过此线

89. (　　) 数据挖掘是指发现并纠正数据文件中可识别错误的最后一道程序，是对数据的完整性、一致性和准确性进行重新审查和校验的过程。

90. (　　) 反映发展趋势的可视化图表有柱型图、面积图、折线图。

91. (　　) 电子商务经营者对收集到的客户数据可以任意使用。

92. (　　) 数据采集人员需要针对数据采集需求、背景等对数据采集渠道进行可信度划分，优先选择可信度更高的数据采集渠道。

93. (　　) 淘数据能够用于跨境电商市场行情数据采集。

94. (　　) 百度指数属于行业趋势及人群数据分析工具。

95. (　　) 市场数据包括行业数据和竞争数据两个部分。

96. (　　) EXCEL 制作的报表可以包含动态图表，交互性较强并且可以实时更新。

97. (　　) 说明型统计分析报告是一种期限短、反映快的统计分析报告。

98. (　　) 计划型统计分析报告是对一定时期内的情况进行总结分析的统计分析报告。

99. (　　) 在数据采集录入的过程中通过平均、求和等操作与原始数据进行比对，发现比对结果不匹配，需要检查出相应的错误数据。

100. (　　) 运营数据采集中商品的采购数量、采购单价等可以通过查阅商品采购合同，调取财务或采购部门的采购记录等方式实施采集。