

第三章 统计与计量分析

一、单项选择题

1. $\bar{x} = \sqrt{x_1 x_2 \dots x_n}$ 是 () 的计算公式。

- A. 调和平均数
- B. 算数平均数
- C. 几何平均数
- D. 指数平均数

【答案】C

2. 四分位数是指将一组数据按照从小到大的顺序排序后, 处于数据中 () 和 () 位次上的数值, 这两个数值分别为下四分位数和上四分位数。

- A. 10%; 90%
- B. 20%; 80%
- C. 25%; 75%
- D. 30%; 70%

【答案】C

【解析】四分位数是指将一组数据按照从小到大的顺序排序后, 处于数据中 25% 和 75% 位次上的数值, 这两个数值分别为下四分位数和上四分位数。

3. 常见的集中趋势度量指标不包括 ()。

- A. 众数
- B. 极差
- C. 中位数
- D. 平均数

【答案】B

【解析】常见的集中趋势度量指标主要有平均数、中位数和众数。(极差是离散程度度量指标)。

4.在假定原假设正确的情况下,如果一个小概率事件发生了,则可以()。

- A. 拒绝备择假设, 接受原假设
- B. 拒绝原假设, 接受备择假设
- C. 拒绝原假设和备择假设
- D. 接受原假设和备择假设

【答案】B

【解析】小概率原理是指小概率事件在一次实验中几乎不可能发生,在假定原假设正确的情况下,如果一个小概率事件发生了,则可以认为原假设是错误的,进而拒绝原假设,接受备择假设。

5.期望()、标准差()的正态分布被称为标准正态分布。

- A. $\mu=1$; $\sigma=1$
- B. $\mu=0$; $\sigma=1$
- C. $\mu=1$; $\sigma=0$
- D. $\mu=0$; $\sigma=0$

【答案】B

6.在应用过程中发现,若对多元线性回归模型增加一个解释变量, R^2 一般会()。

- A. 减小
- B. 增大
- C. 不变

D. 不能确定

【答案】B

【解析】当利用 R^2 度量多元线性回归模型的拟合优度时， R^2 的值会随着解释变量的增多而增大，即便引入一个无关紧要的解释变量也会使 R^2 变大。

7. 根据拟合优度 R^2 与 F 统计量的关系可知，当 $R^2=0$ 时，有（ ）。

A. $F=-1$

B. $F=0$

C. $F=1$

D. $F=\infty$

【答案】B

【解析】拟合度 R^2 为回归平方和 ESS 占总离差平方和 TSS 的比例，计算公式为： $R^2 = ESS/TSS$

F 统计量的计算公式为：

$$F = \frac{ESS/k}{RSS/(n-k-1)} \sim F(k, n-k-1)$$

当 $R^2 = 0$ 时， $ESS = 0$ ，所以有 $F = 0$

8. 如果价格突破了头肩底颈线，期货分析人员通常会认为（ ）。

A. 价格将反转向上

B. 价格将反转向下

C. 价格呈横向盘整

D. 无法预测价格走势

【答案】A

【解析】期货人员通常认为，突破了头肩底颈线，宣告着牛市的来临；跌破头肩顶部的颈线，那么牛市就反转为熊市了。

9.对多元线性回归方程 $y_i = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \mu_i$ 的最小二乘回归结果显示, R^2 为0.92, 总离差平方和为500, 则残差平方和RSS为 ()。

- A.10
- B.40
- C.80
- D.20

【答案】B

【解析】 $R^2 = ESS/TSS = 1 - RSS/TSS$

将数据代入公式得到: $0.82 = 1 - RSS/500$, 解得 $RSS = 40$ 。

10.某种产品产量为1000件时, 生产成本为3万元, 其中固定成本6000元, 建立总生产成本对产量的一元线性回归方程应是 ()。

- A. $y_c = 6000 + 24x$
- B. $y_c = 6 + 0.24x$
- C. $y_c = 24000 - 6x$
- D. $y_c = 24 + 6000x$

【答案】A

【解析】设总生产成本对产量的一元线性回归方程为 $y = a + bx$ 。由题干可知, $x = 1000$ 时, $y_c = 30000$ 元; $x = 0$ 时, $y_c = 6000$ 元。将数据代入方程, 解得 $a = 6000$; $b = 24$ 。所以, 总生产成本对产量的一元线性回归方程为: $y_c = 6000 + 24x$ 。

11.DF 检验在实际应用中存在的一个问题, 即时间序列存在 () 阶滞后相关时 DF 检验才有效。

- A. 1
- B. 2
- C. 3
- D. 4

【答案】 A

【解析】 DF 检验在实际应用中存在一个问题：只适用于存在 1 阶滞后相关的时间序列。由于时间序列可能存在高阶滞后相关，直接使用 DF 检验法会出现偏误。

12. DF 检验回归模型为 $y_t = \alpha + \gamma y_{t-1} + \mu_t$ ，则原假设为（ ）。

- A. $H_0: \gamma=1$
- B. $H_0: \gamma=0$
- C. $H_0: \lambda=0$
- D. $H_0: \lambda=1$

【答案】 A

【解析】 DF 检验的原假设为 $H_0: \gamma=1$ ，备择假设 $H_1: \gamma<1$ 。若拒绝原假设，所检验序列不存在单位根，是平稳时间序列；若不拒绝原假设，所检验序列存在单位根，为非平稳时间序列。

13. 对于两个时间序列 $|x_t|$ 和 $|y_t|$ 均为一阶单整序列，即 $x_t \sim I(1)$ ， $y_t \sim I(1)$ ，若存在一组非零常数 α_0 和 α_1 使得 $\alpha_1 x_t + \alpha_2 y_t \sim I(0)$ ，则称 x_t 和 y_t 之间存在（ ）。

- A. 协整关系
- B. 因果关系
- C. 线性关系
- D. 自相关

【答案】A

【解析】协整是指多个非平稳性时间序列的某种线性组合是平稳的。某些时间序列是非平稳时间序列，但它们之间往往存在长期的均衡关系。具体来讲，对于两个均为1阶单整的时间序列 $\{x_t\}$ 和 $\{y_t\}$ ，即 $x_t \sim I(1)$ 和 $y_t \sim I(1)$ ，若存在一组非零常数 α_0 和 α_1 ，使得 $\alpha_0 x_t + \alpha_1 y_t \sim I(0)$ 则称 x_t 和 y_t 之间存在协整关系。

14. 格兰杰检验是通过（ ）来完成的。

- A. t 检验
- B. F 检验
- C. 单位根检验
- D. 拟合优度检验

【答案】B

15. 下列选项中，（ ）不是 ARMA 模型的分类。

- A. 移动平均模型
- B. 自回归模型
- C. 自回归移动平均
- D. 求和自回归移动平均模型

【答案】D

【解析】具体而言，ARMA 模型可细分为移动平均（MA）模型、自回归（AR）模型以及自回归移动平均（ARMA）。

16. 下列关于相关系数 r 的说法，错误的是（ ）。

- A. r 越接近于 1，相关关系越强
- B. $r=0$ 表示两者之间没有关系

- C. 取值范围为 $-1 \leq r \leq 1$
- D. r 越接近于0, 相关关系越弱

【答案】B

【解析】相关系数 r 的取值范围为： $-1 \leq r \leq 1$ 。当 $|r|$ 越接近于1时，表示二者间的线性相关关系越强；当 $|r|$ 越接近于0时，表示二者之间的相关关系越弱。当 $r > 0$ 时，表示二者之间存在正向相关关系；当 $r < 0$ 时，表示二者之间存在负向相关关系；当 $r = 0$ 时，并不表示二者之间没有关系，而是二者之间不存在线性关系。因此，B项说法错误。

二、多项选择题

1. 平均数是反映一组数据平均水平的统计指标，下列哪些是常用的平均数？（ ）

- A. 简单算术平均数
- B. 加权算术平均数
- C. 平方平均数
- D. 几何平均数

【答案】ABD

【解析】平均数是反映一组数据平均水平的统计指标。常用的平均数包括简单算术平均数、加权算术平均数和几何平均数。

2. 由正态总体得到的（ ），常被称为三大抽样分布。

- A. χ^2 分布
- B. F分布
- C. t分布
- D. Z分布

【答案】ABC

3. 下列关于 F 分布的表述，有误的是（ ）。

- A. F 分布和 t 分布之间有一个确定的等量关系
- B. F 分布由英国统计学家费希尔于 1924 年提出
- C. F 分布是一种对称分布
- D. F 分布是统计学中最常见也是应用最广泛的一种连续型分布

【答案】CD

【解析】C 项，F 分布是一种非对称分布。

D 项，正态分布是统计学中最常见也是应用最广泛的一种连续型分布。

4. 关于数据离散程度度量指标的表述，以下说法正确的有（ ）。

- A. 极差能全面反映数据的离散程度
- B. 标准差是方差的平方根
- C. 离散系数又称变异系数
- D. 在实际中标准差应用更加广泛

【答案】BCD

【解析】A 项，极差的优点是计算简便且易于理解，但该指标只使用了一组数据中的两个极端值，未充分利用中间部分数据的信息，因而无法全面反映数据的离散程度。

5. 多元线性回归模型的基本假定有（ ）。

- A. 零均值假定
- B. 同方差与无自相关假定
- C. 异方差假定
- D. 无多重共线性假定

【答案】ABD

【解析】多元线性回归模型满足如下基本假定：

- (1) 零均值假定。
- (2) 同方差与无自相关假定。
- (3) 无多重共线性假定，即解释变量之间不存在线性关系。
- (4) 随机扰动项与解释变量互不相关。
- (5) 正态性假定，随机扰动项 μ_i 服从正态分布，即 $\mu_i \sim N(0, \sigma^2)$ 。

6.应根据估计量的（ ），来评判一个估计量是否优良。

- A. 无偏性
- B. 有效性
- C. 一致性
- D. 可测性

【答案】ABC

【解析】评判一个估计量是否优良应依据以下三个特性：无偏性、有效性、一致性。

7.利用回归方程，预测期内自变量已知时，对因变量进行的预测通常分为（ ）。

- A. 点预测
- B. 区间预测
- C. 相对预测
- D. 绝对预测

【答案】AB

【解析】当预测期内的自变量已知时，对因变量的预测为无条件预测；如果在预测期内的自变量未知，对因变量预测为有条件预测。一般来说，无条件预测分为点预测与区间预测。

三、判断题

1.当数据呈现出明显的偏态分布时，中位数具有较强的代表性。()

【答案】正确

【解析】中位数是处于一组数据中心位置的数值，它作为一个位置代表值，不易受极端数值影响，当数据呈现出明显的偏态分布时具有较强的代表性。

2.当自由度大于 10 时，t 分布的概率密度函数的图形接近标准正态分布。()

【答案】错误

【解析】当自由度大于 30 时，t 分布的概率密度函数的图形接近标准正态分布。

3.箱线图不仅可以用来反映原始数据的分布特征，还可以用来比较多组数据分布特征的差异。()

【答案】正确

4.一个好的抽样方案选择的样本容量应该是越大越好。()

【答案】错误

【解析】样本容量过小，会使可靠程度和精确程度降低；而样本容量过大，则会带来调查费用增加、耗费人力物力等问题。所以样本容量的确定是十分重要的，一个好的抽样方案应该同时将推断效果和成本考虑在内

5.可决系数越接近于 1，线性回归模型的解释能力越弱。()

【答案】错误

【解析】可决系数越接近于 1，线性回归模型的解释能力越强。可决系数越接近于 0，线性回归模型的解释能力越弱。

6.若一个随机过程的均值和方差不随时间改变，且在任何两期之间的协方差仅依赖于两期的时点，则该随机过程称为平稳随机过程。()

【答案】错误

【解析】若一个随机过程的均值和方差不随时间的改变而改变，且在任意两期之间的协方差值仅依赖两期的距离或滞后长度而不依赖于两期的时点，这样的随机过程称为平稳随机过程；反之，称为非平稳随机过程。

6.ARMA 模型的特点在于可以利用时间序列自身的历史数据来预测未来，但需要依赖其他变量。（ ）

【答案】错误

【解析】ARMA 模型的特点在于可以利用时间序列自身的历史数据来预测未来，而不依赖其他变量。