

# 問 題

2021 年 8 月 2 日

Q1. 近年、データサイエンスが注目されている理由として、適切な記述を全て選べ。

1. コンピュータの演算性能が向上したこと
2. コンピュータの記憶容量が向上したこと
3. 大量のデータが入手可能になったこと
4. AI 技術が進歩したこと

Q2. 適切な記述を全て選べ。

1. アノテーションとは、データにメタデータを付加することである。
2. アノテーションには、大変な労力が必要になる場合がある。
3. オープンデータとは、誰もが簡単に二次利用できるデータのことである。
4. クリエイティブ・コモンズ・ライセンスとは、著作物の再利用を禁止するライセンスのことである。

Q3. 機械学習による予測の例として、適切な記述を全て選べ。

1. 桜の開花日を予測する。
2. 通販サイトの購入履歴データから、購入商品のパターンを発見する。
3. ある文書がある訴訟に関連するかどうかを判断する。
4. マグロの切断面の画像データから、品質等級を判断する。

Q4. 質的データについて、適切な記述を全て選べ。

1. 性別や所属部署は質的データである。
2. サイズ (S、M、L) は、名義尺度ではなく順序尺度である。
3. 性別 (男性、女性) は、順序尺度ではなく名義尺度である。
4. 等級 (1 級、2 級など) は質的データである。

Q5. データの可視化において、適切な記述を全て選べ。

1. 円グラフは、比率を比較するのに適している。
2. 折れ線グラフは、値の変化を表現するのに適している。
3. 値の変化が分かりにくい場合は、グラフの一部を拡大して変化を強調する。
4. 立体円グラフは誤解を招きやすく、用いるべきではない。

Q6. 適切な記述を全て選べ。

1. 累積度数とは、度数をその合計値で割ったものである。
2. 相対度数とは、累積度数を順に足し合わせたものである。
3. 行比率とは、クロス集計表における行ごとの相対度数のことである。
4. 連関係数は、クロス集計表の行と列の変数の関係性の強さを表す指標である。

Q7. 平均値について、適切な記述を全て選べ。

1. 平均値は代表値の一つである。
2. 極端な値を含まないデータであれば、データは概ね平均の周りに分布する。

3. 極端な値を含むデータの場合、平均が実質的な意味を持たない場合もある。
4. 極端な値を含むデータの場合、平均値よりも中央値の方がデータの性質をよく表す。

Q8. 相関関係と因果関係について、適切な記述を全て選べ。

1. 相関関係の有無はデータだけから分かるが、因果関係の有無を調べるのは難しい。
2. 変数 A から変数 B への因果関係があるならば、変数 A を介して、変数 B の値を制御することができる。
3. 変数 A と変数 B の間に相関関係があるならば、変数 A を介して、変数 B の値を制御することができる。
4. ある食品の摂取量が多い人は痩せているというデータ分析結果から、その食品にはダイエット効果があると主張することができる。

Q9. 相関係数について、適切な記述を全て選べ。

1. 相関係数は 0 以上 1 以下の値をとる。
2. 2 つの変数が正の相関を持つとき、相関係数は 1 に近づく。
3. 2 つの変数が負の相関を持つとき、相関係数は 0 に近づく。
4. 2 つの変数が無相関のとき、相関係数は 0.5 に近づく。

Q10. 適切な記述を全て選べ。

1. 確率変数とは、確率的に値が決まる変数のことである。
2. 確率分布とは、確率変数がそれぞれの値をとる確率を表す関数のことである。
3. 独立な確率変数を足し合わせていくと、その和が従う確率分布は正規分布に近づいていく。
4. 正規分布は連続値の確率分布である。

Q11. 適切な記述を全て選べ。

1. 母集団とは、分析対象として想定される全てのデータのことである。
2. 標本とは、データ収集によって得られたデータのことである。
3. 標本に含まれるデータの個数を標本数という。
4. 区間推定とは、母平均や母比率が含まれると考えられる範囲を、データから推測することである。

Q12. 母平均の 95% 信頼区間について、適切な記述を全て選べ。

1. データから推定した母平均の確率分布に基づき、母平均が 95% の確率で含まれる範囲のことである。
2. データ収集と信頼区間の計算を繰り返したとき、95% の確率で母平均が含まれる範囲のことである。
3. 母分散が分からない場合、正規分布を用いて計算する。
4. 母分散が分かる場合、t 分布を用いて計算する。

Q13. 帰無仮説について、適切な記述を全て選べ。

1. 帰無仮説とは、主張したい内容と同じ内容の仮説である。
2. p 値が有意水準を下回るとき、帰無仮説が棄却される。
3. 帰無仮説が棄却されたとき、対立仮説の内容を主張できる。
4. 帰無仮説が棄却されない場合、帰無仮説が正しいといえる。

Q14. 適切な記述を全て選べ。

1. p 値とは、帰無仮説を仮定した場合に、標本平均などの統計量がより極端な値をとる確率である。
2. 「母平均が 0 より大きい」という対立仮説の場合、片側検定を行う。
3. 「母平均が 0 とは異なる」という対立仮説の場合、両側検定を行う。
4. 母平均の検定では、p 値の計算に正規分布や t 分布を利用する。

Q15. 適切な記述を全て選べ。

1. 回帰とは、ある変数  $x$  と  $y$  の関係を回帰モデルと呼ばれる数式で表すことである。
2. 予測したい変数を目的変数といい、予測に使う変数を説明変数という。
3. 回帰のうち、目的変数が連続値ではない場合に、これを分類という。
4. 回帰分析とは、データから回帰モデルを決定し、目的変数の値を決める要因などを考察することである。

Q16. (選択問題: Excel) 分散を計算する関数として、適切なものを一つ選べ。

1. SUM
2. AVERAGE
3. VAR.P
4. STDEV.P

Q17. (選択問題: Excel) クロス集計表を作成する機能の名称として、適切なものを一つ選べ。

1. オートフィル
2. ピボットテーブル
3. ソート (並べ替え)
4. 絶対参照

Q18. (選択問題: Excel) オートフィルについて、適切な記述を全て選ぶ。

1. オートフィル機能によって、複数の行や列にセルの値をコピーすることができる。
2. オートフィルでは、セルの内容をコピーする際に、行の番号や列の記号を自動更新する機能がある。
3. 絶対参照を使うと、オートフィルで行の番号や列の記号が自動更新されるのを防ぐことができる。
4. 絶対参照を使うには、行の番号や列の記号の前に % 記号を付ければよい。

Q19. (選択問題: Excel) 最新版の Excel で作成できるグラフを全て選べ。

1. 棒グラフ
2. 円グラフ
3. ヒストグラム
4. 箱ひげ図

Q20. (選択問題: Excel) RAND 関数が生成する乱数が従う分布として、適切な分布を一つ選べ。

1. 一様分布
2. 二項分布

3. 正規分布
4. t 分布

Q21. (選択問題: Python) 構造化データの処理に最も適したライブラリを一つ選べ。

1. Numpy
2. Scipy
3. Pandas
4. Matplotlib

Q22. (選択問題: Python) Numpy ライブラリで平均値を計算する関数として、適切なものを一つ選べ。

1. sum
2. mean
3. average
4. max

Q23. (選択問題: Python) Python の文法について、適切な記述を一つ選べ。

1. {} 記号によってプログラムの構造を記述する。
2. () 記号によってプログラムの構造を記述する。
3. インデントによってプログラムの構造を記述する。
4. 文末には、; 記号が必要である。

Q24. (選択問題: Python) Numpy ライブラリから sum 関数だけを読み込みたい場合、どのように記述すればよいか。適切なものを一つ選べ。

1. import numpy
2. import numpy as np
3. from numpy import sum
4. import sum from numpy

Q25. (選択問題: Python) リストの作り方として適切なものを一つ選べ。

1. a = [1 2 3]
2. a = [1, 2, 3]
3. a = (1, 2, 3)
4. a = {1, 2, 3}