Фамилия, имя, группа	

ВИШ. Теория вероятностей и математическая статистика 2024/2025

Midterm exam, 27.03.2025

Пожалуйста, напишите свое имя, фамилию и номер группы на каждом листе, который содержит формулировки задач. В задачах 1- 4 вы должны написать **только ответы**. Вы должны дать ответ в виде числа (а не формулы) если вам явно не указано написать формулу. В задаче 5 вы должны написать **полное решение.**

Часть 1

(40% от общей оценки, время: 45 минут)

Задача 1

(a) Дана выборка с измерениями IQ у 20 сомов.

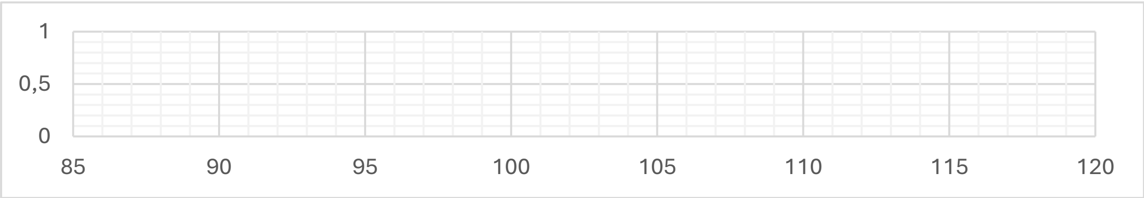
IQ Scores	90	95	100	105	110
	90	95	100	105	110
	90	95	100	105	110
	90	95	100	105	110

Посчитайте среднее значение, медиану, дисперсию. Ответы запишите через запятую.

Ответ:

(b) По данным из задания (a) изобразите график boxplot.

Ответ:



(c) Определите квантиль, который соответствует 72 перцентилю, $x_{0.72}$ — ?

Ответ:

Задача 2

(a) Дана выборка $X_1, X_2, \dots, X_n$, где каждая случайная величина X_i подчиняется распределению Пуассона с математическим ожиданием E[X]=10. Определите функцию плотности вероятности (pdf) первой порядковой статистики $X_{(1)}$.

Ответ:

(b) Из пункта (a) вычислите математическое ожидание и дисперсию.

Ответ:

Фамилия, имя, группа	

(c) Пусть случайные величины X и Y описывают характеристики двух станков на заводе. Совместная плотность вероятности этих величин задана функцией:

$$f(x,y) = c \cdot (x+y), \;\; 0 \le x \le 2, 0 \le y \; \le 3$$

Определите значение c , E[X] и E[Y].

Ответ:

Задача 3

(a) Руководство компании хочет изучить, как возможное сокращение рабочего времени (без уменьшения зарплаты) может повлиять на уровень удовлетворенности сотрудников. Для этого аналитики предложили случайным образом выбрать 600 сотрудников и разделить их на 3 группы: Сотрудники первой группы работают 4 дня в неделю вместо 5. Сотрудники второй группы работают 6 часов в день вместо 8. Сотрудники третьей группы работают в обычном режиме. На презентации проекта один из руководителей заметил, что должность сотрудников может значительно повлиять на результаты эксперимента и предложил более тщательно распределять участников по группам.

Используя только слова из описания эксперимента, укажите:

Ответ:

experimental units:

independent variable:

response variable:

confounded variable:

(b) Пусть T_1 и T_2 - две несмещенные оценки параметра θ, $Var(T_1) = Var(T_2) = 8$. Коэффициент корреляции между оценками равен ρ(T1, T2) = 0,25.Найдите среднеквадратичную ошибку (MSE) оценки $T = (T_1 + 3T_2)/4$?

Ответ:

(c) Пусть $X_1, X_2, \dots, X_n$ — случайная выборка из генеральной совокупности, где каждый X_i — это **время реакции** водителей (в секундах) при внезапном появлении препятствия. Предположим, что время реакции распределено равномерно: $X_i \sim U(a,b)$, где a и b — неизвестные параметры.

$$\hat{\delta} = max(X_1, X_2, \dots, X_n) \cdot \frac{n+1}{n}$$

Докажите, что данная оценка является несмещенной.

Ответ:

Фамилия, имя, группа	

Задача 4

Пусть $\hat{\theta}$ - оценка параметра θ. Докажите формулу разложения среднеквадратичной ошибки MSE($\hat{\theta}$) на смещение и дисперсию. При доказательстве вы должны дать определение среднеквадратичной ошибки, сформулировать формулу разложения и показать, что она справедлива.

Напишите свое решение здесь.

Часть 2

(60% от общей оценки, время: 45 минут)

Задача 5

Пусть $X_1, X_2, \dots, X_n$ — случайная выборка из генеральной совокупности, где каждый X_i — это индекс массы тела (BMI) случайного человека. Известно, что BMI в популяции распределён нормально:

$$X_i \sim N(\mu, \sigma^2)$$

где μ — истинное среднее значение BMI, а σ^2 — истинная дисперсия. Рассмотрим три различные точечные оценки для параметра μ.

Выборочное среднее	Взвешенное среднее с учётом крайних значений	Оценка усечённого среднего
$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{X_{max} - X_{min}}{10}$	$\hat{\mu}_3 = \frac{1}{n-2}\sum_{i=2}^{n-1} X_{(i)}$

(a) Определить, какие из этих оценок **несмещённые**.

(b) Определить, какая из них **эффективная** (имеет наименьшую дисперсию среди несмещённых оценок).

(c) Доказать, что если оценка $\hat{\mu}_3$ является **состоятельной**, то $\hat{\mu}_3 \to \mu$ по вероятности при $n \to \infty$.

(d) Рассмотреть случай, когда $\hat{\mu}_2$ является **асимптотически несмещённой**, и объяснить, чем это отличается от строгой несмещённости.

Фамилия, имя, группа	

Задача 6

В супермаркете работают три кассы. Время обслуживания одного покупателя каждым из кассиров подчиняется **показательному распределению** с плотностью:

$$f(t,\lambda) = \lambda e^{-\lambda t}, \quad t > 0$$

При этом:

Первый кассир **самый опытный**, его среднее время обслуживания **в два раза меньше**, чем у остальных двух. Это означает, что его интенсивность обслуживания λ_1 **в два раза выше**, чем у остальных. Второй и третий кассиры обслуживают покупателей с одинаковой интенсивностью $\lambda_2 = \lambda_3$. Экспериментально было зафиксировано, что: Первый кассир обслужил покупателя за **1 минуту**. Второй кассир обслужил покупателя за **2 минуты**. Третий кассир обслужил покупателя за **1.5 минуты**.

Определите:

(a) **Методом максимального правдоподобия (ММП) оценить параметр λ** — интенсивность обслуживания кассиров.

(b) Проверьте найденную оценку на несмещенность.

(c) Проверить эффективность оценки по границе Рао-Крамера.

(d) Проверить состоятельность оценки.