

Коллоквиум по статистике.

Перечень вопросов:

Раздел 1: Графическое представление данных

1. Что такое гистограмма и чем она отличается от столбчатой диаграммы?
2. Когда применяют boxplot и какие параметры он визуализирует?
3. Чем отличается график накопленной частоты (ogive) от гистограммы?
4. Назовите основные типы графического представления количественных данных. Приведите примеры.
5. Как определить наличие пропусков и выбросов на графике? Как они отражаются в гистограмме и boxplot?

Раздел 2: Описательная статистика

6. Определите и сравните среднее арифметическое, медиану и моду. В каких случаях предпочтительно использовать медиану?
7. Что такое квартиль, перцентиль и квартили размаха? Как они рассчитываются?
8. Что показывает межквартильный размах (IQR) и для чего используется нижняя и верхняя "границы" (fences)?
9. Объясните понятие стандартного отклонения. Почему в формуле выборочной дисперсии делят на $(n-1)$?
10. Что такое Z-оценка (z-score) и зачем она нужна? Как интерпретировать значение $z = 2.5$?

Раздел 3. Точечные оценки

11. Что такое точечная оценка параметра? В чем её отличие от интервальной оценки?
12. Какие параметры могут оцениваться точно? Приведите примеры точечных оценок для:
 - среднего значения (математического ожидания),
 - дисперсии,
 - доли,
 - стандартного отклонения.
13. Перечислите основные свойства хороших точечных оценок. Объясните:
 - несмещенность,
 - состоятельность,

- эффективность.

14. Почему выборочное среднее \bar{X} считается несмещенной оценкой математического ожидания μ ? Докажите это.

15. Является ли выборочная дисперсия

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

несмещенной оценкой генеральной дисперсии? Поясните, почему в формуле делится на $n-1$, а не на n .

16. Что означает эффективность оценки? В каких случаях одна оценка считается более эффективной, чем другая?

17. Приведите пример состоятельной оценки. Как проверить, что оценка является состоятельной?

18. Что такое оценка методом моментов и оценка методом максимального правдоподобия? В чем их различие?

19. Может ли одна и та же точечная оценка быть одновременно несмещенной, состоятельной и эффективной? Приведите пример или опровержение.

20. Почему в реальных задачах важно учитывать свойства оценок, а не только их числовое значение?

21. В чем суть метода максимального правдоподобия (ММП)? Какие шаги включает построение оценки?

22. Запишите общую формулу функции правдоподобия и логарифмической функции правдоподобия. Почему часто берут логарифм?

23. Что такое оценка максимального правдоподобия (MLE)? Какие преимущества и ограничения она имеет?

24. В чем состоит геометрический и статистический смысл информации Фишера? Как она рассчитывается?

25. Сформулируйте теорему Рао–Крамера. При каких условиях она применима и для чего используется?

26. Что означает "нижняя граница дисперсии несмещенной оценки" в теореме Рао–Крамера?

27. В каких случаях достигается равенство в теореме Рао–Крамера? Что это говорит об оценке?

- 28.** Что такое порядковые статистики? Приведите примеры (минимум, максимум, медиана, квартиль и др.).
- 29.** Объясните разницу между выборочной статистикой и порядковой статистикой. Приведите примеры.
- 30.** Приведите пример использования порядковой статистики для построения робастной оценки.

Раздел 4. Интервальные оценки

- 31.** Что такое интервальная оценка? Каков её смысл по сравнению с точечной оценкой?
- 32.** Что такое доверительный интервал? Как связаны уровень доверия и вероятность ошибки первого рода?
- 33.** Как построить доверительный интервал для математического ожидания μ , если:
- (а) дисперсия σ^2 известна;
 - (б) дисперсия неизвестна?
- 34.** Почему при неизвестной дисперсии для оценки среднего используют t-распределение, а не нормальное?
- 35.** Как ширина доверительного интервала зависит от:
- объема выборки,
 - уровня доверия,
 - стандартного отклонения?
- 36.** В чем отличия при построении доверительного интервала для доли (пропорции)? Какие оценки используются?
- 37.** Объясните идею интервальной оценки для дисперсии σ^2 . Какое распределение при этом используется?
- 38.** Почему для построения доверительного интервала для дисперсии необходима нормальность распределения?
- 39.** Что такое доверительный интервал для разности средних? Как построить его при равных и неравных дисперсиях?
- 40.** Как интерпретировать фразу: «с 95% уверенностью среднее значение лежит в интервале от 18 до 22»?
- 41.** Что такое парная выборка?

Раздел 5. Статистические гипотезы

- 41.** Что такое нулевая гипотеза H_0 и альтернативная гипотеза H_1 ? Приведите примеры для среднего, доли и дисперсии.
- 42.** Что такое ошибка первого рода? Как обозначается её вероятность и как она интерпретируется?
- 43.** Что такое ошибка второго рода? Чем она отличается от ошибки первого рода?
- 44.** Как связаны уровень значимости α , вероятность ошибки второго рода β и мощность критерия $1-\beta$?
- 45.** Что такое критическая область и как она определяется для односторонних и двусторонних критериев?
- 46.** Перечислите пошаговый алгоритм проверки статистической гипотезы по Z или t-критерию.
- 47.** Когда применяется Z-критерий для проверки гипотезы о среднем? Какие условия должны выполняться?
- 48.** Когда применяется t-критерий Стьюдента? Каковы его особенности при малых выборках?
- 49.** В чем отличие t-критерия для двух независимых выборок при равных и неравных дисперсиях?
- 50.** Как проверить гипотезу о равенстве средних двух зависимых (связанных) выборок?
- 51.** В каких случаях применим χ^2 -критерий? Приведите пример задачи на проверку дисперсии.
- 52.** Каковы особенности распределения χ^2 ? Как влияет увеличение степени свободы на его форму?
- 53.** В чем суть F-критерия Фишера? Как с его помощью проверить равенство двух дисперсий?
- 54.** Что такое ANOVA и как он связан с F-распределением? Какие гипотезы он проверяет?
- 55.** В каких случаях проверка гипотезы о дисперсии невозможна или требует модификации?
- 56.** Что такое p-value и как его интерпретировать в контексте проверки гипотез?
- 57.** Как влияет увеличение объема выборки на p-value и мощность критерия?
- 58.** Как выбрать правильный критерий для проверки гипотезы: Z, t, χ^2 или F?

59. Как проверить гипотезу о доле (пропорции)? Какую формулу и распределение при этом используют?

60. Что означает фраза: "Гипотеза не отвергается на уровне значимости $\alpha=0,05$ "? Можно ли сказать, что она подтверждена?

Раздел 6. Основные распределения в статистике

61. Что такое статистическое распределение? В чем разница между функцией плотности и функцией распределения?

62. Что такое нормальное распределение? Почему оно считается центральным в математической статистике?

63. Что означает стандартное нормальное распределение? Как его записывают и обозначают?

64. Что такое таблица значений функции Лапласа $\Phi(z)$ и как ею пользоваться?

65. В чем суть центральной предельной теоремы? Почему она позволяет использовать нормальное распределение для больших выборок?

t-распределение Стьюдента

66. Что такое t-распределение? В каких задачах оно применяется?

67. Как форма t-распределения зависит от числа степеней свободы? Что происходит при $n \rightarrow \infty$?

68. Почему при малых объемах выборки нельзя применять Z-критерий и нужно использовать t-критерий?

69. Как устроена таблица критических значений t-распределения? Что в ней отражено?

70. Приведите практический пример, где t-распределение используется для построения доверительного интервала.

Распределение χ^2 (хи-квадрат)

71. В каких задачах применяется распределение χ^2 ? Как оно связано с оценкой дисперсии?

72. Почему χ^2 распределение несимметрично? Как оно изменяется при росте числа степеней свободы?

73. Как используется χ^2 -распределение для построения доверительного интервала дисперсии?
74. Что означает значение χ^2 -статистики в критерии согласия? Как интерпретируется большое значение?
75. Какие предпосылки необходимо учитывать при применении χ^2 -критерия?

F-распределение Фишера

76. Что такое F-распределение? Как оно определяется через две χ^2 -случайные величины?
77. Для чего используется F-распределение в статистике? Приведите примеры задач.
78. В чем особенности формы F-распределения? Как оно зависит от двух степеней свободы?
79. Как использовать таблицу критических значений F-распределения?
80. В чем отличие применения F-распределения в двух задачах: (а) проверка равенства дисперсий, (б) дисперсионный анализ (ANOVA)?

Раздел 7. Общие вопросы и интеграция тем (вопросы 81–100)

Обобщение, логика и методология

81. В чем разница между описательной и выводной (инференциальной) статистикой?
82. Почему важно отличать выборку от генеральной совокупности? Как это влияет на методы анализа?
83. Что такое репрезентативность выборки? Как её обеспечить при сборе данных?
84. Что означает выражение «параметр генеральной совокупности»? Приведите примеры.
85. Почему доверительный интервал — это не вероятность того, что параметр "попал" в интервал?
86. Объясните связь между точечными и интервальными оценками одного и того же параметра.
87. Как связаны доверительный интервал и проверка гипотез? Когда можно сделать вывод об H_0 на основе интервала?
88. Когда выбор метода проверки гипотезы зависит от распределения, а когда — от объема выборки?

89. Какой критерий лучше использовать, если распределение неизвестно и объем выборки мал?

90. Почему важно проверять нормальность распределения перед использованием параметрических критериев?

Прикладные и интерпретационные вопросы

91. Как влияет наличие выбросов на:

- выборочное среднее,
- медиану,
- доверительный интервал?

92. В каких ситуациях лучше использовать медиану вместо среднего значения?

93. Почему нельзя использовать χ^2 -критерий для частот < 5 ? Какие существуют альтернативы?

94. Что такое чувствительность и специфичность критерия в контексте медицинской диагностики?

95. Чем отличается статистическая значимость от практической значимости? Приведите пример.

96. Как различаются $p\text{-value} = 0.049$ и $p\text{-value} = 0.051$? Можно ли проводить "жесткие" выводы?

97. Что такое множественное сравнение и почему оно повышает риск ошибки первого рода?

98. Как оценить надежность статистических выводов при небольшом размере выборки?

99. Как объяснить неспециалисту значение фразы: «Гипотеза не отвергнута на уровне значимости 0,05»?

100. Рустам или Руслан Ансарович?

101. Придумайте собственный пример статистического вывода (на основе оценки или гипотезы), связанного с повседневной жизнью.

103. Сколько всего комментариев под нашими видео в ютубе? А сколько лайков?

104. Мемы – это искусство? Как вам обучение статистики с мемами? Какой был самый смешной?

105. Вы рассказывали своей матери обо мне?