Урок 7. Статистика вывода (часть 1)

Хакимов Р.И. + ChatGPT

Оценка параметров: выборочное среднее и выборочная дисперсия

В статистике вывода оценка параметров является важным аспектом для обобщения результатов выборки на генеральную совокупность. Два ключевых параметра, которые часто оцениваются, это выборочное среднее и выборочная дисперсия.

Оценка параметров: выборочное среднее и выборочная дисперсия

Выборочное среднее (Sample Mean) является оценкой математического ожидания (среднего значения) генеральной совокупности на основе данных выборки.

Формула:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

где: $\bar{X}$ — выборочное среднее, n — размер выборки, X_i — значение i-го наблюдения в выборке.

Свойства:

Смещенность: Выборочное среднее является несмещенной оценкой истинного среднего значения генеральной совокупности.

Эффективность: С увеличением размера выборки выборочное среднее становится более точным.

Оценка параметров: выборочное среднее и выборочная дисперсия

Выборочное среднее (Sample Mean)

Пример: Если в выборке из 10 людей рост составляет: 160, 165, 170, 155, 180, 175, 160, 165, 170, 155 см, то выборочное среднее роста будет:

$$\bar{X} = \frac{160 + 165 + 170 + 155 + 180 + 175 + 160 + 165 + 170 + 155}{10} = 167$$

Оценка параметров: выборочное среднее и выборочная дисперсия

Выборочная дисперсия (Sample Variance) оценивает разброс данных вокруг выборочного среднего и является оценкой дисперсии генеральной совокупности.

Формула:

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

где: S^2 — выборочная дисперсия, $\bar{X}$ — выборочное среднее, n — размер выборки, X_i — значение i-го наблюдения в выборке.

Свойства:

Смещенность: Выборочная дисперсия является несмещенной оценкой дисперсии генеральной совокупности, благодаря делению на $n-1$ вместо n. Этот фактор корректирует смещение, которое возникает из-за оценки средней величины на основе выборки.

Эффективность: С увеличением размера выборки выборочная дисперсия становится более точной.

Оценка параметров: выборочное среднее и выборочная дисперсия

Выборочная дисперсия (Sample Variance)
Пример: Для той же выборки, что и в предыдущем примере (рост): 160, 165, 170, 155, 180, 175, 160, 165, 170, 155 см, выборочная дисперсия будет:

$$S^2 = \frac{1}{10-1} \times ((160-167)^2 + (165-167)^2 + (170-167)^2 + (155-167)^2 + (180-167)^2 +$$
$$(175-167)^2 + (160-167)^2 + (165-167)^2 + (170-167)^2 + (155-167)^2) = \frac{2604}{9} \approx 289.33$$

Оценка параметров: выборочное среднее и выборочная дисперсия

Выборочное стандартное отклонение (Sample Standard Deviation) - это квадратный корень из выборочной дисперсии и предоставляет меру разброса в тех же единицах, что и исходные данные.

$$S = \sqrt{S^2}$$

Оценка параметров: выборочное среднее и выборочная дисперсия

Надежность оценок: Чем больше размер выборки, тем надежнее (меньше ошибка) оценки параметров. Также важно учитывать уровень доверия и строить доверительные интервалы для более полного понимания параметров генеральной совокупности.

Эти оценки являются основой для проведения статистических тестов и построения моделей, которые помогают делать выводы о генеральной совокупности на основе выборки данных.

Введение в тестирование гипотез

Тестирование гипотез — это статистический метод, используемый для проверки предположений (гипотез) о параметрах генеральной совокупности на основе выборочных данных. Этот процесс помогает определить, насколько убедительными являются результаты исследования, и позволяет сделать выводы о популяции, исходя из данных выборки.

Основные концепции тестирования гипотез

Гипотеза: Предположение о параметре генеральной совокупности, которое проверяется на основе данных выборки.

Нулевая гипотеза (H_0): Основное предположение, которое мы тестируем. Обычно это гипотеза о том, что нет эффекта или различий. Например, "средний доход мужчин и женщин одинаков".

Альтернативная гипотеза (H_1 или H_a): Гипотеза, которая предполагает наличие эффекта или различий. Это гипотеза, которая рассматривается в случае, если нулевая гипотеза отвергнута. Например, "средний доход мужчин и женщин различен".

Основные концепции тестирования гипотез

В математической статистике уровень значимости (α) и уровень доверия ($1-\alpha$) связаны следующим образом:

Уровень значимости (α)**:** – это вероятность ошибки первого рода (α), то есть вероятность отклонения нулевой гипотезы, когда она на самом деле верна. Обычно выбирается значение 0.05 или 0.01.

Уровень доверия $(1-\alpha)$ – это вероятность того, что истинное значение параметра действительно попадает в доверительный интервал.
Таким образом, они дополняют друг друга до 1:

$$\text{Уровень доверия} = 1 - \text{Уровень значимости}$$

или

$$\text{Уровень значимости} = 1 - \text{Уровень доверия}$$

Основные концепции тестирования гипотез

Например, если уровень значимости $\alpha = 0.05$ (5%), то уровень доверия составляет $1 - 0.05 = 0.95$ (95%). Это означает, что с вероятностью 95% истинное значение параметра будет находиться в рассчитанном доверительном интервале.

Практическое значение:
- В проверке гипотез α определяет вероятность ложного отклонения нулевой гипотезы. Чем меньше α, тем более строгий критерий проверки.
- В доверительных интервалах $1 - \alpha$ показывает степень уверенности в том, что интервал охватывает истинное значение параметра.

Односторонний и двусторонний критерии

В математической статистике уровень значимости (α) связан с хвостами нормального распределения, так как он определяет области, где отклоняется нулевая гипотеза.

Односторонний критерий (правосторонний или левосторонний)

- Если тест правосторонний, то весь уровень значимости α находится в правом хвосте нормального распределения. Например, если $\alpha = 0.05$, то 5% площади распределения оказывается в правом хвосте.
- Если тест левосторонний, то уровень значимости α находится в левом хвосте.

Например, если уровень значимости $\alpha = 0.05$ (5%), то уровень доверия составляет $1 - 0.05 = 0.95$ (95%). Это означает, что с вероятностью 95% истинное значение параметра будет находиться в рассчитанном доверительном интервале.

Односторонний и двусторонний критерии

Двусторонний критерий

- В этом случае уровень значимости α делится поровну между двумя хвостами распределения. Например, при $\alpha = 0.05$ по 2.5% (0.025) площади оказывается в левом и правом хвостах.

Графическая интерпретация

- В правостороннем тесте α находится справа от критического значения Z_α.
- В левостороннем тесте α находится слева от критического значения Z_α.
- В двустороннем тесте область $\alpha/2$ расположена в обоих хвостах распределения.

Связь с уровнем доверия

Так как уровень доверия $1 - \alpha$ охватывает основную часть нормального распределения (обычно среднюю симметричную область), хвосты представляют исключительные, маловероятные значения.

Основные концепции тестирования гипотез

Ошибка первого рода (α): Ошибка, при которой нулевая гипотеза отвергается, хотя она верна.
Ошибка первого рода важна для понимания, так как её последствия могут быть серьезными, и она учитывается при планировании экспериментов и интерпретации их результатов.

Ошибка второго рода (β): Ошибка, при которой нулевая гипотеза не отвергается, хотя альтернативная гипотеза верна.
Ошибка второго рода особенно критична, когда важно не пропустить реальные эффекты или различия, например, в медицине, правосудии или научных исследованиях.

Ошибка первого рода. Примеры

1. Медицинское исследование

Сценарий: Проводится тест на наличие заболевания у человека. Нулевая гипотеза – пациент здоров.

Ошибка первого рода: Тест показывает, что пациент болен (положительный результат), хотя он здоров.

Последствия: Пациент может получить ненужное лечение, что может повлечь за собой побочные эффекты или затраты.

2. Научное исследование

Сценарий: В эксперименте проверяется гипотеза о наличии нового физического явления. Нулевая гипотеза – нового явления нет.

Ошибка первого рода: Ученые сообщают о наличии нового явления, хотя его нет, и различия в данных вызваны случайностью.

Последствия: Это может привести к ложным научным открытиям и потерям времени на дальнейшие исследования.

Ошибка первого рода. Примеры

3. Клинические испытания нового лекарства

Сценарий: Проводится исследование эффективности нового лекарства против заболевания. Нулевая гипотеза – лекарство не более эффективно, чем плацебо.

Ошибка первого рода: Исследование показывает, что лекарство эффективно, хотя на самом деле оно не лучше плацебо.

Последствия: Лекарство может выйти на рынок и применяться, несмотря на отсутствие реального эффекта.

4. Судебное разбирательство

Сценарий: В суде нулевая гипотеза – обвиняемый невиновен (презумпция невиновности).

Ошибка первого рода: Суд выносит обвинительный приговор (решает, что обвиняемый виновен), хотя он невиновен.

Последствия: Невиновный человек может быть осужден и понести наказание за преступление, которого он не совершал.

Ошибка второго рода. Примеры

1. Медицинское исследование

Сценарий: Проводится тест на наличие заболевания у человека. Нулевая гипотеза – пациент здоров.

Ошибка второго рода: Тест показывает, что пациент здоров, хотя на самом деле он болен.

Последствия: Это может привести к отсутствию своевременного лечения, что ухудшает состояние пациента и усложняет лечение в будущем.

2. Клинические испытания нового лекарства

Сценарий: Проводится исследование эффективности нового лекарства против заболевания. Нулевая гипотеза – лекарство не более эффективно, чем плацебо.

Ошибка второго рода: Исследование не находит доказательств эффективности лекарства, хотя на самом деле оно работает.

Последствия: Эффективное лекарство может быть отклонено, и оно не будет использовано для лечения пациентов, которые могли бы от него выиграть.

Ошибка второго рода. Примеры

3. Судебное разбирательство

Сценарий: В суде нулевая гипотеза – обвиняемый невиновен (презумпция невиновности).

Ошибка второго рода: Суд оправдывает обвиняемого (принимает его невиновность), хотя он на самом деле виновен.

Последствия: Виновный человек остается на свободе и может совершить новые преступления.

4. Эксперименты в маркетинге

Сценарий: Компания тестирует новую маркетинговую стратегию для увеличения продаж. Нулевая гипотеза — новая стратегия не улучшит продажи.

Ошибка второго рода: Эксперимент не показывает значительного роста продаж, хотя на самом деле стратегия эффективна.

Последствия: Компания может отказаться от успешной маркетинговой стратегии, теряя потенциальные доходы.

Доверительные интервалы

Доверительные интервалы в статистике вывода используются для оценки параметров генеральной совокупности на основе данных выборки и предоставляют диапазон, в котором с определённой вероятностью находится истинное значение параметра. Они позволяют учесть неопределённость, связанную с выборкой, и дают представление о надёжности оценок.

Доверительные интервалы

Основные понятия

Доверительный интервал (Confidence interval): Это интервал значений, в котором с определённой вероятностью (уровень доверия) находится истинное значение параметра генеральной совокупности.
Уровень доверия (Confidence Level): Это вероятность того, что доверительный интервал содержит истинное значение параметра. Обычно выбирается 90%, 95% или 99%.

Формулы для доверительных интервалов

Доверительный интервал для среднего значения (при известной дисперсии)
Если дисперсия генеральной совокупности известна и распределение данных нормально, доверительный интервал для среднего значения рассчитывается как:

$$\bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

где:
- $\bar{X}$ — выборочное среднее,
- $Z_{\alpha/2}$ — критическое значение Z для выбранного уровня доверия,
- σ — стандартное отклонение генеральной совокупности (или выборки, если это точная оценка),
- n — размер выборки.

Формулы для доверительных интервалов

Доверительный интервал для среднего значения (при неизвестной дисперсии)
Если дисперсия генеральной совокупности неизвестна и распределение данных нормально, используется распределение Стьюдента (t-распределение):

$$\bar{X} \pm t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}$$

где:
- $t_{\alpha/2,n-1}$ — критическое значение t для выбранного уровня доверия и $n-1$ степеней свободы,
- S — выборочное стандартное отклонение.

Формулы для доверительных интервалов

Доверительный интервал для дисперсии

Если необходимо оценить дисперсию генеральной совокупности, доверительный интервал для дисперсии рассчитывается как (используется χ^2-распределение):

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} \right)$$

где:
- S^2 — выборочная дисперсия,
- $\chi^2_{\alpha/2,n-1}$ и $\chi^2_{1-\alpha/2,n-1}$ — критические значения χ^2 для выбранного уровня доверия и $n-1$ степеней свободы.

Формулы для доверительных интервалов

Доверительный интервал для пропорции

Если оценка касается пропорции, доверительный интервал рассчитывается как:

$$\hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

где:

- $\hat{p}$ — выборочная пропорция,
- n — размер выборки,
- α — уровень значимости,
- $Z_{\alpha/2}$ — критическое значение Z для выбранного уровня доверия.

Как найти критическое значение $Z_{\alpha/2}$?

Критическое значение $Z_{\alpha/2}$ — это значение стандартного нормального распределения, которое отсекает $\alpha/2$ от хвостов распределения.

Для некоторых стандартных уровней доверия:
- Для 90% уровня доверия: $Z_{\alpha/2} \approx 1.645$;
- Для 95% уровня доверия: $Z_{\alpha/2} \approx 1.96$;
- Для 99% уровня доверия: $Z_{\alpha/2} \approx 2.576$.

Алгоритм поиска критического Z-значения в таблице

При проведении **двустороннего Z-теста** необходимо определить критическое значение $Z_{\alpha/2}$, используя таблицу Z-значений (Z-score table).

1. Определите уровень значимости (α)
Например, если $\alpha = 0.05$, то в двустороннем тесте по 2.5% уходит в каждый хвост (так как $\alpha/2 = 0.025$).

2. Найдите вероятность в центральной части таблицы
- В таблице Z-значений даны вероятности слева от Z.
- Нам нужно найти вероятность, соответствующую $1 - \alpha/2$.
- Например, если $\alpha = 0.05$, то $1 - 0.025 = 0.975$.

Алгоритм поиска критического Z-значения в таблице

3. Найдите ближайшее значение в таблице
- В таблице Z-значений ищем вероятность, ближайшую к 0.975.
- Это соответствует $Z \approx 1.96$.

4. Запишите критические значения
- Так как тест двусторонний, критические значения:

$$Z_{\alpha/2} = \pm 1.96$$

- Если расчет выполняется для другого α (например, 0.01), то нужно искать 0.995, что дает $Z \approx \pm 2.576$.

Как найти критическое значение $t_{\alpha/2,df}$?

Для **двустороннего t-теста** критическое значение $t_{\alpha/2,df}$ ищется в t-таблице (t-table).

Примеры для популярных уровней значимости:

df	$t_{0.05/2}$ (95% доверие)	$t_{0.01/2}$ (99% доверие)
10	2.228	3.169
20	2.086	2.845
30	2.042	2.750

Алгоритм поиска критического t-значения в таблице

1. Определите уровень значимости (α)

Например, если $\alpha = 0.05$, то в двустороннем тесте по 2.5% уходит в каждый хвост (так как $\alpha/2 = 0.025$).

2. Определите число степеней свободы df

- Для одновыборочного t-теста:

$$df = n - 1$$

- Для двухвыборочного t-теста (независимые выборки, равные дисперсии):

$$df = n_1 + n_2 - 2$$

- Для двухвыборочного t-теста (неравные дисперсии, Welch's test):

$$df \approx \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

где s_1^2 и s_2^2 — дисперсии выборок.

Алгоритм поиска критического t-значения в таблице

3. Ищем в t-таблице:

- Находим число степеней свободы df в левом столбце.
- Находим уровень значимости $\alpha/2$ в верхней строке.
- На пересечении — критическое значение $t_{\alpha/2,df}$.