**Комп'ютерний практикум 3**

**Тема: Перевірка відповідності теоретичного закону розподілу даних експерименту**

**Мета роботи**

Ознайомитися з поняттям гіпотез, навчитися побудові теоретичного закону розподілу за даними експерименту та перевірці відповідності теоретичного закону розподілу даних експерименту.

**Основні теоретичні викладки**

**3.1. Побудова теоретичного закону розподілу за даними експерименту**

Функція нормального розподілу має вигляд: y(x)=1σ2πe−(x−a)22σ2y(x)=σ2π

​1​e−2σ2(x−a)2​ де aa — математичне очікування, σσ — середнє квадратичне відхилення випадкової величини.

На практиці вводиться величина: λ=x−aσλ=σx−a​

З урахуванням ширини кожного з інтервалу hh, рівняння набуде вигляду: y(λ)=h⋅φ(λ)σy(λ)=σh⋅φ(λ)​ де φ(λ)φ(λ) — щільність нормованого нормального закону розподілу з нульовим математичним очікуванням і σ=1σ=1. Значення φ(λ)φ(λ) табульовані.

**Завдання**

1. Побудувати теоретичну криву розподілу за даними експерименту.
2. Перевірити відповідність теоретичної кривої розподілу даним експерименту для рівня значущості α=0.05α=0.05.

**Приклад розрахунків та побудова кривої**

1. **Збір даних вибірки:**
   * Використовуємо дані з Комп'ютерного практикуму 1.
   * Розраховуємо середнє значення (XˉXˉ) та виправлене середнє квадратичне відхилення (SS).
2. **Розрахунки параметра λλ для кожного інтервалу:**
   * Знаходимо середини інтервалів XˉiXˉi​.
   * Обчислюємо λλ для кожного XˉiXˉi​.
3. **Таблиця значень для розрахунків:**

| **Середина інтервалу XˉiXˉi​** | **λλ** | **φ(λ)φ(λ)** | **y(λ)y(λ)** | **Відносні частоти** | **Абсолютні частоти nемпnемп​** | **Теоретичні частоти nтеор′nтеор′​** |
| --- | --- | --- | --- | --- | --- | --- |

1. **Обчислення теоретичних частот:**
   * За допомогою φ(λ)φ(λ) розраховуємо y(λ)y(λ).
   * Обчислюємо теоретичні частоти nтеор′nтеор′​.
2. **Побудова гістограми:**
   * На основі отриманих значень побудувати гістограму та нанести теоретичні частоти.

**Перевірка відповідності теоретичного закону розподілу даним експерименту**

1. **Використання критерію Пірсона χ2χ2:**

Формула критерію: χ2=∑(nемп−nтеор′)2nтеор′χ2=∑nтеор′​(nемп​−nтеор′​)2​

1. **Знаходження критичного значення χ2χ2:**
   * За таблицею критичних точок розподілу χ2χ2 для рівня значущості α=0.05α=0.05 та відповідної кількості ступенів свободи.
2. **Перевірка гіпотези:**
   * Якщо χспост2<χкр2χспост2​<χкр2​, то приймаємо гіпотезу про відповідність даних експерименту нормальному закону розподілу.

////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

1. **Побудова теоретичної кривої розподілу**:
   * Спочатку потрібно об'єднати всі дані з вибірки в один масив.
   * Після цього використовуються методи функції fit() з бібліотеки scipy.stats для знаходження параметрів нормального розподілу, які найкраще підходять для даних.
   * Побудуйте графік гістограми цих даних разом з теоретичною кривою розподілу.
2. **Перевірка відповідності теоретичної кривої розподілу даним експерименту**:
   * Використовуйте метод kstest() з бібліотеки scipy.stats для проведення тесту Колмогорова-Смірнова.
   * Порівняйте p-значення, отримане з тесту, з заданим рівнем значущості (зазвичай 0.05).
   * Якщо p-значення менше заданого рівня значущості, то ми відхиляємо нульову гіпотезу про відповідність теоретичної кривої розподілу даним експерименту. В іншому випадку, ми приймаємо нульову гіпотезу.

Це базовий опис того, як виконати це завдання без коду. Якщо у вас виникають питання з конкретними кроками або поняттями, будь ласка, запитуйте!

Побудова теоретичного закону розподілу за даними експерименту полягає у тому, що ми намагаємося знайти математичну функцію, яка найкращим чином описує розподіл даних, щоб мати можливість проводити подальший аналіз.

Перевірка відповідності теоретичного закону розподілу даним експерименту - це перевірка того, наскільки добре теоретична модель (наприклад, нормальний розподіл) відповідає фактичним даним. Це може бути зроблено за допомогою різних статистичних методів, таких як критерій Пірсона або критерій Колмогорова-Смірнова.

 **Що таке теоретичний закон розподілу?** Теоретичний закон розподілу - це математична функція, яка описує розподіл випадкової величини. Ця функція визначає ймовірність того, що випадкова величина прийме певне значення або потрапить в певний інтервал.

 **Які поняття пов'язані з гіпотезами в статистиці?** Поняття, пов'язані з гіпотезами в статистиці, включають нульову гіпотезу (H0) та альтернативну гіпотезу (H1). Нульова гіпотеза зазвичай формулюється як ствердження про відсутність ефекту або відсутність зв'язку між явищами, тоді як альтернативна гіпотеза формулюється як протилежне ствердження.

 **Які кроки потрібно виконати для побудови теоретичного закону розподілу за даними експерименту?** Для побудови теоретичного закону розподілу за даними експерименту спочатку необхідно вибрати відповідний закон розподілу, який краще всього пасує до даних. Потім виконують параметризацію цього розподілу, щоб оцінити його параметри за допомогою методів максимальної правдоподібності або методів моментів.

 **Що таке перевірка відповідності теоретичного закону розподілу даним експерименту?** Перевірка відповідності теоретичного закону розподілу даним експерименту - це процес, під час якого перевіряється, наскільки добре теоретичний закон розподілу відображає фактичний розподіл даних експерименту.

 **Які методи можуть бути використані для перевірки відповідності теоретичного закону розподілу даним експерименту?** До таких методів можна віднести критерії відповідності, такі як критерій Колмогорова-Смірнова, критерій Пірсона і інші.

 **Як визначається рівень значущості α при перевірці гіпотези?** Рівень значущості α при перевірці гіпотези - це ймовірність відхилення нульової гіпотези, коли вона насправді є правильною. Зазвичай встановлюється на рівні 0,05 або 0,01, що вказує на ймовірність відхилення нульової гіпотези на 5% або 1% відповідно.

 **Що таке p-значення і як його використовують для прийняття рішення про відхилення гіпотези?** p-значення - це ймовірність отримати спостережувані дані, якщо нульова гіпотеза справді є правильною. Чим менше значення p, тим більша ймовірність того, що ми відхиляємо нульову гіпотезу.

 **Які висновки можна зробити з результатів перевірки відповідності теоретичного закону розподілу даним експерименту?** З результатів перевірки можна зробити висновок про те, наскільки добре теоретичний закон розподілу відповідає даним експерименту. Якщо p-значення менше рівня значущості, можна відхилити нульову гіпотезу про відповідність теоретичного закону даним.

**Що таке критерій Колмогорова-Смірнова?** Критерій Колмогорова-Смірнова - це статистичний тест, який використовується для порівняння емпіричної функції розподілу (ЕФР) з теоретичною функцією розподілу. Він вимірює максимальну відстань між емпіричною та теоретичною функціями розподілу та допомагає визначити, наскільки добре теоретичний закон розподілу відповідає даним.

**ЛАБА 4**

**Які основні кроки потрібно виконати під час вирішення завдань лабораторної роботи?**

1. Основні кроки під час вирішення завдань лабораторної роботи включають аналіз даних, побудову теоретичного закону розподілу, перевірку відповідності теоретичного закону даним експерименту та формулювання висновків на основі отриманих результатів.

**Що таке критерій Пірсона, і як він застосовується для перевірки гіпотез про розподіл?**

1. Критерій Пірсона - це статистичний тест, який використовується для перевірки гіпотез про розподіл випадкових величин. Він порівнює фактичні та теоретично очікувані частоти в кількісних даних.

**Як визначається рівень значущості α, і чому він важливий для статистичних висновків?**

1. Рівень значущості α - це ймовірність відкидання нульової гіпотези, коли вона насправді є правильною. Це важливий показник, який допомагає визначити, наскільки впевненими можуть бути статистичні висновки.

**Що означає поняття "перевірка гіпотези", і які кроки потрібно виконати під час цього процесу?**

1. Перевірка гіпотези - це процес визначення, чи можуть отримані дані випадково виникнути відповідно до певної гіпотетичної моделі чи розподілу. Кроки включають формулювання нульової та альтернативної гіпотез, обчислення тестової статистики, визначення рівня значущості та прийняття рішення на її основі.

**Які параметри потрібно враховувати при використанні критерію Колмогорова для перевірки гіпотез про розподіл?**

1. При використанні критерію Колмогорова для перевірки гіпотез про розподіл, потрібно враховувати емпіричну функцію розподілу та теоретичну функцію розподілу, а також визначити рівень значущості α.

**Які кроки потрібно виконати для побудови теоретичного закону розподілу за даними експерименту?**

1. Для побудови теоретичного закону розподілу за даними експерименту потрібно обрати відповідний статистичний розподіл, підібрати параметри цього розподілу, побудувати теоретичну функцію розподілу та порівняти її з емпіричною функцією розподілу даних.

**Що таке нульова гіпотеза в контексті лабораторної роботи про перевірку гіпотез про розподіл?**

1. Нульова гіпотеза - це ствердження, яке припускається як справжнє для статистичного тестування. В контексті лабораторної роботи про перевірку гіпотез про розподіл, це може бути ствердження про те, що дані відповідають певному теоретичному закону розподілу.

**Які практичні ситуації можуть вимагати використання критерію Колмогорова для перевірки гіпотез про розподіл?**

1. Практичні ситуації, що вимагають використання критерію Колмогорова для перевірки гіпотез про розподіл, можуть включати оцінку якості випадкових вибірок, перевірку адекватності моделей, порівняння емпіричного розподілу з теоретичним.

**Як визначається p-значення, і як його використовують для прийняття рішення про відхилення гіпотези?**

1. P-значення - це ймовірність отримати спостережувані дані, якщо нульова гіпотеза справді є правильною. Використовується для визначення статистичної значущості результатів тестування гіпотези.

**Як перевірити відповідність теоретичного закону розподілу даним експерименту?**

1. Перевірка відповідності теоретичного закону розподілу даним експерименту включає порівняння теоретичної функції розподілу з емпіричною функцією розподілу даних та визначення статистичної значущості цього порівняння.

**Що таке емпірична функція розподілу, і як вона використовується для аналізу даних?**

1. Емпірична функція розподілу - це статистична характеристика, яка відображає частку спостережень, які менше або рівні певного значення випадкової величини.

**Як визначається критичне значення для рівня значущості α при перевірці гіпотези?**

1. Критичне значення для рівня значущості α визначається з таблиць критичних значень або з використанням статистичних програм.

**Які висновки можна зробити з результатів перевірки відповідності теоретичного закону розподілу даним експерименту?**

1. З результатів перевірки відповідності теоретичного закону розподілу даним експерименту можна зробити висновок про те, наскільки добре теоретичний закон розподілу відповідає емпіричним даним, а також прийняти рішення про прийняття чи відхилення нульової гіпотези.

**Що таке статистичний тест, і як він використовується для перевірки гіпотез?**

1. Статистичний тест - це процедура, за допомогою якої виконується перевірка гіпотез про параметри популяції або розподіл випадкових величин на основі спостережуваних даних.

**Які методи можуть бути використані для перевірки відповідності теоретичного закону розподілу даним експерименту?**

1. Для перевірки відповідності теоретичного закону розподілу даним експерименту можна використовувати різні статистичні методи та критерії. Основні методи включають:
2. **Критерій Пірсона (χ²-критерій)**: Цей метод використовується для перевірки відповідності між спостережуваними і сподіваними частотами в категоріях або класах.
3. **Критерій Колмогорова-Смірнова**: Вже обговорюваний критерій, який порівнює емпіричну функцію розподілу з теоретичною.
4. **Критерій Андерсона-Дарлінга**: Цей критерій також використовується для перевірки відповідності між емпіричним і теоретичним розподілами.
5. **Критерій Колмогорова-Лілліфорса**: Цей метод адаптований для випадку, коли параметри теоретичного закону розподілу оцінюються з емпіричних даних.
6. **Інші критерії адекватності моделі**: До цієї категорії входять критерії, які використовуються для порівняння різних моделей розподілу, такі як критерій Акаіке (AIC) або критерій Байєса-Шварца (BIC).

Ці методи дозволяють статистично оцінити відповідність між спостережуваними даними та теоретичним законом розподілу.

1. **Критерій Пірсона** - це статистичний критерій, що використовується для перевірки відповідності емпіричного розподілу даних теоретичному закону розподілу. Він базується на порівнянні фактичних та очікуваних частот спостережень у різних класах або інтервалах.
2. **Критерій Колмогорова**, також відомий як критерій Колмогорова-Смірнова, є статистичним методом для оцінки відповідності між теоретичним розподілом і емпіричним розподілом даних. Він базується на порівнянні функції розподілу кумулятивних частот емпіричної вибірки з функцією розподілу теоретичного закону. Цей критерій використовується для перевірки гіпотези про те, що спостережені дані взяті з певного теоретичного розподілу. Чим більше значення критерію, тим менша відповідність між емпіричним і теоретичним розподілами.

**ЛАБА 5**

**Що таке точкові незміщені статистичні оцінки?**  
Точкові незміщені статистичні оцінки - це оцінки параметрів, які в середньому дорівнюють реальним значенням параметрів у популяції, тобто вони не похиляються від справжніх значень.

**Що таке для параметрів незміщені статистичні оцінки для парної лінійної функції регресії?**  
Для параметрів незміщені статистичні оцінки для парної лінійної функції регресії - це оцінки параметрів регресії, які в середньому дорівнюють їх справжнім значенням у популяції.

**Що таке довірчі інтервали?**  
Довірчий інтервал - це інтервал числових значень, що обмежує теоретично можливі значення параметра з певною ймовірністю (зазвичай вираженою у відсотках).

**Що таке рівень значущості?**  
Рівень значущості - це ймовірність відхилення нульової гіпотези про параметри моделі в разі, якщо вона справді є правильною.

**Як перевірити значущість параметра?**  
Значущість параметра можна перевірити, обчисливши його статистичну значущість або за допомогою статистичних тестів, таких як t-критерій.

**Що таке довірчий інтервал для функції регресії?**  
Довірчий інтервал для функції регресії - це інтервал, що вказує на можливі значення прогнозованих значень відгуку з певною ймовірністю.

**Що таке вибірковий коефіцієнт кореляції?**  
Вибірковий коефіцієнт кореляції - це статистика, яка вимірює силу та напрямок лінійного зв'язку між двома змінними вибірки.

**Що таке парна лінійна регресія?**  
Парна лінійна регресія - це метод аналізу залежності між двома змінними, де одна змінна (незалежна) використовується для прогнозування значень іншої змінної (залежної) за допомогою лінійної функції.

**Які параметри потрібно оцінювати в парній лінійній регресії?**  
У парній лінійній регресії потрібно оцінювати параметри β0β0​ (перетин з віссю yy) та β1β1​ (коефіцієнт нахилу прямої).

**Що таке точкові незміщені статистичні оцінки в парній лінійній регресії?**  
Точкові незміщені статистичні оцінки в парній лінійній регресії - це оцінки параметрів регресії, які в середньому дорівнюють реальним значенням параметрів у популяції.

**Як побудувати довірчі інтервали для параметрів регресії з надійністю 0,99?**  
Довірчі інтервали для параметрів регресії можна побудувати, використовуючи табличні значення квантилів розподілу Стьюдента та стандартні помилки оцінок параметрів.

**Як перевірити значущість параметрів регресії на певному рівні значущості?**  
Значущість параметрів регресії можна перевірити за допомогою t-критерію Стьюдента для кожного параметра.

**Як побудувати довірчий інтервал для функції регресії з певною надійністю?**  
Довірчий інтервал для функції регресії можна побудувати, використовуючи стандартні помилки оцінок параметрів регресії та квантилі розподілу Стьюдента.

**Як обчислити вибірковий коефіцієнт кореляції у парній лінійній регресії?**  
Вибірковий коефіцієнт кореляції rr обчислюється як коефіцієнт кореляції між двома змінними у вибірці.

**Які висновки можна зробити зі значень коефіцієнта кореляції?**  
Значення коефіцієнта кореляції rr показує ступінь лінійної залежності між змінними: близьке до 1 - сильна позитивна залежність, близьке до -1 - сильна негативна залежність, близьке до 0 - відсутність лінійної залежності.

**Які методи можна використовувати для вибору моделі регресії?**  
Для вибору моделі регресії можна використовувати критерії інформаційної наступності, такі як критерій Акаіке чи критерій Шварца.

**Як визначити адекватність моделі регресії?**  
Адекватність моделі регресії можна визначити за допомогою F-критерію Фішера або за допомогою аналізу залишкових діаграм.

**Які можуть бути причини незначущості параметрів регресії?**  
Причини незначущості параметрів регресії можуть бути у недостатньому об'ємі вибірки, неправильному співвідношенні між змінними або неправильному виборі функціональної форми моделі.

**ЛАБА 6**

**Що таке множинна лінійна регресія?**  
Множинна лінійна регресія - це метод статистичного аналізу, який використовується для моделювання зв'язку між однією залежною змінною і кількома незалежними змінними.

**2. Як визначаються статистичні оцінки параметрів у моделі множинної лінійної регресії?**  
Статистичні оцінки параметрів у моделі множинної лінійної регресії визначаються методом найменших квадратів, який мінімізує суму квадратів відхилень спостережуваних значень від прогнозованих.

**3. Що таке коефіцієнт множинної кореляції RR?**  
Коефіцієнт множинної кореляції RR вимірює силу лінійного зв'язку між залежною змінною та набором незалежних змінних. Він приймає значення від 0 до 1, де 1 вказує на ідеальну відповідність.

**4. Як інтерпретувати коефіцієнт множинної кореляції?**  
Коефіцієнт множинної кореляції RR інтерпретується як міра ступеня зв'язку між змінними. Чим ближче значення RR до 1, тим сильніший зв'язок між незалежними змінними та залежною змінною.

**5. Що таке довірчий інтервал для функції регресії і як його побудувати?**  
Довірчий інтервал для функції регресії - це інтервал, який з певною ймовірністю (наприклад, 99%) включає в себе значення залежної змінної для даних значень незалежних змінних. Він будується на основі стандартної похибки регресії та критичних значень t-розподілу.

**6. Які припущення робляться при застосуванні множинної лінійної регресії?**  
Припущення включають лінійність зв'язку між залежною та незалежними змінними, нормальність розподілу залишків, відсутність автокореляції залишків та гомоскедастичність залишків.

**7. Як перевірити відповідність даних припущенням лінійної регресії?**  
Відповідність даних припущенням перевіряється за допомогою діагностичних графіків (наприклад, графіків залишків), статистичних тестів (наприклад, тесту Дарбіна-Уотсона для автокореляції) та аналізу залишків.

**8. Що таке мультиколінеарність і як вона впливає на модель множинної регресії?**  
Мультиколінеарність - це сильний кореляційний зв'язок між незалежними змінними в моделі, який може призвести до нестабільності оцінок коефіцієнтів та збільшення їх стандартних помилок.

**9. Як можна виявити і вирішити проблему мультиколінеарності в моделі регресії?**  
Мультиколінеарність можна виявити за допомогою аналізу матриці кореляцій незалежних змінних або розрахунку факторів інфляції дисперсії (VIF). Для вирішення проблеми можна видалити або об'єднати сильно корельовані змінні, або використовувати методи регресії, стійкі до мультиколінеарності.

**10. Що таке залишки в моделі регресії і як вони використовуються для оцінки якості моделі?**  
Залишки - це різниця між спостережуваними значеннями залежної змінної та значеннями, прогнозованими моделлю. Аналіз залишків допомагає оцінити відповідність моделі припущенням регресії та ідентифікувати можливі проблеми, такі як автокореляція або гетероскедастичність.