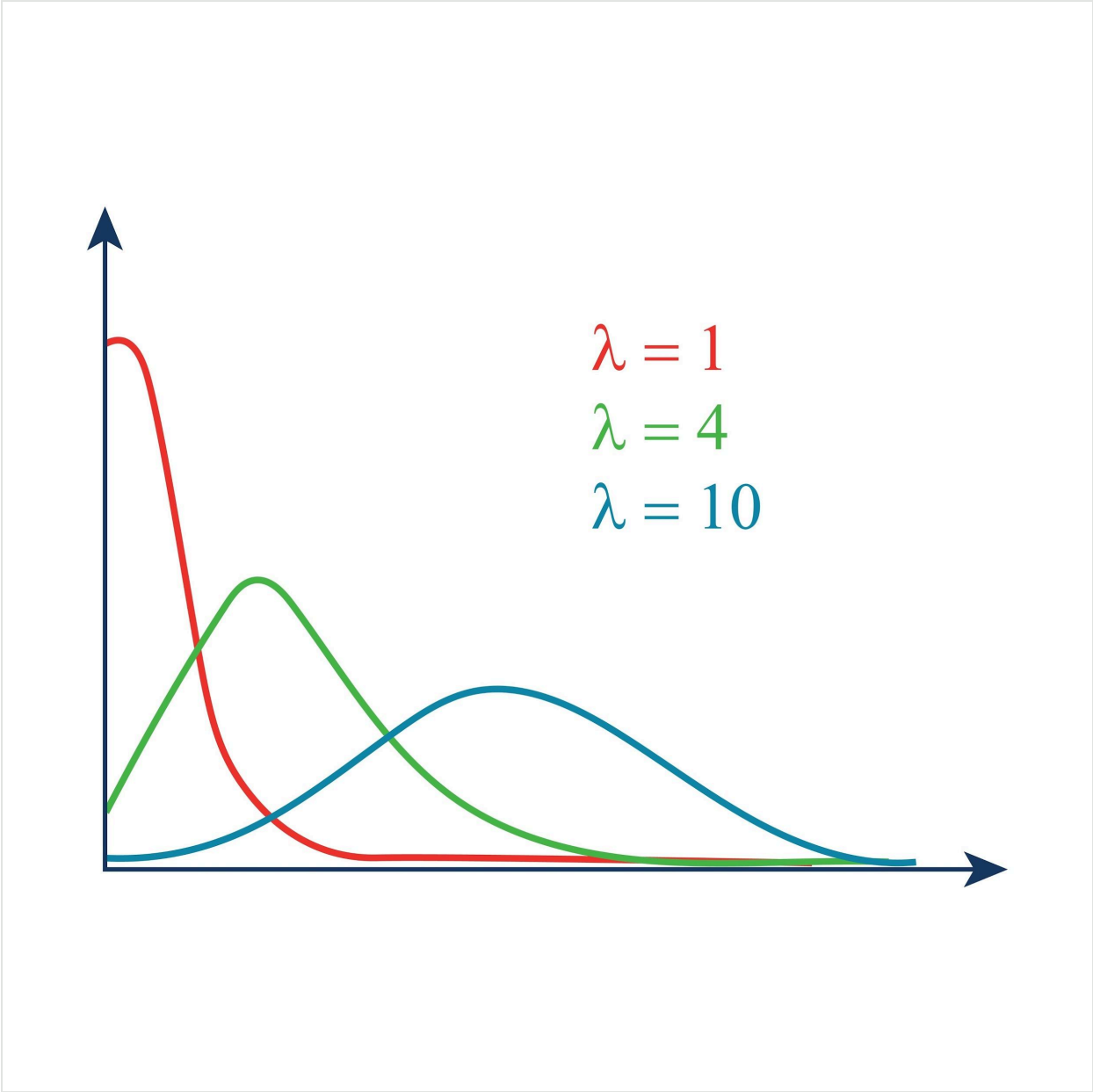
Chi-Square Test (চি-স্কয়ার টেস্ট)



Shutterstock

Explore

- **Basic Concept (প্রাথমিক ধারণা):** * Chi-Square Test (যাকে X^2 টেস্টও বলা হয়) হলো একটি পরিসংখ্যানগত পদ্ধতি (statistical method), যা মূলত **Categorical Data** (যেমন: হ্যাঁ/না, লিঙ্গ, রঙের ধরন) বিশ্লেষণ করতে ব্যবহৃত হয়।
    - এটি একটি Non-parametric test, অর্থাৎ এটি ডেটার নরমাল ডিস্ট্রিবিউশন (Normal Distribution) মেনে চলার ওপর নির্ভর করে না।
    - এর প্রধান উদ্দেশ্য হলো, আমাদের প্রাপ্ত বা পর্যবেক্ষণ করা ডেটা (Observed data) এবং আমরা যা আশা করেছিলাম (Expected data)- এই দুটির মধ্যে কোনো উল্লেখযোগ্য পার্থক্য (significant difference) আছে কি না তা যাচাই করা।
- **Types of Chi-Square Test (চি-স্কয়ার টেস্টের প্রকারভেদ):** এটি মূলত দুটি কাজের জন্য সবচেয়ে বেশি ব্যবহৃত হয়:
  ১) Goodness of Fit Test
  ২) Test of Independence (দুটি ক্যাটাগরিকাল ভেরিয়েবলের মধ্যে সম্পর্ক আছে কি না তা যাচাই করতে)।

Chi-Square Goodness of Fit Test (চি-স্কয়ার গুডনেস অফ ফিট টেস্ট)

- **Definition (সংজ্ঞা):** * এটি এমন একটি টেস্ট যা যাচাই করে যে, একটি স্যাম্পলের ডেটা কোনো নির্দিষ্ট ডিস্ট্রিবিউশন (যেমন: Uniform, Binomial বা Normal distribution) ঠিকঠাক মতো মেনে চলছে কি না।
    - সহজ কথায়, আমাদের Observed Values (বাস্তবে আমরা যা পেয়েছি) এবং Expected Values (তত্ত্ব অনুযায়ী আমাদের যা পাওয়ার কথা ছিল)- এর মধ্যে কতটা মিল বা "Fit" আছে, তা এই টেস্টের মাধ্যমে পরিমাপ করা হয়।
- **Mechanism (কাজের প্রক্রিয়া):**
    - **Null Hypothesis (H0):** Observed এবং Expected frequencies-এর মধ্যে কোনো উল্লেখযোগ্য পার্থক্য নেই (অর্থাৎ, ডেটাটি নির্দিষ্ট ডিস্ট্রিবিউশনটি মেনে চলছে বা "Good Fit")।
    - **Alternate Hypothesis (H1):** Observed এবং Expected frequencies-এর মধ্যে উল্লেখযোগ্য পার্থক্য আছে (অর্থাৎ, ডেটাটি ওই ডিস্ট্রিবিউশন মেনে চলছে না)।
- **The Formula (সূত্র):**
    - Chi-Square = Sum of [(Observed - Expected)^2 / Expected]

    - এখানে প্রতিটি ক্যাটাগরির জন্য Observed (O) থেকে Expected (E) বিয়োগ করে তার বর্গ (square) করা হয় এবং তাকে Expected (E) দিয়ে ভাগ করা হয়। এরপর সবগুলোর যোগফলই হলো Chi-Square মান।
- **Degree of Freedom (dof):**
    - এর সূত্র হলো: dof = k - 1 (যেখানে k হলো মোট ক্যাটাগরি বা গ্রুপের সংখ্যা)।
- **Real World Example (বাস্তব উদাহরণ):**
    - ধরা যাক, একটি কোম্পানির দাবি অনুযায়ী তাদের উৎপাদিত চকলেটের প্যাকেটে লাল, নীল এবং সবুজ চকলেট সমান অনুপাতে (Uniform distribution) থাকে।
    - আপনি একটি প্যাকেট খুলে দেখলেন সেখানে লাল চকলেট ২০টি, নীল ১০টি এবং সবুজ ৩০টি (এগুলো হলো Observed value)। কিন্তু কোম্পানির দাবি অনুযায়ী সবগুলো ২০টি করে থাকার কথা ছিল (এগুলো হলো Expected value)।
    - এখন Chi-Square Goodness of Fit Test ব্যবহার করে আপনি প্রমাণ করতে পারবেন যে, কোম্পানির দাবিটি পরিসংখ্যানগতভাবে (statistically) সত্য কি না।