

TRƯỜNG ĐẠI HỌC MỎ - ĐỊA CHẤT
KHOA TRẮC ĐỊA – BẢN ĐỒ VÀ QUẢN LÝ ĐẤT ĐAI



BÁO CÁO HỌC THUẬT
HỌC KỲ 2 NĂM HỌC 2024-2025

MỘT SỐ VẤN ĐỀ VỀ ỨNG DỤNG XÁC SUẤT THỐNG KÊ
TRONG TRẮC ĐỊA

Người thực hiện: Nguyễn Quang Minh

Bộ môn: Trắc địa phổ thông và Sai số, Khoa Trắc địa – Bản đồ và Quản lý đất đai

Hà nội, 2025

MỤC LỤC

MỤC LỤC.....	1
DANH MỤC BẢNG BIỂU	3
DANH MỤC HÌNH VẼ.....	4
PHẦN 1 NỀN TẢNG CỦA SAI SỐ VÀ THỐNG KÊ TRONG KHOA HỌC ĐO ĐẠC.....	5
1.1 Bản chất của phép đo và sự tồn tại tất yếu của sai số	5
1.2 Phân loại Sai số trong Trắc địa	6
1.3 Vai trò của Xác suất Thống kê	6
PHẦN 2 CÁC KHÁI NIỆM THỐNG KÊ NỀN TẢNG TRONG XỬ LÝ SỐ LIỆU TRẮC ĐỊA ..	7
2.1 Tổng thể và Mẫu (Population and Sample)	7
2.2 Các tham số đặc trưng quan trọng	7
2.3 Luật Phân bố Chuẩn (Normal Distribution)	9
PHẦN 3 ỨNG DỤNG PHÂN BỐ STUDENT (T-DISTRIBUTION) ĐỂ ƯỚC LƯỢNG KHOẢNG TIN CẬY CHO TRỊ TRUNG BÌNH	11
3.1 Giới thiệu Phân bố Student	11
3.2 Xây dựng Khoảng tin cậy (Confidence Interval)	11
3.3 Quy trình thực hiện và ví dụ minh họa chi tiết	13
PHẦN 4 ỨNG DỤNG PHÂN BỐ χ^2 VÀ PHÂN BỐ FISHER-SNEDECOR (F-DISTRIBUTION) ĐỂ ĐÁNH GIÁ VÀ SO SÁNH CÁC SAI SỐ TRUNG PHƯƠNG.....	15
4.1 Giới thiệu Phân bố Chi-bình phương (χ^2).....	15
4.2 Nguyên lý kiểm định giả thuyết thống kê	16
4.3 Đánh giá độ chính xác của một loạt đo so với tiêu chuẩn (Kiểm định χ^2)	16
4.4 Giới thiệu Phân bố Fisher-Snedecor	17
4.5 So sánh độ chính xác của hai loạt đo độc lập (Kiểm định F).....	18
PHẦN 5 MỐI LIÊN HỆ GIỮA CÁC BÀI TOÁN KIỂM ĐỊNH.....	21
5.1 Các Giả định và Điều kiện Áp dụng	22
5.2 Giới thiệu về thống kê	23
PHẦN 6 KẾT LUẬN.....	23

TÀI LIỆU THAM KHẢO	24
--------------------------	----

DANH MỤC BẢNG BIỂU

Bảng 2-1 Các tham số thống kê cơ bản	10
Bảng 3-1 Bảng tra giá trị $t_{\alpha/2, k}$	14
Bảng 4-1 Bảng tra giá trị tới hạn $F_{\alpha/2, k1, k2}$ (trích đoạn cho $\alpha=0.025$)	19
Bảng 4-2 Trình bày ví dụ cho đánh giá độ chính xác của 2 trị đo bằng Máy A và Máy B	20

DANH MỤC HÌNH VẼ

Hình 1 Sai số tuân theo luật phân bố chuẩn và nguyên tắc 3-sigma.....	9
Hình 2 Phân bố Chi-bình phương χ^2	16
Hình 3 Phân bố F.....	18

PHẦN 1 NỀN TẢNG CỦA SAI SỐ VÀ THỐNG KÊ TRONG KHOA HỌC ĐO ĐẠC

1.1 Bản chất của phép đo và sự tồn tại tất yếu của sai số

Trong lĩnh vực khoa học kỹ thuật nói chung và Trắc địa nói riêng, phép đo là hành động cốt lõi nhằm xác định giá trị số của một đại lượng vật lý. Tuy nhiên, một nguyên lý nền tảng và không thể tránh khỏi là mọi kết quả đo đạc, dù được thực hiện bằng những công nghệ tiên tiến và thiết bị hiện đại nhất, đều không bao giờ đạt được giá trị thực tuyệt đối. Luôn tồn tại một sự khác biệt giữa giá trị đo được và giá trị thực của đại lượng, sự khác biệt này được gọi là sai số của phép đo. Sự tồn tại của sai số là một thuộc tính cố hữu của quá trình đo lường, xuất phát từ sự tương tác phức tạp của nhiều yếu tố.

Các nguồn gây ra sai số có thể được phân loại thành ba nhóm chính:

1. **Sai số do điều kiện ngoại cảnh:** Môi trường đo đạc không bao giờ là lý tưởng và luôn biến động. Các yếu tố như nhiệt độ, áp suất khí quyển, độ ẩm, sự khúc xạ của ánh sáng trong khí quyển, hay sự rung động của mặt đất đều ảnh hưởng đến quá trình truyền tín hiệu (sóng điện từ, tia laser) và sự ổn định của thiết bị, từ đó gây ra sai số trong kết quả đo.
2. **Sai số do máy móc, dụng cụ đo:** Không có một thiết bị đo nào là hoàn hảo. Các sai số trong quá trình chế tạo, lắp ráp, hiệu chỉnh các bộ phận của máy (như trục ngắm, trục quay, vạch khắc trên bàn độ, bộ phận đo dài điện tử) đều là nguồn gốc của sai số. Theo thời gian sử dụng, sự hao mòn cũng làm thay đổi các đặc tính kỹ thuật của máy và gia tăng sai số.
3. **Sai số do người đo:** Yếu tố con người luôn mang tính chủ quan. Các thao tác như ngắm chuẩn mục tiêu, đọc số trên vạch khắc, hay thậm chí là trạng thái tâm sinh lý của người đo đều có thể dẫn đến những sai lệch nhỏ, không thể lường trước một cách chính xác.

Thay vì theo đuổi một mục tiêu bất khả thi là tìm ra "giá trị đúng" duy nhất, Trắc địa hiện đại chuyển sang một quan điểm xác suất. Kết quả đo không còn được xem là một hằng số bất định, mà là một thực thể của một biến ngẫu nhiên. Câu hỏi trung tâm không còn là "Giá trị đúng là gì?" mà đã trở thành "Chúng ta có thể phát biểu điều gì về giá trị đúng với một mức độ tin cậy cho trước?". Câu hỏi thứ hai này chỉ có thể được giải quyết bằng các công cụ của lý thuyết xác suất và thống kê toán học. Do đó, chất lượng của một kết quả trắc địa không chỉ nằm ở bản thân con số đo được, mà quan trọng hơn, nó nằm ở độ không đảm bảo (uncertainty) đi kèm với con số đó. Một kết quả đo chiều dài là

100.000 m sẽ thiếu đi ý nghĩa khoa học nếu không có thông tin về độ chính xác của nó, ví dụ như 100.000 ± 0.002 m. Chính xác suất thống kê cung cấp ngôn ngữ và phương pháp luận để định lượng và diễn giải độ không đảm bảo này.

1.2 Phân loại Sai số trong Trắc địa

Để có thể xử lý và đánh giá đúng đắn, các sai số trong trắc địa được phân thành ba loại chính dựa trên nguyên nhân và đặc điểm của chúng:

- **Sai số hệ thống (Systematic Errors):** Đây là loại sai số phát sinh do những nguyên nhân có quy luật rõ ràng, tác động một cách nhất quán lên các lần đo trong cùng một điều kiện. Ví dụ như sai số do thước thép bị giãn nở vì nhiệt, sai số do trục ngắm của máy kinh vĩ không vuông góc với trục quay, hay sai số do hiệu chỉnh tần số của máy đo dài điện tử chưa chính xác. Đặc điểm của sai số hệ thống là chúng thường có trị số và dấu không đổi hoặc thay đổi theo một quy luật đã biết. Do đó, chúng có thể và cần phải được loại trừ hoặc giảm yếu ảnh hưởng bằng cách áp dụng các số hiệu chỉnh (corrections) vào kết quả đo thô.
- **Sai số ngẫu nhiên (Random Errors):** Sau khi đã loại trừ các sai số hệ thống, những sai số còn lại được gọi là sai số ngẫu nhiên. Chúng phát sinh từ vô số các yếu tố nhỏ, biến đổi không ngừng và không thể kiểm soát hoàn toàn, ví dụ như sự rung động nhẹ của chân máy, sự thay đổi tức thời của khúc xạ khí quyển, hay sai sót nhỏ trong thao tác bắt mục tiêu của người đo. Đặc điểm cơ bản của sai số ngẫu nhiên là chúng không tuân theo một quy luật rõ ràng nào; trị số và dấu của chúng xuất hiện một cách ngẫu nhiên trong các lần đo lặp. Chúng không thể bị loại bỏ hoàn toàn, nhưng có thể được mô hình hóa và xử lý bằng các công cụ của lý thuyết xác suất và thống kê. Đây chính là đối tượng nghiên cứu chính của báo cáo này.
- **Sai số thô hay Sai lầm (Blunders/Gross Errors):** Đây là những sai số có giá trị lớn, vượt xa các sai số ngẫu nhiên và hệ thống, gây ra bởi sự nhầm lẫn, cẩu thả hoặc thiếu kinh nghiệm của người đo. Ví dụ như đọc nhầm số trên mia (đọc 1.8 m thành 1.3 m), ghi chép sai kết quả, ngắm sai mục tiêu. Những sai số này phải được phát hiện và loại bỏ khỏi chuỗi số liệu đo trước khi tiến hành các bước xử lý tiếp theo.

1.3 Vai trò của Xác suất Thống kê

Trong bối cảnh tồn tại các loại sai số trên, lý thuyết xác suất và thống kê toán học đóng vai trò là bộ công cụ lý thuyết không thể thiếu. Mục tiêu của việc áp dụng thống kê trong trắc địa không phải là để loại bỏ hoàn toàn sai số ngẫu nhiên—một điều không thể thực hiện được—mà là để:

1. **Tìm ra giá trị ước lượng tin cậy nhất (most probable value)** cho đại lượng cần đo từ một loạt các kết quả đo chứa sai số ngẫu nhiên.
2. Đánh giá một cách định lượng độ chính xác và độ tin cậy của giá trị ước lượng đó.
3. **Xây dựng các tiêu chuẩn khách quan** để so sánh độ chính xác của các phương pháp đo, các thiết bị khác nhau hoặc các kết quả đo từ những người thực hiện khác nhau.
4. **Phát hiện và loại bỏ các sai số thô** dựa trên các quy luật phân bố xác suất.

Tóm lại, xác suất thống kê cung cấp một khung lý thuyết chặt chẽ để xử lý tính không chắc chắn vốn có trong mọi phép đo, biến một tập hợp các số liệu đo đạc thô thành những kết quả khoa học có ý nghĩa, đi kèm với các chỉ tiêu chất lượng rõ ràng.

PHẦN 2 CÁC KHÁI NIỆM THỐNG KÊ NỀN TẢNG TRONG XỬ LÝ SỐ LIỆU TRẮC ĐỊA

Để áp dụng hiệu quả các công cụ thống kê, cần nắm vững một số khái niệm và tham số nền tảng. Các khái niệm này tạo nên ngôn ngữ chung để mô tả và phân tích dữ liệu đo đạc.

2.1 Tổng thể và Mẫu (Population and Sample)

- **Tổng thể (Population):** Trong lý thuyết thống kê, tổng thể là tập hợp tất cả các phần tử mà chúng ta quan tâm. Trong trắc địa, khi đo một đại lượng nào đó, tổng thể được hiểu là tập hợp vô hạn tất cả các kết quả đo có thể có của đại lượng đó nếu ta có thể lặp lại phép đo vô số lần trong cùng một điều kiện không đổi. Các tham số của tổng thể, như giá trị trung bình thực (kỳ vọng toán học) μ và phương sai σ^2 , là những hằng số lý thuyết, cố định nhưng chúng ta không bao giờ biết được giá trị chính xác của chúng.
- **Mẫu (Sample):** Do không thể thực hiện vô hạn các phép đo, trong thực tế, chúng ta chỉ tiến hành một số hữu hạn n lần đo. Tập hợp n kết quả đo này, ký hiệu là $l_1, l_2, \dots, l_n$, được gọi là một mẫu ngẫu nhiên được rút ra từ tổng thể. Từ mẫu này, chúng ta có thể tính toán các đặc trưng thống kê của mẫu, như trung bình mẫu $\bar{x}$ và phương sai mẫu s^2 . Mục tiêu của thống kê suy luận là dùng các đặc trưng của mẫu để ước lượng và kiểm định các giả thuyết về tham số của tổng thể.

2.2 Các tham số đặc trưng quan trọng

Từ một mẫu đo, các tham số sau đây là quan trọng nhất và được sử dụng thường xuyên trong xử lý số liệu trắc địa.

- Giá trị Ước lượng Tốt nhất (Best Estimate): Khi đo lặp một đại lượng n lần, giá trị gần với giá trị thực nhất, hay còn gọi là giá trị có khả năng nhất, được chứng minh là trung bình số học (Arithmetic Mean) của các kết quả đo. Trung bình mẫu, ký hiệu là $\bar{x}$, được tính theo công thức:

$$\bar{x} = \frac{\sum_1^n l_i}{n}$$

Trung bình số học là một ước lượng không chệch (unbiased) và hiệu quả nhất (most efficient) cho kỳ vọng μ của tổng thể.

- Sai số Trung phương (Root Mean Square Error - RMSE) và Phương sai (Variance): Sai số trung phương là một chỉ tiêu thống kê cốt lõi, đặc trưng cho mức độ phân tán của các kết quả đo lặp xung quanh giá trị trung bình của chúng. Nó phản ánh độ chính xác (precision) của một phương pháp đo, một thiết bị, hay một người đo. Sai số trung phương càng nhỏ, các kết quả đo càng tập trung, và do đó độ chính xác càng cao.
- **Phương sai của tổng thể (σ^2):** Là một tham số lý thuyết, không xác định được.
- Ước lượng không chệch cho phương sai từ mẫu (m^2 hoặc s^2): Để ước lượng σ^2 từ một mẫu hữu hạn, người ta sử dụng công thức Bessel:

$$m^2 = s^2 = \frac{\sum_1^n (l_i - \bar{x})^2}{n - 1} = \frac{\sum_1^n v_i^2}{n - 1}$$

Trong đó, $v_i = l_i - \bar{x}$ được gọi là số hiệu chỉnh xác suất của trị đo thứ i .

Việc sử dụng mẫu số là $(n-1)$ thay vì n có một ý nghĩa thống kê sâu sắc. Khi chúng ta tính tổng bình phương các độ lệch, chúng ta đang so sánh các trị đo l_i với trung bình mẫu $\bar{x}$ —một giá trị được tính toán từ chính mẫu đó—chứ không phải giá trị trung bình thực μ . Do tổng các sai số dư so với trung bình mẫu luôn bằng không ($\sum v_i = \sum (l_i - \bar{x}) = 0$). Điều này có nghĩa là chỉ có $(n-1)$ "mảnh thông tin" độc lập về sự biến động trong mẫu. Do đó, $(n-1)$ được gọi là **bậc tự do (degrees of freedom)** của phương sai mẫu. Việc chia cho bậc tự do $(n-1)$ đảm bảo rằng m^2 là một ước lượng không chệch cho phương sai thực σ^2 của tổng thể.

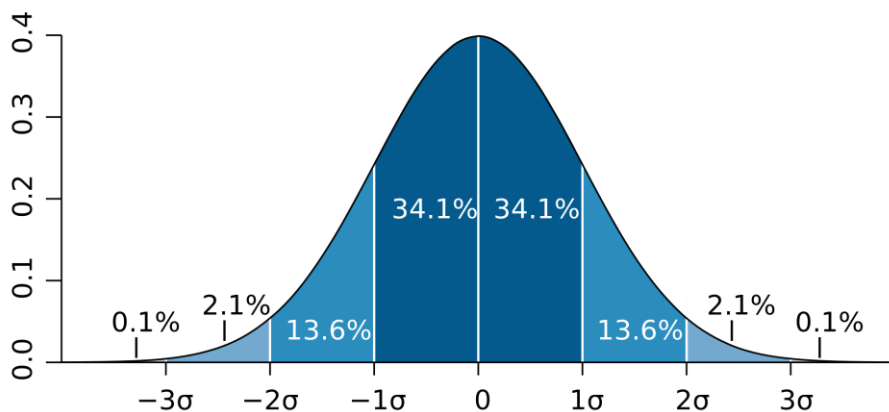
Nó có cùng đơn vị với đại lượng đo và là chỉ tiêu phổ biến nhất để thể hiện độ chính xác.

Một sự phân biệt quan trọng cần được làm rõ là giữa **độ chính xác (precision)** và **độ đúng (accuracy)**.

- **Độ chính xác** liên quan đến mức độ lặp lại hay mức độ phân tán của các kết quả đo. Nó được định lượng bởi sai số trung phương m . Một loạt đo có độ chính xác cao khi các giá trị đo rất gần nhau.
- Độ đúng liên quan đến mức độ gần của giá trị ước lượng (trung bình mẫu $\bar{x}$) so với giá trị thực μ . Một loạt đo có thể rất chính xác (các giá trị đo cụm lại) nhưng lại không đúng (toàn bộ cụm đo lệch xa so với giá trị thực) nếu có sai số hệ thống chưa được loại trừ.
- Lý thuyết xác suất thống kê chủ yếu giải quyết các vấn đề liên quan đến sai số ngẫu nhiên, do đó các chỉ tiêu như m được trực tiếp đánh giá độ chính xác. Để đảm bảo độ đúng, người kỹ sư phải thực hiện các quy trình hiệu chỉnh máy móc và áp dụng các số hiệu chỉnh cần thiết để loại bỏ sai số hệ thống trước khi tiến hành xử lý thống kê.

2.3 Luật Phân bố Chuẩn (Normal Distribution)

Luật phân bố chuẩn, hay phân bố Gauss, là định luật quan trọng nhất trong lý thuyết sai số. Định đề trung tâm của lý thuyết sai số phát biểu rằng: sai số ngẫu nhiên trong trắc địa, vốn là tổng hợp của nhiều nguồn sai số nhỏ, độc lập, sẽ tuân theo luật phân bố chuẩn. Hàm mật độ xác suất của phân bố chuẩn có dạng hình chuông đặc trưng, đối xứng qua giá trị trung bình μ .



Hình 1 Sai số tuân theo luật phân bố chuẩn và nguyên tắc 3-sigma

Các tính chất quan trọng của phân bố chuẩn được áp dụng trong trắc địa:

- **Tính đối xứng:** Sai số ngẫu nhiên có giá trị dương và âm xuất hiện với tần suất như nhau.
- **Tính tập trung:** Sai số có trị tuyệt đối nhỏ xuất hiện thường xuyên hơn sai số có trị

tuyệt đối lớn.

- **Quy tắc 3-sigma (68-95-99.7):** Với một tổng thể có phân bố chuẩn, khoảng 68.3% các sai số sẽ nằm trong khoảng $\pm\sigma$, 95.4% nằm trong khoảng $\pm 2\sigma$, và 99.7% nằm trong khoảng $\pm 3\sigma$. Quy tắc này là cơ sở để xác định sai số giới hạn và phát hiện sai số thô.

Bảng 2-1 Các tham số thống kê cơ bản

Tên tham số	Ký hiệu (Tổng thể)	Ký hiệu (Mẫu)	Công thức tính từ mẫu	Ý nghĩa trong Trắc địa
Giá trị trung bình	μ	$\bar{x}$	$\bar{x} = \frac{\sum_1^n l_i}{n}$	Ước lượng tốt nhất cho giá trị thực của đại lượng đo.
Phương sai	σ^2	m^2 (hoặc s^2)	$m^2 = s^2 = \frac{\sum_1^n (l_i - \bar{x})^2}{n - 1}$	Đặc trưng cho độ phân tán, độ chính xác của phép đo.
Độ lệch chuẩn	σ	m (hoặc s)	$m = s = \sqrt{\frac{\sum_1^n (l_i - \bar{x})^2}{n - 1}}$	Sai số trung phương của một lần đo, chỉ tiêu phổ biến nhất về độ chính xác.
Sai số trung phương của trị trung bình	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$m_{\bar{x}} = \frac{m}{\sqrt{n}}$	$m_{\bar{x}} = \frac{m}{\sqrt{n}}$	Đặc trưng cho độ chính xác của giá trị trung bình mẫu.

PHẦN 3 ỨNG DỤNG PHÂN BỐ STUDENT (T-DISTRIBUTION) ĐỂ ƯỚC LƯỢNG KHOẢNG TIN CẬY CHO TRỊ TRUNG BÌNH

Trong thực tế, chúng ta không bao giờ biết được giá trị thực μ của một đại lượng đo. Chúng ta chỉ có thể tính được giá trị ước lượng của nó là trung bình mẫu $\bar{x}$. Một câu hỏi tự nhiên và quan trọng được đặt ra là: giá trị thực μ có khả năng nằm cách $\bar{x}$ bao xa? Phân bố Student cung cấp một công cụ toán học chặt chẽ để trả lời câu hỏi này thông qua việc xây dựng khoảng tin cậy.

3.1 Giới thiệu Phân bố Student

Phân bố Student (hay t -distribution) được nhà thống kê William Sealy Gosset phát triển vào đầu thế kỷ 20 khi ông làm việc tại nhà máy bia Guinness và công bố dưới bút danh "Student". Vấn đề ông gặp phải là một vấn đề kinh điển trong thống kê: làm thế nào để suy luận về giá trị trung bình của tổng thể khi phương sai tổng thể σ^2 chưa biết và phải được ước lượng từ phương sai mẫu m^2 , đặc biệt là khi kích thước mẫu n nhỏ.

Khi mẫu lớn (thường $n > 30$), theo định lý giới hạn trung tâm, phân bố của đại lượng $(\bar{x} - \mu)/(m/n)$ có thể được xấp xỉ bằng phân bố chuẩn. Tuy nhiên, khi mẫu nhỏ, việc thay thế σ bằng m (một biến ngẫu nhiên) làm tăng thêm sự không chắc chắn, khiến cho đuôi của phân bố trở nên "béo" hơn so với phân bố chuẩn. Phân bố Student mô tả chính xác hình dạng của phân bố này [1].

Đặc điểm của phân bố Student:

- Nó là một họ các phân bố, với hình dạng cụ thể được quyết định bởi một tham số duy nhất gọi là **bậc tự do** ($k = n - 1$).
- Hình dạng của nó đối xứng, hình chuông, tương tự phân bố chuẩn, và cũng có giá trị trung bình bằng 0.
- So với phân bố chuẩn, phân bố Student có phương sai lớn hơn 1 và có đuôi "dày" hơn, điều này có nghĩa là nó gán xác suất cao hơn cho các giá trị cực đoan. Điều này phản ánh sự không chắc chắn tăng lên do phải ước lượng σ^2 từ mẫu nhỏ.
- Khi bậc tự do k tăng lên (tức kích thước mẫu n tăng), phân bố Student sẽ hội tụ về phân bố chuẩn. Khi $k \rightarrow \infty$, phân bố Student chính là phân bố chuẩn.

3.2 Xây dựng Khoảng tin cậy (Confidence Interval)

Khoảng tin cậy là một khoảng giá trị được tính toán từ dữ liệu mẫu, mà chúng ta có

một mức độ tin cậy nhất định rằng nó chứa tham số thực của tổng thể (trong trường hợp này là giá trị thực μ).

- **Độ tin cậy (Confidence Level):** Là xác suất mà khoảng tin cậy được xây dựng sẽ chứa tham số thực, thường được ký hiệu là $(1-\alpha)$. Các mức tin cậy phổ biến trong kỹ thuật là 90%, 95%, 99%. Ví dụ, độ tin cậy 95% có nghĩa là nếu chúng ta lặp lại quy trình lấy mẫu (đo đạc) và xây dựng khoảng tin cậy rất nhiều lần, thì 95% các lần đo đạc sẽ có khoảng tin cậy được tạo ra chứa giá trị thực μ . Đây là một phát biểu về tính đúng đắn của quy trình.
- **Mức ý nghĩa (Significance Level):** Ký hiệu là α , là xác suất mà khoảng tin cậy không chứa tham số thực. Nó là phần bù của độ tin cậy: $\alpha=1-(1-\alpha)$. Ví dụ, độ tin cậy 95% tương ứng với mức ý nghĩa $\alpha=0.05$.

Công thức xây dựng khoảng tin cậy cho μ :

Đại lượng ngẫu nhiên $T = \frac{\bar{x}-\mu}{m_{\bar{x}}} = \frac{\bar{x}-\mu}{m_x/\sqrt{n}}$ tuân theo phân bố Student với $k=n-1$ bậc tự do. Từ đó, ta có thể tìm được một giá trị tới hạn $t_{\alpha/2,k}$ từ bảng phân bố Student sao cho:

$$P(-t_{\alpha/2,k} \leq T \leq t_{\alpha/2,k}) = 1 - \alpha$$

Biến đổi bất đẳng thức trên, ta có:

$$P(\bar{x} - t_{\alpha/2,k} m_x / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2,k} m_x / \sqrt{n}) = 1 - \alpha$$

Trong đó Δ là khoảng từ $\bar{x} - t_{\alpha/2,k} m_x / \sqrt{n}$ đến $\bar{x} + t_{\alpha/2,k} m_x / \sqrt{n}$

Công thức này cho thấy một mối quan hệ tương hỗ quan trọng trong thực hành đo đạc. Để có một khoảng tin cậy hẹp hơn (tức là ước lượng chính xác hơn, Δ nhỏ hơn), người kỹ sư có thể:

1. **Giảm độ tin cậy:** Ví dụ, chuyển từ 95% xuống 90% sẽ làm giảm giá trị $t_{\alpha/2,k}$, từ đó làm giảm Δ . Tuy nhiên, điều này làm giảm sự chắc chắn của chúng ta.
2. **Cải thiện độ chính xác của phép đo:** Sử dụng máy móc tốt hơn, phương pháp đo ổn định hơn để giảm sai số trung phương của một lần đo m .
3. **Tăng số lần đo n :** Đây là biện pháp phổ biến nhất. Vì n nằm dưới dấu căn, việc tăng số lần đo sẽ làm giảm $m_{\bar{x}}$ và thu hẹp khoảng tin cậy. Mối quan hệ này cho phép các kỹ sư lập kế hoạch đo đạc một cách khoa học: dựa trên yêu cầu về độ chính xác (Δ) và độ tin cậy, cùng với ước tính ban đầu về m , họ có thể tính toán số lần đo n cần thiết trước khi ra thực địa.

3.3 Quy trình thực hiện và ví dụ minh họa chi tiết

Quy trình:

1. **Bước 1: Thu thập dữ liệu:** Thực hiện n lần đo lặp độc lập cho một đại lượng, thu được mẫu $l_1, l_2, \dots, l_n$.
2. **Bước 2: Tính toán các đặc trưng mẫu:**
 - Tính giá trị trung bình mẫu: $\bar{x} = \frac{\sum_1^n l_i}{n}$
 - Tính các sai số dư: $v_i = l_i - \bar{x}$.
 - Tính sai số trung phương của một lần đo: $m = s = \sqrt{\frac{\sum_1^n (l_i - \bar{x})^2}{n-1}}$.
3. **Bước 3: Xác định tham số:**
 - Chọn độ tin cậy mong muốn $(1-\alpha)$ (ví dụ: 95%, suy ra $\alpha=0.05$).
 - Xác định bậc tự do $k=n-1$.
4. **Bước 4: Tra bảng phân bố Student:** Tìm giá trị tới hạn $t_{\alpha/2, k}$.
5. **Bước 5: Tính khoảng tin cậy:**
 - Tính sai số trung phương của trị trung bình: $m_{\bar{x}} = \frac{m}{\sqrt{n}}$.
 - Tính sai số giới hạn: $\Delta = t_{\alpha/2, k} m_{\bar{x}}$.
 - Kết luận khoảng tin cậy: $\mu \in$.

Ví dụ thực tiễn:

Một kỹ sư sử dụng máy toàn đạc điện tử để đo lặp một góc 8 lần và thu được các kết quả sau (đơn vị: giây góc, phần độ và phút được giả định là $45^\circ 12'$): 13.4", 11.8", 12.5", 14.1", 12.9", 11.5", 13.7", 12.3". Hãy xác định khoảng tin cậy 95% cho giá trị thực của góc này.

- **Bước 1:** Đã có mẫu đo với $n=8$.
- **Bước 2:** Tính toán đặc trưng mẫu:
 - $\bar{x} = (13.4 + 11.8 + 12.5 + 14.1 + 12.9 + 11.5 + 13.7 + 12.3) / 8 = 102.2 / 8 = 12.775''$
 - Tính các sai số dư và bình phương của chúng:
 - $v_1 = 13.4 - 12.775 = 0.625 \Rightarrow v_1^2 = 0.390625$
 - $v_2 = 11.8 - 12.775 = -0.975 \Rightarrow v_2^2 = 0.950625$
 - $v_3 = 12.5 - 12.775 = -0.275 \Rightarrow v_3^2 = 0.075625$
 - $v_4 = 14.1 - 12.775 = 1.325 \Rightarrow v_4^2 = 1.755625$
 - $v_5 = 12.9 - 12.775 = 0.125 \Rightarrow v_5^2 = 0.015625$
 - $v_6 = 11.5 - 12.775 = -1.275 \Rightarrow v_6^2 = 1.625625$
 - $v_7 = 13.7 - 12.775 = 0.925 \Rightarrow v_7^2 = 0.855625$

- $v_8 = 12.3 - 12.775 = -0.475 \Rightarrow v_8^2 = 0.225625$
 - $\sum v_i^2 = 5.895$
 - $m^2 = 8 - 15.895 = 75.895 \approx 0.8421$
 - $m = \sqrt{0.8421} \approx 0.918''$
- **Bước 3:** Xác định tham số:
 - Độ tin cậy 95% $\Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$.
 - Bậc tự do $k = n - 1 = 8 - 1 = 7$.
- **Bước 4:** Tra bảng phân bố Student (xem Bảng 2), ta tìm được $t_{0.025,7} = 2.365$.
- **Bước 5:** Tính khoảng tin cậy:
 - $m_{\bar{x}} = \frac{m}{\sqrt{n}} = 80.918 \approx 0.3245''$
 - $\Delta = t_{0.025,7} * m_{\bar{x}} = 2.365 \cdot 0.3245 \approx 0.767''$
 - Khoảng tin cậy cho μ là: $[12.775 - 0.767, 12.775 + 0.767] = [12.008'', 13.542'']$.

Diễn giải kết quả: Với độ tin cậy 95%, chúng ta có thể kết luận rằng giá trị thực của góc đo nằm trong khoảng từ $45^\circ 12' 12.008''$ đến $45^\circ 12' 13.542''$.

Bảng 3-1 Bảng tra giá trị $t_{\alpha/2,k}$

Bậc tự do (k)	t0.05 ($\alpha=0.10$)	t0.025 ($\alpha=0.05$)	t0.005 ($\alpha=0.01$)
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499

8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
20	1.725	2.086	2.845
30	1.697	2.042	2.750
∞	1.645	1.960	2.576

PHẦN 4 ỨNG DỤNG PHÂN BỐ χ^2 VÀ PHÂN BỐ FISHER-SNEDECOR (F-DISTRIBUTION) ĐỂ ĐÁNH GIÁ VÀ SO SÁNH CÁC SAI SỐ TRUNG PHƯƠNG

Nếu phân bố Student giúp chúng ta đánh giá độ tin cậy của giá trị trung bình, thì phân bố Chi-bình phương và Fisher-Snedecor là những công cụ thống kê mạnh mẽ để đánh giá và so sánh độ chính xác của các loạt đo thông qua việc kiểm định các giả thuyết về phương sai.

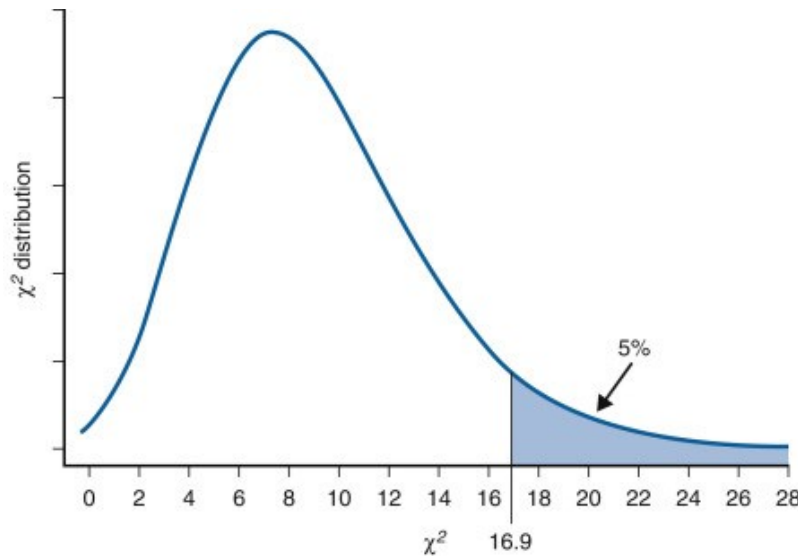
4.1 Giới thiệu Phân bố Chi-bình phương (χ^2)

Phân bố Chi-bình phương (Chi-squared distribution, ký hiệu là χ^2) là một phân bố xác suất liên tục, được định nghĩa là tổng của các bình phương của các biến ngẫu nhiên độc lập có phân bố chuẩn tắc (trung bình 0, phương sai 1). Đây là một trong những phân bố được sử dụng rộng rãi nhất trong thống kê suy luận [2].

Đặc điểm của phân bố Chi-bình phương:

- **Không âm:** Vì là tổng của các bình phương, giá trị χ^2 luôn lớn hơn hoặc bằng 0.³
- **Hình dạng phụ thuộc vào bậc tự do:** Hình dạng của phân bố χ^2 được quyết định bởi một tham số duy nhất là bậc tự do (k). Với bậc tự do thấp, phân bố bị lệch phải mạnh. Khi bậc tự do tăng lên, phân bố trở nên đối xứng hơn và tiệm cận với phân bố chuẩn.³

- **Ứng dụng:** Trong thống kê, phân bố χ^2 là nền tảng cho nhiều kiểm định giả thuyết, bao gồm kiểm định tính phù hợp (goodness-of-fit), kiểm định tính độc lập giữa các biến định tính, và đặc biệt quan trọng trong trắc địa là kiểm định về một phương sai của tổng thể.



Hình 2 Phân bố Chi-bình phương χ^2

4.2 Nguyên lý kiểm định giả thuyết thống kê

Các ứng dụng của phân bố χ^2 , và phân bố F đều dựa trên một quy trình chung của kiểm định giả thuyết thống kê, bao gồm các bước:

1. **Phát biểu giả thuyết:** Đặt ra một giả thuyết không (Null Hypothesis, H_0), thường là một phát biểu về sự "không có khác biệt" hoặc "không có ảnh hưởng", và một giả thuyết đối (Alternative Hypothesis, H_1), là điều chúng ta muốn chứng minh.
2. **Chọn mức ý nghĩa α :** Xác định xác suất sai lầm loại I (bác bỏ H_0 trong khi nó đúng) mà chúng ta chấp nhận được (thường là 0.05 hoặc 0.01).
3. **Tính toán tiêu chuẩn kiểm định (Test Statistic):** Từ dữ liệu mẫu, tính một giá trị theo một công thức nhất định. Giá trị này sẽ tuân theo một phân bố xác suất đã biết (như F , t , χ^2) nếu H_0 là đúng.
4. **Xác định miền bác bỏ:** Dựa vào α và bậc tự do, tìm giá trị tới hạn (critical value) từ bảng phân bố tương ứng. Giá trị này xác định ranh giới của miền bác bỏ.
5. **Đưa ra kết luận:** So sánh tiêu chuẩn kiểm định với giá trị tới hạn. Nếu tiêu chuẩn kiểm định rơi vào miền bác bỏ, chúng ta bác bỏ H_0 và chấp nhận H_1 . Nếu không, chúng ta kết luận "không có đủ bằng chứng để bác bỏ H_0 ".

4.3 Đánh giá độ chính xác của một loạt đo so với tiêu chuẩn (Kiểm định χ^2)

Đây là một bài toán tiền thân của kiểm định F , sử dụng phân bố Chi-bình phương (χ^2), nhưng có cùng mục đích là đánh giá phương sai.

- **Mục tiêu:** Kiểm tra xem độ chính xác của một phương pháp đo hoặc một thiết bị (đặc trưng bởi phương sai thực tế σ^2) có phù hợp với một độ chính xác tiêu chuẩn, yêu cầu kỹ thuật hoặc lý lịch máy (σ_0^2) hay không.
- **Giả thuyết:**
 - $H_0: \sigma^2 = \sigma_0^2$ (Độ chính xác thực tế bằng độ chính xác tiêu chuẩn).
 - $H_1: \sigma^2 \geq \sigma_0^2$ (Độ chính xác thực tế kém hơn tiêu chuẩn - kiểm định một phía bên phải, vì chúng ta thường quan tâm đến trường hợp máy đo kém đi).
- **Tiêu chuẩn kiểm định:** Đại lượng $\chi_{tinh}^2 = \frac{(n-1)m^2}{\sigma_0^2}$ tuân theo phân bố Chi-bình phương với $k=n-1$ bậc tự do.
- **Quy tắc quyết định:** Với mức ý nghĩa α , tra bảng χ^2 để tìm giá trị tới hạn $\chi_{\alpha, n-1}^2$. Nếu $\chi_{tinh}^2 > \chi_{\alpha, n-1}^2$, ta bác bỏ H_0 .

Ví dụ: Một máy toàn đạc có sai số trung phương đo góc theo lý lịch là $\sigma_0 = 1.5''$. Để kiểm tra máy, người ta tiến hành đo lặp một góc 20 lần ($n=20$) và tính được sai số trung phương tính từ việc đo đạc thực tế là $m=1.8''$. Với mức ý nghĩa $\alpha=0.05$, hãy kiểm tra xem độ chính xác của máy có còn đạt tiêu chuẩn kỹ thuật không?

1. **Giả thuyết:** H_0 : sự khác nhau giữa sai số trung phương tính được và sai số trung phương theo lý lịch máy là không có ý nghĩa thống kê (sự khác nhau giữa $m = 1.8''$ và $\sigma_0 = 1.5''$ là không có ý nghĩa về mặt thống kê) hay không đáng kể. Nói cách khác, sai số của máy có thể chấp nhận được không cần hiệu chỉnh. H_1 giả thiết phủ định là: $\sigma^2 = (1.8'')^2 = 3.24$ lớn hơn nhiều so với phương sai cho phép ban đầu của máy theo lý lịch là $\sigma_0^2 = (1.5'')^2 = 2.25$
2. Mức ý nghĩa: $\alpha=0.05$.
3. Tiêu chuẩn kiểm định: $n=20, m^2=(1.8)^2=3.24, \sigma_0^2=2.25$.

$$\chi_{tinh}^2 = \frac{(20 - 1) * 3.24}{2.25} = 27.36$$

4. **Miền bác bỏ:** Bậc tự do $k=19$. Tra bảng χ^2 , ta có $\chi_{0.05, 19}^2=30.14$.
5. **Kết luận:** Vì $\chi_{tinh}^2 = 27.36 < \chi_{bang}^2 = 30.14$, tiêu chuẩn kiểm định không rơi vào miền bác bỏ. Do đó, chúng ta không có đủ bằng chứng thống kê ở mức ý nghĩa 5% để kết luận rằng độ chính xác của máy đã kém hơn so với tiêu chuẩn.

4.4 Giới thiệu Phân bố Fisher-Snedecor

Phân bố F được đặt theo tên của hai nhà thống kê vĩ đại R.A. Fisher và George W.

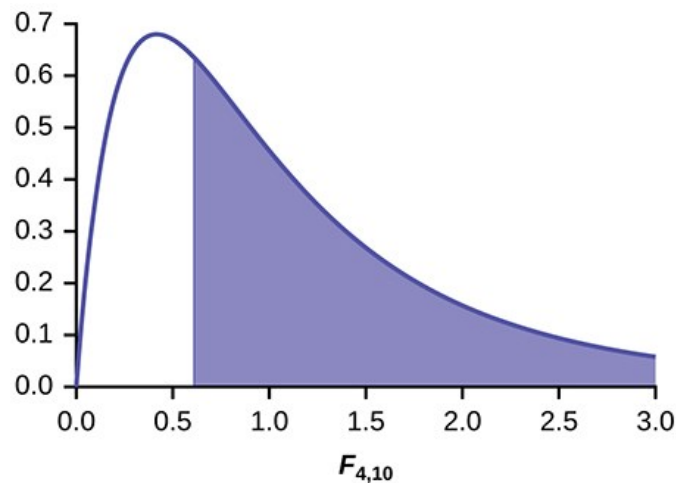
Snedecor. Về mặt toán học, một biến ngẫu nhiên có phân bố F được định nghĩa là tỷ số của hai biến ngẫu nhiên độc lập có phân bố Chi-bình phương (χ^2), mỗi biến được chia cho bậc tự do tương ứng của nó [3].

$$F = \frac{\chi_1^2/k_1}{\chi_2^2/k_2}$$

Trong ứng dụng thực tế, nó được sử dụng để kiểm định tỷ số của hai phương sai mẫu độc lập, $F = \frac{m_1}{m_2}$

Đặc điểm của phân bố F (Hình 3):

- Phân bố F không đối xứng và bị lệch về bên phải.
- Giá trị của F luôn dương ($F \geq 0$).
- Hình dạng của phân bố F được xác định bởi hai tham số bậc tự do: bậc tự do của tử số ($k_1 = n_1 - 1$) và bậc tự do của mẫu số ($k_2 = n_2 - 1$).
- Việc tra bảng giá trị tới hạn của phân bố F phức tạp hơn phân bố t, vì cần có các bảng riêng cho mỗi mức ý nghĩa α .



Hình 3 Phân bố F

4.5 So sánh độ chính xác của hai loạt đo độc lập (Kiểm định F)

Đây là ứng dụng phổ biến và quan trọng nhất của phân bố F trong trắc địa.

- **Mục tiêu:** So sánh độ chính xác của hai phương pháp đo, hai máy đo, hoặc hai người đo khác nhau.
- Giả thuyết:
 - $H_0: \sigma_1^2 = \sigma_2^2$ (Độ chính xác của hai loạt đo là như nhau).

- $H_1: \sigma_1^2 \neq \sigma_2^2$ (Độ chính xác của hai loạt đo là khác nhau - kiểm định hai phía).
- **Tiêu chuẩn kiểm định:** Tính tỷ số $F_{tinh} = \frac{m_1^2}{m_2^2}$, trong đó quy ước m_1^2 luôn là phương sai mẫu lớn hơn để $F_{tinh} \geq 1$. Tiêu chuẩn này tuân theo phân bố F với bậc tự do $k_1 = n_1 - 1$ và $k_2 = n_2 - 1$.
- **Quy tắc quyết định:** Với mức ý nghĩa α , tra bảng F (Bảng 4-1) để tìm giá trị tới hạn $F_{\alpha/2, k_1, k_2}$. Nếu $F_{tinh} > F_{\alpha/2, k_1, k_2}$, ta bác bỏ H_0 .

Bảng 4-1 Bảng tra giá trị tới hạn $F_{\alpha/2, k_1, k_2}$ (trích đoạn cho $\alpha=0.025$)

$k_1 \rightarrow$ $k_2 \downarrow$	10	12	15	20	24
10	3.72	3.54	3.37	3.21	3.14
15	3.06	2.90	2.76	2.62	2.55
20	2.77	2.62	2.49	2.35	2.28
25	2.61	2.46	2.34	2.20	2.14
30	2.51	2.37	2.25	2.11	2.04

●

Ví dụ thực tiễn: Để so sánh độ chính xác của hai máy thủy chuẩn A và B, hai nhóm đo đạc đã tiến hành đo lặp một đoạn chênh cao. Nhóm dùng máy A đo $n_A=16$ lần, tính được sai số trung phương $m_A=1.2$ mm. Nhóm dùng máy B đo $n_B=21$ lần, tính được $m_B=0.9$ mm. Với mức ý nghĩa $\alpha=0.05$, hãy kiểm tra xem độ chính xác của hai máy có khác nhau một cách có ý nghĩa thống kê hay không?

1. **Giả thuyết:** $H_0: \sigma_A^2 = \sigma_B^2$, $H: \sigma_A^2 \neq \sigma_B^2$.
2. **Mức ý nghĩa:** $\alpha=0.05 \Rightarrow \alpha/2=0.025$.
3. Tiêu chuẩn kiểm định:
 - Phương sai mẫu: $m_A^2=1.2^2=1.44$; $m_B^2=(0.9)^2=0.81$ (các ký hiệu σ và m có thể cùng biểu thị một đại lượng là sai số trung phương hay độ lệch chuẩn).

- Vì $m_A^2 > m_B^2$, ta đặt $m_1^2 = m_A^2$ và $m_2^2 > m_B^2$.
 - $F_{tinh} = \frac{m_1^2}{m_2^2} = 0.81/1.44 \approx 1.778$.
4. Miền bác bỏ:
- Bậc tự do của tử số: $k_1 = n_A - 1 = 16 - 1 = 15$.
 - Bậc tự do của mẫu số: $k_2 = n_B - 1 = 21 - 1 = 20$.
 - Tra bảng F (xem Bảng 3), ta tìm giá trị tới hạn $F_{0.025, 15, 20} \approx 2.57$.
5. **Kết luận:** Vì $F_{tinh} = 1.778 < F_{\alpha/2, k_1, k_2} = 2.57$, tiêu chuẩn kiểm định không rơi vào miền bác bỏ. Do đó, chúng ta kết luận rằng: "Với mức ý nghĩa 5%, không có đủ bằng chứng thống kê để nói rằng độ chính xác của máy A và máy B là khác nhau". Mặc dù $m_A > m_B$, sự khác biệt này được coi là không có ý nghĩa thống kê và có thể chỉ là do biến động ngẫu nhiên của mẫu (nhóm các trị đo).

Một trong những ứng dụng sâu sắc nhất của kiểm định F là nó tạo cơ sở cho việc **gán trọng số (weighting)** trong các bài toán bình sai lưới trắc địa theo phương pháp số bình phương nhỏ nhất. Nguyên lý bình sai yêu cầu các trị đo phải được gán trọng số tỷ lệ nghịch với phương sai của chúng ($p_i = C/m_i^2$). Trước khi có thể gán các trọng số khác nhau cho các nhóm đo (ví dụ, các cạnh đo bằng các máy khác nhau, hoặc các góc đo ở các điều kiện khác nhau), chúng ta phải chứng minh được rằng các phương sai của chúng thực sự khác nhau về mặt thống kê. Kiểm định F chính là công cụ để thực hiện điều này. Nếu kiểm định F cho thấy sự khác biệt là có ý nghĩa (H_0 bị bác bỏ), chúng ta mới có cơ sở toán học để gán các trọng số khác nhau. Ngược lại, nếu không thể bác bỏ H_0 , chúng ta phải coi chúng có cùng độ chính xác và gán trọng số bằng nhau. Điều này cho thấy mối liên kết chặt chẽ giữa thống kê cơ bản và các lý thuyết trắc địa cao cấp.

Bảng 4-2 Trình bày ví dụ cho đánh giá độ chính xác của 2 trị đo bằng Máy A và Máy B

Thông số	Máy A	Máy B
Số lần đo (n)	16	21
Sai số trung phương mẫu (m)	1.2 mm	0.9 mm

Phương sai từ hai nhóm đo (m^2)	1.44	0.81
Bậc tự do (k)	15	20
Tính toán kiểm định		
Giả thuyết H_0	$m_A^2 = m_B^2$	
Mức ý nghĩa α	0.05 (kiểm định 2 phía)	
Tiêu chuẩn kiểm định F_{tinh}	$F_{tinh} = \frac{m_1^2}{m_2^2} = 0.81/1.44 = 1.778$	
Bậc tự do	$k_1=15, k_2=20$	
Giá trị tới hạn F_{bang}	$F_{0.025,15,20}=2.57$	
Kết luận	$F_{tinh} = 1.778 < F_{\alpha/2, k_1, k_2} = 2.57.$ Không đủ bằng chứng để bác bỏ H_0 .	

PHẦN 5 MỐI LIÊN HỆ GIỮA CÁC BÀI TOÁN KIỂM ĐỊNH

Trong thực hành xử lý số liệu trắc địa, các bài toán thống kê không tồn tại một cách riêng lẻ mà thường được tiến hành theo một quy trình logic. Một quy trình xử lý chuẩn thường bao gồm:

1. **Phát hiện và loại bỏ sai số thô:** Sử dụng các tiêu chuẩn như quy tắc 3 σ hoặc các kiểm định thống kê chuyên biệt (ví dụ, kiểm định Pope) để loại bỏ các giá trị đo bất thường.

2. **Kiểm định tính chuẩn của dữ liệu:** Trước khi áp dụng các kiểm định t và F , cần phải kiểm tra xem dữ liệu có tuân theo phân bố chuẩn hay không. Đây là giả định nền tảng cho các phương pháp này.
3. **Kiểm định sự đồng nhất về phương sai (Homogeneity of Variances):** Khi có nhiều nhóm đo, cần sử dụng kiểm định F (cho 2 nhóm) hoặc các kiểm định mở rộng như Bartlett's test hoặc Levene's test (cho nhiều hơn 2 nhóm) để xem xét liệu chúng có cùng độ chính xác hay không. Kết quả này quyết định việc có gộp dữ liệu hoặc gán trọng số khác nhau trong các bước sau.
4. **Ước lượng tham số và khoảng tin cậy:** Sau khi đã có một tập dữ liệu "sạch" và đồng nhất, tiến hành tính toán giá trị ước lượng tốt nhất ($\bar{x}$) và xác định khoảng tin cậy cho nó bằng phân bố Student.

5.1 Các Giả định và Điều kiện Áp dụng

Hiệu lực của các kết luận rút ra từ kiểm định t và kiểm định F phụ thuộc rất nhiều vào việc các giả định cơ bản có được thỏa mãn hay không. Người kỹ sư cần nhận thức rõ về các giả định này:

- **Giả định về phân bố chuẩn (Normality Assumption):** Cả hai kiểm định đều yêu cầu các mẫu đo phải được rút ra từ các tổng thể có phân bố chuẩn. Nếu giả định này bị vi phạm nghiêm trọng (ví dụ, dữ liệu bị lệch nặng hoặc có đuôi rất dày), các giá trị p -value và kết luận có thể không còn chính xác. Các phương pháp để kiểm tra tính chuẩn bao gồm các kiểm định thống kê như Shapiro-Wilk, Kolmogorov-Smirnov, hoặc các phương pháp đồ họa như biểu đồ tần suất (histogram) và biểu đồ Q-Q (Quantile-Quantile plot).
- **Giả định về tính độc lập (Independence Assumption):** Các quan sát trong mỗi mẫu phải độc lập với nhau, và nếu so sánh hai mẫu, chúng cũng phải độc lập. Trong trắc địa, điều này thường được đảm bảo bằng cách thiết kế quy trình đo sao cho kết quả của một lần đo không ảnh hưởng đến kết quả của lần đo tiếp theo.

Hậu quả của việc vi phạm các giả định này là rất nghiêm trọng. Nếu dữ liệu chứa sai số thô chưa được loại bỏ hoặc không tuân theo phân bố chuẩn, các phương pháp thống kê cổ điển có thể cho ra kết quả sai lệch. Điều này đã thúc đẩy sự phát triển của một lĩnh vực quan trọng trong xử lý số liệu trắc địa hiện đại, đó là **Bình sai vững (Robust Estimation)**. Các phương pháp bình sai vững (như phương pháp L1-norm, M-estimators của Huber) được thiết kế để ít nhạy cảm hơn với sự tồn tại của các trị đo ngoại lai và sự vi phạm giả định phân bố chuẩn, do đó cung cấp các kết quả đáng tin cậy hơn trong các điều kiện dữ liệu không lý tưởng.

5.2 Giới thiệu về thống kê

Ngày nay, việc thực hiện các phân tích thống kê phức tạp đã trở nên dễ dàng hơn rất nhiều nhờ sự hỗ trợ của các phần mềm máy tính. Các kỹ sư trắc địa có thể sử dụng:

- **Các phần mềm thống kê đa năng:** MATLAB (với Statistics and Machine Learning Toolbox), ngôn ngữ lập trình R (được coi là tiêu chuẩn vàng trong thống kê), hoặc Python (với các thư viện mạnh mẽ như SciPy, Statsmodels, Pandas) đều cung cấp các hàm dựng sẵn để thực hiện kiểm định t , F , χ^2 và nhiều phân tích phức tạp khác một cách nhanh chóng và chính xác.
- **Các phần mềm bình sai trắc địa chuyên dụng:** Các phần mềm thương mại hoặc mã nguồn mở dùng để bình sai lưới trắc địa (như Trimble Business Center, Leica Geo Office, GeoLab, STAR*NET) thường tích hợp sẵn các module phân tích thống kê sau bình sai, cho phép người dùng đánh giá chất lượng của lưới đo một cách toàn diện, bao gồm cả việc kiểm định các phương sai và phát hiện các trị đo có sai số lớn.

PHẦN 6 KẾT LUẬN

Xác suất thống kê không chỉ là một môn khoa học lý thuyết mà còn là bộ công cụ ứng dụng không thể thiếu, là nền tảng cho việc đảm bảo chất lượng và độ tin cậy trong mọi hoạt động của ngành Trắc địa. Báo cáo này đã trình bày một cách hệ thống và chi tiết hai trong số những ứng dụng quan trọng và cơ bản nhất:

- **Phân bố Student (t-distribution)** cung cấp một phương pháp chặt chẽ để ước lượng khoảng tin cậy cho giá trị trung bình của một đại lượng đo khi phương sai tổng thể chưa biết. Nó cho phép chúng ta chuyển từ một ước lượng điểm ($\bar{x}$) sang một ước lượng khoảng, đi kèm với một độ tin cậy định lượng, phản ánh đúng bản chất không chắc chắn của phép đo.
- **Phân bố Fisher-Snedecor (F-distribution)** và phân bố Chi-bình phương (χ^2) là những công cụ thiết yếu để đánh giá và so sánh độ chính xác của các kết quả đo. Chúng cho phép kiểm tra một cách khách quan xem một thiết bị có còn đạt tiêu chuẩn kỹ thuật hay không, hoặc liệu hai phương pháp đo có thực sự khác biệt về độ chính xác hay không, làm cơ sở cho việc lựa chọn công nghệ, quy trình và gán trọng số trong bình sai.

Việc nắm vững và áp dụng đúng đắn các công cụ này giúp các kỹ sư trắc địa không chỉ đưa ra những con số, mà còn cung cấp những kết quả khoa học hoàn chỉnh, có độ tin cậy cao và được đánh giá một cách định lượng, đáp ứng các yêu cầu ngày càng khắt

khe của các tiêu chuẩn kỹ thuật hiện đại.

Lĩnh vực Trắc địa đang không ngừng phát triển với sự ra đời của các công nghệ đo mới (GNSS, LiDAR, InSAR) và khối lượng dữ liệu ngày càng lớn. Điều này đặt ra yêu cầu về việc áp dụng các phương pháp thống kê ngày càng tinh vi và mạnh mẽ hơn. Các hướng nghiên cứu và ứng dụng đang được quan tâm bao gồm:

- **Phân tích phương sai (ANOVA - Analysis of Variance):** Là sự mở rộng của kiểm định F, cho phép so sánh đồng thời độ chính xác của nhiều hơn hai phương pháp, máy móc hoặc điều kiện đo, giúp tối ưu hóa quy trình công nghệ một cách toàn diện.
- **Hồi quy và Tương quan (Regression and Correlation):** Phân tích mối quan hệ phụ thuộc giữa các đại lượng đo khác nhau, ví dụ như sự ảnh hưởng của nhiệt độ đến kết quả đo dài, hoặc xây dựng các mô hình biến dạng công trình theo thời gian.
- **Thống kê Bayes (Bayesian Statistics):** Cung cấp một khung lý thuyết khác biệt để cập nhật kiến thức và độ tin cậy về một tham số khi có thêm dữ liệu mới. Phương pháp này đặc biệt hữu ích trong các ứng dụng định vị động và lọc Kalman.
- **Bình sai vững (Robust Estimation):** Như đã đề cập, đây là một lĩnh vực cực kỳ quan trọng để xử lý các tập dữ liệu lớn trong thực tế, vốn thường xuyên chứa các sai số thô và không hoàn toàn tuân theo các giả định lý tưởng.

Sự kết hợp giữa kiến thức trắc địa truyền thống và các phương pháp thống kê hiện đại sẽ tiếp tục là chìa khóa để nâng cao chất lượng và khai phá những tiềm năng mới trong khoa học đo đạc và xử lý thông tin không gian.

TÀI LIỆU THAM KHẢO

- [1] M. Poutanen and S. Rozsa, "The Geodesist's Handbook 2020," *J. Geod.*, vol. 94, Nov. 2020.
- [2] R. Chaudhari, "Chi-square Test for variance examples," 2024. [Online]. Available: <https://vrcacademy.com/tutorials/chi-square-test-variance-examples/>.
- [3] C. D. Ghilani, *Adjustment computations: spatial data analysis*. John Wiley & Sons, 2017.