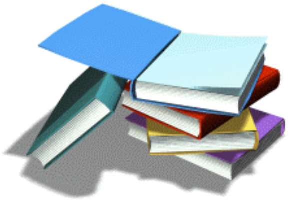
TRƯỜNG ĐẠI HỌC TÀI CHÍNH-MARKETING
KHOA CƠ BẢN
BỘ MÔN TOÁN – THỐNG KÊ

BÀI GIẢNG
LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN

Giảng viên
ThS. Lê Trường Giang



Chương 5

KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

Bài 1: Tổng quan bài toán kiểm định giả thiết thống kê

Bài 2: Kiểm định giả thiết về tham số của một tổng thể

Bài 3: Kiểm định giả thiết về tham số của hai tổng thể

Bài 1. Tổng quan bài toán kiểm định giả thiết thống kê

Bài toán 1. Xét bài toán sau, trọng lượng trung bình của mỗi sản phẩm được đóng gói tự động tại một nhà máy M là 50kg.

Biết rằng nếu quá trình đóng gói không tốt, trọng lượng sản phẩm cao hơn 50kg thì sẽ gây thiệt hại cho nhà sản xuất, ngược lại nếu trọng lượng thấp hơn 50kg sẽ làm mất khách hàng.

Sau một thời gian hoạt động, người ta cho rằng việc đóng gói sản phẩm của nhà máy M không còn tốt. Lấy ngẫu nhiên 100 sản phẩm do nhà máy M đóng gói thu được trọng lượng trung bình là 49,98kg với độ lệch chuẩn là 0,01kg. Vấn đề đặt ra là dựa trên mẫu ta cần phải đưa ra nhận xét, bác bỏ hay không bác bỏ nghi ngờ trên.

Bài 1. Tổng quan bài toán kiểm định giả thiết thống kê

Bài toán 2. Một loại thuốc A ban đầu có tỉ lệ chữa khỏi bệnh B là 90%.

Sau một thời gian, người ta nghi ngờ tỉ lệ chữa khỏi bệnh B của thuốc A đã giảm xuống dưới 90% do bệnh B đã kháng thuốc. Cho ngẫu nhiên 120 người mắc bệnh B chữa trị bằng thuốc A thấy có 15 người không khỏi bệnh. Vai trò của nhà thống kê là dựa trên mẫu quan sát để đưa ra quyết định rằng có bác bỏ nghi ngờ trên được không.

Bài 1. Tổng quan bài toán kiểm định giả thiết thống kê

1. Các khái niệm

a. Kiểm định giả thiết thống kê

Giả thiết thống kê được hiểu là một mệnh đề (hay một khẳng định) về tham số của tổng thể: kỳ vọng, tỉ lệ, phương sai, phân phối xác suất của tổng thể; tính độc lập giữa các biến ngẫu nhiên tổng thể.

Việc tìm ra một kết luận cuối cùng là bác bỏ hay chấp nhận giả thiết được nêu ra từ tổng thể được gọi là **kiểm định giả thiết thống kê**

Bài 1. Tổng quan bài toán kiểm định giả thiết thống kê

1. Các khái niệm

b. Bài toán kiểm định giả thiết thống kê

Trong bài toán kiểm định. Ta đặt ra cặp **giả thiết – đối thiết**

Giả thiết H_0 : mang nghĩa là không có sự thay đổi.

Đối thiết H_1 : là mệnh đề đối của giả thiết.

Bài toán kiểm định giả thiết thống kê được đặt ra như sau:

Dựa vào số liệu mẫu chọn được, với một độ tin cậy cho trước chúng ta cần khẳng định giả thiết đúng hay đối thiết đúng.

Bài 1. Tổng quan bài toán kiểm định giả thiết thống kê

1. Các khái niệm

b. Tiêu chuẩn kiểm định và giá trị quan sát

Tiêu chuẩn kiểm định giả thiết là một thống kê G phụ thuộc vào tham số đã biết trong giả thiết H0, sao cho nếu giả thiết đúng thì quy luật phân phối xác suất của G hoàn toàn được xác định.

Với số liệu mẫu cụ thể chọn được $\left(x_1, x_2, ..., x_n\right)$, ta tính được một giá trị cho thống kê đã chọn $G\left(x_1, x_2, ..., x_n\right)$. Kết quả tính được này gọi là **giá trị quan sát**.

Bài 1. Tổng quan bài toán kiểm định giả thiết thống kê

1. Các khái niệm

c. Nguyên lý kiểm định giả thiết

Nguyên tắc chung của kiểm định giả thiết thống kê là dựa trên nguyên lý xác suất nhỏ: *khi thực hiện một phép thử, một sự kiện có xác suất xuất hiện đủ bé thì coi như không xuất hiện*. Như vậy, chúng ta quyết định bác bỏ giả thiết nếu xác suất xuất hiện của một sự kiện quan sát được, tính trong điều kiện giả thiết đúng là nhỏ..

d. Miền bác bỏ

Xây dựng một miền W_α thỏa mãn điều kiện

$P\left(G \in \mathrm{W}_\alpha \middle| H_0 \text{ đúng}\right) = \alpha$ với α rất bé.

Miền W_α được coi là **miền bác bỏ** giả thiết H_0

Bài 1. Tổng quan bài toán kiểm định giả thiết thống kê

1. Các khái niệm

e. Sai lầm và mức ý nghĩa

Sai lầm loại 1 là sai lầm khi bác bỏ H_0 nhưng thực tế H_0 đúng. Xác suất sai lầm loại 1 cho bởi $P\left(G \in W_\alpha \middle| H_0 \text{ đúng}\right) = \alpha$.

Sai lầm loại 2 là sai lầm khi chấp nhận H_0 nhưng thực tế H_0 sai. Xác suất sai lầm loại 2 được cho bởi $P\left(G \in \overline{W}_\alpha \middle| H_0 \text{ sai}\right) = \beta$.

Khó có thể đồng thời giảm cả hai loại xác suất sai lầm này. Do đó ta cố định xác suất sai lầm loại 1 trước ở mức α rất bé và từ đó tìm miền bác bỏ giả thiết H_0 sao cho xác suất sai lầm loại 2 nhỏ nhất có thể.

Bài 1. Tổng quan bài toán kiểm định giả thiết thống kê

2. Các bước cơ bản của một phép kiểm định giả thiết

1. Đặt giả thiết H_0 và đối thiết H_1.

Tổng thể X có tham số θ chưa biết cần kiểm định

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases};\ \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases};\ \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}.$$

2. Xác định mức ý nghĩa α của phép kiểm định.

3. Chọn tiêu chuẩn kiểm định $G(X_1, X_2, \ldots, X_n)$ dựa trên mẫu.

4. Thiết lập miền bác bỏ giả thiết H_0 là W_α.

5. Kiểm định giả thiết.

Từ mẫu cụ thể tính giá trị kiểm định $g = G(x_1, x_2, \ldots, x_n)$, nếu $g \in W_\alpha$ thì bác bỏ giả thiết H_0, khi $g \notin W_\alpha$ thì không có cơ sở bác bỏ giả thiết H_0.

Bài 2. Kiểm định giả thiết về tham số của một tổng thể

1. Kiểm định giả thiết về trung bình tổng thể

Bài toán. Tổng thể X có kỳ vọng $E(X)=\mu$ chưa biết.

Ta cần kiểm định giả thiết $H_0:\mu=\mu_0$, các đối thiết có thể là

$$\left[\begin{array}{ll} H_1:\mu\neq\mu_0 & \textit{hai phía} \\ H_1:\mu>\mu_0 & \textit{phía phải} \\ H_1:\mu<\mu_0 & \textit{phía trái} \end{array}\right.$$

Thực hành tính toán kiểm định.

Bước 1. Nhận định trường hợp bài toán kiểm định, đặt cặp giả thiết

Bước 2. Với mức ý nghĩa α đã cho xác miền bác bỏ giả thiết W_α.

Bước 3. Dựa vào mẫu cụ thể tính $\overline{x}$, $\hat{s}$ từ đó tính giá trị kiểm định.

Bước 4. Kết luận.

Bài toán được giải theo các trường hợp cho trong bảng sau

T.H	Điều kiện	Chọn cặp giả thiết $H_0 : \mu = \mu_0$; H_1	Miền bác bỏ giả thiết H_0	Giá trị kiểm định
Biết σ	$n \geq 30$ hoặc $\begin{cases} n < 30 \\ X \; ppc \end{cases}$	$H_1 : \mu \neq \mu_0$	$W_\alpha = \left(-\infty, -z_{\frac{\gamma}{2}}\right) \cup \left(z_{\frac{\gamma}{2}}, +\infty\right)$	$z_{qs} = \dfrac{\left(\overline{x} - \mu_0\right)\sqrt{n}}{\sigma}$
		$H_1 : \mu > \mu_0$	$W_\alpha = \left(z_{\gamma-\frac{1}{2}}, +\infty\right)$	
		$H_1 : \mu < \mu_0$	$W_\alpha = \left(-\infty, -z_{\gamma-\frac{1}{2}}\right)$	
Chưa biết σ	$n \geq 30$	$H_1 : \mu \neq \mu_0$	$W_\alpha = \left(-\infty, -z_{\frac{\gamma}{2}}\right) \cup \left(z_{\frac{\gamma}{2}}, +\infty\right)$	$z_{qs} = \dfrac{\left(\overline{x} - \mu_0\right)\sqrt{n}}{s}$
		$H_1 : \mu > \mu_0$	$W_\alpha = \left(z_{\gamma-\frac{1}{2}}, +\infty\right)$	
		$H_1 : \mu < \mu_0$	$W_\alpha = \left(-\infty, -z_{\gamma-\frac{1}{2}}\right)$	
	$\begin{cases} n < 30 \\ X \;\; ppc \end{cases}$	$H_1 : \mu \neq \mu_0$	$W_\alpha = \left(-\infty, -t^{n-1}_{\frac{1-\gamma}{2}}\right) \cup \left(t^{n-1}_{\frac{1-\gamma}{2}}, +\infty\right)$	$z_{qs} = \dfrac{\left(\overline{x} - \mu_0\right)\sqrt{n}}{s}$
		$H_1 : \mu > \mu_0$	$W_\alpha = \left(t^{n-1}_{1-\gamma}, +\infty\right)$	
		$H_1 : \mu < \mu_0$	$W_\alpha = \left(-\infty, -t^{n-1}_{1-\gamma}\right)$	

Bài 2. Kiểm định giả thiết về tham số của một tổng thể

1. Kiểm định giả thiết về trung bình tổng thể

Ví dụ 1. Đo chiều cao (đơn vị cm) của 24 trẻ em 2 tuổi tại 1 huyện ta có số liệu:

84,4; 89,9; 89,0; 91,9; 87,0; 78,5; 84,5; 86,3; 80,6; 80,0; 81,3; 86,8; 83,4; 89,8; 85,4; 80,6; 85,0; 82,5; 80,7; 84,3; 95,4; 85,0; 85,5; 81,6

Biết chiều cao của trẻ em hai tháng tuổi chung của đất nước là biến ngẫu nhiên có phân phối chuẩn N (86, 5; 9, 67) . Hỏi với mức ý nghĩa 1% có sự khác biệt đáng kể về chiều cao trung bình của trẻ em huyện này so với chiều cao trung bình chung của đất nước không?

Bài 2. Kiểm định giả thiết về tham số của một tổng thể

1. Kiểm định giả thiết về trung bình tổng thể

Ví dụ 2. Một trại chăn nuôi gà đã nuôi thí nghiệm bằng khẩu phần thức ăn có bổ sung kháng sinh. Sau 8 tuần lễ nuôi, kiểm tra 81 con gà ta có số liệu:

Trọng lượng (kg)	3,8	3,9	4,0	4,1	4,2	4,3	4,4	4,5	4,6	4,7
Số gà	5	7	9	12	15	10	9	6	5	3

a) Trại chăn nuôi báo cáo trọng lượng trung bình của những con gà nuôi thí nghiệm sau 8 tuần nuôi là 4,3 kg thì có đúng không với độ tin cậy 95%?

b) Giả sử những con gà sau 8 tuần lễ nuôi có trọng lượng lớn hơn 4,3 kg được xếp loại I và trọng lượng của nó có phân phối chuẩn. Với mức ý nghĩa 5%, chúng ta có thể kết luận trọng lượng trung bình của những con gà loại I lớn hơn 4,5 kg được không?

Bài 2. Kiểm định giả thiết về tham số của một tổng thể

2. Kiểm định giả thiết về tỉ lệ tổng thể

Bài toán. Tổng thể gồm hai loại phần tử, phần tử có tính chất A và phần tử không có tính chất A. Trong đó, tỉ lệ phần tử có tính chất A là tham số p chưa biết. Với mức ý nghĩa α ta cần kiểm định một trong các cặp giả thiết – đối thiết (H_0, H_1)

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}; \begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}; \begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases}.$$

Bài 2. Kiểm định giả thiết về tham số của một tổng thể

2. Kiểm định giả thiết về tỉ lệ tổng thể

Điều kiện	Chọn cặp giả thiết $H_0 : p = p_0;\ H_1$	Miền bác bỏ giả thiết H_0	Giá trị kiểm định
$np_0 \geq 5$ $n(1 - p_0) \geq 5$	$H_1 : p \neq p_0$ $H_1 : p > p_0$ $H_1 : p < p_0$	$W_\alpha = \left(-\infty, -z_{\frac{\gamma}{2}}\right) \cup \left(z_{\frac{\gamma}{2}}, +\infty\right)$ $W_\alpha = \left(z_{\gamma-\frac{1}{2}}, +\infty\right)$ $W_\alpha = \left(-\infty, -z_{\gamma-\frac{1}{2}}\right)$	$z_{qs} = \dfrac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}}$

Bài 2. Kiểm định giả thiết về tham số của một tổng thể

1. Kiểm định giả thiết về trung bình tổng thể

Ví dụ 3. Ở một nước, một đảng chính trị tuyên bố rằng 45% cử tri sẽ bỏ phiếu bầu cho ông A là ứng cử viên của họ. Chọn ngẫu nhiên 200 người hỏi ý kiến có 80 người sẽ bầu cho ông A. với mức ý nghĩa 5% hãy cho nhận xét về tuyên bố trên.

Ví dụ 4. Giả sử một huyện năm trước có tỷ lệ trẻ em bị suy dinh dưỡng là 10%, năm nay huyện thực hiện nhiều chính sách nhằm làm giảm tỷ lệ này xuống. chọn 400 đứa trẻ, kiểm tra ta thấy có 32 đứa trẻ vẫn còn bị suy dinh dưỡng. với mức ý nghĩa 1% hãy cho kết luận về việc giảm tỷ lệ trẻ em suy dinh dưỡng của huyện này.

3. Kiểm định trung bình hai tổng thể độc lập (So sánh trung bình của hai tổng thể đ l)

Hai tổng thể X, Y độc lập có kì vọng lần lượt là μ_X, μ_Y chưa biết, cần kiểm định giả thiết về so sánh μ_X và μ_Y

$H_0: \theta_X = \theta_Y \mid H_1: \theta_X \neq \theta_Y$ gọi là kiểm định hai phía;

$H_0: \theta_X = \theta_Y \mid H_1: \theta_X > \theta_Y$ gọi là kiểm định một phía phải;

$H_0: \theta_X = \theta_Y \mid H_1: \theta_X < \theta_Y$ gọi là kiểm định một phía trái.

Bài toán được giải theo từng trường hợp sau:

TH	Điều kiện	Chọn cặp giả thiết $H_0: \mu_X = \mu_Y$; H_1	Miền bác bỏ giả thiết H_0	Giá trị kiểm định
Biết σ_X, σ_Y	$\begin{cases} n_X \geq 30 \\ n_Y \geq 30 \end{cases}$	$H_1: \mu_X \neq \mu_Y$ $H_1: \mu_X > \mu_Y$ $H_1: \mu_X < \mu_Y$	$W_\alpha = \left(-\infty, -z_{\frac{\gamma}{2}}\right) \cup \left(z_{\frac{\gamma}{2}}, +\infty\right)$ $W_\alpha = \left(z_{\gamma-1/2}, +\infty\right)$ $W_\alpha = \left(-\infty, -z_{\gamma-1/2}\right)$	$z_{qs} = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}}$
	$\begin{cases} n_X < 30 \\ n_Y < 30 \\ X;Y\ ppc \end{cases}$			
Chưa biết σ_X, σ_Y	$\begin{cases} n_X \geq 30 \\ n_Y \geq 30 \end{cases}$	$H_1: \mu_X \neq \mu_Y$ $H_1: \mu_X > \mu_Y$ $H_1: \mu_X < \mu_Y$	$W_\alpha = \left(-\infty, -z_{\frac{\gamma}{2}}\right) \cup \left(z_{\frac{\gamma}{2}}, +\infty\right)$ $W_\alpha = \left(z_{\gamma-1/2}, +\infty\right)$ $W_\alpha = \left(-\infty, -z_{\gamma-1/2}\right)$	$z_{qs} = \dfrac{\overline{x} - \overline{y}}{\sqrt{\dfrac{s_X^2}{n_X} + \dfrac{s_Y^2}{n_Y}}}$
	$\begin{cases} n_X < 30 \\ n_Y < 30 \\ X;Y\ ppc \\ \sigma_X \approx \sigma_Y \end{cases}$ (Note: $n_X + n_Y - 2 < 30$)	$H_1: \mu_X \neq \mu_Y$ $H_1: \mu_X > \mu_Y$ $H_1: \mu_X < \mu_Y$	$W_\alpha = \left(-\infty, -t_{\frac{1-\gamma}{2}}^{n_X+n_Y-2}\right) \cup \left(t_{\frac{1-\gamma}{2}}^{n_X+n_Y-2}, +\infty\right)$ $W_\alpha = \left(t_{1-\gamma}^{n_X+n_Y-2}, +\infty\right)$ $W_\alpha = \left(-\infty, -t_{1-\gamma}^{n_X+n_Y-2}\right)$	$z_{qs} = \dfrac{\overline{x} - \overline{y}}{\sqrt{s^2\left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$
Lưu ý:	$s^2 = \dfrac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}$			

Bài 2. Kiểm định giả thiết về tham số của hai tổng thể

Ví dụ 5. So sánh mức thu nhập theo tuần giữa nam và nữ tại một công ty liên doanh ta có số liệu mẫu như sau:

– Nữ: chọn một mẫu 40 người, tính được thu nhập trung bình .

– Nam: chọn một mẫu 50 người, tính được thu nhập trung bình .

Biết rằng phương sai thu nhập theo tuần của nữ là 80 và của nam là 100. Với mức ý nghĩa 1%, có thể kết luận thu nhập trung bình của nữ thấp hơn nam được không?

Ví dụ 6. Khảo sát chiều cao (đơn vị cm) của học sinh nữ tại hai trường phổ thông trung học huyện A và huyện B ta có số liệu:

CC	150-152	152-154	154-156	156-158	158-160	160-162	162-164	164-166	166-168	168-170
A	3	5	7	15	26	25	12	13	10	5
B	5	10	14	18	22	11	9	5	4	2

a) Với mức ý nghĩa 1% có thể xem chiều cao trung bình học sinh trung học nữ của huyện A cao hơn huyện B được không?

b) Những học sinh có chiều cao từ 154 cm trở xuống được xem là nhóm thấp. giả sử chiều cao học sinh nhóm thấp ở hai huyện là biến ngẫu nhiên có phân phối chuẩn có phương sai xấp xỉ bằng nhau. Một người nói chiều cao trung bình học sinh nhóm thấp của hai huyện là như nhau thì có đúng không với độ tin cậy là 95%.

4. Kiểm định tỉ lệ hai tổng thể độc lập (So sánh tỉ lệ của hai tổng thể độc lập)

Hai tổng thể X, Y có tỉ lệ phần tử tính chất A là p_X, p_Y chưa biết, cần kiểm định giả thiết về so sánh p_X và p_Y

$H_0: p_X = p_Y \mid H_1: p_X \neq p_Y$ gọi là kiểm định hai phía;

$H_0: p_X = p_Y \mid H_1: p_X > p_Y$ gọi là kiểm định một phía phải;

$H_0: p_X = p_Y \mid H_1: p_X < p_Y$ gọi là kiểm định một phía trái.

Bài toán được giải như sau:

Bài 2. Kiểm định giả thiết về tham số của hai tổng thể

4. Kiểm định tỉ lệ hai tổng thể độc lập
(So sánh tỉ lệ của hai tổng thể độc lập)

Điều kiện	Chọn cặp giả thiết $H_0: p_X = p_Y$; H_1	Miền bác bỏ giả thiết H_0	Giá trị kiểm định
$n_X; n_Y$ đủ lớn	$H_1: p_X \neq p_Y$	$W_\alpha = \left(-\infty, -z_{\frac{\gamma}{2}}\right) \cup \left(z_{\frac{\gamma}{2}}, +\infty\right)$	$z_{qs} = \dfrac{f_X - f_Y}{\sqrt{f(1-f)\left(\dfrac{1}{n_X} + \dfrac{1}{n_Y}\right)}}$
	$H_1: p_X > p_Y$	$W_\alpha = \left(z_{\gamma-1/2}, +\infty\right)$	
	$H_1: p_X < p_Y$	$W_\alpha = \left(-\infty, -z_{\gamma-1/2}\right)$	
Lưu ý:	$f = \dfrac{n_X f_X + n_Y f_Y}{n_X + n_Y} = \dfrac{n_A}{n}$, $n = n_X + n_Y$, n_A là số phần tử tính chất A của 2 tổng thể		

Bài 2. Kiểm định giả thiết về tham số của hai tổng thể

Ví dụ 7. Kiểm tra 100 đứa trẻ của vùng I phát hiện 42 đứa trẻ bị sâu răng, vùng II có 92 đứa trẻ bị sâu răng khi kiểm tra 200 đứa trẻ. Với mức ý nghĩa 5% có thể xem tỷ lệ trẻ bị sâu răng ở 2 vùng bằng nhau được không?

Ví dụ 8. Kiểm tra chất lượng sản phẩm về một loại hàng do hai nhà máy A và B sản xuất cho kết quả : trong 500 sản phẩm của A có 50 phế phẩm và trong 400 sản phẩm của B có 60 phế phẩm. với mức ý nghĩa 5%, hãy xem chất lượng sản phẩm của A có tốt hơn B không ?

XIN CHÂN THÀNH CẢM ƠN!