TRƯỜNG ĐẠI HỌC FPT HÀ NỘI

FPT UNIVERSITY

COMPUTER PROJECT – MAS291

XÁC SUẤT THỐNG KÊ

Đề tài: Phân tích thống kê GDP của Mỹ và Nhật Bản

Lớp: SE1606

Giảng viên hướng dẫn: Nguyễn Việt Anh

Nhóm sinh viên:

- Trịnh Hoàng Anh
- Lưu Doãn Dưỡng
- Lê Thanh Tùng
- Phạm Quốc Hưng

Hà Nội, ngày 19 tháng 03 năm 2022

MỤC LỤC

1. XÁC ĐỊNH ĐỀ TÀI

1.1 Đặt vấn đề

- Trong thời gian gần đây, cuộc chiến giữa Nga và Ukraine đang nhận nhiều sự chú ý của mọi người không chỉ từ người dân hai ngước mà của nhiều quốc gia trên thế giới.
- Dù muốn dù không, một cuộc chiến giữa hai nước sẽ ảnh hưởng không nhỏ đến kinh tế thị trường trên toàn cầu.
- Ví dự như kể từ khi Nga phát động cuộc chiến tranh với Ukraine thì hầu hết giá cả các loại hàng hoá đều tăng chóng mặt, trong đó có những mặt hàng ảnh hưởng trực tiếp tới ngành công nghiệp sản xuất thang máy như sắt thép, dầu mỏ, inox,…
- Đều này làm chúng ta quan tâm nhiều hơn đến vấn đề kinh tế ảnh hường như thế nào, hãy cùng khảo sát tình hình phát triển tốc độ tăng trưởng kinh tế của các cường quốc lớn trên thế giới qua từng năm.

1.2 Nội dung đề tài

- Nên nhóm chọn chủ đề GDP để thể hiện sự tăng trưởng kinh tế của các nước qua 60 năm từ 1961 đến 2020. Ngoài ra nhóm tập trung thu thập dữ liệu, thống kê và phân tích các dữ liệu GDP của hai nước Mỹ và Nhật Bản.
- Kinh tế Mỹ là một nền kinh tế tư bản chủ nghĩa hỗn hợp với kỹ nghệ, mức độ công nghiệp hóa và trình độ phát triển cao.
- Đối với nền kinh tế của Nhật Bản là một nền kinh tế thị trường tự do phát triển. Nhật Bản là nền kinh tế lớn thứ ba thế giới theo GDP danh nghĩa và lớn thứ tư theo sức mua tương đương (PPP), ngoài ra Nhật Bản là nền kinh tế lớn thứ hai trong số các nước phát triển.

1.3 Lý do chọn đề tài và bài học rút ra

- Cuộc chiến giữa Nga và Ukraine đã có những biến động lớn trên thế giới khi giá cả các mặt hàng đều tăng cao nên sẽ mang lại những ảnh hưởng đến GDP.
- Bởi vì Mỹ không chỉ là một nền kinh tế phát triển mà còn là nền kinh tế lớn nhất trên thế giới theo giá trị GDP danh nghĩa (Nominal) và lớn thứ hai thế giới tính theo ngang giá sức mua (PPP).
- Bên cạnh đó nền kinh tế của Mỹ có nhiều biến động nhưng không quá lớn như siêu lạm phát (Zimbabwe,...), nên vẫn giữ được một nền kinh tế ổn định qua hàng năm.
- Đối với nền kinh tế Nhật Bản cũng là một nước có GDP cao trên thế giới và cũng có những khoảng thời gian xảy ra khủng hoảng và vẫn giữ vị thế của mình trên thế giới.
- Ngoài ra dữ liệu GDP của hai nước được ghi chép lại nhiều, thông tin đầy đủ, chính xác và chi tiết qua từng năm, cũng bởi vì Mỹ và Nhật Bản là hai cường quốc trên thế giới.
- Đề tài này còn mang lại giúp các thành viên trong nhóm phát triển và rèn luyện tư duy thống kê trong nghiên cứu từ các bài toán thực tế.
- Áp dụng các môn đã học và ở đây là môn xác suất thống kê để đưa ra các thông tin đến người dùng và đem lại lợi ích cho xã hội với những vấn đề quan trọng và được quan tâm rộng rãi.

2. NGUỒN DỮ LIỆU

2.1 Thông tin về các nguồn dữ liệu

2.1.1 The World Bank (Ngân hàng thế giới)

- Là một tổ chức tài chính quốc tế nơi cung cấp những khoản vay nhằm thúc đẩy kinh tế cho các nước đang phát triển thông qua các chương trình vay vốn.
- Cũng là opendata hàng đầu với nhiều lĩnh vực như kinh tế, chính trị, xã hội..
- Được thành lập tại hội nghị Bretton Woods năm 1944 cùng 3 tổ chức khác trong đó có Quỹ Tiền tệ Quốc tế (IMF). Cả WB và IMF đều có trụ sở tại Washington DC, và có mối quan hệ gần với nhau.

2.1.2 Link nguồn dữ liệu

- https://data.worldbank.org/

2.1.3 Dữ liệu thu thập được

- GDP của hai nước Mỹ và Nhật Bản trong khoảng thời gian từ 1961 đến 2020.

3. THỐNG KÊ MÔ TẢ

3.1 Số liệu tính toán được

Đặc điểm	Mỹ	Nhật bản
Mean	2.93	3.489666667
Standard Error	0.283224472	0.510850181
Median	3.09	2.955
Mode	3.46	1.37
Standard Deviation	2.193847329	3.957028489
Sample Variance	4.812966102	15.65807446
Kurtosis	0.661164083	0.44350882
Skewness	-0.680158306	0.575467004
Range	10.64	18.57
Minimum	-3.4	-5.69
Maximum	7.24	12.88
Sum	175.8	209.38
Count	60	60

3.2 Cách tính số liệu bằng Excel

- Chọn vùng dữ liệu cần thống kê mô tả. Vào data -> Data Analysis.

3.3 Phân tích số liệu thống kê

- Mean
- Là giá trị trung bình của một tập hợp gồm hai hoặc nhiều số, được tính bằng tổng các giá trị quan sát chia cho số quan sát.

- Giá trị trung bình có đặc điểm là chịu sự tác động của các giá trị ở mỗi quan sát, do đó đây là thang đo nhạy cảm nhất đối với sự thay đổi của các giá trị quan sát.
- Số lượng trung bình GDP tại Mỹ và Nhật Bản lần lượt là xấp xỉ bằng 2.93% và 3.49%

- Sample Variance (Phương sai)

- Phương sai của một bảng số liệu là số đặc trưng cho độ phân tán của các số liệu so với số trung bình của nó.
- Phương sai dùng để đo độ phân tán của một tập các giá trị quan sát xung quanh giá trị trung bình của tập quan sát đó.
- Cụ thể phương sai của GDP của Mỹ và Nhật Bản lần lượt là gần bằng 4.81% và 15.66%.

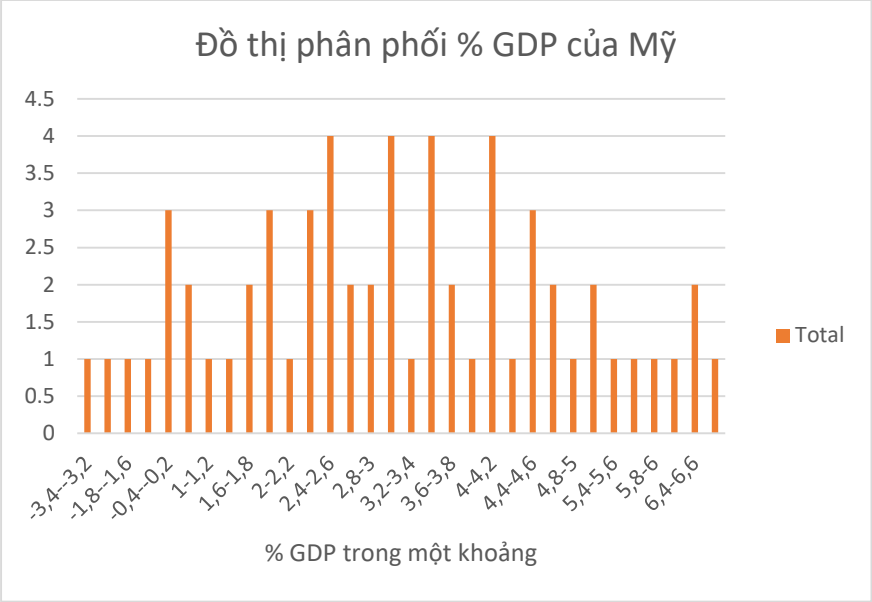
- Standard Deviation (Độ lệch chuẩn)

- Độ lệch chuẩn cũng dùng để đo lường độ phân tán của dữ liệu xung quanh giá trị trung bình của nó.
- Độ lệch chuẩn chính bằng căn bậc hai của phương sai.
- Việc khảo sát phương sai thường cho các giá trị rất lớn, do đó sử dụng độ lệch chuẩn sẽ giúp dễ dàng cho việc diễn giải do các kết quả sai biệt đưa ra sát với dữ liệu gốc.
- Ở đây độ lệch chuẩn GDP của Mỹ và Nhật Bản lần lượt là khoảng 2.19% và 3.96%.

- Max, Min (Giá tị lớn nhất và thấp nhất)

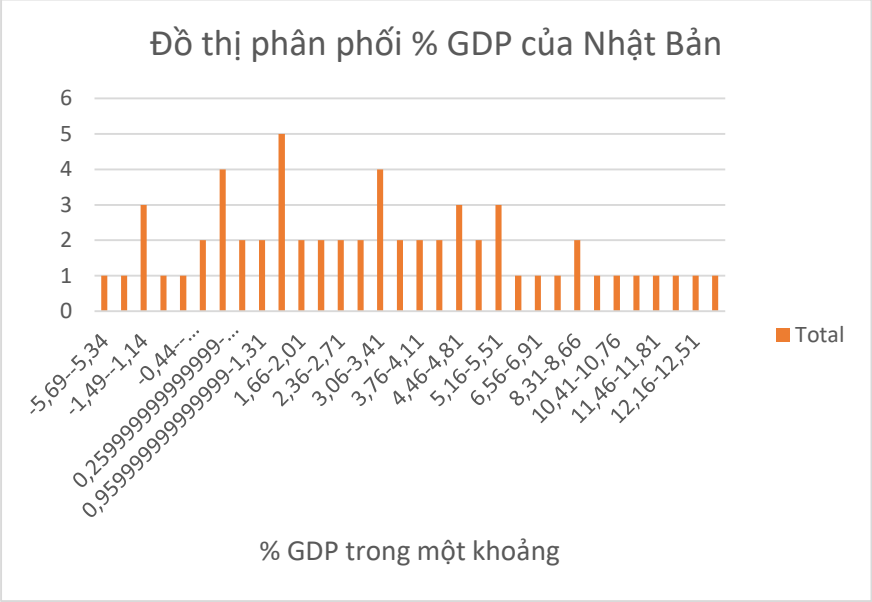
- Thể hiện các giá trị lớn nhất, nhỏ nhất của các đối số hay vùng dữ liệu.
- Giá trị GDP cao nhất của Mỹ là 7.24% còn của Nhật Bản là 12.88%.
- Giá trị GDP thấp nhất của Mỹ là -3.4% còn của Nhật Bản là -5.69%.

3.4 Đồ thị phân phối



Nhận xét:

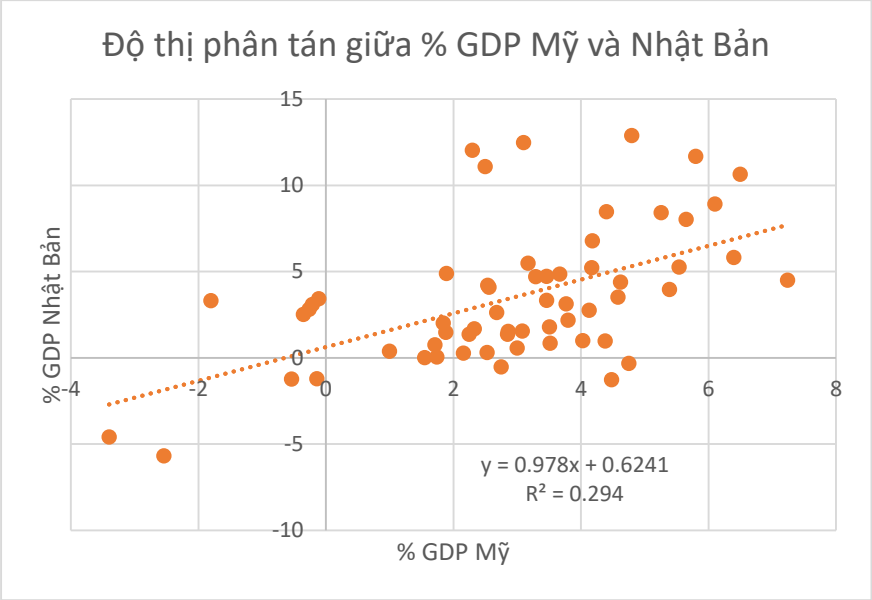
- Biểu đồ thể hiện phân phối phần trăm GDP của Mỹ
- Có thể thấy dữ liệu có nhiều biến động nhưng chủ yếu vẫn lớn hơn một lần xuất hiện và tập trung nhiều nhất vào 4 khoảng dữ liệu trong khoảng từ 2,4 đến 4,2 .
- Hình dạng biểu đồ thuộc dạng phân phối chuẩn khi dữ liệu tập trung ở giữa và thu hẹp ở hai phía.



Nhận xét:

- Từ biểu đồ có thể thấy %GDP của Nhật Bản đều có số lần từ 1 trở lên.
- GDP của Nhật Bản tập trung lớn nhất trong khoảng 1,66% đến 2,01%.
- Hình dạng biểu đồ cũng gần giống với thuộc dạng phân phối chuẩn tuy nhiên dữ liệu tập trung ở bên trái biểu đồ.

3.5 Đồ thị phân tán

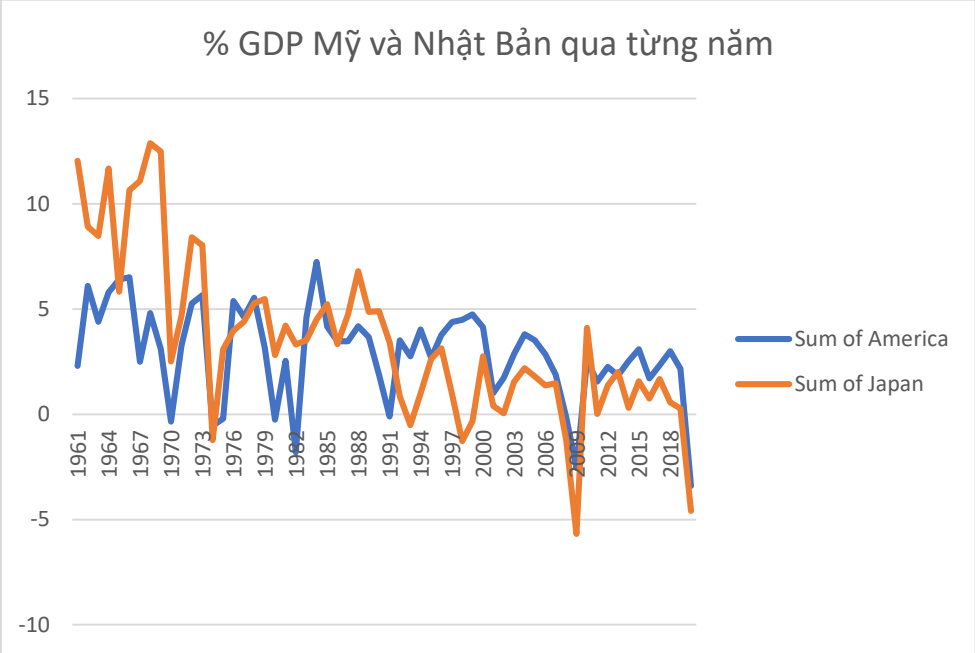


Nhận xét:

- Biểu đồ phân tán cho chúng ta biết mối quan hệ giữa % GDP Nhật Bản và % GDP của Mỹ.
- Hình mẫu dữ liệu dạng tuyến tính và theo bờ dốc dương và hệ số tương quan (correlation coefficient): r = $\sqrt{0.294}$= 0.542 > 0 không tiến gần -1 hoặc 1 nên là Weak positive correlation, có thể thấy chủ yếu là %GDP của Mỹ tăng sẽ chưa chắc dẫn đến của Nhật Bản cũng sẽ tăng.
- Các điểm phân tán rộng, mối quan hệ yếu. Cho nên là sự thay đổi %GDP của Mỹ tăng không dẫn đến %GDP Nhật Bản tăng mà còn phụ thuộc vào một số yếu tố khác.

3.6 Các giả thuyết về kiểm định giả thuyết thống kê

Bài toán 1: Kiểm tra giả thuyết cho rằng phần trăm GDP của Mỹ tăng thì đồng nghĩa kéo theo phần trăm GDP của Nhật Bản cũng tăng lên?

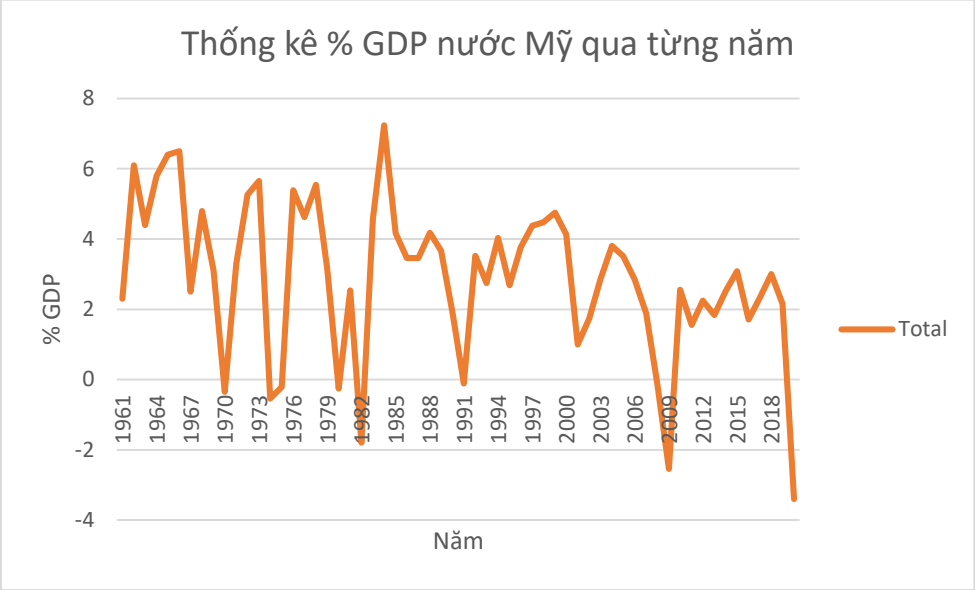


Biểu đồ thể hiện tốc độ tăng trưởng GDP của Mỹ và Nhật Bản qua từng năm

- Từ biểu đồ ta thấy trước năm 1965, tốc độ tăng trưởng GDP của hai nước trái ngược nhau khi Mỹ chiếm khoảng 2.5% thì Nhật Bản chiếm đến hơn 12% cụ thể vào năm 1961.
- Nhưng sau khoảng thời gian đó đến này thì tình trạng biến động của Mỹ và Nhật Bản dường như là giống nhau khi có xu hướng tăng giảm về phần trăm GDP kinh tế.
- Tuy nhiên khi có những cuộc khủng hoảng xảy ra, thì nền tăng trưởng của Nhật Bản lại có sự tụt dốc nhanh và nhiều hơn của Mỹ nhưng lại lấy lại sự ổn định nhanh hơn.
- Vậy ta có thể khẳng định được, phần trăm tăng trưởng GDP của Mỹ tăng lên thì cũng kéo một phần phần trăm GDP của Nhật Bản cũng tăng lên.

Bài toán 2: Nhiều người cho rằng khi các cuộc lạm phát xảy ra hay những biến động lớn như chính trị, chiến tranh, dân số… sẽ đem lại ảnh hưởng đến tốc độ

tăng trưởng kinh tế của Mỹ và đặc biệt là tụt dốc trong tốc độ tăng trưởng, điều này có thực sự đúng?

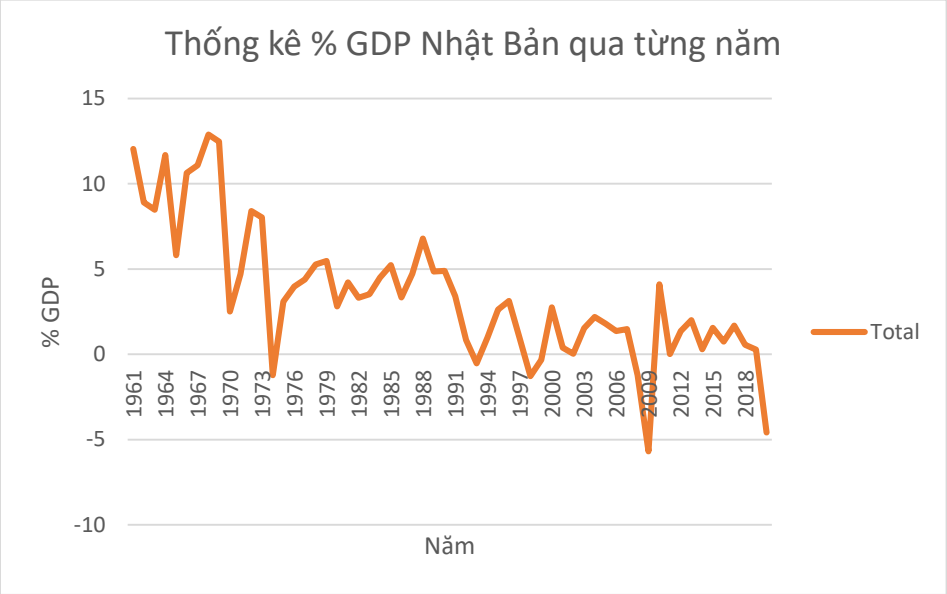


Biểu đồ thể hiện tốc độ tăng trưởng GDP của Mỹ qua từng năm

- Nhìn vào biểu đồ ta thấy qua từng năm tốc độ tăng trưởng GDP của Mỹ có nhiều thay đổi nhưng sau những lần khủng hoảng đều tăng trường trở lại.
- Nước Mỹ xảy ra các cuộc tụt dốc về nền kinh tế nặng nhất là vào các năm 1982, 2009, và 2018.
- Có thể thấy tại Mỹ xuất hiện lạm phát cao nhất từ năm 1982 cũng là lí do dẫn đến tụt dốc về GDP của Mỹ.
- Từ năm 2007 đến năm 2009 tại Mỹ xảy ra cuộc khủng hoảng tài chính nặng nề nên đã ảnh hướng lớn đến tốc độ tăng trưởng kinh tế của đất nước.
- Ngoài ra nước Mỹ đạt được phần trăm GDP lớn nhất khoảng 7% vào năm 1980.

Bài toán 3: Nhiều người cho rằng khi tại Nhật Bản xảy ra các cuộc khủng hoảng năng lượng, khủng hoảng hạt nhân, các cuộc lạm phát xảy ra hay những

biến động lớn như chính trị, chiến tranh, dân số… sẽ làm giảm tốc độ tăng trưởng kinh tế, điều này có thực sự đúng?



Biểu đồ thể hiện tốc độ tăng trưởng GDP của Nhật Bản qua từng năm

- Quan sát biểu đồ ta thấy mặc dù có nhiều biến động trong nền kinh tế của Nhật Bản và những lúc vực dậy sau những lúc khủng hoảng. Nhưng sau năm 1973 tốc độ tăng trưởng đã dưới 7%.
- Đặc biệt Nhật Bản xảy ra các cuộc khủng hoảng nghiêm trọng vào khoảng những năm 1973, 2009, 2018.
- Từ 1973, do tác động khủng hoảng năng lượng, kinh tế Nhật thường khủng hoảng và suy thoái ngắn nên đã dẫn đến sự tụt dốc về nền kinh tế tại Nhật Bản trong thời gian đó.
- Năm 2018, dân số của Nhật Bản giảm mạnh khoảng 260.000 người so với năm trước cũng là lí do ảnh hưởng đến sự phát triển của GDP kinh tế của Nhật Bản.

4. CÁC DẠNG TOÁN

4.1 Kiểm định giả thuyết và khoảng tin cậy của giá trị trung bình (μ) của tổng thể.

4.1.1 Khoảng tin cậy của giá trị trung bình (μ) của tổng thể với phương sai (σ^2) đã biết.

a) Lý thuyết

Nếu $\bar{x}$ là trung bình mẫu của một mẫu ngẫu nhiên có kích thước n từ một quần thể phân phối chuẩn với phương sai đã biết, (1 - α) – khoảng tin cậy trên μ được cho bởi:

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

• Giới hạn tin cậy trên (1 - α) đối với μ là: $\mu \le \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}$

• Giới hạn tin cậy dưới (1 - α) đối với μ là: $\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}} \le \mu$

b) Bài toán

Dựa trên dữ liệu thu thập được và số liệu thống kê được tính toán về tốc độ tăng trường GDP của Nhật Bản từ năm 1961 đến năm 2020, ta có độ lệch chuẩn được biết là 3.957. Trong một mẫu ngẫu nhiên của 20 năm (từ 2001 – 2020), GDP tăng trưởng trung bình được tìm thấy là 0.4965.

Hãy xác định khoảng tin cậy 95% của μ. 95% giới hạn độ tin cậy trên của μ. 95% giới hạn độ tin cậy dưới của μ.

Lời giải

Ta có $\bar{x} = 0.4965, \sigma = 3.957$ (dựa vào số liệu đã cho). α = 1- 0.95 = 0.05

⇒ $z_{\alpha/2} = z_{0.025} = 1.96$, $z_{\alpha} = z_{0,05} = 1.645$.

Khoảng tin cậy 95% của μ: $0.4965 - 1.96\frac{3.957}{\sqrt{20}} \le \mu \le 0.4965 + 1.96\frac{3.957}{\sqrt{20}}$ <-> $-1.238 \le \mu \le 2.231$

Giới hạn khoảng tin cậy 95% trên của μ: $\mu \le 0.4965 + 1.645\frac{3.957}{\sqrt{20}}$ <-> $\mu \le 1.952$

Giới hạn khoảng tin cậy 95% dưới của μ: $0.4965 - 1.645\frac{3.957}{\sqrt{20}} \le \mu <-> -0.959 \le \mu$

4.1.2 Khoảng tin cậy của giá trị trung bình (μ) của tổng thể với phương sai (σ^2) chưa biết.

a) Lý thuyết

Gọi $X_1, X_2, \ldots, X_n$ là một mẫu ngẫu nhiên từ phân phối chuẩn với giá trị trung bình μ chưa biết và phương sai σ^2 chưa biết.

Biến ngẫu nhiên:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

có phân phối t với n - 1 bậc tự do.

Nếu $\bar{x}$ và s là giá trị trung bình và độ lệch chuẩn của một mẫu ngẫu nhiên từ phân phối chuẩn với phương sai chưa biết σ^2

• Khoảng tin cậy phần trăm (1 - α) trên μ được cho bởi: $\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$

• Giới hạn tin cậy trên (1 - α) đối với μ là: $\mu \le \bar{x} + t_{\alpha,n-1}\frac{s}{\sqrt{n}}$

• Giới hạn tin cậy dưới (1 - α) đối với μ là: $\bar{x} - t_{\alpha,n-1}\frac{s}{\sqrt{n}} \le \mu$

b) Bài toán

Cục Dự trữ Liên bang Hòa Kỳ (viết tắt FED) muốn khảo sát sự thay đổi của GDP của Hoa Kỳ trong vài năm qua. Biết rằng tổng thể là một phân phối chuẩn, các giá trị khảo sát được đưa ra như sau:

2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1.55	2.25	1.84	2.53	3.08	1.71	2.33	3.00	2.16	-3.4

Hãy xác định khoảng tin cậy 95% của μ. 95% giới hạn trên độ tin cậy trên của μ. 95% giới hạn độ tin cậy dưới của μ.

Lời giải:

Ta có: $\bar{x} = \frac{1.55 + 2.25 + 1.84 + 2.53 + 3.08 + 1.71 + 2.33 + 3.00 + 2.16 - 3.4}{10} = 1.705$

$$s^2 = \frac{(1.55-\bar{x})^2 + (2.25-\bar{x})^2 + (1.84-\bar{x})^2 + (2.53-\bar{x})^2 + (3.08-\bar{x})^2 + (1.71-\bar{x})^2 + (2.33-\bar{x})^2 + (3.00-\bar{x})^2 + (2.16-\bar{x})^2 + (-3.4-\bar{x})^2}{10-1} = 3.472$$

Do đó: s = 1.86, α = 1- 0.95 = 0.05; n = 10 $\Rightarrow t_{\alpha/2,n-1} = t_{0.025,9} = 2.262$; $t_{\alpha,n-1} = t_{0.05,9} = 1.833$

Khoảng tin cậy 95% của μ: $1.705 - 2.262\frac{1.86}{\sqrt{10}} \le \mu \le 1.705 + 2.262\frac{1.86}{\sqrt{10}}$ <-> $0.375 \le \mu \le 3.035$

Giới hạn khoảng tin cậy 95% trên của μ: $\mu \le 1.705 + 1.833\frac{1.86}{\sqrt{10}}$ <-> $\mu \le 2.783$

Giới hạn khoảng tin cậy 95% dưới của μ: $1.705 - 1.833\frac{1.86}{\sqrt{10}} \le \mu$ <-> $0.627 \le \mu$

4.2 Khoảng tin cậy cho phương sai (σ^2) của một phân phối chuẩn

a) Lý thuyết

• Khoảng tin cậy (1 - α) phần trăm trên σ^2 được cho bởi: $\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}$

• Giới hạn tin cậy trên (1 - α) phần trăm trên σ^2: $\sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}}$

• Giới hạn tin cậy dưới (1 - α) phần trăm trên σ^2: $\frac{(n-1)s^2}{\chi^2_{\alpha,n-1}} \le \sigma^2$

b) Bài toán

Cục Dự trữ Liên bang Hòa Kỳ (viết tắt FED) sau khi lấy 10 mẫu khảo sát sự thay đổi GDP của Mỹ từ năm 2011 tới năm 2020 thì tính toán được phương sai mẫu khảo sát là 3.472.

Dựa vào số liệu đó hãy ước lượng 95% khoảng tin cậy phương sai của tổng thể GDP Mỹ, 95% giới hạn tin cậy trên của tổng thể, 95% giới hạn tin cậy dưới của tổng thể. Biết rằng tổng thể này là xấp xỉ phân phối chuẩn.

Lời giải

Ta có phương sai mẫu $s^2 = 3.472$, $\alpha = 1 - 0.95 = 0.05$; $n = 10 \Rightarrow \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,9} = 2.7$,

$\chi^2_{\alpha/2,n-1} = \chi^2_{0.025,9} = 19.023$, $\chi^2_{1-\alpha,n-1} = \chi^2_{0.95,9} = 3.325$, $\chi^2_{\alpha,n-1} = \chi^2_{0.05,9} = 16.919$

Khoảng tin cậy 95% trên σ^2: $\frac{(10-1)\times 3.472}{19.023} \le \sigma^2 \le \frac{(10-1)\times 3.472}{2.7}$ <-> $1.643 \le \sigma^2 \le 11.573$

Giới hạn khoảng tin cậy 95% trên của σ^2: $\sigma^2 \le \frac{(10-1)\times 3.472}{3.325}$ <-> $\sigma^2 \le 9.398$

Giới hạn khoảng tin cậy 95% dưới của σ^2: $\frac{(10-1)\times 3.472}{16.919} \le \sigma^2$ <-> $1.847 \le \sigma^2$

4.3 Khoảng tin cậy cho tỉ lệ tổng thể (p) với mẫu lớn

a) Lý thuyết

Gọi p là một ước lượng điểm cho tỷ lệ p của tổng thể dựa trên một mẫu ngẫu nhiên có kích thước n, một giá trị xấp xỉ (1 - α) - khoảng tin cậy trên p là:

$$p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le p \le p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

• Giới hạn tin cậy trên (1 - α) phần trăm trên p: $p \le p + z_{\alpha}\sqrt{\frac{p(1-p)}{n}}$

• Giới hạn tin cậy dưới (1 - α) phần trăm trên p: $p - z_{\alpha}\sqrt{\frac{p(1-p)}{n}} \le p$

b) Bài toán

Cục Dự trữ Liên bang Hòa Kỳ (viết tắt FED) sau khi lấy 40 mẫu ngẫu nhiên khảo sát sự thay đổi GDP của Mỹ, tổ chức nhận thấy rằng có 5 năm mà GDP của Mỹ đạt dưới mức âm cho phép.

Dựa vào số liệu đó hãy ước lượng 95% khoảng tin cậy tỉ lệ của tổng thể GDP Mỹ mà những năm GDP đặt dưới mức âm cho phép, 95% giới hạn tin cậy trên của tỉ lệ tổng thể, 95% giới hạn tin cậy dưới của tỉ lệ tổng thể. Biết rằng tổng thể này là xấp xỉ phân phối chuẩn.

Lời giải

Ta có tỉ lệ tổng thể p = 5/40 = 0.125, α = 1- 0.95 = 0.05; α/2=0.025; n = 10; $z_{0,05}$ = 1.645; $z_{0,025}$ = 1.96

Khoảng tin cậy 95% trên p: $0.125-1.96\sqrt{\frac{0.125(1-0.125)}{40}} \le p \le 0.125+1.96\sqrt{\frac{0.125(1-0.125)}{40}}$ <-> $0.023 \le p \le 0.227$

Giới hạn khoảng tin cậy 95% trên của σ^2: $p \le 0.125+1.645\sqrt{\frac{0.125(1-0.125)}{40}}$ <-> $p \le 0.211$

Giới hạn khoảng tin cậy 95% dưới của σ^2: $0.125-1.645\sqrt{\frac{0.125(1-0.125)}{40}} \le p$ <-> $0.039 \le p$

4.4 Kiểm định giả thuyết trên giá trị trung bình (μ)

4.4.1 Kiểm định giả thuyết trên giá trị trung bình (μ) của một phân phối chuẩn với phương sai (σ^2) đã biết

a) Lý thuyết:

Giả thuyết không	$H_0 : \mu = \mu_0$
Thống kê kiểm định	$Z_0 = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$
Giả thuyết thay thế	Vùng bác bỏ
$H_1 : \mu \neq \mu_0$	$\lvert z_0 \rvert > z_{\alpha/2}$
$H_1 : \mu > \mu_0$	$z_0 > z_\alpha$
$H_1 : \mu < \mu_0$	$z_0 < -z_\alpha$

b) Bài toán

Bộ Kinh tế, Thương mại và Công nghiệp Nhật Bản (viết tắt METI) đã khảo sát 10 mẫu về tốc độ tăng tưởng GDP của Nhật bản từ năm 2001 tới năm 2020 thì thấy giá trị tăng trưởng trung bình hằng năm là 0.4965.

Hãy xác định giả thuyết $H_0 : \mu = 2.19$, $H_1 : \mu \neq 2.19$ với mức ý nghĩa α = 0.05. Biết rằng độ lệch chuẩn của tổng thể σ = 3.957.

Lời giải

Ta có thông kê kiểm định: $Z_0 = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} = \frac{0.4965 - 2.19}{3.957 / \sqrt{20}} = -1.914$

Ta có $z_{\alpha/2} = z_{0.025} = 1.96$. Nên $(|z_0| = 1.914) < (z_{\alpha/2} = 1.96)$.

Vậy ta fail to reject H_0, với mức ý nghĩa $\alpha = 0.05$ thì giả thuyết chấp nhận được.

4.4.2 Kiểm định giả thuyết trên giá trị trung bình (μ) của một phân phối chuẩn với phương sai (σ^2) chưa biết

a) Lý thuyết:

Giả thuyết không	$H_0 : \mu = \mu_0$
Thống kê kiểm định	$T_0 = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$
Giả thuyết thay thế	Vùng bác bỏ
$H_1 : \mu \neq \mu_0$	$\lvert t_0 \rvert > t_{\alpha/2,n-1}$
$H_1 : \mu > \mu_0$	$t_0 > t_{\alpha,n-1}$
$H_1 : \mu < \mu_0$	$t_0 < -t_{\alpha,n-1}$

b) Bài toán

Cục Dự trữ Liên bang Hòa Kỳ (viết tắt FED) muốn khảo sát sự thay đổi của GDP của Hoa Kỳ trong vài năm qua. Biết rằng tổng thể là một phân phối chuẩn, các giá trị khảo sát thu tập được đưa ra như sau:

2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1.55	2.25	1.84	2.53	3.08	1.71	2.33	3.00	2.16	-3.4

Hãy xác định giả thuyết $H_0 : \mu = 3.39$, $H_1 : \mu < 3.39$ với mức ý nghĩa $\alpha = 0.05$. Và tính giá trị P-value.

Ta có: $\bar{x} = \frac{1.55 + 2.25 + 1.84 + 2.53 + 3.08 + 1.71 + 2.33 + 3.00 + 2.16 - 3.4}{10} = 1.705$

$$s^2 = \frac{(1.55 - \bar{x})^2 + (2.25 - \bar{x})^2 + (1.84 - \bar{x})^2 + (2.53 - \bar{x})^2 + (3.08 - \bar{x})^2 + (1.71 - \bar{x})^2 + (2.33 - \bar{x})^2 + (3.00 - \bar{x})^2 + (2.16 - \bar{x})^2 + (-3.4 - \bar{x})^2}{10 - 1} = 3.472$$

Do đó: $s = 1.86$, $\alpha = 1 - 0.95 = 0.05$; $n = 10 \Rightarrow t_{\alpha,n-1} = t_{0.05,9} = 1.833$

Ta có thông kê kiểm định: $T_0 = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}} = \dfrac{1.705 - 3.39}{3.472/\sqrt{10}} = -1.533$

Do đó: $(t_0 = -1.533) > (-t_{\alpha,n-1} = -1.833)$. Vậy ta fail to reject H_0, với mức ý nghĩa α = 0.05 thì giả thuyết chấp nhận được.

Ta có P-value = $\phi(T_0) = \phi(-1.533) = 0.063$.

4.5 Kiểm định giả thuyết trên tỉ thể tổng thể (p) với mẫu lớn

a) Lý thuyết

Giả thuyết không	$H_0 : \mu = \mu_0$
Thống kê kiểm định	$Z_0 = \dfrac{p - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$
Giả thuyết thay thế	Vùng bác bỏ
$H_1 : \mu \neq \mu_0$	$\lvert z_0 \rvert > z_{\alpha/2}$
$H_1 : \mu > \mu_0$	$z_0 > z_\alpha$
$H_1 : \mu < \mu_0$	$z_0 < -z_\alpha$

b) Bài toán

Tổ chức Thương mại Thế giới (viết tắt WTO) muốn khảo sát về GDP trong những năm kinh tế Nhật Bản có biến động lớn. Trong một mẫu ngẫu nhiên gồm 40 năm, tổ chức nhận thấy rằng có 5 năm mà GDP của Nhật Bản đạt về dưới mức âm cho phép. Hãy xác định giả thuyết rằng khi kinh tế Nhật Bản có biến động lớn thì tỉ lệ đạt dưới mức âm cho phép luôn về lớn hơn 0.1, với mức ý nghĩa α = 0.05.

Lời giải

Ta có $H_0 : p = 0.1$, $H_1 : p > 0.1$. Do đó $p = \dfrac{5}{40} = 0.125$. $z_\alpha = z_{0,05} = 1.645$.

Thống kê kiểm định: $Z_0 = \dfrac{p - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{0.125 - 0.1}{\sqrt{\dfrac{0.1(1-0.1)}{40}}} = 0.527$

Nên $(z_0 = 0.527) < (z_\alpha = 1.645)$. Vậy ta fail to reject H_0, với mức ý nghĩa 0.05, giả thuyết đó chấp nhận được.

4.6 Kiểm định giả thuyết cho phương sai (σ^2) của một phân phối chuẩn

a) Lý thuyết

Giả thuyết không	$H_0 : \sigma^2 = \sigma_0^2$
Thống kê kiểm định	$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$
Giả thuyết thay thế	Vùng bác bỏ
$H_1 : \sigma^2 \neq \sigma_0^2$	$\chi_0^2 > \chi_{\alpha/2,n-1}^2 \text{or} \chi_0^2 < -\chi_{1-\alpha/2,n-1}^2$
$H_1 : \sigma^2 > \sigma_0^2$	$\chi_0^2 > \chi_{\alpha,n-1}^2$
$H_1 : \sigma^2 < \sigma_0^2$	$\chi_0^2 < \chi_{1-\alpha,n-1}^2$

b) Bài toán

Tổ chức Hiệp ước Bắc Đại Tây Dương (viết tắt NATO) sau khi lấy 10 mẫu khảo sát sự thay đổi GDP của Mỹ từ năm 2011 tới năm 2020 thì tính toán được phương sai mẫu khảo sát là 3.472.

Có thể khẳng định rằng đủ bằng chứng trong dữ liệu mẫu để xác định rằng phương sai của tổng thể lớn hơn 2.5 hay không? Với mức ý nghĩa 0.05 và giả sử rằng mẫu khảo sát là một phân phối chuẩn.

Lời giải

Ta có $H_0 : \sigma^2 = 2.5$, $H_1 : \sigma^2 > 2.5$ phương sai mẫu $s^2 = 3.472$, $\alpha = 1 - 0.95 = 0.05$; $n = 10$

$\Rightarrow \chi_{\alpha,n-1}^2 = \chi_{0.05,9}^2 = 16.919$

Ta có : $\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(10-1)\times 3.472}{2.5} = 12.499$.

Vì $(\chi_0^2 = 12.499) < (\chi_{0.05,9}^2 = 16.919)$ .Vậy ta fail to reject H_0, với mức ý nghĩa 0.05, giả thuyết đó chấp nhận được.

4.7 Phân tích hồi quy (Regression analysis)

Phương trình hồi quy tuyến tính đơn giản giữa hai biến số liên tục được biểu diễn bằng công thức: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$. Trong đó $\hat{\beta}_1$ gọi là độ dốc (slope) và $\hat{\beta}_0$ là

chặn (interception). $\hat{\beta}_1$ và $\hat{\beta}_0$ được tính bằng công thức:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Trong đó:

$$S_{xy} = \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$$

Hệ số tương quan mẫu là: $R = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}$

Bài toán: Khảo sát phần trăm GDP của Mỹ và Nhật Bản trong khoảng 10 từ 2011 đến 2020 thu được kết quả trung bình GDP như sau:

Ngày	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Mỹ	1.55	2.25	1.84	2.53	3.08	1.71	2.33	3	2.16	-3.4
Nhật Bản	0.02	1.37	2.01	0.3	1.56	0.75	1.68	0.56	0.27	-4.59

Hãy tính phương trình hồi quy truyến tính đơn giản và hệ số tương quan mẫu.

Lời giải

Theo số liệu đã cho ta dễ dàng tính được:

$$\overline{x}=\frac{1.55+2.25+1.84+2.53+3.08+1.71+2.33+3+2.16-3.4}{10}=1.705$$

$$\overline{y}=\frac{0.02+1.37+2.01+0.3+11502+1.56+0.75+1.68+0.56+0.27-4.59}{10}=0.393$$

Ta tính được: $\sum_{i=1}^{10}x_i=17.05;\ \sum_{i=1}^{10}y_i=3.93;$

$$\sum_{i=1}^{10}x_i^2=1.55^2+2.25^2+1.84^2+2.53^2+3.08^2+1.71^2+2.33^2+3^2+2.16^2+(-3.4)^2=60.3165$$

$$\sum_{i=1}^{10}y_i^2=0.02^2+1.37^2+2.01^2+0.3^2+1.56^2+0.75^2+1.68^2+0.56^2+0.27^2+(-4.59)^2=33.2805$$

$$\sum_{i=1}^{10}x_iy_i=1.55\times0.02+2.25\times1.37+1.84\times2.01+2.53\times0.3+3.08\times1.56+1.71\times0.75+2.33\times1.68+3\times0.56$$
$$+2.16\times0.27+(-3.4)\times(-4.59)=35.4418$$

Do đó: $S_{XX}=\sum_{i=1}^{n}x_i^2-\frac{(\sum_{i=1}^{n}x_i)^2}{n}=60.3165-\frac{17.05^2}{10}=31.2463$

$$S_{YY}=\sum_{i=1}^{n}y_i^2-\frac{(\sum_{i=1}^{n}y_i)^2}{n}=33.2805-\frac{3.93^2}{10}=31.736$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n} = 35.4418 - \frac{17.05 \times 3.93}{10} = 28.7412$$

Do đó: $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{28.7412}{31.2463} \approx 0.9198$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0,393 - 0.9198 \times 1.705 \approx -1.1753$$

$$R = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} = \frac{28.7412}{\sqrt{31.2463 \times 31.736}} \approx 0.9127 .$$

Vậy ta được phương trình hồi quy tuyến tính đơn giản: $\hat{y} = -1.1753 + 0.9198x$.

Và hệ số tương quan mẫu là: 0.9127.