

**MAS291 Computer Project:**

**Quan hệ giữa dân số của từng vùng với GRDP bình quân đầu người ( triệu đồng) của Đồng bằng sông Hồng và Trung du miền núi phía Bắc.**

**Giảng viên**: Nguyễn Việt Anh

**Môn học**: MAS291 - SE1609

**Thành viên**: Lê Minh Dũng (HE160721)

Lý Minh Nghĩa (HE163464)

Nguyễn Anh Minh (HE151192)

Tống Sỹ Lương (HE153353)

**Mục lục:**

1. **Giới thiệu dự án.**
2. **Thu thập và xử lý dữ liệu.**
3. **Thu thập dữ liệu……………………………………………………….**
4. **So sánh tổng dân số với GRDP………………………………………..**
5. **Tìm khoảng tin cậy…………………………………………………….**
6. **Kiểm định 1 tham số…………………………………………………..**
7. **Hypotheses Tests for a Difference in Means, Variances Unknown: với α=0.05, chứng minh GRDP bình quân đầu người (triệu đồng) của Đồng bằng sông Hồng và Trung du miền núi phía Bắc là như nhau?........................................................................................................**
8. **Confidence Interval on the Difference in Means, Variance Unknown: Xây dựng khoảng tin cậy 90% cho chênh lệch về trung bình dân số giữa giữa Đồng bằng sông Hồng và Trung du miền núi phía Bắc………………………………………………………………….**
9. **Large-Sample Test on the Difference in Population Proportions: 1 tỉnh được coi là đạt chỉ tiêu GRDP nếu doanh thu > 93(triệu), với α=0.05 liệu giả thuyết tỉ lệ số tỉnh ở vùng Đồng bằng sông Hồng sẽ nhiều hơn so với ở Trung du miền núi phía Bắc có đúng hay không?**
10. **Sử dụng hồi quy tuyến tính (Simple Linear Regression ) để kết luận sự tương quan giữa tổng dân số của các tỉnh với GRDP……………..**
11. **Kết luận và nêu ý kiến.**
12. **Nguồn.**

1. **Giới thiệu dự án**
2. **Đặt vấn đề**

* Việt Nam hiện đang là đất nước đang phát triển, cho nên kinh tế luôn là yếu tố được đặt lên hàng đầu và trong đó có GRDP.
* GRDP luôn thay đổi theo từng năm, từng vùng và phụ thuộc rất nhiều vào các yếu tố khác nhau, không vùng nào giống vùng nào.
* Dự án của chúng mình lần này sẽ tổng hợp và phân tích dữ liệu GRDP cùng với tổng số dân của Việt Nam theo từng vùng vào năm 2018.

1. **Mục tiêu dự án**

* Tổng hợp và phân tích dữ liệu dân số và GRDP được của 2 vùng Đồng bằng sông Hồng và Trung du miền núi phía Bắc.
* Xác định xem vùng nào có tỉ lệ GRDP cao hơn, cũng như vùng nào có dân số đông hơn.
* Tìm được xu hướng chung của 2 vùng.

1. **NOTE:**

* Do đối tượng được phân tích ở bài này là các tỉnh/thành phố và số lượng các tỉnh quá ít và dữ liệu đã được xử lý. Vậy nên mọi suy diễn thống kê từ 2 bảng ở trong bài này là hiển nhiên và có thể dễ dàng suy ra được. Thầy và các bạn đọc có thể coi population của bài này là một số lớn nào đó và team bọn mình chọn ra ngẫu nhiên 25 tỉnh từ 2 vùng trong cái population đó, và chỉ nên tham khảo cách sử dụng công thức và sử dụng dữ liệu. Mong thầy và các bạn đọc thông cảm!

1. **Thu thập dữ liệu và xử lý**
2. **Thu thập dữ liệu.**

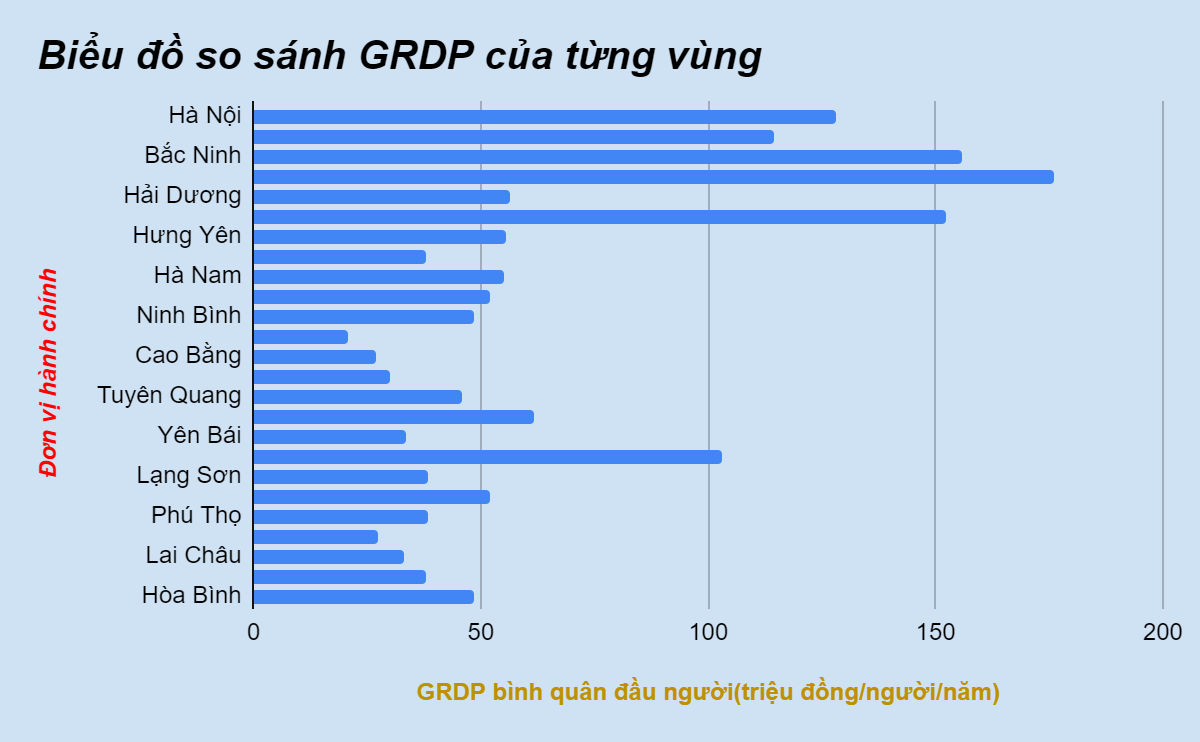
* Bảng GRDP bình quân đầu người theo tổng dân số của từng vùng thuộc khu vực Đồng bằng sông Hồng.

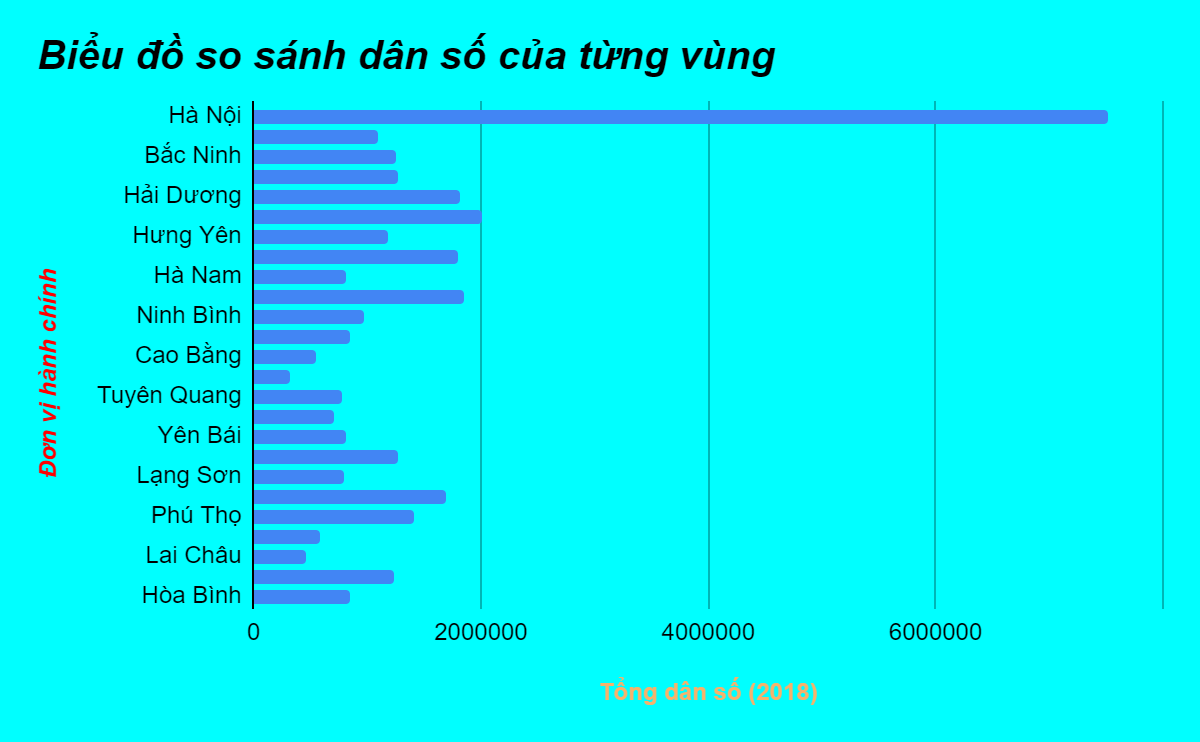
| **STT** | **Đơn vị hành chính** | **Tổng dân số (2018)** | **GRDP bình quân đầu người(triệu đồng/người/năm)** |
| --- | --- | --- | --- |
| 1 | Hà Nội | 7520700 | 128.2 |
| 2 | Vĩnh Phúc | 1092400 | 114.3 |
| 3 | Bắc Ninh | 1247500 | 155.6 |
| 4 | Quảng Ninh | 1266500 | 176 |
| 5 | Hải Dương | 1807500 | 56.3 |
| 6 | Hải Phòng | 2013800 | 152.3 |
| 7 | Hưng Yên | 1188900 | 55.3 |
| 8 | Thái Bình | 1793200 | 38 |
| 9 | Hà Nam | 808200 | 55.2 |
| 10 | Nam Định | 1854400 | 52 |
| 11 | Ninh Bình | 973300 | 48.5 |
| **AVG** | 1960581.818 | | 93.79 |
| **Median** | 1266500 | | 56.3 |
| **Standard Deviation** | 1887396.319 | | 51.85 |

* Bảng GRDP bình quân đầu người theo tổng dân số của từng vùng thuộc khu vực Trung du miền núi phía Bắc.

| **STT** | **Đơn vị hành chính** | **Tổng dân số (2018)** | **GRDP bình quân đầu người(triệu đồng/người/năm)** |
| --- | --- | --- | --- |
| 1 | Hà Giang | 846500 | 20.7 |
| 2 | Cao Bằng | 540400 | 26.7 |
| 3 | Bắc Kạn | 327900 | 30 |
| 4 | Tuyên Quang | 780100 | 45.7 |
| 5 | Lào Cai | 705600 | 61.84 |
| 6 | Yên Bái | 815600 | 33.6 |
| 7 | Thái Nguyên | 1268300 | 102.9 |
| 8 | Lạng Sơn | 790500 | 38.4 |
| 9 | Bắc Giang | 1691800 | 52.1 |
| 10 | Phú Thọ | 1404200 | 38.5 |
| 11 | Điện Biên | 576700 | 27.31 |
| 12 | Lai Châu | 456300 | 33 |
| 13 | Sơn La | 1242700 | 38 |
| 14 | Hòa Bình | 846100 | 48.3 |
| **AVG** | 878050 | | 42.64 |
| **median** | 803050 | | 38.2 |
| **Standard Deviation** | 388543.5739 | | 20.55 |

1. **So sánh dân số với GRDP của từng vùng.**

****

****

1. **Confidence Interval on the Mean of a Normal Distribution, Variance Unknown: xây dựng khoảng tin cậy 95% cho cả 2 vùng Đồng bằng sông Hồng và Trung du miền núi phía Bắc.**

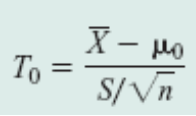
* Ta có: = 93.79; = 42.64; s1 = 51.85; s2 = 20.55
* t∝∕2 ,n-1 = 2.228 (Đồng bằng sông Hồng).
* t∝∕2 ,n-1 = 2.16 (Trung du miền núi phía Bắc).



* Đồng bằng sông Hồng : ( 59.06,128.52).
* Trung du miền núi phía Bắc : (30.77 , 54.50).

1. **Kiểm định 1 tham số:**

* **Tin rằng GRDP bình quân đầu người của đồng bằng Sông Hồng lớn hơn 100 triệu đồng/người/năm.**
* Ta có: n=11; =93.79 ; =100;
* H0: μ=
* H1: μ>



=> = -0.397 < t

=> Không thể bác bỏ H

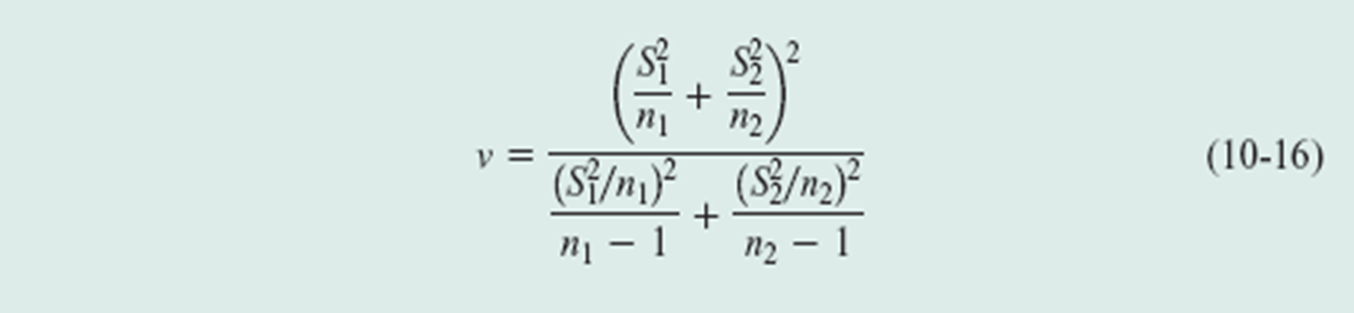
=>.GRDP bình quân đầu người của đồng bằng Sông Hồng nhỏ hơn 100 triệu đồng/người/năm

1. **Hypotheses Tests for a Difference in Means, Variances Unknown: Xây dựng khoảng tin cậy 95% cho trung bình GRDP của vùng Đồng bằng sông Hồng là ít hơn so với của vùng Trung du miền núi phía Bắc.**

* Ta có: = 93.79; = 42.64; s1 = 51.85; s2 = 20.55
* : µ1 - µ2 = 0 (GRDP bình quân đầu người của đồng bằng sông Hồng ít hơn).
* : µ1 > µ2 (GRDP bình quân đầu người của đồng bằng sông Hồng nhiều hơn).
* Do nên ta áp dụng công thức:

= = 3.08

* Ta có công thức tính bậc tự do:

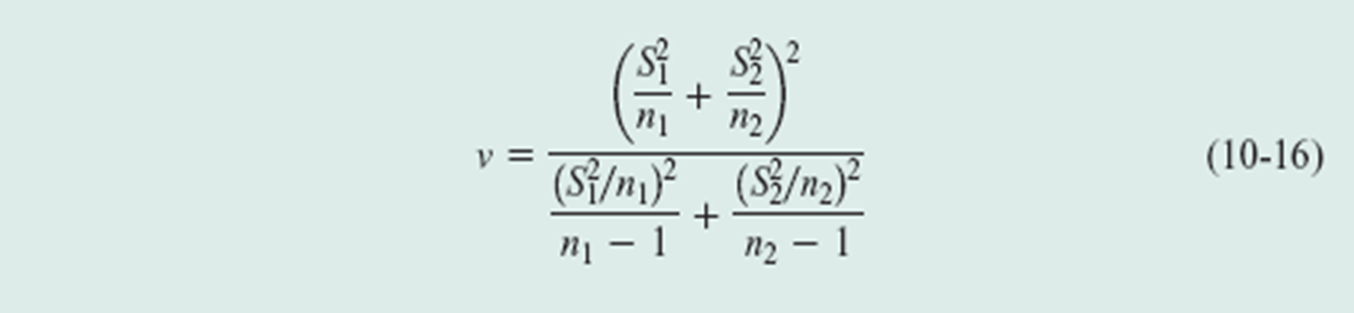


=> v = 12

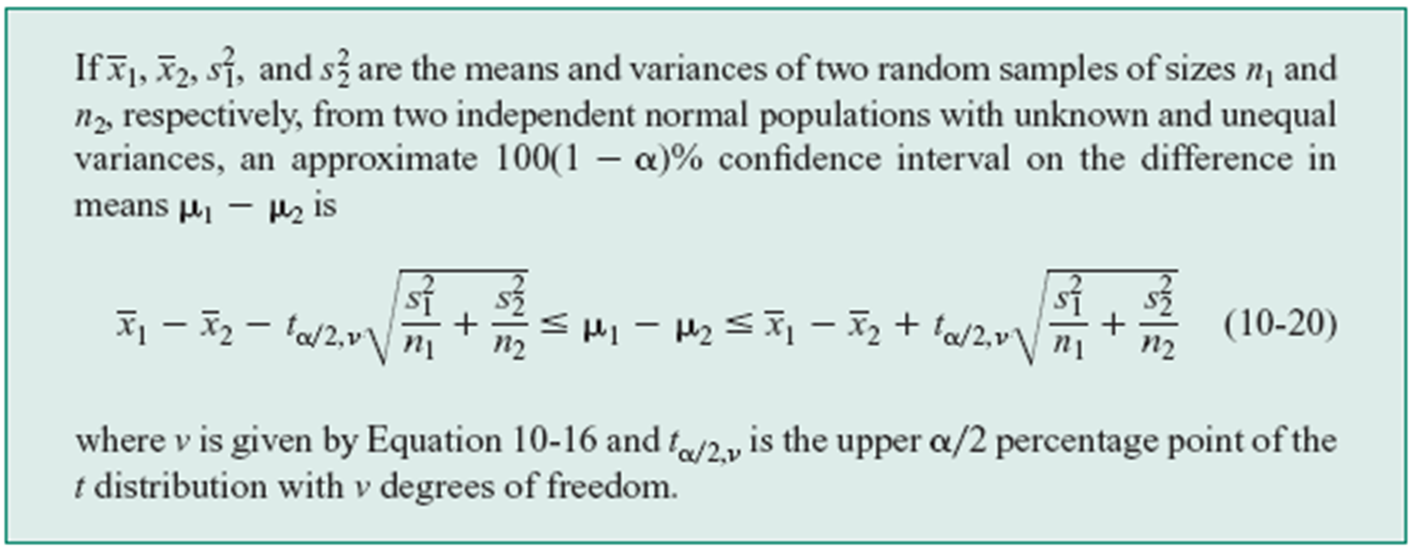
* Do => Không đủ để bác bỏ H0
* Kết luận: GRDP bình quân đầu người (triệu đồng) của Đồng bằng sông Hồng nhiều hơn vùng núi phía Bắc.

1. **Confidence Interval on the Difference in Means, Variance Unknown: Xây dựng khoảng tin cậy 90% cho chênh lệch về trung bình dân số giữa giữa Đồng bằng sông Hồng và Trung du miền núi phía Bắc.**

* Do nên ta áp dụng công thức tính độ tự do:



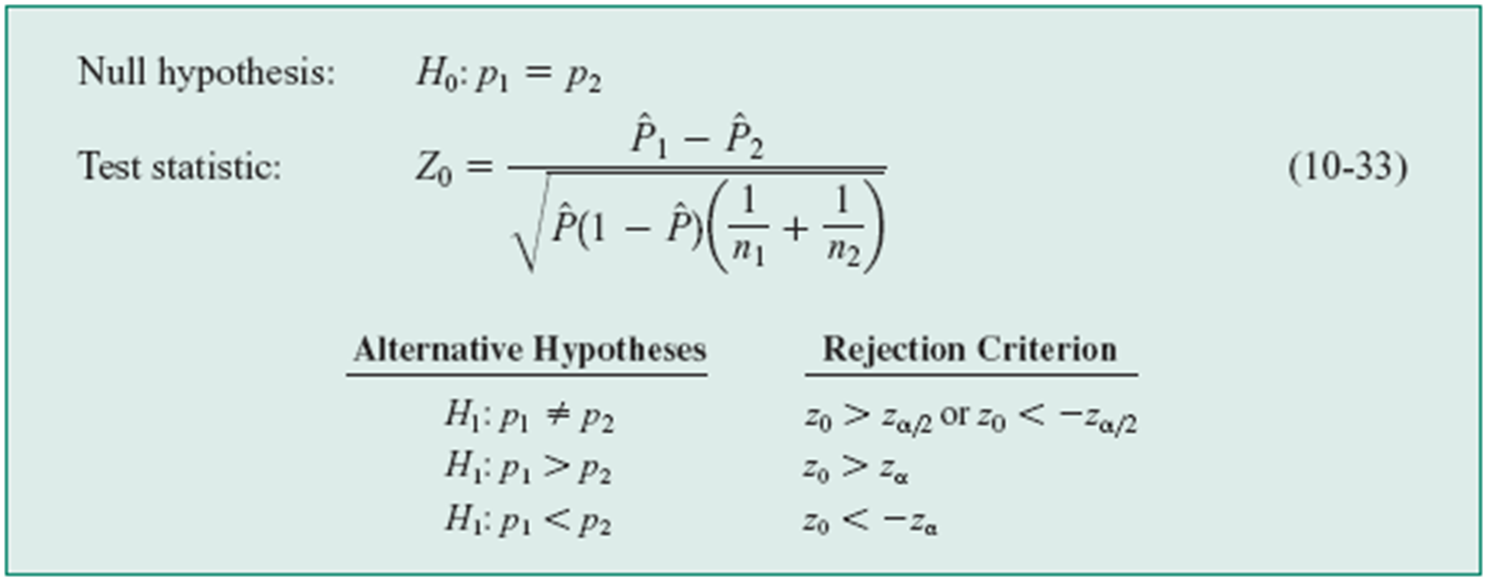
* Từ đó tính ra v = 13. Sau đó ta áp dụng công thức để tìm ra khoảng tin cậy với .



* Khoảng tin cậy: (-43848 ; 2208911.692).

1. **Large-Sample Test on the Difference in Population Proportions: 1 tỉnh được coi là đạt chỉ tiêu GRDP nếu doanh thu > 93(triệu), xây dựng khoảng tin cậy 95% với giả thuyết tỉ lệ số tỉnh ở vùng Đồng bằng sông Hồng sẽ nhiều hơn so với ở Trung du miền núi phía Bắc có đúng hay không?**

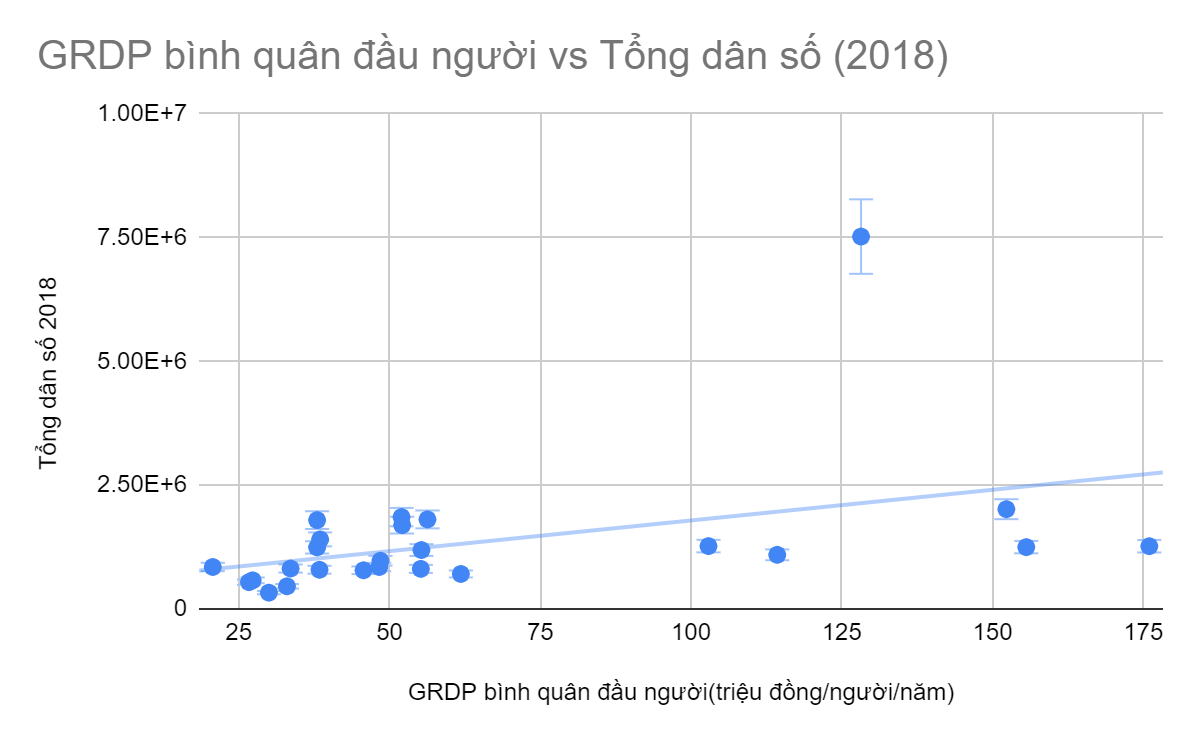
* : = (tỉ lệ các tỉnh đạt chỉ tiêu của Đồng bằng sông Hồng nhiều hơn của Trung du miền núi phía Bắc).
* : < (tỉ lệ các tỉnh đạt chỉ tiêu của Đồng bằng sông Hồng ít hơn của Trung du miền núi phía Bắc).
* Từ cột GRDP ta tìm được các tỉnh đạt chỉ tiêu của 2 vùng lần lượt là: và
* Áp dụng CT: = 0.24
* Sau đó ta kiểm định thống kê bằng CT:



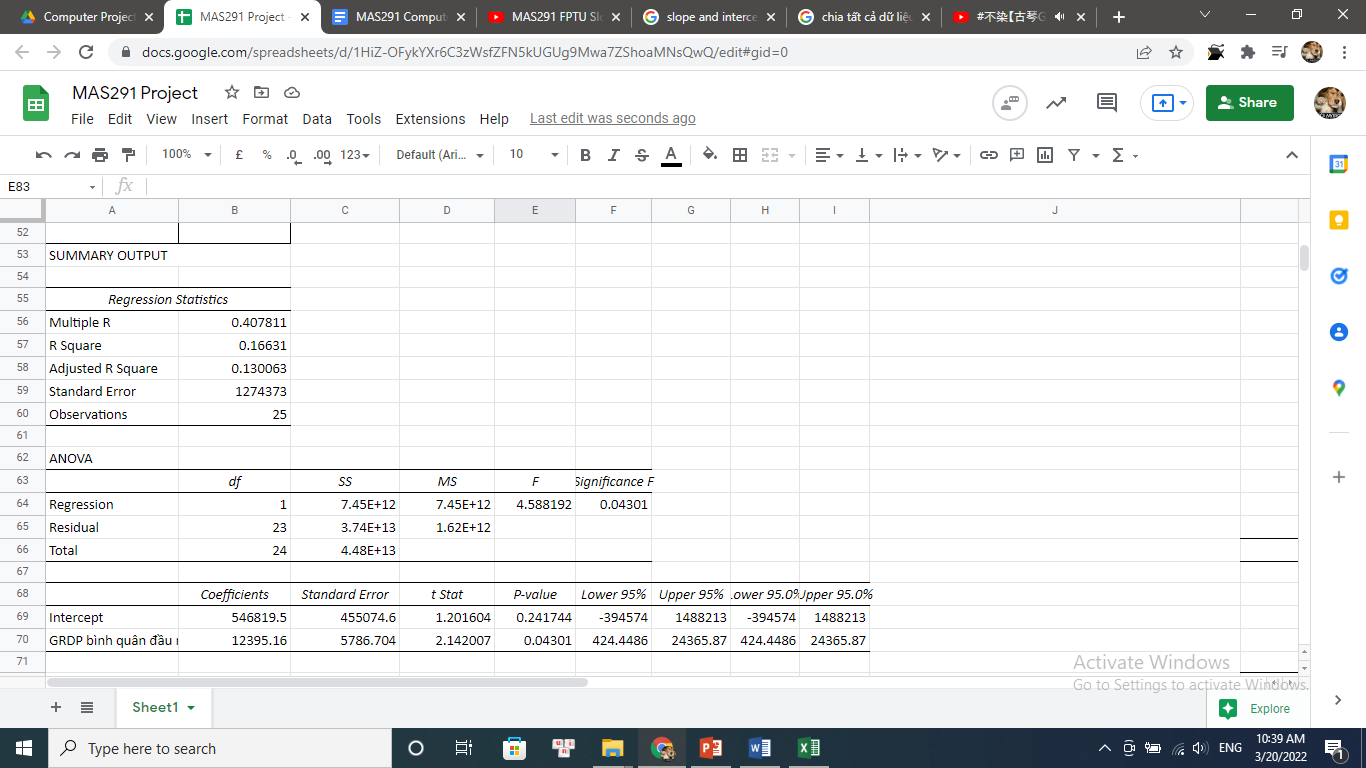
* Với **α=0.05,** ta có miền bác bỏ {Z: Z<}
* Kết luận: do > nên ta thừa nhận và bác bỏ

1. **Sử dụng hồi quy tuyến tính (Simple Linear Regression ) để kết luận sự tương quan giữa tổng dân số của các tỉnh với GRDP.**

* Sử dụng chức năng tạo chart của google sheets ta có thể vẽ được một biểu đồ phân tán như sau:

****

* Tiến hành đặt giả thuyết:
* (dân số và GRDP không có sự tương quan).
* (dân số và GRDP có sự tương quan).

****

* Sử dụng Regression trong Excel ta có thể tính được regression line là :

y = 546819.5 + 12395.16.x

* =
* . Do > ta có thể bác bỏ và kết luận được rằng giữa tổng số người dân và GRDP của các vùng có sự tương quan. Có thể lấy ví dụ cụ thể ở từ trong biểu đồ phân tán ở trên, ngoại trừ Hà Nội với mức dân số ở 7 triệu người và GRDP ở mức 128 triệu/năm là không nằm trên trend line.

**C. Kết luận.**

* Từ những dữ liệu thu được cũng như quá trình phân tích dữ liệu, ta có thể đi tới những kết luận như sau:
* GRDP bình quân đầu người ở khu vực đồng bằng Sông Hồng lớn hơn GRDP bình quân đầu người ở khu vực Trung du miền núi phía Bắc
* Những khu vực có đông dân cư sinh sống thì GDP bình quân đầu người ở khu vực đó cao hơn
* Tổng dân số có liên quan tới GRDP bình quân đầu người

**D. Nguồn.**

* <https://vi.wikipedia.org/wiki/Danh_s%C3%A1ch_%C4%91%C6%A1n_v%E1%BB%8B_h%C3%A0nh_ch%C3%ADnh_Vi%E1%BB%87t_Nam_theo_GRDP_b%C3%ACnh_qu%C3%A2n_%C4%91%E1%BA%A7u_ng%C6%B0%E1%BB%9Di?fbclid=IwAR1SaCGHPNUTDRqK0m2JZMEyn-jN96V1HAVe9CooHebNV4ibDZufSiP9ngQ>