**Laporan Projek 4**

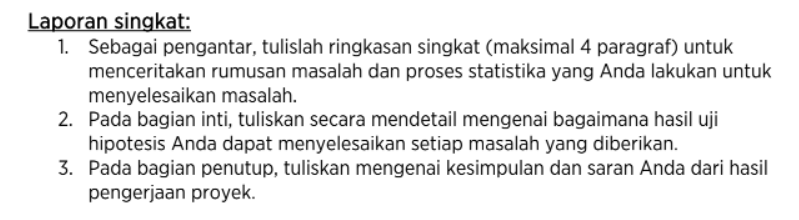
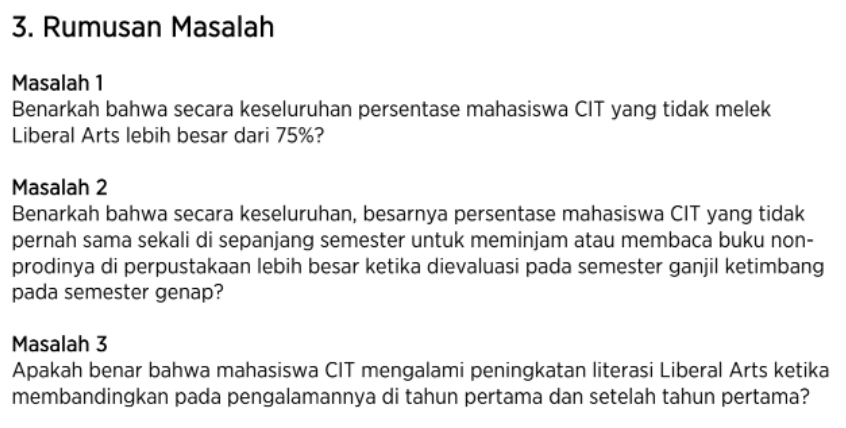
**Literasi Liberal Arts**

Kelompok 8:

* Bryan Christopher Wijaya
* James Patrick Oentoro
* Noel Christevent Mandak
* Stefannus Christian
* Tiffany Sondakh

Link tugas pemrograman :

<https://datalore.jetbrains.com/notebook/CPsQxEPW7yVpS665a3lC8b/HQnffFKXGwSjP7woAxXKRS/>



* Ringkasan Singkat
  1. Rumusan Masalah:
     1. Dr. Calvin memiliki kekhawatiran mengenai literasi Liberal Arts pada mahasiswa CIT. Literasi Liberal Arts yang dimaksudkan adalah saat mahasiswa memiliki rasa ingin tahu terhadap hal-hal di luar bahasa program studinya sendiri. Dr. Calvin merasa sangat prihatin jika ternyata sebagian besar mahasiswa CIT tidak pernah datang ke perpustakaan untuk membaca ataupun meminjam buku yang di luar cakupan Prodinya. Dari kekhawatiran Dr. Calvin, terdapat 3 rumusan masalah. Dr. Calvin meminta kami untuk menguji dugaan-dugaan tersebut secara statistika. 3 Rumusan masalah tersebut adalah:

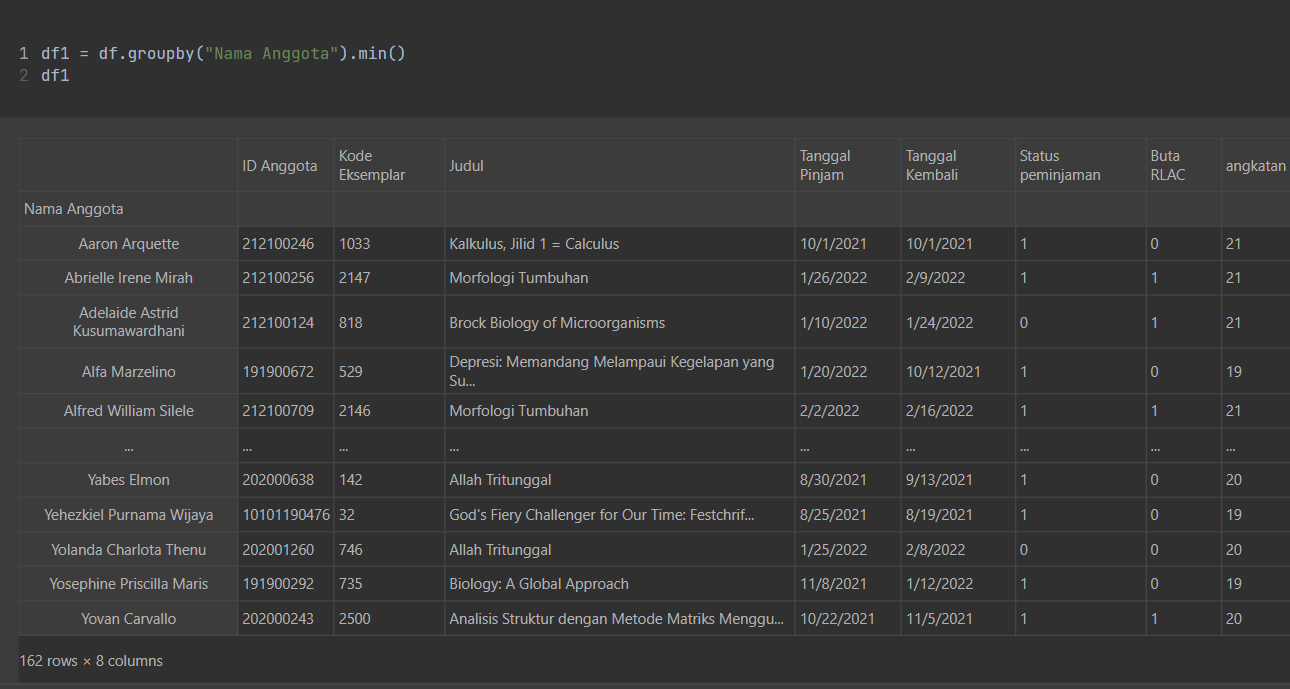
1) Apakah keseluruhan persentase mahasiswa CIT yang **tidak** melek RLAC >= 75%.

2) Apakah besarnya persentase mahasiswa CIT yang tidak pernah sama sekali di sepanjang semester untuk meminjam atau membaca buku non-prodinya di perpustakaan lebih besar di semester ganjil ketimbang pada semester genap.

3) Apakah Mahasiswa CIT mengalami peningkatan literasi RLAC ketika membandingkan pada pengalamannya di tahun pertama dan setelah tahun pertama.

* 1. Proses Statistika
     1. Hal pertama yang dilakukan adalah mengumpulkan data sampel. Kami meminta data sampel dari pihak perpustakaan CIT secara langsung. Hal ini menyebabkan data sampel yang kami miliki untuk membuat hipotesis adalah data sampel yang saling bebas dan otentik.
     2. Kemudian kami melakukan data preprocessing dengan cara menambahkan kolom Buta RLAC. Kolom ini akan memberikan nilai 1 jika orang tersebut membaca buku yang non - RLAC atau membaca buku sesuai dengan prodinya, dan akan memberikan nilai 0 jika orang tersebut membaca buku RLAC atau membaca buku di luar prodinya. Langkah ini membuat suatu masalah yaitu jika misalnya orang tersebut meminjam atau membaca suatu buku lebih dari sekali. Katakanlah misal orang A pernah membaca buku non prodi dan prodinya. Maka data orang ini akan menjadi bias. Untuk mengatasi hal ini, kami menggunakan fungsi group by berdasarkan nama dan mengambil data yang minimal. Setelah dilakukan proses ini, maka data yang tersisa adalah 0 saja (Melek RLAC) karena kami mengambil yang minimal setelah di group by berdasarkan nama.

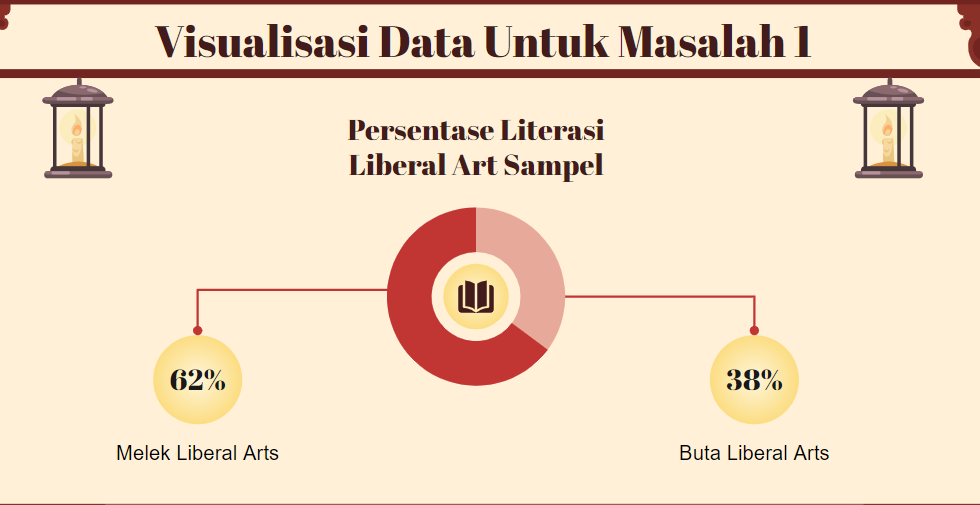
Kode Pra-Pemrosesan Data



* + 1. Setelah itu, kami menentukan hipotesis null dan hipotesis alternatif untuk ketiga masalah dari Dr. Calvin. Kami memiliki tingkat signifikansi 0.01 untuk ketiga masalah. Setelah itu kami memilih statistik yang cocok untuk membangun critical region atau p-value berdasarkan nilai alpha. Kemudian, kami membuat kesimpulan hasil hipotesis, kesimpulan detail akan dijelaskan di nomor 2. Langkah terakhir yang dilakukan adalah menerjemahkan kesimpulan ke dalam bahasa sehari-hari sehingga orang awam yang tidak mengerti statistika dapat mengerti kesimpulan yang dibuat.
* Hasil Uji Hipotesis
* **Masalah 1 :**

*Benarkah bahwa secara keseluruhan persentase mahasiswa CIT yang tidak melek Liberal Arts lebih besar dari 75% ?*

* 1. Visualisasi Data Masalah 1

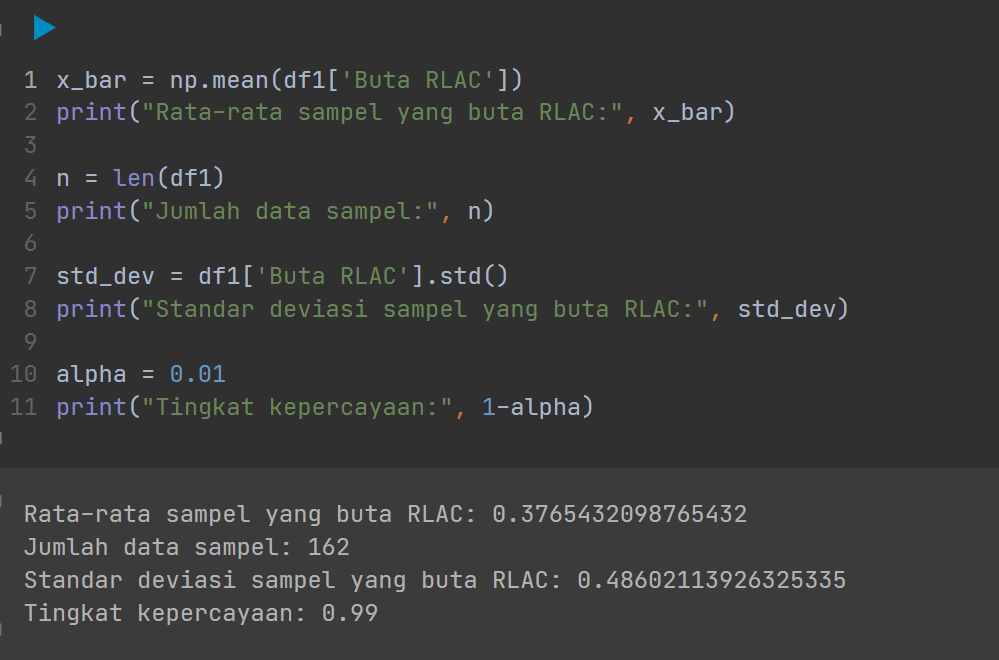


Dapat dilihat dari visualisasi data diatas bahwa secara kasar, 62% dari data sampel Melek Liberal Arts dan 38% Buta Liberal Arts. Dapat ditebak bahwa dari data diatas bahwa seharusnya hasil statistika hipotesis yang didapatkan adalah bahwa mahasiswa Mahasiswa CIT yang tidak melek Liberal Arts tidak lebih dari 75%. Hal ini akan dibuktikan dengan langkah-langkah statistika dibawah ini.

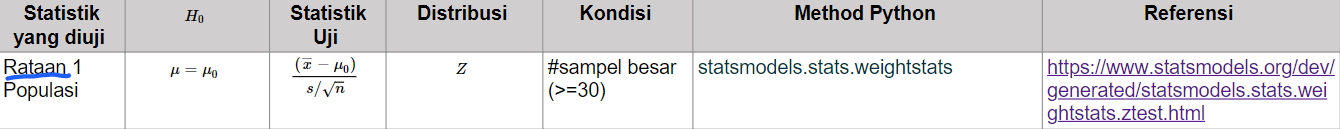
Langkah pertama adalah menentukan hipotesis null dan hipotesis alternatif. Untuk masalah satu, kami memilih tingkat signifikansi 0.01 (tingkat kepercayaan 99%) dengan hipotesis null yaitu dan .

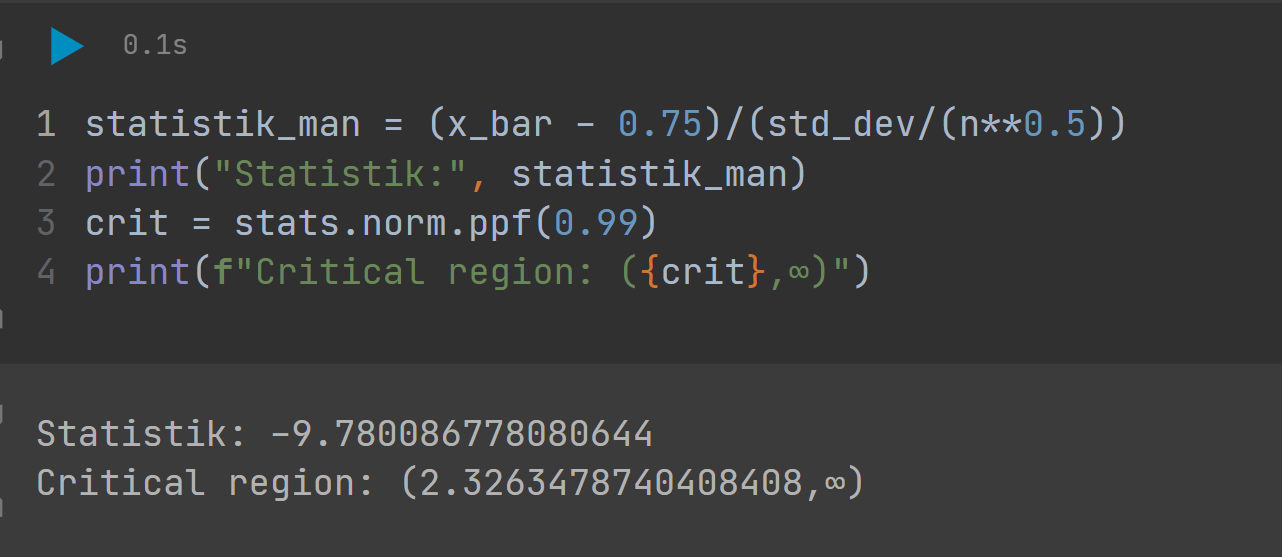
Statistik yang dipakai dalam masalah satu adalah rataan satu populasi.

Perhitungan Statistika. Kami melakukan perhitungan statistika dengan membagi ke dalam dua cara yaitu cara manual dan cara modern.

Perhitungan Cara Manual Masalah 1

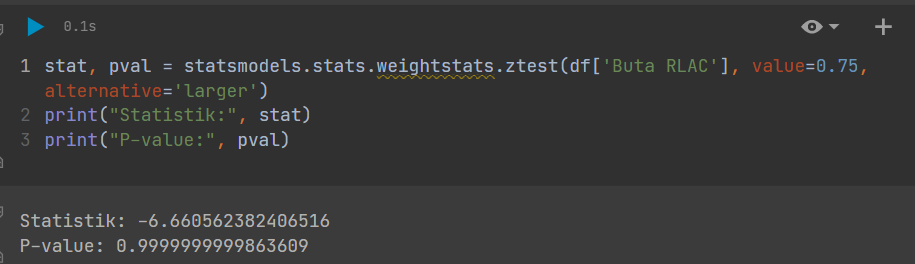
Karena sampel yang kami kumpulkan lebih besar dari 30, maka data sampel kami termasuk data sampel yang besar. Karena itu kami menggunakan distribusi normal (Z). Referensi rumus yang dipakai adalah sebagai berikut.



Berikut merupakan hasil yang statistika yang didapatkan.

Dapat dilihat bahwa hasil statistik tidak jatuh pada critical region, maka dari itu Hipotesis null tidak ditolak sehingga data tidak cukup signifikan untuk menyimpulkan apapun.

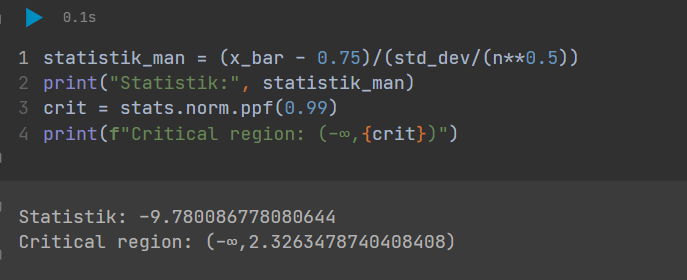
Perhitungan Cara Modern Masalah 1



Library python yang dipakai adalah statsmodel dan method yang digunakan adalah stats.weightstats.ztest. Dapat dilihat dari hasil diatas bahwa p value > alpha sehingga H0 tidak ditolak. Karena H0 tidak ditolak, maka data tidak cukup signifikan untuk menyimpulkan apapun.

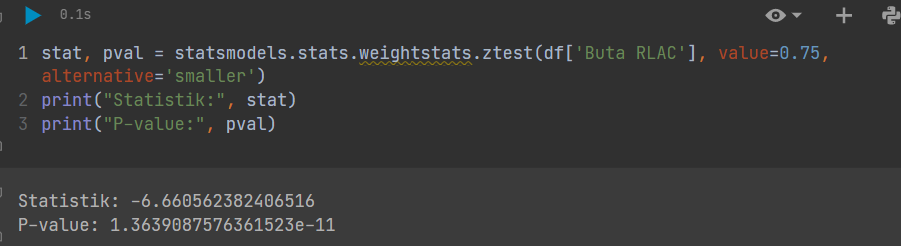
Hal yang dapat dilakukan agar Hipotesis null ditolak adalah dengan menukar hipotesis. Hipotesis null sekarang adalah dan Hipotesis Alternatif . Tingkat signifikansi yang digunakan masih sama yaitu 0.01.

Cara Manual (Setelah Hipotesis Masalah 1 Dibalik)



Dapat dilihat, setelah hipotesis dibalik, maka hasil statistik sekarang berada di dalam critical region. Artinya, H0 ditolak sehingga dapat disimpulkan bahwa dengan tingkat kepercayaan 99%, Mahasiswa CIT yang tidak melek Liberal Arts tidak lebih dari 75%.

Cara Modern (Setelah Hipotesis Masalah 1 Dibalik)



Dapat dilihat sekarang bahwa setelah hipotesis dibalik, p value lebih kecil dari alpha. Hal ini menyebabkan H0 ditolak dan H1 dapat dipertimbangkan. Maka dari itu, Dengan tingkat kepercayaan 99%, Mahasiswa CIT yang tidak melek Liberal Arts tidak lebih dari 75%.

**Masalah 2:**

*Benarkah bahwa secara keseluruhan, besarnya persentase mahasiswa CIT yang tidak pernah sama sekali di sepanjang semester untuk meminjam atau membaca buku non-prodinya di perpustakaan lebih besar ketika dievaluasi pada semester ganjil ketimbang pada semester genap?*

Misalkan:

adalah mahasiswa CIT yang membaca buku non-prodinya pada semester ganjil

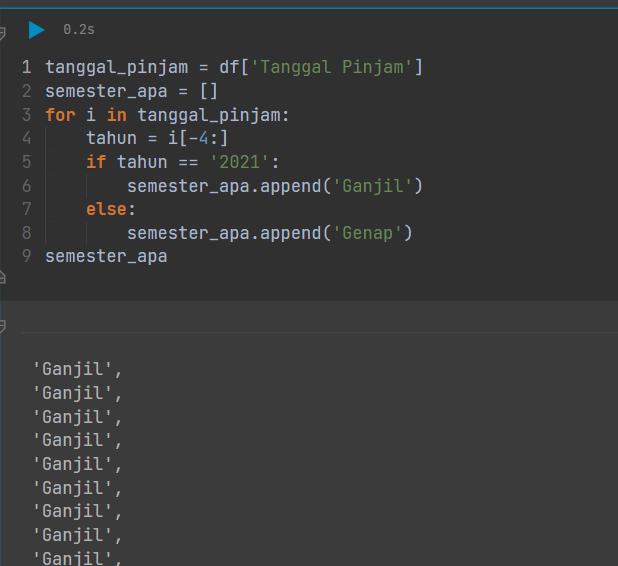
adalah mahasiswa CIT yang membaca buku non-prodinya pada semester genap

Hipotesis null dan alternatif yang dipilih adalah sebagai berikut.

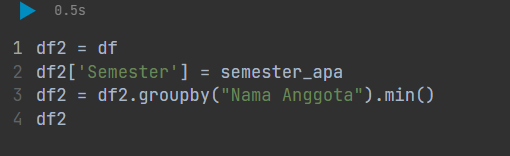
Kami memilih tingkat signifikansi = 0.01 untuk masalah ini dan statistik yang dipakai adalah selisih rataan dua populasi.

Pemrosesan Data

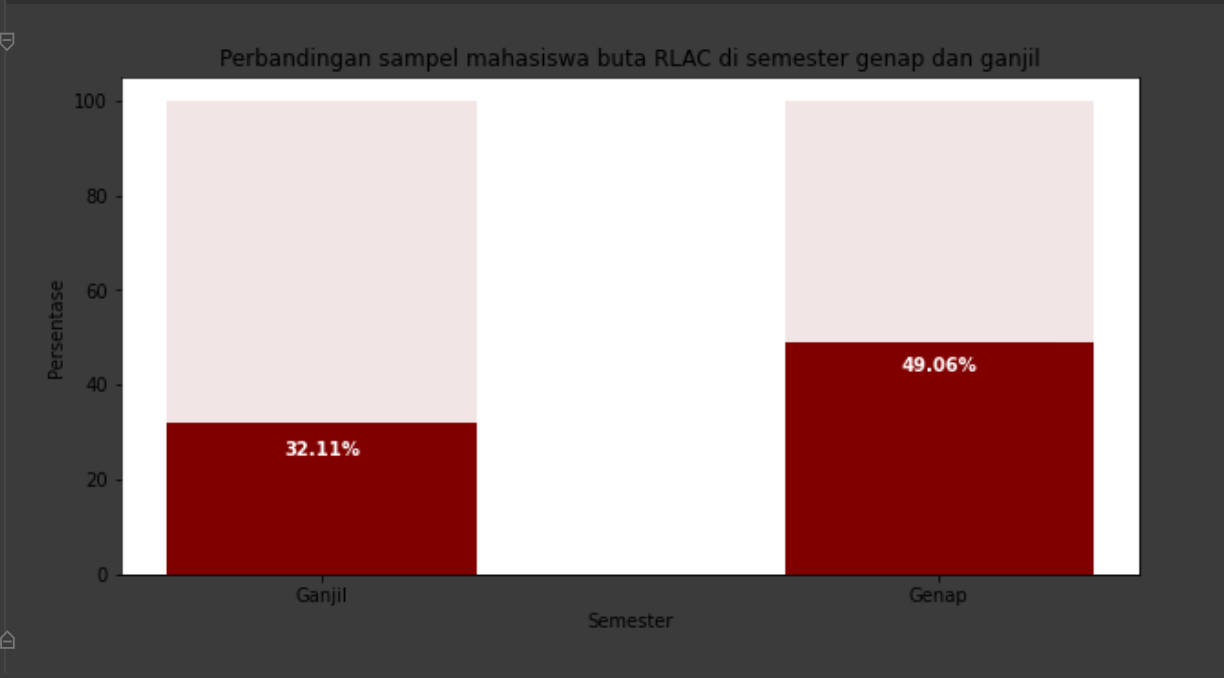
Sebelum masuk ke dalam perhitungan statistika terdapat pemrosesan data yang harus dilakukan.



Data sampel yang digunakan dalam masalah 2 adalah data sampel yang didapat dari akhir 2021 sampai awal 2022. Artinya, untuk memisahkan yang mana yang data semester ganjil dan yang mana data semester genap. Hal yang perlu dilakukan adalah melihat tahun nya, jika tahun nya adalah 2021, maka mahasiswa tersebut tergolong meminjam di semester ganjil, sebaliknya jika tahun nya adalah 2022, maka mahasiswa tersebut tergolong meminjam di semester genap.

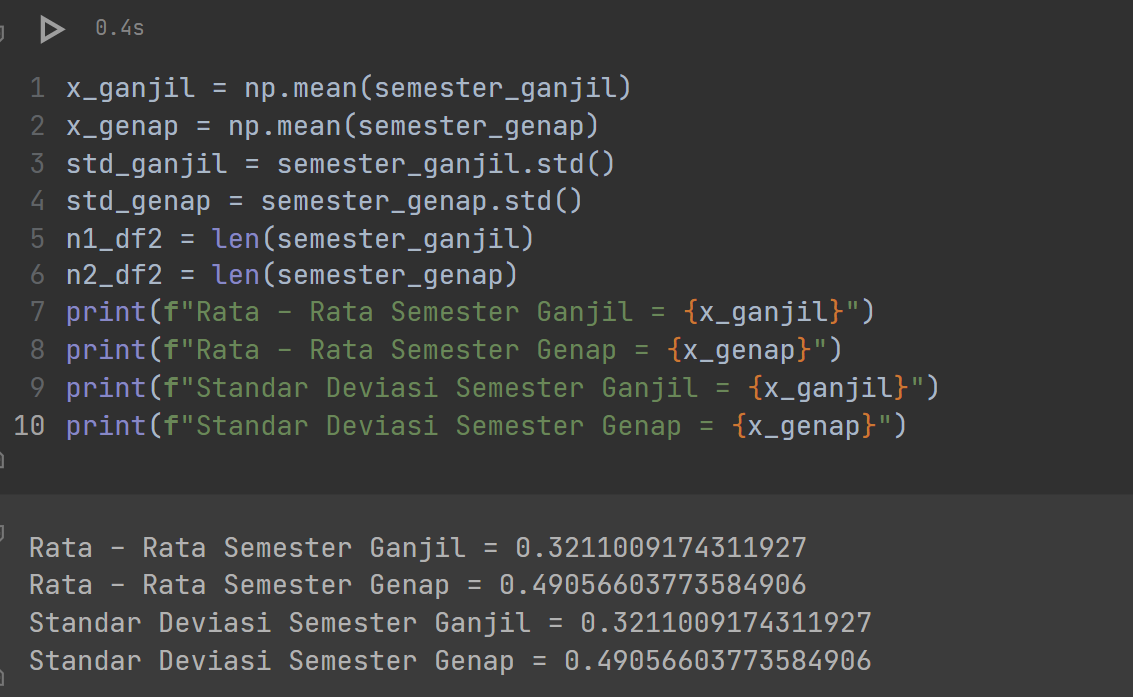


Setelah, dipisah berdasarkan semester, perlu dilakukan group by berdasarkan nama lagi untuk mencegah masalah yang sudah dijelaskan pada masalah 1.



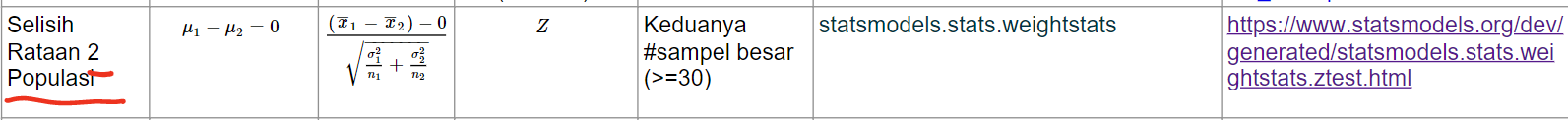
Berikut merupakan visualisasi data dari masalah 2. Dapat dilihat bahwa terdapat 32.11 % mahasiswa di semester ganjil yang buta RLAC, sedangkan terdapat 49.06% mahasiswa di semester genap yang buta RLAC. Dari hal ini dapat diambil inferensi bahwa lebih banyak mahasiswa yang buta RLAC pada semester genap dibandingkan dengan semester ganjil. Hal ini juga dapat dibuktikan dengan hasil perhitungan statistika hipotesis dibawah ini.

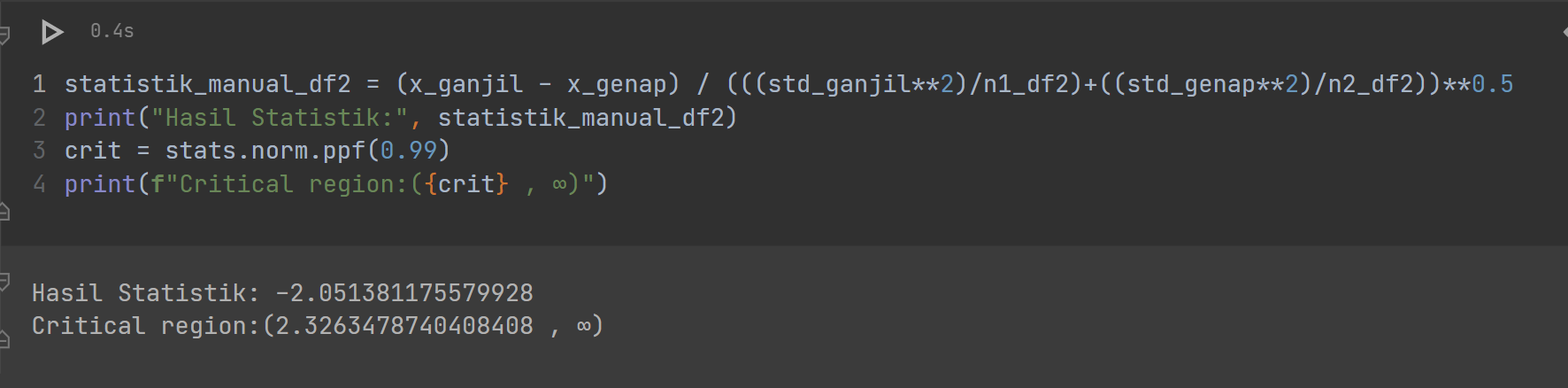
Berikut merupakan statistika deskriptif dari masalah 2



Cara Manual

Karena sampel yang dipilih >= 30, maka data sampel termasuk data sampel yang besar. Maka dari itu, distribusi yang digunakan adalah distribusi normal (Z). Berikut merupakan perhitungan dengan cara manual. Rumus dan referensi library python yang digunakan adalah sebagai berikut.

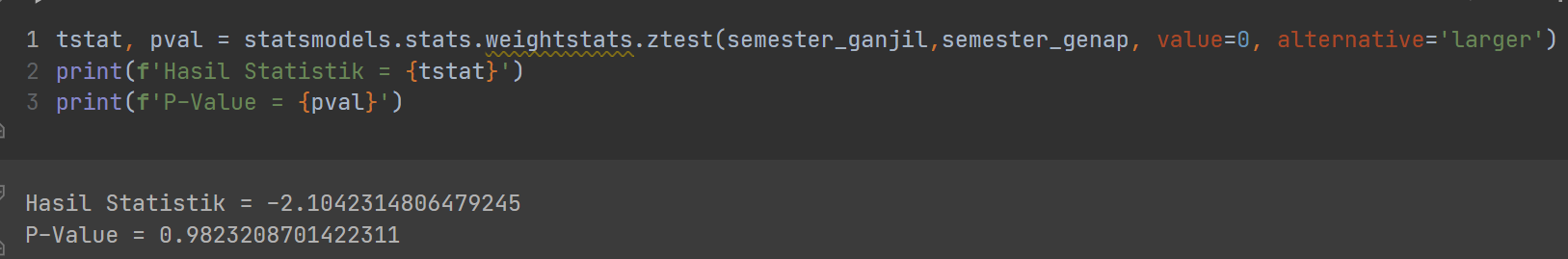




Dapat dilihat dari hasil statistik diatas bahwa hasil statistik berada diluar critical region. Maka dari itu, H0 tidak ditolak sehingga data tidak cukup signifikan untuk menyimpulkan apapun.

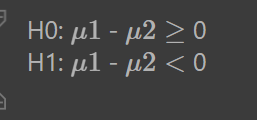
Cara Modern

Berikut merupakan hasil perhitungan dengan cara modern.

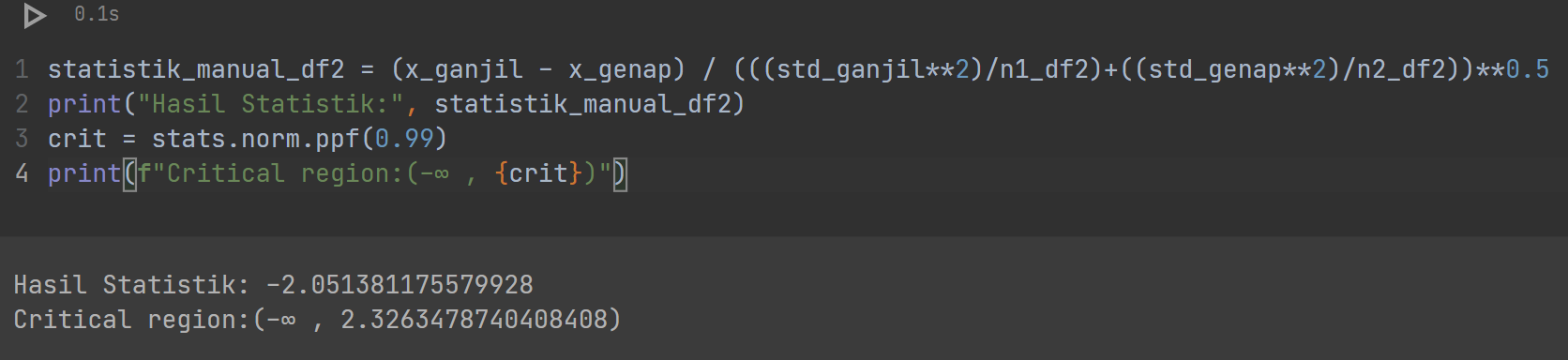


Dapat dilihat dari hasil perhitungan diatas bahwa p value lebih besar dari alpha. Maka dari itu H0 tidak ditolak sehingga data tidak cukup signifikan untuk menyimpulkan apapun.

Agar diperoleh hasil yang diinginkan yaitu H0 ditolak, maka hal yang dapat dilakukan adalah untuk membalik hasil uji hipotesis. Sekarang uji hipotesis nya adalah sebagai berikut.

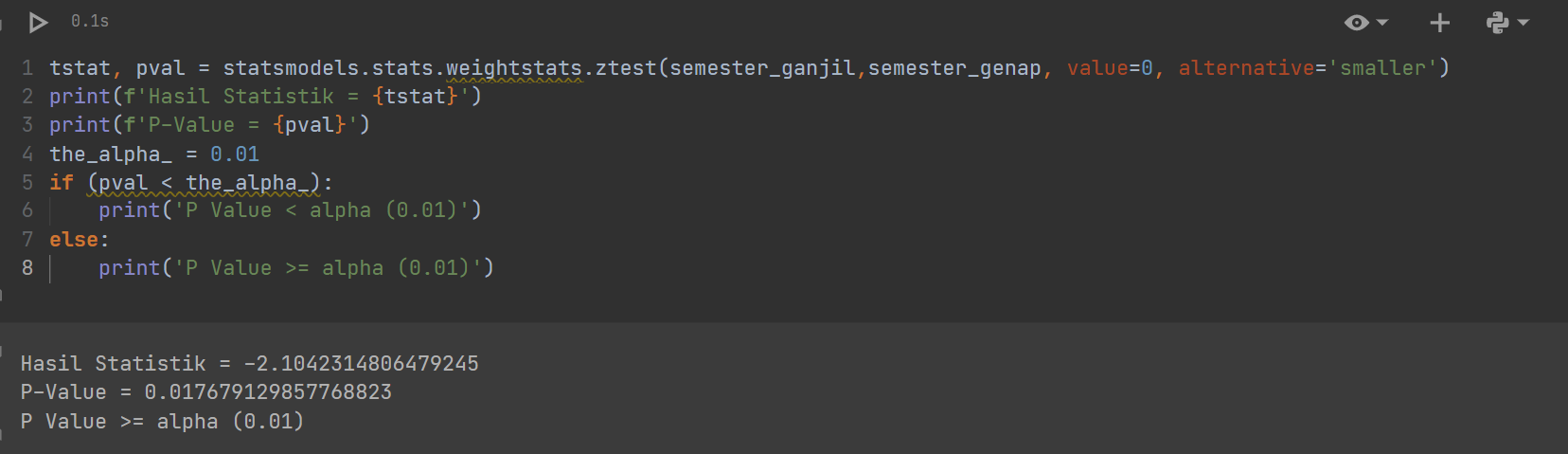


Cara Manual (Setelah Uji Hipotesis dibalik)



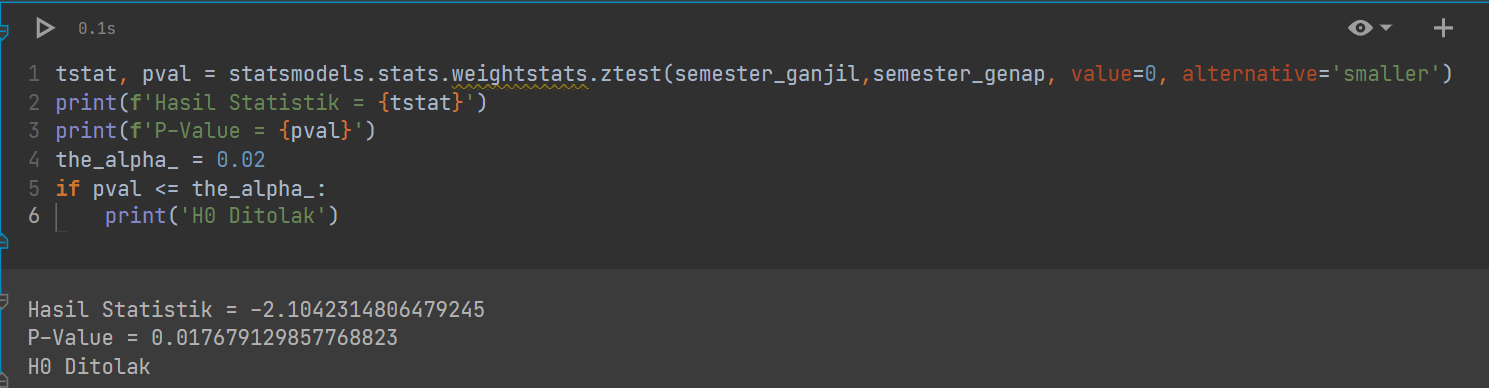
Dapat dilihat dari hasil diatas bahwa hasil statistik berada di critical region, maka dari itu H0 ditolak dan H1 dapat dipertimbangkan. Maka dari itu, dapat ditarik kesimpulan bahwa dengan tingkat kepercayaan 99%, rata-rata mahasiswa yang buta RLAC pada semester ganjil tidak lebih besar dibandingkan dengan semester genap.

Cara Modern (Setelah Uji Hipotesis dibalik)



Dapat dilihat di perhitungan di atas bahwa dengan menggunakan method library statsmodel dari python yaitu statsmodels.stats.weightstats.ztest, p value masih >= alpha. Hal ini berarti bahwa H0 masih tidak ditolak. Kami masih bingung mengapa hasil yang diperoleh demikian, karena dengan menggunakan cara manual, setelah uji hipotesis dibalik, hasilnya menjadi H0 ditolak. Maka dari itu, untuk mengatasi hal ini, kami menaikkan tingkat signifikansi menjadi 0.02

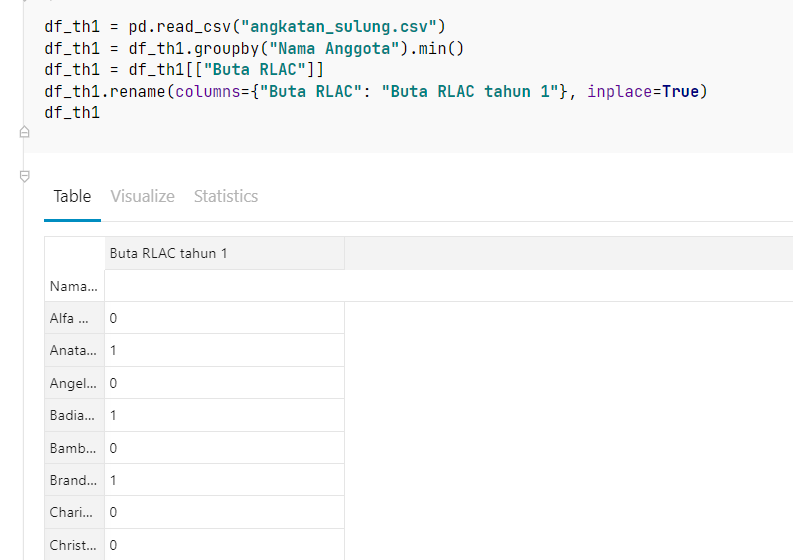
Berikut merupakan hasil perhitungan setelah hasil signifikansi dinaikkan.

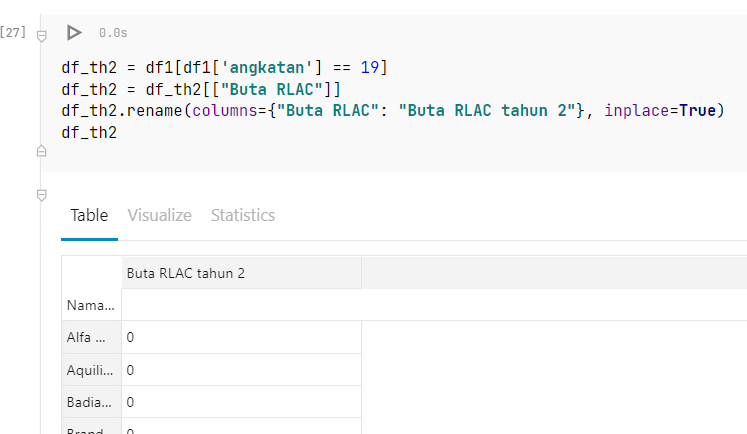


Dapat dilihat setelah menaikkan tingkat signifikansi, sekarang p value <= alpha. Maka dari itu H0 ditolak dan H1 dapat dipertimbangkan. Maka dari itu, dapat ditarik kesimpulan bahwa dengan tingkat kepercayaan 98%, rata-rata mahasiswa yang buta RLAC pada semester ganjil tidak lebih besar dibandingkan dengan semester genap.

**Masalah 3 :**

**Kode pre-prosessing data**

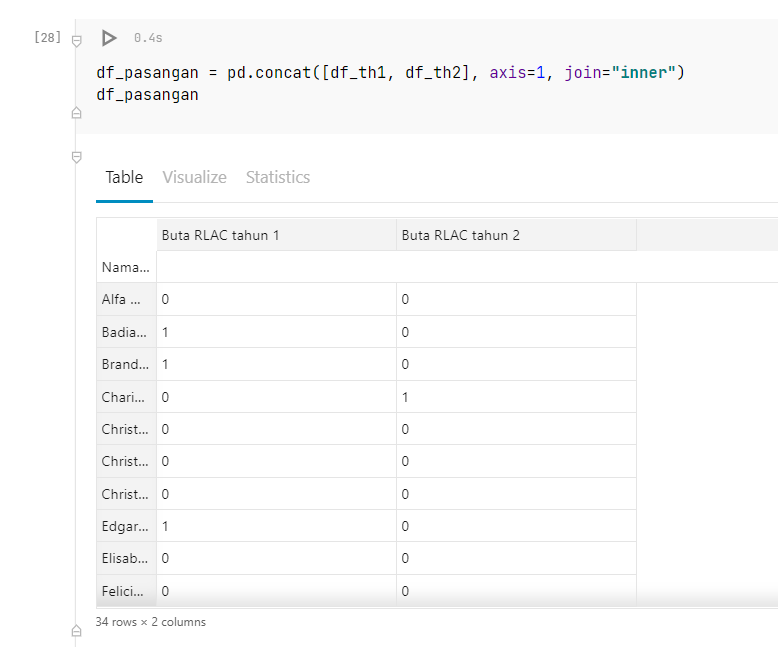




Data sampel untuk tahun pertama kami ambil dari data peminjaman buku angkatan pertama pada tahun pertamanya, dari awal semester ganjil (26-08-2019) sampai akhir semester genap (16-05-2020). Sedangkan untuk data sampel tahun kedua dari pertama kalinya perpustakaan buka di semester antara ditahun ke 3 angkatan pertama (30-09-2021) sampai tanggal 12-04-2022).

Kami membuat kolom “Buta RLAC” dan mengisinya secara manual dengan membandingkan judul buku dan prodi dari mahasiswa tersebut, “1” untuk buta RLAC dan “0” untuk melek RLAC. Jika mahasiswa pernah melakukan lebih dari satu peminjaman buku, kolom “Buta RLAC” akan ditandai dengan “0” jika setidaknya mahasiswa tersebut pernah sekali meminjam buku RLAC dan ditandai dengan “1” jika tidak pernah meminjam buku RLAC dalam 1 tahun akademik. Strategi tersebut kami lakukan dengan melakukan group by “nama” dan mengambil nilai paling kecil dari kolom “Buta RLAC”.

Kami hanya mengambil kolom “nama” dan “Buta RLAC” untuk tahun pertama dan kedua. Kemudian dilakukan penggabungan inner join berdasarkan kolom nama agar hanya nama yang peminjaman yang terdaftar di kedua tahun yang diambil sebagai data pasangan. Dari 53 baris dari data tahun pertama dan 40 baris dari data tahun kedua, hasil penggabungan dengan inner join menghasilkan 34 baris data berpasangan.



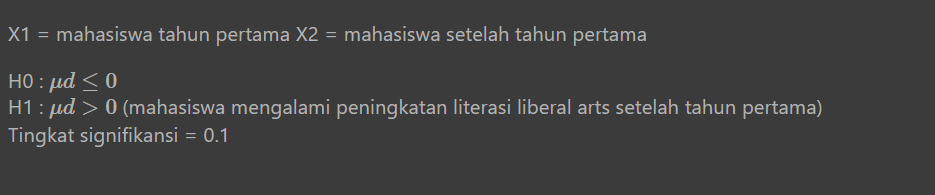
*Apakah benar bahwa mahasiswa CIT mengalami peningkatan literasi Liberal Arts ketika membandingkan pada pengalamannya di tahun pertama dan setelah tahun pertama?*

(rata2 mahasiswa setelah tahun pertama lebih besar dari mahasiswa tahun pertama)

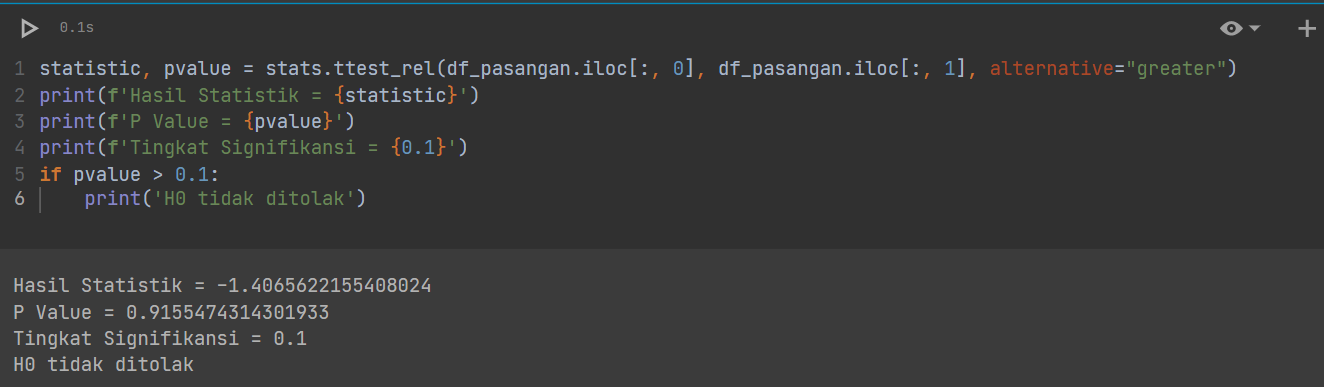


Dapat dilihat dari visualisasi diatas bahwa dari 33 sampel data, 20 orang tidak berubah, 4 orang memiliki peningkatan, dan 9 orang memiliki penurunan.

Misalkan

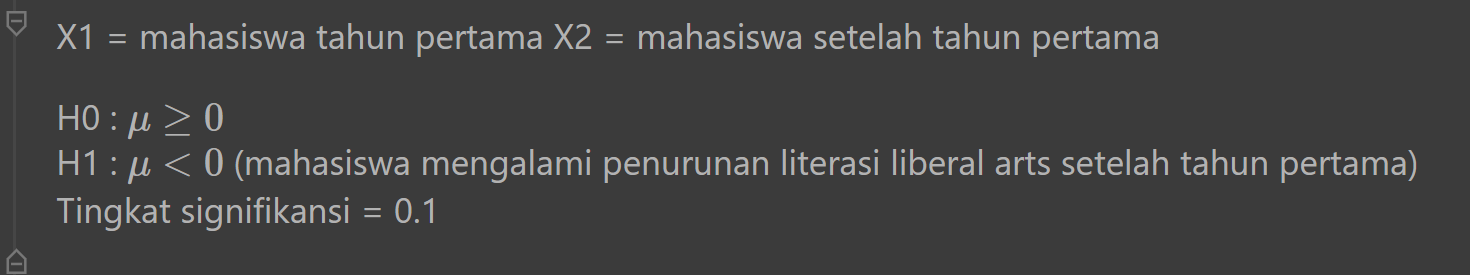


Cara Modern

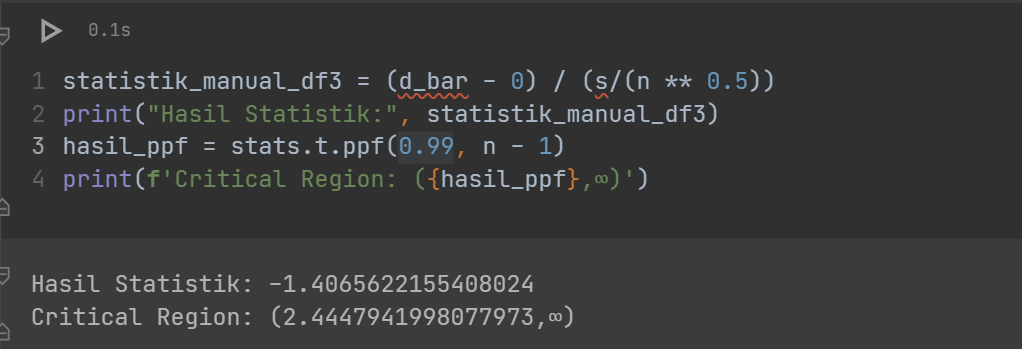


Dapat dilihat dari hasil diatas bahwa p value lebih besar dari tingkat signifikansi, maka dari itu H0 tidak ditolak dan hasil uji hipotesis tidak cukup untuk menyimpulkan apapun.

Membalik Uji Hipotesis

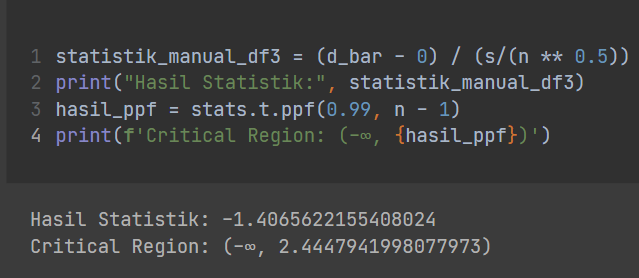


Cara Manual (Sebelum Dibalik)

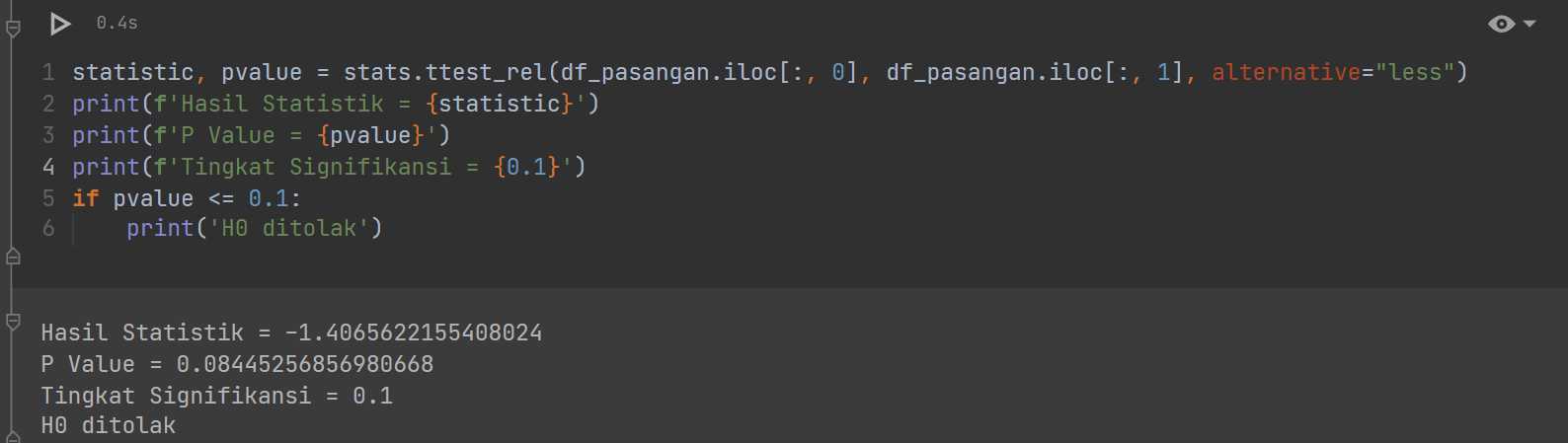


Karena statistik tidak ada di dalam critical region maka kesimpulan Uji Hipotesis : H0 tidak ditolak. Maka dari itu, data tidak cukup signifikan untuk menyimpulkan apapun.

Cara Manual (Setelah Uji Hipotesis Dibalik)



Cara Modern (Setelah Uji Hipotesis Dibalik)



Dapat dilihat setelah hipotesis dibalik, maka sekarang p value lebih kecil dari tingkat signifikansi (0.1) maka dari itu H0 ditolak dan H1 dapat dipertimbangkan. Maka dari itu, dengan tingkat kepercayaan 90% mahasiswa CIT mengalami penurunan literasi liberal arts setelah tahun pertama.

Kesimpulan dan Saran

Kesimpulan

1. Persentase mahasiswa CIT yang tidak melek Liberal Arts kurang dari 75%.
2. Persentase mahasiswa buta RLAC pada semester ganjil lebih banyak daripada semester genap.
3. Mahasiswa CIT mengalami penurunan literasi liberal arts setelah tahun pertama.

Saran

1. Data sampel yang diambil dapat lebih random.
2. Jika dari hasil hipotesis tidak dapat ditarik kesimpulan apapun maka hal-hal berikut dapat dilakukan
   * 1. Tambah data sampel, semakin banyak data sampel maka akan semakin baik
     2. Tambahkan tingkat signifikansi, kurangi tingkat kepercayaan. Tingkat kepercayaan yang semakin rendah artinya range data yang semakin kecil, mungkin setelah tingkat kepercayaan diperkecil, kesimpulan uji hipotesis alternatif dapat dipertimbangkan.
     3. Menukar Hipotesis Null dengan Alternatif.
   1. Gunakan Google untuk memastikan teori dan pemakaian syntax. Untuk kasus statistika hipotesis, tautan berikut dapat digunakan untuk mengecek hasil pekerjaan statistika. <https://www.statssolver.com/hypothesis-testing.html>