Praktikum Eksplorasi dan Visualisasi Data Pertemuan 7

Uji Hipotesis Rata-Rata Populasi Dua Angkatan

Uji hipotesis rata-rata populasi dua angkatan dibagi menjadi dua macam, yaitu :

- Uji hipotesis rata-rata populasi dua angkatan independen
- Uji hipotesis rata-rata populasi dua angkatan dependen (berpasangan)

A. UJI RATA-RATA 2 ANGKATAN INDEPENDEN

Teori

Uji 2 Rata-rata Independen

Uji sampel independen (Unpaired Sample Test) adalah bentuk uji yang bertujuan untuk membandingkan dua set data yang independen atau tidak saling mempengaruhi. Biasanya digunakan untuk menguji hal-hal berikut:

- 1) Perbedaan statistik antara rata-rata dua kelompok
- 2) Perbedaan statistik antara sarana dua intervensi
- 3) Perbedaan statistik antara rata-rata dua skor perubahan

Misalnya:

- 1) Untuk melihat apakah nilai tes rata-rata pria dan wanita berbeda
- 2) Tes menjawab pertanyaan, "Mungkinkah perbedaan ini terjadi secara kebetulan?"

Catatan:

Uji Z atau T Sampel Independen hanya dapat membandingkan mean untuk dua (dan hanya dua) kelompok. Itu tidak dapat membuat perbandingan di antara lebih dari dua kelompok. Jika ingin membandingkan rata-rata di lebih dari dua kelompok, dapat digunakan ANAVA (Analisis Variansi).

Prosedur Uji

Uji Normalitas

Asumsi normalitas dapat dilihat dari bentuk boxplot dua angkatan. Jika dari bentuk boxplot tidak bisa dikatakan mendekati distribusi normal, maka dilakukan transformasi untuk beberapa angkatan.

Uji Kesamaan 2 Variansi

Secara eksplorasi:

- Tampilkan variansi pada setiap variabel
 Catatan: Syntax sudah dipelajari pada pertemuan Ringkasan Numerik.
- Apabila nilai variansi yang tertinggi memiliki nilai 3 kali lebih besar dari nilai variansi yang terendah, maka dianggap terdapat perbedaan variansi. Jadi, asumsi terpenuhi apabila nilai variansi maksimum tidak lebih besar dari 3 kali variansi minimum.

Secara inferensial:

Syntax:

```
var.test(x, y, ratio = 1, conf.level = 0.95, ...)
```

Keterangan argumen:

```
x = data sampel 1
y = data sampel 2
ratio = rasio hipotesis variansi populasi x dan y
(default)
```

conf.level = interval konfidensi

• Hipotesis

 H_0 : Variansi data dan data sama H_1 : Variansi data dan data tidak sama

• Tingkat Signifikansi

α atau (1-Tingkat Kepercayaan)

• Statistik Uji

Gunakan nilai P-value

• Daerah kritik

 H_0 ditolak jika P-value $\leq \alpha$

• Kesimpulan

Karena nilai P-Value = (</>) $0.05 = \alpha$, maka H_0 (ditolak / diterima) Sehingga dapat disimpulkan ... (kesimpulan harus memberikan jawaban dari persoalan).

• Interpretasi

Berdasarkan pengujian yang telah dilakukan dengan *software R*. dilakukan uji ...(nama uji) dengan data ... (nama datanya), ingin diuji apakah ... (persoalan), dengan hipotesis H_0 : ... (kalimat isi H0). Dengan tingkat signifikansi (α) sebesar ... (nilai alpha). Didapatkan hasil P-Value = (</>) $0.05 = \alpha$, maka H_0 (ditolak / diterima) Sehingga dapat disimpulkan ... (kesimpulan harus memberikan jawaban dari persoalan).

* Gunakan secara deskriptif dan inferensi untuk menjawab soal

Uji Hipotesis

```
Syntax:
```

Keterangan argumen:

```
x = Sampel 1
y = Sampel 2
```

alternative = karakter string yang menspesifikasikan hipotesis alternatif (Hanya dipilih 1 saja). Contoh: alternative = "less"

mu = nilai rata-rata selisih antar populasi. Defaultnya adalah 0. Apabila pada kasus mu bukan 0, maka isi dari argumen mu harus diubah.

paired = logika yang menentukan apakah uji digunakan untuk sampel berpasangan (TRUE) atau tidak (FALSE).

conf.level = tingkat konfidensi.

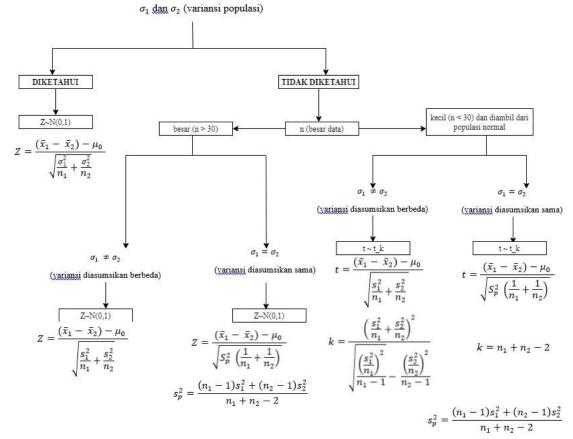
Hipotesis

A.
$$H_0: \mu 1 - \mu 2 = \mu 0$$
 (Uji dua Sisi)
 $H_1: \mu 1 - \mu 2 \neq \mu 0$
B. $H_0: \mu 1 - \mu 2 \leq \mu 0$ (Uji Satu Sisi)
 $H_1: \mu 1 - \mu 2 > \mu 0$
C. $H_0: \mu 1 - \mu 2 \geq \mu 0$ (Uji Satu Sisi)
 $H_1: \mu 1 - \mu 2 \leq \mu 0$

Catatan:

*Saat pengerjaan jangan lupa ditentukan terlebih dahulu sampel 1 (μ 1) dan sampel 2 (μ 2) itu yang mana. *Nilai μ 0 sesuai dengan yang ingin diuji

- Tingkat Signifikansi
 α atau (1-Tingkat Kepercayaan)
- Statistik Uji
 Berdasarkan nilai pvalue atau thitung dan Zhitung berikut ini



Daerah Kritik

 H_0 ditolak jika P-value $< \alpha$ Jika menggunakan perhitungan Z atau t.

 H_0 ditolak jika,

A.
$$|Z| > Z_{a/2}$$
 $|t| > t_{(a/2;k)}$

B. $Z > Z_a$
 $t > t_{(a;k)}$

C. $Z < -Z_a$
 $t < -t_{(a;k)}$

* $k = (n_1-1) + (n_2-1)$

Kesimpulan

Karena nilai P-Value = (</>) $0.05 = \alpha$, dan $Z/t = ... (</>)...=Zcrit/tcrit maka <math>H_0$ (ditolak / diterima) Sehingga dapat disimpulkan ... (kesimpulan harus memberikan jawaban dari persoalan).

*contoh minimal kesimpulan (Gunakan bahasa dan pemahaman sendiri dalam menjelaskan uji yang telah dilakukan)

• Interpretasi

Berdasarkan pengujian yang telah dilakukan dengan *software R*. dilakukan uji ...(nama uji) dengan data ... (nama datanya), ingin diuji apakah ... (persoalan), dengan hipotesis H_0 : ... (kalimat isi H0). Dengan tingkat signifikansi (α) sebesar ... (nilai alpha). Didapatkan hasil P-Value = (</>) $0.05 = \alpha$, dan $Z = ... (</>)...=Zcrit maka <math>H_0$ (ditolak / diterima) Sehingga dapat disimpulkan ... (kesimpulan harus memberikan jawaban dari persoalan).**contoh minimal interpretasi (Gunakan bahasa dan pemahaman sendiri dalam menjelaskan uji yang telah dilakukan*)

CONTOH SOAL

1. Suatu sekolah mencoba menerapkan metode pembelajaran baru. Untuk menguji apakah metode baru tersebut lebih efektif dibandingkan dengan metode standar yang telah diterapkan sebelumnya, sekolah mengambil data dari masing-masing 9 siswa dari kelas A yang sudah menerapkan metode pembelajaran baru dan kelas B yang masih menerapkan metode pembelajaran standar. Berikut merupakan tabel data nilai siswa kelas A (baru) dan B (standar):

Baru	82	87	85	78	91	94	85	81	84
Standar	85	91	79	75	84	95	77	84	81

Penyelesaian:

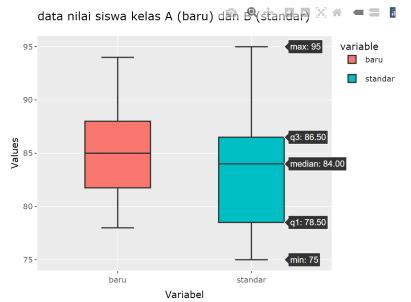
Dari soal di atas, diketahui bahwa terdapat dua metode pembelajaran, yaitu metode baru dan metode standar. Sampel dari tiap-tiap metode diambil dari dua populasi yang berbeda. Dalam hal ini sampel metode baru diambil dari kelas A dan sampel metode lama diambil dari kelas B. Oleh karena itu, data yang diambil bukan data berpasangan (saling independen). Maka jenis analisis yang tepat digunakan adalah uji rata-rata dua populasi independen. Ada 2 asumsi yang harus dipenuhi dalam uji rata-rata dua populasi independen yaitu asumsi normalitas dan asumsi kesamaan variansi.

a. Uji Asumsi Normalitas

```
Syntax:
```

```
# Uji 2 Rata-Rata Independen
# library
library(tidyverse)
library(plotly) # untuk plot
library(reshape) # untuk reshape/melt data
# Data
baru = c(82,87,85,78,91,94,85,81,84)
standar = c(85, 91, 79, 75, 84, 95, 77, 84, 81)
# Make data frame
data = data.frame(baru, standar)
# melted data
data2 <- melt(data)</pre>
# Boxplot
x=ggplot(data2, aes(x=variable, y=value, fill=variable)
)+geom boxplot()+
   ggtitle("data nilai siswa kelas A (baru)
                                                  dan B
(standar)")+
  xlab("Variabel") +
  ylab("Values")
ggplotly(x)
```

Output:



Interpretasi:

Dari boxplot di atas, dapat dilihat distribusi data nilai siswa dengan metode pembelajaran baru dan standar. Pada boxplot metode pembelajaran baru dapat dilihat letak median berada di tengah. Hal ini menunjukkan sebaran nilai atas dan nilai bawah merata. Selain itu, tidak terdapat outlier. Oleh karena itu dapat disimpulkan bahwa data nilai siswa dengan metode pembelajaran baru berdistribusi normal. Sedangkan boxplot data nilai dengan metode pembelajaran standar, dapat dilihat bahwa median sedikit lebih dekat dengan kuartil atas atau dengan kata lain nilai rendah lebih menyebar. Hal ini menunjukkan bahwa distribusi data nilai dengan metode standar cenderung menjurai ke bawah. Selain itu, tidak ada outlier. Karena letak median yang tidak begitu jauh dari tengah dan tidak adanya outlier, maka dapat diasumsikan data nilai dengan metode standar berdistribusi normal. Karena kedua data berdistribusi normal, maka asumsi normalitas terpenuhi.

b. Uji Asumsi Kesamaan Variansi

Syntax:

```
#Asumsi Kesamaan Dua Variansi
# Secara deskriptif
#Variansi dari variabel baru
var(baru)
#Variansi dari variabel standar
var(standar)

#Secara inferensi
var.test(baru, standar, ratio=1, alternative="two.sided", conf.level = 0.95)
```

Secara deskriptif:

Output:

```
> var (baru)
[1] 24.44444
> var (standar)
[1] 41.52778
```

Interpretasi:

Dari output tersebut terlihat bahwa variansi tertinggi yaitu metode standar sebesar 41.52778. Dimana nilai tersebut tidak lebih tinggi dari 3 kali variansi metode baru yaitu 3 x 24.44444 = 73.33332. Maka, dapat disimpulkan bahwa terdapat kesamaan variansi antara kedua data.

Secara Inferensi:

```
F test to compare two variances

data: baru and standar
F = 0.58863, num df = 8, denom df = 8, p-value = 0.47
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
0.1327756 2.6095443
sample estimates:
ratio of variances
0.5886288
```

• Hipotesis

 H_0 : Variansi data baru dan data standar tidak berbeda / sama H_1 : Variansi data baru dan data standar berbeda / tidak sama

• Tingkat Signifikansi

 $\alpha = 0.05$

• Statistik Uji

P-value = 0,47

• Daerah kritik

 H_0 ditolak jika P-value $< \alpha$

• Kesimpulan

Karena nilai P-Value = $0.47 > 0.05 = \alpha$, maka H_0 diterima. Sehingga dapat disimpulkan Variansi data baru dan data standar adalah tidak berbeda atau sama .

Interpretasi

Berdasarkan pengujian yang telah dilakukan dengan *software R*. dilakukan uji kesamaan variansi dengan data nilai siswa kelas A (baru) dan B (standar), ingin diuji apakah data nilai siswa kelas A (baru) dan B (standar) memiliki variansi yang sama, dengan hipotesis H_0 : Variansi data baru dan data standar tidak berbeda / sama . Dengan tingkat signifikansi (α) sebesar 0.05. Didapatkan hasil P-Value = 0,47 > 0,05 = α , maka H_0 diterima. Sehingga dapat disimpulkan Variansi data baru dan data standar adalah tidak berbeda atau sama. maka asumsi kesamaan variansi terpenuhi sehingga dapat dilakukan pengujian lebih lanjut.

<u>Uji Hipotesis Rata-rata Dua Populasi Independen</u>

Syntax:

Output:

```
> # tcrit
> qt(0.05,16,lower.tail = F)
[1] 1.745884
```

Hipotesis

 H_0 : $\mu 1 \le \mu 2$ (Rata-rata nilai dengan metode pembelajaran baru kurang dari sama dengan rata-rata nilai dengan metode pembelajaran standar) H_1 : $\mu 1 > \mu 2$ (Rata-rata nilai dengan metode pembelajaran baru lebih dari rata-rata nilai dengan metode pembelajaran standar)

- Tingkat Signifikansi
 - $\alpha = 0.05$
- Statistik Uji P-Value = 0.2604 t = 0.65663
- Daerah Kritik H_0 ditolak jika P-value $< \alpha$ atau $t > tcrit_{(a;k)}$
- Kesimpulan

Karena nilai P-Value = $0.2604 > 0.05 = \alpha$, dan t = 0.65663 < 1.745884= tcrit maka H_0 diterima. Sehingga dapat disimpulkan Rata-rata nilai dengan metode pembelajaran baru kurang dari sama dengan rata-rata nilai dengan metode pembelajaran standar.

Interpretasi

Berdasarkan pengujian yang telah dilakukan dengan *software R*. dilakukan Uji Rata-rata Dua Populasi Independen dengan data nilai siswa kelas A (baru) dan B (standar), ingin diuji apakah apakah metode baru tersebut lebih efektif dibandingkan dengan metode standar yang telah diterapkan sebelumnya, dengan hipotesis H_0 : $\mu 1 \le \mu 2$ (Rata-rata nilai dengan metode pembelajaran baru kurang dari sama dengan rata-rata nilai dengan metode pembelajaran standar). Dengan tingkat signifikansi (α) sebesar 0.005. Didapatkan hasil P-Value = 0.2604 > 0,05 = α , dan t = 0.65663 < 1.745884= tcrit maka H_0 diterima. Sehingga dapat disimpulkan Rata-rata nilai dengan metode

pembelajaran baru kurang dari sama dengan rata-rata nilai dengan metode pembelajaran standar.

B. UJI RATA-RATA 2 ANGKATAN DEPENDEN

Dalam uji hipotesis rata - rata 2 angkatan dependen, kedua angkatan mempunyai kaitan erat. Kaitan yang erat ini terjadi bisa disebabkan oleh efek suatu perlakuan pada objek yang sama tetapi diukur sebelum dan sesudah diberi perlakuan. Contoh sederhana yang menandakan populasi dependen adalah jika terdapat kata data berpasangan (paired), data sebelum dan sesudah (before and after), serta data sampel yang diulang pengukurannya. Adapun, terdapat contoh lain untuk populasi yang diasumsikan merupakan dependen seperti bayi kembar dan pasangan suami istri.

Langkah pengujian:

1. Uji Normalitas

Asumsi normalitas dapat dilihat dari bentuk boxplot dua angkatan. Jika dari bentuk boxplot tidak bisa dikatakan mendekati distribusi normal, maka dilakukan transformasi untuk beberapa angkatan.

2. Uji Hipotesis Rata-Rata Dua Populasi Dependen

Syntax:

```
t.test(x, y, alternative = c("two.sided", "less", "greater"),
mu = 0, paired = TRUE, conf.level = 0.95, ...)
```

Keterangan argumen:

x = Sampel 1

y = Sampel 2

alternative = karakter string yang menspesifikasikan hipotesis alternatif (Hanya dipilih 1 saja). Contoh: alternative = "less"

mu = nilai rata-rata selisih antar populasi. Defaultnya adalah 0. Apabila pada kasus mu bukan 0, maka isi dari argumen mu harus diubah.

paired = logika yang menentukan apakah uji digunakan untuk sampel berpasangan (TRUE) atau tidak (FALSE).

conf.level = tingkat konfidensi.

Uji Hipotesis

a. Hipotesis

```
\begin{split} H0: \mu D &= \mu 0 \text{ vs } H1: \mu D \neq \mu 0 \\ H0: \mu D &\leq \mu 0 \text{ vs } H1: \mu D > \mu 0 \\ H0: \mu D &\geq \mu 0 \text{ vs } H1: \mu D < \mu 0 \end{split}
```

b. Tingkat signifikansi

 $\alpha = \dots$ (Jika tidak diketahui pada soal, maka gunakan $\alpha = 0.05$)

c. Statistik uji

P-value = ...

d. Daerah Kritik

H0 ditolak jika P-value $< \alpha$

e. Kesimpulan

Karena P – Value = α , maka H0 ditolak/tidak ditolak sehingga dapat disimpulkan bahwa ada / tidak ada perbedaan antara ...

Interpretasi:

Setelah dilakukan Uji Hipotesis Mean Dua Populasi Dependen dengan aplikasi RStudio dengan data ..., ingin diuji apakah Dengan tingkat signifikansi ...%, didapatkan nilai P – Value sebesar Adapun daerah kritik H0 ditolak jika P – Value $< \alpha = \ldots$ maka dengan tingkat konfidensi ...%, H0 ditolak/tidak ditolak sehingga dapat disimpulkan bahwa ...

Contoh Kasus

Sebuah obat diklaim mampu menurunkan berat badan orang yang meminumnya secara teratur dalam waktu 3 minggu. Untuk itu Yayasan Lembaga Konsumen Indonesia (YLKI) mengambil data dan diperoleh sebagai berikut:

No Sampel	Sebelum minum obat	Setelah minum obat		
1	78	78		
2	77	73		
3	73	74		
4	65	61		
5	59	59		
6	53	49		
7	61	60		
8	62	62		
9	62	57		
10	60	58		
11	61	65		

Bagaimana kesimpulan yang bisa Anda ambil?

Penyelesaian:

Pertama, perlu menentukan jenis analisis yang tepat digunakan.

Dari soal di atas, ingin diketahui pengaruh obat baru terhadap penurunan berat badan. Data didapatkan dari 11 objek yang sama dan diukur sebelum dan sesudah meminum obat. Oleh karena itu, data yang diambil merupakan data berpasangan sehingga jenis analisis yang tepat digunakan adalah uji rata-rata dua populasi dependen.

*Kedua, melakukan uji asumsi yang digunakan dan jangan lupa beri interpretasi.*Perlu dilakukan uji asumsi, yaitu asumsi normalitas sebelum melakukan uji rata-rata dua populasi dependen.

a. Uji Normalitas

Syntax:

#Asumsi Normalitas

#menginput data

```
sebelum = c(78,77,73,65,59,53,61,62,62,60,61)
sesudah = c(78,73,74,61,59,49,60,62,57,58,65)
```

#membentuk data frame

data d = data.frame(sebelum, sesudah)

#memanggil library

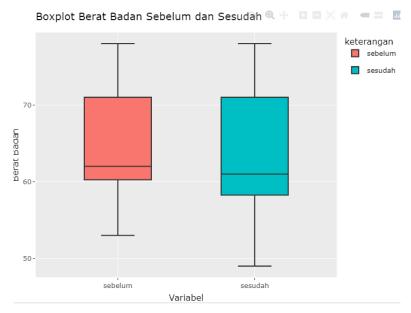
library(tidyverse)

library(plotly)

#menyatukan variabel baru dan standar dari data frame sebagai data d2

```
data_d2 = data_d %>%
  gather(key="keterangan", value="bb")
```

#membuat boxplot variabel baru dan standar



Interpretasi:

Diberikan data berat badan orang sebelum dan sesudah minum obat, ingin diuji apakah kedua data tersebut berdistribusi normal menggunakan boxplot. Dapat dilihat, boxplot dari data variabel berat badan orang sebelum minum obat tidak simetris karena letak garis median cenderung mendekati kuartil bawah (boxplot menjurai ke atas) dan juga tidak memiliki nilai outlier. Oleh karena itu, data berat badan sebelum minum obat dapat dikatakan tidak berdistribusi normal. Selain itu, boxplot data berat badan sesudah minum obat tidak simetris karena letak garis median cenderung mendekati kuartil bawah (boxplot menjurai ke atas) dan tidak memiliki nilai outlier. Oleh karena itu, data berat

badan sesudah minum obat dapat dikatakan tidak berdistribusi normal. Untuk memenuhi asumsi normalitas ini, dilakukan transformasi pada data menggunakan metode Box-cox.

Transformasi menggunakan metode Box-cox. #melakukan transformasi box-cox

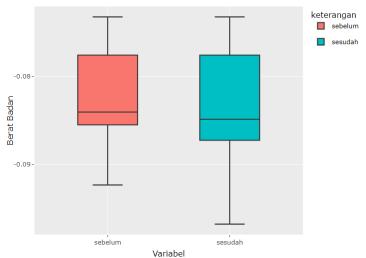
```
library(car)
out = powerTransform(data d2$bb, family = "bcPower")
out$lambda
#melakukan transformasi terhadap data
transformasi.bc=function(data, lambda) {
  for(i in 1:length(data)){
    data[i] = -1*data[i]^lambda
  }
  return(data)
}
data d$sebelum=transformasi.bc(data d$sebelum,-0.6)
data d$sesudah=transformasi.bc(data d$sesudah,-0.6)
datad.bc=gather(data d, key="keterangan", value="tfbb")
                        > out$lambda
                        data_d2$bb
                        -0.6097006
```

Berdasarkan hasil transformasi box-cox, diperoleh nilai λ = -0.6. Transformasi dilakukan dengan memangkatkan tiap datum pada data dengan -0.6 dan setiap dantum dikalikan dengan -1 karena diperoleh nilai lamda negatif.

Kemudian akan dibuat boxplot dengan data hasil transformasi, sebagai berikut #membuat boxplot hasil transformasi

```
b = ggplot(datad.bc, aes(x=keterangan,
y=tfbb,fill=keterangan))+
    geom_boxplot()+ggtitle("Boxplot Transformasi Berat Badan
Sebelum dan Sesudah")+xlab("Variabel")+ylab("Berat Badan")
ggplotly(b)
```





Interpretasi:

Setelah dilakukan transformasi, dapat dilihat pada boxplot bahwa data berat badan sebelum minum obat tidak simetris (data menjurai ke atas) dan tidak memiliki nilai outlier. Selain itu, boxplot data berat badan sesudah minum obat masih tidak simetris (data menjurai ke atas) dan tidak memiliki nilai outlier. Namun, karena sudah dilakukan transformasi, kedua data diasumsikan sudah berdistribusi normal dan dapat dilanjutkan ke tahap berikutnya.

b. Uji Hipotesis Rata-Rata Dua Populasi Dependen

Syntax:

Output

Paired t-test

Hipotesis

H0: μ D \leq 0 H1: μ D > 0

Tingkat signifikansi

 $\alpha = 0.05$

Statistik uji

P-value = 0.05744

Thit = 1.7269

Daerah kritik

H0 ditolak jika P-value $< \alpha$ atau thit $> t(\alpha, n-1)$

```
> #t tabel
> qt(0.95,10)
[1] 1.812461
```

Kesimpulan

Karena P-Value = $0.05744 > \alpha = 0.05$ atau Thit = 1.7269 < tcrit = 1.812461 maka H0 tidak ditolak dan dapat disimpulkan bahwa obat baru tidak menurunkan berat badan secara signifikan.

Interpretasi:

Setelah dilakukan Uji Hipotesis Mean Dua Populasi Dependen dengan aplikasi RStudio dengan data berat badan sebelum dan sesudah minum obat, ingin diuji apakah terdapat penurunan berat badan setelah minum obat. Dengan tingkat signifikansi 5%, didapatkan P-Value = $0.05744 > \alpha = 0.05$ atau Thit = 1.7269 < tcrit=1.812461, maka dengan tingkat konfidensi 95%, H0 tidak ditolak sehingga dapat disimpulkan bahwa obat baru tidak menurunkan berat badan secara signifikan.

Latihan:

1. Dalam perusahaan dua orang asesor menaksir nilai pajak (dalam 10.000 rupiah) dari 8 buah properti.

Properti	Penilai A	Penilai B	
1	76.3	75.1	
2	88.4	86.8	
3	80.2	77.3	
4	94.7	90.6	
5	68.7	69.1	
6	82.8	81.0	
7	76.1	75.3	
8	79.0	79.1	

Apakah benar ada kecenderungan asesor A menaksir lebih tinggi dari asesor B?

2. In a study conducted by the Department of Human Nutrition and Foods at Virginia Tech, the following data were recorded on sorbic acid residuals, in parts per million, in ham immediately after dipping in a sorbate solution and after 60 days of storage:

Before Storage	After Storage
224	116
270	96
400	239
444	329
590	437
660	597
1400	689
680	576

Is there sufficient evidence, at the 0.05 level of significance, to say that the length of storage influences sorbic acid residual concentrations?