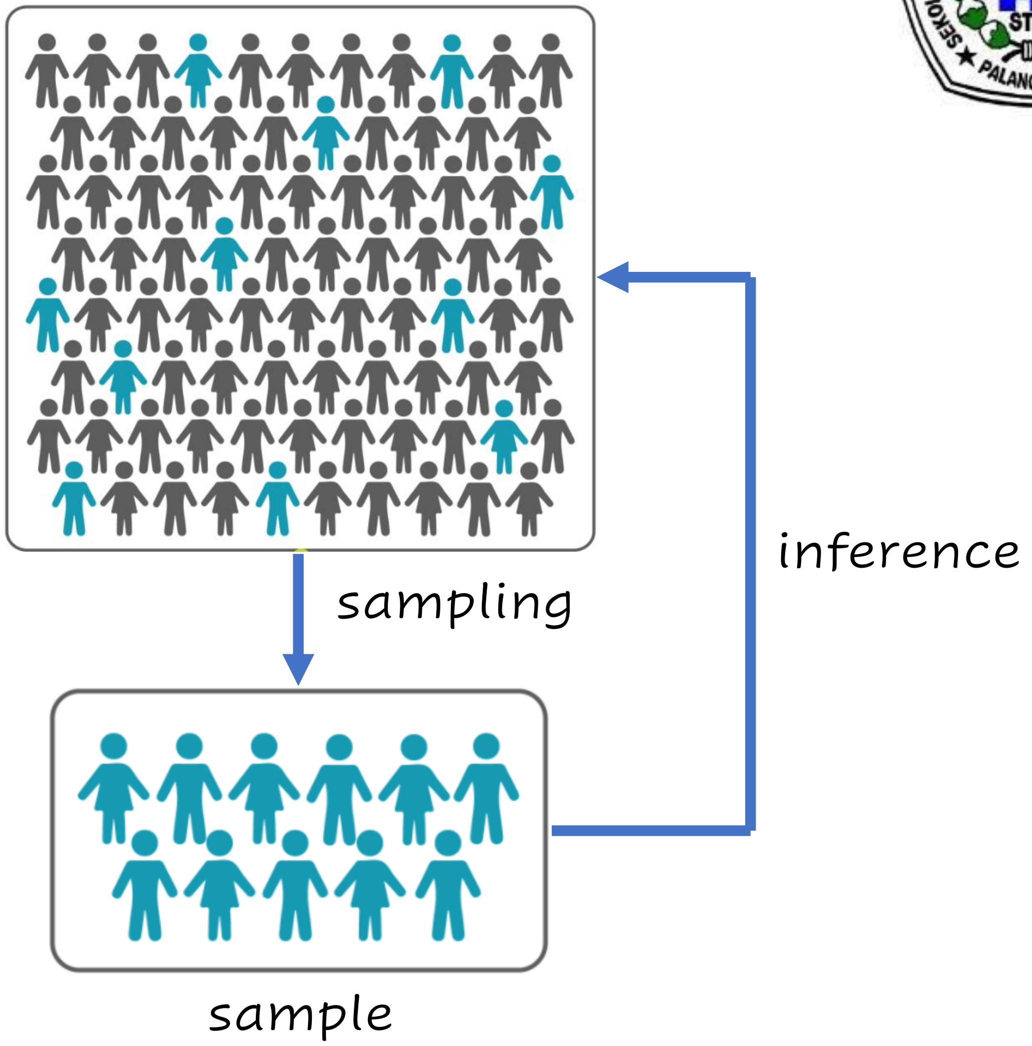
R for Inferential Statistics

Elok Faiqotul Himmah
STMIK Palangkaraya

Statistika Inferensial

- Karakteristik populasi dapat diketahui dari sampel yang kita miliki melalui statistika inferensial

- Statistik Inferensial adalah statistika yang digunakan untuk menganalisis data sampel dan hasilnya diberlakukan untuk populasi. Beberapa hal yang dapat dilakukan adalah menguji hipotesis dengan statistik uji, seperti chi-square test, student-t test, f-test, z-score test, regression, correlation. Statistik Inferensial dapat digunakan untuk konfirmasi dari hasil statistik deskriptif.



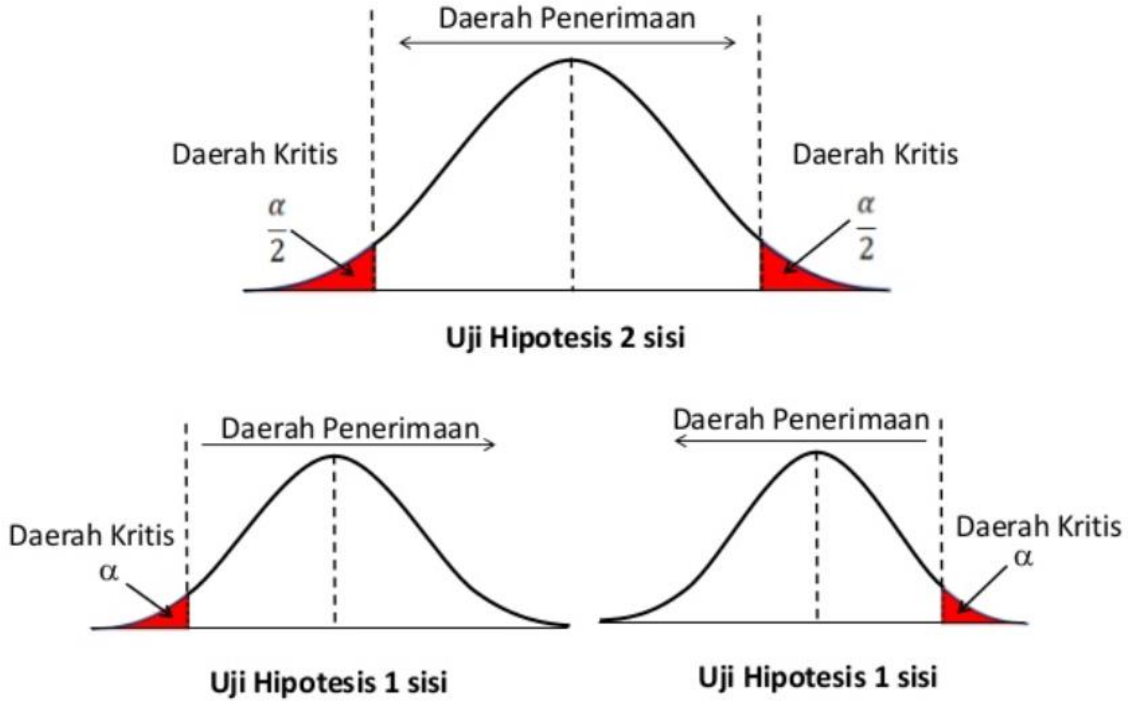
Uji Hipotesis

Ada 2 jenis hipotesis yaitu hipotesis null (hipotesis nihil) dan hipotesis alternatif.

Hipotesis nihil (H0) yaitu hipotesis yang berlawanan dengan teori yang akan dibuktikan. *Hipotesis alternatif* (H1) adalah hipotesis yang berhubungan dengan teori yang akan dibuktikan.

P-value dan alpha (α)

- *P-value* adalah peluang terkecil dalam menolak Ho. Sedangkan alpha (α) adalah tingkat kesalahan. Nilai α biasanya adalah 1%, 5%, dan 10%. Dalam prakteknya α =5% sering digunakan, karena lebih moderat.
- Hipotesis H0 ditolak jika p-value < α, sedangkan jika p-value > α maka H0 diterima.



Uji-t

Untuk membandingkan nilai rata-rata populasi dengan nilai tertentu atau nilai tengah populasi lainnya bisa dilakukan dengan uji-z. Namun uji-z hanya bisa digunakan jika data berdistribusi normal dan ragam (varian) populasi diketahui.

Pada kenyataannya,jarang sekali kita bisa mengetahui nilai parameter suatu populasi dengan pasti, sehingga kita hanya bisa menduga parameter populasi dari sampel yang kita amati.

Dengan demikian,untuk sampel kecil kita bisa gunakan uji-t.

Uji-t

Dalam statistika terdapat tiga macam uji-t yaitu:

- Uji Mean Satu Sampel (*One Sample t-test*)

  Uji hipotesis untuk rata-rata sampel sama dengan nilai tertentu

- Uji Mean Dua Sampel (*Independent Sample t-test*)

  Uji hipotesis perbedaan dua nilai rata-rata sampel acak dengan varian sama

- Uji Mean Dua Sampel Berpasangan (*Paired Sample t-test*)

  Uji hipotesis untuk rata-rata dua sampel berpasangan. Dua sampel berpasangan adalah sampel dengan subjek sama tetapi menjalani dua perlakuan atau pengukuran yang berbeda.

One Sample t-test

One Sample t-test merupakan suatu uji hipotesis terhadap nilai rata-rata suatu populasi dengan nilai tertentu.

Berikut adalah format sintaks one sample t-test menggunakan R:

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

- x merupakan data input
- Alternative adalah argumen bernilai karakter yang menandakan arah hipotesis alternatif: “two sided”->uji hipotesis dua arah; “less”->arah kiri; “greater”->arah kanan.
- Mu adalah argumen yang bernilai numerik sebagai pembanding nilai rata-rata
- Conf.level adalah tingkat kepercayaan uji hipotesis.

One Sample t-test

Uji-t Satu Arah (Hipotesis Alternatif (H1): lebih dari)

Hipotesis pada uji-t satu arah “lebih dari” dapat diekspresikan sebagai berikut:

$$H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0$$

Pengambilan keputusan:

tolak H_0 apabila t hitung > t tabel atau $\left(t_0 > t_{\alpha;v=n-1}\right)$

Atau

tolak H_0 apabila $\boxed{\textbf{p-value} < \alpha}$

One Sample t-test

Uji-t Satu Arah (Hipotesis Alternatif (H1): lebih dari)

Latihan 1.

Seorang guru ingin menguji apakah rata-rata tinggi badan 14 siswa lebih dari 159 cm. Lakukan uji t satu arah dengan tingkat kepercayaan 90%. Data tinggi badan siswa diasumsikan berdistribusi normal.

$$H_0: \mu = \mu_0$$
$$H_1: \mu > \mu_0$$

One Sample t-test

Uji-t Satu Arah (Hipotesis Alternatif (H1): lebih dari)

Data diberikan oleh Tabel 1.

Tabel 1. Data Tinggi dan Berat Badan Siswa

No.	Tinggi	Berat	Gender
1	174,5	65,8	pria
2	178,6	65,8	pria
3	170,8	66,4	pria
4	168,2	68,9	pria
5	159,7	67,8	pria
6	167,8	67,8	pria
7	165,5	65,8	pria
8	154,7	48,7	wanita
9	152,7	45,7	wanita
10	155,8	46,2	wanita
11	154,8	43,8	wanita
12	157,8	58,1	wanita
13	156,7	54,7	wanita
14	154,7	49,7	wanita

1. Menetapkan hipotesis uji dan hipotesis alternatifnya

$$H_0: \mu = 159$$
$$H_1: \mu > 159$$

2. Meng-import data

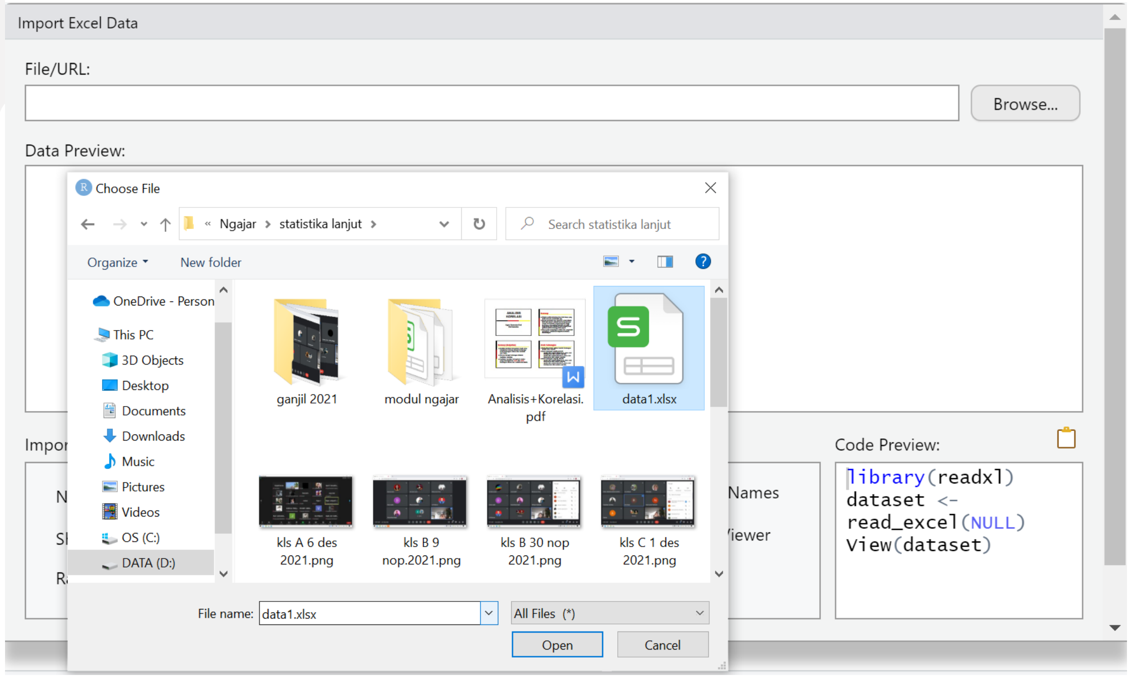
   Import data siswa yang telah diketik di Ms. Excel dan disimpan dengan nama "data1"

Import Data

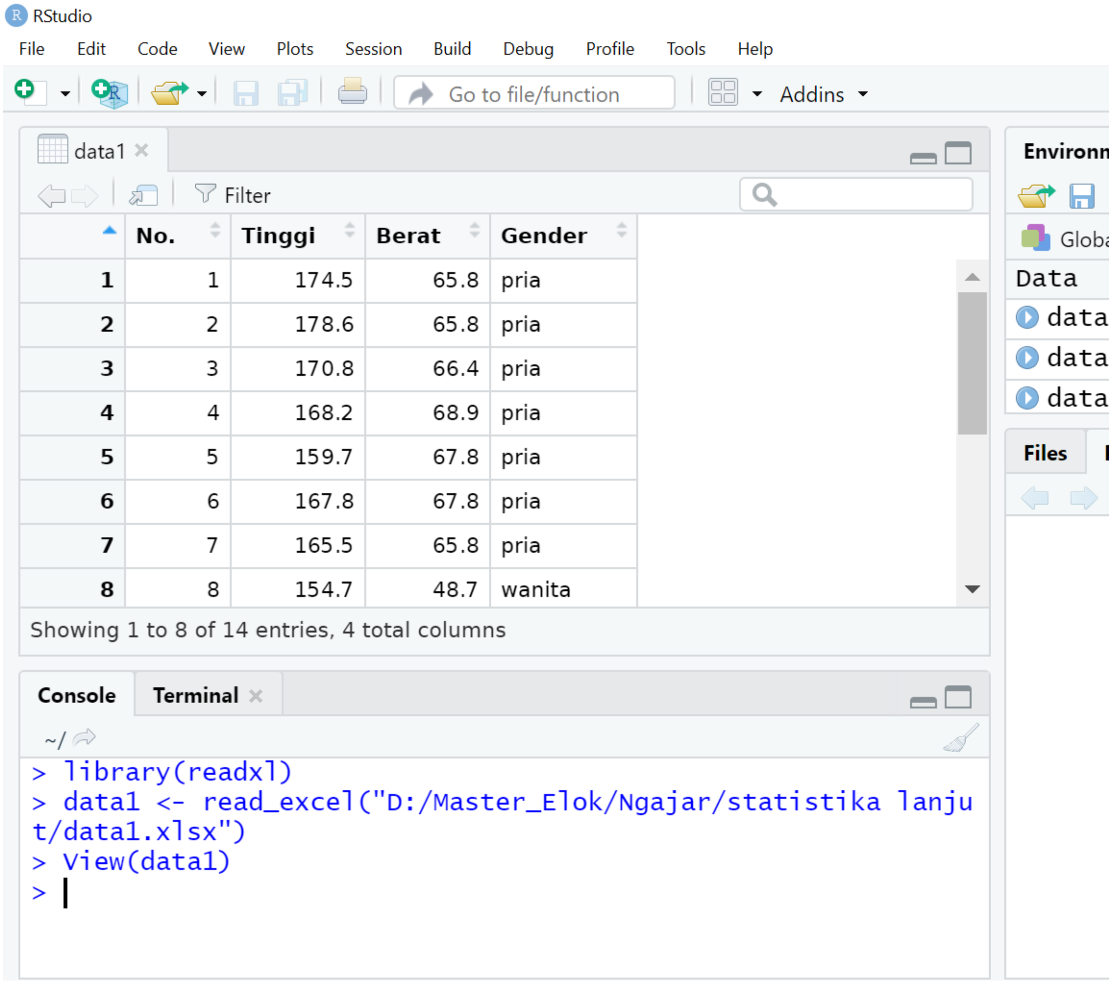
Klik Import Dataset->pilih From Excel

Setelah muncul kotak dialog Import Excel Data, klik Browse

Setelah muncul kotak dialog Choose File, pilih file data1.xlsx ->klik Open->klik Import



- Data hasil import akan muncul di jendela editor seperti berikut ini



One Sample t-test

Uji-t Satu Arah (Hipotesis Alternatif (H1): lebih dari)

3. Menghitung statistik uji

```
> #mencari t hitung dan p-value
> t.test(data1$Tinggi,alternative="greater",mu=159,conf.lev
el=0.90)

        One Sample t-test

data:  data1$Tinggi
t = 1.4657, df = 13, p-value = 0.08325
alternative hypothesis: true mean is greater than 159
90 percent confidence interval:
 159.2606      Inf
sample estimates:
mean of x
 162.3071

> #mencari nilai t tabel dengan alpha 10%, v=n-1
> qt(0.10,df=13,lower.tail = FALSE)
[1] 1.350171
```

4. Uji Hipotesis

Pengambilan keputusan:

tolak H_0 apabila t hitung > t tabel

atau

tolak H_0 apabila *p-value*< α

Karena nilai t hitung (1.4657)>t tabel (1.350171) atau p-value (0.08325)<alpha(0.10) maka H0 ditolak->H1 diterima. Jadi, berdasarkan data tersebut, rata-rata tinggi badan siswa saat ini dengan kepercayaan 90% lebih dari 159 cm.

One Sample t-test

Uji-t Satu Arah (Hipotesis Alternatif (H1): kurang dari)

Latihan 2.

Seorang guru ingin menguji apakah rata-rata tinggi badan 14 siswa kurang dari 168 cm. Lakukan uji t satu arah dengan tingkat kepercayaan 90%. Data tinggi badan siswa diasumsikan berdistribusi normal. Gunakan data pada Latihan 1.

One Sample t-test

Uji-t Satu Arah (Hipotesis Alternatif (H1): kurang dari)

Penyelesaian.

1. Menetapkan hipotesis uji dan hipotesis alternatifnya

$$H_0: \mu = 168$$
$$H_1: \mu < 168$$

2. Menghitung statistik uji

```
> #mencari t hitung dan p-value
> t.test(data1$Tinggi,alternative="less",mu=168,conf.level=
0.90)

        One Sample t-test

data:  data1$Tinggi
t = -2.523, df = 13, p-value = 0.01273
alternative hypothesis: true mean is less than 168
90 percent confidence interval:
     -Inf 165.3537
sample estimates:
mean of x
 162.3071

> #mencari nilai t tabel dengan alpha 10%, v=n-1
> qt(0.10,df=13,lower.tail = FALSE)
[1] 1.350171
```

3. Uji Hipotesis

Pengambilan keputusan:

tolak H_0 apabila t hitung > t tabel

atau

tolak H_0 apabila *p-value*< α

Karena nilai t hitung (-2.523)<-t tabel (-1.350171) atau p-value (0.01273)<alpha(0.10) maka H0 ditolak->H1 diterima. Jadi, berdasarkan data tersebut, rata-rata tinggi badan siswa saat ini dengan kepercayaan 90% kurang dari 168 cm.

One Sample t-test

Uji-t Dua Arah

(Hipotesis Alternatif (H1): tidak sama dengan)

Latihan 3.

Seorang guru ingin menguji apakah rata-rata berat badan 14 siswa sama dengan 58 kg. Lakukan uji t dua arah dengan tingkat kepercayaan 90%. Data tinggi badan siswa diasumsikan berdistribusi normal. Gunakan data pada Latihan 1.

One Sample t-test

Uji-t Dua Arah
(Hipotesis Alternatif (H1): tidak sama dengan)

Penyelesaian.

1. Menetapkan hipotesis uji dan hipotesis alternatifnya

$$H_0: \mu = 58$$
$$H_1: \mu \neq 58$$

2. Menghitung statistik uji

```
> t.test(data1$Berat,alternative="two.sided",mu=58,conf.level=0.90)

        One Sample t-test

data:  data1$Berat
t = 0.088222, df = 13, p-value = 0.931
alternative hypothesis: true mean is not equal to 58
90 percent confidence interval:
 53.64032 62.81682
sample estimates:
mean of x
 58.22857

> #mencari nilai t tabel dengan fungsi qt([alpha=10%]/2,v=n-1
> alpha<-0.10
> t.table<-abs(qt(alpha/2,df=13))
> c(-t.table,t.table)
[1] -1.770933  1.770933
```

3. Uji Hipotesis

Pengambilan keputusan:

tolak H_0 apabila t hitung > t tabel

atau

tolak H_0 apabila *p-value*< α

Karena nilai t hitung (0.088222)<t tabel (1.770933) atau p-value (0.931)>alpha(0.10) maka H0 diterima. Jadi, berdasarkan data tersebut, rata-rata berat badan siswa saat ini dengan kepercayaan 90% sama dengan 58 kg.

Independent Sample t-test

- *Independent Sample t-test* digunakan untuk mengetahui ada tidaknya perbedaan rata-rata hitung antaradua kelompok sampel yang tidak berhubungan.
- Kedua kelompok sampel itu harus diketahui ragam (varian)nya apakah sama atau tidak. Oleh karena itu, sebelum melakukan uji rataan dua populasi sebaiknya terlebih dahulu dilakukan uji kesamaan ragam (varian) mnggunakan Fisher's Test (Uji-F) untuk melihat kehomogenan kedua kelompok sampel tersebut.

Berikut adalah sintaks dasar uji ragam (varians) dua sampel:

```
var.test(x,y)
```

Sedangkan sintaks dasar independent sample t-test adalah:

```
t.test(x,y,var.equal = ,conf.level = )
```

- x merupakan data sampel 1
- y merupakan data sampel 2
- Var.equal adalah argumen yang bernilai TRUE dan FALSE. TRUE bila varian kedua sampel sama.
- Conf.level adalah tingkat kepercayaan uji hipotesis.

Independent Sample t-test

Latihan 4.

Seorang dosen menduga bahwa nilai ujian Statistika kelas A dan kelas B adalah sama. Kedua sampel diasumsikan berasal dari populasi berdistribusi normal. Lakukan uji hipotesis dengan tingkat kepercayaan 95%. Gunakan data pada Tabel 2.

Tabel 2. Data Nilai Statistika

No.	Kelas A	Kelas B
1	58	96
2	56	85
3	84	45
4	56	85
5	75	87
6	65	86
7	68	32
8	95	58
9	84	65
10	76	65
11	56	69
12	63	98
13	54	85
14	68	74
15	67	68
16	75	80
17	88	75

Independent Sample t-test

Penyelesaian.

Melakukan uji perbedaan varians (uji homogenitas)

1. Menetapkan hipotesis uji dan hipotesis alternatifnya

$$H_0: {\sigma_1}^2 = {\sigma_2}^2$$
$$H_1: {\sigma_1}^2 \neq {\sigma_2}^2$$

2. Meng-import data

   Import data siswa yang telah diketik di Ms. Excel dan disimpan dengan nama "data2"

Independent Sample t-test

Melakukan uji perbedaan varians (uji homogenitas)

3. Menghitung statistika uji

```
> #uji homogenitas (uji-F)
> var.test(data2$`Kelas A`,data2$`Kelas B`)

        F test to compare two variances

data:  data2$`Kelas A` and data2$`Kelas B`
F = 0.4258, num df = 16, denom df = 13, p-value =
0.108
alternative hypothesis: true ratio of variances is not equa
l to 1
95 percent confidence interval:
 0.1406716 1.2137684
sample estimates:
ratio of variances
         0.4258003

> #mencari nilai F tabel dengan fungsi qf(p,df.num,df.den)
> qf(0.95,16,13)
[1] 2.51492
```

4. Uji Hipotesis

Pengambilan keputusan:

tolak H_0 apabila F hitung > F tabel

atau

tolak H_0 apabila *p-value*< α

Karena Fhitung(0.4258)<Ftabel (2.51492) atau p-value(0.108)>alpha(0.05) maka H0 diterima.

Jadi, dapat disimpulkan bahwa varians kedua populasi adalah sama.

Independent Sample t-test

5. Menghitung statistik uji

   Karena varians kedua populasi sama, maka kita tambahkan argumen var.equal=TRUE dalam fungsi T.test().

```
> #mencari nilai t hitung
> t.test(data2$`Kelas A`,data2$`Kelas B`,var.equal=TRUE)

        Two Sample t-test

data:  data2$`Kelas A` and data2$`Kelas B`
t = -0.64483, df = 29, p-value = 0.5241
alternative hypothesis: true difference in means is not equ
al to 0
95 percent confidence interval:
 -15.389877   8.011725
sample estimates:
mean of x mean of y
 69.88235  73.57143

> #mencari t tabel dengan fungsi qt(alpha/2,v=n1+n2-2)
> alpha<-0.05
> t.table<-abs(qt(alpha/2,df=29))
> c(-t.table,t.table)
[1] -2.04523  2.04523
```

4. Uji Hipotesis

   Pengambilan keputusan:

   tolak H_0 apabila t hitung > t tabel

   atau

   tolak H_0 apabila *p-value*< α

   Karena t hitung(-0.64483)<-t tabel (-2.04523) atau p-value(0.5241)>alpha(0.05) maka H0 diterima.

   Jadi, dapat disimpulkan bahwa dengan tingkat kepercayaan 95%, rata-rata nilai statistika kelas A dan B adalah sama.

Paired Sample t-test

- *Paired Sample t-test atau* uji-t sampel berpasangan merupakan uji-t untuk data sampel berpasanan membandingkan rata-rata dua variabel untuk suatu populasi
- *Paired Sample t-test* ini biasanya digunakan pada penelitian yang membandingkan reaksi respon sebelum dan sesudah diberi perlakuan.

Sintaks dasar:

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

- x merupakan data input
- y merupakan data input
- Alternative adalah arah uji hipotesis
- Paired adalah argumen bernilai TRUE dan FALSE. Untuk paired-t test harus bernilai TRUE
- Conf.level adalah tingkat kepercayaan uji hipotesis

Paired Sample t-test

Latihan 5.

Seorang peneliti ingin melihat kehandalan suatu obat pelangsing. Peneliti tersebut menduga bahwa setelah meminum obat pelangsing selama seminggu, berat badan akan turun. Data diasumsikan berasal dari populasi berdistribusi normal. Lakukan uji hipotesis dengan tingkat kepercayaan 95%.

No.	Sebelum	Sesudah
1	65,8	60
2	65,8	62
3	66,4	64
4	68,9	65
5	67,8	66
6	67,8	60
7	65,8	63
8	48,7	48
9	45,8	45
10	55,4	50
11	65,1	60
12	58,1	55
13	49,7	48
14	48,5	46

Paired Sample t-test

Penyelesaian.

1. Menetapkan hipotesis uji dan hipotesis alternatifnya

$$H_0: \mu_{sesudah} = \mu_{sebelum}$$
$$H_1: \mu_{sesudah} < \mu_{sebelum}$$

2. Meng-import data

   Import data siswa yang telah diketik di Ms. Excel dan disimpan dengan nama “data3”.

Paired Sample t-test

3. Menghitung statistik uji untuk sampel berpasangan

```
> #paired sample t-test
> #menghitung statistik uji
> t.test(data3$Sesudah,data3$Sebelum,alternative = "less",pair
ed = TRUE,conf.level = 0.95)

        Paired t-test

data:  data3$Sesudah and data3$Sebelum
t = -6.2315, df = 13, p-value = 1.53e-05
alternative hypothesis: true difference in means is less than
 0
95 percent confidence interval:
      -Inf -2.433758
sample estimates:
mean of the differences
                   -3.4
> #mencari t tabel
> qt(0.05,13,lower.tail = TRUE)
[1] -1.770933
```

4. Uji Hipotesis

Pengambilan keputusan:

tolak H_0 apabila t hitung > t tabel

atau

tolak H_0 apabila *p-value*< α

Berdasarkan output diperoleh nilai t hitung= -6.2315<t tabel=-1.770933 atau p-value=1.53e-05<alpha=0.05 sehingga dapat disimpulkan bahwa tidak cukup bukti untuk menerima H0, dengan kata lain Ho ditolak->H1 diterima.
Jadi, berdasarkan data tersebut, setelah meminum obat pelangsing selama seminggu, rata-rata berat badan akan turun dengan tingkat kepercayaan 95%.

Uji ANOVA (satu arah)

- *Anova satu arah (One way anova/Anava) digunakan untuk menguji apakah rata-rata dua atau lebih kelompok data sama atau berbeda, dilihat dari variansnya.*
- Anova satu arah merupakan pengamatan dengan satu kriteria, terdiri dari satu variable bebas dengan dua atau lebih krtiteria.
- Uji anova dengan R dilakukan dengan menggunakan fungsi **aov()** dan **summary()**.

Contoh.

Seorang petani ingin mengetahui apakah ada perbedaan signifikan dalam tinggi tanaman yang menggunakan tiga jenis pupuk berbeda (A,B, dan C). Data diberikan sebagai berikut.

Responden	Tinggi Tanaman (cm)		
	Metode A	Metode B	Metode C
1	15	12	20
2	18	14	22
3	17	13	19
4	16	15	21
5	19	11	23

Lakukan uji anova satu arah untuk membantu petani tersebut!

Uji ANOVA (satu arah)

Penyelesaian.

1. Menetapkan hipotesis uji dan hipotesis alternatifnya

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_n$$
$$H_1: \mu_1 \neq \mu_2 = \cdots = \mu_n$$

2. Meng-import data

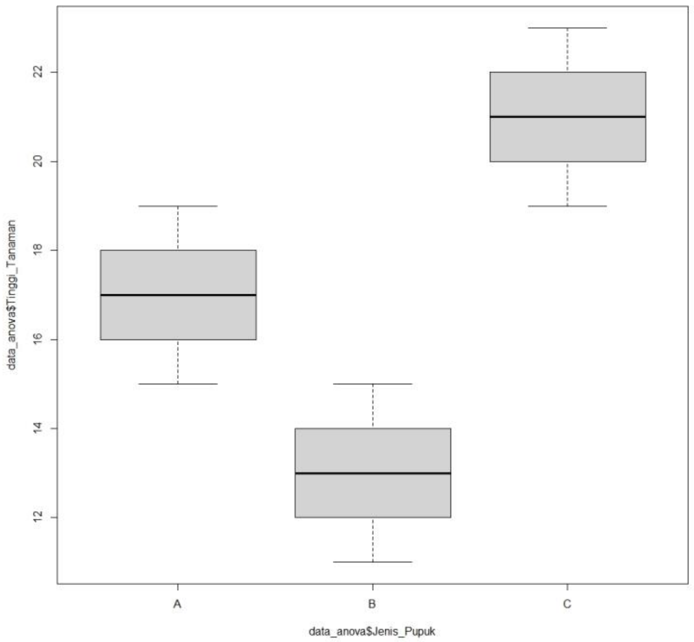
Import data siswa yang telah diketik di Ms. Excel dan disimpan dengan nama "data_anova". Data disusun dalam excel seperti tampak pada gambar, dimana jenis pupuk merupakan variable bebas dengan 3 kategori (A,B,C) yang mempengaruhi tinggi tanaman.

	Tinggi_Tanaman	Jenis_Pupuk
1	15	A
2	18	A
3	17	A
4	16	A
5	19	A
6	12	B
7	14	B
8	13	B
9	15	B
10	11	B
11	20	C
12	22	C
13	19	C
14	21	C
15	23	C

Uji Anova (Satu Arah)

3. Lakukan eksplorasi data dengan menggunakan fungsi boxplot()

```
> boxplot(data_anova$Tinggi_Tanaman~data_anova$Jenis_Pupuk)
```



Boxplot tersebut menunjukkan bahwa rata-rata tinggi tanaman dengan jenis pupuk B lebih rendah daripada tinggi tanaman dengan pupuk jenis A, sedangkan tinggi tanaman dengan jenis pupuk C lebih tinggi dibandingkan dengan rata-rata lainnya.

Uji Anova (Satu Arah)

4. Lakukan uji Anova satu arah dengan variable bebas Jenis_Pupuk.

   1) Hipotesis:

   $H_0: \mu_A = \mu_B = \mu_C$ (rata-rata tinggi tanaman untuk ketiga jenis adalah sama)

   $H_1: \mu_1 \neq \mu_j$ (minimal ada 1 pasang jenis pupuk (i,j) yang mempunyai rata-rata tinggi tanaman terhadap jenis pupuk yang berbeda

   2) Kriteria pengambilan Keputusan: Tolak H_0 jika p-value<α.

   3). Gunakan fungsi aov() dan summary() untuk menguji apakah terdapat perbedaan signifikan secara keseluruhan antara rata-rata tinggi tanaman untuk jenis pupuk berbeda.

```
> model<-aov(data_anova$Tinggi_Tanaman~data_anova$Jenis_Pupuk)#untuk melihat keragaman
> summary(model)
                       Df Sum Sq Mean Sq F value   Pr(>F)    
data_anova$Jenis_Pupuk  2    160    80.0      32 1.55e-05 ***
Residuals              12     30     2.5                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Berdasarkan output diperoleh p-value=$1.55 \times 10^{-5} < \alpha = 0.05$, maka tidak cukup bukti untuk menerima H_0, artinya minimal ada 1 pasang kelompok pupuk yang rata-rata tinggi tanamannya terhadap jenis pupuk berbeda. Karena terdapat perbedaan itu (H_0 ditolak), maka perlu dilakukan uji lanjut, jika H_0 diterima maka tidak perlu dilakukan uji lanjut.

Uji Anova (Satu Arah)

5. Lakukan uji asumsi normalitas terhadap residual dengan Metode Shapiro Wilk.

1) Hipotesis:

H_0: Data berdistribusi normal

H_1:Data tidak berdistribusi normal

2) Kriteria pengambilan Keputusan: Tolak H_0 jika p-value<α.

3) Uji normalitas residual dengan Shapiro-Wilk:

```
> #variabel residual
> residual<-resid(model)
> #uji normalitas variabel residual dgn shapiro wilk
> shapiro.test(residual)

        Shapiro-Wilk normality test

data:  residual
W = 0.90219, p-value = 0.1028
```

Berdasarkan output diperoleh p-value=$0.1028 > \alpha = 0.05$, maka terima H_0 artinya data berdistribusi normal.

Uji Anova (Satu Arah)

6. Lakukan uji asumsi kehomogenan ragam (varians) (uji heteroskedastisitas) dengan Uji Breusch-Pagan).

   1) *Hipotesis:*

      H_0: ketiga populasi memiliki varians yang sama

      H_1:ketiga populasi memiliki varians yang berbeda

   2) Kriteria pengambilan Keputusan: Tolak H_0 jika p-value<α.

   3) Uji heteroskedastisitas Breusch-Pagan dengan package lmtest :

```
> #Uji keragaman varian (uji heteroskedatisitas) dgn Uji Breusch-Pagan
> install.packages("lmtest")
> library(lmtest)
> bptest(model,studentize = F,data=data_anova)

        Breusch-Pagan test

data:  model
BP = 3.0199e-31, df = 2, p-value = 1
```

Berdasarkan output diperoleh p-value=$1 > \alpha = 0.05$, maka terima H_0 artinya ketiga populasi memiliki varians yang sama.

Uji Anova (Satu Arah)

7. Tentukan rata-rata hitung dan standar deviasi tinggi tanaman untuk setiap kelompok jenis pupuk dengan package RcmdrMisc dan fungsi numSummary().

```
> install.packages("RcmdrMisc")
> library(RcmdrMisc)
> numSummary(data_anova$Tinggi_Tanaman, groups=data_anova$Jenis_Pupuk,statistics = c("mean","sd"))
  mean        sd data:n
A   17 1.581139      5
B   13 1.581139      5
C   21 1.581139      5
```

8. Lakukan uji lanjut ANOVA dengan metode TukeyHSD.

```
> #identifikasi kelompok mana yang memiliki perbedaan signfikan dengan metode TukeyHSD
> TukeyHSD(model)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = data_anova$Tinggi_Tanaman ~ data_anova$Jenis_Pupuk)

$`data_anova$Jenis_Pupuk`
    diff       lwr       upr     p adj
B-A   -4 -6.667864 -1.332136 0.0046341
C-A    4  1.332136  6.667864 0.0046341
C-B    8  5.332136 10.667864 0.0000104
```

Berdasarkan output, semua p-value$<\alpha = 0.05$, artinya rata-rata tinggi tanaman untuk setiap jenis pupuk tersebut berbeda secara nyata dengan rata-rata terbesar adalah tanaman dengan jenis pupuk A

Regresi

- Secara kuantitatif hubungan antara suatu variabel bebas dan variabel tak bebas dapat dimodelkan dalam suatu persamaan matematis yaitu persamaan regresi.
- Tujuan utama dilakukan analisis regresi adalah:
  1. ingin melihat atau menguji besarnya pengaruh variabel bebas terhadap variabel tak bebas (pendugaan/estimasi parameter)
  2. Ingin melakukan prediksi terhadap data baru
  3. Kombinasi keduanya yaitu menduga parameter dan melakukan prediksi

Analisis Regresi Linier Sederhana

- Regresi linier sederhana adalah metode statistika untuk menggambarkan hubungan kausalitas antara satu variabel bebas (X) dan satu variabel tak bebas (Y), hubungan keduanya dapat digambarkan sebagai suatu garis lurus (garis regresi).
- Bentuk umum model regresi linier sederhana:

$$\boldsymbol{Y = a + bX + \varepsilon}$$

dengan:
Y = variabel dependen
a = konstanta (intercept)
b = koefisien regresi (slope)
X = variabel independen
ε = error

Analisis Regresi Linear Sederhana dengan R

Akan digunakan fungsi lm() untuk melakukan regresi dengan R.
Sintaks dasar:

```
lm(formula,data,...)
```

- Formula adalah simbol yang menggambarkan hubungan antara variabel bebas dan tak bebas
- Data adalah kumpulan variabel yang akan digunakan pada bagian formula.

Analisis Regresi Linier Sederhana

Latihan 6.
Suatu perusahaan dalam beberapa bulan mempromosikan sejumlah peralatan elektronik dengan membuka outlet-outlet di berbagai daerah. Perusahaan tersebut ingin melihat apakah ada pengaruh biaya promosi (juta rupiah) dengan hasil penjualan (juta rupiah).

No.	Daerah	Penjualan	Promosi
1	Jakarta	205	26
2	Tangerang	206	28
3	Bekasi	254	35
4	Bogor	246	31
5	Bandung	201	21
6	Semarang	291	49
7	Solo	235	30
8	Yogyakarta	209	24
9	Purwokerto	204	31
10	Madiun	216	32
11	Surabaya	245	47
12	Malang	286	54
13	Kudus	312	40
14	Pekalongan	265	42
15	Indamayu	322	45

Regresi Linier Sederhana

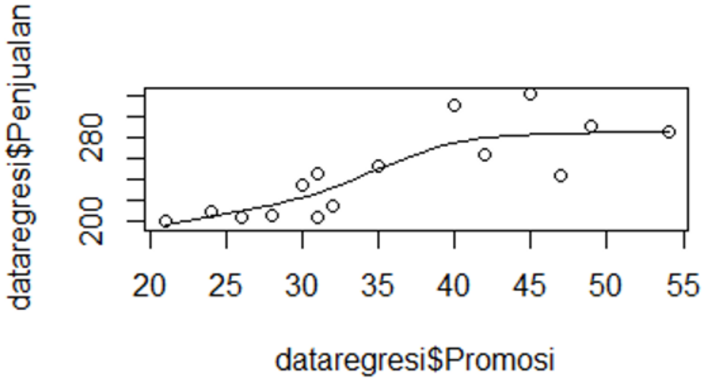
Penyelesaian.

1. Meng-import data

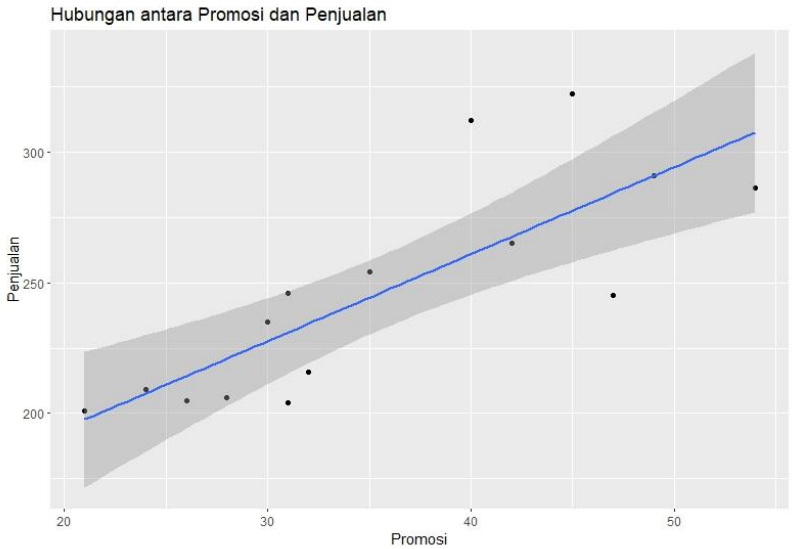
   Import data promosi dan penjualan yang telah diketik di Ms. Excel dan disimpan dengan nama "dataregresi".

2. Mengecek pola hubungan antara variabel Promosi dan variabel Penjualan dengan scatterplot

```
> #plot hubungan antara variabel Promosi dan Penjualan
> scatter.smooth(dataregresi$Promosi,dataregresi$Penjualan)
```




```
> install.packages("ggplot2")
> library(ggplot2)
> ggplot(dataregresi, aes(x = Promosi, y = Penjualan)) +
+   geom_point() +
+   geom_smooth(method = lm) + #Menambahkan garis regresi linear
+   labs(title = "Hubungan antara Promosi dan Penjualan",
+        x = "Promosi",
+        y = "Penjualan") +
+   theme_grey() #mengatur tema plot
```



Dari bentuk kedua scatterplot di atas dapat diketahui bahwa ada hubungan linar antara kegiatan Promosi dan Penjualan. Hubungan yang terjadi adalah hubungan positif yang **sangat kuat** dengan **koefisien korelasi 0.8041674** yang berarti bahwa pertambahan hasil penjualan seiring dengan pertambahan kegiatan promosi.

Analisis Regresi Linier Sederhana

4. Melakukan uji regresi linier sederhana menggunakan fungsi lm() dan fungsi summary() untuk melihat hasil uji regresi.

```
> #uji regresi linier sederhana
> model<-lm(Penjualan~Promosi,data=dataregresi)
> summary(model)

Call:
lm(formula = Penjualan ~ Promosi, data = dataregresi)

Residuals:
    Min      1Q  Median      3Q     Max 
-39.170 -16.615   0.177   8.568  51.117 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  127.813     25.189   5.074 0.000213 ***
Promosi        3.327      0.682   4.878 0.000302 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 25.35 on 13 degrees of freedom
Multiple R-squared:  0.6467,	Adjusted R-squared:  0.6195 
F-statistic: 23.79 on 1 and 13 DF,  p-value: 0.0003017
```

Analisis Regresi Linier Sederhana

Dari output uji regresi tersebut dapat dilakukan beberapa interpretasi sebagai berikut:

1. Persamaan regresi

   persamaan regresi yang menunjukkan hubungan linier antara variabel Promosi (X) dan Penjualan (Y) adalah:

$$\boldsymbol{Y = 127.813 + 3.327X}$$

Persamaan ini menunjukkan adanya pengaruh positif variabel biaya promosi terhadap hasil penjualan. Nilai penduga $\boldsymbol{b}$ sebesar 3.327 menunjukkan bahwa setiap kenaikan 1 satuan biaya promosi akan meningkatkan rata-rata hasil penjualan sebesar 3.327 juta rupiah, sedangkan nilai penduga $\boldsymbol{a}$ sebesar 127.813 menunjukkan bahwa rata-rata hasil penjualan jika kegiatan promosi tidak dilakukan.

Analisis Regresi Linier Sederhana

2. **Uji Nyata Regresi (Uji-F)/ *Goodness of Fit Test***

Uji ini digunakan untuk menguji apakah model fit (tepat) atau tidak untuk memprediksi variable Y.

Hipotesis:

$H_0: b = 0$ *(model tidak fit (tepat)untuk memprediksi variable Y)*
$H_1: b \neq 0$ *(model fit (tepat) untuk memprediksi variable Y)*

Pengambilan keputusan: Ho ditolak jika p-value<α.
Berdasarkan output, diperoleh p-value=0.000302< α=0.05,sehingga Ho ditolak. Ini berarti model regresi yang diperoleh tepat digunakan untuk memprediksi variable Y.

Analisis Regresi Linier Sederhana

2. **Uji Parsial (Uji-t)**

   Uji ini digunakan untuk menentukan apakah terdapat pengaruh nyata (signifikan) variable X terhadap variable Y.

   Hipotesis:

   $H_0 : b = 0$ *(tidak ada pengaruh nyata biaya promosi terhadap hasil penjualan)*
   $H_1 : b \neq 0$ *(ada pengaruh nyata biaya promosi terhadap hasil penjualan)*

   Pengambilan keputusan: Ho ditolak jika p-value< α.
   Berdasarkan output, diperoleh p-value=0.000213< α=0.05,sehingga Ho ditolak. Ini berarti ada pengaruh nyata biaya promosi terhadap hasil penjualan.

Analisis Regresi Linier Sederhana

3. Koefisien Determinasi

Pada pengujian sebelumnya diperoleh Kesimpulan awal yaitu model regresi yang diperoleh tepat untuk memprediksi hasil penjualan berdasarkan biaya promosi yang digunakan, dan terbukti ada pengaruh nyata biaya promosi terhadap hasil penjualan. Selanjutnya akan ditentukan berapa besar pengaruh biaya promosi tersebut terhadap hasil penjualan berdasarkan nilai koefisien determinasi (R^2).

Koefisien determinasi pada regresi linier sering diartikan sebagai seberapa besar kemampuan semua variabel bebas dalam menjelaskan variasi dari variabel terikatnya. Sederhananya, seberapa besar pengaruh X terhadap Y.

Berdasarkan output diperoleh koefisien determinasi (R-squared) sebesar 0,6467=64.67% yang menunjukkan biaya Promosi mempengaruhi variabel Penjualan sebesar 64.67%, ini berarti terdapat sebesar 35.33% faktor-faktor lain yang tidak diamati, mempengaruhi hasil Penjualan.

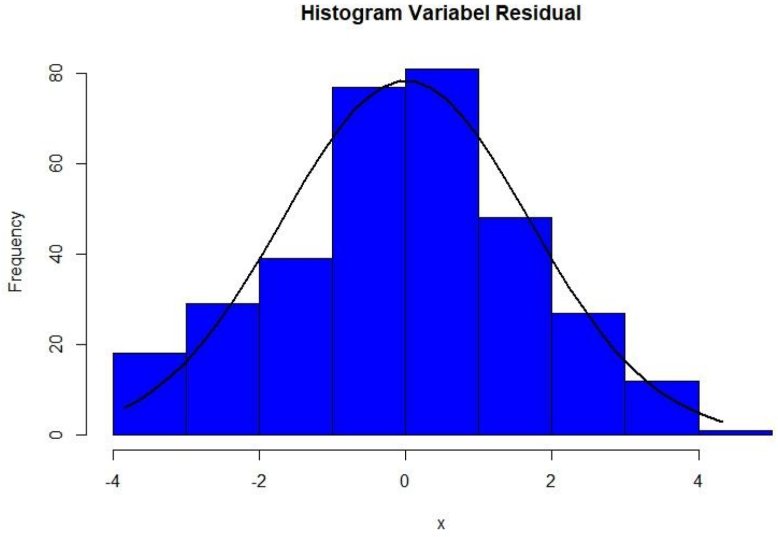
Uji Normalitas

Dalam analisis regresi linear, uji normalitas dilakukan terhadap variable residual. Oleh sebab itu, setelah model didefinisikan kemudian dicari terlebih dahulu variable residualnya dengan R sbb:

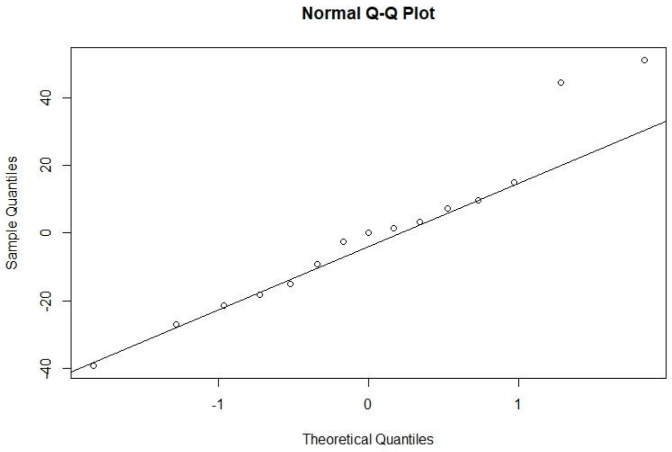
```
> modelreg=lm(Penjualan~Promosi,dataregresi)
> residual=resid(modelreg)
```

selanjutnya, gambarkan histogram dan kurva normal untuk melihat distribusi variable residual secara visual.

```
> #histogram dan kurva normal
> hist(residual1)
> x<-residual1
> h<-hist(x,breaks=10,col="blue",main="Histogram Variabel Residual")
> xfit<-seq(min(x),max(x),length=40)
> yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
> yfit<-yfit*diff(h$mids[1:2])*length(x)
> lines(xfit,yfit,col="black",lwd=2)
```




```
> qqnorm(residual)
> qqline(residual)
```



Uji Normalitas

1. Uji Normalitas

Uji normalitas dengan beberapa jenis uji dilakukan sebagai berikut. Jenis uji yang dilakukan harus memenuhi syarat2 yang ditentukan (pilih yang paling sesuai, jangan lupa baca referensi ya). Untuk sampel kecil (n<30) bisa menggunakan Uji Shapiro-Wilk dan untuk sampel besar bisa menggunakan Uji Kolmogorov-Smirnov.

Hipotesis:

H_0 : data berdistribusi normal
H_1 : data tidak berdistribusi normal

a. Uji Shapiro-Wilk

```
> #Uji Shapiro-Wilk
> shapiro.test(residual1)

        Shapiro-Wilk normality test

data:  residual1
W = 0.98911, p-value = 0.01396
```

diperoleh p-value<2.2e-16. Untuk α =0.05, p-value< α, oleh sebab itu H_0 ditolak yang berarti menerima H_1. Dengan demikian, data tidak berdistribusi normal.

Uji Normalitas

b. Uji Kolmogorov-Smirnov

```
> #Uji normalitas data
> #uji kolmogorov-smirnov
> ks.test(model2$residuals,ecdf(model$residuals))

        Asymptotic one-sample Kolmogorov-Smirnov test

data:  model2$residuals
D = 0.28661, p-value < 2.2e-16
alternative hypothesis: two-sided
```

diperoleh p-value<2.2e-16. Untuk α =0.05, p-value< α, oleh sebab itu H_0 ditolak yang berarti menerima H_1. Dengan demikian, data tidak berdistribusi normal.