

Analisa Regresi Dua Variabel: Konsep Dasar

Review Statistik: Uji Hipotesa Satu Sampel

Tjipto Juwono, Ph.D.

May 2018



SURYA
UNIVERSITY

Mean Value Vs Fixed Values

Y ↓ \ X →	X →									
	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y, \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	-	88	-	113	125	140	-	160	189	185
	-	-	-	115	-	-	-	162	-	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y, $E(Y X)$	65	77	89	101	113	125	137	149	161	173

Gambar 1: Populasi terdiri dari 60 keluarga terbagi atas 10 income groups

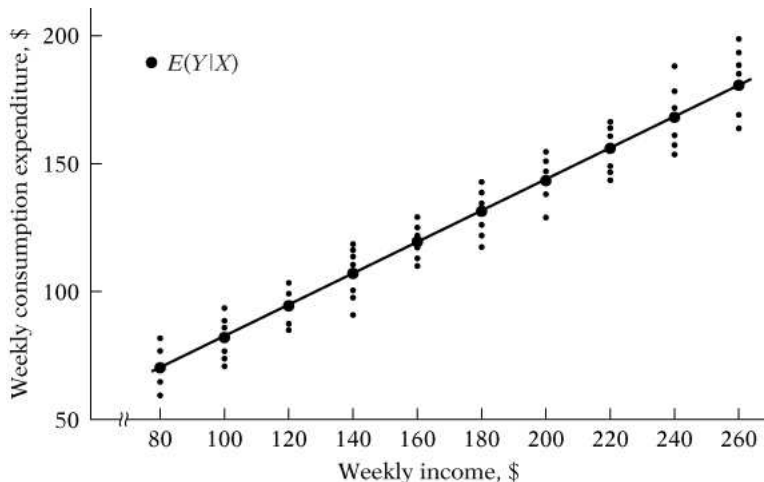
Mean Values Vs Fixed Values

- Perhatikan tabel pada Gbr. (1). Untuk setiap income group, ada variasi pengeluaran per minggu, tetapi *mean*-nya mengalami kenaikan dengan kenaikan income.
- Mean untuk pengeluaran pada Gbr. (1) disebut *conditional mean* dengan symbol:

$$E(Y|X) \quad (1)$$

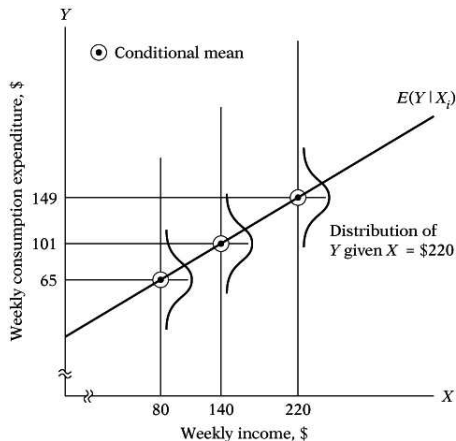
Arti simbol ini: Mean untuk variabel Y jika diberikan satu X tertentu.

Mean Values Vs Fixed Values



Gambar 2: Grafik $E(Y|X)$ vs X . Populasi terdiri dari 60 keluarga terbagi atas 10 income groups

Mean Values Vs Fixed Values



Gambar 3: Grafik $E(Y|X)$ vs X . Populasi terdiri dari 60 keluarga terbagi atas 10 income groups

Population Regression Line

- Pada Gbr. (2), ditampilkan $E(Y|X)$ vs X . Jika kita menghubungkan semua titik $E(Y|X)$, kita memperoleh *Population Regression Line (PRL)*, atau secara umum: *Population Regression Curve*.
- Kata "population" digunakan sebab data-datanya mencakup seluruh populasi.
- Jika data-datanya mencakup sampel saja, maka kita akan memperoleh *Sample Regression Line (SRL)*

Population Regression Function

- Dari Gbr. (2) dan Gbr. (3) jelaslah bahwa setiap $E(Y|X_i)$ adalah suatu fungsi dari X_i . Dapat kita tuliskan:

$$E(Y|X_i) = f(X_i) \quad (2)$$

- Pers. (2) disebut *Population Regression Function (PRF)*
- Apa bentuk fungsional dari PRF? Ini adalah persoalan empiris.
- Misalnya, kita dapat mengasumsikan bahwa $E(Y|X_i)$ adalah **fungsi linear** dari X_i , yang dapat kita tulis:

$$E(Y|X_i) = \beta_1 + \beta_2 X_i \quad (3)$$

Arti Kata Linear

Linear dalam variabel: Jika kurva-nya berbentuk linear. Contoh yang tidak linear dalam variabel:

$$E(Y|X) = \beta_1 + \beta_2 X_i^2 \quad (4)$$

Linear dalam parameter: Suatu fungsi bisa jadi linear dalam parameter tetapi tidak linear dalam variabel. Contohnya adalah Pers. (4). Contoh yang tidak linear dalam parameter:

$$E(Y|X) = \beta_1 + \beta_2^2 X_i \quad (5)$$

Stochastic Disturbance

- Kita sudah mengetahui bahwa untuk setiap X_i ada sejumlah Y
- Rata-rata dari sejumlah Y untuk suatu X_i tertentu adalah $E(Y|X_i)$
- Berarti untuk setiap X_i ada sejumlah Y_i yang masing-masingnya mempunyai selisih:

$$u_i = Y_i - E(Y|X_i) \quad (6)$$

- Sehingga dapat kita tuliskan:

$$Y_i = E(Y|X_i) + u_i \quad (7)$$

- Suku u_i kita sebut suku **Stochastic Disturbance** yang merupakan suku **Random**.
- Suku $E(Y|X_i)$ kita sebut suku **Systematic** atau suku **Deterministic**.

Asumsi: Linear Model

- Jika kita asumsikan model linear pada Pers. (7) maka:

$$\begin{aligned} Y_i &= E(Y|X_i) + u_i \\ &= \beta_1 + \beta_2 X_i + u_i \end{aligned} \quad (8)$$

- Dapat dibuktikan bahwa:

$$E(u_i|X_i) = 0 \quad (9)$$

- Artinya? $E(Y|X_i) = \beta_1 + \beta_2 X_i$ adalah equivalent dengan $Y_i = \beta_1 + \beta_2 X_i + u_i$ jika $E(u_i|X_i) = 0$

Apa Arti u_i ?

Makna Suku Stokastik u_i

Suku stokastik u_i pada Pers. (8) mengandung arti bahwa sebenarnya ada faktor-faktor lain, selain income, yang mempengaruhi konsumsi. Besar konsumsi tidak dapat sepenuhnya dijelaskan dengan variabel-variabel yang ada pada model regresi.

Apa arti Y_i ?

Y ↓ \ X →	X →									
	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y , \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	-	88	-	113	125	140	-	160	189	185
	-	-	-	115	-	-	-	162	-	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y , $E(Y X)$	65	77	89	101	113	125	137	149	161	173

Apa arti Y_i ?

Misalkan kita ambil $X = \$80$. Maka ada sejumlah Y , yaitu: $Y = 55, 60, 65, 70$, dan 75 . Dengan conditional mean $E(Y|80) = \$65$. Dapat kita tuliskan:

$$\begin{aligned} Y_1 &= 55 = \beta_1 + \beta_2(80) + u_1 \\ Y_2 &= 60 = \beta_1 + \beta_2(80) + u_2 \\ Y_3 &= 65 = \beta_1 + \beta_2(80) + u_3 \\ Y_4 &= 70 = \beta_1 + \beta_2(80) + u_4 \\ Y_5 &= 75 = \beta_1 + \beta_2(80) + u_5 \end{aligned} \tag{10}$$

Dengan:

$$\beta_1 + \beta_2(80) = E(Y|80) = 65 \tag{11}$$

Sample Regression Function

- Tabel pada Gbr. (1) menampilkan **populasi**, dan bukan **sampel**.
- Seringkali data yang kita miliki adalah data sampel, dan bukan data populasi.
- Jika demikian, kita tidak dapat memperoleh *Population Regression Function, PRF*, melainkan hanya dapat memperoleh *Sample Regression Function, SRF*.
- SRF merupakan pendekatan terhadap PRF.
- Seberapa akurat SRF dalam mendekati PRF? Kita tidak tahu secara pasti, tetapi itulah manfaatnya statistik! Yaitu untuk memperkirakan Confidence Interval dari perkiraan kita (akan dibahas pada kuliah pada kesempatan lain).

Populasi → Sampel

Y ↓ \ X →	X →									
	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y , \$	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	-	88	-	113	125	140	-	160	189	185
	-	-	-	115	-	-	-	162	-	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y , $E(Y X)$	65	77	89	101	113	125	137	149	161	173

Contoh Sampel

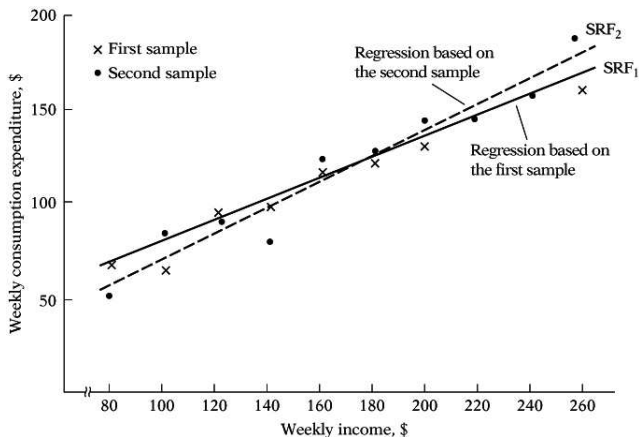
Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

Gambar 4: Sampel-1, dari data pada tabel Gbr (1)

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260

Gambar 5: Sampel-2, dari data pada tabel Gbr (1)

Sample Regression Function



Gambar 6: SRF, diperoleh dari dua sample pada Gbr. (4,5)

Bagaimana memperoleh SRF?

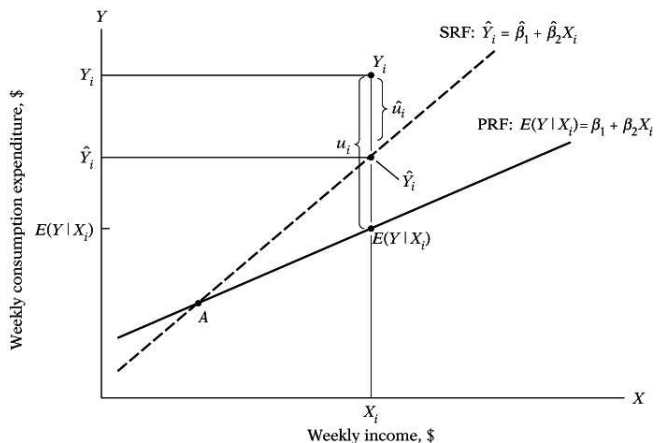
Dari kuliah pertama:

$$\begin{aligned}\hat{Y} &= \beta_1 + \beta_2 X \\ \beta_2 &= r \frac{s_y}{s_x} \\ \beta_1 &= \bar{Y} - \beta_2 \bar{X}\end{aligned}\tag{12}$$

$$s_x = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}}\tag{13}$$

$$s_y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{n - 1}}\tag{14}$$

PRF vs SRF



Gambar 7: Perbandingan antara PRF dan SRF

Review Statistik: Uji Hipotesa Satu Sampel

Uji Hipotesa

Contoh

Sebuah perusahaan mebel menghasilkan meja tulis, dengan rata-rata produksi 200 meja/minggu, dengan standard deviasi 16 meja/minggu. Kemudian perusahaan menerapkan cara baru dalam membuat meja. Setelah cara baru itu ditetapkan, produksi rata-rata per minggu menjadi 203.5 meja/minggu.

Apa Tujuan Uji Hipotesis?

- Misalkan kita mempunyai sebuah nilai, yang kita sebut saja "*nilai acuan*"
- Kemudian kita mempunyai sebuah nilai lain yang ingin kita uji (sebut saja "*nilai uji*").
- Apakah nilai uji itu berbeda dengan nilai acuan? atau,
- Apakah nilai uji itu lebih kecil dari nilai acuan? atau,
- Apakah nilai uji itu lebih besar dari nilai acuan?

Yang dimaksud dengan "berbeda, lebih kecil, atau lebih besar"
"berbeda, lebih kecil, atau lebih besar secara signifikan"

Uji Hipotesa

Contoh

Sebuah perusahaan mebel menghasilkan meja tulis, dengan rata-rata produksi 200 meja/minggu, dengan standard deviasi 16 meja/minggu. Kemudian perusahaan menerapkan cara baru dalam membuat meja. Setelah cara baru itu ditetapkan, produksi rata-rata per minggu menjadi 203.5 meja/minggu.

Pertanyaan: Apakah 203.5 meja/minggu lebih besar daripada 200 meja/minggu? Dengan kata lain: Dapatkah kita mengatakan bahwa terjadi peningkatan yang signifikan setelah diterapkannya metode baru itu?

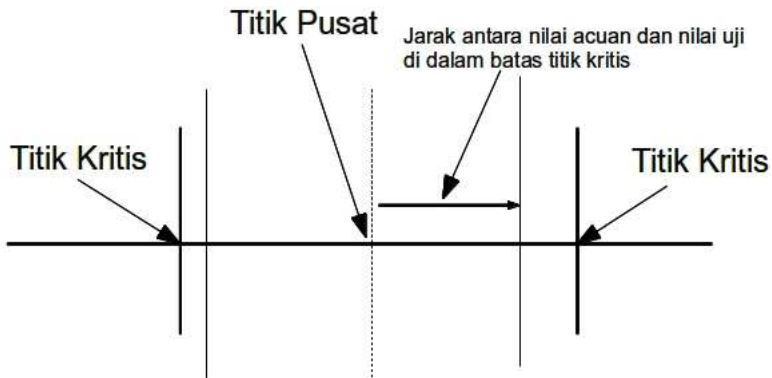
Uji Hipotesa Two-Tailed

Apa bila pertanyaannya adalah: "Apakah nilai uji berbeda secara signifikan dengan nilai acuan?". Maka berarti ada dua kemungkinan: (1) Nilai uji lebih kecil daripada nilai acuan, atau (2) nilai uji lebih besar daripada nilai acuan. Karena adanya dua kemungkinan ini, maka uji hipotesa semacam ini disebut uji hipotesa **two tailed**.

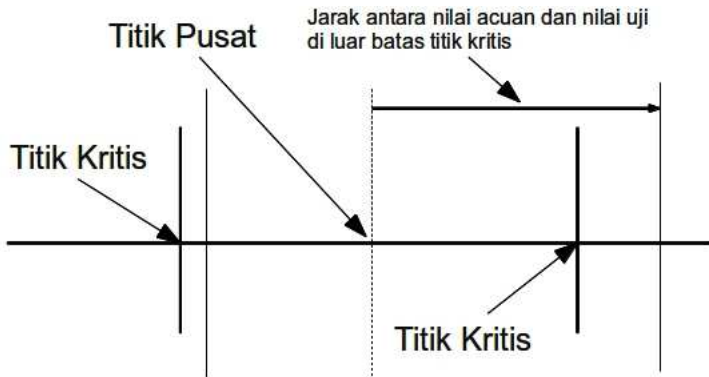
Uji Hipotesa Two-Tailed

- Untuk melakukan uji hipotesa two tailed, pertama-tama kita harus menentukan jarak antara nilai acuan dan nilai uji.
- Kemudian kita menentukan dua titik kritis (satu negatif dan lainnya positif).
- Jika jarak antara nilai acuan dan nilai uji kurang dari titik kritis, maka itu artinya tidak ada perbedaan signifikan antara nilai acuan dan nilai uji. Jika jarak tersebut melampaui titik kritis, berarti ada perbedaan yang signifikan.
- Jarak antara nilai acuan dan nilai uji dapat negatif atau positif, sebab nilai ujian dapat lebih kecil atau lebih besar daripada nilai acuan. Itulah sebabnya kita mempunyai dua titik kritis, satu negatif dan lainnya positif.

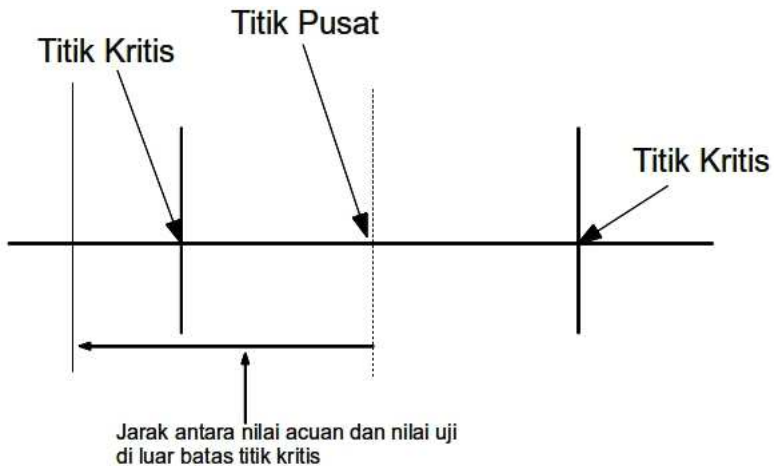
Two-Tailed



Two-Tailed



Two-Tailed



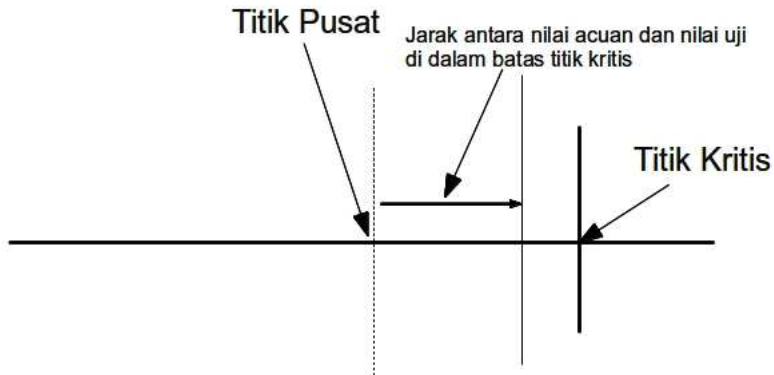
Uji Hipotesa One-Tailed

Apa bila pertanyaannya adalah: "Apakah nilai uji lebih kecil (atau lebih besar) secara signifikan dengan nilai acuan?". Maka berarti ada satu kemungkinan: Nilai uji lebih kecil (lebih besar) daripada nilai acuan. Karena adanya satu kemungkinan ini, maka uji hipotesa semacam ini disebut uji hipotesa **one-tailed**.

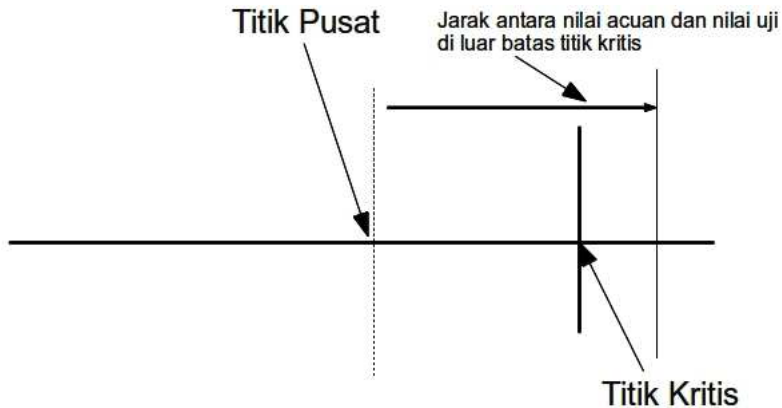
Uji Hipotesa One-Tailed

- Untuk melakukan uji hipotesa one tailed, pertama-tama kita harus menentukan jarak antara nilai acuan dan nilai uji.
- Kemudian kita menentukan satu titik kritis.
- Jika jarak antara nilai acuan dan nilai uji kurang dari titik kritis, maka itu artinya tidak ada perbedaan signifikan antara nilai acuan dan nilai uji. Jika jarak tersebut melampaui titik kritis, berarti ada perbedaan yang signifikan.

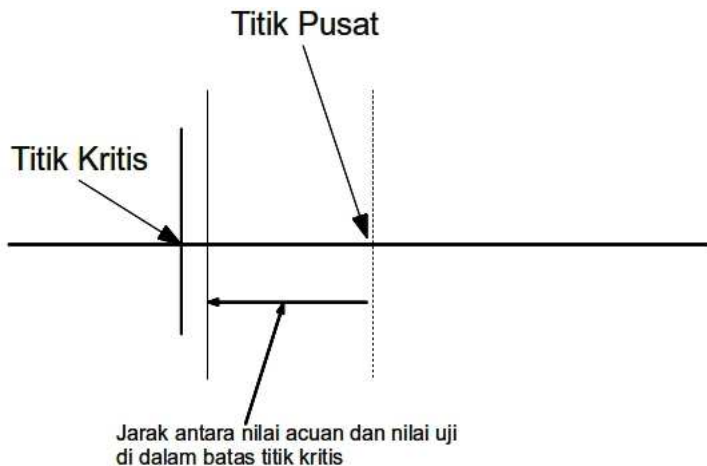
One-tailed



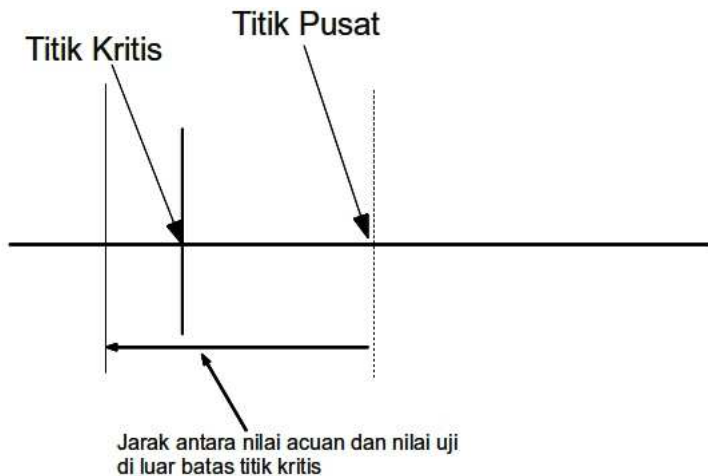
One-tailed



One-tailed



One-tailed



Menentukan Jarak Antara Nilai Uji dan Nilai Acuan

Jarak antara nilai uji dan nilai acuan ditentukan berdasarkan statistik yang kita pilih, misalnya apakah menggunakan distribusi- Z atau distribusi- t .

Distribusi- Z :

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad (15)$$

Distribusi- t :

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad (16)$$

Menentukan Titik Kritis

Titik kritis ditentukan dengan memilih level of significance. Level of significance sama dengan $(1 - \textit{level of confidence})$. Jika level of confidence 95%, maka level of significance adalah $1 - 0.95 = 0.05$. Level of confidence biasanya ditulis dengan simbol α .

Prosedur Uji Hipotesa

Langkah 1. Penentuan Hipotesa Pada prinsipnya kita membuat dua hipotesa.

- 1 Hipotesa pertama (disebut **Null Hypothesis**) menyatakan bahwa **tidak ada perubahan yang signifikan**. Null Hypothesis biasanya ditulis dengan simbol H_0
- 2 Hipotesa kedua (disebut **Alternate Hypothesis**) menyatakan bahwa **ada perubahan yang signifikan**. Alternate Hypothesis biasanya ditulis dengan simbol H_1

Prosedur Uji Hipotesa

Tujuan dari prosedur uji hipotesa adalah untuk menentukan apakah (1) kita tidak menolak Null Hypothesis, atau (2) kita menolak Null Hypothesis.

Catatan: Pada nomor (1) di atas, kita tidak mengatakan "menerima Null Hypothesis", melainkan "tidak menolak Null Hypothesis"

Prosedur Uji Hipotesa

Simbol Yang Digunakan Untuk Hipotesa

1 $H_0 \rightarrow =, \leq, \geq$

2 $H_1 \rightarrow \neq, >, <$

Prosedur Uji Hipotesa

Contoh:

Two-tailed

$$\begin{aligned}H_0 : \rho &= 0 \\H_1 : \rho &\neq 0\end{aligned}\tag{17}$$

One-tailed

$$\begin{aligned}H_0 : \mu &\leq 20 \\H_1 : \mu &> 20\end{aligned}\tag{18}$$

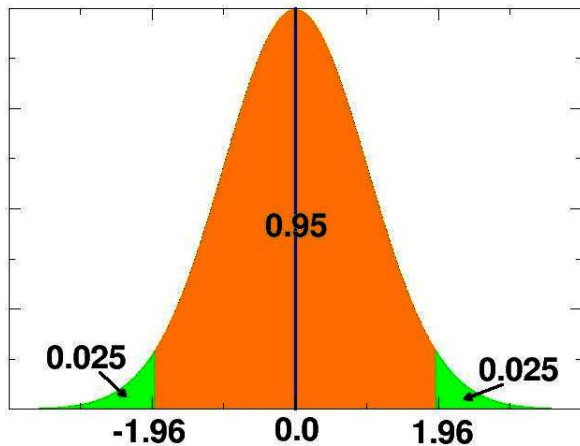
One-tailed

$$\begin{aligned}H_0 : \mu &\geq 50 \\H_1 : \mu &< 50\end{aligned}\tag{19}$$

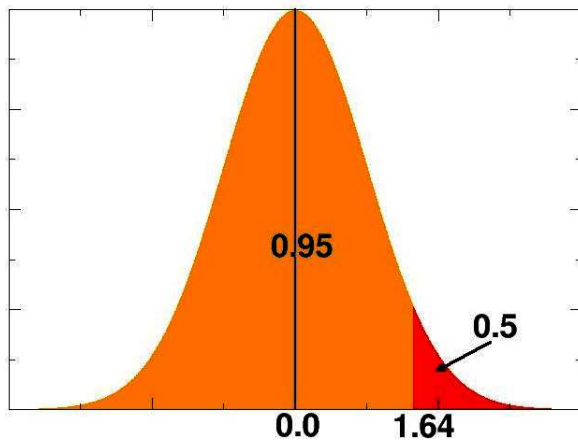
Prosedur Uji Hipotesa

Langkah 2: Menentukan level of significance. Pada slide sebelumnya, kita sudah membahas bahwa kita perlu menentukan di mana lokasi titik kritis. Lokasi titik kritis ini ditentukan berdasarkan level of significance.

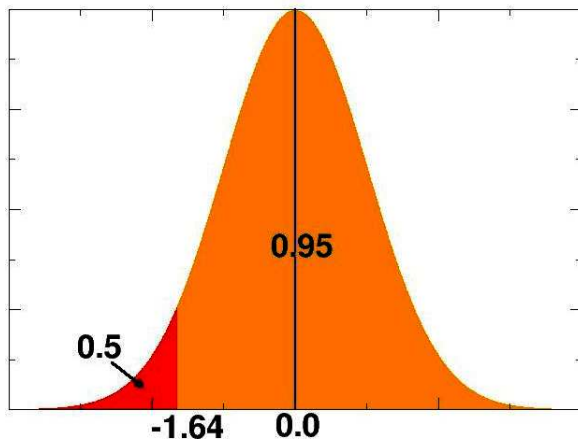
Two Tailed $\alpha = 0.05$



One Tailed $\alpha = 0.05$



One Tailed $\alpha = 0.05$



Prosedur Uji Hipotesa

Langkah 3: Menentukan Statistik. Tergantung dari problemnya, kita dapat menggunakan Z, t, χ^2 , dll

Langkah 4: Menentukan aturan pengambilan keputusan

Aturan ini diperoleh setelah kita menghitung statistiknya (misalkan nilai Z), dan lalu menghubungkannya dengan hipotesa yang telah kita tulis.

Langkah 5: Mengambil keputusan dan menafsirkan hasilnya

Uji Hipotesa: Contoh

Contoh

Sebuah perusahaan mebel menghasilkan meja tulis, dengan rata-rata produksi 200 meja/minggu, dengan standard deviasi 16 meja/minggu. Kemudian perusahaan menerapkan cara baru dalam membuat meja. Setelah cara baru itu ditetapkan, produksi rata-rata per minggu menjadi 203.5 meja/minggu. Pertanyaan: Apakah ada perubahan yang signifikan dalam jumlah produksi per minggu? (level of significance 0.01) Anggap bahwa mereka menerapkan metode baru ini selama 50 minggu.

Uji Hipotesa: Contoh

Langkah 1: Penentuan Hipotesa. Pertanyaan pada soal tersebut adalah "apakah ada perubahan?". Artinya, tidak dipersoalkan apakah lebih besar atau lebih kecil. Dengan demikian, kita mempunyai hipotesa two-tailed. Ingat bahwa yang dipertanyakan oleh soal biasanya diletakkan pada H_1 . Jadi H_0 adalah hipotesa di mana tidak ada perubahan.

$$\begin{aligned}H_0 : \mu &= 200 \\H_1 : \mu &\neq 200\end{aligned}\tag{20}$$

Langkah 2: Tentukan level of significance

$$\alpha = 0.01\tag{21}$$

Uji Hipotesa: Contoh

Langkah 3: Tentukan Statistik Gunakan distribusi- Z , karena σ diketahui.

Uji Hipotesa: Contoh

Langkah 4: Menentukan aturan pengambilan keputusan
Pertama-tama hitunglah Z :

$$\begin{aligned} Z &= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \\ &= \frac{203.5 - 200}{16/\sqrt{50}} \end{aligned} \quad (22)$$

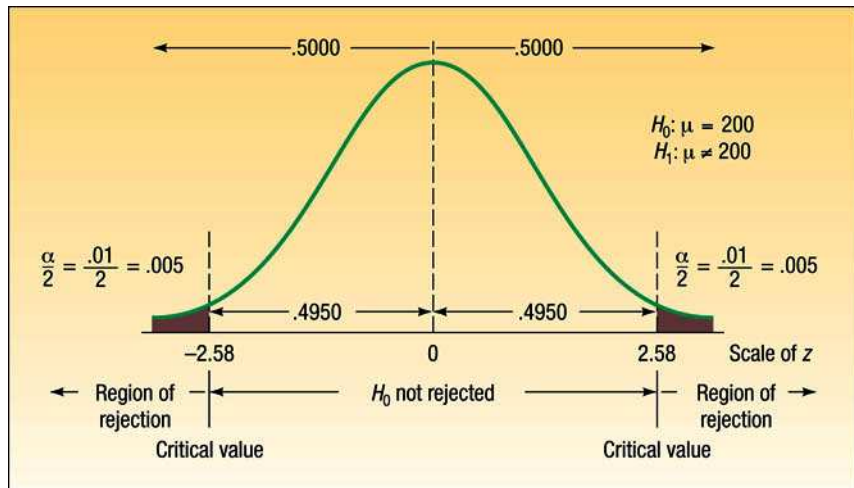
$$= 1.55 \quad (23)$$

Berdasarkan level of significance, diperoleh:

$$Z_{\alpha/2} = 2.58 \quad (24)$$

Maka aturan pengambilan keputusan adalah:
 H_0 ditolak jika $|Z| > 2.58$.

Uji Hipotesa: Contoh



Uji Hipotesa: Contoh

Langkah 5: Mengambil keputusan dan menafsirkan hasilnya

Diperoleh hasil $Z = 1.55$, sedangkan H_0 ditolak jika $|z| > 2.58$. Karena 1.55 tidak berada pada daerah ditolak, maka keputusannya adalah H_0 tidak ditolak. Berarti, dari sample yang tersedia, tidak menunjukkan bahwa metode baru menghasilkan perubahan yang signifikan.