# BUKU AJAR

MATA KULIAH
STATISTIKA EKONOMI LANJUTAN



oleh

Drs. WIRATNO, M.EC
Drs. SUGIONO, MSIE

**DEPARTEMEN PENDIDIKAN NASIONAL**
**FAKULTAS EKONOMI**
**UNIVERSITAS DIPONEGORO**
**SEMARANG**
**2007**

# PENGANTAR

Buku Pegangan Kuliah Mahasiswa(BPKM) untuk mata kuliah Statistika Ekonomi lanjutan ini merupakan edisi revisi terhadap BPKM untuk mata kuliah Statistika II yang kami susun pada tahun 1990. Revisi ini kami lakukan guna lebih menyaesuaikan dengan kemajuan-kemajuan yang telah terjadi dalam statistika serta teknik pengohan data dengan penggunaan paket program komputer.

Ada satu pokok bahasan yang kami tambahkan dalam edisi revisi ini, yaitu tentang statistika non-parametrik. Walaupun dalam hal ini tidak banyak perubahan yang terjadi, namun pengetahuan tentang statistika non-parametrik tetap diperlukan oleh mahasiswa sebagai kelengkapan pemahaman terhadap statistika.

Dalam kesempatan ini, kami ingin menyampaikan terima kasih kepada ketua jurusan IESP FE Undip yang telah mendorong kami guna melakukan revisi terhadap BPKM ini. Akhirnya, kami tetap menrima kritik dan saran demi perbaikan BPKM ini. Semoga BPKM ini dapat bermanfaat. Amien.

Semarang, Desember 2007.

DAFTAR ISI

Pokok Bahasan V PENGUJIAN HIPOTESIS
5.1. Pendahuluan
5.2. TIU dan sasbel
5.3. Arti dan kegunaan
5.4. Langklah-langkah pengujian hipotesis
5.5. Soal
5.6. Rangkuman

Pokok Bahasan VI UJI – Z
6.1. Pendahuluan
6.2. TIU dan sasbel
6.3. Uji-Z untuk parameter rata-rata populasi
6.4. Uji-Z untuk selisih dua rata-rata populasi
6.5. Uji-Z untuk parameter proporsi populasi
6.6. Uji-Z untuk selisih dua proporsi populasi
6.7. Soal
6.8. Rangkuman

Pokok Bahasan VII UJI – t
7.1. Pendahuluan
7.2. TIU dan sasbel
7.3. Uji-t untuk parameter rata-rata populasi
7.4. Uji-t untuk parameter proporsi
7.5. Uji-t untuk selisih dua rata-rata populasi
7.6. Soal
7.7. Rangkuman

Pokok Bahasan VIII UJI – F
8.1. Pendahuluan
8.2. TIU dan sasbel
8.3. AAnava satu arah
8.4. Anva dua arah
8.5. Perbandingan berganda
8.6. Soal
8.7. Rangkuman

Pokok Bahasan IX UJI - $\chi$ KUADRAT
9.1. Pendahuluan
9.2. TIU dan sasbel
9.3. Pengujian kompatibilitas
9.4. Pengujian independensi
9.5. Pengujian homogenitas
9.6. Penyesuaian bilangan $n_{ij} < 5$
9.7. Pengujian terhadaap deviasi standar
9.8. Soal
9.9. Rangkuman

# POKOK BAHASAN I
# PROBABILITA

1.1. Pendahuluan

Untuk bisa lebih mudah mengikuti pokok bahasan tentang probabilita ini, mahasiswa harus sudah memahami tentang teori himpunan, permutasi dan kombinasi. Secara umum, probabilita berarti kemungkinan tentang terjadinya suatu peristiwa. Probabilita berupa proporsi, sehingga rentang nilainya antara 0 (nol) dan 1 (satu). Pokok bahasan ini terdiri dari 5 sub-pokok bahasan, yaitu :

a. Probabilita objektif dan subjektif;
b. Ruang cuplikan;
c. Peristiwa;
d. Perhitungan probabilita; dan
e. Probabilita bersyarat.

1.2. Tujuan Instruksional Umum dan Sasaran Belajar

**T I U :**

Mahasiswa diharapkan mampu memahami berbagai jenis probabilita, melakukan perhitungan probabilita , dan menerapkan konsep probabilita dalam bidang ekonomi.

**Sasaran Belajar** :

Setelah mempelajari bab ini, mahasiswa diharapkan mampu :

1. Menjelaskan konsep untung-untungan (chance) melalui paling tidak 3 contoh.
2. Menjelaskan pengertian probabilita melalui tiga cara: perumusan klasik, perumusan secara frekuensi relative, dan probabilita subjektif.
3. Melukiskan pohon hasil percobaan (*outcome trees*).
4. Menjelaskan arti ruang cuplikan dengan 3 contoh.
5. Mengidentifikasi 3 sifat elementer probabilita.
6. Mengidentifikasi peristiwa (event) sebagar bagian dari ruang cuplikan.
7. Memberi paling sedikit 3 contoh peristiwa tunggal dan kombinasi.
8. Menunjukkan besarnya probabilita suatu peristiwa.
9. Menggambar diagram Venn yang menunjukkan gabungan dan perpotongan dari peristiwa.
10. Menyebutkan rumus dasar dalam probabilita tentang gabungan dan perpotongan.
11. Menghitung probabilita dengan kaidah penjumlahan.
12. Menghitung probabilita dengan kaidah penggandaan.
13. Mengidentifikasi peristiwa yang bebas stokastik.
14. Mengerjakan dengan teliti persoalan probabilita dalam ekonomi.

1.3. Probabilita Objektif dan Subjektif.

Keadaan di dunia ini penuh dengan ketidakpastian. Dari keadaan yang serba tidak pasti orang cenderung ingin mengetahui seberapa pasti hal itu akan terjadi. Probabilita hanya merupakan suatu ukuran tentang tingkat kepastian dari ketidakpastian yang berkaitan dengan peristiwa (*event*). Hal itu dapat menunjukkan peristiwa mana yang lebih mungkin terjadi atau peristiwa mana yang lebih tidak mungkin terjadi, tetapi probabilita tidak dapat meramalkan yang akan datang secara pasti. Sebagai contoh, bila BMG (Badan Meteorologi dan Geofisika) menyatakan bahwa probabilita untuk hujan pada hari ini hanya sebesar 15 % dan ternyata bahwa hari ini hujan, biasanya ada yang menyalahkan BMG karena ramalannya atau perkiraannya terhadap cuaca tidak tepat atau tidak benar. Ini suatu persepsi yang salah terhadap makna probabilita, probabilita lebih menunjukkan suatu kecenderungan dari pada menentukan secara pasti terntang terjadinya peristiwa.

Teori probabilita sebetulnya memberikan cara pengukuran kuantitatif tentang peluang atau tingkat kepastian tentang terjadinya peristiwa (AD:67). Ada 3 cara untuk menentukan atau mengukur tingkat kepastian tersebut, yaitu: perumusan klasik, perumusan atas dasar frekuensi relatif dan perumusan atas dasar subjetivitas.

a. Perumusan Klasik

Pada kondisi-kondisi yang diketahui, jika terdapat sejumlah n kejadian yang mungkin timbul dan jika kejadian tersebut lengkap terbatas jumlahnya (exhaustive), ekslusif secara bersama (*mutually exclusive*) dan memiliki kesempatan yang sama untuk timbul, maka jika sejumlah m dari kejadian di atas merupakan peristiwa E, probabilita peristiwa E tersebut dapat dirumuskan sebagai

$$P(E) = m/n$$

atau

$$\text{Prob. suatu peristiwa} = \frac{\text{Jumlah outcome yang sesuai dg peristiwa tertentu}}{\text{Jumlah seluruh alternatif outcome}}$$

Contoh : Dalam pelemparan sebuah koin (mata uang logam) ratusan yang seimbang (*fair coin*), probabilita untuk keluar angka 100 adalah sebesar 0.5, yaitu satu dari dua sisi koin tersebut.

b. Perumusan atas dasar konsep frekuensi relatif atau konsep empiris

Jika n merupakan jumlah perwujudan kejadian yang khusus, katakanlah peristiwa E dalam serangkaian n percobaan dalam jumlah besar, maka probabilita peristiwa E merupakan frekuensi relatif m/n dan dinyatakan sebagai

$$P(E) = \lim_{n \to \infty} m/n$$

atau

$$\text{Prob.terjadinya peristiwa} = \frac{\text{Jumlah frekuensi peristiwa terjadi pada waktu lalu}}{\text{Jumlah seluruh observasi}}$$

Contoh : Bila kita melempar satu mata uang ratusan dalam jumlah yang tidak terbatas, maka probabilita keluarnya angka 100 akan mendekati 0.5

c. Perumusan atas dasar subjektivitas

Dalam hal ini probabilita dirumuskan sebagai pengukuran dari keyakinan pribadi (seseorang) terhadap suatu hipotesis yang tertentu atau terjadinya suatu peristiwa tertentu. Keadaan ini mungkin terjadi bila tidak ada pengalaman masa lalu sebagai dasar perhitungan probabilita, atau karena ada pertimbangan-pertimbangan yang lebih bersifat irrasional dari para pelakunya. Missal perhitungan probabilita oleh seseorang berdasar "wangsit"

Contoh : Jika ada dua peristiwa, A dan B, mempunyai kondisi yang sama untuk terjadi tetapi kita dua kali lebih yakin akan terjadinya peristiwa A daripada peristiwa B, maka probabilita peristiwa A atau P(A) sebesar 2/3 dan $P(B) = P(\overline{A})$ sebesar 1/3.

Probabilita yang dirumuskan atas dasar perumusan klasik dan konsep frekuensi relatif merupakan probabilita objektif, sedangkan probabilita yang dirumuskan atas dasar subjektivitas disebut probabilita subjektif. Sebagai suatu kesimpulan adalah adanya kesan bahwa probabilita selalu dirumuskan sebagai proporsi atau rasio. Sehingga alternative nilainya antara 0 dan 1 dan tidak pernah bernilai negative., dan apabila probabilita suatu peristiwa telah dapat ditentukan, maka dasar perhitungan probabilita dari berbagai peristiwa mudah dilakukan. Dalam pokok bahasan ini, kita akan memakai pengertian probabilita atas dasar konsep frekuensi relatif.

## 1.4. Ruang Cuplikan (*sample space*).

Suatu eksperimen adalah sesuatu proses yang direncanakan dan diikuti observasi dan / atau pengumpulan data. Hasil suatu eksperimen disebut outcome (kejadian) dari eksperimen tersebut. Set dari seluruh outcome suatu eksperimen disebut ruang cuplikan (sample space) eksperimen itu. Huruf S biasanya digunakan untuk melambangkan ruang cuplikan, yang merupakan set dari seluruh outcome yang mungkin terjadi.

Contoh : Ada suatu keluarga merencanakan untuk mempunyai tiga anak, anak laki-laki dilambangkan dengan P dan anak wanita dilambangkan dengan W. Ruang cuplikan (*sample space*) yang menunjukkan semua kemungkinan / alternatif dari ketiga anak yang akan dipunyai adalah sebagai berikut :

S = {(PPP), (PPW), (PWP), (PWW), (WPP), (WPW), (WWP), (WWW)}

## 1.5. Peristiwa (*event*)

Satu atau beberapa *outcome* dapat membentuk suatu peristiwa (*event*). Pada umumnya peristiwa diartikan sebagai terjadinya sesuatu. Tetapi dalam pembahasan tentang probabilita ini, peristiwa mempunyai arti yang lebih lua, karena peristiwa mencakup tentang terjadinya sesuatu ataupun tidak terjadinya sesuatu. Misal terjadinya gempa adalah peristiwa. Tidak ada gempa juga sebuah peristiwa yaitu peristiwa aman tidak ada bencana. Lambang dari peristiwa biasanya berupa huruf besar seperti A, B, C, dst. Suatu peristiwa juga dapat didefinisikan secara verbal, dengan kata atau kalimat, atau dilakukan dengan membuat daftar *outcome* yang sesuai dengan definisi dari peristisa tersebut.

Suatu peristiwa, misalnya peristiwa E, merupakan subset dari set S (ruang cuplikan, *universe*). Ruang cuplikan tentang adanya tiga anak dalam suatu keluarga seperti di atas dapat ditulis kembali dalam bentuk tabel seperti Tabel 1.1. berikut.

Tabel 1.1. Set *outcome* dari 3 anak dalam satu keluarga, dengan asumsi probabilita untuk punya anak laki-laki dan probabilita untuk punya anak wanita sama., P(Pria) = P(Wanita) = 0,5.

| *Outcome* | Lambang | Probabilita |
|---|---|---|
| P P P | e1 | 1/8 |
| P P W | e2 | 1/8 |
| P W P | e3 | 1/8 |
| P W W | e4 | 1/8 |
| W P P | e5 | 1/8 |
| W P W | e6 | 1/8 |
| W W P | e7 | 1/8 |
| W W W | e8 | 1/8 |

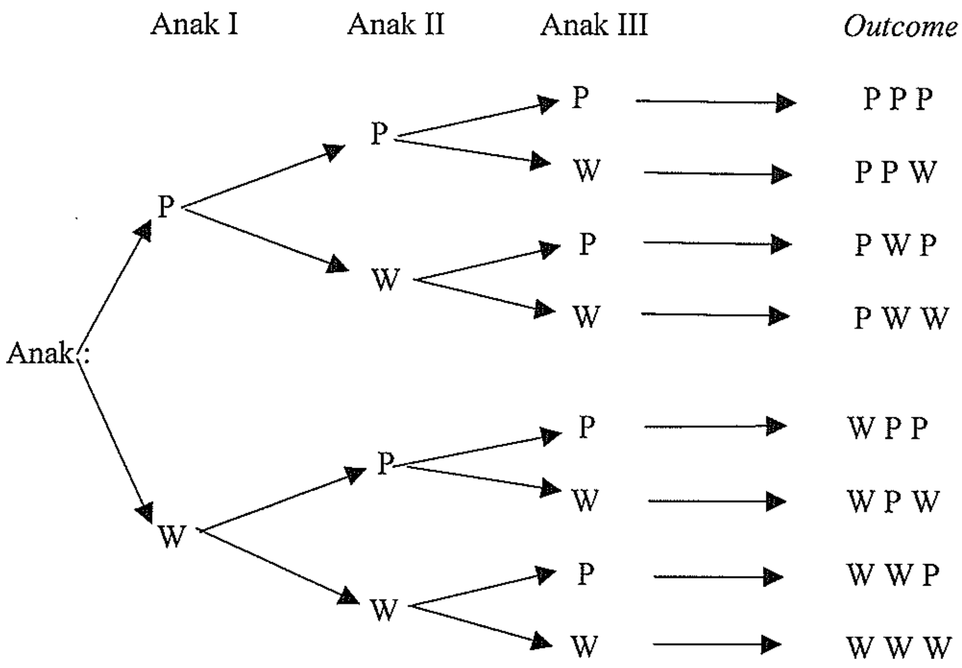
Beberapa contoh peristiwa (*event*) tentang rencana keluarga untuk mempunyai 3 anak dapat ditunjukkan dengan contoh seperti dalam tabel 1.2.

Tabel 1.2. Tiga cara alternatif untuk memberi nama / lambang bagi suatu peristiwa.

| Simbol peristiwa | Deskripsi dg kata-kata | *Outcome* | Daftar Probabilita |
|---|---|---|---|
| E | Paling sedikit 2W | $\{e_4,e_6,e_7,e_8\}$ | 4/8 |
| F | Anak kedua W diikuti satu P | $\{e_3,e_7\}$ | 2/8 |
| G | Kurang dari 2W | $\{e_1,e_2,e_3,e_5\}$ | 4/8 |
| H | Semua dengan sex sama | $\{e_1,e_8\}$ | 2/8 |
| I | Tanpa W | $\{e_1\}$ | 1/8 |
| I1 | Tepat 1W | $\{e_2,e_3,e_5\}$ | 3/8 |
| I2 | Tepat 2W | $\{e_4,e_6,e_7\}$ | 3/8 |
| I3 | Tepat 3 W | $\{e_8\}$ | 1/8 |

Tiga kolom pertama dari Tabel 1.2. menunjukkan 3 cara untuk memberi nama suatu peristiwa. Misalnya peristiwa E menggambarkan satu keluarga yang menghendaki adanya paling sedikit 2 anak laki-laki dari 3 anak yang direncanakan, maka *outcome* yang sesuai dengan kriteria tersebut adalah $e_1$, $e_2$, $e_3$ dan $e_5$. Dengan kata lain peristiwa E = paling sedikit 2 P, dapat ditunjukkan atau disusun *outcome*-nya sebagai
E = {e1, e2, e3, e5}
dan probabilitanya
P(E) = 1/8 + 1/8 + 1/8 + 1/8 = 4/8 = 1/2.

Cara lain untuk menunjukkan hasil percobaan adalah dengan menggambarkan pohon hasil percobaan (outcome trees). Dari contoh tentang rencana untuk mempunyai tiga anak, pohon hasil percobaannya dapat digambarkan sebagai berikut :




1.6. Perhitungan Probabilita.

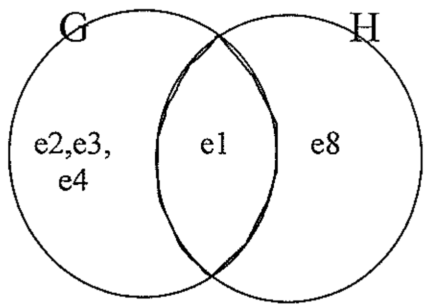
Antara dua peristiwa atau lebih dapat ditemukan gabungan, potongan, atau komplemen peristiwa.

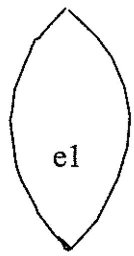
a. Gabungan peristiwa.
   Secara umum gabungan antara dua peristiwa, misalnya G atau H dilambangkan dengan $G \cup H$, adalah set *outcome* yang termasuk dalam G atau termasuk H atau kedua-duanya.
   Dari Tabel 1.2. yang termasuk dalam $G \cup H$ adalah {e1,e2,e3,e5 dan e8} sehingga probabilitanya sebesar 5/8.

b. Potongan peristiwa.
   Potongan dua peristiwa, misalnya G dan H dilambangkan dengan $G \cap H$, adalah set outcome yang termasuk baik dalam G maupun dalam H.
   Dari Tabel 1.2. yang termasuk dalam $G \cap H$ adalah {e1}.

   Kombinasi antara dua peristiwa atau lebih yang berupa gabungan atau potongan dapat pula disajikan dengan diagram Venn.

$G \cup H$ $\qquad\qquad$ $G \cap H$




Sekarang kita dapat memberikan rumus untuk kombinasi antara dua peristiwa, sebagai berikut (aturan umum penambahan dalam probabilita) :

$$G \cup H = G + H - G \cap H,$$

atau dalam bentuk probabilita

$$P(G \cup H) = P(G) + P(H) - P(G \cap H).$$

Jika $G \cup H = G + H$, aturan khusus penambahan dalam probabilita, yang berarti antara keduanya tidak ada tumpang tindihnya (potongan/irisan/*intersection*), maka kedua peristiwa tersebut disebut terpisah secara bersama (*mutually exclusive*).

c. Komplemen peristiwa.
   Komplemen dari suatu peristiwa adalah set *outcome* yang tidak termasuk dalam peristiwa tersebut.
   Dari Tabel 1.2. komplemen peristiwa $E = E^c = \overline{E}$ = bukan E mempunyai anggota {e5,e6,e7,e8}.

De Morgan's Laws (hukum De Morgan) :

$\overline{E \cap F} = \overline{E} \cup \overline{F}$ (bukan E dan F sama dengan bukan E atau bukan F)

$\overline{E \cup F} = \overline{E} \cap \overline{F}$ (bukan E atau F sama dengan bukan E dan bukan F)

Buktikan dengan diagram Venn adanya hukum tersebut.

Latihan :

1. Buatlah ruang sampel tentang suatu keluarga yang merencanakan untuk mempunyai 4 orang anak. Dengan asumsi bahwa probabilita untuk punya anak laki-laki 0,6, berapakah probabilita keluarga tersebut untuk :
   a. punya dua anak perempuan
   b. punya paling banyak satu anak laki-laki
   c. punya anak semua laki-laki

2. Mahasiswa dari suatu fakultas mempunyai kesenangan terhadap berbagai jenis olah raga dengan proporsi sebagai berikut : Sepakbola 30 % dari seluruh mahasiswa, basket 20 %, bisbol 20%, sepakbola dan basket 5 %, sepakbola dan bisbol 10 %, basket dan bisbol 5 % , tiga jenis olah raga tersebut 2 %. Jika seorang mahasiswa dipilih dengan undian untu suatu wawancara, berapa probabilita :
   a. dia seorang atlit (senang salah satu olah raga tersebut)
   b. dia hanya senang sepakbola
   c. dia senang sepakbola atau bisbol.

## 1.7. Probabilita Bersyarat.

Probabilita bersyarat adalah probabilita yang berhubungan dengan peristiwa dalam sub-himpunan yang dibatasi oleh syarat-syarat tambahan yang tidak terdapat dalam populasinya. Bila P(B) > 0, probabilita bersyarat dari peristiwa A dengan ketentuan (syarat) peristiwa B menjadi :

$$P(A|B = \frac{P(A \cap B}{P(B)}$$

Contoh : Dengan merujuk pada Tabel 1.2. Misalnya peristiwa G telah terjadi, berapa probabilita terjadinya peristiwa H ? Dengan asumsi probabilita P = probabilita W = 0,50, maka $P(H \cap G)$ adalah sebesar 1/4 = 0,25. Karena dari 4 *outcome* yang termasuk G hanya ada satu *outcome* yang masuk dalam H.

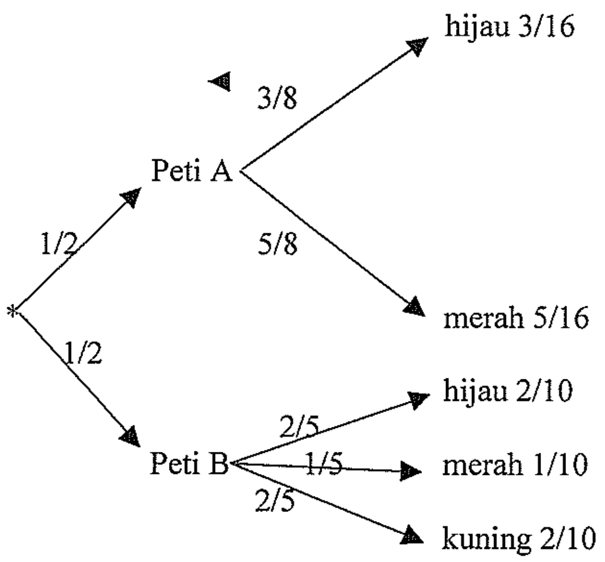
Peristiwa F dikatakan independen secara statistik dari E bila $P(F | E) = P(F)$.

Bagi peristiwa-peristiwa independen, $P(E \cap F) = P(E) * P(F)$.

$P(A \cap B) = P(A) * P(B | A) = P(B) * P(A | B)$ merupakan rumus probabilita berganda (*compound probability*). Rumus ini penting untuk menghitung probabilita peristiwa yang terdiri dari serangkaian percobaan ganda. Pada pelemparan sekeping uang logam, kita dapat lemparkan uang logam tersebut sekali, kemudian melemparkannya untuk kedua kali dan untuk ketiga kalinya. Percobaan seperti itu dinamakan percobaan ganda dengan 3 kali percobaan.

Contoh : Peti A terisi dengan 3 bola hijau dan 5 bola merah.. Peti B terisi dengan 2 bola hijau, 1 bola merah dan 2 bola kuning. Bila kita memilih sebuah peti secara acak dan kemudian memilih satu bola dari dalamnya secara acak pula, berapa probabilita kita akan memilih bola hijau ?

Keadaan tersebut dapat digambarkan dengan diagram pohon seperti berikut :




Latihan :
Di suatu daerah terpencil, 75 % penduduk dewasa adalah pria dan sisanya (25 %) adalah wanita. Tingkat persetujuan pria terhadap KB sebesar 83 % dan tingkat persetujuan wanita terhadap KB sebesar 48 %. Berapa proporsi yang setuju terhadap KB adalah wanita ?

## 1.8. <u>Soal.</u>

Gunakan petunjuk A untuk menjawab soal no. 1 s.d. no. 2.

1. Dalam pelemparan 3 buah dadu, ruang sampel yang terbentuk terdiri dari :
   A. 18 titik sampel
   B. 27 titik sampel
   C. 108 titik sampel
   D. 216 titik sampel
   E. 729 titik sampel

2. Apabila $P(A \cap B) = P(A) \times P(B)$, maka peristiwa A dan peristiwa B merupakan :
   A. Peristiwa saling lepas
   B. Peristiwa independent
   C. Peristiwa biasa
   D. Peristiwa gabungan
   E. Peristiwa potongan

Gunakan petunjuk B untuk menjawab soal no. 3 s.d. no. 4.

3. Bila probabilita penderita penyakit paru-paru dan merokok 0,4 ; probabilita penyakit paru-paru dan tidak merokok 0,2 ; probabilita bukan penderita penyakit paru-paru dan merokok 0,3 ; dan probabilita bukan penderita penyakit paru-paru dan tidak merokok 0,1 , maka
   (1) Probabilita penderita penyakit paru-paru 0,6
   (2) Probabilita perokok 0,6
   (3) Probabilita perokok atau penderita penyakit paru-paru 0,9
   (4) Probabilita tidak merokok atau bukan penderita penyakit paru-paru 0,1

4. Hubungan antara dua peristiwa dapat merupakan peristiwa :
   (1) Saling lepas
   (2) Independen
   (3) Komplemen
   (4) Bersyarat

Gunakan petunjuk C untuk menjawab soal no. 5 s.d. no. 6.

5. Probabilita tidak sama dengan kesempatan
   SEBAB
   Kesempatan tidak dapat diukur probabilitanya.

6. Probabilita H dengan syarat G selalu lebih kecil daripada probabilita H sendiri
   SEBAB
   Peristiwa H dan peristiwa G tidak punya potongan (irisan).

7. Lima lampu pijar yang rusak tercampur menjadi satu dengan 10 buah lampu yang masih baik. Buruh perusahaan diinstruksikan guna mencari kembali lampu pijar yang rusak. Bila buruh tersebut secara acak mengambil 3 buah lampu pijar, berapakah probabilita :
   a. Tidak ada satupun dari ketiga lampu tersebut ternyata rusak
   b. Satu saja yang rusak
   c. Paling sedikit satu lampu rusak.

## 1.9. Rangkuman.

Probabilita merupakan suatu ukuran tentang ketidakpastian yang biasanya diwujudkan dalam bentuk proporsi atau prosentase, dengan nilai antara 0 dan 1. Perumusan probabilita yang banyak dipakai adalah perumusan atas dasar frekuensi relatif. Setiap peristiwa yang merupakan subset dari ruang cuplikan selalu dapat ditentukan probabilitanya, demikian pula untuk gabungan, potongan maupun komplemen peristiwa. Probabilita bersyarat dan probabilita berganda merupakan konsep penting dalam menentukan probabilita yang merupakan hasil eksperimen berganda.

# POKOK BAHASAN II
# DISTRIBUSI PROBABILITA

## 2.1. Pendahuluan

Setelah anda mempelajari pokok bahasan tentang probabilita dengan baik, maka anda diharapkan tidak menemui kesulitan yang berarti dalam mempelajari distribusi probabilita ini.

Pokok bahasan ini mencakup sub-pokok bahasan tentang :

a. Variabel Rambang : diskrit dan sinambung
b. Probabilita Diskrit
c. Probabilita Sinambung
d. Hubungan Antar Distribusi

## 2.2. T I U dan Sasaran Belajar

**T I U** :
Setelah mempelajari pokok bahasan ini mahasiswa diharapkan mampu mengetahui dan memahami berbagai macam distribusi probabilita.

**Sasaran Belajar** :
Mahasiswa setelah mempelajari pokok bahasan ini dapat :

1. Membedakan variabel diskrit dan variabel sinambung
2. Menggambarkan distribusi tepritis variabel rambang diskrit.
3. Menggambarkan distribusi teoritis variabel rambang sinambung
4. Melakukan kembali percobaan Bernouli dengan menggunakan sekeping uang logam.
5. Menjelaskan fungsi probabilita peristiwa "sukses" atau "gagal"
6. Menghitung probabilita peristiwa dengan rumus probabilita binomial.
7. Menghitung distribusi binomial secara rekursif.
8. Menggunakan tabel distribusi binomial dan distribusi binomial kumulatif.
9. Menggunakan rumus rata-rata distribusi binomial.
10. Menggunakan rumus varians dan simpangan baku distribusi binomial.
11. Menunjukkan 2 ciri distribusi Poisson.
12. Menunjukkan kembali fungsi Poisson
13. Menunjukkan hubungan antara distribusi binomial dan distribusi Poisson.
14. Menggunakan tabel distribusi normal untuk menunjukkan probabilita peristiwa.
15. Menggunakan rumus rata-rata dan simpangan baku distribusi normal.
16. Menggunakan tabel distribusi t untuk menunjukkan probabilita peristiwa.
17. Menggunakan rumus rata-rata dan simpangan baku distribusi t.
18. Menggunakan tabel distribusi kai-kuadrat untuk menunjukkan probabilita peristiwa.
19. Menunjukkan hubungan antara distribusi binomial dan distribusi normal.
20. Menunjukkan hubungan antara distribusi normal dan distribusi t.

2.3. Variabel Rambang Diskrit dan Sinambung

Variabel yang nilainya merupakan suatu bilangan ditentukan oleh terjadinya hasil suatu percobaan dinamakan variabel rambang (random,acak). Variabel rambang tersebut, berdasar skala pengukurannya, dapat dibedakan menjadi variabel rambang diskrit dan variabel rambang sinambung.Suatu variabel rambang disebut diskrit apabila variabel tersebut hanya dapat mengambil nilai-nilai terbatas jumlahnya. Misalnya suatu variabel yang hanya dapat dinyatakan atau mengambil nilai berupa bilangan bulat 0 dan 1. Sedangkan apabila variabel tersebut dapat mengambil nilai yang dianggap dapat dinyatakan dengan sembarang angka yang terdapat dalam suatu interval atau kelompok interval tertentu, maka variabel tersebut disebut variabel sinambung. Misalnya di antara 0 dan 1 ada nilai 0,5; 0,25; 0,6; dst.

Latihan : Berikan paling sedikit 3 contoh untuk masing-masing variabel diskrit dan variabel sinambung.

2.4. Probabilita Diskrit

Variabel rambang diskrit yang hanya dapat mengambil dua alternatif nilai, misalnya: 0 dan 1, sukses dan gagal, pria dan bukan pria, memenuhi standar dan tidak memenuhi standar, disebut variabel binomial.

**Percobaan Bernoulli**

Satu atau serangkaian percobaan dinamakan percobaan binomial bila dan hanya bila percobaan yang bersangkutan terdiri dari percobaan-percobaan Bernoulli. Suatu percobaan dinamakan percobaan Bernoulli bila dan hanya bila memiliki ciri-ciri sebagai berikut :

a. Tiap percobaan hanyalah mempunyai dua hasil, sukses atau gagal. Dirumuskan dalam ruang cuplikan {S,G}. S bila sukses, dan G bila gagal.
b. Probabilita sukses untuk tiap-tiap percobaan haruslah sama dan dinyatakan dengan $\pi$ (phi).
c. Setiap percobaan harus bersifat independen.
d. Jumlah percobaan yang merupakan komponen percobaan binomial harus tertentu.

Percobaan Bernoulli ini mendasari percobaan binomial, yang mempunyai penggunaan yang luas di bidang ekonomi.

**Distribusi Binomial**

Contoh klasik variabel binomial adalah S = jumlah "*heads*" (H) dalam pelemparan satu coin.

Misalkan :

1. Ada sejumlah n pelemparan coin
2. Dalam tiap lemparan, peristiwa tertentu dapat terjadi, yaitu berhasil (bila muncul H) atau gagal (bila yang muncul bukan H). Jika probabilita berhasil $P(H)= \pi$, maka probabilita gagal, $P(T) = P(H^c) = (1 - \pi)$. T = *tail* atau "ekor" dari suatu coin.
3. Pelemparan adalah independen (bebas) secara statistik.

Maka, distribusi binomial mempunyai probabilita

$$P(s) = f(s) = {}_nC_s\, \pi^s (1-\pi)^{n-s} = C\binom{n}{s}\pi^s (1-\pi)^{n-s}$$

dengan koefisien binomial =

$${}_nC_s = C\binom{n}{s} = \frac{n!}{s!(n-s)!}$$

Dari contoh tentang rencana untuk mempunyai tiga anak, probabilita suatu keluarga untuk mempunyai satu anak wanita, dengan asumsi P(anak wanita) = P(W) = 0,5, adalah
P(satu wanita) =

$$P(1) = f(1) = {}_3C_1\, \pi^1 (1-\pi)^{3-1} = C\binom{3}{1}\pi^1 (1-\pi)^{3-1}$$

$$= 3\ (0{,}5)^1 (0{,}5)^2 = 0{,}375$$

Latihan : Carilah P(3 wanita), P(2 wanita), P(3 pria).

Contoh soal :
Setelah diadakan penyelidikan bertahun-tahun lamanya terhadap hasil copian mesin photocopy "C", diketahui bahwa pada tiap pengcopian 1350 lembar kertas HVS ukuran kwarto akan terdapat 135 lembar yang rusak. Dalam mengcopi 5 lembar kertas HVS ukuran kwarto, berapakah probabilita untuk menemukan 0, 1, 2, 3, 4, 5 lembar yang rusak?

Jawab : Probabilita kerusakan 135/1350 = 0,1 berarti

$\pi = 0{,}1.$

$n = 5$

$s = \{0, 1, 2, 3, 4, 5\}$

$$P(s=0) = C\binom{5}{0}(0{,}1)^0 (1-0{,}1)^5$$

$$= 1\ (0{,}5)^0 (0{,}5)^5 = 0{,}59049$$

$$P(s=1) = C\binom{5}{1}(0{,}1)^1 (1-0{,}1)^4$$

$$= 5\ (0{,}1)^1 (0{,}9)^4 = 0{,}32805$$

$$P(s=2) = C\binom{5}{2}(0{,}1)^2 (1-0{,}1)^3$$

$$= 10(0{,}1)^2 (0{,}9)^3 = 0{,}07290$$

$$P(s=3) = C\binom{5}{3}(0{,}1)^3 (1-0{,}1)^2$$

$$= 10(0{,}1)^3 (0{,}9)^2 = 0.00810$$

$$P(s = 4) = C\binom{5}{4}(0,1)^4(1-0,1)^1$$
$$= 5(0,1)^4\,(0,9)^1 = 0,00045$$
$$P(s = 5) = C\binom{5}{5}(0,1)^5(1-0,1)^0$$
$$= 1\,(0,1)^5(0,9)^0 = 0,00001$$

Dari contoh soal di muka kita juga dapat mencari berapa probabilita kerusakan paling sedikit 2 lembar, 3 lembar dan seterusnya. Persoalan seperti ini dapat dipecahkan dengan mencari distribusi binomial kumulatif

Misalnya probabilita kerusakan paling sedikit 2 lembar berarti sebesar jumlah probabilita kerusakan 0 lembar, 1 lembar dan 2 lembar; sehingga probabilitanya menjadi

$P(s \leq 2) = 0,59049 + 0,32805 + 0,07290 = 0,99144$

**Rata-rata hitung dan Varians variabel binomial**

Bila nilai parameter n dan $\pi$ (phi) telah diketahui, maka menghitung rata-rata distribusi binomial dapat dilakukan dengan mudah sekali.

Rumus rata-rata hitung variabel binomial adalah

$$\mu_s = n\,\pi$$

Contoh : Bila satu dadu dilempar sebanyak 4 kali, berapa rata-rata probabilita memperoleh hasil mata dadu 6 ?

Jawab :

Probabilita timbulnya jumlah mata dadu 6 dalam percobaan pelemparan dadu sebanyak 4 kali adalah sebagai berikut :

| s | nCs | $\pi^s(1-\pi)^{n-s}$ | P(s) |
|---|---|---|---|
| 0 | 1 | $(1/6)^0(5/6)^4$ | 0,386 |
| 1 | 4 | $(1/6)^1(5/6)^3$ | 0,482 |
| 2 | 6 | $(1/6)^2(5/6)^2$ | 0,116 |
| 3 | 4 | $(1/6)3(5/6)^1$ | 0,015 |
| 4 | 1 | $(1/6)^4(5/6)^0$ | 0,001 |

Oleh karena itu, rata-rata probabilita timbulnya mata dadu 6, $\mu_s$ = 0 x 0,482) + 1 x (0,386) + 2 x (0,116) + 3 x (0,015) + 4 x (0,001) = 0,667

Dengan rumus, $\mu_s = n\,\pi$

$= 4\,(1/6) = 0,667$

Rumus varians bagi variabel binomial adalah

$$\sigma^2 = n\,\pi(1 - \pi)$$

Sedangkan deviasi standarnya adalah akar pangkat dua dari varians tersebut.

$$\sigma = \sqrt{n\,\pi\,(1-\pi)}$$

Contoh : Dari soal pelemparan dadu sebanyak 4 kali, berapa varians dan deviasi standar probabilita timbulnya mata dadu 6 ?

Jawab :

Varians $= \sigma^2 = n\,\pi(1 - \pi)$

$= 4\,(1/6)(5/6)$

$= 20/36 = 0{,}555556$

Deviasi standarnya =

$$\sigma = \sqrt{n\,\pi\,(1-\pi)}$$

$$= \sqrt{4(1/6)(5/6)}$$

$$= 0{,}745356$$

**Distribusi Poisson**

Salah satu bentuk lain dari perhitungan probabilita variabel diskrit adalah pendekatan distribusi Poisson. Sebenarnya distribusi ini tetap merupakan distribusi binomial, hanya saja proporsi *sukses* dan *gagal*-nya sangat tidak seimbang dan dengan n yang sangat besar. Dalam keadaan seperti ini penggunaan distribusi binomial sangat sulit diterapkan.

Bila kita mempersamakan $n\pi = \mu_X$ maka distribusi Poisson yang aproksimatif dapat diberikan sebagai berikut:

$$P(s) = f(x) = \frac{\mu_x e^{-\mu_x}}{x!}$$

dimana X dapat merupakan nilai 0, 1, 2,.... n dan
e = 2,71828 ....

Contoh :

Menurut hasil penelitian dan pengalaman para karyawan Perusahaan XYZ, sebuah mesin cetak merk "R" pada tiap mencetak 2000 lembar kertas ukuran folio akan membuat kerusakan 1 lembar kertas. Perusahaan yang memakai mesin tersebut ingin sekali mengetahui berapakah probabilita kerusakan 0,1,2,3,4 dan 5 lembar kertas pada tiap pencetakan 1000 lembar kertas ukuran folio.

(Di sini $\pi < 0{,}1$ dan $n > 50$, maka pendekatan Poisson bisa dipakai).

Jawab :

Karena $\pi = 1/2000$, $n = 1000$, maka $\mu_X = 0{,}5$. Distribusi Poisson secara berturut dapat diberikan sebagai berikut

$$P(0) = f(0) = \frac{(0{,}5)^0 e^{-0{,}5}}{0!}$$

$$= 0{,}606531$$

$$P(4) = f(4) = \frac{(0{,}5)^4 e^{-0{,}5}}{4!}$$

$= (1/384) \times 0{,}6066$

$= 0{,}00158$

$$P(5) = f(5) = \frac{(0,5)^5 e^{-0,5}}{5!}$$

$= (1/3840) \times 0,6066$

$= 0,000184$

Distribusi Poisson mempunyai rata-rata hitung ($\mu_x$) dan Varians ($\sigma_x^2$) yang dapat dirumuskan sebagai berikut :

Rata-rata hitung $= \mu_X = n\pi$

Sedangkan variansnya $= \sigma_2^2 = \mu_X = n\pi$

Oleh karena itu deviasi standarnya $= \sigma_x = \sqrt{n\pi}$

Contoh :

Sebuah perusahaan radio telah melakukan pengawasan kualitas terhadap 1000 buah radio yang baru diterima dari divisi produksi. Jumlah radio yang tidak memenuhi spesifikasi kualitas standar (X) diberikan sebagai berikut :

| X | 0 | 1 | 2 | 3 | 4 | Jumlah |
|---|---|---|---|---|---|---|
| f | 950 | 47 | 2 | 1 | 0 | 1000 |

a, Buat distribusi Poisson-nya

b. Carilah rata-rata, varians dan deviasi standarnya

Jawab :

a. Distribusi Poisson

| X | f | fX |
|---|---|---|
| 0 | 950 | 0 |
| 1 | 47 | 47 |
| 2 | 2 | 4 |
| 3 | 1 | 3 |
| 4 | 0 | 0 |
| | 1000 | 54 |

$$\overline{X} = \frac{\sum fX}{n} = 54/1000 = 0,054$$

$\sum fX = 54$

$n = 1000$

$$f(0) = \frac{(0,054)^0 (2,71828)^{-0,054}}{0!} = 0,947432$$

$$f(1) = \frac{(0,054)^1 (2,71828)^{-0,054}}{1!} = 0,051161$$

$$f(2) = \frac{(0,054)^2 (2,71828)^{-0,054}}{2!} = 0,001381$$

$$f(3) = \frac{(0{,}054)^3(2{,}71828)^{-0{,}054}}{3!} = 2{,}48644E\text{-}05 = 0{,}0000248644$$

$$f(4) = \frac{(0{,}054)^4(2{,}71828)^{-0{,}054}}{4!} = 3{,}3567E\text{-}07 = 0{,}00000033567$$

Rata-rata distribusi Poisson = $\mu_X = 0{,}054$
Varians distribusi Poisson = $\sigma_X{}^2 = 0{,}054$
Deviasi standar distribusi Poisson $\sigma = \sqrt{\sigma_x^2} = \sqrt{0{,}054}$
$= 0{,}232379$

2.5. <u>Distribusi Probabilita Sinambung</u>
Distribusi probabilita sinambung terjadi baik karena variabelnya memang bersifat sinambung maupun karena pengulangan atau repetisi secara banyak kali dari variabel diskrit. Variabel bersifat sinambung bila variable tersebut dapat mengambil alternative nilai berupa angka bulat maupun angka pecahan. Termasuk dalam distribusi probabilita sinambung antara lain distribusi normal atau distribusi Z (normal standar), distribusi t, distribusi F dan distribusi $\chi^2$.

**Distribusi normal**
Merupakan distribusi sinambung yang sangat penting, karena sering dijadikan asumsi dasar dalam menggambarkan distribusi suatu variabel.

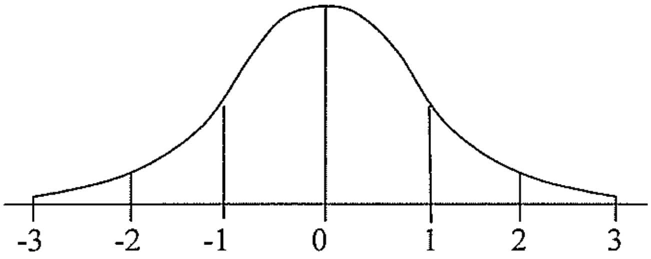
Rata-rata hitung untuk distribusi normal dapat diperoleh dengan rumus $\mu_X = \frac{\sum X_i}{n}$, sedangkan deviasi standarnya adalah $\sigma_X = \sqrt{\frac{\sum(X_i - \mu_X)^2}{N}}$ yang memerlukan adanya cuplikan atau sampel besar. (n $\geq$ 25).
Distribusi normal merupakan distribusi yang bersifat simetris, berbentuk genta, kontinyu dan memiliki fungsi kepekatan normal (normal density function) sebagai berikut.

$$f(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-(1/2\sigma_X^2)(x-\mu_X)^2}$$

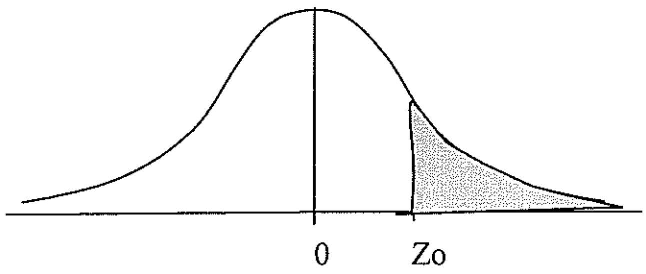
Dari fungsi kepekatan normal tersebut dapat diubah menjadi fungsi kepekatan normal standar f(Z) dengan rumus:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-(1/2z)^2} \quad \text{di mana} \quad z = \frac{X - \mu_X}{\sigma_X}$$

Gambar kurve normal standar




Setiap nilai Z menunjukkan nilai deviasi standar dari rata-rata hitungnya. Bagi distribusi variabel normal standar ini sudah disusun suatu table yang biasa disebut Tabel Z yaitu table yang menunjukkan luas daerah / probabilita yang dibatasi oleh nilai-nilai Z terentu. Untuk nilai Z negatif, dapat diperoleh dengan menggunakan azas simetri.




Penggunaan dalam ekonomi

Contoh : Rata-rata berat ayam potong usia 35 hari sebesar 2,75 kg, dengan deviasi standar sebesar 0,98 kg. Dengan cuplikan sebanyak 49 ayam potong, ada berapa banyak ayam potong tersebut yang beratnya kurang dari 2,4 kg?

Jawab : $\mu = 2{,}75 \quad \sigma = 0{,}46 \quad n = 49$

$P(X) < 2{,}4$ ?

$$z = \frac{X - \mu_X}{\sigma_X} \rightarrow \quad z = \frac{2{,}4 - 2{,}75}{0{,}98 / \sqrt{49}} = \frac{-0{,}35}{0{,}14} = -2{,}5$$

$P(X) < 2{,}4 = P(Z<-2{,}5) = 0{,}006$

Berarti ayam potong usia 35 hari yang beratnya kurang dari 2,4 kg sebesar 0,6 %..

**Distribusi t**

Distribusi t ini merupakan pengembangan dari distribusi Z, yaitu lebih bermanfaat dalam penggunaan cuplikan kecil (n < 25) atau bila deviasi standar populasi ($\sigma_X$) tidak diketahui.. Dalam hal cuplikan kecil atau dalam pengambilan cuplikan tanpa dengan pengembalian, diperlukan perhatian terhadap d.f (*degree of freedom*). Distribusi t yang disusun dalam suatu table yang biasa disebut tabel t, biasanya terdapat pada lampiran tabel pada buku-buku statistika inferensia. Besarnya d.f pada dasarnya sama dengan pembagi atau penyebut dalam rumus untuk menghitung deviasi standarnya.

Rata-rata hitung untuk distribusi t dapat diperoleh dengan rumus $\overline{X}_X = \frac{\sum X_i}{n}$ dengan

deviasi standar $s_X = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}}$

Variabel t standar dapat diperoleh dengan rumus $t = \frac{X - \overline{X}_X}{s/\sqrt{n}}$

Penggunaan dalam ekonomi

Contoh : Dari suat survey dengan cuplikan acak sebanyak 16 kaleng susu diketahui bahwa rata-rata berat susu dalam kaleng tersebut sebesar 1,01 kg dengan deviasi standar sebesar 0,02 kg. Ada berapa prosen dari susu dalam kaleng tersebut yang kurang memenuhi standar (berat = 1,00 kg)?

Jawab : $n = 16$ $\overline{X} = 1{,}011$ $s = 0{,}02$

Distribusi t

$$t = \frac{X - \overline{X}_X}{s/\sqrt{n}} = \frac{1 - 1{,}011}{0{,}02/\sqrt{16}}$$

$$= (-0{,}011)/(0{,}005) = -2{,}11$$

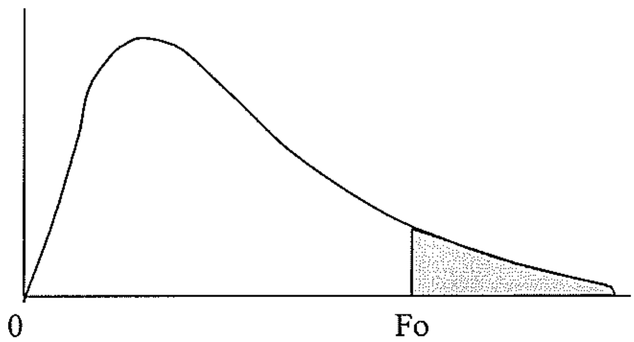
dengan menggunakan tabel t, pada d.f sebesar n-1 atau sebesar 15, $t < -2{,}11$ sekitar 0,025 atau sekitar 2,5 %.

**Distribusi F**

Distribusi F merupakan salah satu distribusi probabilita, yang diberi nama demikian untuk menghormati Sir Ronald Fieher, salah satu pelopor dari statistika modern. Beberapa karakteristik dari distribusi F adalah sebagai berikut :

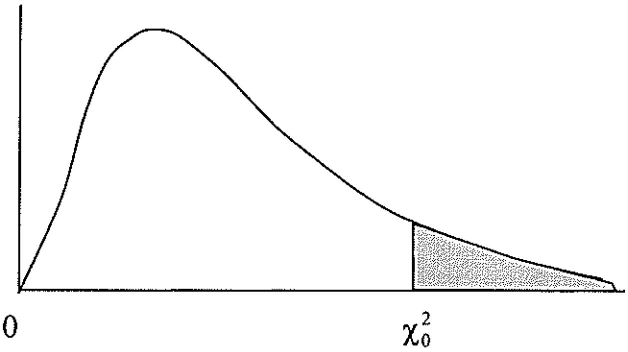
1. Ada suatu keluarga dari distribusi F, yang nilainya ditentukan oleh dua parameter yang d.f (*degree of freedom*) pembilang dan d.f penyebut
2. Nilai F tidak mungkin negative daan merupakan distribusi yang bersifat kontinyu.
3. Bentuk distribusi F tidak simetris, melainkan menceng ke kiri (*positively skewed)*
4. Nilai F antara 0 daan ∞. Ketika nilai F makin besar, kurvenya makin mendekati sumbu datar (absis( tetapi tidak pernah menyentuhnya.

Gambar kurve distribusi F secara umum adalah sebagai berikut :



Pada luas daerah (probabilita) tertentu, nilai F mungkin berbeda-beda tergantung pada kombinasi antara d.f pembilang (*numerator*) dan d.f penyebut (*denumerator*).
Distribusi F ini sangat berguna dalam analisis regresi dan analisis varians seperti akan diuraikan dalam Pokok Bahasan VIII.

**Distribusi Kai-kuadrat ($\chi^2$)**
Walaupun sebenarnya distribusi Kai-kuadrat termasuk distribusi diskrit, tetapi pada umumnya didekati sebagai distribusi kontinyu untuk lebih mempermudah perhitungan.
Nilai Kai-kuadrat yang membatasi suatu luas daerah (probabilita) di bawah kurve Kai-kuadrat yang tertentu akan berubah tergantung pada besar kecilnya d.f.
Bentuk distribusi Kai-kuadrat yang kontinyu sangat mirip dengan distribusi F, hanya saja dalam distribusi ini hanya ada satu d.f. Nilai distribusi Kai-kuadrat tersebut telah disusun dalam suatu tabel yang disebut tabel distribusi Kai-kuadrat,
Secara umum gambar distribusi Kai-kuadrat ($\chi^2$) adalah sebagai berikut.




Ada beberapa penggunaan dari distribusi Kai-kuadrat, yang sebenarnya termasuk dalam kelompok uji statistika non-parametrik, seperti uji kebaikan suai (*goodness of fit*), uji independensi, uji homogenitas. uji kontingensi, dan uji parameter deviasi standar.

## 2.6. <u>Hubungan Antar Distribusi</u>

**Distribusi binomial dan distribusi normal**
Distribusi normal dapat digunakan sebagai pendekat bagi distribusi binomial, terutama bila cuplikan yang digunakan adalah cuplikan besar. Memang antara hasil perhitungan dengan pendekatan diskrit dan pendekatan kontinyu ada perbedaan, tetapi perbedaan tersebut umumnya tidaklah besar terutama pada ordinat sentralnya.

**Distribusi normal dan distribusi t**
Distribusi normal standar (distribusi Z) hanya dapat digunakan bila cuplikan yang digunakan adalah cuplikan besar,. Bila cuplikan yang digunakan termasuk cuplikan kecil, di mana d.f harus diperhatikan maka distribusi yang tepat adalah distribusi t. Distribusi t akan sama dengan distribusi Z pada d.f. tidak terhingga ($\infty$).

**Distribusi t dan distribusi F**
Dalam keadaan tertentu nilai F sama dengan nilai $t^2$. Pada umumnya distribusi F mempunyai penggunaan yang lebih banyak dibanding dengan distribusi t.

## 2.7. Soal

1. Sebuah perusahaan industri yang memproduksi alat-alat pertanian mengetahui secara teknis bahwa 25 persen dari alat-alat yang diproduksi dengan mesin tertentu akan tidak memenuhi kualitas standard an dianggap sebagai produk BS. Jika 10 buah alat-alat pertanian yang dihasilkan dengan mesin di atas dipilih secara acak dari seluruh hasil produksi, berapakah probabilita:
   a. tiga dari 10 alat pertanian tadi rusak (sebagai BS)?
   b. tiga atau lebih yang rusak?
   c. lebih dari 4 yang rusak?
2. Suatu toko swalayan menerima kiriman kedelai kualitas bagus dengan kemasan 1 kg dalam partai besar. Toko tersebut ingin mengetahui butir pasir atau kerikil yang ada pada tiap kemasan kedelai. Untuk itu diambillah cuplikan acak sebanyak 50 kemasan kedelai isi 1 kg. Ternyata jumlah butir pasir atau kerikil (X) dalam kemasan dapat ditabulasikan sebagai berikut.

| Butir pasir (X) | 0 | 1 | 2 | 3 | 4 | Jumlah |
|---|---|---|---|---|---|---|
| Frekuensi (f) | 44 | 4 | 1 | 1 | 0 | 50 |

   a. Buatlah distribusi Poissonnya.
   b. Gambarkan grafiknya.
   c. Hitung rata-rata dan deviasi standarnya.
3. Variabel acak atau random X disitribusikan secara normal dengan rata-rata sebesar 82 dan deviasi standarnya sebesar 14.
   a. Berapakah $P(68 < X < 80)$?
   b. Tentukanlah $P(72 \leq X)$
   c. Hitunglah $P(70 < X < 100)$
4. Dengan cuplikan acak sebanyak 20 alumni dari program D3 XYZ ternyata rata-rata gaji awalnya pada saat dia bekerja adalah Rp 625.000,00 per bulan dengan deviasi standar sebesar Rp 174,500,00. Berapakah gaji pertama minimal dari 10 % alumni dengan gaji awal tertinggi?

## 2.8. Rangkuman

Probabilita atau peluang dari suatu peristiwa berupa proporsi atau persentase tertentu dari selaruh alternatif outcome. Oleh karena itu probabilita tersebut dapatlah digambarkan distribusinya sehingga terbentuklah suatu distribusi probabilita. Ada dua bentuk distribusi probabilita, yaitu distribusi probabilita diskrit dan distribusi probabilita kontinyu. Termasuk distribusi probabilita diskrit adalah distribusi binomial dan distribusi Poisson, dan yang termasuk distribusi probabilita kontinyu adalah distribusi normal (Z dan t), distribusi F dan distribusi $\chi^2$. Distribusi probabilita diskrit dapat pula didekati dengan distribusi probabilita kontinyu, terutama bila digunakan cuplikan besar. Distribusi probabilita kontinyu dapat ditunjukkan dalam suatu grafik dengan mengukur luas daerah di bawah kurve tertentu. Probabilita atau luas daerah tersebut sudah disusun dalam beberapa tabel misalnya Tabel Z, Tabel t, Tabel F dan Tabel $\chi^2$. Ada hubungan tertentu di antara distribusi-distribusi probabilita tersebut.

# POKOK BAHASAN III
# DASAR-DASAR TEORI PENCUPLIKAN

## 3.1. Pendahuluan

Dalam mempelajari pokok bahasan ini tidak banyak pengetahuan awal yang dikehendaki. Hal-hal yang perlu dikuasai sebelumnya hanyalah kemampuan untuk menafsirkan dan menghitung proporsi, rata-rata, median, dan varians, serta penggunaan Tabel Z.
Selengkapnya pokok bahasan ini mencakup :

a. Arti dan kegunaan pencuplikan
b. Rata-rata dan Jumlah Cuplikan
c. Dalil Batas Tengah.
d. Distribusi Pencuplikan :
   d.1. Tentang Rata-rata
   d.2. Tentang Proporsi
   d.3. Tentang Beda antara Dua Rata-rata
   d.4. Tentang Beda antara Dua Proporsi
   d.5. Tentang Varians
   d.6. Tentang Median

## 3.2. T I U dan Sasaran Belajar

T I U : Setelah mempelajari pokok bahasan ini mahasiswa diharapkan mampu memahami jenis-jenis cuplikan.

Sasaran Belajar :
Setelah mempelajari pokok bahasan ini mahasiswa dapat :

1. Menjelaskan arti cuplikan secara acak.
2. Menjelaskan 4 alasan untuk menggunakan cuplikan.
3. Membedakan cuplikan dengan pengembalian dan tanpa pengembalian, terutama dalam kaitannya dengan besar cuplikan.
4. Menghitung nilai rata-rata dan varians jumlah cuplikan.
5. Menghitung nilai rata-rata dan varians rata-rata cuplikan.
6. Menjelaskan arti dalil batas tengah.
7. Menghitung distribusi cuplikan rata-rata hitung bagi populasi terbatas dan tidak terbatas
8. Menghitung distribusi cuplikan proporsi bagi populasi terbatas dan tidak terbatas.
9. Menghitung distribusi cuplikan beda antara dua rata-rata.
10. Menghitung distribusi cuplikan beda antara dua proporsi
11. Menghitung distribusi cuplikan varians.
12. Menghitung distribusi cuplikan median.

## 3.3. Arti dan Kegunaan Pencuplikan.

Suatu populasi merupakan keseluruhan dari sesuatu yang sejenis dengan ciri-ciri tertentu, terdiri dari seluruh observasi aktual maupun hipotetis yang mungkin dilakukan terhadap fenomena (gejala) tertentu. Pada umumnya penyelidikan hanya dilakukan terhadap sebagian saja dari populasi, walaupun sebenarnya yang ingin digambarkan adalah keadaan dari seluruh populasinya. Sebagian dari populasi yang benar-benar diselidiki atau diobservasi tersebut dinamakan cuplikan (sampel, contoh). Cara untuk mengambil sebagian dari populasi disebut pencuplikan, pengambilan sampel atau *sampling*. .
Pencuplikan itu harus dilakukan secara acak (*random*) agar supaya dapat menggambarkan populasinya dengan baik atau merupakan wakil yang tidak bias. Pencuplikan dikatakan acak apabila setiap anggota populasi mempunyai kesempatan yang sama untuk dipilih sebagai anggota cuplikan. Beberapa bentuk pencuplikan acak adalah : pencuplikan acak sederhana, pencuplikan sistematis, pencuplikan acak bertingkat, dan pencuplikan acak berkelompok. Di samping pencuplikan acak ada pula pencuplikan non-acak (*non-random, purposive*) yang memang ditujukan untuk menemukan ciri-ciri tertentu saja dari populasi. Pencuplikan tipe kedua tersebut tidak dapat digunakan untuk melakukan generalisasi, penggambaran populasi secara keseluruhannya.
Populasi darimana cuplikan diambil dapat bersifat terbatas (diketahui jumlahnya secara pasti) atau bersifat tidak terbatas.
Dalam pencuplikan dikenal adanya pencuplikan dengan pengembalian dan pencuplikan tanpa pengembalian.
Contoh pencuplikan tanpa pengembalian adalah sistim arisan.

## 3.4. Rata-rata dan Jumlah Cuplikan.

Pengambilan cuplikan dari populasi terbatas.
Misalkan kita mempunyai 9 kartu yang diberi nomor 1,2,3 sampai dengan 9 dan dikocok secara merata betul. Kemudian kita mengambil 3 kartu dari 9 kartu tersebut secara acak, maka kita melakukan pengambilan cuplikan/sampel dari populasi yang hanya terdiri dari 9 kartu (terbatas). Bila kartu-kartu yang telah kita ambil tadi dikembalikan atau dicampur dengan kartu yang tidak terambil, maka kita dapat melakukan percobaan mengambil kartu-kartu secara berulang kali, bahkan secara tidak terbatas.
Jika pada tiap-tiap pengambilan, kita jumlahkan angka-angka yang terdapat pada ketiga kartu tersebut dan kemudian dibagi dengan jumlah kartu yang diambil yaitu sebanyak 3, maka kita akan memperoleh rata-rata dari angka yang terdapat pada ketiga kartu yang membentuk cuplikan/sampel eksperimental tersebut. Rata-rata sedemikian tadi diberi nama rata-rata cuplikan/sampel (*sample mean*) dan merupakan statistik cuplikan/sampel yang dapat dipakai untuk mengestimasi rata-rata populasinya.
Bila kita melakukan cuplikan/sampel eksperimental tersebut sebanyak 100 kali, kita akan memperoleh 100 cuplikan/sampel eksperimental dan 100 rata-rata cuplikan/sampel.
Penjumlahan angka-angka yang terdapat pada ketiga kartu yang terpilih dinamakan hasil penjumlahan nilai-nilai cuplikan (*sample sum*) dan misalnya diberi tanda S(X).
Bila k merupakan komponen suatu cuplikan eksperimental, maka hubungan antara rata-ratanya, $\overline{X}$, dan jumlah cuplikan S(X) dapat dirumuskan sebagai berikut :

$$k\overline{X} = S(X)$$

3.5. Dalil Batas Tengah.

Andaikan kita melakukan pengambilan cuplikan/sampel dari populasi yang tidak memiliki distribusi normal, bagaimana kiranya bentuk distribusi rata-rata cuplikan pengambilan cuplikan tersebut ?

Secara matematis, distribusi rata-rata cuplikan akan menyerupai distribusi normal jika besarnya cuplikan **n** bertambah secara tak terbatas. Hal sedemikian dinyatakan oleh suatu dalil matematik yang sangat terkenal dan berguna sekali dalam analisis statistika inferensia, yaitu Dalil Batas Tengah (*the central limit theorem*) Jika cuplikan acak dipilih dari populasi dengan rata-rata $\mu_X$ dan varians $\sigma_X^2$ dan jika besarnya cuplikan $n$ bertambah secara tidak terhingga, maka rata-rata cuplikannya akan memiliki distribusi pengambilan cuplikan yang mendekati distribusi normal dengan rata-rata $\mu_{\overline{X}} = \mu_X$ dan deviasi standar $\sigma_{\overline{X}} = \sigma_X / \sqrt{n}$. Jika populasinya terbatas, rata-rata cuplikannya tetap sebesar $\mu_{\overline{X}} = \mu_X$ sedangkan deviasi standarnya menjadi

$$\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Apabila populasinya adalah populasi normal maka distribuasi pengambilan cuplikannya akan normal. Dengan demikian, dalam menentukan distribusi probabilita dari berbagai keadaan tersebut dapat dipergunakan pendekatan normal, di mana rumusnya adalah

$$Z = \frac{\text{Stat} - \text{Par}}{\sigma_{\text{Stat}}}$$

di mana

Z adalah nilai variabel normal standar

Stat adalah statistik, nilai yang berasal dari sampel

Par adalah parameter, nilai yang berasal dari populasi (yang dianggap tertentu)

$\sigma_{\text{stat}}$ adalah deviasi standar statistik

3.6. Distribusi Pencuplikan :

3.6.1. Tentang Rata-rata

Rata-rata hitung populasi biasanya dilambangkan dengan $\mu$, sedangkan rata-rata hitung cuplikan/sampel dilambangkan dengan $\overline{X}$. Jika cuplikan/sampel acak sebanyak **n** dipilih dari suatu populasi yang terbatas dengan pengembalian, akan diperoleh :

Rata-rata hitung cuplikan/sampel

$$\overline{X} = \frac{\sum X_i}{n} \quad \text{atau} \quad \overline{X} = \frac{\sum f_i X_i}{\sum f_i}$$

dan deviasi standar cuplikan sebesar

$$s_x = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}} \quad \text{atau} \quad s_x = \sqrt{\frac{\sum f_i (X_i - \overline{X})^2}{\sum f_i - 1}}$$

Tetapi bila pengambilan cuplikan/sampel tersebut dilakukan dengan tanpa pengembalian, maka akan diperoleh :
Rata-rata hitung tetap sebesar

$$\overline{X} = \frac{\sum X_i}{n}$$

dan deviasi standar cuplikan/sampel menjadi sebesar

$$s_x = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}} \sqrt{\frac{N-n}{N-1}}$$

Contoh :
Ada 9 buah blanko kartu yang kemudian diberi nomor urut mulai dari angka 1 sampai dengan angka 9. Bila setiap kali kita mengambil 3 buah kartu dari 9 kartu tersebut, dan pengambilan dilakukan sebanyak 200 kali, jumlah angka yang mungkin muncul dan frekuensinya secara hipotetis dapat dilihat pada tabel berikut.

| (1) s(X) | (2) $\overline{X}$ | (3) fi | (4) fi/n | (5) Xfi | (6) $f_i(X_i - \overline{X})^2$ |
|---|---|---|---|---|---|
| 6 | 2,00 | 1 | 0,005 | 2,000 | 9,590 |
| 7 | 2,33 | 3 | 0,015 | 6,990 | 22,966 |
| 8 | 2,67 | 4 | 0,020 | 10,680 | 23,557 |
| 9 | 3,00 | 6 | 0,030 | 18,000 | 26,379 |
| 10 | 3,33 | 11 | 0,055 | 36,630 | 34,337 |
| 11 | 3,67 | 13 | 0,065 | 47,710 | 26,465 |
| 12 | 4,00 | 16 | 0,080 | 64,000 | 19,248 |
| 13 | 4,33 | 12 | 0,060 | 51,960 | 7,056 |
| 14 | 4,67 | 24 | 0,120 | 112,080 | 4,372 |
| 15 | 5,00 | 18 | 0,090 | 90,000 | 0,169 |
| 16 | 5,33 | 16 | 0,080 | 85,280 | 0,870 |
| 17 | 5,67 | 13 | 0,065 | 73,710 | 4,271 |
| 18 | 6,00 | 15 | 0,075 | 90,000 | 12,237 |
| 19 | 6,33 | 13 | 0,065 | 82,290 | 19,770 |
| 20 | 6,67 | 12 | 0,060 | 80,040 | 29,699 |
| 21 | 7,00 | 10 | 0,050 | 70,000 | 36,222 |
| 22 | 7,33 | 7 | 0,035 | 51,310 | 34,910 |
| 23 | 7,67 | 4 | 0,020 | 30,680 | 26,485 |
| 24 | 8,00 | 2 | 0,010 | 16,000 | 16,857 |
| Jumlah | | 200 | 1,000 | 1.019,360 | 355,461 |

$$\overline{X} = \frac{1.019,36}{200} = 5,0968 \text{ dan } s_X = \sqrt{\frac{355,461}{199}} = 1,3365$$

3.6.2 Tentang Proporsi

Proporsi populasi dinyatakan dengan $\pi$ = X/N, sedangkan proporsi cuplikan/sampel dinyatakan dengan p = x/n.

Jika cuplikan/sampel acak sebesar **n** dipilih dari populasi binomial dengan pemulihan, cuplikan tersebut akan mempunyai :

Rata-rata hitung $$E(p) = \mu_p = \frac{np}{n} = p$$

dan Varians

$$Var(p) = \sigma_p^2 = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

sehingga deviasi standarnya sebesar

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$

Jika cuplikan/sampel acak tersebut dilakukan tanpa pemulihan, akan mempunyai Rata-rata

$$E(p) = \mu_p = p$$

dan deviasi standar

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Faktor koreksi kontinuitas yang besarnya $\pm \frac{1}{2n}$ perlu diaplikasikan, terutama bila besarnya cuplikan relatif kecil bila dibanding dengan populasinya.

Contoh :

Dari sekali pembakaran produk keramik berupa benda souvenir kecil yang dilakukan oleh seorang pengusaha keramik diperoleh hasil 24.000 barang souvenir tersebut. Dengan sampel acak sebanyak 50 buah keramik, ternyata terdapat 7 buah keramik yang tidak mememnuhi standar karena ada cacat dalam pembakaran. Berapa besar rata-rata dan deviasi standar proporsi produk cacat dari sampel tersebut?

Jawab :

Populasi = N = 24.000
Sampel = n = 50
Produk cacat = x = 7
Jadi proporsi produk cacat dari sampel = x/n = 0,14
Rata-rata produk cacat =

$$E(p) = \mu_p = \frac{np}{n} = \frac{50x0,14}{50} = 0,14$$ atau sebesar 14 %, dan

deviasi standarnya =

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{0,14(1-0,14)}{5}}\sqrt{\frac{24.000-50}{50}}$$

= (0,155177)(0,998979) = 0,155019 atau sebesar 15,50 %.

3.6.3. Tentang Beda antara Dua Rata-rata

Jika dari dua populasi normal diambil dua cuplikan secara acak dan independen yang pertama sebesar $n_1$ dipilih dari populasi pertama dengan rata-rata $\mu_1$ dan deviasi standar $\sigma_1$, sedangkan cuplikan kedua sebesar $n_2$ dipilih dari populasi normal lainnya dengan rata-rata $\mu_2$ dan deviasi standar $\sigma_2$, maka selisih antara kedua rata-rata cuplikan dapat dinyatakan dengan $\Delta\overline{X}$ dan didistribusikan secara normal dengan rata-rata

$$E(\Delta\overline{X}) = \mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$$

dan deviasi standar

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

Jika besaran sampel **n** → ∞ dan rata-rata sampel didistribusikan secara normal, maka rata-rata ($\Delta\overline{X}$) atau $\overline{X}_1 - \overline{X}_2$ juga berdistribusi normal jika $n_1$ dan $n_2$ → ∞. Jadi kita dapat mengubah variabel acaknya dalam bentuk variabel normal standar sebagai berikut :

$$Z = \frac{(\overline{X}_1 - \overline{X}_2)-(\mu_1-\mu_2)}{s_{\overline{X}_1-\overline{X}_2}}$$

Contoh :

Dari data yang terkumpul, lama pakai ban mobil merek A rata-rata 10.000 jam perjalanan dengan dengan varians sebesar 4.000 jam, sedangkan untuk lama pakai ban mobil merek B rata-rata 9.500 jam dengan varians 3.000 jam. Untuk itu diambil samepl ban merek A 50 buah dan merek B 50 buah. Berapa probabilitanya lama pakai ban merek A labih lama 525 jam dari pada ban merek B?

Jawab :

$\mu_1$- $\mu_2$ = 10.000 jam – 9.500 jam = 500 jam

$$s_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} = \sqrt{\frac{4.000}{50}+\frac{3.000}{50}} = 11,832$$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2)-(\mu_1-\mu_2)}{s_{\overline{X}_1-\overline{X}_2}} = \frac{525-500}{11,832} = 2,1129$$

Kemudian, kita cari probabilitanya dari Tabel Luas Daerah di bawah kurve normal standar (Tabel Z) sehingga diperoleh perhitungan sebagai berikut.

$P(\Delta\overline{X} > 525) = P(Z>2,11) = 0,5000 - 0,4826 = 0,0174$

Jadi probabilita lama pakai ban merek A lebih lama dari pada ban merek B lebih dari 525 jam adalah sebesar 0,0174 atau sebesar 1,74 %.

3.6.3. Tentang Beda antara Dua Proporsi

Jika dua cuplikan acak yang independen dipilih dari dua populasi binomial dan bila cuplikan pertama sebasar $n_1$ dipilih dari populasi binomial dengan $p_1$ sedangkan cuplikan kedua sebesar $n_2$ dipilih dari populasi lainnya dengan $p_2$, maka beda antara kedua cuplikan proporsi $\Delta$ p merupakan statistik cuplikan dan didistribusikan dengan rata-rata

$E(\Delta p) = \mu_{p1-p2} = P_1 - P_2$

dan deviasi standar $\sigma_{p_1-p_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

Bila cuplikan yang dipergunakan besar, distribusi pemilihan cuplikan seperti ini akan mendekati distribusi normal, sehingga variabel acaknya dapat diubah menjadi variabel normal standar dengan cara

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{p_1-p_2}}$$

Oleh karena itu probabilitanya dapat dicari dengan menggunakan tabel normal standar atau Tabel Z.

3.6.4. Tentang Varians

Jika sebuah cuplikan acak dengan **n** yang besar sekali dipilih dari populasi yang tertentu, secara umum varians cuplikannya dapat diberikan sebagai $s^2 = \frac{1}{n}\sum(X - \overline{X})^2$ dan memiliki distribusi dengan rata-rata $E(s^2) = \frac{n-1}{n}\sigma^2$

Sedangkan jika jumlah cuplikannya cukup kecil atau kecil, varians cuplikannya adalah $s^2 = \frac{1}{n-1}\sum(X - \overline{X})^2$

Distribusi varians cuplikannya memiliki rata-rata $E(s^2) = \mu_s^2 = \sigma_X^2$

Selisih standar varians populasi, bila populasinya normal, akan didistribusikan dengan rumus $\sigma_{s^2}^2 = \sigma_X^2\sqrt{\frac{2}{n-1}}$

3.6.5. Tentang Median

Jika populasi didistribusikan secara normal dan jika cuplikan acak **n** > 30 dipilih dari populasi di atas, maka median cuplikan akan memiliki distribusi yang mendekati distribusi normal dengan

Rata-rata $= E(md) = \mu_{md} = \mu_X$

Varians $= \sigma^2_{md} = (\pi/2)(\sigma^2_X / n)$

Di mana $\pi = 22/7 \approx 3,14...$

3.7. Soal

1. Berat kantong semen yang diisi oleh mesin A rata-rata 50 kg dengan deviasi standar 5 kg, sedangkan berat kantong semen yang diisi dengan mesin B rata-rata 50,5 kg dengan deviasi standar 5,5 kg. bila diambil sampel untuk mesin A sebanyak 40 dan mesin B 50 buah kantong semen, berapa probabilita kantong semen dari mesin B lebih berat 0,75 kg?

2. Amir dan Anang bermain taruhan melalui pelemparan sekeping uang logam setimbang (*fair coin*). Masing-masing melempar 50 kali. Amir dinyatakan menang bila ia mendapatkan 5 sisi kepala (H=Head) lebih banyak dari hasil yang diperoleh Anang. Berapa probabilita Amir akan menang?

## 3.8. Rangkuman

Penggunaan cuplikan dalam suatu penelitian sering mempunyai beberapa keuntungan antara lain menghemat biaya, waktu, dan tenaga. Apalagi bila penelitian dengan sensus bersifat merusak. Pencuplikan dapat dilakukan dengan pengembalian dan tanpa pengembalian. Besar cuplikan dibaedakan menjadi cuplikan besar dan cuplikan kecil yang dalam beberapa hal membawa konsekuensi adanya perbedaan dalam rumus yang tepat untuk dipergunakan. Batas antara cuplikan besar dan cuplikan kecil ada dua pendapat, yaitu $n \geq 25$ dan $n \geq 30$ untuk masuk dalam kategori cuplikan besar. Bila cuplikan yang digunakan termasuk cuplikan besar, maka distribusinya akan mendekati distribusi normal sehingga probabilitanya dapat dicari dengan bantuan Tabel Z. Jika populasinya bersifat terbatas perlu ditambahkan factor koreksi karena terbatasnya populasi.

# POKOK BAHASAN IV
# PENDUGAAN/ESTIMASI

Daftar Isi :



## 4.1. Pendahuluan

Untuk dapat mempelajari pokok bahasan tentang pendugaan dengan baik, maka anda harus telah benar-benar memahami pokok bahasan tentang distribusi probabilita dan dasar-dasar teori pencuplikan.

Sedangkan pokok bahasan ini pada dasarnya mencakup uraian tentang :

a. Arti dan kegunaan estimasi
b. Ciri-ciri penduga/estimator yang baik
c. Pendugaan/estimasi dengan cuplikan besar
d. Pendugaan/estimasi dengan cuplikan kecil.
e. Besar cuplikan dan ketepatan pendugaan/estimasi.

Guna memberikan pemahaman yang lebih baik di dalam tiap sub-pokok bahasan dipergunakan pola penguraian teori, contoh dan latihan. Contoh soal yang mencakup pokok bahasan ini juga diberikan dan diakhiri dengan adanya rangkuman.

## 4.2. T I U dan Sasaran belajar

T I U :

Setelah mempelajari pokok bahasan ini anda diharapkan mampu menemukenali dan memecahkan masalah-masalah estimasi atau pendugaan pada berbagai situasi dan kondisi

Sasaran Belajar (Sasbel) :

Mahasiswa setelah mempelajari pokok bahasan ini diharapkan dapat :

a. Menjelaskan arti dan kegunaan pendugaan secara statistik
b. Menghitung dan menentukan pendugaan titik terhadap parameter distribusi normal dan distribusi binomial
c. Menjelaskan tiga ciri penduga yang baik.
d. Memahami kapan dan bagaimana menggunakan distribusi t daripada distribusi Z.
e. Menghitung pendugaan/estimasi interval dengan cuplikan besar bagi empat keadaan yang berbeda.
f. Menghitung pendugaan/estimasi interval dengan cuplikan kecil bagi tiga keadaan yang berbeda.
g. Menentukan besar cuplikan berkait dengan ketepatan pendugaan/estimasi.

## 4.3. Arti dan Kegunaan Pendugaan/Estimasi.

Apabila kita melakukan pengumpulan data kuantitatif melalui pencuplikan, sebenarnya kita berharap agar dapat mengetahui gambaran tentang populasi darimana cuplikan tadi diambil. Besaran-besaran (*magnitudes*) atau nilai-nilai yang diperoleh dari cuplikan disebut statistik, sedangkan besaran-besaran yang diperoleh dari populasi disebut parameter. Dengan besaran dari cuplikan, kita akan menarik kesimpulan tentang parameter. Penarikan kesimpulan tersebut dapat berupa pendugaan/estimasi tentang parameter atau berupa penolakan /penerimaan suatu hipotesis. Besaran atau nilai cuplikan yang digunakan untuk menduga atau mengestimasi parameter populasi disebut penduga atau estimator.

Contoh: $\overline{X}$ merupakan estimator bagi $\mu$ (rata-rata hitung populasi).
s merupakan estimator bagi $\sigma$ (deviasi standar populasi).
p merupakan estimator bagi P.(proporsi populasi)

Seluruh proses penggunaan penduga/estimator untuk menghasilkan pendugaan terhadap parameter populasi disebut pendugaan/estimasi. Suatu nilai tunggal digunakan untuk menduga satu parameter populasi disebut penduga/estimator titik dan proses pendugaannya disebut pendugaan/estimasi titik. Misalnya dari nilai Rp 800.000,00 yang menunjukkan rata-rata GDP per capita/tahun dari cuplikan digunakan untuk menyatakan bahwa rata-rata GDP per capita/tahun dari Indonesia diduga sebesar Rp 800.000,00.. Ketepatan pendugaan/estimasi titik bisa tinggi sekali kalau memang pendugaan/estimasi kita tadi benar, akan tetapi kemungkinan atau probabilita untuk menduga/mengestimasi dengan benar tersebut biasanya sangat kecil. Hal itu disebabkan karena pendugaan/estimator titik hanyalah merupakan satu nilai dari skala nilai yang mungkin terjadi. Maka dari itu, untuk menduga/mengestimasi satu parameter populasi digunakan suatu jangka (interval) nilai.

Suatu jangka nilai yang digunakan untuk menduga/mengestimasi satu parameter populasi disebut penduga/estimator interval yang juga merupakan interval kepercayaan (*confidence interval*) dan proses pendugaannya disebut estimasi interval. Dalam pembahasan selanjutnya kita hanya akan membahas tentang pendugaan/estimasi interval.

Latihan :

Berikan rumus dan contoh perhitungannya untuk pendugaan/estimasi titik bagi :

a. Rata-rata hitung populasi, $\mu$,
b. Proporsi Populasi, P, dan
c. Deviasi standar populasi, $\sigma$.

## 4.4. Ciri-ciri Penduga/Estimator yang baik

Agar supaya suatu penduga/estimator dapat menjadi penduga/estimator yang baik, dalam arti bahwa penduga/estimator tersebut tidak berbeda dari apa yang diduga (parameter), ada 3 ciri yang harus dipenuhi yaitu : tidak bias, efisien dan konsisten.

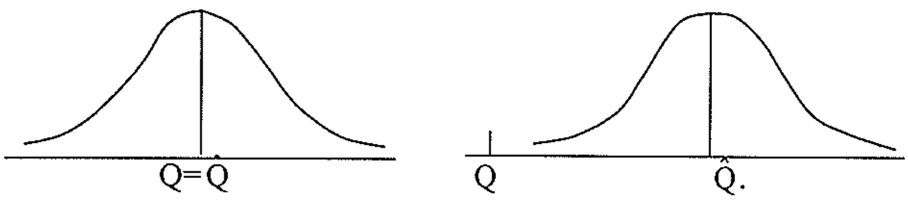
a. Tidak bias.

Suatu penduga/estimator dikatakan tidak bias (*unbiased*) apabila yang diduga tidak berbeda dengan penduganya. $\hat{Q}$ merupakan penduga/estimator tidak bias bagi Q jika $E(\hat{Q}) = Q$.

Dalam diagram berikut dapat dilihat perbedaan antara penduga/estimator yang bias dengan yang tidak bias.

Gambar 4.1

(a) Penduga/Estimator yang tidak bias (b) Penduga/Estimator yang bias




Suatu penduga/estimator dianggap bias jika $E(\hat{Q}) \neq Q$ dan besarnya bias dapat dirumuskan sebagai :

$$\text{Bias} = E(\hat{Q}) - Q \qquad \rightarrow \text{RUMUS : 4-1}$$

Contoh penduga/estimator yang tidak bias : $\overline{X}$ dan Md cuplikan/sampel bagi $\mu$ (rata-rata hitung populasi).

Contoh penduga/estimator yang bias : s yang dihitung dengan pembagi atau penyebut n, bagi $\sigma$ (deviasi standar populasi).

b. Efisien.

Penduga/estimator dikatakan merupakan penduga/estimator yang makin efisien jika varians dari penduga/estimator tersebut makin kecil. Hal ini berarti bahwa distribusinya makin terkonsentrasi di sekitar nilai penduga/estimator tersebut. Efisiensi relatif penduga/estimator A jika dibandingkan dengan penduga/estimator B adalah sebesar

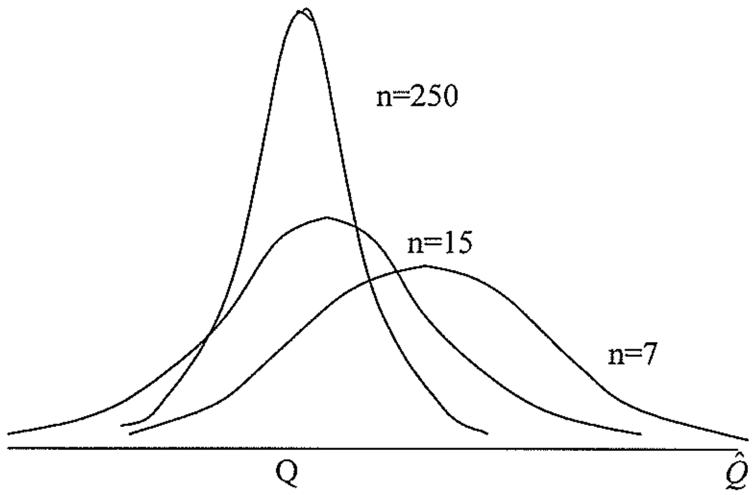
$$\frac{\text{Varians A}}{\text{Varians B}}$$

Dilihat dari efisiensinya, maka $\overline{X}$ merupakan pendugaestimator yang lebih efisien bila dibandingkan dengan Md cuplikan/sampel. Coba cari pembuktiannya, misal dalam bukunya Anto Dayan.

c. Konsisten.

Secara kasar, penduga/estimator yang konsisten merupakan penduga/estimator yang berkonsentrasi secara sempurna pada parameter jika besarnya cuplikan/sampel ditambah secara tidak berhingga. Penyebaran dari distribusi statistik akan makin mengecil dengan semakin besarnya cuplikan/sampel yang digunakan. Gambaran tentang suatu penduga/estimator yang konsisten dapat dilihat dalam gambar 4.2. berikut ini.

Gambar 4.2.




## 4.5. Pendugaan/Estimasi dengan Cuplikan Besar

Pada sub pokok bahasan pendugaan/estimasi dengan cuplikan besar akan diuraikan pendugaan/estimasi terhadap $\mu$, $\mu_1 - \mu_2$, P, dan $P_1 - P_2$.
Rumus dasar pendugaan/estimasi dengan cuplikan besar adalah :

$$\text{Parameter} = \text{Statistik} \pm Z_{\alpha/2}\,\sigma_{\text{Statistik}} \qquad \rightarrow \text{ RUMUS : 4-2}$$

a. Pendugaan/Estimasi terhadap $\mu$

Dalam pendugaan/estimasi terhadap $\mu$, rumus dasar tersebut berubah menjadi

$$\mu = X \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \rightarrow \text{ RUMUS 4-3}$$

dimana $\alpha$ menunjukkan tingkat signifikansi atau tingkat keyakinan yang diinginkan. Bila populasinya terbatas maka perlu diberi faktor koreksi yang besarnya

$$\sqrt{\frac{N-n}{N-1}} \qquad \rightarrow \text{ RUMUS 4-4}$$

Sedangkan bila $\sigma$ tidak diketahui, maka digunakan deviasi standar cuplikan yaitu s. Oleh karena itu ada 4 alternatif rumus untuk melakukan estimasi terhadap $\mu$ dengan sampel besar pada situasi dan kondisi yang berbeda.

a1. Estimasi terhadap $\mu$, dengan $n \geq 25$, $\sigma$ diketahui, dan $N = \infty$ (N tidak terbatas)

$$\mu = X \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \rightarrow \text{RUMUS 4-3}$$

a2. Estimasi terhadap $\mu$, dengan $n \geq 25$, $\sigma$ diketahui, dan $N \neq \infty$ (N terbatas)

$$\mu = X \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad \rightarrow \text{RUMUS 4-5}$$

a3. Estimasi terhadap $\mu$, dengan $n \geq 25$, $\sigma$ tidak diketahui, dan $N = \infty$

$$\mu = X \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} \quad \rightarrow \text{RUMUS 4-6}$$

a4. Estimasi terhadap $\mu$, dengan $n \geq 25$, $\sigma$ tidak diketahui, dan $N \neq \infty$

$$\mu = X \pm Z_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad \rightarrow \text{RUMUS 4-7}$$

Contoh :

Tiga puluh enam nilai diambil secara acak dari sekelompok nilai yang jumlahnya besar sekali dan mempunyai deviasi standar sebesar 12. Jika ketiga puluh enam nilai tadi mempunyai rata-rata sebesar 58, dapatkan interval keyakinan 95 % guna menduga rata-rata nilai dari seluruh anggota kelompok.

Jawab : Data yang diketahui $n = 36$, $\sigma = 12$ dan $\overline{X} = 58$.
Diminta membuat interval keyakinan 95 %, berarti $\alpha = 5$ %.
Interval keyakinan 95 % bagi $\mu$ adalah

$$\mu = \overline{X} \pm Z_{0,025} \frac{\sigma}{\sqrt{n}} \quad \rightarrow \text{RUMUS 4-3}$$

$$= 58 \pm 1{,}96 \frac{12}{\sqrt{36}}$$

$$= 58 \pm 3{,}92$$

Jadi interval keyakinan 95% tersebut adalah

$$54{,}08 \leq \mu \leq 61{,}92$$

b. Estimasi terhadap $\mu_1$ - $\mu_2$

Rumus estimasi terhadap selisih antara dua rata-rata hitung populasi adalah :

$$\mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$ → RUMUS : 4-8

Bila deviasi standar populasi bagi kedua populasi sama, maka rumus di atas dapat diubah menjadi

$$\mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$ → RUMUS : 4-9

Sedangkan bila deviasi standar populasi tidak diketahui, kita harus menggunakan deviasi standar cuplikan. Dalam keadaan ini kita bisa menggunakan asumsi :

$\sigma_1 = \sigma_2$ atau $\sigma_1 \neq \sigma_2$

Dengan asumsi bahwa deviasi standar kedua populasi sama, maka kita harus mencari deviasi standar cuplikan gabungan, sp, yang dapat dicari dengan rumus

$$\mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm Z_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$ → RUMUS :4-10

di mana

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$ → RUMUS 4-11a

atau

$$s_p = \sqrt{\frac{\sum (X_{i1} - \overline{X}_1)^2 + \sum (X_{i2} - \overline{X}_2)^2}{n_1 + n_2 - 2}}$$ → RUMUS 4-11b

Dengan asumsi bahwa deviasi standar kedua populasi tidak diketahui dan tidak sama, maka kita bisa langsung menggunakan RUMUS : 4-9 dengan mengganti $\sigma_1$ dan $\sigma_2$ menjadi $s_1$ dan $s_2$.

Contoh :

Seorang importir menerima kiriman 2 macam lampu mercuri yang bermerk A dan B dalam jumlah yang besar sekali. Importir memilih dari lampu mercuri merk A sebanyak 50 lampu secara acak sebagai sampel, dari lampu mercuri merk B juga diambil 50 lampu secara acak, serta menguji daya tahan lampu-lampu tersebut secara cermat sekali. Ternyata, lampu merk A memiliki daya tahan rata-rata 1.282 jam dan untuk lampu B daya tahan rata-ratanya 1.208 jam. Dari pengalaman importir tersebut, deviasi standar daya tahan dari seluruh lampu A dan B secara berturut-turut sebesar 80 dan 94 jam. Buatlah pendugaan tentang beda rata-rata daya tahan kedua macam lampu mercuri tersebut dengan menggunakan interval keyakinan 95 %.

Jawab : Data yang diketahui adalah

$$n_A = 50 \qquad n_B = 50$$
$$\overline{X}_A = 1.282 \qquad \overline{X}_B = 1.208$$
$$\sigma_A = 80 \qquad \sigma_B = 94$$
$$\alpha = 5\ \%$$

Dengan $\alpha$ = 5 %, berarti yang ditanyakan adalah interval keyakinan 95 % bagi $\mu_A$-$\mu_B$ yaitu

$$\mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm Z_{0,025}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad \rightarrow \text{RUMUS 4-8}$$

$$\mu_1 - \mu_2 = (1.282 - 1.208) \pm 1{,}96\sqrt{\frac{80^2}{50} + \frac{94^2}{50}}$$
$$= 74 \pm 33{,}95$$

Jadi interval keyakinan 95 % bagi beda rata-rata daya tahan lampu A dan B adalah

$$40{,}05 \leq (\mu_1 - \mu_2) \leq 107{,}95$$

Latihan :
Dari suatu survei tentang rata-rata pendapatan keluarga dari dua kota A dan B diperoleh catatan sebagai berikut :

| Perihal | Kota A | Kota B |
|---|---|---|
| Besar cuplikan | 100 | 120 |
| $\overline{X}$ | 5.900,- | 5.800,- |
| Varians | 9.050,- | 8.700,- |

Berapa beda rata-rata pendapatan antara kedua kota tersebut?
Jelaskan makna hitungan Saudara.

c. Estimasi terhadap P
Dalam melakukan estimasi terhadap parameter yang berupa P yaitu proporsi populasi dengan p yaitu proporsi cuplikan, ada rumus umum sebagai berikut :

$$P = p \pm Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}} \qquad \rightarrow \text{RUMUS : 4-12}$$

Rumus ini digunakan apabila populasinya tidak terbatas, dan proporsi populasi diketahui. Pada umumnya proporsi populasi tersebut tidak diketahui, untuk itu dalam rumus estimasi di atas P dapat diganti dengan : (a) proporsi cuplikan, p = x/n atau (b) proporsi yang dapat memaksimumkan nilai P(1-P), yaitu p = ½ atau 0,5. Oleh karena itu RUMUS 4-12 dapat menjadi dua alternatif rumus lagi, yaitu

$$P = p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$ → RUMUS 4-13a

$$P = p \pm Z_{\alpha/2}\sqrt{\frac{(0,5)(1-0,5)}{n}}$$ → RUMUS 4-13b

Contoh :

Seperempat dari 300 orang konsumen yang diwawancarai menyatakan tidak menyukai sabun mandi cap "D". Tentukan interval keyakinan sebesar 99 % guna menduga proporsi populasi konsumen yang tidak menyukai sabun cap "D" tersebut. Gunakan pendekatan secara normal.

Dari soal di atas dapat dilakukan pemecahan sebagai berikut:

Diketahui : n = 300, x = 1/4 x 300 = 75

p = x/n = 0,25     P tidak diketahui, digunakan p (proporsi sampel)

N tidak terbatas

Diminta : Interval keyakinan 99 % bagi P

Jawab : Interval keyakinan 99 % bagi P adalah

$$P = p \pm Z_{0,005}\sqrt{\frac{p(1-p)}{n}}$$ → RUMUS 4-13a

$$= 0,25 \pm 1,96\sqrt{\frac{0,25(1-0,25)}{300}}$$

$= 0,25 \pm 0,0037$

Jadi interval keyakinan 99 % bagi P adalah

$0,2463 < P < 0,2537$

Jika kita ingin lebih teliti dalam estimasiterhadap P ini, kita bisa menambahkan faktor koreksi kontinuitas yang besarnya 1/2n dan bersifat memperlebar interval keyakinan. Oleh karena itu, rumus estimasinya menjadi :

$$P = p \pm \frac{1}{2n} \pm Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$ → RUMUS 4-14

atau

$$P = p \pm \frac{1}{2n} \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$ → RUMUS 4-14a

atau

$$P = p \pm \frac{1}{2n} \pm Z_{\alpha/2}\sqrt{\frac{(0,5)(1-0,5)}{n}}$$ → RUMUS 4-14b

Latihan :

Kerjakan soal contoh tersebut dengan mengetrapkan faktor koreksi kontinuitas.

d. Estimasi terhadap $P_1$ - $P_2$

Dengan cuplikan besar kita dapat melakukan estimasi terhadap selisih antara dua proporsi populasi, dan rumus umumnya adalah

$$P_1 - P_2 = (p_1 - p_2) \pm Z_{\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}} \quad \rightarrow \text{ RUMUS : 4-15}$$

Contoh :

Dua mesin digunakan untuk menghasilkan barang yang sama. Dari 400 unit barang yang dihasilkan mesin A ternyata ada 16 unit kurang memenuhi kualitas standar. Dalam periode yang sama, mesin B menghasilkan 600 unit barang dan dari jumlah tersebut ada 60 yang kurang memenuhi standar.

Buatlah interval keyakinan 95 % bagi perbedaan proporsi kerusakan barang yang dihasilkan mesin A dan mesin B.

Jawab :

Dari soal tersebut, hal-hal yang diketahui adalah :

$n_A = 400$ $\quad$ $n_B = 600$

$x_A = 16$ $\quad$ $x_B = 60$

$p_A = 16/400 = 0{,}025$ $\quad$ $p_B = 60/600 = 0{,}10$

$\alpha = 5$ %

P tidak diketahui, sehingga dalam menghitung deviasi standar bisa digunakan p cuplikan atau nilai p yang memaksimumkan P(1-P), yaitu p = 0,5.

Diminta : Interval keyakinan 95 % bagi $P_A$ - $P_B$

Jawab :

Interval keyakinan tersebut dapat diperoleh melalui modifikasi RUMUS 4-15, menjadi

$$P_A - P_B = (p_A - p_B) \pm Z_{\alpha/2}\sqrt{\frac{p_A(1-p_B)}{n_A} + \frac{p_B(1-p_B)}{n_B}} \quad \rightarrow \text{ RUMUS 4-15a}$$

$$= (0{,}025 - 0{,}100) \pm 1{,}96\sqrt{\frac{(0{,}025)(1-0{,}025)}{400} + \frac{(0{,}100)(1-0{,}100)}{600}}$$

$$= -\,0{,}075 \pm 0{,}01545$$

Jadi interval keyakinan 95 % bagi $P_A$ - $P_B$ adalah

$$-\,0{,}09045 < (P_A - P_B) < -\,0{,}05955$$

Latihan :

Kerjakan soal di atas dengan menggunakan nilai p yang memaksimumkan P(1-P).

## 4.6. Estimasi dengan Cuplikan Kecil

Estimasi dengan cuplikan kecil pada umumnya digunakan untuk mengestimasi $\mu$, $\mu_1$- $\mu_2$ dan P.Dalam estimasi dengan cuplikan kecil distribusi Z tidak bisa digunakan, dan untuk itu digunakan distribusi t. Dalam distribusi t, yang perlu diperhatikan adalah $\alpha$ dan d.f (*degree of freedom*) atau derajat kebebasan.

a. Estimasi terhadap $\mu$.

Dari rumus dasar estimasi, dalam estimasi dengan cuplikan kecil terhadap $\mu$, rumus tersebut berubah menjadi :

$$\mu = \overline{X} \pm t_{\alpha/2;d.f} \frac{\sigma}{\sqrt{n}} \qquad \rightarrow \text{ RUMUS : 4-16}$$

Rumus ini dipakai bila populasinya tidak terbatas dan deviasi standar populasi diketahui. Sedangkan bila populasinya terbatas perlu ditambahkan faktor koreksi karena terbatasnya N. Bila deviasi standar populasi tidak diketahui, dipakai deviasi standar cuplikan.

Contoh : Dari suatu penelitian dengan cuolikan acak sebesar 25 responden yang terdiri dari para SPG (*sales promotion girl*), diperoleh hasil pngolahan data setengah jadi sebagai berikut. Rata-rata penjualan sebesar 46 unit barang. Dari pengalaman sebelumnya dan masih dapat dipercaya, deviasi standar penjualan oleh seluruh SPG tersebut sebesar 14 unit. Buatlah pendugaan interval dengan tingkat kepercayaan 95% bagi rata-rata penjualan seluruh SPG.

Yang diketahui n = 25, termasuk sampel kecil, $\overline{X} = 46$, $\sigma = 14$. $N = \infty$

Ditanya : Interval Keyakinan (*Confidence Interval, CI*) 95 % bagi $\mu$.

Jawab :

CI 95 % bagi $\mu$. Dapat diperoleh dengan rumus 4-16,

$$\mu = \overline{X} \pm t_{0,025;d24} \frac{\sigma}{\sqrt{n}} = 46 \pm 2{,}064 \frac{14}{\sqrt{25}}$$

$$= 46 \pm 5{,}7792$$

Jadi CI 95 % bagi $\mu$ adalah $40{,}2208 < \mu < 51{,}7792$.

b. Estimasi terhadap $\mu_1$-$\mu_2$.

Seperti halnya dengan pendugaan/estimasi terhadap beda antara dua $\mu$ dengan cuplikan besar, maka rumus dasar pendugaan/estimasi berubah menjadi :

$$\mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2;d.f} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad \rightarrow \text{ RUMUS 4-17}$$

Juga analog dengan pendugaan/estimasi dengan cuplikan besar, dalam hal deviasi standar populasi tidak diketahui, perlu digunakan deviasi standar cuplikan. Ada dua keadaan yang bisa terjadi, yaitu :

- Deviasi standar populasi dianggap sama
- Deviasi standar populasi dianggap tidak sama

Bila deviasi standar populasi yang tidak diketahui tersebut dianggap sama, maka perlu dicari deviasi standar gabungan dengan rumus seperti RUMUS 4.11a atau 4-11b. Sedangkan d.f nya adalah sebesar n1 + n2 - 2. Oleh karena itu rumus pendugaan/estimasi yang relevan menjadi

$$\mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2;d.f} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \qquad \rightarrow \text{ RUMUS 4.18}$$

Bila deviasi standar populasi dianggap tidak sama, maka $\sigma_1$ dan $\sigma_2$ langsung disubstitusi dengan $s_1$ dan $s_2$ dalam RUMUS : 4.17 tersebut, sehingga rumus estimasinya adalah

$$\mu_1 - \mu_2 = (\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2;d.f}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad \rightarrow \text{RUMUS 4-19}$$

Sedangkan d.f nya agak lebih sukar ditentukan, yaitu misalnya bisa diperoleh dari rumus :

$$\text{d.f.} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1}{n_1}\right)^2}{n_1} + \frac{\left(\frac{s_2}{n_2}\right)^2}{n_2}} - 2 \qquad \rightarrow \text{RUMUS 4-20}$$

Contoh : Dari 20 karyawan sampel PT A rata-rata dapat menyelesaikan pembungkusan yang cukup rumit dalam waktu 12 menit dengan deviasi standar 3 menit. Dengan tugas pembungkusan sejenis, 15 karyawan sample dari PT B dapat menyelesaikan tugas tersebut rata-rata dalam waktu 10 menit dan deviasi standarnya 2 menit. Susunlah interval keyakinan 98 % bagi beda antara dua rata-rata penyelesaian tugas pembungkusan tersebut.

Yang diketahui $\overline{X}_A = 12$ $\quad s_A = 3$ $\quad n_A = 20$

$\overline{X}_B = 10$ $\quad s_A = 2$ $\quad n_B = 15$

$\sigma_A$ dan $\sigma_B$ tidak diketahui, diasumsikan $\sigma_A \neq \sigma_B$

$\alpha$ tidak diketahui, misal dipilih $\alpha = 5\%$

Ditanyakan CI 95 % bagi $\mu_A$ - $\mu_B$

Jawab CI 95 % bagi $\mu_A$ - $\mu_B$ adalah

$$\mu_A - \mu_B = (\overline{X}_A - \overline{X}_B) \pm t_{0,025;d.f}\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

d.f dapat diperoleh dengan rumus

$$\frac{\left(\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}\right)^2}{\frac{\left(\frac{s_A^2}{n_{A1}}\right)^2}{n_A} + \frac{\left(\frac{s_B^2}{n_B}\right)^2}{n_B}} - 2 = \frac{\left(\frac{9}{20} + \frac{4}{15}\right)^2}{\frac{\left(\frac{9}{20}\right)^2}{20} + \frac{\left(\frac{4}{15}\right)^2}{15}} - 2 = 32{,}55$$

d.f = 33 (dibulatkan)

$$\mu_A - \mu_B = (\overline{X}_A - \overline{X}_B) \pm t_{0,025;33}\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$$

$$= (12-10) \pm 2{,}064\sqrt{\frac{9}{20} + \frac{4}{15}} =$$

$$= 2 \pm 1{,}747303$$

Jadi CI 95 % bagi $\mu_A$ - $\mu_B$ adalah

$\underline{0{,}025 < (\mu_A - \mu_B) < 3{,}75}$

c. Estimasi terhadap P.

Walaupun pada umumnya estimasi terhadap P dilakukan dengan cuplikan besar, estimasi dengan cuplikan kecil juga dapat digunakan dalam estimasi terhadap P asalkan dipenuhi syarat sebagai berikut : np > 5 bagi p < 1/2, atau n(1-p) > 5 bagi p > 1/2. dimana : n adalah besar cuplikan.

Adapun rumus estimasi terhadap P dengan cuplikan kecil adalah :

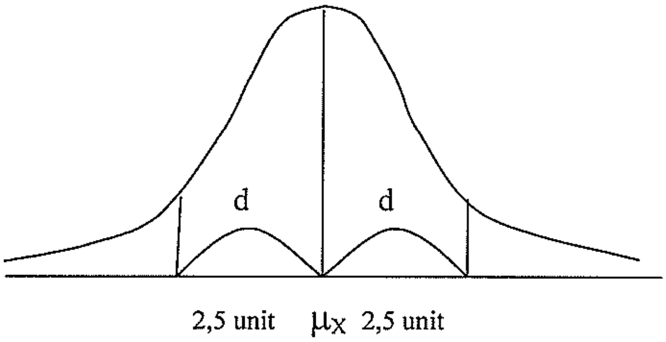
$$P = p \pm t_{\alpha/2;d.f}\sqrt{\frac{P(1-P)}{n}} \qquad \rightarrow \text{ RUMUS 4.21}$$

atau bila ingin lebih teliti bisa ditambahkan faktor koreksi kontinuitas seperti halnya dengan estimasi cuplikan besar, sehingga rumusnya menjadi

$$P = p \pm \frac{1}{2n} \pm t_{\alpha/2;d.f}\sqrt{\frac{P(1-P)}{n}} \qquad \rightarrow \text{ RUMUS 4-22}$$

## 4.7. Besar Cuplikan/Sampel dan Ketepatan Estimasi

Pada umumnya, sebelum kita melakukan pengambilan cuplikan secara acak dalam pendugaan parameter populasi, kita seharusnya ingin mengetahui berapa besar cuplikan acak sedemikian itu yang harus digunakan agar kita dapat menduga parameter tersebut dengan ketelitian yang kita inginkan dan yang diukur atas dasar lebar interval keyakinannya. Misalnya, kita ingin mengetahui besarnya cuplikan acak yang harus dipergunakan agar kita 95 % yakin bahwa selisih atau kesalahan rata-rata cuplikan dari rata-rata populasi yang sesungguhnya tidak akan lebih dari 2,5 unit. Dengan lain perkataan, interval keyakinan 95 % seharusnya tidak melebihi 5 unit lebarnya dari rata-rata secara keseluruhan. Keadaan tersebut dapat dijelaskan melalui diagram berikut :



Telah kita ketahui bahwa bahwa pada $\alpha$ tertentu, lebar interval keyakinan akan tergantung pada varians dan besarnya cuplikan/sampel acak. Jika varians diketahui, setengah lebar interval keyakinan dapat dirumuskan sebagai

$$n = \frac{Z^2\sigma^2}{d^2}$$

dimana :
Z adalah nilai Z-tabel pada $\alpha$ tertentu,
d adalah selisih atau kesalahan yang direncanakan,
$\sigma$ adalah deviasi standar

Contoh : Jika deviasi standar populasi normal diketahui sebesar 10 unit dan jika kita ingin interval keyakinan sebesar 95% yang mencakup rata-rata parameter tidak melebihi 5 unit lebarnya, berapakah besarnya cuplikan yang harus digunakan ?
Jawab :
Dketahui : $d = 2{,}5$ $\sigma = 10$ dan $Z_{0{,}025} = 1{,}96$
Jadi besarnya cuplikan yang harus digunakan adalah

$$n = \frac{Z^2\sigma^2}{d^2} = \frac{(1{,}96)^2(10)^2}{(2{,}5)^2}$$

$= 61{,}4656$ dibulatkan menjadi 61

## 4.8. Soal

1. Sebauh cuplikan terdiri dari 200 kendaraan bermotor masing-masing telah dipilih dari populasi yang terdiri dari kendaraan bermotor di dua kota A dan B. Di kota A, 144 buah kendaraan ternyata telah melunasi pajak kendaraan sedangkan di kota B hanya 128 buah kendaraan yang pajaknya telah lunas.
   a. Buatlah interval keyakinan 95 % guna menduga proporsi kendaraan yang telah lunas pajaknya di kota A.
   b. Buatlah interval keyakinan 95 % guna menduga proporsi kendaraan yang telah lunas pajaknya di kota B.
   c. Buatlah interval keyakinan 95 % guna menduga beda proporsi kendaraan yang telah lunas pajaknya antara kota A dan B.

2. Survei tentang rata-rata pendapatan keluarga dari dua kota Y dan Z menghasilkan catatan sebagai berikut :

| | Kota Y | Kota Z |
|---|---|---|
| Besar cuplikan (n) | 100 | 120 |
| Rata-rata ($\overline{X}$) | \$ 5.800,00 | \$ 5.900,00 |
| Varians cuplikan ($s^2$) | 9.050,00 | 8.700,00 |

   a. Buatlah pendugaan interval bagi rata-rata pendapatan di kota Y.
   b. Buatlah pendugaan interval bagi rata-rata pendapatan di kota Z.
   c. Buatlah pendugaan interval bagi beda rata-rata pendapatan antara kedua kota Y dan Z tersebut.

   Gunakan $\alpha = 5$ %.

## 4.9. Rangkuman

Pendugaan/estimasi pada dasarnya dilakukan untuk menduga atau menaksir nilai parameter populasi yang biasanya tidak diketahui dengan nilai dari statistik cuplikan yang sesuai.. Pendugaan/estimasi tersebut dapat dilakukan pada berbagai keadaan, misalnya : Pendugaan/estimasi terhadap $\mu$, $\mu_1 - \mu_2$, P, $P_1 - P_2$.
Pendugaan/estimasi tersebut berupa estimasi titik atau estimasi interval.
Pendugaan/estimasi dilakukan dengan n yang besar atau yang kecil.
Pendugaan/estimasi dilakukan pada N yang terbatas atau tidak terbatas.
Pendugaan/estimasi dilakukan dalam keadaan deviasi standar populasi diketahui atau tidak diketahui.
Karena ada berbagai keadaan dalam melakukan pendugaan/estimasi, maka rumus pendugaan/estimasi tersebut juga bermacam-macam. Namun, rumus umum dalam pendugaan/estimasi interval adalah :

$$\text{Parameter} = \text{Statistik} \pm Z_{\alpha/2}\,\sigma_{\text{Statistik}}$$

# POKOK BAHASAN V
# PENGUJIAN HIPOTESIS

## 5.1. Pendahuluan.

Pokok bahasan ini merupakan salah satu pokok bahasan terpenting dalam Statistika Ekonomi Lanjutan. Untuk dapat mempelajarinya dengan baik, pengetahuan tentang pendugaan/estimasi dan distribusi probabilita sangat diperlukan. Dalam pokok bahasan V ini baru dibicarakan segi teori dari pengujian hipotesis, sedangkan pembicaraan secara rinci untuk masing-masing jenis pengujian akan dibicarakan dalam pokok bahasan VI s.d. IX.
Adapun isi pokok bahasan ini mencakup :

a. Arti dan kegunaan pengujian hipotesis
b. Kesalahan Tipe I dan Tipe II
c. ,Langkah-langkah pengujian hipotesis.

## 5.2. T I U dan Sasbel.

T I U :
Setelah mempelajari pokok bahasan ini mahasiswa mampu memahami arti pengujian hipotesis dan langkah-langkah atau proses pengujiannya.

Sasbel :
Mahasiswa setelah mempelajari pokok bahasan ini diharapkan dapat :

a. Menjelaskan dengan kata-kata sendiri arti pengujian hipotesis
b. Menjelaskan secara ringkas paling sedikit empat macam penggunaan pengujian hipotesis.
c. Menjelaskan arti Ho dan Ha.
d. Memberikan paling sedikit tiga contoh perumusan Ho dan Ha.
e. Menjelaskan arti kesalahan Tipe I dan Tipe II
f. Menjelaskan sembilan langkah yang diperlukan dalam pengujian hipotesis.

## 5.3. Arti dan Kegunaan.

Sebagian besar penelitian menggunakan hipotesis. Salah satu alasannya adalah agar supaya penelitian tersebut lebih terarah lebih terfokus pada pemecahan masalah penelitian. Hipotesis adalah jawaban sementara sebagai pemecahan terhadap suatu masalah. Dikatakan sementara, karena masih menunggu hasil pengujian terhadap jawaban tersebut. Pengujian hipotesis dapat dianggap sebagai suatu prosedur guna menentukan apakah hipotesis tersebut diterima atau ditolak. Dilihat dari bentuk penyajiannya, hipotesis dapat dibedakan menjadi :

Ho : Hipotesis nol, hipotesis penguji
Ha : Hipotesis alternatif, hipotesis kerja.

Dalam statistika, kedua hipotesis tersebut selalu harus dirumuskan dan dinyatakan. Hipotesis nol (Ho) merupakan hipotesis yang akan diuji dan yang nantinya, sebagai hasil uji hipotesis, akan diterima atau ditolak. Sedangkan hipotesis alternative (Ha) merupakan komplemen hipotesis nol, dalam arti kalau Ho diterima berarti Ha ditolak, dan sebaliknya. Oleh karena itu dalam penolakan atau penerimaan hipotesis harus ditegaskan yang diterima atau ditolak itu Ho atau Ha nya.

Melalui ungkapan yang mudah dipahami, rumusan umum dari Ho misalnya dapat dicontohkan dengan

"Ini" sama dengan "Itu",
Tidak ada perbedaan antara "Ini" dan "Itu", atau
Tidak ada hubungan antara "Ini" dengan "Itu".
Tidak ada pengaruh dari "Ini" terhadap "Itu."

Sedangkan rumusan Ha yang merupakan komplemen bagi Ho misalnya :

"Ini" tidak sama dengan "Itu."
"Ini" lebih kecil dari "Itu."
"Ini" lebih besar dari "Itu."
Ada perbedaan antara "Ini" dan "Itu," atau
Ada hubungan antara "Ini" dengan "Itu," atau
Ada pengaruh dari "Ini" terhadap "Itu.".

Perumusan Ha tersebut sesuai dengan arah uji yang dikehendaki, yaitu uji satu arah atau uji dua arah. Contoh rumusan Ha yang pertama menunjukkan uji dua arah dan contoh rumusan Ha kedua dan ketiga menunjukkan uji satu arah.

Penentuan arah uji biasanya tergantung pada tujuan pengujian hipotesis, untuk melihat adanya perubahan yang terjadi setelah adanya perlakuan (*treatment*) tertentu umumnya digunakan uji satu arah. Sedangkan bila kita hanya sekadar ingin melihat ada tidaknya perbedaan antara dua kadaan, padat digunakan uji dua arah. Kalau di dalam soal ujian, biasanya ada tersurat atau tersirat dari perintah yang ada. Misal bila pertanyaannya berbunyi apakah $\mu_1 > \mu_2$, apakah vaksin itu efektif meningkatkan daya tahan tubuh terhadap suatu penyakit, maka kita diminta untuk melakukan uji satu arah.

Pengujian hipotesis biasanya memerlukan observasi cuplikan (sampel), yang dapat dilakukan secara berulang kali atau hanya sekali saja. Atas dasar nilai statistik sampel, keputusan diambil guna menentukan apakah Ho diterima atau ditolak. Dalam hal ini, salah satu tahap prosedur yang terpenting ialah menentukan statistik cuplikan (sampel) yang dianggap sebagai dasar guna menerima atau menolak Ho. Secara umum, interval keyakinan (*confidence interval*) merupakan set penerimaan Ho. Kita juga cenderung dapat menolak Ho dengan cara mengubah besarnya $\alpha$ yang kita pakai.

Latihan : Buatlah tiga buah rumusan Ho beserta rumusan Ha yang relevan. Berikan alasan mengapa rumusan Ha yang anda pilih seperti itu.

## 5.4. Kesalahan Tipe I dan Tipe II.

Dalam setiap proses pengambilan keputusan tentang apakah kita harus menerima atau menolak Ho, kita selalu dihadapkan pada 2 macam kesalahan pengambilan keputusan yang berbeda. Kesalahan-kesalahan tersebut dalam statistika biasanya disebut kesalahan Tipe I dan kesalahan Tipe II. Secara skematis kemungkinan hasil pengujian hipotesis dan kesalahan Tipe I dan Tipe II dapat dilihat dalam Tabel 5.1. berikut. Dari tabel tersebut kita juga dapat memberikan definisi bagi tingkat signifikansi ($\alpha$) yaitu besarnya kesediaan / toleransi / keberanian kita (peneliti) untuk menerima kesalahan Tipe I atau kesalahan karena kita menolak Ho yang benar. Sedangkan $\beta$ menunjukkan kesediaan kita untuk menerima Ho yang tidak benar, dan (1 - $\beta$) disebut *power* atau kuasa dari pengujian. Walaupun tidak secara proporsional, semakin tinggi tingkat signifikansi atau semakin kecil $\alpha$, maka $\beta$ akan semakin besar, dan sebaliknya. Untuk memahami tentang hubungan antara $\alpha$ dan $\beta$ misalnya anda dapat membaca Bab XIII bukunya Anto Dayan Jilid II. Secara teoritis kedua jenis kesalahan tersebut harus diusahakan sekecil mungkin melalui pemilihan daerah kritis yang setepat-tepatnya.

Tabel 5.1. *Possible results of Hypothesis Testing*
(Wonnacott & Wonnacott)

| *State of the World* | *Decision* | |
|---|---|---|
| | *Accept Ho* | *Reject Ho* |
| *If Ho is True* | *Correct decision*<br>*Prob = 1 - $\alpha$*<br>*corresponds to*<br>*'confidence level'* | *Type I Error*<br>*Level of Significance*<br>*($\alpha$)* |
| *If Ho is False* | *Type II Error*<br>*Prob. = $\beta$* | *Correct decision*<br>*Power of the Test*<br>*(1-$\beta$)* |

Latihan : Bila kita memakai $\alpha$ = 5 % uji dua arah, gambarkan daerah kritis yang menunjukkan penerimaan atau penolakan Ho berdasar kurve normal standar..

## 5.5. Langkah-langkah Pengujian Hipotesis.

Langkah-langkah pengujian hipotesis seperti yang diuraikan dalam buku-buku statistika dapat berbeda-beda, namun semuanya menuju kepada suatu kesimpulan sama yaitu untuk menerima atau menolak Ho. Dalam pengujian hipotesis, Ho berfungsi sebagai hipotesis penguji karena hipotesis dalam bentuk Ho itulah yang selalu akan diuji. Aturan keputusan umum untuk menolak Ho dapat dinyatakan dengan ungkapan apabila " | yang dihitung | " > " | yang di tabel" | maka Ho ditolak. Sebagai kesimpulan atau hasil dari pengujian hipotesis adalah penolakan atau penerimaan Ho. Lebih lanjut hal tersebut dapat pula diartikan sebagai penerimaan atau penolakan hipotesis alternatif (Hia. Harus pula diingat, bahwa penolakan atau penerimaan Ho itu dilakukan pada kondisi tertentu, yaitu tingkat signifikansi yang dipilih ($\alpha$) dan dalam beberapa kasus *degree of freedom* (d.f) tertentu. Jika kita menolak Ho pada $\alpha$ tertentu, misalnya $\alpha$ = 5%, belum tentu kita juga menolak Ho tersebut pada $\alpha$ yang lain, misalnya $\alpha$ = 1%.

Sebagai contoh langkah-langkah pengujian hipotesis yang dikemukakan oleh Sanders dan oleh Anto Dajan adalah sebagai berikut :

| Sanders | Anto Dajan |
|---|---|
| 1. Nyatakan Ho dan Ha | 1. Nyatakan Ho dan Ha |
| 2. Pilih tingkat keyakinan (α) yang dikehendaki | 2. Pilih tingkat keyakinan (α) dan besar cuplikan |
| 3. Tentukan distribusi uji yang akan dipakai | 3. Pilih statistik uji yang sesuai sebagai dasar pengujian |
| 4. Definisikan daerah penolakan Hipotesis | 4. Tentukan daerah kritis |
| 5. Nyatakan aturan keputusan | |
| 6. Lakukan perhitungan secukupnya terhadap data sampel (cuplikan) | 5. Kumpulkan data cuplikan/sampel dan hitung statistik cuplikan/sampel serta ubah ke dalam variabel standar Z |
| 7. Hitung rasio kritis | |
| 8. Bandingkan rasio kritis dengan aturan keputusan | 6. Membandingkan nilai yang diobservasi bagi karakteristik tertentu dengan nilai teoritis yangdi nyatakan oleh hipotesis nya, untuk menerima atau menolak Ho. |
| 9. Ambil kesimpulan secara statistik tentang Ho | |

5.6. Soal

1. Ada suatu penelitian yang bertujuan untuk mengetahui pengaruh dari suatu perlakuan (*treatment*) terhadap sekumpulan ayam buras. Coba anda bantu peneliti tersebut dengan membuat
   a. Rumusan Ho dan Ha yang cocok.
   b. Memberikan alur langkah pengujian Ho tersebut.
2. Apa perbedaan uji hipotesis satu arah dengan uji hipotesis dua arah.

5.7. Rangkuman

Hipotesis merupakan jawaban sementara sebagai pemecahan terhadap suatu masalah. Jawaban tadi bersifat sementara, karena menunggu hasil pembuktian atau pengujiannya. Dalam melakukan pengujian hipotesis selalu terkandung adanya kemungkinan kesalahan. Kesalahan tersebut dapat berupa kesalahan Tipe I dan Tipe II.Kedua kesalahan tersebut dapat kita tekan sekecil mungkin dengan memilih daerah kritis yang setepat-tepatnya. Hipotesis yang diuji selalu dalam rumusan Ho. Penolakan Ho pada tingkat signifikansi tertentu juga berarti penerimaan Ha pada tingkat signifikansi tersebut. Secara teoritis kita selalu dapat menerima Ho atau sebaliknya menolak Ho dengan jalan mengubah tingkat signifikansi yang dipakai.

# POKOK BAHASAN VI
# UJI - Z

6.1. Pendahuluan.

Setelah kita mempelajari pokok bahasan V tentang pengujian hipotesis, maka dapatlah sekarang kita melangkah pada salah satu aplikasi pengujian hipotesis yaitu pengujian dengan cuplikan besar yang sering disebut uji-Z. mengenai batas antara cuplikan/sampel besar dan cuplikan/sampel kecil ada dua pendapat yaitu 25 atau 30. Sama atau di atas batas tersebut termasuk cuplikan/sampel besar. Yang sering diuji dengan uji-Z adalah satu rata-rata, beda antara dua rata-rata, satu proporsi dan beda/selisih antara dua proporsi.
Secara rinci pokok bahasan ini mencakup tentang :
a. Uji-Z untuk rata-rata hitung populasi
b. Uji-Z untuk selisih dua rata-rata hitung populasi
c. Uji-Z untuk proporsi populasi
d. Uji-Z untuk selisih dua proporsi populasi.

6.2. T I U dan Sasbel.

T I U :
Setelah mempelajari pokok bahasan ini mahasiswa mampu memahami dan menerapkan uji-Z dengan cuplikan/sampel besar serta melakukan pengambilan kesimpulannya.

Sasaran belajar :
Setelah mempelajari pokok bahasan ini mahasiswa diharapkan mampu :
a. Melakukan pengujian hipotesis tentang rata-rata hitung populasi pada tiga keadaan yang berbeda.
b. Melakukan pengujian hipotesis tentang proporsi pada dua keadaan yang berbeda.
c. Melakukan pengujian hipotesis tentang beda antara dua rata-rata hitung pada tiga keadaan yang berbeda.
d. Melakukan pengujian hipotesis tentang beda antara dua proporsi.

6.3. Uji-Z untuk Parameter Rata-rata Hitung Populasi.

Ada tiga kemungkinan keadaan yang akan kita bicarakan dalam sub pokok bahasan ini, yaitu uji terhadap $\mu$ dengan :

$\sigma$ diketahui atau tidak diketahui,
Populasi (N) tidak terbatas atau terbatas,
Tingkat signifikansi ($\alpha$) yang berbeda-beda.

Uji-Z ini hanya digunakan apabila kita dalam melakukan pengujian hipotesis menggunakan cuplikan besar. Suatu cuplikan dikatakan besar apabila $n \geq 25$. Sedangkan jika populasinya terbatas, perlu adanya faktor koreksi terbatasnya populasi yang besarnya :

$$\sqrt{\frac{N-n}{N-1}}$$

Pengujian terhadap parameter rata-rata $\mu$ dengan $\sigma$ diketahui.
Rumus statistik uji-Z adalah :

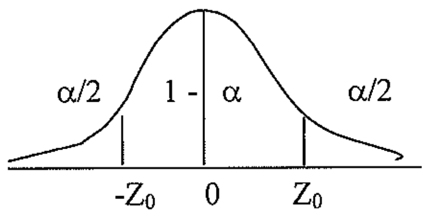
$$Z_{hit} = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Contoh :
Secara teknis populasi yang terdiri dari seluruh tube TV yang dihasilkan oleh sebuah perusahaan industri TV memiliki rata-rata umur 1200 jam dan deviasi standar sebesar 300 jam. Setelah perusahaan berjalan selama 3 tahun, diambil cuplikan terhadap tube TV sebanyak 100 buah, dan diperoleh rata-rata hidup 1265 jam. Para teknisi masih percaya bahwa deviasi standar tube TV tersebut masih tetap. Apakah ada alasan guna meragukan bahwa rata-rata umur tube TV yang dihasilkan perusahaan tersebut tidak sama dengan 1200 jam ?

Persoalan di atas dapat diselesaikan langkah demi langkah sebagai berikut :

1. Ho : $\mu = 1200$, Hi : $\mu \neq 1200$
2. Tingkat signifikansi ($\alpha$) = 5 %, uji dua arah
3. Dengan cuplikan/sampel lebih dari 25 maka statistik uji yang dapat digunakan adalah uji-Z, dimana

$$Z_{hit} = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$




4. Statistik uji di atas merupakan distribusi normal.
5. Dengan tingkat signifikansi 5 % dalam uji dua arah ialah $Z_0 = Z_{\alpha/2} = Z_{0,025} = 1{,}96$, sehingga daerah kritis atau daerah penolakan Ho adalah < -1,96 atau . > 1,96
6. Dari data yang diketahui dapay dihitung besarnya $Z_{hit}$ yaitu

$$Z_{hit} = \frac{1265 - 1200}{\frac{300}{\sqrt{100}}} = 2{,}17$$

7. Jadi Z hit > Zo, berarti Ho ditolak. Bahwa tidak ada keraguan untuk menyatakan rata-rata hidup tube TV tersebut tidak sama dengan 1200 jam.

Pada pengujian terhadap parameter rata-rata dengan $\sigma$ tidak diketahui, rumus $Z_{hit}$ yang lalu tidak dapat digunakan, dan perlu diganti dengan Rumus statistik uji-Z berikut :

$$Z_{hit} = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

s dalam hal ini adalah deviasi standar cuplikan, yang dihitung dengan rumus

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}}$$

Latihan :
Dalam suatu penelitian tentang rata-rata berat roti kaleng diperoleh informasi sebagai berikut Besar cuplikan = 64 kaleng, berat rata-ratanya = 0,53 kg dengan varians sebesar 0,4 kg. Apakah ada alasan secara nyata untuk mengatakan bahwa rata-rata berat roti kaleng tersebut sebesar 0,5 kg ?

6.4. Uji-Z untuk Selisih Dua Rata-rata Hitung Populasi.

Dalam hal ini pengujian Ho: $\mu_1 = \mu_2$ mempunyai empat keadaan alternatif, yaitu :
a. $\sigma$ diketahui dan $\sigma_1 = \sigma_2$,
b. $\sigma$ diketahui tetapi $\sigma_1 \neq \sigma_2$
c. $\sigma$ tidak diketahui tetapi dianggap $\sigma_1 = \sigma_2$, dan
d. $\sigma$ tidak diketahui dan $\sigma_1 \neq \sigma_2$.

ad a. Dalam hal ini Z hit dapat dicari dengan rumus :

$$Z_{hit} = \frac{(X_1 - X_2) - 0}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ad b. Z hitung dapat dicari dengan rumus

$$Z_{hit} = \frac{(X_1 - X_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

ad c. Jika $\sigma$ tidak diketahui maka dapat digunakan s sebagai penduga bagi $\sigma$. Karena $\sigma$ yang tidak diketahui berupa deviasi standar yang berlaku baik bagi populasi 1 maupun populasi 2, maka perlu dicari $s_p$($s_{pooled}$) yaitu deviasi standar gabungan dari $s_1$ dan $s_2$.

Dalam hal ini rumus sp sebagai deviasi standar gabungan adalah :

$$s_p = \sqrt{\frac{\sum(X_{i1} - \overline{X}_1)^2 - \sum(X_{i2} - \overline{X}_2)^2}{n_1 + n_2 - 2}}$$

atau

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Oleh karena itu, rumus Z hitungnya menjadi :

$$Z_{hit} = \frac{(X_1 - X_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ad d. Dalam hal ini $s_1$ dan $s_2$ digunakan langsung sebagai penduga bagi $\sigma_1$ dan $\sigma_2$. Oleh karena itu rumus Z hit menjadi :

$$Z_{hit} = \frac{(X_1 - X_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Contoh : Cuplikan acak masing-masing sebesar 60 mahasiswa diambil dari dua universitas yang berbeda. Cuplikan pertama mempunyai nilai rata-rata sebesar 77 dengan varians 36. Cuplikan kedua mempunyai rata-rata sebesar 66 dengan deviasi standar 10. Apakah ada perbedaan secara sangat nyata rata-rata nilai mahasiswa dua universitas tersebut ?

Jawab :

Dari soal tersebut, hal-hal yang diketahui adalah :

$n_1 = n_2 = 60$, berarti cuplikan besar.

$N_1$ maupun $N_2$ tidak tidak diketahui, dianggap jumlahnya besar atau tak terhingga.

$\overline{X}_1 = 77$, varians, $s_1^2 = 289$ atau deviasi standarnya, $s_1 = 17$

$\overline{X}_2 = 66$, deviasi standar, $s_2 = 10$.

Deviasi standar populasi tidak diketahui, maka dapat dibuat asumsi kedua deviasi standar populasi tersebut sama atau tidak sama.

Perbedaan secara sangat nyata, berarti $\alpha$ sebesar 1 %.

Bila dianggap bahwa deviasi standar kedua populasi tidak sama, maka :

Ho : $\mu_1 - \mu_2 = 0$

Ha: $\mu_1 - \mu_2 \neq 0$

$$Z_{hit} = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$Z_{hit} = \frac{(77-66)-0}{\sqrt{\frac{289}{60}+\frac{100}{60}}} = \frac{11}{2,546} = 4.3205$$

$$Z_{0,005} = 2,575$$

Berarti Z hit > Z tabel, maka Ho ditolak. Jadi ada perbedaan secara sangat nyata nilai rata-rata dari mahasiswa dua universitas tersebut.

Bila deviasi standar kedua populasi dianggap sama, maka :

Ho : $\mu_1 - \mu_2 = 0$

Hi : $\mu_1 - \mu_2 \neq 0$

Kemudian deviasi standar gabungannya dicari dengan rumus

$$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

$$s_p = \sqrt{\frac{(60-1)(289)+(60-1)(100)}{60+60-2}}$$

$= 13,9463$

oleh karena itu besarnya $Z_{hit}$ dapat dicari, yaitu sebesar

$$Z_{hit} = \frac{(X_1-X_2)-0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$Z_{hit} = \frac{(77-66)-0}{13,9463\sqrt{\frac{1}{60}+\frac{1}{60}}} = \frac{11}{2,546}$$

$$= 4,3205$$

Kesimpulan seperti jawaban sebelumnya..

Latihan :

Sebuah cuplikan acak yang terdiri dari 100 buruh perusahaan A telah dipilih dan mereka ternyata dapat menyelesaikan suatu tugas yang cukup rumit secara rata-rata dalam 12 menit dengan deviasi standar sebesar 2 menit. Sebuah cuplikan acak lainnya yang terdiri dari 50 buruh perusahaan B yang sejenis telah dipilih dan mereka ternyata dapat menyelesaikan tugas yang sama seperti yang diselesaikan buruh perusahaan A, secara rata-rata dalam 11 menit dengan deviasi standar sebesar 3 menit.

Apakah ada perbedaan secara sangat nyata dalam rata-rata kemampuan menyelesaikan tugas dua kelompok buruh tersebut ?

## 6.5. Uji-Z untuk Parameter Proporsi Populasi.

| | | |
|---|---|---|
| Hipotesis nolnya berbentuk | Ho : | $P = Po$ |
| Sedangkan Ha dapat berbentuk | Ha : | $P \neq Po$ atau |
| | | $P > Po$ atau |
| | | $P < Po$ |

Statistik uji Z-nya dapat diberikan sebagai :

$$Z_{hit} = \frac{(p_1 - p_2) - 0}{\sqrt{\frac{P(1-P)}{n}}}$$

Bila P tidak diketahui maka bisa diganti dengan $p = x/n$ yaitu proporsi cuplikan., atau dengan nilai P yang memaksimalkan P(1-P)

Contoh : Sebuah cuplikan acak yang terdiri dari 400 unit tube TV telah dipilih dari suatu populasi tube TV yang jumlahnya besar sekali. Setelah diteliti secara seksama, ternyata 12 tube di atas dinyatakan rusak atau tidak memenuhi kualitas standar.

Apakah hasil cuplikan di atas merupakan suatu bukti yang cukup guna menarik kesimpulan bahwa persentase tube rusak adalah lebih dari 2 % ?

Prosedur pengujian :

1. Ho : $P \leq 0{,}02$ Hi : $P > 0{,}02$
2. $\alpha = 5$ %, uji satu arah.
3. $Z_{hit} = \dfrac{\frac{x}{n} - P_o}{\sqrt{\frac{P(1-P)}{n}}}$
4. Statistik uji di atas merupakan distribusi binomial yang dapat didekati dengan distribusi normal.
5. Daerah kritis dengan tingkat signifikansi 5% uji dsatu arah ialah Z hit > Zo atau Z hit > 1,645
6. $Z_{hit}$

$$Z_{hit} = \frac{\frac{12}{400} - 0{,}02}{\sqrt{\frac{(0{,}2)(0{,}8)}{400}}} = \frac{\frac{12}{400} - \frac{8}{400}}{\sqrt{\frac{0{,}16}{400}}} = \frac{4}{400} x \frac{100}{2} =$$

$$= 0{,}50$$

7. Jadi Z hit < Z tabel, berarti Ho diterima. Pernyataan bahwa proporsi kerusakan sama dengan 2 % dapat diterima, adanya kerusakan sebesar 12 dari 400 tube (0,03) hanya terjadi secara kebetulan.

Latihan :

Apabila dari suatu cuplikan acak sebesar 256 konsumen produk Y ternyata ada 13 orang yang kurang menyukai bentuk produk Y tersebut, apakah kita yakin bahwa proporsi konsumen yang kurang menyukai bentuk produk Y kurang dari 7,5 % ?,

Bila n semakin besar, maka x/n kurang lebih akan didistribusikan secara normal dengan rata-rata hitung P dan deviasi standar $\sqrt{\frac{P(1-P)}{n}}$

Statistik uji terhadap selisih antara dua proporsi populasi dapat dirumuskan sebagai :

$$Z_{hit} = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{(P_1)(1-P_1)}{n_1} + \frac{(P_2)(1-P_2)}{n_2}}}$$

Pada hakekatnya, $P_1$ dan $P_2$ umumnya tidak diketahui sehingga harus diduga. Karena pengujian dilakukan terhadap $P_1 = P_2$, maka $P_1 = P_2 = P$ diduga dengan $p_p$ yang merupakan penduga bagi gabungan proporsi populasi, dan dirumuskan sebagai :

$$P_p = \frac{p_1 + p_2}{n_1 + n_2}$$

Contoh : Suatu penelitian mengenai preferensi konsumen terhadap sabun mandi cap Camay telah dilakukan oleh perusahaan industri yang bersangkutan. Penelitian tersebut dilakukan pada 200 keluarga responden. Berdasarkan pendapatan rata-rata per bulan, responden dapat dibagi dalam 2 golongan yang berbeda. Golongan I merupakan golongan mampu meliputi 30 % dari seluruh responden dan golongan II merupakan golongan kurang mi 70 % dari responden. Pada golongan pertama, 40 orang menyatakan suka pada sabun Camay sedangkan pada golongan kedua, 80 orang menyatakan suka pada sabun Camay.

Adakah alasan guna menyangsikan pernyataan yang menganggap bahwa proporsi kedua golongan konsumen yang menyukai sabun Camay tersebut sama ?

Prosedur pengujian adalah sebagai berikut :

1. Ho : $P_1 = P_2$ ; $H_a : P_1 > P_2$ ; $n_1 = 60$ ; $n_2 = 140$
   $p_1 = 40/60 = 0{,}67$ : $p_2 = 80/140 = 0{,}57$
2. $\alpha = 5$ %, uji satu arah, $Z_{0,05} = 1{,}645$
3. Z hit = $$Z_{hit} = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{(P_1)(1-P_1)}{n_1} + \frac{(P_2)(1-P_2)}{n_2}}}$$

   Bila menggunakan prinsip kehati-hatian, digunakan P yang memaksimalkan P(1-P) yaitu 0,50. maka $Z_{hit}$ =

$$Z_{hit} = \frac{(0{,}67-0{,}57)-0}{\sqrt{\frac{(0{,}5)(1-0{,}5)}{60}+\frac{(0{,}5)(1-0{,}5)}{140}}} = \frac{0{,}10}{\sqrt{\frac{1}{240}+\frac{1}{640}}}$$

$$= \frac{0{,}10}{0{,}075691} = 1{,}321157$$

Jadi $Z_{hit} < Z_{0,05}$ atau 1,321157 < 1,645, maka Ho diterima. Berarti tidak ada perbedaan yang signifiakan proporsi antara kedua kelompok tersebut.

5. $p_p = \frac{x_1 + x_2}{n_1 + n_2} = \frac{40+80}{60+140} = 0{,}60$
6. Bila menggunakan $p_p$ maka $Z_{hit}$

$$Z_{hit} = \frac{(0{,}67-0{,}57)-0}{\sqrt{\frac{(0{,}6)(1-0{,}6)}{60}+\frac{(0{,}6)(1-0{,}6)}{140}}} = \frac{0{,}10}{\sqrt{\frac{24}{6000}+\frac{24}{14000}}}$$

$$= \frac{0{,}10}{0{,}080178} = 1{,}247219$$

7. Kesimpulan : Z hit < Z tabel, berarti Ho diterima. Jadi tidak ada alasan guna menyangsikan pernyataan bahwa proporsi kedua kelompok konsumen tersebut sama.

Latihan :
Guna menguji efektif tidaknya suatu vaksin, maka 150 ekor tikus diberi vaksinasi sedangkan 150 ekor lainnya tidak diberi vaksinasi di atas. Ketigaratus ekor tikus tersebut ditulari dengan semacam penyakit. Di antara tikus yang menerima vaksinasi, 10 ekor mati karena penyakit itu sedangkan di antara tikus yang tidak diberi vaksinasi ternyata 30 ekor mati. Dapatkah kita menarik kesimpulan bahwa vaksin tersebut memeang efektif guna memperkecil kematian dari penyakit itu ? Gunakan tingkat signifikansi 5% Berilah evaluasi terhadap hasil pengujian saudara.

## 6.7. Soal

1. Dua merek ban truck akan dibandingkan daya tahannya oleh suatu perusahaan angkutan darat. Suatu cuplikan ban merek A sejumlah 40 buah mempunyai rata-rata sebesar 45.000 mil, sedangkan dari sejumlah cuplikan yang sama untuk ban merek B diperoleh rata-rata sebesar 46.500 mil. Dengan menganggap bahwa deviasi standar kedua populasi ban adalah tetap yaitu masing-masing sebesar 2.000 mil dan 1.500 milm apakah ada perbedaan yang signifikan antara rata-rata daya tahan kedua merek ban tersebut ? Gunakan $\alpha$ = 1 %.

2. TVRI percaya bahwa suatu program baru yang ditayangkan akan lebih disukai oleh pemirsa di kota daripada pemirsa di desa. Untuk menguji kepercayaan tersebut, dilakukan pengambilan cuplikan acak terhadap 300 pemirsa di kota dan 100 pemirsa

di desa. Ternyata 65 pemirsa di kota dan 18 pemirsa di desa menyukai program baru itu. Pada $\alpha$ = 5 %, dapatkah kepercayaan TVRI tersebut kita terima ?

## 6.8. Rangkuman

Pengujian hipotesis dengan cuplikan besar, yaitu uji-Z, dapat dilakukan untuk menguji $\mu$, $\mu_1 - \mu_2$, P, dan $P_1$-$P_2$. Batas cuplikan besar dan kecil dapat dipakai n = 25.
Keadaan dalam pengujian dapat berupa : (1) Populasi terbatas atau tidak terbatas; (2) Deviasi standar populasi diketahui atau tidak
Bila deviasi standar populasi tidak diketahui, dapat diasumsi atau bahkan diuji (dengan uji F) bahwa keduanya sama atau tidak sama. Bila sama, hitung dulu deviasi standar gabungan, dan bila tidak sama gunakan deviasi standar masing-masing cuplikan.
Penerimaan atau penolakan Ho semata-mata tergantung pada besarnya tingkat signifikansi ($\alpha$) yang dipilih.

# POKOK BAHASAN VII
# UJI - t

'.1. Pendahuluan.

Uji-t merupakan salah satu bentuk pengujian hipotesis terhadap variabel yang berskala interval, dengan menggunakan cuplikan/sampel kecil, tentang proporsi, rata-rata dan beda antara dua rata-rata.. Untuk dapat memahami pokok bahasan ini dengan baik diperlukan pemahaman tentang langkah-langkah pengujian hipotesis dan cara penggunaan tabel t seperti yang diuraikan dalam pokok bahasan distribusi probabilita.

Secara berurutan, pokok bahasan uji-t ini mencakup :

a. Uji-t terhadap parameter rata-rata hitung populasi
b. Uji-t terhadap parameter proporsi populasi
c. Uji-t terhadap selisih antara dua parameter rata-rata hitung populasi.

'.2. T I U dan Sasbel.

T I U :

Setelah mempelajari pokok bahasan ini mahasiswa dapat memahami dan mampu menerapkan uji-t dengan cuplikan kecil serta melakukan pengambilan kesimpulannya.

Sasaran belajar :

Mahasiswa diharapkan mampu :

a. Melakukan pengujian hipotesis tentang rata-rata hitung populasi.
b. Melakukan pengujian hipotesis tentang proporsi populasi.
c. Melakukan pengujian hipotesis tentang beda antara dua rata-rata hitung populasi.

'.3. Uji-t untuk Rata-rata Hitung Populasi.

Pada prinsipnya, pengujian terhadap $\mu$ dengan cuplikan/sampel kecil ini sama dengan pengujian terhadap $\mu$ dengan cuplikan besar. Hanya saja, dalam pengujian dengan cuplikan kecil masalah derajat kebebasan atau *degree of freedom* (d.f) menjadi mengemuka. Nilai t tabel pada tingkat signifikansi ($\alpha$) tertentu akan berubah sesuai dengan berubahnya d.f tersebut. Nilai t hitung atau t rasio dapat diperoleh dengan rumus :

$$t_{hit} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

Nilai $t_{hit}$ ini kemudian diperbandingkan dengan nilai t tabel pada $\alpha$ dan d.f. yang sesuai.. Yang merupakan d.f dalam hal ini adalah pembagi/penyebut dalam rumus untuk menghitung deviasi standar cuplikan, yaitu n-1. Bila deviasi standar populasi ($\sigma$) tidak diketahui, dapat digunakan deviasi standar cuplikan yang dilambangkan dengan s. Sedangkan bila populasi (N) terbatas, maka perlu dipakai faktor koreksi terbatasnya N yang besarnya sama dengan :

$$\sqrt{\frac{N-n}{N-1}}$$

Uji t dengan cuplikan kecil bagi μ dapat berupa uji satu arah maupun uji dua arah. Prosedur pengujian dalam uji-t langkah demi langkah sama dengan prosedur pengujian dengan uji-Z. Ada 4 alternatif rumus $t_{hit}$ dalam uji hipotesis terhadap μ dengan cuplikan kecil ini, yaitu:

1). $$t_{hit} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$

2). $$t_{hit} = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$

3). $$t_{hit} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n} \sqrt{\frac{N-n}{N-1}}}$$

4). $$t_{hit} = \frac{\overline{X} - \mu_0}{s / \sqrt{n} \sqrt{\frac{N-n}{N-1}}}$$

Contoh :
Berdasar spesifikasi dari pabriknya, mesin photo copy "R" dapat mengkopi 65 helai kertas per menit. Sebuah usaha photo copy ingin membuktikan keabsahan spesifikasi di atas. Perusahaan mengadakan observasi secara empiris dengan menggunakan 13 buah mesin photo copy "R" dan hasil observasi sedemikian itu diberikan sebagai berikut :

60 59 62 62 55 61 58 64 56 65 54 62 67

Apakah ada alasan bagi perusahaan guna mempercayai hipotesis di atas ?

Jawab :
Dari data tersebut secara deskriptip diperoleh n = 13 ; $\overline{X}$ = 60 dan s = 4,1833.

Prosedur pengujian langkah demi langkah adalah sebagai berikut :

1. Ho : $\mu o = 65$, Hi : $\mu o \neq 65$
2. $\alpha = 5\ \% = 0{,}05$ uji dua arah.
3. $$t_{hit} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$$
4. Daerah kritis dengan tingkat signifikansi 5 % secara dua arah adalah sebagai berikut :
   $t_{hit} > t_{0,025;\ 12}$ dan $t_{hit} < -t_{0,025;12}$
   atau
   $t_{hit} > 2{,}179$ dan $t_{hit} < -\ 2{,}179$

5. $t_{hit} = \dfrac{60-65}{4{,}1833/\sqrt{13}}$

$= -\ 4{,}3095$

6. Karena $-4{,}3095 < -2{,}179$, maka kita tolak Ho yang menyatakan bahwa $\mu o = 65$. Jadi, rata-rata kemampuan mesin photo copy ($\mu$) adalah tidak sama dengan 65 lembar per menit..

Latihan :
Sebuah cuplikan yang terdiri dari 19 kaleng penambal atap (*roof seal*) memiliki isi berat kotor seperti yang diberikan di bawah ini. Isi berat kotor dinyatakan dalam kg per kaleng.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1,21 | 1,21 | 1,23 | 1,20 | 1,21 | 1,24 | 1,22 | 1,24 | 1,22 | 1,18 |
| 1,21 | 1,19 | 1,19 | 1,18 | 1,19 | 1,23 | 1,18 | 1,19 | 1,20 | |

Jika kita menggunakan taraf nyata sebesar 1 %, dapatkah kita diyakinkan bahwa populasi *roof seal* dalam kaleng secara rata-rata memiliki isi berat kotor 1,2 kg per kaleng ? Berilah evaluasi terhadap hasil perhitungan Saudara !

.4. Uji-t untuk Proporsi Populasi.

Proporsi sebetulnya merupakan variabel yang bersifat diskrit. Tetapi dengan cuplikan yang cukup, pendekatan normal yang berlandaskan pada variabel yang bersifat kontinu/sinambung dapat dipergunakan. Untuk itu kalau kita menginginkan suatu pengujian yang cermat faktor koreksi kontinuitas dapat pula diaplikasikan di sini.
Rumus $t_{hit}$ dalam pengujian terhadap proporsi populasi adalah sebagai berikut :

$$t_{hit} = \frac{p - P_0}{\sqrt{\dfrac{P_0(1-P_0)}{n}}}$$

Sedangkan bila koreksi kontinuitas diaplikasikan, maka rumusnya menjadi

$$t_{hit} = \frac{p - \dfrac{1}{2n} - P_0}{\sqrt{\dfrac{P_0(1-P_0)}{n}}}$$

Ketentuan tentang seberapa besar n yang harus dipakai, seperti dalam pendugaan/estimasi terhadap proporsi populasi masih tetap berlaku. Oleh karena itu dalam pengujian terhadap proporsi populasi ini praktis jarang digunakan cuplikan kecil.
Bila populasinya terbatas, maka perlu faktor koreksi terbatasnya populasi, sehingga runus $t_{git}$ menjadi

$$t_{hit} = \frac{p - P_0}{\sqrt{\dfrac{P_0(1-P_0)}{n}}\sqrt{\dfrac{N-n}{N-1}}} \quad \text{atau} \quad t_{hit} = \frac{p - \dfrac{1}{2n} - P_0}{\sqrt{\dfrac{P_0(1-P_0)}{n}}\sqrt{\dfrac{N-n}{N-1}}}$$

Contoh :
Dalam sebuah cuplikan acak yang terdiri dari ban mobil yang diproduksi perusahaan industri ban, 25% ternyata tidak memenuhi kualitas standar. Apakah kita yakin bahwa proporsi ban yang tidak memenuhi kualitas standar lebih dari 20 %, bila cuplikan yang dipergunakan sebesar 25 buah ban ?

Jawab : Dari soal tersebut dapat diketahui hal-hal berikut :

$n = 25$ (Cuplikan kecil)

$p = \frac{x}{n} = 0{,}25$ $Po = 0{,}20$

$Ho : P \leq 0{,}20$

$Ha : P > 0{,}20$

Syarat besar cuplikan yang dapat digunakan dalam pengujian proporsi yaitu $nP \geq 5$ dipenuhi, karena $25(0{,}2) = 5$.

$\alpha = 5$ %, uji satu arah.

d.f = 24.

$$t_{hit} = \frac{p - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

$$= \frac{0{,}25 - 0{,}20}{\sqrt{\frac{0{,}20(1-0{,}20)}{25}}} = 0{,}625$$

$t_{0,05;24} = 1{,}711$

Karena t hit < t tabel pada $\alpha = 5$ % dan d.f 24, maka Ho diterima. Berarti kita tidak yakin bahwa proporsi produk yang tidak memenuhi standar lebih dari 20 %.

7.5. Uji-t untuk Beda antara Dua Rata-rata Hitung Populasi.

Dalam hal ini pengujian Ho: $\mu_1 = \mu_2$ mempunyai dua keadaan alternatif, yaitu :

a. Deviasi standar populasi $\sigma$ diketahui, dan
b. Deviasi standar populasi tidak diketahui, yang dapat dibedakan lagi menjadi dua alternatif, yaitu : b1. $\sigma$ tidak diketahui tetapi dianggap $\sigma_1 = \sigma_2$, dan
 b2. $\sigma$ tidak diketahui tetapi dianggap $\sigma_1 \neq \sigma_2$.

ad a. Jika deviasi standar kedua populasi diketahui, maka kita bisa langsung menggunakan nilai dari kedua devaiasi standar tadi dalam rumus untuk mencari nilai $t_{hit}$ Dalam kasus ini $t_{hit}$ dapat dihitung dengan rumus :

$$t_{hit} = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

ad b1. Jika $\sigma$ tidak diketahui tetapi dianggap kedua deviasi standar populasi tersebut sama, maka dapat digunakan deviasi standar cuplikan gabungan, $s_p$ sebagai penduga bagi $\sigma$ kedua populasi..

Dalam hal ini rumus deviasi standar gabungan, $s_p$ sebagai deviasi standar gabungan yang berlaku untuk kedua populasi, adalah :

$$s_p = \sqrt{\frac{\sum(X_{i1} - \overline{X}_1)^2 - \sum(X_{i2} - \overline{X}_2)^2}{n_1 + n_2 - 2}}$$

atau

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Kedua rumus $s_p$ di atas walaupun penampilannya berbeda, tetapi sesungguhnya kedua sama. Selanjutnya rumus $t_{hit}$ nya menjadi :

$$t_{hit} = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Contoh :

Dua jenis pupuk buatan telah digunakan di atas lahan pertanian padi yang memiliki tingkat kesuburan, cara budi daya maupun kondisi iklim yang kurang lebih sama. Tujuan penggunaan pupuk buatan di atas ialah untuk menguji apakah daya hasil salah satu jenis pupuk buatan tersebut betul-betul berbeda dari yang lain. Peneliti memilih secara acak 12 petak lahan pertanian dan memberinya dengan pupuk buatan $X_1$ dan 12 petak lahan pertanian lainnya untuk diberi pupuk buatan $X_2$. Hasil pertambahan padi dalam kg diberikan sebagai berikut.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Hasil penggunaan $X_1$ | 31 | 34 | 29 | 26 | 32 | 35 | 38 | 34 | 30 | 29 | 32 | 31 |
| Hasil penggunaan $X_2$ | 26 | 24 | 28 | 29 | 30 | 29 | 32 | 26 | 31 | 29 | 32 | 28 |

Apakah hasil produksi padi dengan penggunaan $X_1$ berbeda dari $X_2$ ?

Dari soal tersebut diperoleh : $\overline{X}_1 = 31{,}75$ $s_1^2 = 10{,}2045$

$\overline{X}_2 = 18{,}67$ $s_2^2 = 6{,}0580$

Langkah-langkah pengujian hipotesis :

1. Ho : $\mu_1 = \mu_2$ Ha : $\mu_1 \neq \mu_2$
2. $\alpha$ 5 %

3. Cuplikan kecil, uji t dua arah. Deviasi standar populasi tidak diketahui, dianggap kedua deviasi standar tersebut sama, sehingga perlu dicari dulu deviasi standar gabungan.

$$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

$$s_p = \sqrt{\frac{(12-1)(10{,}2045)+(12-1)(6{,}0580)}{12+12-2}}$$

$$t_{hit} = \frac{(\overline{X}_1-\overline{X}_2)-0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$t_{hit} = \frac{(31{,}75-18{,}67)-0}{s_p\sqrt{\frac{1}{12}+\frac{1}{12}}}$$

4. Daerah kritis dengan $\alpha$ = 5 % secara dua arah, dengan d.f (n1+n2-2) = 22, adalah t > 2,074 dan t < - 2,074

5. $$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

$$s_p = \sqrt{\frac{(12-1)(10{,}2045+(12-1)(6{,}0580}{12+12-2}}$$

$$= \sqrt{\frac{112{,}2495+66{,}6280}{22}}$$

$$= 2{,}8515$$

$$t_{hit} = \frac{(31{,}75-18{,}67)-0}{2{,}8515\sqrt{\frac{1}{12}+\frac{1}{12}}}$$

$$= 3{,}08 \,/\, 1{,}1641$$

$$= 2{,}6458$$

Karena $t_{hit} > t_{0,025;22}$ atau 2,6458 > 2,074, maka Ho ditolak. Jadi ada perbedaan yang berarti (signifikan) antara daya hasil pupuk $X_1$ dan $X_2$.

Rumus lain untuk mencari $t_{hit}$ adalah seperti yang terdapat dalam bukunya Anto Dayan, yaitu :

$$t_{hit} = \frac{(\overline{X}_1-\overline{X}_2)-0}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Kemudian bila populasinya terbatas, atau dengan cuplikan tanpa pengembalian, ditambahkan faktor koreksi terbatasnya populasi dengan rumus

$$\sqrt{\frac{(N_1 + N_2) - (n_1 + n_2)}{(N_1 + N_2 - 2)}}$$

ad b2. Dalam hal ini $s_1$ dan $s_2$ digunakan langsung sebagai penduga bagi $\sigma_1$ dan $\sigma_2$. Oleh karena itu rumus $t_{hit}$ menjadi :

$$t_{hit} = \frac{(\overline{X}_1 - \overline{X}_2) - 0}{\sqrt{\frac{s_1^2}{(n_1 - 1)} + \frac{s_2^2}{(n_2 - 1)}}}$$

Namun, dalam pengujian terhadap beda antara dua $\mu$ dengan cuplikan kecil di sini terdapat masalah d.f. Untuk ini d.f dapat dicari dengan rumus :

$$\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} - 2$$

Contoh :

Dari soal contoh seperti pada ad b1, tetapi dengan anggapan bahwa deviasi standar kedua populasi yang tidak diketahui dianggap tidak sama.

Penyelesaian terhadap soal tersebut menjadi :

1. Ho : $\mu_1 = \mu_2$
   $H_a$ : $\mu_1 \neq \mu_2$

2. $\alpha = 5\ \%$

3. $t_{hit} = \dfrac{(\overline{X}_1 - \overline{X}_2) - 0}{\sqrt{\frac{s_1^2}{(n_1 - 1)} + \frac{s_2^2}{(n_2 - 1)}}}$

4. Daerah kritis dengan $\alpha = 5\ \%$ secara dua arah, dengan d.f sebesar :

$$\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} - 2 =$$

$$\frac{\left(\frac{10,2045}{12}+\frac{6,0580}{12}\right)^2}{\frac{\left(\frac{10,2045}{12}\right)^2}{12-1}+\frac{\left(\frac{6,0580}{n_2}\right)^2}{12-1}}-2=$$

= 24.41 - 2 = 22,41 dibulatkan menjadi 22,
adalah $t_{0,025;22} > 2,074$ dan $- t_{0,025;22} < - 2,074$

5. $$t_{hit}=\frac{(31,75-28,67)-0}{\sqrt{\frac{10,2045}{(12-1)}+\frac{6,0580}{(12-1)}}}$$

= 3,08 / 1.1641 = 2.6458

Karena $t_{hit} > t_{0,025;22}$ atau 2,6458 > 2,074, maka Ho ditolak. Jadi ada perbedaan yang berarti antara daya hasil pupuk $X_1$ dan $X_2$.

Latihan :
Cuplikan acak berupa produksi barang per jam dari dua perusahaan sejenis yang masing-masing dipilih dari 2 populasi menghasilkan data sebagai berikut.

| Cuplikan | n | d.f | $\overline{X}$ | $\sum(X-\overline{X})^2$ |
|---|---|---|---|---|
| $X_1$ | 15 | 14 | 21 | 1224 |
| $X_2$ | 9 | 8 | 29 | 756 |

Apakah ada alasan guna mempercayai pendapat bahwa rata-rata produksi per jam populasi pertama lebih kecil daripada rata-rata produksi per jam populasi kedua ?
Gunakan taraf nyata 5 %.

## 7.6. Soal

1. Andaikan dua cuplikan acak masing-masing sebesar 10 dan 12 dipilih dari dua populasi normal yang independen dan andaikan hasil X-ba $\overline{X_1} = 20$, $\overline{X_2} = 24$, $s_1 = 5$ dan $s_2 = 6$.
   Coba uji Ho : $\mu_1 = \mu_2$, jika kita menganggap :
   a. $\sigma_1 = \sigma_2$
   b. $\sigma_1 \neq \sigma_2$.
   Gunakan alpha 5 %.

2. Suatu proses produksi dianggap dapat diawasi dan memenuhi ketentuan-ketentuan yang telah ditetapkan jika jumlah rata-rata kopi yang diisikan dalam kaleng ialah 0,50 kg. Deviasi standarnya ialah sebesar 0,02 kg. Sebuah cuplikan yang terdiri dari 100 kaleng yang telah terisi dipilih secara acak dan rata-rata cuplikannya ternyata sebesar 0,51 kg. Apakah proses produksi di atas sudah memenuhi ketentuan ? Gunakan tingkat signifikansi 5 %.

## 7.7. Rangkuman

Pengujian hipotesis dengan cuplikan kecil atau uji-t dapat digunakan untuk menguji $\mu$, P dan $\mu_1 - u_2$. Uji-t jarang digunakan untuk menguji beda antara dua proporssi populasi, karena adanya syaarat besarnya cuplikan minimal yang harus digunakan dalam menguji proporsi tersebut. Dalam memperbandingkan $t_{hit}$ dengan $t_{tabel}$ ada dua hal yang perlu diperhatikan yaitu besarnya $\alpha$ yang digunakan dan d.f yang gayut. Untuk menguji selisih antara dua rata-rata hitung populasi, bila deviasi standar populasinya tidak diketahui maka dapat digunakan deviasi standar cuplikan atau deviasi standar gabungan Ada dua asumsi yang dapat diambil tentang deviasi standar populasinya, yaitu deviasi standar populasi dianggap sama atau dianggap tidak sama.

# POKOK BAHASAN VIII
# UJI - F

## 8.1. Pendahuluan.

Agar supaya dapat memahami pokok bahasan ini dengan baik, mahasiswa harus sudah menguasai pengetahuan tentang pengujian hipotesis, rata-rata hitung dan varians. Kalau dalam pembicaraan tentang uji-Z dan uji-t ada pengujian terhadap satu atau dua rata-rata hitung populasi, maka dalam uji-F ini kita akan menguji atau memperbandingkan beberapa rata-rata hitung secara serempak. Di samping itu, kita juga dapat menguji perbandingan antara dua varians populasi. Distribusi probabilita yang digunakan dalam pokok bahasan ini adalah distribusi F, yang diberi nama demikian untuk menghormat Sir Ronald Fisher salah seorang pelopor dalam statistika modern. Beberapa karakteristik dari distribusi F adalah : (1) ada satu "keluarga" distribusi F. Anggota tertentu dari keluarga tersebut ditentukan oleh dua parameter yaitu derajat kebebasan (*degree of freedom* = d.f) untuk pembilang dan derajat kebebasan untuk penyebut. (2) nilai F tidak mungkin negatif, (3) Distribusi F merupakan distribusi yang tidak simetris, menceng ke kanan (*skewed to the right / positively skewed*), (4) alternatif nilai F antara 0 sampai dengan ∞. Ketika nilai F makin besar, kurvenya makin mendekati sumbu datar (sumbu x) tetapi tidak pernah menyentuhnya.

Pokok bahasan ini mencakup :

a. Analisis Varians (anava) satu arah
   - a1. Dengan cuplikan sama
   - a2. Dengan cuplikan berbeda

b.. Anava dua arah

c.. Perbandingan berganda.

d... Perbandingan antara dua varians populasi.

## 8.2. T I U dan Sasaran Belajar (Sasbel)

T I U :

Setelah mempelajari pokok bahasan ini mahasiswa diharapkan dapat memahami uji-F dan mampu menerapkannya serta melakukan pengambilan keputusannya pada berbagai keadaan..

Sasaran belajar :

Mahasiswa diharapkan dapat :

a. Melakukan uji-F dalam analisis varians (anava) satu arah dengan cuplikan sama.
b. Melakukan uji-F dalam analisis varians satu arah dengan cuplikan yang berbeda.
c. Melakukan uji-F dalam analisis varians dua arah.
d. Melakukan perbandingan berganda dalam analisis varians.
e. Melakukan perbandingan antara dua varians populasi.

## 8.3. Anava Satu Arah. (*One-way Analysis of Variance*)

Anava di sini digunakan untuk membandingkan beberapa rata-rata untuk menentukan apakah rata-rata tersebut diperoleh dari populasi yang sama atau tidak. Asumsi dalam penggunaan Anava adalah sebagai berikut :

1. Populasi populasi tersebut terdistribusi secara normal;
2. Populasi-populasi tersebut mempunyai deviasi standar yang sama; dan
3. Sampel yang digunsksn sdalah sampel independen

Dalam Anava satu arah ini, kita dapat memisahkan adanya dua macam Anava yaitu :

Anava satu arah dengan cuplikan (n) sama dan

Anava satu arah dengan cuplikan tidak sama.

a. Anava satu arah dengan cuplikan (n) sama.

Misalnya kita akan memperbandingkan produktivitas tiga buah mesin sejenis. Apakah produktivitas ketiga mesin tersebut memang berbeda ataukah perbedaan itu hanya karena adanya faktor kebetulan Mesin-mesin tersebut mungkin dilayani oleh karyawan yang berbeda, atau sebab-sebab lain yang tidak bisa dijelaskan, sehingga produktivitas ketiga mesin tersebut mungkin berbeda . Untuk menghindari hal tersebut, dari tiap mesin diambil cuplikan secara acak, misalnya lima kali agar supaya perbedaan secara kebetulan dapat dikurangi. Hasil observasi terhadap ketiga mesin tersebut dalam lima kali pengambilan sampel misalnya sebagai berikut :

Tabel 8.1. Produktivitas cuplikan tiga mesin (dalam unit)

| Mesin 1 | Mesin 2 | Mesin 3 | |
|---|---|---|---|
| 47 | 55 | 54 | |
| 53 | 54 | 50 | |
| 49 | 58 | 51 | |
| 50 | 61 | 51 | |
| 46 | 52 | 49 | |
| $\bar{X}_1 = 49$ | $\bar{X}_2 = 56$ | $\bar{X}_3 = 51$ | $\bar{\bar{X}} = 52$ |

Pengenalan notasi :

Xij adalah hasil observasi pada mesin ke i dan cuplikan ke j

j = 1,2, ....,c dimana c menunjukkan jumlah kolom (mesin)

i = 1,2, ....,n atau r dimana n menunjukkan banyaknya baris atau banyaknya cuplikan dari tiap mesin.

$\bar{X}_j$ adalah rata-rata hitung dari mesin ke j

$\bar{\bar{X}}$ adalah rata-rata hitung total (*grand mean*)

Hipotesis nol kita adalah

Tidak ada perbedaan rata-rata produktivitas antar ketiga mesin., atau

Ho : $\mu_1 = \mu_2 = \mu_3$

Sedangkan hipotesis alternatifnya adalah

Ada perbedaan rata-rata produktivitas antar ketiga mesin, atau

Hi : $\mu_1 \neq \mu_2 \neq \mu_3$

Untuk melakukan pengujian, kita harus menyusun tabel kerja yang juga disebut Tabel Anava sebagai berikut :

Tabel 8.2. Tabel Anava Satu Arah (Untuk cuplikan sama)

| Sumber Variasi | Variasi SS (*sum of squares*) | d.f. | Varians MSS (*mean sum of squares*) | F-*ratio* (F-hitung) |
|---|---|---|---|---|
| Antar kolom, dijelaskan | $SS_c = n\sum(\overline{X}_j - \overline{\overline{X}})^2$ | c-1 | $MSS_c = \frac{SS_c}{d.f.}$ | F *ratio* = $\frac{MSS_c}{MSS_u}$ |
| Dalam kolom, tidak dijelaskan, *residual* | $SS_u = \sum\sum(X_{ij} - \overline{X}_j)^2$ | c(n-1) | $MSS_u = \frac{SS_u}{d.f.}$ | |
| Total | $SS_T = \sum\sum(X_{ij} - \overline{\overline{X}}_j)^2$ | cn-1 | | |

$SS_T = SS_c + SS_u$ d.f $(SS_c)$ + d.f $(SS_u)$ = d.f $(SS_T)$

Variasi antar kolom disebut variasi yang dijelaskan, karena bila kita bertanya mengapa antar kolom ada perbedaan kita dapat mencari jawabnya, yaitu karena antar kolom menunjukkan mesin-mesin yang berbeda. Sedangkan variasi dalam kolom disebut variasi yang tidak dijelaskan atau *residual*, karena bila kita mengajukan pertanyaan seperti di atas, kita sementara ini tidak tahu jawabnya, mungkin terjadi secara kebetulan atau karena sebab yang lainnya. Jumlah dari variasi antar kolom dan variasi dalam kolom merupakan variasi total, yaitu seluruh variasi yang mungkin terjadi yang dihitung dari seluruh penyimpangan dari setiap angka hasil observasi terhadap rata-rata total (*grand mean*)-nya.

Berdasar data pada Tabel 8.1. kita dapat melakukan analisis varians satu arah sebagai berikut.
Dalam hal ini kita hanya ingin melihat ada tidaknya perbedaan produktivitas antar mesin tanpa memperhatikan siapa yang menjadi operator dari mesin-mesin tersebut atau waktu pengambilan sampelnya..
Langkah-langkah pengujian hipotesis :
Hipotesis nol (Ho) adalah

$$Ho : \mu_1 = \mu_2 = \mu_3$$

Sedangkan hipotesis alternatifnya adalah

$$Hi : \mu_1 \neq \mu_2 \neq \mu_3$$

Tingkat signifikansi yang kita pilih misalnya 5%, sehingga $F_{tabel}$ atau $F_{0,05:(2;12)} = 3,89$ (lihat Tabel Distribusi F yang biasanya ada pada Lampiran di buku-buku statistika inferensia).
Aturan keputusan dalam penolakan Ho adalah bila $F_{hit}$ atau F-ratio > $F_{tabel}$ pada $\alpha$ dan d.f tertentu.

Dengan demikian Tabel Anava satu arah-nya menjadi sebagai berikut.

Tabel 8.3. Tabel Anava Satu Arah dengan Data dari Tabel 8.1.

| Sumber Variasi | Variasi SS (*sum of squares*) | d.f. | Varians MSS (*mean sum of squares*) | F-*ratio* |
|---|---|---|---|---|
| Antar kolom, dijelaskan | $SS_c = n\sum(\overline{X}_j - \overline{\overline{X}})^2$ $= 5(26) = 130$ | c-1 2 | $MSS_c = \frac{SS_c}{d.f.}$ $= 65$ | F *ratio* = $\frac{MSS_c}{MSS_u}$ $= 8,3$ |
| Dalam kolom, tidak dijelas-kan, *residual* | $SS_u = \sum\sum(X_{ij} - \overline{X}_j)^2$ $= 94$ | c(n-1) 12 | $MSS_u = \frac{SS_u}{d.f.}$ $= 7,83$ | |
| Total | $SS_T = \sum\sum(X_{ij} - \overline{\overline{X}}_j)^2$ $= 224$ | cn-1 14 | | |

Perhitungan :

$SS_c = n\sum(\overline{X}_j - \overline{\overline{X}})^2 \quad = 5[(49\text{-}52)^2 + (56\text{-}52)^2 + (51 - 52)^2 = 5\ (26) \ = 130$

$SS_u = \sum\sum(X_{ij} - \overline{X}_j)^2 = [(47\text{-}49)^2 + (53\text{-}49)^2 + (49\text{-}49)^2 + (50\text{-}49)^2 + (46\text{-}49)^2] +$
$[(55\text{-}56)^2 + (54\text{-}56)^2 + (58\text{-}56)^2 + (61\text{-}56)^2 + (52\text{-}56)^{2]} +$
$[(54\text{-}51)^2 + (50\text{-}50)^2 + (51\text{-}51)^2 + (51\text{-}51)^2 + (49\text{-}51)^2]$
$= 94$

$SS_T = SS_c + SS_u \quad = 130 + 94 \quad = 224$

F-tabel dengan tingkat signifikansi 5 % d.f. (2;12) sebesar 3,89 atau $F_{0,05\ (2;12)} = 3,89$. Karena F *ratio* > F-tabel, maka Ho ditolak. Berarti ada perbedaan rata-rata produktivitas yang signifikan antara ketiga mesin tersebut. (Tanpa memperhatikan siapa operatornya) Dengan anava ini, kita juga dapat melakukan pembandingan antara dua rata-rata sepasang demi sepasang, yaitu dengan menyusun interval keyakinan secara simultan (simultaneous confidence interval), seperti yang dijelaskan pada butir 8.5.

b. Anava satu arah dengan cuplikan (n) tidak sama.

Kembali pada contoh tentang pembandingan rata-rata produktivitas tiga mesin, kita dapat pula mengguna-kan cuplikan yang tidak sama dari tiap mesin. Dari contoh pada Tabel 8.1. kita mengetahui bahwa $n_1 = n_2 = n_3 = 5$; sekarang misalnya kita mengubah n masing-masing mesin menjadi $n_1 = 3$, $n_2 = 5$ dan $n_3 = 4$. Keadaan seperti ini bisa terjadi misalnya karena ada kehilangan kesempatan untuk menggunakan sample yang sama. Berarti, cuplikan yang digunakan adalah tidak sama. Walaupun proses pengujian pada dasarnya sama, tetapi ada beberapa perubahan penyesuaian yang perlu dilakukan. Tabel Anava yang perlu kita kerjakan, yaitu Tabel Anava Satu Arah dengan Sampel tidak Sama, seperti yang terlihat pada Tabel 8.4 berikut.

**Catatan :**
**Perhatikan perubahan yang terjadi pada rumus untuk SSc, d.f. bagi SSu dan $SS_T$, dan rumus *grand mean* (rata-rata total).**

Tabel 8.4. Tabel Anava Satu Arah (untuk cuplikan tidak sama)

| Sumber Variasi | Variasi (SS) | d.f. | Varians (MSS) | F-*ratio* |
|---|---|---|---|---|
| Antar kolom, dijelaskan | $SS_c = \sum n_j(\overline{X}_j - \overline{\overline{X}})^2$ | c-1 | $MSS_c = \frac{SS_c}{d.f.}$ | F *ratio* = $\frac{MSS_c}{MSS_u}$ |
| Dalam kolom, tidak dijelaskan, *residual* | $SS_u = \sum\sum(X_{ij} - \overline{X}_j)^2$ | $\sum(n_j - 1)$ | $MSS_u = \frac{SS_u}{d.f.}$ | |
| Total | $SS_T = \sum\sum(X_{ij} - \overline{\overline{X}}_j)^2$ | $\sum n_j - 1$ | | |

$$\overline{\overline{X}} = \frac{\sum X_{ij}}{\sum n_j}$$

Contoh :
Suatu cuplikan tentang besar pendapatan dari 11 keluarga Amerika tahun 1971 yang berasal dari empat daerah menghasilkan data sebagai berikut :
Pendapatan keluarga menurut daerah (dalam ribuan dollar)

| Timur Laut | Utara | Selatan | Barat |
|---|---|---|---|
| 8 | 13 | 7 | 7 |
| 14 | 9 | 14 | 7 |
| | | 8 | 16 |
| | | 7 | |

Susun Tabel Anava, termasuk nilai probabilita bagi Ho.
Jawab :
Tabel Kerja :

| Daerah | Pendapatan ($X_{ij}$) | | | | $\overline{X}_j$ |
|---|---|---|---|---|---|
| TL | 8 | 14 | | | 11 |
| U | 13 | 9 | | | 11 |
| S | 7 | 14 | 8 | 7 | 9 |
| B | 7 | 7 | 16 | | 10 |

$$\overline{\overline{X}} = \frac{\sum X_{ij}}{\sum n_j} = \frac{110}{11} = 10$$

$SS_c = \sum n_j(\overline{X}_j - \overline{\overline{X}})^2$
$= [2(11\text{-}10)^2 + (2(11\text{-}10)^2 + 4(9\text{-}10)^2 + 3(10\text{-}10)^2] = 2 + 2 + 4 + 0 = 8$
$SS_u = \sum\sum(X_{ij} - \overline{X}_j)^2$
$= \{(8\text{-}11)^2 + (14\text{-}11)^2\} + \{(13\text{-}11)^2 + (9\text{-}11)^2\} + \{(7\text{-}9)^2 + (14\text{-}9)^2 + (8\text{-}9)^2 + (7\text{-}9)^2\}$
$+ \{(7\text{-}10)^2 + (7\text{-}10)^2 + (16\text{-}10)^2\}$
$= 18 + 8 + 34 + 54 = 114$
SSt $= SSc + SSu = 8 + 114 = 122$ √
$SS_T = \sum\sum(X_{ij} - \overline{\overline{X}}_j)^2$
$= (8\text{-}10)^2 + (14\text{-}10)^2 + (13\text{-}10)^2 + (9\text{-}10)^2 + (7\text{-}10)^2 + (14\text{-}10)^2 + (8\text{-}10)^2 + (7\text{-}10)^2 +$
$(7\text{-}10)^2 + (7\text{-}10)^2 + (16\text{-}10)^2$
$= 4 + 16 + 9 + 1 + 9 + 16 + 4 + 9 + 9 + 9 + 36$
$= 122$ √

Tabel 8.5. Tabel Anava Satu Arah dengan Data Pendapatan Keluarga

| Sumber Variasi | Variasi SS (Sum of Squares) | d.f. | Varians MSS (Mean Sum of Squares) | F-rasio |
|---|---|---|---|---|
| Antar kolom, dijelaskan | $SS_c = \sum n_j(\overline{X}_j - \overline{\overline{X}})^2$ $= 8$ | c-1<br>3 | $MSS_c = \frac{SS_c}{d.f.}$<br>$= 8/3 = 2{,}67$ | F rasio = $\frac{MSS_c}{MSS_u}$<br>$= 2{,}67/16{,}29$<br>$= 0{,}1639$ |
| Dalam kolom, tidak dijelas-kan, residual | $SS_u = \sum\sum(X_{ij} - \overline{X}_j)^2$ $= 114$ | $\sum(n_j - 1)$<br>7 | $MSS_u = \frac{SS_u}{d.f.}$<br>$= 114/7$<br>$= 16{,}29$ | |
| Total | $SS_T = \sum\sum(X_{ij} - \overline{\overline{X}}_j)^2$ $= 122$ | $\sum n_j - 1$<br>10 | $\overline{\overline{X}} = \frac{\sum X_{ij}}{\sum n_j}$ | |

Nilai Probabilita (*Probability Value*) bagi Ho bila benar: > 0,25
Berarti pada tingkat signifikansi 5 %, Ho diterima. Dengan kata lain tidak ada perbedaan yang signifikan antar keempat rata-rata tersebut.
Interval keyakinan simultan juga dapat dilakukan di sini, dengan beberapa modifikasi yang perlu sebagai penyesuaian karena sampel yang tidak sama.

Latihan :
Di tahun 1977, 50 pekerja wanita dicuplik secara acak dari tiap kelompok pendidikan, dan pendapatan mereka (dalam ribuan dollar per tahun) dilaporkan sebagai berikut :

| Kelompok pendidikan | $\overline{X}$ | $\sum(X_i - \overline{X})^2$ |
|---|---|---|
| Sekolah dasar | 7,8 | 1835 |
| Sekolah menengah atas | 9,7 | 2442 |
| Akademi | 14,0 | 4707 |

Buatlah tabel Anavanya ! Apakah ada perbedaan secara nyata dalam rata-rata pendapatan antar ketiga kelompok pendidikan tersebut ?

## 8.4. Anava Dua Arah.

Dalam anava dua arah ini, kita secara sekaligus ingin membandingkan beberapa rata-rata baik antar kolom maupun antar baris. Dengan demikian sumber variasi yang dijelaskan menjadi bertambah. Misalnya dalam hal produktivitas mesin, kita dapat melihat adanya perbedaan rata-ratanya karena mesinnya yang berbeda dan juga karena operatornya berbeda.
Tabel Anava Dua Arah secara umum adalah sebagai berikut :

Tabel 8.6. Anava Dua Arah

| Sumber Variasi | Variasi SS (Sum of Squares) | d.f. | Varians MSS (Mean Sum of Squares) | F-rasio |
|---|---|---|---|---|
| Antar kolom, dijelaskan | $SS_c = r\sum(\overline{X}_j - \overline{\overline{X}})^2$ | c-1 | $MSS_c = \frac{SS_c}{d.f.}$ | F rasio = $\frac{MSS_c}{MSS_u}$ |
| Antar baris, dijelaskan | $SS_r = c\sum(\overline{X}_i - \overline{\overline{X}})^2$ | r-1 | | |
| Dalam kolom, tidak dijelas-kan, residual | $SS_u = \sum\sum(X_{ij} - \overline{X}_j - \overline{X}_i + \overline{\overline{X}})^2$ | (r-1)(c-1) | $MSS_u = \frac{SS_u}{d.f.}$ | |
| Total | $SS_T = \sum\sum(X_{ij} - \overline{\overline{X}}_j)^2$ | rc-1 | | |

$$\overline{\overline{X}} = \frac{\sum\sum X_{ij}}{\sum n_{ij}} = \frac{\sum \overline{X}_i}{\sum r} = \frac{\sum \overline{X}_j}{\sum c} \qquad SS_T = SSc + SSr + SSu$$

Contoh :
Data cuplikan tentang produktivitas per jam dari tiga mesin sejenis dengan lima operator (dalam unit) adalah sebagai berikut :

| Operator (r=1,2,…,r) | Mesin (j = 1,2, …, c) | | | $\overline{X}_i$ |
|---|---|---|---|---|
| | 1 | 2 | 3 | |
| 1 | 53 | 61 | 51 | 55 |
| 2 | 47 | 55 | 51 | 51 |
| 3 | 46 | 52 | 49 | 49 |
| 4 | 50 | 58 | 54 | 54 |
| 5 | 49 | 54 | 50 | 51 |
| $\overline{X}_j$ | 49 | 56 | 51 | 52 |

Tabel anava dua arahnya adalah :

Tabel 8.7. Anava Dua Arah untuk data pada contoh soal

| Sumber Variasi | Variasi SS (*Sum of Squares*) | d.f. | Varians MSS (*Mean Sum of Squares*) | F-*ratio* |
|---|---|---|---|---|
| Antar kolom, dijelaskan | $SS_c = r\sum(\overline{X}_j - \overline{\overline{X}})^2$ $= 130$ | c-1 $= 2$ | $MSS_c = \frac{SS_c}{d.f.}$ $= 65$ | F *ratio* (c ) $= \frac{MSS_c}{MSS_u}$ $= 23{,}6$ |
| Antar baris, dijelaskan | $SS_r = c\sum(\overline{X}_i - \overline{\overline{X}})^2$ $= 72$ | r-1 $= 4$ | $MSS_r = \frac{SS_r}{d.f.}$ $= 18$ | |
| Dalam kolom, tidak dijelaskan, *residual* | $SS_u = \sum\sum(X_{ij} - \overline{X}_j - \overline{X}_i + \overline{\overline{X}})^2$ $= 22$ | (r-1) x (c-1) $= 8$ | $MSS_u = \frac{SS_u}{d.f.}$ $= 2{,}75$ | F*ratio* (r) $= \frac{MSS_r}{MSS_u}$ $= 6{,}50$ |
| Total | $SS_T = \sum\sum(X_{ij} - \overline{\overline{X}}_j)^2$ $= 224$ | rc-1 $= 14$ | | |

Nilai probabilita (*probability value*) antar mesin < 0,0001 Nilai probabilita (*probability value*) antar operator <0,05 Jadi bila tingkat signifikansi yang digunakan bagi Ho adalah sebesar 5 %, maka kesimpulan yang dapat kita tarik adalah : Ada perbedaan yang signifikan dalam produktivitas baik antar mesin maupun antar operator...
Tetapi bila tingkat signifikansi yang digunakan sebesar 1 %, maka kesimpulannya adalah Ada perbedaan yang sangat signifikan dalam produktivitas antar mesin tetapi tidak ada perbedaan yang sangat signifikan antar operator.
Bisa juga kesimpulan tersebut dipertajam yaitu ada **perbedaan yang sangat-sangat signifikan ($\alpha$ = 0,0001) antar mesin**, dan **perbedaan yang signifikan ($\alpha$= 0,05) antar operator.**

## 8.5. Perbandingan Berganda.

Contoh :
Dari contoh pada sub 8.3 dapat dilakukan pengolahan lebih lanjut sebagai berikut. Buatlah interval keyakinan bagi $\mu_1 - \mu_2$ ; $\mu_1 - \mu_3$ dan $\mu_2 - \mu_3$. Tandailah interval keyakinan yang menunjukkan adanya perbedaan yang signifikan. Gunakan a = 5 %.

Telah dihitung dalam contoh pada sub 8.3 hal-hal berikut :
MSSu = 7,83
k = 3
$F_{0,05;(2,12)} = 3,89$

Perhitungan lanjutan yang perlu kita lakukan adalah :

$$\mu_1 - \mu_2 = (\overline{X}_i - \overline{X}_2) \pm \sqrt{(3-1)(3,89)}\sqrt{7,83}\sqrt{\frac{1}{5}+\frac{1}{5}}$$

$$= (49 - 56) \pm 4,9$$
$$= -7 \pm 4,9 \text{ *)}$$

$$\mu_1 - \mu_3 = (\overline{X}_i - \overline{X}_3) \pm \sqrt{(3-1)(3,89)}\sqrt{7,83}\sqrt{\frac{1}{5}+\frac{1}{5}}$$

$$= (49 - 51) \pm 4,9$$
$$= -2 \pm 4,9$$

$$\mu_2 - \mu_3 = (\overline{X}_2 - \overline{X}_3) \pm \sqrt{(3-1)(3,89)}\sqrt{7,83}\sqrt{\frac{1}{5}+\frac{1}{5}}$$

$$= (56 - 51) \pm 4,9$$
$$= 5 \pm 4,9 \text{ *)}$$

*) berarti antara dua $\mu$ ada perbedaan yang signifikan.
Perbandingan antara $\mu_2$-$\mu_1$, $\mu_3$-$\mu_1$, dan $\mu_3$-$\mu_2$ akan menghasilkan kesimpulan yang sama seperti di atas.

## 8.6. Perbandingan antara dua varians populasi

Distribusi F dalam hal ini digunakan untuk menguji hipotesis bahwa varians suatu populasi normal sama dengan varians dari populasi normal lainnya. Jadi, uji ini juga dapat

Langkah 3 : Tentukan statistik uji yang relevan yaitu $s_1^2/s_2^2$ yang menuruti distribusi F bila Ho benar.

Langkah 4 : Menentukan daerah kritis atau daerah penolakan Ho

Langkah 5 : Menentukan aturan keputusan untuk menolak atau menerima Ho

Langkah 6 : Menghitung rasio kritis

## 8.7. Soal.

1. Dari suatu percobaan tentang penggunaan tiga jenis pupuk masing-masing pada empat petak lahan untuk bertanam kentang, diperoleh data sebagai berikut :
   Angka dalam sel tabel menunjukkan produksi kentang yang diukur dalam satuan kilogram.

| | Pupuk | | |
|---|---|---|---|
| Petak | A | B | C |
| 1 | 69 | 72 | 60 |
| 2 | 75 | 74 | 64 |
| 3 | 70 | 78 | 65 |
| 4 | 66 | 68 | 55 |

   Apakah ada perbedaan secara nyata dalam produksi kentang karena pemakaian jenis pupuk maupun karena petak tanah ?

2. Tiga pekerja menjalankan tugas yang sama yaitu membungkus kotak. Jumlah kotak yang berhasil dibungkusnya dalam tiga periode waktu yang terpilih adalah sebagai berikut :

| | Pekerja | | |
|---|---|---|---|
| Periode waktu | X | Y | Z |
| 11 - 12 pagi | 24 | 19 | 20 |
| 1 - 2 siang | 23 | 17 | 14 |
| 4 - 5 sore | 25 | 21 | 17 |

   Dari data di atas, dengan $\alpha = 5\ \%$ :
   a. Apakah ada perbedaan produktivitas antar pekerja ?
   b. Apakah ada perbedaan produktivitas antar periode ?
   c. Buatlah interval keyakinan berganda baik antar pekerja maupun antar periode waktu.

# POKOK BAHASAN IX
# UJI - CHI KUADRAT

## 9.1. Pendahuluan.

Pokok bahasan ini merupakan salah satu aplikasi atau rincian dari pengujian hipotesis. Oleh karena itu pemahaman tentang langkah-langkah umum dalam pengujian hipotesis sangat diperlukan. Uji chi-kuadrat menyediakan uji sederhana berdasar pada perbedaan frekuensi observasi dan frekuensi harapannya. Karena uji ini sangat mudah untuk dipahami dan dihitung, maka uji ini sangat populer dalam pengujian hipotesis. Juga, karena hanya sedikit asumsi tentang populasinya, uji ini umumnya dikelompokkan dalam uji non parametrik.

Adapun pokok bahasan ini mencakup :

a. Pengujian kompatibilitas
b. Pengujian independensi,
c. Pengujian homogenitas, dan
d. Pengujian parameter deviasi standar.

## 9.2. T I U dan Sasaran belajar

T I U :

Setelah mempelajari pokok bahasan ini mahasiswa diharapkan mampu memahami dan menerapkan uji-chi kuadrat pada berbagai keadaan serta melakukan pengambilan kesimpulannya.

Sasbel :

Mahasiswa diharapkan mampu :

a. Menjelaskan lima langkah yang menunjukkan prosedur pengujian kompatibilitas.
b. Melakukan pengujian kompatibilitas (*test of goodness of fit*).
c. Menyusun distribusi frekuensi harapan dari distribusi frekuensi observasi.
d. Melakukan pengujian hipotesis tentang independensi (*test of independency*).
e. Melakukan pengujian hipotesis tentang homogenitas.
f. Melakukan pengujian hipotesis tentang deviasi standar populasi.

Dimana : $o_i$ adalah frekuensi observasi $e_i$ adalah frekuensi harapan

Prosedur pengujian kompatibilitas ini adalah sebagai berikut

1. Nyatakan Ho dan Ha
2. Tentukan tingkat signifikansi ($\alpha$)
3. Tentukan statistik uji Kai-kuadrat dan d.f nya
4. Tentukan daerah penolakan Ho
5. Hitung $\chi^2$ dan tentukan ditolak/diterimanya Ho.

Contoh :

Misal kita akan menguji hipotesis nol tentang tingkat kelahiran yang merata sepanjang tahun dari negara X. Untuk itu data yang tersedia hanyalah cuplikan acak sebanyak 88 kelahiran dikelompokkan dalam musiman yang tidak sama panjang. Data observasinya adalah sebagai berikut :

| Musim | Spring Apr-June | Summer July-Aug | Fall Sept-Oct | Winter Nov-March | |
|---|---|---|---|---|---|
| Jumlah (hari) | (91) | (62) | (61) | (151) | (365) |
| Frekuensi Observasi | 27 | 20 | 8 | 33 | 88 |

Bagaimana tingkat kompatibilitas atau kesesuaian data tersebut dengan hipotesis nol di atas ?

Jawab :

Ho : Tidak ada perbedaan kelahiran antar musim (Kelahiran adalah merata sepanjang tahun)

Ha : Ada perbedaan kelahiran antar musim. (Kelahiran adalah tidak merata sepanjang tahun)

$\alpha = 5$ %

Kai-kuadrat tabel pada tingkat signifikansi ($\alpha$) sebesar 5% dengan d.f.= k-1= 3 adalah 7,81

Tabel kerja untuk mencari chi-kuadrat adalah sebagai berikut.

| | Frekuensi | Probabi | Frekuensi | | Kuadrat simpang |
|---|---|---|---|---|---|

Frekuensi harapan ($e_i$) dihitung dengan cara :

Spring = 0,25 x 88 = 22,0

Summer = 0,17 x 88 = 15,0 dst

Pada $\alpha$ = 5 % ternyata bahwa $\chi^2$hitung lebih kecil dari pada $\chi^2$ tabel, berarti Ho diterima.

Jadi Ho yang menyatakan tidak ada perbedaan kelahiran antar musim diterima.

## 9.4. Pengujian Independensi.

Pengujian independensi bertujuan untuk menguji apakah antara dua atribut dari populasi atau cuplikan independen atau tidak. Pengujian ini tidak menyatakan derajat asosiasi atau arah dependensinya. Tabel silang dipergunakan sebagai tabel kerja dalam pengujian ini. Sebuah pertanyaan dalam penelitian yang sering muncul adalah "apakah dua variabel yang diamati saling berhubungan ?" Misal : Apakah harga/nilai perabot rumahtangga yang dibeli berkaitan dengan mutunya.
Dua variabel disebut independen atau tidak berkaitan bila distribusi yang satu sama sekali tidak bergantung (tidak dipengaruhi/ oleh tidak berhubungan dengan) distribusi yang lain.

Tabel kontingensi r X c

| Klasifikasi B | Klasifikasi A | | | | Jumlah |
|---|---|---|---|---|---|
| | A1 | A2 | A.. | Ac | |
| B1 | $n_{11}$ | $n_{12}$ | | $n_{1c}$ | $\sum n_{1.}$ |
| B2 | $n_{21}$ | | | | $\sum n_{2.}$ |
| B.. | | $n_{.2}$ | | | |
| Br | | | | $n_{rc}$ | $\sum n_{r.}$ |
| Jumlah | $\sum n_{.1}$ | $\sum n_{.2}$ | | $\sum n_{.c}$ | $\sum\sum n_{rc}$ |

Kai-kuadrat =

$$\chi^2 = \sum\sum\frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Dimana : $e_{ij} = \frac{(\sum n_{i.})(\sum n_{.j})}{\ldots}$

cuplikan sebanyak 400 orang konsumen secara acak, dan hasil observasinya dapat dilihat dalam tabel berikut.

| Merek barang | Tingkat pendapatan | | | |
|---|---|---|---|---|
| | Tinggi | Sedang | Rendah | Jumlah |
| ABC | 40 | 50 | 10 | 100 |
| ARC | 50 | 90 | 10 | 150 |
| IBM | 70 | 60 | 20 | 150 |
| Jumlah | 160 | 200 | 40 | 400 |

Ujilah hubungan tersebut dengan $\alpha$ 5 %.

Pemecahan :

1. Ho : Tidak ada hubungan antara tingkat pendapatan dan merek barang yang dibelinya.

   Hi : Ada hubungan antara tingkat pendapatan dan merek barang yang dibelinya.

2. $\alpha = 5\,\%$ ; d.f. (3-1)(3-1) = 4.

   Chi-kuadrat tabel = 9,4877

3. $\chi^2 = \sum\sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$

4. Tabel kontingensi 3 X 3

| Merek barang | Tingkat pendapatan | | | Jumlah |
|---|---|---|---|---|
| | Tinggi | Sedang | Rendah | |
| ABC | 40 (40) | 50 (50) | 10 (10) | 100 |
| ARC | 50 (60) | 90 (75) | 10 (15) | 150 |
| IBM | 70 (60) | 60 (75) | 20 (15) | 150 |
| Jumlah | 160 | 200 | 40 | 400 |

Probabilita $P_{i.}$ dan $P_{.j}$ tidak diketahui, diestimasikan dengan

$$P_{.} = \frac{N_i}{\quad} \qquad P_{:} = \frac{N_j}{\quad}$$

$$\chi^2 = \frac{(40-40)^2}{40} + \frac{(50-50)^2}{50} + \frac{(10-10)^2}{10} + \frac{(50-60)^2}{60} + \frac{(90-75)^2}{75} +$$

$$\frac{(10-15)^2}{15} + \frac{(70-60)^2}{60} + \frac{(60-75)^2}{75} + \frac{(20-15)^2}{15} =$$

$$= 0{,}0 + 0{,}0 + 0{,}0 + 1{,}67 + 3{,}0 + 1{,}67 + 1{,}67 + 3{,}0 + 1{,}67 = 12{,}67.$$

6. Kai-kuadrat hit = 12,67 > Kai-kuadrat tabel = 9,4877 , berarti Ho ditolak. Jadi pernyataan bahwa ada hubungan antara tingkat pendapatan dan merek barang yang dibelinya dapat diterima.

## 9.5. Pengujian homogenitas.

Kegunaan uji kai-kuadrat yang lain adalah untuk mendeteksi apakah suatu cuplikan yang diambil dari masing-masing populasi yang berbeda dapat dipandang homogen atau tidak. Walaupun dalam melakukan perhitungan adalah sama dengan perhitungan dalam uji independensi, tetapi pola pemikiran yang mendasarinya berbeda.

Pertama : Dalam uji independensi, frekuensi harapan $e_{ij}$ diasumsikan kedua populasi bebas;
Dalam uji homogenitas, frekuensi harapan $e_{ij}$ diasumsikan populasi-populasi homogen, maka :
Bila benar HOMOGEN cuplikan-cuplikan tersebut dapat dianggap berasal dari populasi yang sama, jadi dapat dianggap sebagai cuplikan tunggal.

Kedua : Dalam pengumpulan data, Untuk uji independensi diambil hanya cuplikan tunggal dari satu populasi, baru kemudian diklasifikasikan subjek-subjek tersebut secara silang. Untuk uji homogenitas ditentukan dahulu dua atau lebih populasi dan dari tiap populasi tersebut diambil cuplikan, kemudian masing-masing cuplikan diuji homogenitasnya setelah ditentukan kategori klasifikasinya.

Contoh :
Seorang peneliti lingkungan hidup ingin mengetahui reaksi terhadap pertanyaan :"Adakah masalah pencemaran udara di sekitar anda ?". Untuk itu peneliti tersebut mengambil 3 daerah yang berbeda (3 populasi) dari mana diambil

Gunakan $\alpha$ 5 % untuk membantu peneliti tersebut dalam melakukan pengujian hipotesis.

.

Pemecahan :

1. Ho: Ketiga populasi masyarakat (semarang, Jakarta dan Surabaya) homogen dalam hal pengetahuannya tentang ada tidaknya masalah pencewmaran udara.

   Hi : Ketiga populasi tersebut tidak homogen.

2. $\alpha = 5$ % ; d.f.(3-1)(4-1) = 6. Kai-kuadrat tabel = 12,5916

3. $$\chi^2 = \sum\sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

4. Tabel kontingensi 3 X 4 adalah sebagai berikut

Pengetahuan Masyarakat tentang Pencemaran Udara

| Daerah | Pebgetahuan tentang adanya pencemaran | | | | Jumlah |
|---|---|---|---|---|---|
| | Tidak | Ya | Ragu-ragu | Tidak tahu | |
| Semarang | 5 (8,667) | 31 (24) | 2 (4,333) | 2 (3) | 40 |
| Jakarta | 10 (8,667) | 21 (24) | 4 (4,333) | 5 (3) | 40 |
| Surabaya | 11 (8,667) | 20 (24) | 7 (4,333) | 2 (3) | 40 |
| Jumlah | 26 | 72 | 13 | 9 | 120 |

$$\chi^2 = \frac{(5-8{,}667)^2}{8{,}667} + \frac{(31-24)^2}{24} + \frac{(2-4{,}333)^2}{4{,}333} + \frac{(2-3)^2}{3} + \frac{(10-8{,}667)^2}{8{,}667} + \frac{(21-24)^2}{24}$$

## 9.6. Penyesuaian bilangan nij < 5

Rumus chi-kuadrat yang berikut

$$\chi^2 = \sum\sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

hanya berlaku bila nij lebih besar atau sama dengan 5 untuk semua i dan j. Toleransi adanya $n_{ij} < 5$ maksimum sebanyak 20% dari seluruh $n_{ij}$. Oleh karena itu bila ada nij < 5 yang melebihi batas toleransi tersebut, kita harus mengadakan penggabungan suatu kelas/kategori dengan kelas yang lain yang sesuai sehingga dapat diperoleh kelas/kategori baru dengan frekuensi masing-masing $\geq 5$..

Contoh :

Penggabungan kolom.

| | A1 | A2 | A3 | A4 | Jumlah |
|---|---|---|---|---|---|
| B1 | 40 | 50 | 6 | 4 | 100 |
| B2 | 50 | 90 | 7 | 3 | 150 |
| B3 | 70 | 60 | 17 | 3 | 150 |
| | 160 | 200 | 30 | 10 | 400 |

Dimensi : 3 X 4

| A1 | A2 | A3 & A4 | Jumlah |
|---|---|---|---|
| 40 | 50 | 10 | 100 |
| 50 | 90 | 10 | 150 |
| 70 | 60 | 20 | 150 |
| 160 | 200 | 40 | 400 |

Dimensi : 3 X 3

Kita dapat menggabungkan A4 dengan A3 atau A yang lain, asal dapat menjadi kelas yang baru dan secara teoritis ada maknanya. .

Penggabungan baris.

| | A1 | A2 | A3 | A4 | Jumlah |
|---|---|---|---|---|---|
| B1 | 40 | 50 | 6 | 4 | 100 |
| B2 | 50 | 90 | 7 | 3 | 150 |
| B3 | 70 | 60 | 17 | 3 | 150 |
| | 160 | 200 | 30 | 10 | 400 |

Hasil penggabungan misalnya dapat menjadi :

| | A1 | A2 | A3 | A4 | Jumlah |
|---|---|---|---|---|---|
| B1&B2 | 90 | 140 | 13 | 7 | 250 |
| B3 | 70 | 60 | 17 | 3 | 150 |
| | 160 | 200 | 30 | 10 | 400 |

## 9.7. Pengujian Parameter Deviasi Standar, $\sigma_X$.

Bila deviasi standar populasi tidak diketahui, kita biasanya menggunakan deviasi standar cuplikan sebagai penduga bagi deviasi standar populasi tersebut. Dalam sub pokok bahasan ini kita akan menguji apakah deviasi standar populasi sama dengan suatu nilai yang diperkirakan atau tidak. Statistik uji dujudkan dalam rumus sebagai berikut :

$$u = \frac{(n-1)s^2}{\sigma_X^2} \qquad d.f = (n\text{-}1).$$

Statistik uji tersebut yang kita perbandingkan dengan $\chi^2$ tabel pada $\alpha$ dan d.f tertentu, untuk menentukan apakah Ho yang diajukan diterima atau ditolak. Kriteria penolakan Ho sama seperti uji hipotesis yang lain.

Contoh :

Berdasarkan pengalaman yang bertahun-tahun lamanya, manajemen suatu Badan Penerbitan mengetahui bahwa deviasi standar kerusakan pencetakan sebanyak 1.000 lembar kertas HVS ukuran kwarto ialah sebesar 9 lembar. Pada akhir-akhir ini, ketika diadakan pengambilan cuplikan secara acak sebesar n = 25 ternyata s = 12 (dari tiap 1000 lembar pencetakan kertas HVS). Apakah ada alasan guna menolak Ho : Deviasi standar populasi, $\underline{\sigma_X} = 9$ ?

Pemecahan :

1. Ho : $\underline{\sigma_X} = 9$ Hi : $\underline{\sigma_X} \neq 90$

2. $\alpha = 5$ %, uji dua arah d.f = n-1 = 24.
   $\chi^2_{0,05;\,24} = 39,4$

3. Statistik uji u adalah
$$u = \frac{(n-1)s^2}{\sigma_X^2} = \frac{(25-1)(12)^2}{9^2}$$

banyak perusahaan yang mengklaim bahwa tingkat kecelakaan yang berat dan fatal lebih banyak dialami oleh mobil-mobil berukuran lebih besar. Untuk itu dilakukan penelitian, dan dari data kecelakaan yang tercatat di Polantas selama 1 tahun sbb.

Kondisi kecelakaan dan ukuran mobil

| Kondisi kecelakaan | Ukuran mobil | | | |
|---|---|---|---|---|
| | Kecil | Sedang | Besar | Super besar |
| Fatal atau luka kritis | 36 | 16 | 14 | 34 |
| Tidak fatal atau luka kritis | 64 | 34 | 20 | 82 |
| Tidak fatal atau luka ringan | 50 | 50 | 16 | 84 |

Dengan menggunakan $\alpha = 5$ %, ujilah pendapat perusahaan asuransi : kondisi kecelakaan dipengaruhi ukuran mobil.

2 . Sebuah toko serba ada ingin sekali mengetahui pola pembungkus yang disukai pembelanja. Seksi pemasaran mengadakan wawancara dengan 200 pembelanja serta menunjukkan 4 macam pola pembungkus yang berbeda. Hasil observasi sedemikian itu diberikan dalam tabel di bawah ini.

| Pola pembungkus | A | B | C | D |
|---|---|---|---|---|
| Jumlah pembelanja yang menyukai | 33 | 42 | 67 | 58 |

Apakah ada alasan guna menganggap bahwa preferensi pembelanja terhadap keempat pola pembungkus di atas tidak berbeda ? Gunakan $\alpha = 5$ % dan 1 %.

## 9.9. Rangkuman.

Uji chi-kuadrat merupakan bentuk pengujian yang bebas sebaran (non parametrik) namun mempunyai banyak manfaat, antara lain untuk menguji kompatibilitas, independensi, homogenitas dan parameter deviasi standar. Pada umumnya dalam pengujian ini diperlukan tabel kontingensi sebagai tabel kerjanya. Secara umum statistik uji dapat dirumuskan sebagai berikut :

$$\text{Kai} - \text{kuadrat} = \chi^2 = \sum\sum \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

## POKOK BAHASAN X
## STATISTIKA NON-PARAMETRIK

1. Pendahuluan

Pokok bahasan ini menyediakan alat analisis statistika yang menyangkut pengujian hipotesis. Oleh karena itu pemahaman tentang langkah-langkah pengujian hipotesis yang berlaku umum harus telah dikuasai dengan baik oleh para mahasiswa.

Secara terperinci uraian dalam statistika non-parametrik ini mencakup berbagai macam uji yang dapat dikelompokkan ke dalam :

a. Kasus satu cuplikan
b. Kasus dua cuplikan
c. Kasus k-cuplikan, dan
d. Pengukuran korelasi dan uji signifikansinya

2. TIU dan Sasbel

T I U :

Setelah mempelajari pokok bahasan ini mahasiswa dapat memahami dan menerapkan berbagai pengujian hipotesis yang tercakup dalam statistika non-parametrik.

Sasbel :

Mahasiswa setelah mengikuti pokok bahasan ini diharapkan mampu :

1. Menjelaskan beberapa macam pengujian statistika non-parametrik dengan kasus satu cuplikan.
2. Melakukan aplikasi pengujian binomial pada kasus satu cuplikan.
3. Melakukan aplikasi pengujian kai-kuadrat pada kasus satu cuplikan.
4. Melakukan aplikasi pengujian Kolmogorov-Smirnov pada kasus satu cuplikan.
5. Melakukan aplikasi pengujian urutan pada kasus satu cuplikan.
6. Menjelaskan beberapa macam pengujian statistika non-parametrik dengan kasus dua cuplikan baik yang berpasangan maupun yang bebas.
7. Melakukan aplikasi pengujian statistika non-parametrik kasus dua cuplikan berpasangan.
8. Melakukan pengujian statistika non-parametrik kasus dua cuplikan bebas.
9. Menjelaskan beberapa macam pengujian statistika non-parametrik dengan kasus k-cuplikan.
10. Melakukan aplikasi pengujian statistika non-parametrik kasus k-cuplikan berpasangan.
11. Melakukan pengujian statistika non-parametrik kasus k-cuplikan bebas.
12. Menjelaskan dan mengaplikasi pengukuran koefisien kontingensi dan uji siginifikansinya.
13. Menjelaskan dan mengaplikasi korelasi tata jenjang Spearman dan uji signifikansinya.

statistika yang tidak memerlukan anggapan tertentu mengenai bentuk sebaran atau parameter dari variabel random yang diselidiki disebut metode statistika non-parametrik atau statistika bebas sebaran.

10.4.. Kasus satu cuplikan

a. Uji Binomial

Arti dan kegunaan

Uji binomial adalah suatu uji non-parametrik yang berkaitan dengan variable yang bersifat diskrit dan hanya mempunyai dua alternative nilai, seperti rusak dan tidak rusak, lulus dan tidak lulus, pria dan wanita, dst.

Metoda

Probabilita untuk memperoleh x obyek dalam satu kategori dan n-x obyek dalam kategori lain adalah sebesar

$$p(x) = nCx.P^{x}Q^{N-x}$$

Di mana : P = proporsi kasus dalam satu kategori, dan

Q = 1-P = proporsi kasus dalam kategori lainnya.

$$nCx = \frac{n!}{x!(n-x)!}$$

Prosedur pengujian

1. Nyatakan Ho dan Hi
2. Tentukan statistiK uji yang dipakai dan besar sampel (n)
3. Tentukan tingkat signifikansi ($\alpha$) yang dipilih
4. Tentukan daerah kritis sebagai daerah penolakan Ho
   Sampel kecil → Gunakan Tabel D. ***Table of Probabilities Associated with Values as small as Observed Values of x in the Binimial Test.***
   Sampel besar → Gunakan pendekatan normal, dengan rata-rata $\mu = np$ dan $\sigma = \sqrt{np(1-p)}$
5. Lakukan perhitungan dan bandingkan hasil perhitungan dengan daerah kritis, dengan langkah-langkah sebagai berikut
   5a. Tentukan besar sample n, jumlah kasus yang akan diamati;
   5b. Hitung frekuensi observasi utk masing-masing kategori;
   5c. Cari besar probabilita terjadinya Ho berdasar hasil observasi, seperti pada butir 4.

Sampel tersebut termasuk besar yaitu sebesar 243 + 682 = 925 buah, bisa digunakan pendekatan normal (Tabel Z).

Ho : P = ¾

Hi : P ≠ ¾

Tingkat signifikansi ($\alpha$) dipilh sebesar 5 %

T = jumlah tanaman yang masuk "Klas I" yaitu sebesar 682.

Daerah kritis berhubungan dengan nilai T lebih kecil atau sama dengan $t_1$, di mana

$$t_1 = np^* - Z_{0,025}\sqrt{np^*(1-p^*)}$$

$$= (925)(3/4)\ - 1{,}96\surd(925(3/4)(1/4)$$

$$= 667{,}95$$

dan semua nilai T lebih besar dari $t_2$, di mana

$$t_2 = np^* + Z_{0,025}\sqrt{np^*(1-p^*)}$$

$$= (925)(3/4) + 1{,}96\surd(925)(3/4)(1/4)$$

$$= 719{,}55$$

berarti nilai T berada di dalam interval 667,95 < T < 719,55, sehingga Ho kita terima. Jadi pernyataan teori sederhana tersebut beralasan.

Kekuatan uji

Uji binomial ini sangat cocok untuk menguji variable yang diukur dengan skala nominal dan bersifat dikotomis, terlabih lagi bila sample yang digunakan sangat kecil.

b. Uji Kai-kuadrat

Arti dan kegunaan

Uji Kai-kuadrat mrupakan salah satu uji non-parametrik yang berkaitan dengan variable yang diukur dengan skala nominal dengan dua alternatif nilai atau lebih. Uji ini dapat digunakan untuk menguji homogenitas suatu populasi, pemerataan antar kategori, hubungan antara dua variable dan uji terhadap varians atau deviasi standar.

Metoda

Metoda dalam uji Kai-kuadrat ini tergantung pada penggunaannya. Namun secara umum besarnya nilai Kai-kuadrat dapat diperoleh dengan mencari kuadraat dari selisih frekuensi observasi dengan frekuensi harapannya secara relative (dibandingkan dengan frekuensi harapan).

Rumus Kai-kuadrat, selanjutnya disebut Kai-kuadrat hitung ($\chi^2_{hit}$), dalam uji homogenitas dan

Kai-kuadrat hitung ini kemudian dibandingkan dengan Kai-kuadrat table pada $\alpha$ dan d.f yang tertentu. Bila $\chi^2_{hit} > \chi^2_{tabel}$ maka Ho ditolak, yang berarti Hi yang diterima.

Sedangkan rumus Kai-kuadrat untuk menguji hubungan antara dua variable dengan table silang adalah

$$\chi^2 = \sum\sum\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Prosedur pengujian

1. Nyatakan Ho dan Hi nya, yaitu
   Ho : tidak ada hubungan antara variable baris dan variable kolom, atau variable baris independent dari variable kolom.
   Hi : ada hubungan antara variable baris dan variable kolom, atau variable baris tidak independen dari variable kolom.
2. Tentukan tingkat signifikansi yang akan dipakai, umumnya 5% atau 1%.
3. Gambarkan daerah kritis atau daerah penolakan Ho
4. Nyatakan aturan keputusan untuk menolak Ho atau menerima Hi.
5. Kumpulkan data sample dan lakukan perhitungan secukupnya
6. Hitung rasio kritis (
7. Bandingkan rasio kritis ($\chi^2_{hit}$) dengan aturan keputusan.
8. Tarik kesimpulan.

Contoh
Kekuatan uji

c. Uji Kolmogorov-Smirnov
Arti dan kegunaan
Metoda
Prosedur pengujian
Contoh
Kekuatan uji

A. Cuplikan berpasangan
  a. Uji McNemar
  b. Uji tanda
  c. Uji Wilcoxon
  d. Uji Walsh
  e. Uji Randomisasi
B. Cuplikan bebas
  a. Uji Fisher
  b. Uji Kai-kuadrat
  c. Uji Median
  d. Uji U Mann-Whitney
  e. Uji Kolmogorov-Smirnov
  f. Uji Urutan Wald-Wolfowitz
  g. Uji Reaksi ekstrim Moses
  h. Uji Randomisasi

10.6. Kasus k-cuplikan
A. Cuplikan berpasangan
  a. Uji Q Cochran
  b. Uji Anava tata jenjang Friedman
B. Cuplikan bebas
  a. Uji Kai-kuadrat

  b. Uji Median (Perluasan)
  c. Uji Anava tata jenjang satu arah Kruskal-Walis

10.7. Pengukuran korelasi dan uji signifikansinya
a. Koefisien Kontingensi

$$KK = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

b. Korelasi tata jenjang Spearman

$$R_S = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

## DAFTAR PUSTAKA


1. Anderson David R., Dennis J. Sweeney and Thomas A. Williams, **Statistics for business and economics**, South-Western Thomson Learning, 2002.

2. Andi Hakim Nasution dan Barizi, **Metode Statistika untuk penarikan kesimpulan**, PT Gramedia, Jakarta, 1979.

3. Anto Dayan, **Pengantar Metode Statistik Jilid II**, LP3ES,1986

4. Blalock Jr. Hubert M., **Social Statistics**, Rev. Second Ed.,McGraw Hill, 1979.

5. Conover William Jay, **Practical Nonparametric Statistics**, John Wiley & Sons, 1971.

6. Iman Ronald L. dan W. J. Connover, **Modern Business Statistics**, John Wiley & Sons, 1983.

7. Mason Robert D. Douglas A Lind, William G Marchal, **Statistical techniques in business and economics**, tenth ed., international Ed., the Mc Graw-Hill Companies, Inc. 1999

8. Lind, Douglas A., Robert D. Mason and William G. Marchal, **Basic Statistics for business and economics**, third ed., Mc Graw-Hill International Edition, 2000

9. Keller Gerald, **Statistics for management and economics**, seventh ed., Thomson Book/Cole, 2005.

10. Sanders Donald H., **Statistics A Fresh Approach**, Fourth Ed, McGraw-Hill Publishing Company, 1990.

11. Siegel Sydney, **Non Parametric Statistics for the behavioral sciences**, McGrawhill International Book Company, 1956.

12. Wonnacott Thomas H. dan Ronald J. Wonnacott, **Introductory Statistics for Business and Economics**, Third Ed., John Wiley & Sons, 1984.

13. Zanten Wim Van, **Statistik untuk Ilmu-ilmu Sosial**, PT Gramedia, 1980.