

MODUL PEMBELAJARAN

BIOSTATISTIK

KODE MK: WP102

Dosen Pengampu:

Dr. Mitra, S.KM., M.KM.

Dr. drg. Oktavia Dewi, M.Kes

**PROGRAM STUDI MAGISTER ILMU KESEHATAN MASYARAKAT
SEKOLAH TINGGI ILMU KESEHATAN HANG TUAH PEKANBARU**



VISI, MISI, DAN TUJUAN
SEKOLAH TINGGI ILMU KESEHATAN HANG TUAH PEKANBARU

VISI

Terwujudnya institusi yang unggul dan kompeten dalam penyelenggaraan pendidikan tinggi kesehatan yang berbasis teknologi tepat guna dan mampu bersaing di tingkat nasional dan regional tahun 2036.

MISI

1. Melaksanakan pendidikan kesehatan yang bermutu untuk menghasilkan lulusan yang profesional berorientasi kepada peningkatan kesehatan masyarakat;
2. Mengembangkan ilmu dan teknologi kesehatan melalui kegiatan penelitian dan pengabdian kepada masyarakat untuk peningkatan kesehatan masyarakat;
3. Melaksanakan penjaminan mutu yang konsisten dan berkelanjutan;
4. Mewujudkan admosfir akademik yang kondusif melalui kinerja akademik;
5. Menjaln kerjasama dengan berbagai pihak dalam penyelenggaraan Tridharma Perguruan Tinggi;
6. Mengembangkan teknologi tepat guna dalam bidang kesehatan.

TUJUAN

1. Menghasilkan lulusan di bidang ilmu kesehatan yang mandiri dan profesional berdaya saing tinggi melalui sistem pendidikan ilmu kesehatan;
2. Menghasilkan lulusan yang mempunyai kemampuan di bidang penelitian dan hasil penelitiannya digunakan untuk pengabdian kepada masyarakat dalam rangka menunjang kesehatan masyarakat;
3. Terwujudnya pengembangan ilmu dan teknologi melalui publikasi ilmiah;
4. Terwujudnya tata kelola penyelenggaraan institusi yang transparan, akuntabel, dan adil melalui sistem penjaminan mutu yang konsisten dan berkelanjutan;
5. Tersedianya infrastruktur dan iklim akademik yang kondusif bagi penyelenggaraan Tridharma Perguruan Tinggi;
6. Tercapainya kerjasama dengan berbagai pihak dalam penyelenggaraan Tridharma Perguruan Tinggi.

VISI, MISI, DAN TUJUAN
PROGRAM STUDI MAGISTER ILMU KESEHATAN MASYARAKAT

VISI

Menjadi Program Studi Magister Kesehatan Masyarakat yang unggul dan kompeten berbasis teknologi tepat guna dalam mengelola kesehatan masyarakat serta dapat bersaing di Tingkat Nasional dan Regional tahun 2036.

MISI

Program Studi Magister Kesehatan Masyarakat Hang Tuah Pekanbaru mendidik sarjana kesehatan strata 1 atau setingkat menjadi tenaga kesehatan strata 2 dengan kompetensi kesehatan masyarakat yang profesional, inovatif dan teruji, tepat guna melalui pendidikan, penelitian, dan pengabdian kepada masyarakat.

TUJUAN

1. Menghasilkan lulusan yang professional dalam pengembangan ilmu kesehatan masyarakat menggunakan teknologi tepat guna;
2. Menghasilkan karya penelitian yang berkualitas, inovatif dan teruji dan di publikasikan pada jurnal nasional dan internasional;
3. Menghasilkan karya pengabdian kepada masyarakat sesuai dengan kebutuhan dan permasalahan yang ada pada masyarakat;
4. Terwujudnya penjaminan mutu yang konsisten dan berkelanjutan untuk meningkatkan mutu program studi;
5. Terjalinnnya kerja sama dengan intitusi dalam dan luar negeri untuk meningkatkan kualitas lulusan dalam rangka peningkatan kesehatan masyarakat.

BAB I

PENDAHULUAN

A. IDENTITAS MATA KULIAH

Nama Mata Kuliah : Biostatistik
Kode Mata Kuliah : WP102
Beban SKS : 2 (Dua) SKS
Semester : II
Peminatan : Kelas Umum
Mata kuliah prasyarat : -

B. DESKRIPSI MATA KULIAH

Biostatistik adalah salah satu ilmu dasar dari ilmu kesehatan. Seorang ahli kesehatan, baik ilmuwan maupun praktisi dapat dipastikan ia akan menggunakan metode-metode biostatistik dalam menjalankan dan mengembangkan profesinya sehari-hari. Mata ajaran ini memberikan pemahaman tentang konsep dasar Biostatistik dan kemampuan dalam aplikasi Biostatistik. Pada mata kuliah ini menjelaskan tentang konsep dan peranan statistic dibidang kesehatan, memahami konsep sampling dan estimasi, memahami konsep pengujian hipotesis, memahami dan melakukan analisis bivariate baik statistic parametric dan non parametric, melakukan analisis multivariate dan melakukan telaah kritis terhadap artikel hasil penelitian kuantitatif .

C. CAPAIAN PEMBELAJARAN LULUSAN

1. Sikap

- a. Bertaqwa kepada Tuhan Yang Maha Esa.
- b. Memiliki moral, etika dan kepribadian yang baik di dalam menyelesaikan tugasnya.
- c. Mampu bekerja sama dan memiliki kepekaan sosial dan kepedulian yang tinggi terhadap masyarakat dan lingkungannya.
- d. Menghargai keanekaragaman budaya, pandangan, kepercayaan, dan agama serta pendapat/temuan original orang lain.
- e. Menjunjung tinggi penegakan hukum serta memiliki semangat untuk mendahulukan kepentingan bangsa serta masyarakat luas.
- f. Mampu melaksanakan tugas sederhana, terbatas, bersifat rutin

- g. Menunjukkan sikap bertanggungjawab atas pekerjaan di bidang keahliannya secara mandiri.

2. Pengetahuan

Menguasai konsep teoritis desain dan analisis epidemiologi untuk mampu memformulasi penyelesaian masalah kesehatan secara prosedural.

3. Keterampilan Umum

- a. Mampu menerapkan pemikiran logis, kritis, sistematis, dan inovatif dalam konteks pengembangan atau implementasi ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora yang sesuai dengan bidang kesehatan masyarakat.
- b. Mampu mengambil keputusan dalam konteks menyelesaikan masalah pengembangan ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora berdasarkan kajian analisis terhadap informasi dan data.

4. Keterampilan Khusus

Kemampuan untuk mengembangkan pengetahuan ilmu kesehatan masyarakat dan kajian biostatistik sebagai metode dan alat dalam pengambilan keputusan dalam riset.

D. CAPAIAN PEMBELAJARAN MATA KULIAH

Setelah mengikuti kuliah ini, mahasiswa :

1. Mampu memahami konsep dan peranan statistik di bidang kesehatan
2. Mampu memahami konsep probabilitas
3. Mampu memahami konsep dan melakukan perhitungan distribusi sampling dan estimasi
4. Mampu memahami konsep hipotesis dan melakukan pengujian hipotesis parametrik dan non parametrik
5. Mampu menganalisis confounding dan interaksi
6. Mampu menganalisis data melalui regresi logistic ganda
7. Mampu menganalisis data melalui regresi linear ganda
8. Mampu melakukan telaah kritis penelitian kesehatan

BAB II

KEGIATAN PEMBELAJARAN

A. KEGIATAN PEMBELAJARAN 1

Pembelajaran 1 dilakukan melalui pendekatan *contextual learning*. Dosen :Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata. Pembelajaran dilakukan selama 2 x 50 menit.

B. KEGIATAN PEMBELAJARAN 2

Pembelajaran 2 dilakukan melalui pendekatan Contextual Instruction dan Cooperative Learning. Dosen : Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, Menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa secara kelompok, Memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, Mambahas/menyimpulkan masalah/tugas yang diberikan dosen secara berkelompok Pembelajaran dilakukan selama 2 x 50 menit.

C. KEGIATAN PEMBELAJARAN 3.

Pembelajaran 3 dilakukan melalui pendekatan Contextual Instruction dan Cooperative Learning. Dosen : Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, Menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa secara kelompok, Memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, Mambahas/menyimpulkan masalah/tugas yang diberikan dosen secara berkelompok. Pembelajaran dilakukan selama 2 x 50 menit.

D. KEGIATAN PEMBELAJARAN 4

Pembelajaran 4 dilakukan melalui pendekatan Contextual Instruction dan Cooperative Learning. Dosen :Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, Menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa secara kelompok, Memeriksa dan memberi usulan terhadap hasil belajar

mandiri mahasiswa. Mahasiswa :Membahas konsep teori dan kaitannya dengan situasi nyata, Mencari, mengumpulkan dan menyusun informasi terkait dengan perhitungan yang tepat. Pembelajaran dilakukan selama 2 x 50 menit.

E. KEGIATAN PEMBELAJARAN 5

Pembelajaran 5 dilakukan melalui pendekatan Contextual Instruction. Dosen : Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata. Pembelajaran dilakukan selama 2 x 50 menit.

F. KEGIATAN PEMBELAJARAN 6

Pembelajaran 6 dilakukan melalui pendekatan Contextual Instruction dan Cooperative Learning. Dosen :Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, Menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa, Memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, Mencari, mengumpulkan dan menyusun informasi terkait dengan perhitungan yang tepat. Pembelajaran dilakukan selama 2 x 50 menit.

G. KEGIATAN PEMBELAJARAN 7

Pembelajaran 7 dilakukan melalui pendekatan Contextual Instruction dan Cooperative Learning. Dosen :Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, Menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa, Memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, Mencari, mengumpulkan dan menyusun informasi terkait dengan perhitungan yang tepat. Pembelajaran dilakukan selama 2 x 50 menit.

H. KEGIATAN PEMBELAJARAN 8

Pembelajaran 8 dilakukan melalui pendekatan Contextual Instruction dan Cooperative Learning. Dosen :Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, Menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa, Memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa.

Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, Mencari, mengumpulkan dan menyusun informasi terkait dengan perhitungan yang tepat. Pembelajaran dilakukan selama 2 x 50 menit.

I. KEGIATAN PEMBELAJARAN 9

Pembelajaran 9 dilakukan melalui pendekatan Contextual Instruction dan Simulasi. Dosen : Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa, memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, mencari, mengumpulkan dan menyusun informasi terkait dengan perhitungan yang tepat. Pembelajaran dilakukan selama 2 x 50 menit.

J. KEGIATAN PEMBELAJARAN 10

Pembelajaran 10 dilakukan melalui pendekatan Contextual Instruction dan Simulasi. Dosen : Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa, memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, mencari, mengumpulkan dan menyusun informasi terkait dengan perhitungan yang tepat. Pembelajaran dilakukan selama 2 x 50 menit.

K. KEGIATAN PEMBELAJARAN 11

Pembelajaran 11 dilakukan melalui pendekatan Contextual Instruction dan Simulasi. Dosen : Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa, memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, mencari, mengumpulkan dan menyusun informasi terkait dengan perhitungan yang tepat. Pembelajaran dilakukan selama 2 x 50 menit.

L. KEGIATAN PEMBELAJARAN 12

Pembelajaran 12 dilakukan melalui pendekatan Small Group Discussion. Dosen : Membuat rancangan bahan diskusi dan aturan diskusi, Menjadi moderator dan sekaligus

mengulas pada sesi akhir diskusi. Mahasiswa : Memilih bahan diskusi, Mempresentasikan bahan diskusi Pembelajaran dilakukan selama 2 x 50 menit.

M. KEGIATAN PEMBELAJARAN 13

Pembelajaran 13 dilakukan melalui pendekatan Contextual Instruction dan Simulasi. Dosen : Menjelaskan bahan kajian yang bersifat teori dan mengkaitkan dengan situasi yang nyata, menyiapkan suatu masalah atau tugas untuk diselesaikan oleh mahasiswa, memeriksa dan memberi usulan terhadap hasil belajar mandiri mahasiswa. Mahasiswa : Membahas konsep teori dan kaitannya dengan situasi nyata, mencari, mengumpulkan dan menyusun informasi terkait dengan perhitungan yang tepat. Pembelajaran dilakukan selama 2 x 50 menit.

N. KEGIATAN PEMBELAJARAN 14

Pembelajaran 14 dilakukan melalui pendekatan Small Group Discussion. Dosen :Membuat rancangan bahan diskusi dan aturan diskusi, Menjadi moderator dan sekaligus mengulas pada sesi akhir diskusi. Mahasiswa : Membentuk kelompok, 1 kelompok 4-5 mahasiswa, Memilih bahan diskusi, Mempresentasikan bahan diskusi

BAB III
SISTEM PENILAIAN

A. INDIKATOR PENILAIAN

Inikator penilaian pencapaian kompetensi yang dikembangkan mengacu pada aktivitas pembelajaran didasarkan pada pencapaian aspek kognitif, psikomotor, dan afektif yang terdiri dari:

No.	Indikator Penilaian	Persentase Penilaian
1.	Penugasan : a. Kelompok b. Individu	20% 25%
2.	Ujian Tengah Semester (UTS)	25%
3.	Ujian Akhir Semester (UAS)	25%
4.	<i>Softskill</i>	5%
	TOTAL	100%

B. NILAI LULUS MATA AJAR

Sistem penilaian berdasarkan acuan STIKes Hang Tuah Pekanbaru dalam nilai angka mutu, huruf mutu, dan bobot. Nilai batas lulus untuk setiap matakuliah adalah B (Minimal nilai 70).

Nilai Angka	Nilai Mutu	Angka Mutu	Sebutan Mutu
85-100	A	4,00	Sangat Baik
80-84	A ⁻	3,70	Sangat Baik
75-79	B ⁺	3,30	Baik
70-74	B	3,00	Baik
65-69	B ⁻	2,70	Cukup
60-64	C ⁺	2,30	Cukup
55-59	C	2,00	Cukup
50-54	C ⁻	1,70	Kurang
40-49	D	1,00	Sangat Kurang
0-40	E	0,00	Gagal

C. KRITERIA BOLEH MENGIKUTI UJIAN

1. Kehadiran minimal 75%;
2. Telah mengumpulkan semua tugas yang telah diberikan baik tertulis maupun tidak tertulis.

BAB IV

CENTRAL TENDENCY (NILAI TENGAH)

1. Pengertian:

Nilai Tengah adalah nilai yang dapat mewakili sekelompok nilai hasil pengamatan dan disebut juga rata-rata. Nilai tengah mempunyai kecenderungan berada ditengah-tengah suatu distribusi dan disebut juga *central tendency*.

2. Alasan Perlunya Nilai Tengah:

Jika kita ingin membandingkan satu kelompok pengamatan dengan kelompok lain tidak mungkin membandingkan satu persatu, tetapi dibandingkan masing-masing kelompok secara keseluruhan. Untuk itu diperlukan satu nilai yang dapat mewakili masing-masing kelompok agar dapat dibandingkan yaitu Nilai Tengah.

3. Macam-macam Ukuran Nilai Tengah antara lain:

- a. Rata-rata hitung (*arithmetic hitung*) atau *mean*
- b. Rata-rata ukur (*geometric mean*)
- c. *Median*
- d. *Modus (mode)*

a. Rata-rata hitung (*arithmetic hitung*) atau *mean*

Merupakan nilai tengah yang paling sering digunakan untuk menganalisis data. Simbol yang digunakan untuk rata-rata populasi adalah μ (miu) dan rata-rata sampel digunakan symbol \bar{x} (x bar).

Mean adalah jumlah semua pengamatan ($\sum X$) dibagi banyaknya pengamatan (n). Nilai mean dipengaruhi nilai ekstrim.

Rumus

$$\bar{x} = \frac{\sum x}{n}$$

Contoh:

Berat Badan 10 ibu hamil yang periksa di Rumah Bersalin adalah sebagai berikut: 62,60,55,54,70,67,53,61,64,75 dan 50. Ukuran berat dinyatakan dalam kg.

$$\text{Rata-rata} : \frac{62 + 60 + 55 + 54 + 70 + 66 + 53 + 61 + 64 + 75}{10} = \frac{620}{10} = 62$$

Jumlah aljabar perbedaan masing-masing pengamatan dengan nilai rata-rata sama dengan nol. Untuk membuktikan data perlu diurutkan dulu.

Contoh:

Berat Badan	$x - \bar{x}$	
53	53-62	= -9
54	54-62	= -8
55	55-62	= -7
60	60-62	= -2
61	61-62	= -1
		<hr/>
		= -27
62	62-62	= 0
64	64-62	= 2
66	66-62	= 4
70	70-62	= 8
75	75-62	= 13
		<hr/>
		= +27
	Jumlah	= 0

Untuk rata-rata yang banyak, penghitungan rata-rata dapat diselesaikan dengan beberapa rumus:

- 1) Data disusun dalam distribusi frekuensi tanpa pengelompokan
- 2) Distribusi frekuensi dikelompokkan dengan interval kelas yang sama
- 3) Distribusi frekuensi dikelompokkan dengan interval kelas tidak sama
- 4) Penghitungan rata-rata menggunakan kode.

Rata-rata hitung dengan pembebanan (*weighted mean*)

Digunakan untuk menghitung rata-rata beberapa kelompok dengan jumlah pengamatan tidak sama. Merupakan jumlah hasil kali antara banyaknya pengamatan dengan rata-rata tiap kelompok dibagi dengan jumlah pengamatan masing-masing kelompok.

Contoh: Pengukuran rata-rata berat badan 3 kelompok ibu hamil yang masing-masing terdiri atas 3, 5 dan 10 orang.

Kelompok 1:	50 kg	$\bar{x}_1 = 53$ kg
	55 kg	
	54 kg	

Kelompok 2:	50 kg	$\bar{x}_2 = 53.5$ kg
-------------	-------	-----------------------

	53 kg		
	52 kg		
	55 kg		
	57 kg		
Kelompok 3:	51 kg	48 kg	$\bar{x}_3=54,9$ kg
	55 kg	47 kg	
	57 kg	58 kg	
	60 kg	59 kg	
	52 kg	62 kg	

Hasil diatas dapat disusun:

Kelompok	n_i	\bar{x}_i	$\overline{n_1 x_1}$
1	3	53,0	159
2	5	53,5	267
3	10	54,9	549
Jumlah	18	161,3	975
\bar{x}			$975/18= 54,17$ kg

Bila rata-rata ketiga kelompok dihitung tanpa pembebanan maka akan dihasilkan rata-rata:
 $= 53+53,5+54,9/3=53,8$ kg

Bila dalam distribusi frekuensi terdapat nilai ekstrim maka penghitungan rata-rata harus dilakukan dengan pembebanan.

Contoh:

Tekanan darah sistolik penderita adalah 170, 180, 190 dan 90. Maka tekanan darah rata-rata dipengaruhi oleh tekanan darah sistolik 90 mmHg. Bila dihitung rata-rata tanpa mengikutsertakan nilai ekstrim hasilnya 180 mmg, sedangkan rata-rata dengan mengikutsertakan nilai ekstrim adalah 162 mmHg.

b. Rata-rata ukur (*geometric mean*)

Adalah penghitungan rata-rata akar pangkat n dari hasil perkalian setiap pengamatan

$$\bar{x}_u = \sqrt[n]{x_1 \times x_2 \times x_3 \dots x_n}$$

Contoh:

Hasil pengobatan suatu penyakit terhadap 3 orang penderita adalah sebagai berikut:

- Penderita 1: sembuh dalam waktu 3 hari
- Penderita 2: sembuh dalam waktu 9 hari

- Penderita 3: sembuh dalam waktu 27 hari

Rata-rata kesembuhan :

$$X_1=3$$

$$X_2=9, \text{ dan}$$

$$X_3=27$$

$$n=3$$

Rata-rata=

$$\bar{x}_u = \sqrt[3]{3 \times 9 \times 27} = 9 \text{ hari}$$

Rata-rata ukur hanya bisa digunakan untuk n kecil. Bila n besar maka cara penghitungan yang digunakan dengan mengambil logaritmanya.

Logaritma rata-rata ukur sama dengan jumlah logaritma setiap pengamatan dibagi banyaknya pengamatan. Logaritma ukur sama dengan antilogaritma dari jumlah logaritma setiap pengamatan dibagi dengan banyaknya pengamatan.

Rumus:

$$\log \bar{x}_u = \frac{\sum \log x}{n}$$
$$\bar{x}_u = \text{anti log} \left(\frac{\sum \log x}{n} \right)$$

c. Median

Median merupakan ukuran nilai tengah, menyatakan posisi tengah dari sederet angka hasil pengamatan. Ini berarti 50% nilai terletak dibawah median dan 50% terletak diatas median.

Rumus:

$$Me = \frac{(n+1)}{2}$$

Me = Median

n = banyaknya pengamatan

Rumus diatas berlaku bila jumlah pengamatan (n) ganjil. Bila n genap maka akan terdapat 2 posisi median yaitu antara $1/2n$ dan $1/2n + 1$. Dalam hal demikian median ditentukan dengan menghitung nilai rata-ratanya.

Keuntungan dan Kerugian:

- 1) Median mempunyai keuntungan bila dibandingkan mean, terutama bila data tidak simetris. Median merupakan nilai tengah yang paling sedikit dipengaruhi nilai ekstrim.
- 2) Median dapat digunakan untuk menghitung data dalam bentuk distribusi frekuensi dengan kelas terbuka yang tidak dapat dilakukan dengan mean.
- 3) Median dapat digunakan untuk menghitung nilai tengah data kualitatif.

Kerugian dan Kelemahan:

- 1) Perhitungan median lebih rumit dibanding mean, karena data harus disusun lebih dahulu terutama bila data banyak.
- 2) Median tidak memperhitungkan setiap nilai hasil pengamatan seperti mean.
- 3) Nilai median gabungan dua kelompok tidak sama dengan jumlah nilai kelompok satu ditambah nilai kelompok dua.

Penghitungan Median data tidak dikelompokkan:

Cara menghitung nilai median:

Data disusun secara berurutan dari yang kecil ke yang besar atau sebaliknya.

Tentukan posisi median dengan rumus. Cocokkan nilai median dengan nilai data yang telah disusun dan akan diperoleh nilai median.

Contoh:

Akan mengukur HB 5 orang wanita hamil yang diperiksa di Puskesmas B, dan kita kan menentukan nilai mediannya.

Posisi median terletak pada $(5+1)/2=3$

Kadar HB yang diperoleh disusun secara teratur : 8, 9, 10,11,12

Data disusun sebagai berikut:

Posisi Median	1	2	3	4	5
Kadar HB (mg%)	8	9	10	11	12

Berdasarkan perhitungan median terletak pada posisi ke-3 yang sesuai dengan kadar HB 10 mg%. Bila data yang diperoleh merupakan bilangan genap, misalnya 6 orang maka posisi median terletak pada posisi ke-3 dan ke-4.

Posisi Median	1	2	3	4	5	6
Kadar HB (mg%)	8	9	10	11	12	13

Median terletak antara posisi ke-3 dan ke-4 sehingga nilai mediansama dengan $(10+11)/2 = 10,5g\%$.

d.Modus (mode)

Modus merupakan ukuran nilai tengah yang dinyatakan dala frekuensi terbanyak dari data kualitatif maupun kuantitatif. Modus dapat dinyatakan sebagai puncak dari suatu kurva. Oleh karena itu dikenal unimodal bila puncaknya satu, bimodal bila puncaknya dua dan multimodal bila puncaknya lebih dari dua.

Penghitungan modus dapat dilakukan untuk data distribusi frekuensi yang tidak dikelompokkan maupun yang dikelompokkan. Modus jarang digunakan sebagai nilai tengah karena sangat dipengaruhi oleh nilai ekstrim, dan letaknya akan berubah mengikuti nilai ekstrim.

Penghitungan modus untuk data tidak dikelompokkan:

Modus jarang digunakan untuk data yang tidak dikelompokkan:

Contoh penghitungan:

Menilai banyaknya kesalahan seorang laboran dalam menentukan jumlah leukosit di RS B selama 4 minggu.

Minggu 1	Jumlah Kesalahan	0	2	5	7	15
Minggu 2		0	2	5	7	15
Minggu 3		1	4	6	8	15
Minggu 4		1	4	6	12	19

Dari data tersebt diketahui bahwa modulusnya adalah 15, tetapi bila kita hitung meannya adalah 6,7. Bila data dikelompokkan maka frekuensi terbanyak adalah 4,7 kali kesalahan. Ini mendekati mean. Oleh karena itu modus sebaiknya diubah dahulu menggunakan distribusi frekuensi.

Penghitungan modus pada data dikelompokkan

Bila data berupa distribusi frekuensi yang telah dikelompokkan maka modusnya terletak pada interval kelas dengan frekuensi terbanyak. Untuk menghitung satu nilai modus dengan frekuensi terbanyak pada distribusi frekuensi yang telah dikelompokkan digunakan rumus:

Rumus:

$$M_o = L_{M_o} + (d_1 / (d_1 + d_2)) \times i$$

Keterangan:

M_o = Modus

L_{M_o} = tepi bawah kelas dimana modus berada

d_1 = selisih antara frekuensi kelas modus dengan kelas tepi dibawahnya

d_2 = selisih antara frekuensi kelas modus dengan kelas tepi diatasnya

i = lebar interval kelas modus

Contoh:

Distribusi umur dari 80 penderita stroke di RS C pada tahun 2012

Distribusi Frekuensi	f	
21-30	6	
31-40	7	
41-50	40	Modus
51-60	10	
61-70	10	
71-80	7	
Jumlah	80	

Nilai median=

$L_{M_o} = 40,5$

$d_1 = 40 - 7 = 33$

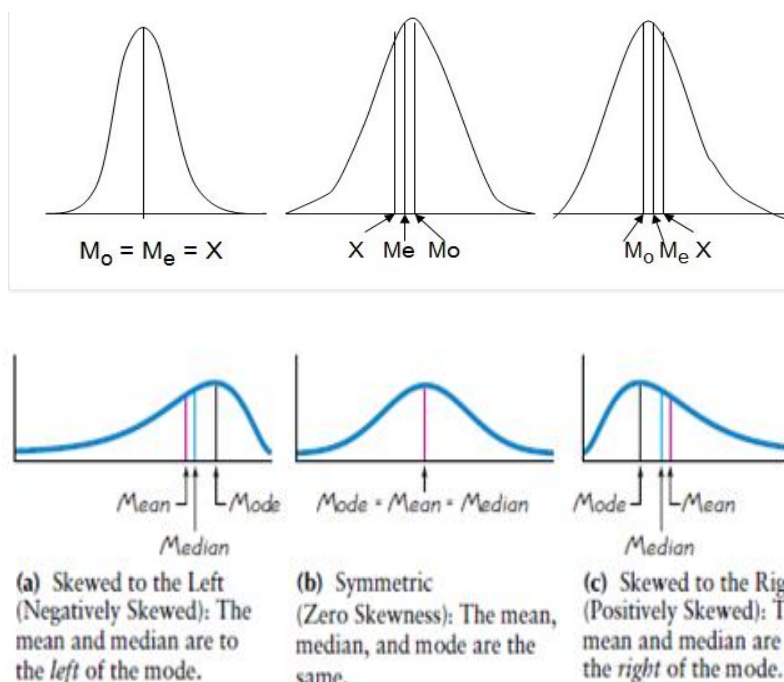
$d_2 = 40 - 10 = 30$

$i = 10$

$$\begin{aligned} M_o &= 40,5 + 10 \left(\frac{33}{33 + 30} \right) \\ &= 40,5 + 5,2 \\ &= 45,7 \text{ tahun} \end{aligned}$$

e. Hubungan antara Mean, Median dan Modus

- 1) Pada kurva yang simetris, mean, median dan modus terletak pada satu titik (mean=median=modus)
- 2) Pada distribusi miring ke kanan, modus akan bergeser ke kiri mengikuti nilai dengan frekuensi terbanyak, sedangkan mean akan bergeser ke kanan karena terpengaruh oleh nilai ekstrim dan median terletak antara mean dan modus
- 3) Bila distribusi miring ke kiri maka modus akan bergeser ke kanan mengikuti nilai dengan frekuensi terbanyak sedangkan mean bergeser ke kiri mengikuti nilai ekstrim dan median terletak antara mean dan modus
- 4) Secara empiris dapat dikatakan bahwa jarak antara modus dan median merupakan $\frac{2}{3}$ jarak antara modus dan mean
- 5) Modus mengalami pergeseran terbesar diikuti mean dan median. Median relative stabil dibanding modus dan mean, tetapi rata-rata dari sampel ke sampel maka mean mempunyai fluktuasi terkecil.



Gambar. Hubungan Mean, Median, Modus

BAB V

UKURAN VARIASI/DISPERSI/UKURAN PENYIMPANGAN

Pada ukuran nilai tengah, beberapa distribusi dengan mean yang sama kemungkinan mempunyai variasi yang berbeda. Bila hanya menghitung nilai tengah saja tanpa memperhatikan variasi yang mungkin berbeda akan kehilangan informasi penting dan akan menimbulkan interpretasi yang berbeda. Untuk mendapatkan informasi yang lengkap perlu menghitung variasi dan penyimpangan terhadap rata-rata.

Penghitungan variasi penting karena:

1. Akan mendapatkan informasi tambahan tentang penyimpangan yang terjadi pada suatu distribusi
2. Dapat menilai ketepatan nilai tengah dalam mewakili distribusinya.
3. Mengadakan analisis melalui penghitungan statistic yang lebih mendalam

Mengapa Terjadi Variasi

Terjadinya variasi merupakan peristiwa alamiah. Perbedaan atau variasi bisa terjadi karena perbedaan pengukuran antar individu yang disebut variasi eksterna. Bila seseorang melakukan penghitungan berkali-kali pada waktu yang berbeda dengan hasil yang berbeda-beda, maka variasi ini terjadi intra individu disebut variasi interna.

Macam-macam ukuran variasi atau disperse meliputi:

1. Rentang (*range*)
2. *Inter-quartile Range*
3. Deviasi rata-rata (*mean deviation*)
4. Varians dan Deviasi standar (*standard deviation*)
5. Koefisien variasi (*Coefficient of variation*)

1. Rentang (*range*)

Ukuran dispersi yang paling sederhana pada suatu data numerik adalah dengan cara menghitung selisih nilai terbesar (nilai maksimum) dengan nilai terkecil (nilai minimum). Cara ini dikenal dengan sebutan *Range*.

Rumus:

$$\text{Range} = \text{Nilai maksimum} - \text{Nilai minimum}$$

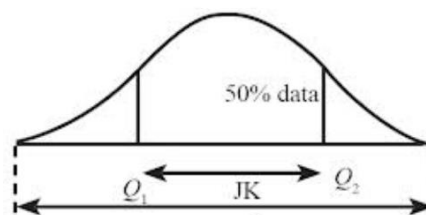
Manfaat range sangat terbatas karena:

- 1) Range hanya melibatkan nilai terbesar dan nilai terkecil tanpa melibatkan nilai-nilai dalam distribusi
- 2) Range sangat dipengaruhi oleh nilai ekstrim
- 3) Range tidak dapat ditentukan pada distribusi dengan kelas interval yang terbuka

Range dapat digunakan untuk memperkirakan simpangan baku atau deviasi standar (standard deviation (SD) karena pada distribusi normal, simpangan 1 SD kira-kira sama dengan 1/6 dari range. Bila hasil penghitungan standard baku sangat jauh menyimpang dari 1/6 rangemaka kemungkinan penghitungan kita yang salah atau distribusinya tidak normal.

2. Rentang antar Kuartil (*Inter quartile range*)

Selisih antara K3-K1 disebut rentang antar-kuartil (interquartile range) yang sama dengan 50% bagian tengah dari seluruh distribusi. Setengah dari rentang antar kuartil disebut simpangan kuartil (*quartile deviation*)



Cara perhitungan:

Untuk menghitung rentang antar kuartil ditentukan letak K3 dan K1. Perhitungan rentang antar kuartil dan simpangan kuartil dapat dilakukan pada data yang tidak dikelompokkan dan data yang dikelompokkan.

Data yang tidak dikelompokkan:

Menentukan letak K₃ dan K₁

Rumus:

$$K_3 = \frac{3}{4} (n+1) \text{ dan } K_1 = \frac{1}{4} (n+1)$$

Menghitung nilai K_3 dan K_1 dengan rumus:

$$K_3 \text{ dan } K_1 = L + b (S - L)$$

L = nilai sebelum K_3 dan K_1

b = kekurangan unit untuk mencapai letak K_3 dan K_1

S = nilai dimana K_3 dan K_1 berada

Contoh:

Ingin mengetahui nilai rentang antar kuartil kolesterol darah dari 10 penderita hipertensi.

Hasil pemeriksaan:

Kolesterol = 150, 152, 160, 165, 167, 169, 171, 174, 175 dan 593.

Data ke = 1 2 3 4 5 6 7 8 9 10

Letak = $K_3 = \frac{3}{4} (10+1)$ $K_3 = 8,25$

Antara data ke 8 dan 9

= $K_1 = \frac{1}{4} (10+1)$ $K_1 = 2,75$

Antara data ke 2 dan 3

Nilai $K_3 = 174 + 0,25 (175-174)$
 $= 174,25$

$K_1 = 152 + 0,75 (160-152)$
 $= 158$

Rentang antar kuartil = $K_3 - K_1 = 174,25 - 158 = 16,25$

Data dengan pengelompokan:

Untuk menghitung rentang antar kuartil pada data yang dikelompokkan digunakan rumus.

Letak kuartil diubah menjadi unit dengan rumus

$$x = \frac{(K \times n)}{4}$$

$$\text{Nilai Kuartil } K_k = L + i \frac{(x - f_{\text{kum}})}{f}$$

Keterangan:

L = tepi bawah kelas dimana kuartil berada

i = interval kelas

f_{kum} = frekuensi kumulatif sebelum kuartil

f = frekuensi dimana kuartil berada

x = letak kuartil

Contoh:

Ingin mengetahui rentang antar kuartil penderita demam berdarah yang dirawat di RS C pada Tahun 2012. Hasil distribusi umur disajikan dalam bentuk distribusi frekuensi kumulatif.

umur	f	f _{kum}
10-19	2	2
20-29	23	25
30-39	15	40
40-49	11	51
50-59	9	60
60-69	5	65
70-79	2	67
Jumlah	67	

Letak = $K3 = (3 \times 67)/4$

$$K3 = 50,25$$

Antara kelas ke 4 dan 5

=K1 = (1 x 67)/4

$$K1 = 16,75$$

Antara kelas ke 1 dan 2

Nilai $K3 = 39,5 + 10 (50,25 - 40)/11$

$$= 39,5 + 9,3 = 48,8$$

$K1 = 19,5 + 10 (50,25 - 40) / 11$

$$= 19,5 + 6,4 = 25,9$$

Rentang antar kuartil = $48,8 - 25,9 = 22,9$

Deviasi Kuartil (Quartile Deviation)

Dispersi Kuartil (D_k) ialah ukuran disperse berdasarkan rentang antar kuartil dengan median sebagai posisi tengah.

Deviasi kuartil merupakan setengah selisih antara K_3 dan K_1

Rumus:

$$D_k = (K_3 - K_1) / 2$$

Median atau K_2 sama dengan median dengan rumus:

$$Me = (K_3 + K_1) / 2$$

Contoh:

1) Untuk data yang tidak dikelompokkan:

Digunakan contoh untuk perhitungan rentang antar kuartil

Rentang antar kuartil = 16,25 maka deviasi kuartil = 8,1

Nilai median = $(174,25 + 158) / 2 = 166,1$

2) Untuk data yang dikelompokkan:

Rentang antar kuartil = 22,9, deviasi kuartil = 11,45

Median = 37,35

3. Deviasi Rata-rata (*mean deviation*)

Simpangan atau deviasi rata-rata merupakan ukuran yang sangat penting dari dasar dua ukuran disperse simpangan baku (standard deviasi) dan varians (variance).

Simpangan rata-rata adalah jumlah selisih antara hasil setiap pengamatan dengan rata-rata dibagi dengan banyaknya pengamatan.

Simpangan rata-rata bermanfaat untuk mengetahui variasi yang terjadi dalam kelompok pengamatan atau membandingkan tingkat variabilitas dua kelompok atau lebih.

Rumus Deviasi rata-rata (Mean Deviation=MD)

(Populasi) MD=

$$\frac{\sum(x_i - \mu)}{N}$$

(Sampel) MD=

$$\frac{\sum(x - \bar{x})}{n}$$

Contoh:

Berat badan 2 kelompok ibu hamil yang masing-masing terdiri 5 orang.

Kelompok 1			Kelompok 2		
1	2	Selisih 1 & 2	1	2	Selisih 1 & 2
Berat Badan	Mean		Berat Badan	Mean	
40	50	10	25	50	40
45		5	35		15
50		0	55		5
55		5	60		10
60		10	75		35
250		30	250		105

Kelompok 1

$$\bar{x} = 50$$

$$x - \bar{x} = 30$$

$$MD = 30/5 = 6$$

Kelompok 2

$$\bar{x} = 50$$

$$x - \bar{x} = 105$$

$$MD = 105/5 = 21$$

Dari perhitungan diatas dapat dinyatakan bahwa variabilitas kelompok 2 adalah 3,5 kali lebih besar daripada kelompok 1.

4. Varians dan Deviasi Standar (*Standard Deviation*)

Varians merupakan teknik statistic yang digunakan untuk menjelaskan homogenitas kelompok. Varians merupakan jumlah kuadrat semua deviasi nilai-nilai individu terhadap nilai rata-rata kelompok. Akar varians disebut standar deviasi atau simpangan baku.

Simpangan baku (*standard deviation*) adalah ukuran dispersi yang penting. Mula-mula digunakan oleh Karl Pearson. Penyimpangan atau selisih nilai hasil pengamatan dengan rata-rata dapat menghasilkan nilai yang negative. Untuk menghindari dapat dilakukan dengan mengambil nilai absolutnya, tetapi dengan nilai absolute tidak dapat mengetahui arah penyimpangan yaitu ke kiri atau ke kanan. Untuk mengatasi kelemahan tersebut maka selisih hasil pengamatan dengan rata-rata dipangkatkan dua hingga hasilnya menjadi positif.

Jumlah seluruh selisih hasil pengamatan dengan rata-rata yang telah dipangkatkan dua dibagi jumlah pengamatan disebut *varians*.

Bila varians ditarik akarnya maka akan menghasilkan deviasi standar (*standard deviation*)

Rumus

Varians =

$$\sigma^2 = \sum (x - \mu)^2 / N$$

Deviasi standar =

$$\sigma = \sqrt{\sum (x - \mu)^2 / N}$$

Keterangan:

x = hasil pengamatan

N = banyaknya pengamatan

σ = deviasi standar

μ = rata-rata

Cara penghitungan

- 1) Data mentah disusun secara berurutan
- 2) Jumlahkan hasil pengamatan ($\sum X$)
- 3) Bagi sigma x dengan banyaknya pengamatan ($\sum X/N = \mu$)
- 4) Kurangi hasil tiap pengamatan dengan rata-rata
- 5) Pangkatkan hasil no 4.
- 6) Jumlahkan hasil no 5.
- 7) Bagilah hasil no 5 dengan banyaknya pengamatan
- 8) Hasil no 7 ditarik akarnya

Contoh:

Hasil pemeriksaan glukosa darah puasa 10 orang dewasa normal adalah:

Glukosa darah	$x - \bar{x}$	$(x - \bar{x})^2$
70	-8,4	70,56
72	-6,4	40,96
76	-2,4	5,76
77	-0,4	0,16
78	0,6	0,36
79	1,6	2,56
80	2,6	6,76
85	6,6	43,56
86	7,6	57,76
Jumlah 784		230,40

$$\bar{x} = 784/10 = 78,4$$

$$SD = \sqrt{230,4/10} = \sqrt{23} = 4,8 \text{ mg\%}$$

Deviasi standar Populasi dan Sampel

Pada suatu populasi selalu terjadi variasi dari hasil pengamatan, baik eksterna maupun interna. Oleh karena itu varians dan deviasi standar populasi harus dihitung sebagai parameter. Bila menghitung varians dan standar deviasi sampel maka akan diekstrapolasikan pada parameter populasi. Namun kita jarang melakukan penghitungannya pada tingkat populasi karena sering besarnya populasi tidak diketahui, membutuhkan waktu dan biaya besar. Berdasar hal tersebut maka penghitungan dilakukan pada sampel yang diekstrapolasikan pada parameter populasi.

Rumus varians sampel

$$\sigma^2 = \sum (x - \bar{x})^2 / n - 1$$

Rumus standar deviasi sampel:

$$\sigma = \sqrt{\sum (x - \bar{x})^2 / n - 1}$$

Alasan n dikoreksi dengan n-1 adalah:

Varians dan standar deviasi sampel menunjukkan suatu kecenderungan untuk lebih kecil dari varians dan standar deviasi populasinya hingga nilai estimasinya lebih kecil (under estimate) dan makin kecil sampel, makin besar under estimasinya. Untuk mengatasi kelemahan tersebut maka dilakukan koreksi menjadi n-1.

5. Koefisien variasi (*Coefficient of variation*)

Salah satu cara membandingkan data secara relative ialah dengan koefisien variasi dimana deviasi standar dinyatakan sebagai persentase dari rata-rata atau merupakan perbandingan antara simpangan baku terhadap rata-rata yang dinyatakan dalam persen.

Rumus:

$$KV = (SD/\bar{x}) \times 100\%$$

Contoh:

Seorang analis A dalam sehari memeriksa 40 sampel darah dengan deviasi standar 5.

Analis B dalam sehari memeriksa 160 sampel darah dengan deviasi standar 15.

Secara sepintas analisis B mempunyai variasi yang lebih besar disbanding A, tetapi analis B dapat memeriksa sampel darah 4 kali lebih banyak disbanding A.

Dengan perhitungan KV diperoleh hasil:

Analisis A : $KV = (5/40) \times 100 = 12,5\%$

Analisis B : $KV = (15/160) \times 100 = 9,4\%$

Dengan membandingkan secara relative maka analisis B mempunyai variasi yang lebih kecil dari A, karena B mempunyai kemampuan 4 kali lebih besar dari A.

Koefisien variasi dapat digunakan untuk membandingkan satu variable dari dua kelompok dengan satuan yang sama maupun berbeda. KV bisa digunakan untuk mengetahui homogenitas suatu kelompok apabila KV kurang dari 10% maka kelompok dianggap homogeny.

Kerugian:

KV tidak menyatakan suatu kemungkinan, perhitungan KV tidak sama bila data semula mengalami perubahan yang non linier.

BAB VI

UKURAN POSISI

Bila data kita susun mulai dari data yang terkecil sampai terbesar, maka kita dapat membagi pengamatan menjadi beberapa bagian. Pembagian pengamatan ini disebut sebagai nilai letak atau posisi.

Jadi Ukuran Letak, yakni ukuran yang menggambarkan posisi skor di dalam distribusi.

Adapun ukuran posisi yang sering digunakan adalah :

1. Median
2. Kuartil/Kuartil (Q)/K,
3. Desil (D),
4. Persentil (P)

1. Kuartil/Kuartil (Q)/K,

Kuartil adalah nilai yang membagi data menjadi 4 bagian sama banyak. Nilai-nilai Kuartil diberi simbol K1, K2 (sama dengan Median) dan K3.

Kuartile 1 (K1) merupakan nilai yang membagi jumlah pengamatan menjadi dua bagian, yaitu sebesar 25% dibawah K1 dan 75% diatas K1

K2 atau sama dengan Median merupakan nilai yang membagi jumlah pengamatan menjadi dua bagian yaitu sebesar 50% dibawah K2 dan 50% diatas K2

K3 atau sama dengan Median merupakan nilai yang membagi jumlah pengamatan menjadi dua bagian yaitu sebesar 75% dibawah K2 dan 25% diatas K2

Perhitungan Kuartile:

$$Posisi - Kuartile = \frac{k_i}{4} (n)$$

Nilai kuartil adalah nilai observasi pada posisi kuartil tersebut

2. Desil (D) adalah nilai yang membagi data menjadi 10 bagian sama banyak. simbolnya adalah D1, D2, ..., D9.

Prinsipnya sama dengan kuartile, hanya dengan desil kita dapat memperoleh informasi yang lebih detail.

Perhitungan desil

$$\text{Posisi} - \text{Desil} = \frac{d_i}{10} (n)$$

Nilai desil adalah nilai observasi pada posisi desil tersebut

3. Persentil (P) adalah nilai yang membagi data menjadi 100 bagian sama banyak. simbolnya adalah P1, P2, ..., P99.

Prinsipnya sama dengan kuartile/desil, hanya dengan persentil kita dapat memperoleh informasi yang lebih detail lagi.

$$\text{Posisi} - \text{Persentil} = \frac{P_i}{100} (n)$$

Nilai persentil adalah nilai observasi pada posisi persentil tersebut.

PROPORSI

Proporsi diperoleh melalui perbandingan frekuensi suatu data dengan frekuensi total. Proporsi dapat dinyatakan dalam pecahan di antara 0 sampai 1 dan dapat juga dinyatakan dalam persentase dari 0% sampai 100%.

1. Parameter Proporsi

- a. Parameter adalah ciri tertentu pada kelompok populasi
- b. Salah satu parameter pada populasi adalah proporsi (dapat dinyatakan dalam pecahan dan dapat juga dinyatakan dalam persentase)
- c. Proporsi berkaitan dengan frekuensi data

- d. Proporsi suatu data di dalam populasi adalah rasio frekuensi data itu terhadap seluruh frekuensi
- e. Parameter proporsi pada populasi diberi notasi π

2. Rumus Proporsi

Jika suatu data memiliki frekuensi f sedangkan seluruh frekuensi adalah Σf , maka proporsi untuk data itu adalah

$$\pi = \frac{f}{\Sigma f}$$

Batas nilai proporsi adalah $0 \leq \pi \leq 1$

Jumlah seluruh proporsi adalah $\Sigma \pi = \Sigma f / \Sigma f = 1$

Apabila data bersifat dikotomi yang dinyatakan dengan 0 dan 1 (misalnya, 0 adalah gagal dan 1 adalah sukses), maka proporsi sukses : π Proporsi gagal : $1 - \pi$

$$\text{Rumus } P = \frac{X}{X + Y} * K$$

3. Kumulasi Proporsi

a. Kumulasi Umum

Seperti halnya pada frekuensi, proporsi untuk sejumlah data tertentu dapat dijumlahkan. Jumlah proporsi ini dikenal sebagai kumulasi proporsi

b. Kumulasi Bawah

Jika kumulasi itu dilakukan bertahap menurut urutan data dari kecil ke besar maka kumulasi proporsi itu adalah kumulasi bawah

c. Kumulasi Atas

Jika kumulasi itu dilakukan bertahap menurut urutan data dari besar ke kecil maka kumulasi proporsi itu adalah kumulasi atas.

4. Grafik dan Ojaif Proporsi

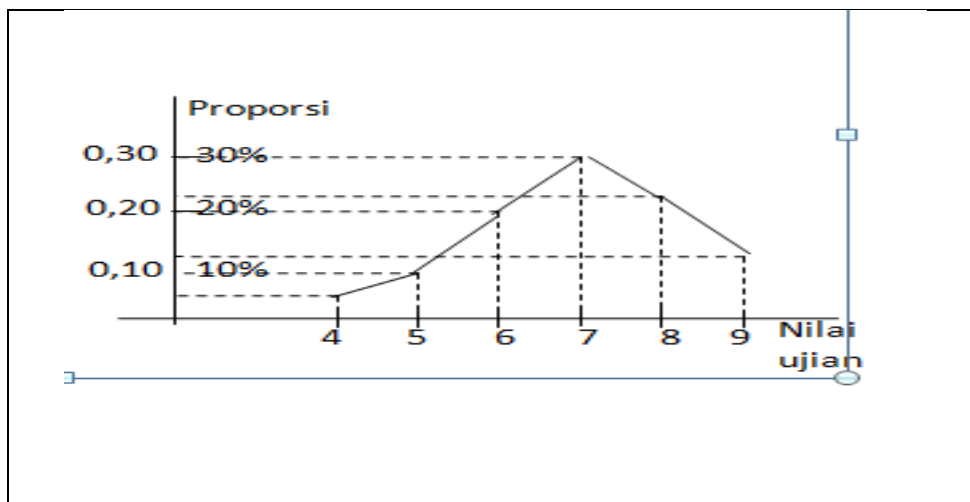
Seperti halnya frekuensi, proporsi dapat ditampilkan dalam bentuk grafik. Ada banyak macam cara untuk menampilkan proporsi ke dalam grafik. Contoh grafik poligon dan histogram. Grafik poligon menghubungkan letak titik proporsi setiap data sehingga membentuk segi banyak. Grafik histogram menampilkan proporsi setiap data dalam ukuran luas. Ojaif proporsi adalah grafik poligon dari kumulasi bawah dan kumulasi atas

a. Grafik Poligon Proporsi

Setiap data dilukiskan sebagai titik di dalam grafik. Titik-titik dihubungkan sehingga membentuk gambar segi banyak atau poligon

Contoh

Nilai ujian	Frekuensi	Proporsi
4	3	0,06 (6%)
5	5	0,10 (10%)
6	10	0,20 (20%)
7	15	0,30 (30%)
8	11	0,22 (22%)
9	6	0,12 (12%)



b. Grafik Histogram

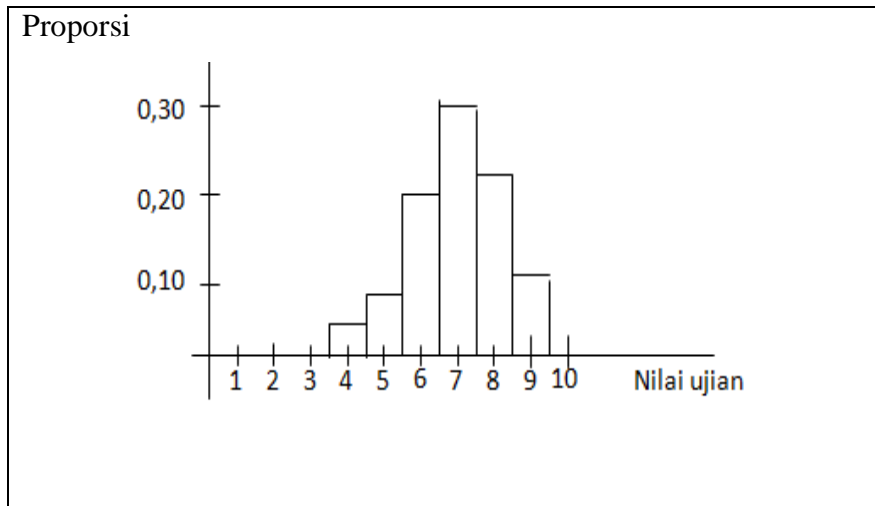
Proporsi setiap data dilukis sebagai suatu luas di dalam grafik. Kumulasi proporsi diperoleh melalui penjumlahan luas di dalam grafik

Contoh:

Nilai ujian : 4 5 6 7 8 9

Frekuensi: 3 5 10 15 11 6

Proporsi: 0,06 0,10 0,20 0,30 0,22 0,12



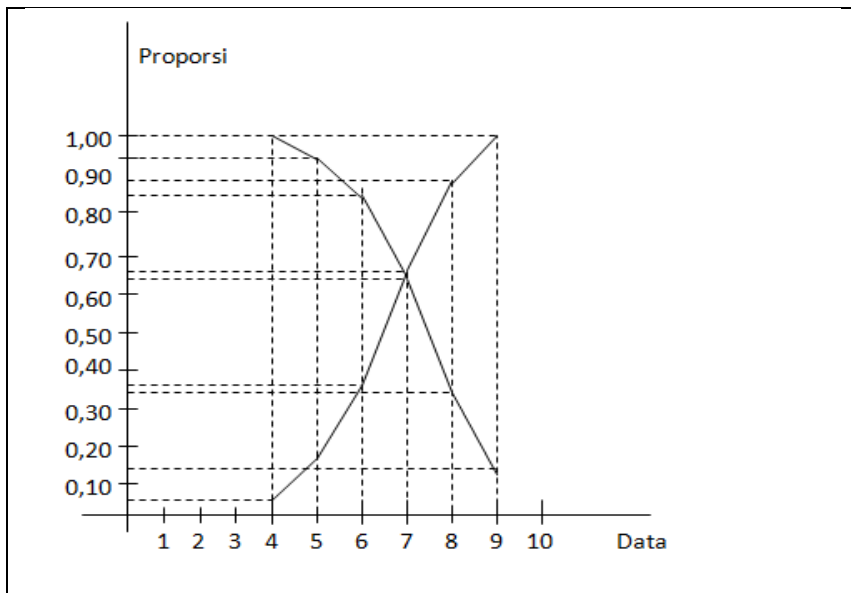
c. Grafik Ojaif

Grafik ojaif (ogive) merupakan grafik poligon untuk kumulasi proporsi bawah dan kumulasi proporsi atas.

Contoh 13

Data	Frek	Prop	Kum bawah	Kum atas
4	3	0,06	0,06	1,00
5	5	0,10	0,16	0,94
6	10	0,20	0,36	0,84
7	15	0,30	0,66	0,64
8	11	0,22	0,88	0,34
9	6	0,12	1,00	0,12

Yang digambar di dalam grafik adalah kumulasi bawah dan, di sini, juga, kumulasi atas



DAFTAR PUSTAKA

Budiarto.2002.Biostatistika. Jakarta: EGC

Dahlan, Sopiudin. 2008. *Statistik untuk Kedokteran dan Kesehatan*. Seri Evidence Based Medicine 2 Edisi 3. Jakarta: Penerbit Salemba Medika.

Kuzma, J.W. 1984. *Basic Statistics for the Health Sciences*. California: Meyfield Publishing Company.

Sugiyono.1999.Statistika untuk Penelitian.Bandung: Alfabeta

BAB VII

DISTRIBUSI SAMPLING DAN ESTIMASI STATISTIK

A. Distribusi Sampel

Distribusi sampel adalah distribusi dari rata-rata atau proporsi sampel yang diambil secara berulang-ulang (n kali) dari populasi.

Ada sebanyak " n " nilai rata-rata atau sebanyak " n " nilai proporsi.

Distribusi dari nilai rata-rata atau nilai proporsi tersebut disebut sebagai distribusi sampel (*sampling distribution*)

Sifat-sifat dari distribusi sampel tersebut dikenal dengan *Central Limit Theorem*

1. Bentuk distribusi dari rata-rata sampel akan mendekati distribusi normal meskipun distribusi populasi tidak normal.
2. Rata-rata dari rata-rata sampel sama dengan rata-rata populasi (μ)

$$\mu_x = \mu$$

$$\sigma_x = \sigma / \sqrt{n}$$

$$\sigma_x = SE$$

μ_x = rata-rata dari rata-rata sampel yang mungkin diambil dari suatu populasi

μ = rata-rata populasi

σ_x = sd dari mean sampel yang mungkin diambil dari suatu populasi

σ = sd populasi

3. Standar deviasi dari rata-rata sampel, = sama dengan standar deviasi populasi (σ) dibagi dengan akar jumlah sampel. Dikenal dengan istilah *Standard Error* (SE)

$$SE = \sigma / \sqrt{n}$$

SE \rightarrow ukuran yang menunjukkan ketepatan sampel mewakili populasi

$\ll SE \rightarrow$ penelitian semakin baik \rightarrow mendekati populasi

$\gg SE \rightarrow$ penelitian semakin tdk baik

$$\text{Margin of error (MoE)} = z * \sigma / \sqrt{n}$$

MoE ; nilai yang dapat menunjukkan seberapa besar kesalahan yang terjadi saat pengambilan sampel.

Jika Margin of Error bernilai kecil, menunjukkan semakin tingginya Level Kepercayaan akan data sampel.

Saat Margin of Error semakin besar, maka data semakin beragam sehingga Standar Deviasi juga akan semakin besar.

Apabila data sampel semakin banyak, Margin of Error akan semakin kecil karena kemungkinan kesalahan pengambilan data semakin sedikit.

B. Estimasi (Pendugaan)

Estimasi dan Pengujian Hipotesis

Estimasi (pendugaan) adalah : pendugaan karakteristik populasi (parameter) melalui karakteristik sampel (statistik).

Uji hipotesis adalah pengujian karakteristik populasi (parameter) terhadap nilai tertentu atau perbandingan nilai tertentu melalui karakteristik sampel (statistik).

Interpretasi frekuensi relatif

Dari 100 kali pengambilan sampel akan diperoleh sebanyak sampel yang memuat parameter populasi (μ , p).

- 90%, 95%, 99% CI (Confidence Interval)
- Dalam statistik tidak ada kepercayaan 100% (*Jika mau percaya 100% silakan ukur seluruh populasi /sensus*)

Suatu penelitian bertujuan untuk mengestimasi 'uric acid' pada populasi Lansia (umur di atas 65 tahun). Dari 100 orang sampel Lansia yang diambil secara random dari populasi Lansia tersebut, peneliti melaporkan bahwa rata-rata 'uric acid' pada Lansia adalah 5.9 mg/100 (point estimate) dengan standar deviasi 1.5 mg/100 ml. 95% tingkat kepercayaan peneliti bahwa rata-rata 'uric acid' pada populasi Lansia berkisar antara 5.6-6.2 mg/100 ml (interval estimate).

Estimasi (Pendugaan)

Dari 10.000 wanita berusia 50-54 tahun yang diteliti ditemukan sebanyak 400 menderita kanker payudara. Dari data tersebut, maka prevalensi kanker payudara pada populasi adalah sebesar $400/10.000=0.04$ atau 4% (point estimate) dengan 95% tingkat kepercayaan peneliti bahwa prevalensi kanker payudara pada populasi wanita tersebut berkisar antara 3.6%-4.4% (interval estimate).

Estimasi Interval terhadap μ di populasi bila σ diketahui

Beberapa asumsi

- a. Standard deviation populasi diketahui.
- b. Populasi berdistribusi normal.
- c. Jika populasi tidak normal, gunakan sampel besar

Estimasi Interval terhadap nilai μ

Estimasi Interval terhadap μ di populasi bila σ (std dev) di populasi diketahui

Data Lansia. Bila standar deviasi 'urid acid' di populasi diketahui sebesar 1.5 mg/100 ml dengan rata-rata dari 100 sampel Lansia adalah 5.9 mg/100ml. Maka estimasi interval terhadap rata-rata 'uric acid' pada populasi Lansia adalah sebagai berikut:

- Pada tingkat kepercayaan 90%

$$= 5.9 \pm (1.64)(1.5/\sqrt{100})$$

$$= 5.9 \pm 0.246$$

$$5.654 < \mu < 6.146 \text{ mg/100 ml}$$

- Pada tingkat kepercayaan 95%

$$= 5.9 - (1.96)(1.5/\sqrt{100}) < \square < 5.9 + (1.96)(1.5/\sqrt{100})$$

$$= 5.9 - 0,294 < \square < 5.9 + 0,294$$

$$5.606 < \mu < 6.194 \text{ mg/100 ml}$$

- Pada tingkat kepercayaan 99%

$$5.9 - (2.58)(1.5/\sqrt{100}) < \square < 5.9 + (2.58)(1.5/\sqrt{100})$$

$$5.51 < \mu < 6.29 \text{ mg/100 ml}$$

Estimasi Interval terhadap μ di populasi bila σ di populasi tidak diketahui

Beberapa asumsi

- Standart deviasi populasi tidak diketahui
- Populasi berdistribusi normal
- Jika populasi tidak berdistribusi normal gunakan sampel besar
- Menggunakan distribusi student's t
- Estimasi Interval terhadap nilai μ

Estimasi Interval terhadap μ di populasi bila σ di populasi tidak diketahui

Data Lansia. Bila standar deviasi populasi tidak diketahui maka digunakan standar deviasi sampel, misalnya 1.5 mg/100 ml dengan rata-rata dari 100 sampel Lansia adalah 5.9 mg/100ml. Maka estimasi interval terhadap rata-rata 'uric acid' pada populasi Lansia adalah sebagai berikut:

- Pada tingkat kepercayaan 90%

$$5.9 - (1.66)(1.5/\sqrt{100}) < \mu < 5.9 + (1.66)(1.5/\sqrt{100})$$

$$5.65 < \mu < 6.15 \text{ mg/100 ml}$$

- Pada tingkat kepercayaan 95%

$$5.9 - (1.98)(1.5/\sqrt{100}) < \mu < 5.9 + (1.98)(1.5/\sqrt{100})$$

$$5.60 < \mu < 6.20 \text{ mg/100 ml}$$

- Pada tingkat kepercayaan 99%

$$5.9 - (2.63)(1.5/\sqrt{100}) < \mu < 5.9 + (2.63)(1.5/\sqrt{100})$$

$$5.51 < \mu < 6.29 \text{ mg/100 ml}$$

Estimasi Interval terhadap proporsi di populasi

Beberapa asumsi

- Data berupa dua kategori
- Populasi mengikuti distribusi binomial
 - Rata-rata p
 - SD $\sqrt{p(1-p)}$
 - SE $\sqrt{p(1-p)/n}$

Estimasi Interval Proporsi

p_s adalah proporsi sampel

Estimasi Interval terhadap proporsi di populasi

Data Kanker Payudara. $p_s=0.04$, $q_s=1-p_s=1-0.04=0.96$, $n=10.000$. Maka estimasi interval terhadap proporsi kanker payudara pada populasi adalah sebagai berikut:

■ Pada tingkat kepercayaan 90%

$$= 0.04 \pm \{1.64\sqrt{[(0.04)(0.96)/10.000]}\}$$

$$= 0.04 \pm 0,00321$$

$$= 0.037 < p < 0.043$$

■ Pada tingkat kepercayaan 95%

$$0.04 - [1.96\sqrt{(0.04)(0.96)/10.000}] < p < 0.04 + [1.96\sqrt{(0.04)(0.96)/10.000}]$$

$$0.036 < p < 0.044$$

Faktor yang Berpengaruh Terhadap Lebar Interval

- Variasi data
Diukur dengan Ukuran sampel
- Tingkat kepercayaan

Ciri-ciri Estimator yang Baik

- a. Tidak Bias: hasil estimasinya mengandung nilai parameter yang diestimasi
- b. Efisien: dalam rentang (interval) yang kecil atau sempit saja sudah mengandung nilai parameter yang diestimasi
- c. Konsisten: berapapun besarnya sampel pada rentangnya (interval) mengandung nilai parameter yang diestimasi

LATIHAN

Guna mengetahui berapa rata-rata berat bayi baru lahir, dilakukan pengamatan terhadap bayi yang lahir di rumah sakit di suatu Kabupaten. Bayi yang lahir dipilih secara random dari 3 rumah sakit. Hasilnya seperti berikut (berat dalam kg):

Rumah sakit 1 :	3,5	3,0	2,0	2,6	2,7	2,8	2,4
Rumah sakit 2 :	4,2	3,0	1,9	3,8	3,6	3,5	
Rumah sakit 3 :	2,9	2,8	2,6	3,0	2,4		

1. Tentukan berat lahir bayi di masing-masing rumah sakit ? (Mean & SD)
2. Bila data ketiga rumah sakit disatukan, berapa rata-rata berat lahir bayi di Kabupaten tersebut?
3. Hitung perkiraan rata-rata berat lahir di Kabupaten tsb dengan confidence interval 95%.
4. Bila untuk studi di atas, menurut aturan WHO ditentukan bahwa bayi yang lahir dengan berat <2500 gram dikategorikan BBLR. Hitunglah:
5. Berapa proporsi BBLR di masing-masing rumah sakit tersebut?
6. Berapa perkiraan proporsi BBLR di kabupaten tsb dengan confidence interval 95%.
7. Sajikan hasilnya dalam bentuk Tabel (Memuat hasil pada soal no 1 s.d no 5)

BAB VIII

UJI-HIPOTESIS

HIPOTESIS

Hipotesis berasal dari kata hupo dan thesis. Hupo artinya sementara/lemah kebenarannya dan thesis artinya pernyataan/teori. Dengan demikian hipotesis berarti pernyataan sementara yang perlu diuji kebenarannya. Untuk menguji kebenaran sebuah hipotesis digunakan pengujian yang disebut pengujian hipotesis. Didalam pengujian hipotesis dijumpai dua jenis hipotesis, yaitu hipotesis nol (H_0) dan hipotesis alternatif (H_a).

a. Hipotesis Nol (H_0)

Hipotesis yang menyatakan tidak ada perbedaan suatu kejadian antara kedua kelompok (kejadian pada kedua kelompok adalah sama). Atau hipotesis yang menyatakan tidak ada hubungan antara variabel satu dengan variabel yang lain.

Contoh:

1. Tidak ada perbedaan berat badan bayi antara mereka yang dilahirkan dari ibu yang merokok dengan mereka yang dilahirkan dari ibu yang tidak merokok.
2. Tidak ada hubungan antara ibu merokok dengan berat badan bayi.

b. Hipotesis Alternatif (H_a)

Hipotesis yang menyatakan ada perbedaan kejadian antara kedua kelompok. Atau hipotesis yang menyatakan ada hubungan variabel satu dengan variabel yang lain.

Contoh :

1. Ada perbedaan berat badan bayi antara mereka yang dilahirkan dari ibu yang merokok dengan mereka yang dilahirkan dari ibu yang tidak merokok. (*2-tailed; BBBayi ibu perokok bisa lebih tinggi atau lebih rendah*) → *arahnya tidak/belum jelas*).
2. Ada hubungan merokok dengan berat badan bayi. (*2-tailed; BBBayi ibu perokok bisa lebih tinggi atau lebih rendah*) → *arahnya tidak/belum jelas*
3. Rata-2 berat bayi yang lahir dari ibu perokok lebih rendah dibandingkan dengan ibu yang

tidak perokok. (1-tailed) → *arahnya jelas*

ARAH UJI HIPOTESIS

Bentuk pernyataan hipotesis alternatif akan menentukan arah uji statistik, apakah satu arah (one tail) atau dua arah (two tail).

1. **One tail** (satu arah/sisi): bila hipotesis alternatifnya menyatakan adanya arah, hal yang satu lebih tinggi atau lebih rendah dari hal yang lain. ($A > B$ atau $B < A$)

Contoh :

Rata-rata berat badan bayi dari ibu hamil yang merokok lebih rendah dibandingkan rata-rata berat badan bayi dari ibu hamil yang tidak merokok.

Proporsi BBLR dari ibu hamil yang merokok lebih tinggi dibandingkan dengan ibu hamil yang tidak merokok.

2. **Two Tail** (dua sisi) merupakan hipotesis alternatif yang hanya menyatakan adanya perbedaan, **tanpa** bisa menduga apakah hal yang satu lebih tinggi atau lebih rendah dari hal yang lain.

Contoh

Berat badan bayi dari ibu hamil yang peminum alkohol berbeda dibandingkan berat badan bayi dari ibu yang tidak peminum alkohol.

Atau dengan kata lain: Ada perbedaan berat badan bayi antara mereka yang dilahirkan dari ibu yang peminum alkohol dibandingkan dari mereka yang tidak peminum alkohol.

Ada perbedaan rata-rata berat badan bayi lahir menurut tingkat pendidikan ibu (SD, SMP, SMA+).

Ada hubungan antara berat ibu (kg) dengan berat bayi lahir (gram)

Hipotesis Two Tail Biasanya dipakai untuk masalah-2 yang masih baru atau belum jelas pengaruhnya.

Contoh penulisan hipotesis:

Suatu penelitian ingin mengetahui hubungan antara jenis kelamin dengan tekanan darah, maka hipotesisnya adalah sbb:

$H_0 : \mu_A = \mu_B :$

Tidak ada perbedaan mean tekanan darah antara laki-laki dan perempuan, atau

Tidak ada hubungan antara jenis kelamin dengan tekanan darah

$H_a : \mu_A \neq \mu_B :$

Ada perbedaan mean tekanan darah antara laki-laki dan perempuan, atau

Ada hubungan antara jenis kelamin dengan tekanan darah

KESALAHAN PENGAMBILAN KEPUTUSAN

Dalam pengujian hipotesis kita selalu dihadapkan suatu kesalahan pengambilan keputusan.

Ada dua jenis kesalahan pengambilan keputusan dalam uji statistik, yaitu:

a. **Kesalahan Tipe I (α)**

Merupakan kesalahan yang terjadi karena menolak H_0 padahal sesungguhnya H_0 benar. Artinya: menyimpulkan adanya perbedaan padahal sesungguhnya tidak ada perbedaan. Peluang kesalahan tipe satu (I) adalah α atau sering disebut Tingkat signifikansi (significance level). Sebaliknya peluang untuk tidak membuat kesalahan tipe I adalah sebesar $1-\alpha$, yang disebut dengan Tingkat Kepercayaan (confidence level).

b. **Kesalahan Tipe II (β)**

Merupakan kesalahan yang terjadi karena tidak menolak H_0 padahal sesungguhnya H_0 salah. Artinya: menyimpulkan tidak ada perbedaan padahal sesungguhnya ada perbedaan. Peluang untuk membuat kesalahan tipe kedua (II) ini adalah sebesar β . Peluang untuk tidak membuat kesalahan tipe kedua (II) adalah sebesar $1-\beta$, dan dikenal sebagai Tingkat Kekuatan Uji (power of the test).

ILUSTRASI KESALAHAN AKIBAT PENGAMBILAN SAMPEL

POPULASI BUMIL (A)

Desa		Kota	
No	Kadar Hb	No	Kadar Hb
1	9.2	1	9.2
2	9.6	2	9.6
3	9.8	3	9.8
4	10.2	4	10.2
5	10.6	5	10.6
6	10.8	6	10.8
7	11.2	7	11.2
8	11.6	8	11.6
9	11.8	9	11.8
10	12.0	10	12.0
Mean	10.68		10.68

POPULASI

SAMPEL BUMIL

Desa		Kota	
No	Kadar Hb	No	Kadar Hb
1	9.2	5	10.6
3	9.8	7	11.2
5	10.6	10	12.0
Mean	9.87		11.27

→ **P-value = 0.010** → Ho ditolak

Ho: Rata-rata Hb Bumil di Kota **sama** dengan di Desa

Ha: Rata-rata Hb Bumil di Kota **lebih tinggi** dari di Desa (simpulan)

- Di populasi **Ho betul (tidak ada perbedaan)**, tetapi hasil sampel menyimpulkan **Ho ditolak**/Ho salah (**ada perbedaan**).

Kesalahan yang terjadi akibat **menyimpulkan ada perbedaan padahal sebetulnya tidak ada perbedaan** (menolak Ho yang betul) = Type-I error (alpha)

POPULASI BUMIL (B)

Buruh Pabrik		Ibu rumah tangga	
No	Kadar Hb	No	Kadar Hb
1	9.2	1	9.2
2	9.6	2	9.6
3	9.8	3	9.8
4	9.4	4	10.2
5	9.5	5	10.6
6	9.7	6	10.8

POPULASI

	7	9.8	7	11.2
	8	10.2	8	11.6
	9	10.6	9	11.8
	10	10.8	10	12.0
Mean		9.86		10.68

SAMPEL BUMIL

	Buruh Pabrik		Ibu rumah tangga	
	7	9.8	2	9.6
	8	10.2	4	10.2
	9	10.6	6	10.8
Mean		10.2		10.2

P-value = 0.999 → Ho diterima

Ho: Rata-rata Hb Bumil yang Pekerja **sama** dengan Bukan Pekerja

Ha: Rata-rata Hb Bumil yang Pekerja **lebih rendah** dari Bukan Pekerja

- Pada populasi **Ho salah (ada perbedaan)**, tetapi hasil sampel menyimpulkan **Ho gagal ditolak /Ho betul (tidak ada perbedaan)**.

Kesalahan yang terjadi akibat menyimpulkan **tidak ada perbedaan padahal sebenarnya DI POPULASI ada perbedaan** (gagal menolak Ho yang salah) = Type-II error (betha)

KESALAHAN YANG MUNGKIN TERJADI PADA UJI HIPOTESIS

Simpulan hasil Sampel	Populasi	
	H ₀ Salah (Ada perbedaan)	H ₀ Benar (Tdk ada perbedaan)
Ho ditolak (Ada perbedaan)	1-β(Power) <i>Rejecting fals Ho</i>	α (type-I error) <i>Rejecting true Ho</i>
Ho gagal ditolak/Ho diterima (Tdk ada perbedaan)	β (type-II error) <i>Do not reject true Ho</i>	1-α (Interval kepercayaan) <i>Do not reject fals Ho</i>
Jumlah	1.0	1.0

α Batas kesalahan tipe 1 yg boleh diperbuat (1% (klinis), 5% (kesmas), 10% (sosbud))
(Menghukum orang yg difitnah telah membunuh)

HARUS KECIL

β Batas kesalahan tipe-2 yg boleh diperbuat (5% (klinis), 10% (kesmas), 20% (sosbud))
(Membebaskan seorang pembunuh)

BOLEH AGAK BESAR DIKIT DARI ALPHA

PENENTUAN α dan β DISESUAIKAN DENGAN KEBIASAAN (penelitian sebelumnya dibidang yang ingin kita teliti)

Hakim membebaskan pembunuh \rightarrow XXX (type-2: alpha)

\rightarrow boleh lebih besar probabilitas salahnya (5%, 10%, 20%)

Hakim menghukum org yg tidak bersalah \rightarrow XXXX (type-1: alpha)

\rightarrow harus lebih kecil probabilitas salahnya (1%, 5%, 10%)

Semakin kecil tingkat kesalahan yang diinginkan, semakin besar jumlah sampel yang dibutuhkan

KEMAKNAAN STATISTIK VERSUS KEMAKNAAN SUBSTANSI

Tabel 1

Kebiasaan	PJK		n
	Ya	Tidak	
minum teh			
Ya	12 (12%)	88	100
Tidak	10 (10%)	90	100
Jumlah	22	178	200
$X^2 = 0,20$ $p = 0,6513$			

\rightarrow Ho diterima

Ho = Kebiasaan minum teh tidak berhubungan dengan PJK.

Ha = Kebiasaan minum teh berhubungan dengan PJK

Pada tabel 2, sampel penelitian ditambah menjadi 20 kali lipat dari tabel 1, sedangkan proporsi penyakit jantung koroner pada kedua kelompok yang diuji tetap sama (12% dan 10%).

Tabel 2

Kebiasaan minum teh	Penyakit Jantung Koroner		Jumlah
	Ya	Tidak	
Ya	240 (12%)	1760	2000
Tidak	200 (10%)	1800	2000
Jumlah	440	3560	4000

$X^2 = 4,09$ $p = \mathbf{0,0432} \rightarrow$ Ho ditolak

Ho = Kebiasaan minum teh tidak berhubungan dengan PJK.

Ha = Kebiasaan minum teh berhubungan dengan PJK

- Uji statistik menunjukkan $p=0,0432$, sehingga pada $\alpha=0,05$ kita menyimpulkan ada perbedaan proporsi penyakit jantung koroner pada orang yang memiliki kebiasaan minum teh dengan yang tidak memiliki kebiasaan minum teh.
 - Perbedaan proporsi sebesar 2% yang tidak terdeteksi pada tabel 1 dapat terdeteksi pada tabel 2. Namun peneliti perlu mempertimbangkan apakah perbedaan kejadian penyakit jantung koroner sebesar **2%** memang **bermakna secara ilmu kedokteran?**
- Untuk itu perlu sampel minimum untuk mendeteksi secara substansi

+ Disease -	Analysis of Single Table
+-----+-----+	Odds ratio = 1.23 (0.47 <OR< 3.26)
+ 12 88 100	Cornfield 95% confidence limits for OR
+-----+-----+	Relative risk = 1.20 (0.54 <RR< 2.65)
- 10 90 100	Taylor Series 95% confidence limits for RR
+-----+-----+	Ignore relative risk if case control study.
22 178 200	

Chi-Squares P-values

----- -----

Uncorrected : 0.20 0.6512801
Mantel-Haenszel: 0.20 0.6520953
Yates corrected: 0.05 0.8212074

+ Disease -	Analysis of Single Table
+-----+-----+	Odds ratio = 1.23 (1.00 <OR< 1.50)
+ 240 1760 2000	Cornfield 95% confidence limits for OR
+-----+-----+	Relative risk = 1.20 (1.01 <RR< 1.43)
- 200 1800 2000	Taylor Series 95% confidence limits for RR
+-----+-----+	Ignore relative risk if case control study.
440 3560 4000	

Chi-Squares P-values

----- -----

Uncorrected : 4.09 0.0432449
Mantel-Haenszel: 4.08 0.0432711
Yates corrected: 3.88 0.0487464

LANGKAH-LANGKAH PENGUJIAN HIPOTESIS

1. Nyatakan H_0 dan H_a

- H_0 : Tidak ada perbedaan A dg B *atau* $A = B$
- H_a : Ada perbedaan A dg B *atau* $A \neq B$ (2-tailed)
- H_a : A lebih besar dari B *atau* $A > B$ (1-tailed)
- H_a : P lebih kecil dari Q *atau* $P < Q$ (1-tailed)

H_a merupakan hypothesis penelitian yang diyakini oleh peneliti

2. Tentukan derajat kemaknaan (α): 0.01, 0.05, 0.1

0.01 → Berhubungan dengan obat + nyawa manusia

0.05 → Berhubungan dengan kesehatan/kesakitan

0.10 → Berhubungan dengan sosial budaya

3. Tentukan jenis uji statistik yg cocok:

a → (uji beda proporsi, beda rata², uji-korelasi)?

b → Tentukan area penolakan H_0 dg melihat nilai-tabel pada α tertentu

4. Hitung nilai-uji-statistik

5. Hitung p-value *atau* Bandingkan nilai-statistik dg nilai-tabel

6. Keputusan:

- jika | nilai uji-statistik | \geq nilai-tabel *atau* p-value $\leq \alpha$ → H_0 ditolak
 - jika | nilai uji-statistik | $<$ nilai-tabel *atau* p-value $> \alpha$ → H_0 gagal ditolak
- Uji 2-sisi (2-tailed) menggunakan tanda nilai mutlak | |
Uji 1-sisi (1-tailed) tanpa menggunakan tanda nilai mutlak
→ Lebih jelasnya, dapat digambarkan area penolakan

7. Kesimpulan: - jika H_0 ditolak → simpulkan H_a

- jika H_0 gagal ditolak → simpulkan H_0

Penentuan jenis uji statistik yang cocok tergantung pada:

1. Skala pengukuran: NOIR → proporsi (kategorik), rata-rata (numerik)
2. Tujuan uji statistik:
 - uji 1-sampel
 - uji 2-sampel
 - uji k-sampel
 - uji korelasi atau hubungan
3. Jenis sampel:
 - independent
 - paired (berpasangan) / dependent
4. Asumsi uji statistik:
 - Parametrik: jika - sampel random
 - distribusi normal

- varians sama (pd kondisi ttt)
- Non-parametrik (tidak perlu asumsi)

GUIDE TO THE CHOICE OF STATISTICAL TEST FOR HYPOTHESIS TESTING

TYPE OF SAMPLE	SCALE OF MEASUREMENT		
	NOMINAL (kategorik)	ORDINAL atau numerik tetapi Distribusi tidak normal)	RATIO/INTERVAL (Numerik & Distribusi normal)
<p>I. One-sample case (Determine if sample is from a population with pre-specified Proportion or Mean)</p> <p>(Determine if sample is from a population with pre-specified distribution)</p> <p>(Test for randomness)</p>	<ul style="list-style-type: none"> Binomial test χ^2 goodness of fit test 	<ul style="list-style-type: none"> Kolmogorov-Smirnov one sample test Runs test 	<ul style="list-style-type: none"> t-test for one mean z-test for one mean
<p>II. Two-sample cases (Comparison of 2 groups)</p> <p>- Related</p>	<ul style="list-style-type: none"> McNemar change test or Marginal χ^2 (sebelum-sesudah) <p>Nominal</p>	<ul style="list-style-type: none"> Sign test Wilcoxon signed rank test (sebelum-sesudah) <p>O atau I-R, tapi Distribusi tdk normal</p>	<ul style="list-style-type: none"> Pared t-test <p>I-R: (sebelum-sesudah) Distribusi normal</p>
<p>- Independent</p>	<ul style="list-style-type: none"> Fisher exact test for 2 x 2 tables χ^2 test for r x 2 tables 	<ul style="list-style-type: none"> Wilcoxon-Mann-Whitney test Median test Kolmogorov-Smirnov two sample test 	<ul style="list-style-type: none"> t-test for independent samples z-test for independent samples
<p>III. k-sample cases (Comparison of more than 2 groups)</p> <p>- Related</p>	<ul style="list-style-type: none"> Cochran Q test 	<ul style="list-style-type: none"> Friedman two-way ANOVA by ranks Page test for ordered relative 	<ul style="list-style-type: none"> Two-way ANOVA
<p>- Independent</p>	<ul style="list-style-type: none"> χ^2 test for r x k tables 	<ul style="list-style-type: none"> Extention of Median test Kruskal-Wallis one-way ANOVA 	<ul style="list-style-type: none"> One-way ANOVA

		<ul style="list-style-type: none"> ▪ Jonckheere test for ordered alternatives 	
IV. Measures of Association	<ul style="list-style-type: none"> ▪ Cramer V coefficient ▪ Phi coefficient ▪ Kappa coefficient 	<ul style="list-style-type: none"> ▪ Spearman rank-order correlation coefficient ▪ Kendall rank-order correlation coefficient 	<ul style="list-style-type: none"> ▪ Pearson correlation coefficient (simple and multiple)

BAB IX

UJI BEDA DUA MEAN

Pengambil kebijakan maupun penelitian dalam bidang kesehatan seringkali ingin melihat apakah dua populasi memiliki karakteristik yang sama atau berbeda satu sama lain dan jika berbeda, perbedaannya bukanlah suatu kebetulan (*by chance*). Untuk menguji hal tersebut sampel diambil dari dua populasi yang berbeda dengan asumsi mereka mewakili nilai yang ada di masing-masing populasi untuk selanjutnya nilai kedua sampel tersebut dibandingkan. Ilustrasi gambarnya adalah sebagai berikut :



Contoh :

1. Ingin diketahui apakah ada perbedaan rata-rata berat badan bayi lahir antara ibu yang anc baik dan ibu yang anc kurang.
2. Ingin diketahui apakah ada perbedaan rata-rata hb ibu sebelum dan sesudah pemberian tablet Fe.

Sebelum kita melakukan pengujian statistik dua kelompok data, perlu diperhatikan apakah dua kelompok data tersebut berasal dari kelompok yang sama/independen atau berasal dari dua kelompok yang sama/dependen/berpasangan. Kelompok data independen adalah bila data kelompok yang satu tidak tergantung pada data kelompok kedua, misalnya membandingkan rata-rata berat badan bayi pada kelompok ibu yang anc baik dan rata-rata berat badan bayi pada kelompok ibu yang anc kurang baik, sedangkan kelompok dikatakan dependen apabila dua kelompok data yang dibandingkan datanya

saling mempunyai ketergantungan, misalnya data hb ibu sebelum dan sesudah pemberian tablet Fe. Kita harus mencocokkan dan memastikan setiap data ibu yang diambil sebelum pemberian tablet Fe sama dengan data ibu sesudah pemberian tablet Fe (subjek yang sama), jika tidak ada data yang bisa dibandingkan.

Alasan pentingnya untuk mengenali karakteristik data adalah karena terdapat dua uji beda 2 rata-rata/mean yang berbeda. Uji beda dua mean dibagi dalam dua kelompok, yaitu: uji beda dua mean independen/*independen sample t-test* (untuk 2 sampel yang independen) dan uji beda mean dependen / *paired t-test* (untuk 2 sampel dependen/ data sampel sebelum dan sesudah).

Asumsi berikut harus dipenuhi untuk uji beda 2 mean :

1. Data pada kedua sampel berdistribusi normal
2. Kedua kelompok data independen
3. Variabel yang dihubungkan adalah katagorik dan numerik.

Prosedur untuk menguji perbedaan 2 sampel :

1. Buat Hipotesis nol dan Hipotesis alternatif
2. Pilih level signifikansi, (biasanya $\alpha=0.05$ atau $\alpha=0.01$)
3. Hitung melalui rumus statistik
4. Tentukan area kritis (critical region)
5. H_0 akan ditolak jika nilai t berdasarkan uji statistik berada pada area kritis (tail) dan H_0 akan ditolak jika nilai t berdasarkan uji statistik berada pada area kritis (tail) dan H_0 gagal ditolak jika nilai t berdasarkan uji statistik berada diluar pada area kritis (tail).
6. Hitung nilai p (p value)
7. Bandingkan nilai p dengan nilai α
8. Hitung confidence interval (95% atau 99%)
9. Buat kesimpulan

Prinsip pengujiannya adalah melihat perbedaan variasi kedua kelompok data. Oleh karena itu dalam pengujian ini diperlukan informasi apakah varian kedua kelompok yang diuji sama atau tidak. Bentuk varian ini akan berpengaruh pada standar error yang akhirnya akan membedakan rumus pengujiannya. Untuk itu perlu dilakukan pengujian

varians dengan uji F. Tujuan uji ini adalah untuk mengetahui varian antara kelompok data satu, apakah sama dengan kelompok data yang kedua.

Adapun rumus untuk uji Varians (Uji F) :

$$F = \frac{S_1^2}{S_2^2}$$

Dengan :

Varian yang lebih besar sbg pembilang (numerator)

Varian yang lebih kecil sebagai penyebut (denominator)

Jika $p \text{ value} < \alpha$ H_0 gagal ditolak : Varian sama

Jika $p \text{ value} > \alpha$ H_0 ditolak : Varian berbeda

$df_1 = n_1 - 1$ dan $df_2 = n_2 - 1$

1. Uji untuk Varian Sama.

Menggunakan uji Z atau uji t. Uji Z digunakan apabila standar populasi deviasi populasi (σ) diketahui dan jumlah sampel besar (>30). Apabila dua syarat tersebut tidak dipenuhi maka dilakukan uji t. Pada umumnya σ sulit diketahui sehingga digunakan uji t.

$$t = \frac{x_1 - x_2}{S_p \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}}$$

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

dengan :

x_1 atau x_2 = rata-rata sampel klp 1 atau 2

n_1 atau n_2 = jumlah sampel klp 1 atau 2

S_1 atau S_2 = standar deviasi sampel klp 1 atau 2

df = degree of freedom (derajat kebebasan)

df = $n_1 + n_2 - 2$

2. Uji untuk Varian Berbeda

$$t = \frac{x_1 - x_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Dengan degree of freedom :

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

Uji Beda Dua Mean Dependen

Bertujuan untuk menguji perbedaan mean antara dua kelompok data yang dependen.

Contoh kasus :

- Apakah ada perbedaan perbedaan rata-rata berat badan bayi sebelum dan sesudah pemberian ASI Eksklusif.
- Apakah ada perbedaan berat badan balita sebelum dan sesudah pemberian makanan tambahan.

Syarat :

1. Distribusi data normal
2. Kedua kelompok data dependen / pair
3. Jenis variabel yang dihubungkan numerik dan katagori

Formula uji T dependen :

$$t = \frac{d}{Sd_d/\sqrt{n}}$$

df = n - 1

d = rata-rata deviasi/selisih sampel 1 dengan sampel 2

Sd_d = standar deviasi dari deviasi/selisih sampel 1 dan sampel 2

Contoh Data:

50 ibu / responden diambil untuk menjadi sampel, diperoleh data sebagai berikut :

Tabel.1.1. Distribusi data Ibu di Kabupaten X Tahun 2012

No_resp	Umur	BB_Bayi	ANC	Hb 1	Hb 2	Pendidikan	Sosek
1	19	3600	0	9.8	14.6	3	3
2	18	3300	1	9.6	12.4	3	3
3	17	2800	1	9.1	12.5	3	2
4	19	2500	0	9.8	14.7	1	2
5	19	2400	0	9.3	13.4	1	2
6	19	2100	1	9.4	13.5	1	1
7	20	2300	1	9.4	13.5	1	1
8	20	2600	0	9.6	14.1	2	3

9	20	2700	0	9.5	13.5	2	3
10	20	2300	0	9.5	13.6	2	2
11	21	2000	0	9.5	13.7	1	1
12	21	3000	0	9.9	14.8	1	2
13	21	2200	1	9.6	14.2	1	1
14	22	2800	0	9.7	14.3	2	1
15	22	2900	0	9.7	14.3	2	2
16	22	3000	0	9.8	14.6	2	2
17	22	2900	1	9.7	15.5	3	3
18	22	2900	0	10.1	15.5	3	3
19	23	2600	1	9.4	13.4	3	1
20	23	2400	1	9.2	12.7	3	2
21	23	3400	1	9.3	13.3	3	2
22	23	3100	1	9.5	13.7	1	2
23	23	2600	1	9.5	13.8	1	2
24	24	3500	0	9.6	14.2	4	3
25	24	3000	0	9.7	14.5	4	2
No_resp	Umur	BB_Bayi	ANC	Hb 1	Hb 2	Pendidikan	Sosek
26	24	3100	0	9.8	10.2	4	2
27	24	3200	0	9.8	10.2	4	2
28	25	2700	1	9.2	9.1	3	2
29	25	2800	0	9.2	9.1	3	3
30	26	2200	1	11.2	10	1	1
31	26	2900	0	11.2	10	1	3
32	26	2500	1	11.2	10	1	3
33	26	2600	1	13.2	12.3	3	3
34	26	3400	0	13.2	12.3	3	3
35	27	2700	1	11.2	10	4	3
36	27	2500	1	11.2	10	4	1
37	27	2900	0	11.2	10	4	2
38	28	3300	0	11.2	10	3	3
39	28	3700	0	11.2	10	3	2
40	29	3500	0	11.2	10	3	2
41	29	3000	0	13.2	12.3	4	2
42	30	2800	0	13.2	12.3	4	3
43	31	2800	0	13.2	12.3	4	3
44	32	2400	1	13.2	12.3	4	2
45	32	3300	0	13.2	12.3	4	3
46	32	3200	0	13.2	12.3	4	3
47	34	3100	0	11.1	11.5	2	2
48	34	2700	1	11.1	11.5	2	2
49	35	3100	0	10.2	9.8	3	2
50	35	3200	0	10.2	9.8	3	3

Kasus 1.

Dari data diatas ingin diketahui apakah ada perbedaan rata-rata berat badan bayi lahir antara ibu dengan pemeriksaan kehamilan (anc) yang baik dan tidak.

Jawaban

Diketahui

anc baik		anc kurang	
X_1	= 3135.71	X_2	=2486.364
Sd_1	= 257.069	Sd_2	= 233.596
n_1	= 28	n_2	= 22

Prosedur untuk menguji perbedaan 2 sampel Independen:

1. Hipotesis nol dan Hipotesis alternatif

H_0 : Tidak ada perbedaan rata-rata berat badan bayi lahir antara ibu dengan pemeriksaan kehamilan (anc) yang baik dan tidak.

H_a : Ada perbedaan rata-rata berat badan bayi lahir antara ibu dengan pemeriksaan kehamilan (anc) yang baik dan tidak.

2. Two Tail (dua arah), Level signifikansi $\alpha=0.05$

3. Uji Homogenitas Varians

H_0 : Varian berat badan bayi dengan ibu dengan anc baik sama dengan varian berat badan ibu dengan anc kurang

H_a : Varian berat badan bayi dengan ibu dengan anc baik berbeda dengan varian berat badan ibu dengan anc kurang

$$F = \frac{S_1^2}{S_2^2} = \frac{257,0694^2}{233,596^2} = 1,21$$

$$df_1 = 28-1 = 27 \text{ (Numerator)}$$

$$df_2 = 22-1 = 21 \text{ (Denominator)}$$

Tabel F :

Denominator	.. 24	∞
	P	
20	0,100	1,61
	0,050	1,84
	0,025	2,09
	0,010	2,42
	0,005	2,69
	0,001	3,35

F=1,21

- Hasilnya dari tabel terlihat bahwa nilai F=1,21 berada di atas nilai 1,61 karena tabel memiliki pola semakin ke atas nilainya semakin kecil sehingga diperoleh p value lebih besar dari 0,10 (karena untuk pola p value semakin ke atas semakin besar).
- Perbandingan p value dan α : p value $>$ α (lebih besar dari 0,10 $>$ 0,05) sehingga keputusannya H_0 gagal ditolak atau **varians sama**.
- Uji beda dua mean dengan varian sama :

$$t = \frac{x_1 - x_2}{s_p \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}} = \frac{3135,71 - 2486,36}{247,07 \sqrt{\frac{28 + 22}{28 \cdot 22}}} = \frac{649,35}{70,39} = 9,225$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(28 - 1)257,069^2 + (22 - 1)233,596^2}{(28 - 1) + (22 - 1)} = 61.045,617$$

$$s_p = 247,07$$

- Diperoleh nilai t hitung sebesar 9,225
- df = $n_1 + n_2 - 2 = 28 + 22 - 2 = 48$
- Hitung nilai p (p value) dengan tabel t

df \ α	0.10	0.05	0.025	0.01	0.005	P
40	1,303	1,684	2,021	2,423	2,704	
60						t=9,225

Nilai t = 9,225 dalam tabel berada di sebelah kanan dari 2,704 karena terlihat pola tabel nilai t semakin ke kanan nilainya semakin besar, sehingga dari tabel diatas terlihat bahwa

p value berada pada posisi lebih kecil dari 0,005 (memiliki pola semakin ke kanan semakin kecil nilainya)

9. Bandingkan nilai p dengan nilai α

0,05 < 0,005 , Ho ditolak sehingga kesimpulannya ada perbedaan rata-rata berat badan bayi lahir antara ibu dengan pemeriksaan kehamilan (anc) yang baik dan tidak.

Kasus 2

Dari data yang tersedia ingin diketahui apakah ada perbedaan rata-rata hb pada ibu sebelum dan sesudah pemberian tablet Fe

Jawaban

Prosedur untuk menguji perbedaan 2 sampel Dependen:

1. Hipotesis nol dan Hipotesis alternatif

Ho : Tidak ada perbedaan rata-rata kadar hb ibu sebelum dan sesudah pemberian tablet Fe

Ha : Ada perbedaan rata-rata kadar hb ibu sebelum dan sesudah pemberian tablet Fe.

2. Level signifikansi $\alpha=0.05$

3. Hitung dengan rumus :

$$t = \frac{d}{Sd_d/\sqrt{n}}$$

No_resp	Hb 1	Hb 2	deviasi	No_resp	Hb 1	Hb 2	deviasi
1	9.8	14.6	-4.8	26	9.1	12.6	-3.5
2	9.6	12.4	-2.8	27	9.2	12.8	-3.6
3	9.1	12.5	-3.4	28	9.4	12.2	-2.8
4	9.8	14.7	-4.9	29	9.4	12.3	-2.9
5	9.3	13.4	-4.1	30	9.4	12.9	-3.5
6	9.4	13.5	-4.1	31	9.6	14.2	-4.6
7	9.4	13.5	-4.1	32	9.8	14.7	-4.9
8	9.6	14.1	-4.5	33	9.9	15.2	-5.3
9	9.5	13.5	-4	34	9.6	15.6	-6
10	9.5	13.6	-4.1	35	9.5	13.8	-4.3
11	9.5	13.7	-4.2	36	9.5	13.9	-4.4
12	9.9	14.8	-4.9	37	9.5	13.7	-4.2
13	9.6	14.2	-4.6	38	9.2	13.1	-3.9

14	9.7	14.3	-4.6	39	9.6	14.2	-4.6
15	9.7	14.3	-4.6	40	9.7	14.2	-4.5
16	9.8	14.6	-4.8	41	9.5	15.5	-6
17	9.7	15.5	-5.8	42	10	15.4	-5.4
18	10.1	15.5	-5.4	43	10	15.3	-5.3
19	9.4	13.4	-4	44	9	12.4	-3.4
20	9.2	12.7	-3.5	45	9.7	15.2	-5.5
21	9.3	13.3	-4	46	9.9	15.1	-5.2
22	9.5	13.7	-4.2	47	9.6	14.1	-4.5
23	9.5	13.8	-4.3	48	9.6	14.1	-4.5
24	9.6	14.2	-4.6	49	9.3	13.2	-3.9
25	9.7	14.5	-4.8	50	9.3	13.2	-3.9

$$df = n - 1 = 50 - 1 = 49$$

$$d = \text{rata-rata deviasi/selisih sampel 1 dengan sampel 2} = -4,39$$

$$Sd_d = \text{standar deviasi dari deviasi/selisih sampel 1 dan sampel 2} = 0,759$$

$$t = \frac{d}{Sd_d/\sqrt{n}} = \frac{-4,39}{0,759/\sqrt{50}} = -40.9$$

4. Hitung nilai p (p value) dengan tabel t

df \ α	0.10	0.05	0.025	0.01	0.005	P
40	1,303	1,684	2,021	2,423	2,704	t=40,9
60						

Dari tabel diatas terlihat bahwa nilai 40.9 berada di sebelah kanan dari 2,704 karena terlihat pola tabel nilai t semakin ke kanan nilainya semakin besar, sehingga dari tabel diatas terlihat bahwa p value berada pada posisi lebih kecil dari 0,005 (memiliki pola semakin ke kanan semakin kecil nilainya).

5. Bandingkan nilai p dengan nilai α

$0,05 < 0,005$, H_0 ditolak sehingga kesimpulannya ada perbedaan rata-rata kadar hb ibu sebelum dan sesudah pemberian tablet Fe.

BAB X

ANALISIS OF VARIANS (ANOVA)

A. Pengertian dan Dasar Penggunaan Anova

Analysis of Varians atau disingkat Anova merupakan salah satu uji statistik yang digunakan dalam menganalisis perbedaan lebih dari dua mean atau uji beda mean lebih dari dua kelompok data. Uji ini merupakan perluasan dari uji t yang menguji beda mean dua kelompok data. Sebenarnya, kelompok data lebih dari dua dapat dilakukan dengan uji t, yaitu dengan melakukan pengujian berulang kali sesuai kombinasi yang mungkin. Namun kelemahan dari uji t untuk menguji lebih dari dua kelompok data adalah dapat meningkatkan peluang hasil yang keliru ($\alpha = 1 - (1 - \alpha)^n$).⁽¹⁾

Uji Anova dikenal juga dengan Uji F, berasal dari Sir Ronald Fisher yang pertama kali mengembangkan analisis ini pada awal abad 20.⁽²⁾ Pada dasarnya Uji Anova melakukan telaah variabilitas data menjadi dua sumber variasi, yaitu variasi dalam kelompok (*within*) dan variasi antar kelompok (*between*). Jika variasi *within* dan *between* sama (nilai perbandingan kedua varian sama dengan 1), mean-mean yang dibandingkan tidak ada perbedaan. Sebaliknya, jika variasi *within* dan *between* berbeda (nilai perbandingan kedua varian lebih dari 1), mean-mean yang dibandingkan menunjukkan adanya perbedaan.⁽¹⁾

Anova terdiri dari dua jenis, yaitu analisis varian satu faktor (*one way*) dan analisis varian dua faktor (*two way*) atau dua variabel independen.⁽¹⁾⁽³⁾ Analisis varian satu faktor (*one way*), misalnya ingin menguji perbedaan mean berat badan bayi diantara 4 tingkat pendidikan ibu yang berbeda. Sedangkan analisis varian dua faktor (*two way*), misalnya ingin menguji perbedaan mean berat badan bayi berdasarkan tingkat pendidikan ibu dan sosial ekonomi.

Uji Anova dapat digunakan apabila terpenuhi beberapa asumsi, sebagai berikut:⁽¹⁾

1. Varian homogen.
2. Sampel/kelompok independen.
3. Data berdistribusi normal.

Cara untuk mengetahui data berdistribusi normal, yaitu dilihat dari grafik histogram dan kurva normal, hasil bagi nilai skewness dan standar error menghasilkan angka ≤ 2 atau hasil uji signifikan dengan Uji Kolmogorov Smirnov (untuk jumlah data > 50) atau Uji Shapiro Wilk (untuk jumlah data ≤ 50) diperoleh nilai $p > 0,05$.

4. Jenis data yang dihubungkan adalah numerik dengan kategori (untuk kategori yang lebih dari dua kelompok).

B. Prosedur Uji Anova

Prosedur Uji Anova One Way dan Two Way secara manual sebagai berikut:

1. Uji Anova One Way

Prosedur Uji Anova One Way sebagai berikut:

- 1) Menentukan hipotesis nol (H_0) dan hipotesis alternatif (H_a)
 - a. $H_0: \mu_1 = \mu_2 = \mu_3 \dots = \mu_k$, k merupakan jumlah dari kelompok
Artinya variasi antara mean tidak lebih besar dari variasi diantara pengamatan individu dalam kelompok.
 - b. $H_a: \mu_1 \neq \mu_2$ atau $\mu_1 \neq \mu_3$ atau $\mu_1 \neq \mu_k$ atau $\mu_2 \neq \mu_3$ atau $\mu_2 \neq \mu_k$
Artinya satu atau lebih nilai mean berbeda dengan yang lain.
- 2) Melakukan uji statistik, dengan perhitungan sebagai berikut:

$$F = \frac{Sb^2}{Sw^2}$$

Keterangan :

Sb = Varian *between*

Sw = Varian *within*

$df_1 = k-1 \rightarrow$ untuk pembilang

$df_2 = N - k \rightarrow$ untuk penyebut

N = Jumlah seluruh data ($n_1 + n_2 + \dots + n_k$)

n = Jumlah data di masing-masing kelompok data

X = mean

Langkah-langkah dalam uji statistik dengan Anova, yaitu :

- a. Menentukan nilai X dengan perhitungan berikut :

$$X = \frac{n_1 \cdot X_1 + n_2 \cdot X_2 + \dots + n_k \cdot X_k}{N}$$

- b. Menentukan nilai Sb^2 dengan perhitungan berikut :

$$Sb^2 = \frac{n_1(X_1 - X)^2 + n_2(X_2 - X)^2 + \dots + n_k(X_k - X)^2}{k - 1}$$

- c. Menentukan nilai Sw^2 dengan perhitungan berikut :

$$Sw^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2 + \dots + (n_k - 1)S_k^2}{N - k}$$

- d. Setelah diperoleh nilai Sb^2 dan Sw^2 , selanjutnya nilai tersebut dimasukkan dalam perhitungan/rumus F sehingga diperoleh nilai F hitung.

- 3) Membuat asumsi berdasarkan hasil perbandingan antara nilai F yang diperoleh dari hasil perhitungan (F hitung) dengan nilai F tabel (mengacu pada Tabel Distribusi F/terlampir).

Sebelum menentukan Nilai F tabel pada Tabel Distribusi F, terlebih dahulu ditentukan nilai df (*degree of freedom*) numerator, df denominator dan area. Untuk df numerator diperoleh melalui pengurangan jumlah kelompok data dengan angka 1 atau $df = k - 1$ (sebagai pembilang) sedangkan df denominator diperoleh melalui pengurangan jumlah seluruh data dengan jumlah kelompok data atau $df = n - k$ (sebagai penyebut). Kemudian tentukan posisi nilai F hitung pada Tabel Distribusi F dengan melihat df numerator pada bagian atas kolom dan df denominator pada bagian kiri baris sehingga dapat dilihat nilai F tersebut berada di bagian area yang menunjukkan F tabel dan nilai p. Selanjutnya dapat dibuat asumsi sebagai berikut:

- Apabila F hitung lebih kecil dari F tabel maka H_0 diterima atau nilai $p \geq 0,05$ maka Hipotesis Nol (H_0) diterima atau gagal ditolak, artinya tidak adanya perbedaan mean antar kelompok data.
- Apabila F hitung lebih besar dari F tabel maka H_0 ditolak atau nilai $p < 0,05$ maka Hipotesis Nol (H_0) ditolak atau Hipotesis alternatif diterima, artinya adanya perbedaan mean antar 1 atau lebih kelompok data.

- 4) Penyajian dalam tabel sebagai berikut :

Tabel.....Analisis Perhitungan Uji Anova One-Way

Sumber Variasi	Jumlah Kuadrat	Derajat Kebebasan	Mean Kudrat (S^2)	F Hitung	F Kritis	Nilai P
Between	SSb	k - 1	$MSb = SSb/k - 1$	MSb/MSw	$F_{k-1,N-k}$	
Within	SSw	N - k	$MSw = SSw/N - k$			
Total	SSt	N - 1				

Contoh kasus (Uji Anova One Way) :

Suatu penelitian ingin mengetahui perbedaan Berat Badan Lahir (BBL) bayi (dalam gram) pada empat tingkat pendidikan ibu (SD, SMP, SMA dan Perguruan Tinggi/PT). Data yang berhasil dikumpulkan sebagai berikut :

Tabel 1. Berat Badan Lahir Bayi berdasarkan Tingkat Pendidikan Ibu

No.	Berat Badan Lahir Bayi (gram)			
	Tingkat Pendidikan Ibu			
	SD (kelompok	SMP	SMA	PT

	1)	(kelompok 2)	(kelompok 3)	(kelompok 4)
1.	2500	2600	3600	3500
2.	2400	2700	3300	3000
3.	2100	2300	2800	3100
4.	2300	2800	2900	3200
5.	2000	2900	2900	2700
6.	3000	3000	2600	2500
7.	2200	3100	2400	2900
8.	3100	2700	3400	3000
9.	2600	-	2700	2800
10.	2200	-	2800	2800
11.	2900	-	2600	2400
12.	2500	-	3400	3300
13.	-	-	3300	3200
14.	-	-	3700	-
15.	-	-	3500	-
16.	-	-	3100	-
17.	-	-	3200	-

Pertanyaan : Apakah ada perbedaan berat badan lahir bayi pada keempat kelompok ibu dengan tingkat pendidikan berbeda dengan alpha 5%.

Jawab :

Sebelum dilakukan Uji Anova, data BBL ini sudah diperiksa terlebih dahulu untuk mengetahui apakah memenuhi beberapa asumsi seperti yang dijelaskan pada bagian paragraf pertama pada bab ini. Setelah terpenuhi beberapa asumsi kemudian dilakukan langkah-langkah sebagai berikut:

1. Hipotesis :

- a. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, artinya tidak ada perbedaan mean berat badan lahir bayi pada keempat tingkat pendidikan ibu
- b. $H_1 : \mu_1 \neq \mu_2$ atau $\mu_1 \neq \mu_3$ atau $\mu_1 \neq \mu_4$ atau $\mu_2 \neq \mu_3$ atau $\mu_2 \neq \mu_4$ atau $\mu_3 \neq \mu_4$
Minimal ada 1 pasang mean berat badan lahir bayi yang berbeda pada keempat tingkat pendidikan ibu

2. Perhitungan Uji Anova :

$$F = \frac{Sb^2}{Sw^2}$$

Jumlah Data = N = 50

Kelompok 1 : mean (X) = 2483,33, standar deviasi = 358,87, n = 12

Kelompok 2 : mean (X) = 2762,5, standar deviasi = 250,36, n = 8

Kelompok 3 : mean (X) = 3070,59, standar deviasi =390,14, n = 17

Kelompok 4 : mean (X) = 2953,85, standar deviasi = 315,21, n = 13

$$a. \quad X = \frac{n1.X1 + n2.X2 + n3.X3 + n4.X4}{N}$$

$$X = \frac{(12)(2483,33) + (8)(2762,5) + (17)(3070,59) + (13)(2953,85)}{50} = 2850$$

$$b. \quad Sb^2 = \frac{n1(X1-X)^2 + n2(X2 - X)^2 + n3 (X3 - X)^2 + n4 (X4 - X)^2}{k - 1}$$

$$Sb^2 = \frac{(12)(2483,33-2850)^2 + (8)(2762,5-2850)^2 + (17)(3070,59-2850)^2 + (13)(2953,85-2850)^2}{4-1}$$

$$Sb^2 = 880678,17$$

$$c. \quad Sw^2 = \frac{(n1-1) S1^2 + (n2 - 1)S2^2 + (n3 - 1)S3^2+ (n4 - 1)S4^2}{N - k}$$

$$Sw^2 = \frac{(12-1)(358,87)^2 + (8-1)(250,36)^2 + (17-1)(390,14)^2 + (13-1)(315,21)^2}{50-4} = 119196,98$$

$$d. \quad F = \frac{880678,17}{119196,98} = 7,39$$

3. Asumsi :

F hitung = 7,39, df1 = 4-1 = 3, df2 = 50-4 = 46, pada Tabel Distribusi F menunjukkan F hitung terdapat pada area berikut :

Denominator (df2)	Area	Numerator (df1)				
		1	2	3	4	---
40	0,100	---	---	2,23	---	---
	0,050	---	---	2,84	---	---
	0,025	---	---	3,46	---	---
	0,010	---	---	4,31	---	---
	0,005	---	---	4,98	---	---
	0,001	---	---	6,59	---	---

← **F=7,39**

Pada Tabel Distribusi F dapat dilihat nilai F hitung terdapat setelah nilai F tabel 6,59 dan area setelah 0,001 atau < 0,001, artinya nilai p = < 0,001 sehingga diputuskan H0 ditolak.

Dengan demikian dengan alpha 5% dapat disimpulkan bahwa ada perbedaan mean berat badan lahir bayi diantara keempat kelompok tingkat pendidikan ibu atau 1 atau lebih kelompok pendidikan ibu memiliki perbedaan mean berat badan lahir bayi.

4. Hasil

Tabel....Analisis Perhitungan Uji Anova One-Way

Sumber Variasi	Jumlah Kuadrat	Derajat Kebebasan	Mean Kudrat (S ²)	F Hitung	F Kritis	Nilai P
Between	938,44455	3	880678,17	7,39	6,59	< 0,001
Within	345,24916	46	119196,98			
Total	1283,6937	49				

Untuk mengetahui lebih lanjut kelompok mana saja yang berbeda mean-nya maka dapat dilakukan **Analisis Multi Comparison**.

2. Uji Anova Two Way

Uji Anova Two Way pada prinsipnya langkah-langkah dalam perhitungan secara manual hampir sama dengan Uji Anova One Way. Dalam Uji Anova Two Way digunakan formula perhitungan sebagai berikut:

Tabel....Analisis Perhitungan Uji Anova Two-Way

Sumber Variasi	Jumlah Kuadrat	Derajat Kebebasan	Mean Kudrat (S ²)	F Hitung	F Kritis	Nilai P
Between	SSb	r-1	MSb = SSb/r-1	F1 = MSb/MSw		
- Faktor A	SSbA	a-1	MSbA = SSbA/a-1	F2 = MSbA/MSw		
- Faktor C	SSbB	c-1	MSbB = SSbB/k-1	F3 = MSbC/MSw		
- Interaksi AC	SSbAC	(a-1)(c-1)	MSbAC = SSbAC/(a-1)(c-1)	F4 = MSbAC/MSw		
Within	SSw	N-ac	MSw = SSw/N-ac			
Total	SSt	N-1				

$$SS_{Total} = \sum X^2 - \frac{(\sum X)^2}{N}$$

$$SS_{Between} = \frac{(\sum X_1)^2}{n_1} + \frac{(\sum X_2)^2}{n_2} + \dots + \frac{(\sum X_{ac})^2}{n_{ac}} - \frac{(\sum X)^2}{N}$$

$$SS_A = \sum \frac{(\sum \text{for each row})^2}{n \text{ for each row}} - \frac{(\sum X)^2}{N}$$

$$SS_C = \sum \frac{(\sum \text{for each column})^2}{n \text{ for each column}} - \frac{(\sum X)^2}{N}$$

$$SS_{AC} = SS_{Between} - SS_A - SS_C$$

$$SS_{Within} = SS_{Total} - SS_{Between}$$

C. Analisis Multi Comparison

Banyak metoda yang dapat digunakan dalam Analisis Multi Comparison (*Post Hoc Test*). Kebanyakan metoda tersebut digunakan untuk perbandingan pasangan kelompok mean, yaitu untuk menentukan kelompok mean mana yang berbeda secara bermakna dengan kelompok lainnya atau dilakukan untuk mengetahui lebih lanjut kelompok mana yang berbeda mean-nya apabila hasil uji Anova diperoleh asumsi H_0 ditolak. Metoda Analisis Multi Comparison lainnya digunakan untuk tujuan tertentu, misalnya membandingkan tiap beberapa perlakuan dengan kontrol atau menguji hipotesis yang lebih umum untuk membandingkan kelompok mean (seperti sering dilakukan pada perbandingan yang belum direncanakan).

Berbagai metoda berbeda dalam hal seberapa baik metoda tersebut mengontrol secara keseluruhan tingkat kemaknaan dan kekuatan relatif. Beberapa jenis Analisis Multi Comparison, antara lain Bonferroni, *Tukey's Honestly Significant Different*, *Scheffe's Procedure*, *The Newman-Keuls Procedure*, *Dunnett's Procedure*, *Duncan's New Multiple Range Test* dan *Fisher's Least Significant Difference (LSD) method*.⁽³⁾

1. Bonferroni

Bonferroni atau disebut juga prosedur *Dunn's Multi Comparison* merupakan Analisis Multi Comparison yang sering digunakan. Metoda Bonferroni adalah metoda yang sangat umum, sederhana dan lebih praktis digunakan sepanjang nilai k (jumlah kelompok) yang dibandingkan tidak terlalu besar.⁽⁴⁾ Namun dikatakan tingkat kekuatan (*power*) metoda ini lemah. Perhitungan metoda Bonferroni sebagai berikut:

$$t_{ij} = \frac{x_i - x_j}{\sqrt{S_w 2 \left[\frac{1}{n_i} - \frac{1}{n_j} \right]}}$$

$$df = N - k$$

Level of Significance (α) pada metoda Bonferroni sebagai berikut :

$$\alpha^* = \frac{\alpha}{k_2}$$

Dari contoh kasus 1 dapat ditelusuri lebih lanjut dari keempat kelompok tingkat pendidikan ibu, kelompok mana saja yang berbeda mean berat badan lahir bayinya:

Kombinasi uji t yang mungkin adalah $\binom{4}{2} = \frac{4!}{2!(4-2)!} = 6$

Pada soal diatas digunakan alpha 5% maka α bonferroni adalah menjadi :

$$\alpha^* = \frac{0,001}{6} = 0,000167$$

a. Uji kelompok I dan II :

$$t_{12} = \frac{2483,33 - 2762,5}{\sqrt{119196,98 \left[\frac{1}{12} + \frac{1}{8} \right]}} = -1,77$$

Kemudian lihat nilai p dengan menggunakan tabel t dengan df = 50 - 4 = 46

Degree of freedom (df)	Area untuk dua pihak				
	0,10	0,05	0,025	0,01	0,005
40	1,303	1,684	2,021	2,423	2,704

← $t = 1,77$

Dengan nilai $t = 1,77$ dan $df = 46$, maka nilai $p < 0,05$ dan $> 0,025$ sehingga nilai ini lebih besar dari nilai $\alpha^* = 0,000167$ maka hipotesis nol diterima. Dengan demikian dapat disimpulkan bahwa tidak ada perbedaan mean berat badan lahir bayi antara kelompok ibu berpendidikan SD (kelompok I) dengan pendidikan SMP (kelompok II).

b. Uji kelompok I dan III :

$$t_{13} = \frac{2483,33 - 3070,59}{\sqrt{119196,98 \left[\frac{1}{12} + \frac{1}{17} \right]}} = -4,51$$

Kemudian lihat nilai p dengan menggunakan tabel t dengan df = 50 - 4 = 46

Degree of freedom (df)	Area untuk dua pihak				
	0,10	0,05	0,025	0,01	0,005
40	1,303	1,684	2,021	2,423	2,704

← t = 4,5

Dengan nilai t = - 4,51 dan df = 46, maka nilai p < 0,005 sehingga nilai ini kemungkinan lebih kecil dari nilai $\alpha^* = 0,000167$ maka hipotesis nol ditolak. Dengan demikian dapat disimpulkan bahwa ada perbedaan mean berat badan lahir bayi antara kelompok ibu berpendidikan SD (kelompok I) dengan pendidikan SMA (kelompok III).

c. Uji kelompok I dan IV :

$$t_{14} = \frac{2483,33 - 2953,85}{\sqrt{119196,98 \left[\frac{1}{12} + \frac{1}{13} \right]}} = -3,40$$

Kemudian lihat nilai p dengan menggunakan tabel t dengan df = 50 - 4 = 46

Degree of freedom (df)	Area untuk dua pihak				
	0,10	0,05	0,025	0,01	0,005
40	1,303	1,684	2,021	2,423	2,704

← t = 3,40

Dengan nilai t = - 3,40 dan df = 46, maka nilai p < 0,005 sehingga nilai ini kemungkinan lebih kecil dari nilai $\alpha^* = 0,000167$ maka hipotesis nol ditolak. Dengan demikian dapat disimpulkan bahwa ada perbedaan mean berat badan lahir bayi antara kelompok ibu berpendidikan SD (kelompok I) dengan pendidikan perguruan tinggi (kelompok IV).

d. Uji kelompok II dan III :

$$t_{23} = \frac{2762,5 - 3070,59}{\sqrt{119196,98 \left[\frac{1}{8} + \frac{1}{17} \right]}} = -2,08$$

Kemudian lihat nilai p dengan menggunakan tabel t dengan $df = 50 - 4 = 46$

Degree of freedom (df)	Area untuk dua pihak				
	0,10	0,05	0,025	0,01	0,005
40	1,303	1,684	2,021	2,423	2,704

← $t = 2,08$

Dengan nilai $t = -2,80$ dan $df = 46$, maka nilai $p < 0,05$ dan $> 0,025$ sehingga nilai ini kemungkinan lebih besar dari nilai $\alpha^* = 0,000167$ maka hipotesis nol diterima. Dengan demikian dapat disimpulkan bahwa tidak ada perbedaan mean berat badan lahir bayi antara kelompok ibu berpendidikan SMP (kelompok II) dengan pendidikan SMA (kelompok III).

e. Uji kelompok II dan IV :

$$t_{24} = \frac{2762,5 - 2953,85}{\sqrt{119196,98 \left[\frac{1}{8} + \frac{1}{13} \right]}} = -1,23$$

Kemudian lihat nilai p dengan menggunakan tabel t dengan $df = 50 - 4 = 46$

Degree of freedom (df)	Area untuk dua pihak				
	0,10	0,05	0,025	0,01	0,005
40	1,303	1,684	2,021	2,423	2,704

← $t = 1,23$

Dengan nilai $t = -1,23$ dan $df = 46$, maka nilai $p > 0,10$ sehingga nilai ini kemungkinan lebih besar dari nilai $\alpha^* = 0,000167$ maka hipotesis nol diterima. Dengan demikian dapat disimpulkan bahwa tidak ada perbedaan mean berat badan lahir bayi antara kelompok ibu berpendidikan SMP (kelompok II) dengan pendidikan perguruan tinggi (kelompok IV).

f. Uji kelompok III dan IV :

$$t_{34} = \frac{3070,59 - 2953,85}{\sqrt{119196,98 \left[\frac{1}{17} + \frac{1}{13} \right]}} = 0,92$$

Kemudian lihat nilai p dengan menggunakan tabel t dengan $df = 50 - 4 = 46$

Degree of freedom (df)	Area untuk dua pihak				
	0,10	0,05	0,025	0,01	0,005
40	1,303	1,684	2,021	2,423	2,704

40	1,303	1,684	2,021	2,423	2,704
----	-------	-------	-------	-------	-------

Dengan nilai $t = 0,92$ dan $df = 46$, maka nilai $p > 0,10$ sehingga nilai ini kemungkinan lebih besar dari nilai $\alpha^* = 0,000167$ maka hipotesis nol diterima. Dengan demikian dapat disimpulkan bahwa tidak ada perbedaan mean berat badan lahir bayi antara kelompok ibu berpendidikan SMA (kelompok III) dengan pendidikan perguruan tinggi (kelompok IV).

2. Prosedur Tukey's HSD

Uji *Tukey* atau *Tukey's Honestly Significant Different* (HSD) dapat digunakan untuk membandingkan mean semua kelompok pasangan. Stoline (1981) menemukan bahwa prosedur ini paling akurat dan paling kuat dalam melihat perbedaan nilai mean antar kelompok pasangan ketika besar sampel tidak sama atau dibutuhkannya interval kepercayaan bahkan dengan besar sampel sama tanpa interval kepercayaan. Uji *Tukey* hampir sama dengan Bonferroni yaitu berdasarkan jumlah tingkat perlakuan dan derajat kebebasan (*degree of freedom* = df) bagi kesalahan mean kuadrat (Sw) serta dapat digunakan untuk membentuk interval kepercayaan (*confidence intervals*) bagi perbedaan mean antar kelompok. Formula untuk perhitungan pada Metoda Scheffe's sebagai berikut:

$$\bar{Y}_i - \bar{Y}_j \pm Q_{1-\alpha, k, r} \sqrt{\frac{MSE}{n}}$$

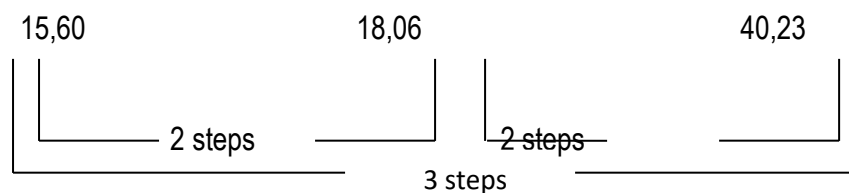
3. Prosedur Scheffe's

Prosedur Scheffe's merupakan prosedur paling fleksibel dari semua prosedur Post Hoc karena memungkinkan peneliti untuk membuat semua jenis perbandingan, bukan hanya perbandingan berpasangan. Contohnya membandingkan rata-rata keseluruhan dari dua atau lebih tingkat dosis dengan plasebo. Karena fleksibilitas metoda ini maka nilai kritis yang lebih tinggi digunakan untuk menentukan signifikansi. Dengan demikian, prosedur Scheffe adalah juga yang paling konservatif dari prosedur perbandingan ganda. Formula untuk perhitungan pada Metoda Scheffe's sebagai berikut:

$$S = \sqrt{(k - 1)F_{\alpha, k}} \sqrt{MS_E \sum \frac{C_j^2}{n_j}}$$

4. Prosedur The Newman-Keuls

Prosedur Newman-Keuls menggunakan pendekatan bertahap untuk menguji perbedaan antar mean dan hanya dapat digunakan untuk membuat perbandingan berpasangan. Prosedur ini menyusun rangking nilai mean dari nilai terendah ke tertinggi, dan jumlah langkah yang memisahkan pasangan nilai mean dicatat. Contohnya:



Gambar 1. Ranking mean dan Langkah-langkah untuk Tes Newman-Keuls

Nilai kritis untuk prosedur ini, sebagai berikut:

$$Newman - Keuls = Multiplier \sqrt{\frac{MS_E}{n}}$$

Nilai kritis untuk tiga tahapan ini sama dengan nilai kritis pada uji Tukey's HSD, tapi nilai kritis untuk dua tahapan lebih lemah. Walaupun kesimpulan sama dengan tes Tukey's, namun dengan prosedur Newman-Keuls pada beberapa kelompok, peneliti dapat menyatakan bahwa adanya perbedaan antara dua cara yang signifikan ketika perbedaan tidak akan signifikan dalam uji Tukey's HSD. Kelemahan utama prosedur ini adalah tidak dapat membentuk *confidence interval* bagi perbedaan nilai mean.

5. Prosedur Dunnett's

Prosedur Dunnett's berlaku hanya pada situasi dimana peneliti ingin membandingkan nilai mean beberapa perlakuan dengan nilai mean kontrol tunggal. Perbandingan antar nilai mean beberapa perlakuan tidak dilakukan jadi prosedur ini memiliki aplikasi yang sangat khusus. Pada situasi tertentu, Uji Dunnett's sangat mudah digunakan karena memiliki nilai kritis relatif rendah. Besaran *multiplier* (pengganda) tergantung jumlah kelompok, termasuk grup kontrol dan derajat kebebasan (*degree of freedom*) untuk eror mean kuadrat. Formula untuk prosedur ini sebagai berikut :

$$\bar{Y}_i - \bar{Y}_j \pm D_{1-\alpha, k-1, v} \sqrt{\frac{2MSE}{n}}$$

6. Uji *Duncan's New Multiple Range* dan *Fisher's Least Significant Difference* (LSD)

Uji *Duncan's New Multiple Range* dan *Fisher's Least Significant Difference* (LSD) merupakan dua prosedur yang muncul dalam literatur medis namun tidak direkomendasikan oleh ahli statistik. Uji *Duncan's New Multiple Range* menggunakan prinsip yang sama dengan tes Newman-Keuls. Tapi pengganda (*multipliers*) pada formula ini sangat kecil, jadi secara statistik perbedaan kemaknaan didapatkan perbedaan nilai mean yang kecil. Duncan, ahli yang menemukan uji ini berargumen bahwa ada banyak kemungkinan dalam menemukan perbedaan jika jumlah kelompok banyak, dan dia meningkatkan kekuatan uji-nya dengan menggunakan pengganda terkecil. Namun uji tersebut, sebagai hasilnya menolak hipotesis nol terlalu sering.

Uji *Fisher's Least Significant Difference* (LSD) adalah salah satu prosedur multi comparison yang paling 'tua'. Hampir sama dengan prosedur multi comparison lain, prosedur ini membutuhkan F rasio yang signifikan dari Anova dalam rangka membuat perbandingan antar pasangan kelompok data. Karena prosedur ini menggunakan penyesuaian untuk membuat nilai kritis lebih besar, seperti uji lain gunakan, Uji *Fisher's LSD* menggunakan distribusi t terkait jumlah derajat kebebasan bagi eror mean kuadrat. Alasan ahli statistik tidak merekomendasikan uji ini adalah dengan jumlah perbandingan yang besar, derajat alpha tiap perbandingan turun dan perbedaan menjadi kecil karena tidak benarnya tingkat kemaknaan yang turun. Formula untuk perhitungan pada Uji *Fisher's LSD* sebagai berikut:

$$\bar{Y}_i - \bar{Y}_j \pm t_{1-\alpha/2, v} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Masing-masing prosedur pada Multi Comparison ini memiliki keterbatasan dan kelebihan, seperti terangkum dalam tabel 1 berikut:

Tabel 1. Perbedaan Keterbatasan dan Kelebihan Masing-Masing Metoda Multi Comparison⁽⁵⁾

Procedure:	Equal Sample Size?	False Positive Rate (Type I Error)	False Negative Rate (Type II Error)	Comments:
Bonferroni ¹	No	Low	High	Adjusts α for t -tests
Fisher's LSD ²	No	High	Low	Most liberal
Duncan MRT ³	Yes	Moderate	Moderate	Not widely recommended
SNK ⁴	Yes	Moderate	Moderate	
Hochberg (GT2)	No	Low	Moderate	
Tukey-Kramer	No	Low	Moderate	More powerful than GT2
Scheffè	No	Low	High	Most conservative
Dunnnett	No	Low	Moderate	Compares treatment means to a control mean

¹Also called Bonferroni-Dunn. ²LSD = Least Significant Differences. ³MRT = Multiple Range Test. ⁴SNK = Student-Newman-Keuls test.

DAFTAR PUSTAKA

1. Kuzma, Jan W dan Bhnenblust, Stephen E, Basic Statistic for the Health Science, McGraw-Hill Company, New York, 2005
2. Saunders, Beth Dawson dan Trapp, Robert G, Basic & Clinical Biostatistics, Appleton & Lange, California, 1994.
3. Priyohastomo, Sutanto dan Sabri, Luknis, Statistik Kesehatan, Rajawali Press, Jakarta, 2010
4. Sanders, Donald h, Statistics A First Course, Mc.Graw Hill, 1995

BAB XI

UJI CHI SQUARE

Pendahuluan

Seringkali dalam suatu penelitian, kita menemui data yang tidak dapat dinyatakan dalam bentuk angka-angka pengukuran (data numerik). Sebaliknya justru yang kita jumpai adalah data hasil dari menghitung jumlah pengamatan yang diklasifikasi atas beberapa kategori. Data seperti ini disebut data kategori (kualitatif). Misalnya jenis kelamin yang mempunyai kategori laki-laki dan perempuan, status merokok yang mempunyai kategori: perokok berat, perokok ringan dan tidak merokok.

Banyak data hasil pengamatan yang dapat digolongkan ke dalam beberapa faktor (variabel), karakteristik atau atribut dengan tiap faktor (variabel) atau atribut terdiri dari beberapa klasifikasi, kategori, golongan atau mungkin tingkatan. Berdasarkan hasil pengamatan terhadap fenomena tersebut akan diselidiki mengenai hubungan atau kaitan antar faktor (variabel). Dengan kata lain akan dipelajari apakah terdapat hubungan atau tidak di antara variabel-variabel itu. Jika ternyata tidak terdapat hubungan antar variabel maka dikatakan variabel-variabel itu bersifat independen atau bebas.

Misalnya kita ingin mengetahui hubungan jenis pekerjaan dengan perilaku menyusui ibu, apakah ada perbedaan proporsi kejadian menyusui eksklusif antara ibu yang bekerja dengan ibu yang tidak bekerja. Dari contoh ini terlihat bahwa variabel jenis pekerjaan (bekerja/tidak bekerja) merupakan variabel kategori, dan perilaku menyusui (eksklusif/non eksklusif) juga merupakan variabel kategori.

Tujuan Uji Chi Square

Secara spesifik uji chi square dapat digunakan untuk menentukan/menguji:

- a. Ada tidaknya hubungan/asosiasi antara 2 variabel (*test of independency*)
- b. Apakah suatu kelompok homogen dengan sub kelompok lain (*test of homogeneity*)
- c. Apakah ada kesesuaian antara pengamatan dengan parameter tertentu yang dispesifikasikan (*Goodness of fit*)

Asumsi Penggunaan Uji Chi Square

Penggunaan Uji Chi Square menggunakan asumsi:

- a. Jenis data kategori
- b. Sampel independen
- c. Distribusi tidak normal/tidak diketahui distribusinya (*free distribution*)

Prinsip Dasar Uji Chi Square

Proses pengujian chi square adalah membandingkan frekuensi yang terjadi (observasi) dengan frekuensi harapan (ekspektasi). Bila nilai frekuensi observasi dengan nilai harapan sama, maka dikatakan tidak ada perbedaan yang bermakna (signifikan). Sebaliknya, bila nilai frekuensi observasi dan nilai frekuensi harapan berbeda, maka dikatakan ada perbedaan yang bermakna (signifikan).

Pembuktian dengan uji chi square dengan menggunakan formula:

$$\chi^2 = \sum \frac{(O - E)^2}{E} \dots\dots\dots(1)$$

Dalam taraf kemaknaan = α dan derajat kebebasan (*degree of freedom*=df) untuk distribusi chi square adalah (B-1)(K-1)

O= nilai observasi (pengamatan)

E = nilai ekspektasi (harapan)

B = jumlah baris

K = jumlah kolom

Untuk mempermudah analisis chi square, nilai data antar variabel disajikan dalam bentuk tabel silang:

Tabel Kontingensi B X K

Untuk hasil pengamatan Terdiri Atas Dua Variabel

Variabel I	Variabel II				Jumlah
	1	2	...	K	
1	O ₁₁	O ₁₂		K _{1K}	N _{1O}
2	O ₂₁	O ₂₂		O _{2K}	N _{2O}
.
.
B	O _{B1}	O _{B2}	...	O _{BK}	N _{BO}
Jumlah	N _{O1}	N _{O2}	...	N _{OK}	N

Nilai ekspektasi adalah nilai yang kita harapkan terjadi sesuai dengan hipotesis penelitian. Dihitung dengan perkalian antara nilai marginal kolom dan baris yang bersangkutan dibagi dengan jumlah seluruhnya (N) atau *grand total* yang terletak pada sudut kanan tabel kontingensi.

Contoh untuk mencari nilai ekspektasi setiap sel adalah:

$$E_{ij} = \frac{(N_{i0} \times N_{0j})}{N} \dots\dots\dots(2)$$

Dengan N_{i0} = jumlah baris ke-I (marginal baris)

N_{0j} = jumlah kolom ke-j (marginal kolom)

Misalnya:

$$E_{11} = (N_{10} \times N_{01})/N$$

$$E_{21}=(N_{20} \times N_{01})/N \text{ dan seterusnya}$$

Jika tabelnya berukuran 2 x 2, maka untuk pengujian hipotesis digunakan distribusi chi-square dengan derajat kebebasan satu. Untuk hal ini koreksi kontinuitas perlu digunakan dan telah ditemukan koreksi Yates (*Yate's Correction*), yaitu setiap harga mutlak dikurangi dengan setengah. $|O - E|$

Hasil pengamatan dapat disajikan dalam tabel kontingensi 2 x 2 di bawah ini:

Variabel I	Variabel II		Jumlah
	Ya	Tidak	
Ya	a	b	a + b
Tidak	c	d	c + d
Jumlah	a + c	b + d	N

Khususnya untuk tabel 2x2 dapat mencari nilai χ^2 dengan menggunakan Formula Chi Square Yate's Correction adalah:

$$\chi^2 = \sum \frac{(|O - E| - 0,5)^2}{E} \dots\dots\dots(3)$$

atau:

$$\chi^2 = \frac{N(|ad - bc| - \frac{1}{2}N)^2}{(a + b)(c + d)(a + c)(b + d)} \dots\dots\dots(4)$$

Contoh:

Suatu penelitian ingin mengetahui efektivitas penggunaan helm pengaman kendaraan bermotor dalam mencegah terjadinya cedera kepala. Dari 793 sampel random dalam satu tahun didapatkan hasil sebagai berikut:

Cedera kepala	Penggunaan helm		Jumlah
	Ya	Tidak	
Ya	17	218	235
Tidak	130	428	558
Jumlah	147	646	793

Karena tabelnya adalah 2x2, maka dihitung dengan formula:

$$\chi^2 = \frac{(|17 - 43,6| - 0,5)^2}{43,6} + \frac{(|130 - 103,4| - 0,5)^2}{103,4} + \frac{(|218 - 191,4| - 0,5)^2}{191,4} + \frac{(|428 - 454,6| - 0,5)^2}{454,6}$$

$$= 15,62 + 6,59 + 3,56 + 1,50 = 27,27$$

Dengan menggunakan formula 4 didapatkan:

$$\chi^2 = \frac{N(|ad - bc| - \frac{1}{2}N)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$\chi^2 = \frac{793(|17 \times 428 - 218 \times 130| - \frac{1}{2} \times 793)^2}{235 \times 558 \times 147 \times 646}$$

$$\chi^2 = \frac{793(|50076 - 28340| - 396,5)^2}{12452367060}$$

$$\chi^2 = \frac{793(21339,5)^2}{12452367060}$$

$$\chi^2 = 28,99$$

Selama nilai $\chi^2 \geq \chi^2$ tabel dengan α dan $df=1$, kita menolak H_0 dan menyimpulkan bahwa proporsi individu yang mengalami cedera kepala tidak sama pada kedua populasi.

OR dan RR

Meskipun uji chi square dapat untuk mengetahui hubungan antara dua variabel, namun tidak dapat menjelaskan kekuatan hubungan tersebut. Dalam bidang kesehatan, derajat hubungan dikenal dengan ukuran Risiko Relatif (RR) dan Odds Rasio (OR) khusus untuk tabel kontingensi 2x2.

Penyakit/ Kejadian	Paparan (<i>Exposure</i>)		Jumlah
	Ya	Tidak	
Ya	a	b	a+b
Tidak	c	d	c+d
Jumlah	a+c	b+d	n

RR membandingkan risiko pada kelompok ter-ekspose dengan kelompok tidak ter-ekspose. Dihitung dengan formula:

$$RR = \frac{a(b+c)}{b(a+c)} \dots\dots\dots(5)$$

Sedangkan **Odds rasio** membandingkan odds pada kelompok ter-ekspose dengan odds kelompok tidak ter-ekspose. Dihitung dengan formula:

$$OR = \frac{ad}{bc} \dots\dots\dots(6)$$

Ukuran RR pada umumnya digunakan pada desain kohort, sedangkan ukuran OR biasanya digunakan pada desain kasus kontrol atau potong lintang (*cross sectional*).

Contoh:

Pada penelitian hubungan antara konsumsi kopi dengan kanker pankreas diambil sebanyak 69 pasien kanker dan 114 sebagai kelompok kontrol. Hasil penelitian ditunjukkan dalam tabel berikut:

Kanker pankreas	Kebiasaan minum kopi (cangkir/hari)		Jumlah
	≥ 5	0	
Kasus	60	9	69
Kontrol	82	32	114
Jumlah	142	41	

$$OR = \frac{ad}{bc} = \frac{(60)(32)}{(9)(82)} = 2,6$$

Dari hasil tersebut kita dapat memperkirakan bahwa kebiasaan minum kopi akan meningkatkan risiko kanker pankreas pada laki-laki sebesar 2,6 kali dibandingkan pada laki-laki yang tidak minum kopi.

RR = 2 mengindikasikan bahwa kebiasaan meminum kopi berat berhubungan dengan kenaikan risiko kanker pankreas. Jadi, kopi menjadi faktor etiologi terjadinya kanker pankreas.

Keterbatasan Uji Chi Square

Uji chi square menuntut frekuensi harapan/ekspektasi (E) dalam masing-masing sel tidak boleh terlampau kecil. Jika frekuensi sangat kecil, penggunaan uji ini menjadi tidak tepat. Oleh karena itu harus diperhatikan keterbatasan uji chi square. Adapun keterbatasan Uji Chi Square adalah:

- a. Untuk tabulasi silang $> 2 \times 2$ dengan memperhatikan persyaratan:
 - 1) Tidak boleh ada frekuensi harapan yang lebih kecil 1 ($E < 1$)
 - 2) Boleh ada frekuensi harapan antara 1 dan 5 ($1 \leq E < 5$) maksimal 20% dari jumlah keseluruhan sel.

Apabila kedua persyaratan di atas tidak dipenuhi, maka penggabungan kategori perlu dilakukan agar diperoleh nilai harapan yang berharga besar
- b. Untuk tabulasi silang 2×2 dengan memperhatikan persyaratan:
 - 1) Bila terdapat nilai $E < 5$ maka digunakan *Uji Fisher Exact*
 - 2) Bila tidak ada nilai $E < 5$, maka dipakai *Continuity Correction*

Jika keterbatasan tersebut terjadi, peneliti harus menggabungkan kategori-kategori yang berdekatan dalam rangka memperbesar frekuensi harapan dari sel-sel tersebut. Penggabungan ini dapat dilakukan untuk analisis tabel silang lebih dari 2×2 , misal 3×2 , 3×4 dll. Asalkan tidak sampai membuat datanya kehilangan makna.
- c. Apabila keterbatasan tersebut terjadi pada tabel 2×2 (tidak bisa menggabung kategori-kategorinya lagi), maka dianjurkan menggunakan **Uji Fisher Exact** yang akan dibahas pada modul selanjutnya.

DAFTAR PUSTAKA

1. Kuzma, J.W. 1984. *Basic Statistics for the Health Sciences*. California: Meyfield Publishing Company.
2. Hastono, S.P. 2001. *Modul Analisis Data*. Jakarta : FKM UI.
3. Pagano, M. & Gauvreau, K. 1993. *Principles of Biostatistics*. California: Wadsworth Publishing Company
4. Dahlan, Sopiudin. 2008. *Statistik untuk Kedokteran dan Kesehatan*. Seri Evidence Based Medicine 2 Edisi 3. Jakarta: Penerbit Salemba Medika.

KORELASI DAN REGRESI LINEAR

Pada uji *chi square*, dapat diketahui apakah variabel yang diteliti mempunyai hubungan bermakna atau tidak; tetapi tidak diketahui seberapa besar dan bagaimana arah hubungan. Untuk mengetahui besar dan arah hubungan, uji yang digunakan adalah uji korelasi; karena variabel yang dihubungkan adalah variabel numerik dengan variabel numerik. Contohnya, suatu penelitian bertujuan untuk mengetahui hubungan tinggi badan dengan berat badan. tinggi badan dan berat badan merupakan variabel berjenis numerik. Uji regresi linear sederhana dapat dilanjutkan apabila variabel yang diteliti mempunyai korelasi yang signifikan pada uji korelasi.

Uji Regresi Linear sederhana bertujuan untuk memprediksi variabel dependen berdasarkan variabel independen. Contohnya adalah kita ingin memprediksi berat badan bayi yang dilahirkan dari seorang ibu bila berat badan ibu diketahui melalui persamaan garis regresi.

A. Persyaratan Uji Korelasi dan Regresi Linear Sederhana

Persyaratan uji korelasi dan regresi linear sederhana adalah sebagai berikut :

1. Data yang dikumpulkan berdistribusi normal
2. Kedua variabel yang dihubungkan berjenis numerik
3. Uji regresi linear sederhana dapat dilanjutkan bila uji korelasi signifikan.

B. Kekuatan Hubungan

Besar/kekuatan hubungan dapat dilihat dari nilai *r pearson produk moment*. Rumus *r pearson produk moment* adalah sebagai berikut :

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2 (n \sum Y^2 - (\sum Y)^2)}}$$

Nilai korelasi (*r*) berkisar 0 s.d 1 atau bila dengan disertai arahnya nilainya antara -1 s.d.+1.

$r=0$ → tidak ada hubungan linier

$r=-1$ → hubungan linier negatif sempurna

$r=+1$ → hubungan linier positif sempurna

Colton (1980) dalam Hastono (2007), membagi kekuatan hubungan dalam empat kategori, yaitu:

$r = 0,00-0,25 \rightarrow$ tidak ada hubungan - hubungan lemah

$r = 0,26-0,50 \rightarrow$ hubungan sedang

$r = 0,51-0,76 \rightarrow$ hubungan kuat

Contoh :

Tinggi Badan (in) dan Berat badan (lb) dari 12 mahasiswa adalah sebagai berikut :

Tinggi Badan	70	63	72	60	66	70	74	65	62	67	65	68
Berat Badan	155	150	180	135	156	168	178	160	132	145	139	152

Tentukan nilai *r pearson produk moment* :

Jawab :

Tinggi Badan	70	63	72	60	66	70	74	65	62	67	65	68	ΣX	802
Berat Badan	155	150	180	135	156	168	178	160	132	145	139	152	ΣY	1850
X^2	4900	3969	5184	3600	4356	4900	5476	4225	3844	4489	4225	4624	ΣX^2	53792
Y^2	24025	22500	32400	18225	24336	28224	31684	25600	17424	21025	19321	23104	ΣY^2	287868
XY	10850	9450	12960	8100	10296	11760	13172	10400	8184	9715	9035	10336	ΣXY	124258

$$r = \frac{12 * 124.258 - (802 * 1850)}{\sqrt{12 * 53.792 - (802)^2 (12 * 287.868 - (1850)^2)}} = 0,863$$

Diperoleh r pearson = 0,863, artinya hubungan tinggi badan dan berat badan mempunyai hubungan yang sangat kuat, arah hubungan adalah positif artinya semakin bertambah tinggi badan, semakin bertambah pula berat badan.

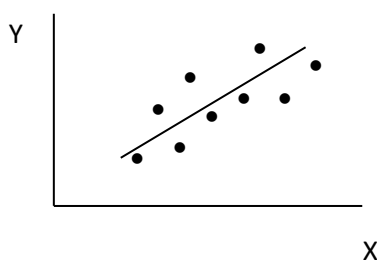
C. Arah hubungan

Arah hubungan dalam uji korelasi, yaitu arah hubungan positif dan arah hubungan negatif. Arah hubungan dapat dilihat dari tanda positif (+) atau (-) dari nilai koefisien pearson (r). Dikatakan kedua variabel mempunyai hubungan yang positif apabila semakin meningkat variabel X akan diikuti peningkatan variabel Y, sedangkan hubungan negatif adalah apabila terjadi penurunan variabel X, maka variabel Y akan meningkat atau sebaliknya. Contoh hubungan positif adalah semakin bertambahnya umur maka akan

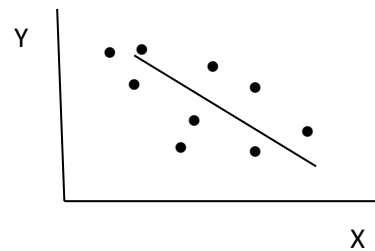
bertambah pula pengetahuan ibu. Contoh hubungan yang negatif adalah semakin bertambah kunjungan ibu hamil ke tenaga kesehatan, maka akan menurunkan Angka Kematian Bayi.

Secara visual, arah hubungan kedua variabel dapat dilihat dari grafik tebar (*scatter plot*). Scatter plot adalah grafik yang menunjukkan titik-titik perpotongan variabel X dengan variabel Y. Variabel X diletakkan pada garis horizontal, dan variabel Y diletakkan pada garis vertikal.

Hubungan positif



Hubungan Negatif



Tidak ada hubungan



D. Uji Korelasi pada SPSS

Kita akan menganalisis hubungan umur dengan tekanan darah sistole. Variabel Umur dan tekanan darah sistole merupakan variabel berjenis numerik.

Langkah-langkah Uji Korelasi dengan SPSS adalah sebagai berikut :

1. Buka data ASI.sav
2. Dari menu utama SPSS, klik 'Analyze', pilih 'Correlate', kemudian pilih 'Bivariate', maka akan muncul menu "bivariate Correlations"
3. Pindahkan variabel 'BB_ibu' dan 'BB_Bayi', lalu masukkan ke kotak sebelah kanan 'variables'.
4. Klik 'OK' dan hasil output adalah sebagai berikut :

Correlations

		bbibu	bbbayi
bbibu	Pearson Correlation	1	.693**
	Sig. (2-tailed)		.000
	N	46	46
bbbayi	Pearson Correlation	.693**	1
	Sig. (2-tailed)	.000	
	N	46	46

** . Correlation is significant at the 0.01 level

Berdasarkan tabel Correlations, dapat diketahui nilai r pearson adalah 0,693, artinya besarnya hubungan BB ibu dengan BB Bayi tergolong hubungan yang kuat dan mempunyai arah positif artinya semakin bertambah berat badan ibu, semakin bertambah pula berat badan bayi yang dilahirkan. Uji hipotesis diperoleh pvalue=0,000, keputusan uji adalah pvalue< α (0,05) maka Ho ditolak, kesimpulannya adalah terdapat hubungan yang signifikan berat badan ibu dengan berat badan bayi yang dilahirkan.

E. Regresi Linear Sederhana

Uji Regresi Linear sederhana dapat dilanjutkan apabila uji korelasi mempunyai hubungan yang signifikan (pvalue \leq 0,05). Analisis regresi merupakan suatu model matematis yang dapat digunakan untuk mengetahui bentuk hubungan antar dua atau lebih variabel. Tujuan analisis regresi adalah untuk membuat perkiraan atau prediksi nilai suatu variabel dependen (Y) melalui variabel independen (X). Contoh : suatu penelitian bertujuan untuk mengetahui hubungan umur dengan tekanan darah sistole. Variabel umur (X) dan variabel tekanan darah sistole (Y) merupakan variabel numerik. Dengan uji regresi, kita dapat memprediksi besarnya nilai tekanan darah sistole apabila data umur diketahui.

Untuk melakukan prediksi digunakan persamaan garis yang dapat diperoleh dengan berbagai cara/metode. Salah satu cara yang sering digunakan oleh peneliti adalah dengan menggunakan metode kuadrat terkecil (*least square*). Metode *least square* merupakan suatu metode pembuatan garis regresi dengan cara meminimalkan jumlah

kuadrat jarak antara nilai Y yang teramati dan Y yang diramalkan oleh garis regresi itu. Secara matematis persamaan garis regresi adalah sebagai berikut:

$$Y = a + bX + e$$

Y = Variabel Dependen

X = Variabel Independen

a = *Intercept* , perbedaan besarnya rata-rata variabel Y ketika variabel X=0

b = *Slope*, perkiraan besarnya perubahan nilai variabel Y bila nilai variabel X berubah satu unit pengukuran

e = nilai kesalahan (error) yaitu selisih antara nilai Y individual yang teramati dengan nilai Y yang sesungguhnya pada titik X tertentu

F. Kesalahan Standar Estimasi (Standar Error of Estimate/Se)

Besarnya kesalahan standar estimasi (Se) menunjukkan ketepatan persamaan estimasi untuk menjelaskan nilai variabel dependen yang sesungguhnya. Semakin tinggi ketepatan persamaan estimasi yang dihasilkan untuk menjelaskan nilai variabel dependen yang sesungguhnya.

G. Koefisien Determinasi (R^2)

Ukuran yang penting dan sering digunakan dalam analisis regresi adalah koefisien determinasi atau disimbolkan R^2 (*R Square*). Koefisien determinasi dapat dihitung dengan mengkuadratkan nilai r, atau dengan formula $R^2 = r^2$. Koefisien determinasi berguna untuk mengetahui seberapa besar variasi variabel dependen (Y) dapat dijelaskan oleh variabel independen (X) atau dengan kata lain R^2 menunjukkan seberapa tepat variabel independen dapat memprediksi variabel dependen. Semakin besar nilai R square semakin baik /semakin tepat variabel independen dapat memprediksi variabel dependen. Besarnya nilai R square antara 0 s.d. 1 atau antara 0% s.d 100%.

Regresi Linear Ganda

Analisis regresi digunakan untuk menganalisis hubungan variabel independen berjenis numerik dan variabel dependen yang juga berjenis numerik. Analisis regresi linear sederhana menghubungkan satu variabel independen numerik dan satu variabel dependen numerik, sedangkan pada Regresi Linear Ganda menganalisis hubungan antara beberapa variabel independen dengan satu variabel dependen.

Dalam regresi linier ganda variabel dependennya harus numerik sedangkan variabel independen boleh semuanya numerik dan boleh juga campuran numerik dan katagorik. Model persamaan regresi linier ganda merupakan perluasan regresi linier sederhana , yaitu:

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k + e$$

H . Diagnostik Regresi Linier

Persamaan garis linear mempunyai asumsi-asumsi yang harus dipatuhi. Asumsi tersebut dapat disingkat dengan istilah HEIL-Gauss, yang terdiri dari asumsi *Homocedasticity, Existence, Independence, Linearity* serta sebaran Gauss of Residual.

Apabila data yang kita olah, **tidak memenuhi ketentuan asumsi** tersebut diatas, maka persamaan garis linear yang dihasilkan dapat dikatakan tidak akurat. **Untuk itu, dianjurkan untuk menggantinya dengan jenis analisis statistik yang lain.** Langkah-langkah untuk memeriksa apakah asumsi HEIL-Gauss terpenuhi atau tidak, disebut sebagai **Diagnostik Regresi Linear.**

Tabel 1. Diagnostik regresi linear

Asumsi	Arti	Cara Diagnosis
<i>Homoscedasticity</i>	Variasi nilai variabel Y sama untuk semua nilai variabel X.	Homoscedasticity dapat diketahui dengan melakukan pembuatan plot residual. Bila titik tebaran tidak berpola tertentu dan menyebar merata disekitar garis titik nol maka dapat disebut varian homogen pada setiap nilai X dengan demikian asumsi homoscedasticity terpenuhi. Sebaliknya bila titik tebaran membentuk pola tertentu misalnya mengelompok dibawah atau digaris tengah nol, maka diduga variannya terjadi heterosdasticity
<i>Existence</i> (Eksistensi/Variasi Random)	Untuk tiap nilai dari variabel X (Variabel independen), Variabel Y (dependen) adalah variabel random yang mempunyai mean dan varian tertentu.	Cara mengetahui asumsi eksistensi dengan cara melakukan analisis deskriptif variabel residual dari model, bila residual menunjukkan adanya mean dan sebaran (varian atau standar deviasi) maka asumsi eksistensi terpenuhi.
<i>Independence</i> (Independensi)	Suatu keadaan dimana masing-masing nilai Y bebas satu sama lain. Jadi nilai dari tiap-tiap individu saling berdiri sendiri. Tidak diperbolehkan nilai observasi yang berbeda yang diukur dari satu individu diukur dua kali	Untuk mengetahui asumsi ini dilakukan dengan cara mengeluarkan uji Durbin Watson, bila nilai durbin Watson -2 s.d +2 berarti asumsi independensi terpenuhi, sebaliknya bila nilai durbin Watson <-2 atau >+2 berarti asumsi tidak terpenuhi
Linearity (Linearitas)	Koefisien a dan b berhubungan linear. Nilai mean dari variabel Y untuk suatu kombinasi $A = \pi r^2 X_1, X_2, X_3 \dots, X_K$ terletak pada garis/bidang linier yang dibentuk dari persamaan regresi	Untuk mengetahui asumsi linieritas dapat diketahui dari uji ANOVA (overall F test) bila hasilnya signifikan ($p \text{ value} < \alpha$) maka model berbentuk linier
Gauss of Residual (Normalitas)	Variabel Y mempunyai distribusi normal untuk setiap pengamatan variabel X.	dapat diketahui dari normal P-P Plot residual, bila data menyebar disekitar garis diagonal mengikuti arah garis diagonal maka model regresi memenuhi asumsi model regresi tidak memenuhi asumsi normalitas.

I. Kegunaan Analisis Regresi Ganda

Tujuan analisis regresi linier ganda adalah untuk menemukan model regresi yang paling sesuai menggambarkan faktor-faktor yang berhubungan dengan variabel dependen. Pada prinsipnya, model regresi ganda dapat berguna untuk dua hal :

a. Prediksi

Memperkirakan variabel dependen dengan menggunakan informasi yang ada pada sebuah atau beberapa variabel independen . Disini dapat diketahui secara probabilitas nilai variabel dependen bila seseorang/individu mempunyai suatu set variabel dengan independen tertentu. Misalnya kita melakukan analisis variabel independen umur, BB dan jenis kelamin dihubungkan dengan variabel dependen tekanan darah. Dari hasil regresi seseorang individu dapat diperkirakan tekanan darahnya pada umur, berat badan , dan jenis kelamin tertentu.

b. Estimasi

menguantifikasi hubungan sebuah atau beberapa variabel independen dengan sebuah variabel dependen. Pada fungsi ini regresi dapat digunakan untuk mengetahui variabel independen apa saja yang berhubungan dengan variabel dependen. Selain itu kita juga dapat mengetahui seberapa besar hubungan masing-masing independen terhadap variabel independen lainnya. Dari analisis ini dapat diketahui variabel mana yang paling besar/dominan mempengaruhi variabel dependen, yang ditunjukkan dari koefisien regresi (b) yang sudah distandardisasi yaitu nilai beta.

J. Pemodelan

Satu hal yang penting dalam regresi ganda adalah bagaimana memilih variabel independen sehingga terbentuk sebuah model yang paling sesuai menjelaskan/menggambarkan variabel dependen yang sesungguhnya dalam alam (populasi).

Dalam pembuatan model sering kali dijumpai pandangan yang kurang tepat yaitu “memasukkan semua/sebanyak mungkin variabel independen ke dalam model. Alasannya, dengan memasukkan sebanyak mungkin variabel independen ke dalam model, maka variabel dependen diharapkan diprediksi dengan sempurna. Perlu diketahui bahwa penambahan variabel independen tidak selalu meningkatkan kemampuan prediksi variabel independen terhadap variabel dependen, sebab semakin banyak variabel independen (lebih-lebih variabel yang tidak relevan) mengakibatkan

makin besarnya nilai standar error (Se). Disamping itu, model dengan banyak variabel seringkali malah menyulitkan dalam interpretasi.

Berdasarkan pertimbangan tersebut pemilihan variabel independen hendaknya dengan memperhatikan aspek statistik dan substansi. Model yang dihasilkan diharapkan model yang PARSIMONI, artinya variabel yang masih dalam model sebaiknya yang sedikit jumlahnya, namun cukup baik untuk menjelaskan faktor-faktor penting yang berhubungan dengan variabel dependen.

Banyak kriteria yang dapat digunakan untuk memilih variabel masuk dalam model, salah satu kriteria yang sering digunakan adalah melihat R^2 (*R Square*). Namun penggunaan kriteria ini perlu hati-hati, karena setiap penambahan satu variabel independen akan meningkatkan R^2 walaupun variabel tersebut tidak cukup penting. Oleh karena itu model yang digunakan adalah model dengan nilai R^2 yang besar namun variabel independennya dengan jumlah sedikit.

Berikut langkah-langkah dalam pemodelan regresi linier ganda (Hastono, 2007):

- 1) Melakukan analisis bivariat untuk menentukan variabel yang menjadi kandidat pemodelan multivariat. Masing-masing variabel independen dihubungkan dengan variabel dependen (bivariat), bila hasil uji bivariat mempunyai nilai $p < 0,25$, maka variabel tersebut dijadikan kandidat model multivariat. Untuk variabel yang p value-nya $> 0,25$ namun secara substansi penting, maka variabel tersebut dapat dijadikan kandidat untuk masuk ke pemodelan multivariat.
- 2) Lakukan analisis secara bersamaan semua variabel kandidat. Lakukan pemilihan variabel, dengan cara mengeluarkan variabel yang mempunyai p value $> 0,05$. Pengeluaran dari model satu persatu, dimulai dari variabel yang p valuenya paling besar. Bila variabel tersebut setelah dikeluarkan dari model mengakibatkan yang koefisien dari variabel yang masih dalam model berubah besar (merubah koefisien lebih dari 10 %) maka variabel tersebut tidak jadi dikeluarkan tapi dimasukkan kembali dalam model karena dianggap sebagai variabel *confounding*. Lakukan proses ini berulang-ulang sampai semua variabel yang p valuenya $> 0,05$. Dalam pemilihan/ penentuan variabel independen yang masuk model digunakan metode ENTER.

BAB XIII
REGRESI LOGISTIK GANDA

Sebelumnya telah diperkenalkan analisis regresi linier yang digunakan bila variabel tergantung atau responsnya berskala numerik atau kontinu. Namun bila kita mempunyai variabel respons yang bukan numerik melainkan dikotom (yaitumempunyai dua nilai atribut), maka analisis yang tepat digunakan bukanlah analisis regresi linier melainkan analisis regresi logistik.

Definisi Regresi Logistik

Hosmer & Lomeshow (1989), pernah menyatakan tentang regresi logistik seperti yang dikutip oleh Sounders & Trapp (1994), yaitu :

“Regresi logistik adalah sebuah model yang muncul dengan peningkatan frekuensi dalam literatur medis; hal ini secara umum digunakan pada variabel independen, baik pengukuran numerik dan nominal dan variabel outcome secara binary, atau dikotomi, memiliki hanya 2 kategori atau lebih.”

“ Logistic regression is a model appearing with increasing frequency in the medical literature ; it is commonly used when the independent variables include both numerical and nominal measures and the outcome variable is binary, or dichotomous, having only two values. Logistic regression can also be used when outcome had more than values.”

Klainbaum (1992) memberikan definisi regresi logistik seperti ini, yaitu :

“Pendekatan model matematika yang dapat digunakan untuk menggambarkan hubungan antara beberapa variabel menjadi variabel dependen yang dikotomi.”

“ Logistic regression is a mathematical modelling apporoach that can be used to describe the relationship of saveral X’s to a dichotomous dependent variable “

Fungsi Logistik

Fungsi logistik dapat digambarkan dalam bentuk matematis dibawah ini :

$$F(z) = \frac{1}{1 + e^{-z}}$$

Dari fungsi regresi logistik diperoleh model regresi logistik. Model regresi logistik dituliskan dalam bentuk penjumlahan linear :

$$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots \dots \beta_k X_k$$

Model ini digunakan untuk data longitudinal (Klainbaum,1992)

Di dalam konsep epidemiologi, terdapat pengertian perbandingan antara probabilitas sakit dan sehat, sebagai $p/(1-p)$ untuk kondisi variabel X tertentu (misal kondisi terpajan), yang disebut sebagai ODD. Untuk hal diatas, maka dapat menghitung ODD (dapat dicatat bahwa penulisan e (= bilangan natural euler) berpangkat suatu bilangan bisa ditulis dengan exp bilangan itu).

$$ODD = \frac{\exp(a+bX)}{(1-p)}$$

Dari persamaan diatas, maka dapat dilakukan transformasi, yang disebut sebagai transformasi logit atau logistik, yaitu dengan transformasi logaritma nilai e, sehingga persamaannya disebut sebagai regresi logit atau regresi logistik.

$$\text{Log-e ODD} = \ln ODD = \text{Logit}(Y=y) = a+bX$$

Bentuk persamaan ini lebih sederhana, dan dapat digunakan untuk membandingkan dua ODD, yang disebut sebagai ODDS RATIO atau di singkat OR.

Pada kondisi X terpajan (misal X=1), maka

$$\ln ODD_1 = a + (b \cdot 1) = a + b$$

$$ODD_1 = \exp(a + b)$$

Pada kondisi X tak terpajan (misal X =0), maka:

$$\ln ODD_0 = a + (b \cdot 0) = a$$

$$ODD_0 = \exp(a)$$

Odds Ratio dihitung dengan membandingkan dua ODD diatas :

$$\frac{ODD_1}{ODD_0} = \frac{\exp(a+b)}{\exp(a)} = \exp(b)$$

Jelaslah bahwa dari persamaan regresi logistik Logit ($Y=y) = a + bX$, hasil utama yang dapat dibaca adalah Odds Ratio.

Bila kembali kepada tabel silang 2x2, maka Odds Ratio dapat dihitung seperti berikut :

Risiko sakit= $a/a+b$; bila sakit jarang,

maka $=a/b=ODD_1$

Risiko sakit = $c/c+d$; bila sakit jarang, maka $=c/d=ODD_2$

$ODDS\ RATIO = ODD_1 / ODD_2 = (a/b) / (c/d) = ad / bc$

Ukuran Faktor Risiko Berdasarkan Desain Penelitian

Regresi Logistik:

Model matematis untuk menganalisis hubungan antara satu atau beberapa variabel independen yang bersifat kontinu maupun binary dengan satu variabel dependen yang bersifat binary/binomial atau dikotom (sehat/sakit, hidup/mati).

Regresi Logistik:

- a. Sederhana (simple) → hanya satu variabel independen
- b. Ganda (multiple) → lebih dari satu variabel independen

Kegunaan Regresi Logistik

A. Model Prediksi

1. Analisis ini bertujuan untuk memperoleh model yang terdiri dari beberapa variabel independen (prediktor) dari kumpulan besar variabel, yang dianggap terbaik untuk memprediksi kejadian variabel dependen (outcome).
2. Pada model ini semua variabel dianggap sama penting, sehingga dapat dilakukan estimasi beberapa koefisien regresi logistik sekaligus.

Prosedur pemodelan:

Agar diperoleh model regresi yang hemat dan mampu menjelaskan hubungan variabel independen dan dependen dalam populasi, diperlukan prosedur pemilihan variabel

1. Melakukan analisis bivariat antara masing-masing variabel independen dengan variabel dependennya. Bila hasil uji bivariat mempunyai nilai $p < 0,25$, maka variabel tersebut dapat masuk model multivariat. Namun bisa saja $p \text{ value} > 0,25$ tetap diikuti ke multivariat bila variabel tsb secara substansi penting
2. Memilih variabel yang dianggap penting yang masuk dalam model, dengan cara mempertahankan variabel yang mempunyai $p \text{ value} < 0,05$ dan mengeluarkan variabel yang $p \text{ valuenya} > 0,05$. Pengeluaran variabel tidak serentak semua yang $p \text{ valuenya} > 0,05$, namun dilakukan secara bertahap dimulai dari variabel yang mempunyai $p \text{ value}$ terbesar
3. Identifikasi linearitas variabel numerik dengan tujuan untuk menentukan apakah variabel numerik dijadikan variabel katagorik atau tetap variabel numerik.

Caranya :

dengan mengelompokkan variabel numerik ke dalam 4 kelompok berdasarkan nilai kuartilnya. Kemudian lakukan analisis logistik dan dihitung nilai OR-nya. Bila nilai OR masing-masing kelompok menunjukkan bentuk garis lurus, maka variabel numerik dapat dipertahankan. Namun bila hasilnya menunjukkan adanya patahan, maka dapat dipertimbangkan dirubah dalam bentuk katagorik.

4. Setelah memperoleh model yang memuat variabel-variabel penting, maka langkah terakhir adalah memeriksa kemungkinan interaksi variabel ke dalam model. Penentuan variabel interaksi sebaiknya melalui pertimbangan logika substantif. Pembuktian interaksi dilihat dari kemaknaan uji statistik. Bila variabel mempunyai nilai bermakna, maka variabel interaksi penting dimasukkan dalam model.

2. Model Faktor Resiko

Analisis ini bertujuan untuk mengestimasi secara valid hubungan satu variabel utama dengan variabel dependent dengan mengontrol beberapa variabel *confounding*. Pada model ini yang diutamakan adalah nilai koefisien regresi suatu determinan yang memang ingin dipelajari. Sedangkan variabel lain-kovariat dipertimbangkan sebagai variabel kontrol, karena variabel tersebut ikut berpengaruh (*confounding*) atau sebagai faktor modifikasi (interaksi) yang nilai pengaruhnya tak homogen dalam hubungan antara determinan dan outcome.

Kekuatan dan Kelemahan Analisis Regresi Logistik

1. Regresi logistik mempunyai batasan yang longgar dengan arti regresi logistik mentoleransi asumsi- asumsi yang harus dipenuhi pada regresi linear. Seperti, tidak diperlukan asumsi linearitas, normalitas, homoscedasticity, dll.
2. Pada analisis regresi, variabel dependennya berupa data katagorik yang bisa saja berasal dari data numerik, yang mengakibatkan terjadi perubahan nilai dari data numerik tersebut sehingga banyak kehilangan informasi dari data yang didapat.

Tahapan pemodelan:

1. Lakukan pemodelan lengkap, mencakup variabel utama, semua kandidat konfonding dan kandidat interaksi (interaksi dibuat antara variabel utama dengan semua variabel konfonding).
2. Lakukan penilaian interaksi, dengan cara mengeluarkan variabel interaksi yang nilai p Wald-nya tidak signifikan dikeluarkan dari model secara berurutan satu per satu dari nilai p Wald yang terbesar.
3. Lakukan penilaian konfonding, dengan cara mengeluarkan variabel kovariat/ konfonding satu per satu dimuali dari yang memiliki nilai p Wald terbesar, bila setelah dikeluarkan diperoleh selisih OR faktor/variabel utama antara sebelum dan sesudah variabel kovariat (X^1) dikeluarkan lebih besar dari 10%, maka variabel tersebut dinyatakan sebagai konfonding dan harus tetap berada dalam model.

Asumsi Yang Harus Diperhatikan dalam Analisis Regresi Logistik

1. Meaning coding
2. Inclusion of all relevant variable in the regression model
3. Exclusion of all irrelevant variabels
4. Low error in the explanatory variables.
5. Linearity
6. Additivity
7. No multicollinearity
8. No outliers
9. Large samples
10. Sampling adequacy
11. The dependent variable need not be normally distributed
12. The dependent variable need not be homoscedastic for each level of the independents; that is, there no homogeneity of variance assumption.
13. Expected dispersion

DAFTAR PUSTAKA

1. Hastono, Sutanto Priyo. Analisis Data Kesehatan. Fakultas Kesehatan Masyarakat Universitas Indonesia, Jakarta. 2007
2. Hastono, SP dan Sabri, L. *Statistik Kesehatan*. Rajawali Press, Jakarta, 2010
3. Budiarto, Eko *Biostatistik untuk Kedokteran dan Kesehatan Masyarakat*, .EGC, Jakarta, 2001
4. Usman, M; Warsono; Ruswandi, R. *Statistika, Pengantar Tehnik Analisa Data*, Sinar Baru Algensindo, Bandung, 2009