



PUSAT PENDIDIKAN DAN PELATIHAN
BADAN PUSAT STATISTIK

MODUL PELATIHAN FUNGSIONAL STATISTISI

PENGUMPULAN DAN PENGOLAHAN DATA



PUSDIKLAT BPS
2022

Hak Cipta © pada:
Pusat Pendidikan dan Pelatihan Badan Pusat Statistik
Edisi Tahun 2022

**Pusat Pendidikan dan Pelatihan Badan Pusat Statistik
Jl. Raya Jagakarsa NO. 70 Jakarta Selatan 12620**

Modul Pengumpulan dan Pengolahan Data

TIM PENGARAH SUBSTANSI:

1. Dr. Eni Lestariningsih, S.Si, MA
2. Tri Nugrahadi Ph.D

PENULIS MODUL:

1. Budiyanto
2. Eko Yulian
3. Muhammad Irkham
4. Erya Afrianus
5. Cepy Ramdhani

EDITOR:

1. Eko Yulian
2. Budiyanto

COVER: Else Haslijah

JAKARTA – PUSDIKLAT BPS – 2022
ISBN: nomor ISBN

KATA PENGANTAR



Puji syukur kehadiran Allah SWT yang telah memberikan petunjuk sehingga Modul 3 tentang Pengumpulan dan Pengolahan Data ini dapat disusun. Sub Modul Pengumpulan data memberikan penjelasan kepada para peserta pelatihan mengenai merancang kerangka sampel, menentukan pengambilan sampel, melakukan pelatihan petugas, melakukan pengumpulan data. Sedangkan Sub Modul pengolahan data memberikan penjelasan tentang arti dan pengertian data, statistik dan informasi, Kegunaan data dan statistik; serta pengolahan data itu sendiri. Pengumpulan data dan Pengolahan data merupakan bagian dari proses kegiatan statistik.

Modul ini merupakan salah satu bahan ajar yang digunakan dalam Pelatihan Fungsional Statistisi. Adapun modul yang digunakan terdiri dari:

1. Modul 1: Penguatan Sistem Statistik Nasional
2. Modul 2: GSBPM (Umum, Specify Needs, dan Design & Build)
3. Modul 3: Pengumpulan dan Pengolahan Data
4. Modul 4: Analisis dan Diseminasi Data

Ucapan terima kasih dan apresiasi kami sampaikan kepada seluruh pihak yang telah membantu dan memberikan masukan dalam penyusunan modul ini. Tanggapan dan saran yang konstruktif kami harapkan guna perbaikan dan pengembangan di masa mendatang. Semoga modul ini dapat bermanfaat bagi pengembangan kompetensi bidang statistik bagi para peserta pelatihan.

Jakarta, Maret 2022
Kepala Pusdiklat BPS

Eni Lestariningsih, S.Si, M.A.
NIP. 197003101994012001

Sub Modul Pengumpulan Data

DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI.....	ii
BAB I.....	1
PENDAHULUAN	1
A. Latar Belakang.....	1
B. Deskripsi Singkat	2
C. Tujuan Pembelajaran	2
D. Materi Pokok dan Sub Materi Pokok	2
BAB II.....	4
MERANCANG KERANGKA SAMPEL.....	4
A. Kerangka Sampel	4
B. Merancang Jumlah Sampel	4
C. Alokasi Sampel Unit Analisis.....	7
BAB III.....	9
MENENTUKAN PENGAMBILAN SAMPEL.....	9
A. Probability Sampling.....	9
B. Non-Probability Sampling	11
BAB IV.....	13
MELAKUKAN PELATIHAN PETUGAS	13
BAB V.....	14
MELAKUKAN PENGUMPULAN DATA	14
A. Metode Pengumpulan Data	14
B. Pelaksanaan Pengumpulan Data.....	16

BAB I

PENDAHULUAN

A. Latar Belakang

Pengumpulan data merupakan aspek fundamental dalam penyelenggaraan kegiatan statistik. Sesuai dengan UU No.16 Tahun 1997, cara pengumpulan data dalam kegiatan statistik adalah sensus, survei, kompilasi produk administrasi, dan cara lain sesuai dengan perkembangan ilmu pengetahuan dan teknologi.

Sensus dimaksudkan sebagai suatu cara pengumpulan data yang dilakukan melalui pencacahan semua unit populasi untuk memperoleh karakteristik suatu populasi pada saat tertentu. Sedangkan Survei adalah cara pengumpulan data yang dilakukan melalui pencacahan sampel untuk memperkirakan karakteristik suatu populasi pada saat tertentu. Kompilasi produk administrasi adalah cara pengumpulan data, pengolahan, penyajian, dan analisis data yang didasarkan pada catatan administrasi yang ada pada pemerintah atau masyarakat.

Di dalam melaksanakan pengumpulan data, maka metode yang dapat digunakan dalam pengumpulan data adalah wawancara, swacacah (responden mengisi kuesioner sendiri), dan pengamatan (observasi). Pemilihan metode tersebut menyesuaikan kebutuhan dan kemampuan penyelenggara kegiatan statistik.

Tidak termasuk dalam kegiatan pengumpulan data adalah Big Data. Karena Big Data, merupakan fenomena pengumpulan data dengan volume yang sangat besar dikumpulkan karena adanya kemajuan teknologi dan informasi terutama pada era industri 4.0 ini. Seperti diketahui, bahwa dalam Big Data, baik data yang terstruktur seperti data penjualan e-commerce (TokoPedia, Buka lapak dls) atau data yang tidak terstruktur seperti teks, audio, video, dan lain lain tidak lagi mengikuti kaedah kaedah statistik seperti yang dilakukan dalam kegiatan statistik. Namun demikian, kemajuan teknologi tersebut masih dapat digunakan dalam kegiatan pengumpulan data yaitu pengumpulan data dengan memanfaatkan seperti CAWI, CAPI, CATI, dll.

Sesuai dengan tahapan pelaksanaan kegiatan statistik dalam kerangka General Statistics Business Process Model (GSBPM), maka pengumpulan data dalam bab ini dapat dibagi dalam

empat tahap: merancang kerangka sampel, melakukan pengambilan sampel, melakukan pelatihan petugas dan melakukan pengumpulan data.

B. Deskripsi Singkat

Dalam mata diklat pengumpulan data, pembahasan berfokus pada kemampuan merancang kerangka sampel, menentukan besarnya sampel, pelaksanaan pelatihan petugas, dan pengumpulan data. Peserta diklat akan dibekali metode penarikan sampel dan pengumpulan data. Mata diklat ini disajikan dengan banyak latihan yang terencana sehingga peserta diklat memiliki kemampuan merancang kerangka sampel dan pengumpulan data.

C. Tujuan Pembelajaran

Setelah mengikuti mata diklat pengumpulan data, peserta diharapkan mampu:

1. Menjelaskan perancangan kerangka sample
2. Mampu menentukan besarnya sampel
3. Menjelaskan pengumpulan data
4. Mampu mengumpulkan data

D. Materi Pokok dan Sub Materi Pokok

- I. Pendahuluan
 - A. Latar Belakang
 - B. Deskripsi Singkat
 - C. Tujuan Pembelajaran
 - D. Materi Pokok dn Sub Materi Pokok
- II. Merancang Kerangka Sampel
 - A. Kerangka Sampel
 - B. Merancang Jumlah Sampel
- III. Menentukan Pengambilan Sampel
 - A. Probaility Sampling
 - B. Non Probability Sampling
- IV. Melakukan Pelatihatam petugas

- V. Melakukan Pengumpulan Data
 - A. Metode pengambilan Sampel;
 - B. Pelaksanaan Pengambilan Sampel

BAB II

MERANCANG KERANGKA SAMPEL

A. Kerangka Sampel

Kerangka sampel adalah kumpulan unit dalam populasi yang menjadi dasar pemilihan sampel. Populasi adalah sekumpulan unsur atau elemen atau unit analisis yang menjadi objek penelitian, seperti kumpulan semua kota, semua rumah tangga, semua perusahaan, dan sebagainya. Sementara itu, populasi target merupakan populasi yang ingin disimpulkan dan ditentukan sesuai dengan masalah penelitian. Populasi survei adalah populasi yang terliput dalam penelitian yang dilakukan.

Kerangka sampel yang ideal adalah lengkap (mencakup seluruh populasi), akurat (sesuai kondisi sebenarnya), dan terkini (*up to date*). Kerangka sampel yang tidak memenuhi syarat akan berdampak pada hasil survei yang bias (sering kali berupa *underestimate* populasi target). Dalam suatu rancangan survei, dimungkinkan menggunakan beberapa jenis kerangka sampel sesuai desain sampel yang diterapkan. Sebagai contoh, pada survei yang menerapkan desain sampel dua tahap, kerangka sampel yang digunakan adalah kerangka sampel area (*area frame*) dan kerangka sampel daftar unit analisis (*list frame*).

B. Merancang Jumlah Sampel

Sampel adalah unsur-unsur yang diambil dari populasi. Penentuan jumlah sampel bertujuan memperoleh jumlah sampel yang cukup untuk penyajian estimasi karakteristik yang merepresentasikan populasi pada suatu tingkat wilayah tertentu. Lebih lanjut, jumlah sampel dapat digunakan untuk penentuan volume kegiatan survei, seperti jumlah petugas, jumlah dokumen, anggaran yang diperlukan, dan sebagainya. Adapun keuntungan penggunaan sampel dalam suatu kegiatan statistik adalah dapat memberikan gambaran tentang populasi, dapat menentukan presisi, sederhana sehingga relatif mudah dilaksanakan, dan dapat memberikan keterangan sebanyak mungkin.

Ukuran sampel tergantung pada derajat keseragaman karakteristik unit-unitnya, presisi yang dikehendaki, rencana analisis data, dan sumber daya yang tersedia (Singarimbun dan Effendi, 1982). Semakin besar sampel, semakin tinggi tingkat presisi yang didapatkan. Bila unsur populasi benar-benar seragam, jumlah sampel sedikit saja cukup untuk mewakili populasi.

Beberapa hal yang harus diperhatikan dalam penentuan ukuran (jumlah) sampel adalah:

1. Indikator dan Variabel Survei

Dalam setiap kegiatan survei, setiap estimasi yang akan dihasilkan dari survei tersebut memerlukan ukuran sampel yang berbeda agar menghasilkan pengukuran yang *reliabel*. Namun, untuk rancangan survei, biasanya hanya menggunakan satu ukuran sampel. Oleh karena itu, penghitungan jumlah sampel minimum suatu survei harus berdasarkan salah satu indikator kunci yang akan diukur dalam survei. Contohnya, jika indikator kunci adalah tingkat pengangguran, maka penghitungan jumlah sampel akan didasarkan pada indikator kunci tersebut. Jika terdapat banyak indikator kunci, konversi yang dapat diterapkan adalah menghitung jumlah sampel yang dibutuhkan bagi masing-masing indikator kunci kemudian menggunakan indikator kunci yang menghasilkan sampel paling besar.

2. Populasi Target

Penghitungan ukuran sampel harus memperhitungkan setiap populasi target. Penentuan ukuran sampel dapat difokuskan pada populasi target yang paling kecil. Semakin sedikit kasus pada suatu variabel yang dijadikan dasar penentuan jumlah sampel, akan semakin besar ukuran sampel yang diperlukan. Contohnya, jika anak-anak berusia di bawah 5 tahun merupakan kelompok target dalam survei, maka ukuran sampel harus didasarkan pada kelompok tersebut. Perancang survei dapat menentukan ukuran sampel sebesar 10% dari anak berusia di bawah 5 tahun.

3. Presisi dan Tingkat Keyakinan

Estimasi indikator-indikator kunci harus reliabel. Oleh karena itu, penentuan ukuran sampel juga sangat bergantung pada derajat presisi yang diharapkan dari indikator. Semakin tepat dan reliabel estimasi survei, maka semakin besar pula ukuran sampelnya. Dengan mempertimbangkan indikator kunci, nilai yang digunakan dalam penentuan jumlah sampel minimum adalah relatif *margin of error* (MoE) sebesar 5-10 persen pada tingkat keyakinan sebesar 95 persen. Namun, biasanya akan memerlukan anggaran yang lebih besar disbanding toleransi error yang lebih besar. *Relatif* MoE sebesar 20% secara umum dianggap sebagai yang paling maksimum yang diperbolehkan untuk indikator-indikator penting. Semakin besar toleransi *error* yang ditetapkan, hasil survei tidak dapat bermanfaat banyak bagi keperluan analisis atau pengambilan kebijakan.

4. Grup/Domain Analisis

Domain analisis secara umum didefinisikan sebagai subkelompok analisis yang menjadi tujuan dihasilkannya data yang reliabel. Semakin banyak domain estimasi yang ingin dilakukan secara reliabel, maka semakin tinggi jumlah sampel yang dibutuhkan. Jumlah sampel yang dibutuhkan untuk melakukan estimasi yang reliabel pada tingkat nasional saja, tentunya akan lebih sedikit dibandingkan dengan jumlah sampel yang dibutuhkan agar hasil estimasi juga reliabel pada domain provinsi/kabupaten.

5. Anggaran Survei

Komponen anggaran bukan merupakan parameter penting dalam penghitungan matematis ukuran sampel. Namun pada praktiknya, anggaran memainkan peranan yang sangat penting pada kegiatan survei dan terkait erat dengan total jumlah sampel yang diperlukan.

Pada bagian ini, disediakan rumus untuk menghitung ukuran sampel dengan mempertimbangkan parameter-parameter yang telah dibahas sebelumnya. Pada contoh ini, unit analisis sebagai unit sampling utama (rancangan pemilihan sampel satu tahap):

- 1) Rumus penghitungan jumlah sampel minimum berdasarkan grup/domain analisis:

$$m = \frac{Z^2 p(1-p)}{e^2} \cdot \frac{1}{r}$$

Penyesuaian ukuran sampel minimum tersebut karena populasi terbatas (*finite*

population) dilakukan dengan rumus:
$$m = \frac{m_0}{1 + \frac{(m_0 - 1)}{M}},$$

dengan:

Z^2 : menunjukkan tingkat presisi yang diharapkan. Nilai Z^2 adalah absis kurva normal yang memotong area pada ujung (ekor) kurva ($1-\alpha$ sama dengan selang kepercayaan yang diinginkan).

p : proporsi populasi *eligible* terhadap populasi target.

Misal: proporsi orang yang puas dengan pelayanan kesehatan terhadap jumlah orang yang menerima layanan kesehatan.

r : toleransi tingkat respons minimum

Misal: ditentukan 90 persen.

m : jumlah unit sampel yang harus diambil.

m_0 : hasil penghitungan ukuran sampel minimum awal.

- 2) Simplifikasi penghitungan ukuran sampel yang didasarkan pada pendugaan proporsi populasi dapat dilakukan dengan cara (Yamane, 1967) di bawah ini. Rumus ini digunakan apabila tidak ada informasi awal terkait kondisi populasi.

$$m = \frac{M}{M(e)^2 + 1} .$$

dengan:

e : margin *of error* yang ditoleransi (ditentukan 10 atau 20 persen),

M : populasi unit observasi yang diteliti.

- 3) Pada survei penduduk yang dilakukan dengan pendekatan rumah tangga, ukuran sampel yang dihitung adalah jumlah rumah tangga yang harus dipilih. Variabel yang digunakan adalah variabel dengan karakteristik individu (anggota rumah tangga). Secara umum, ketika suatu proporsi u disertakan, rumus perkiraan ukuran sampel n adalah:

$$m = \frac{z^2 u(1-u) \times deff}{pne^2} \times \frac{1}{r} ,$$

dengan:

u : Estimasi indikator kunci yang akan diukur oleh survei;

\bar{n} : Rata-rata ukuran rumah tangga (jumlah ART per rumah tangga)

Pada rancangan survei yang menerapkan pemilihan sampel dua tahap (misal memilih area terlebih dahulu, kemudian memilih unit analisis), diperlukan penentuan jumlah sampel untuk pemilihan sampel tahap pertama. Ini dapat dilakukan dengan beberapa cara. Salah satu contohnya, bila ditentukan *fixed size* dalam area terpilih sebanyak sepuluh unit analisis, maka setelah diperoleh jumlah sampel unit analisis dapat dihitung jumlah sampel area sebanyak

$$n = \frac{m}{10} .$$

C. Alokasi Sampel Unit Analisis

Alokasi sampel merupakan proses lanjutan setelah diperoleh jumlah sampel unit analisis untuk keperluan estimasi menurut domain tertentu. Alokasi sampel diperlukan agar sampel dapat terdistribusi secara proporsional di setiap subdomain yang ditentukan. Rumus ini menggunakan proporsi variabel tertentu pada suatu domain terhadap agregat variabel

tersebut pada domain yang melingkupinya (tingkat atasnya). Rumus *proportional allocation* ditulis sebagai berikut:

$$n_h = \frac{M_h}{\sum_{h=1}^H M_h} \times n,$$

dengan:

n_h	:	jumlah unit sampel dalam subdomain (kelompok) h ,
M_h	:	jumlah unit observasi/agregat variabel pokok dalam kelompok h ,
n	:	Jumlah (target) unit analisis (sampel) untuk estimasi tingkat domain tertentu.

Rumus lain yang menerapkan proporsi populasi adalah penggunaan proporsi akar kuadrat ukuran populasi, yaitu dengan cara memberi pangkat 0,5 untuk angka populasi (M_h) pada rumus di atas menjadi $M_h^{0,5}$ atau $\sqrt{M_h}$. Hasil alokasi sampel ntuk setiap subdomain lebih

merata.

Beberapa contoh kasus untuk penerapan alokasi sampel adalah:

- Jumlah sampel nelayan untuk estimasi indeks kesejahteraan nelayan tingkat kecamatan diperoleh 80 orang nelayan. Agar sampel sebanyak 80 orang nelayan terdistribusi secara proporsional di setiap desa nelayan, maka perlu dilakukan alokasi sampel nelayan untuk setiap desa nelayan secara proporsional. Dasar hitung proporsi yang digunakan adalah banyaknya nelayan masing-masing desa nelayan terhadap total nelayan pada seluruh desa nelayan di kabupaten.
- Jumlah sampel rumah tangga untuk estimasi kepuasan layanan kesehatan tingkat kecamatan diperoleh 100 orang penerima layanan. Survei dilakukan di semua puskesmas dalam kecamatan. Agar sampel sebanyak 100 orang tersebut merata ke setiap puskesmas desa, dilakukan alokasi sampel untuk setiap puskesmas desa dalam kecamatan. Alokasi ini dapat dilakukan berdasarkan proporsi jumlah penerima layanan kesehatan selama tiga bulan yang lalu.

BAB III

MENENTUKAN PENGAMBILAN SAMPEL

Metode pengambilan sampel adalah teknik memilih sebagian unit populasi yang kemudian digunakan untuk melakukan generalisasi (estimasi) populasi tempat sampel tersebut diambil. Penetapan metode pengambilan sampel tidak terlepas dari ketersediaan kerangka sampel, operasional di lapangan, anggaran yang tersedia, serta toleransi *sampling error* yang terjadi. Metode pengambilan sampel terdiri dari dua jenis, yaitu *probability sampling* dan *non-probability sampling*.

A. Probability Sampling

Probability sampling adalah metode pengambilan sampel yang menerapkan bahwa setiap unit dalam populasi memiliki peluang (lebih dari nol) untuk terpilih dalam sampel dan peluang ini dapat ditentukan secara akurat. Berdasarkan rancangan ini, hasil survei dengan jumlah sampel yang cukup mewakili populasinya dapat digunakan untuk estimasi populasi.

Yang termasuk dalam *probability sampling* adalah:

1) Simple Random Sampling (SRS)

Pada rancangan ini, setiap unit sampel dalam kerangka sampel memiliki peluang terpilih yang sama. Penentuan unit terpilih menggunakan angka acak (*random*) yang dapat diperoleh dari tabel angka random atau dihasilkan dari *random number generator* (RNG) pada komputer. Pemilihan sampel dengan angka random yang berbeda akan menghasilkan gugus sampel yang memuat unit-unit sampel yang berbeda. Teknik ini menghasilkan banyaknya kemungkinan sampel (*all possible sample*) yang paling banyak. SRS dapat diterapkan jika populasi yang akan diteliti kecil, homogen, dan kerangka sampel sampai unit sampling terkecil sudah tersedia.

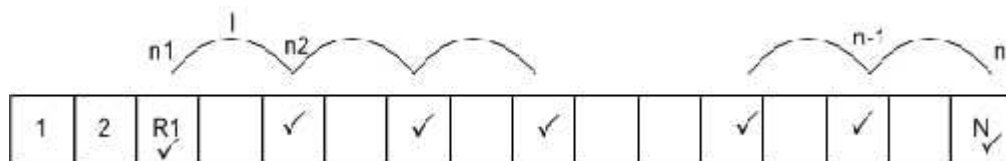
Contoh:

Suatu lembaga mengadakan survei terhadap pegawai yang ada di kantornya tentang pemahaman pegawai terhadap reformasi birokrasi dan *progress* yang sudah dilakukan oleh lembaga tersebut. Kerangka sampel berupa daftar seluruh pegawai yang lengkap dan *up to date* dimiliki oleh kantor tersebut. Dari daftar tersebut, dipilih sejumlah pegawai secara acak (menggunakan angka random) untuk diwawancarai. Dengan demikian, pemilihan sampel ini dinamakan *simple random sampling* karena setiap pegawai memiliki peluang terpilih yang sama.

2) Systematic Sampling

Systematic sampling merupakan metode pemilihan sampel yang menggunakan angka random pada pemilihan unit pertama dan menggunakan pola yang teratur untuk pemilihan sejumlah unit sampel berikutnya. Teknik ini memerlukan pengurutan populasi target berdasarkan kriteria tertentu. Sementara itu, pola tertentu dibangun dengan interval yang dihitung dari perbandingan antara populasi terhadap jumlah sampel yang diperlukan. Seperti halnya pada SRS, semua unit dalam populasi memiliki peluang yang sama untuk terpilih.

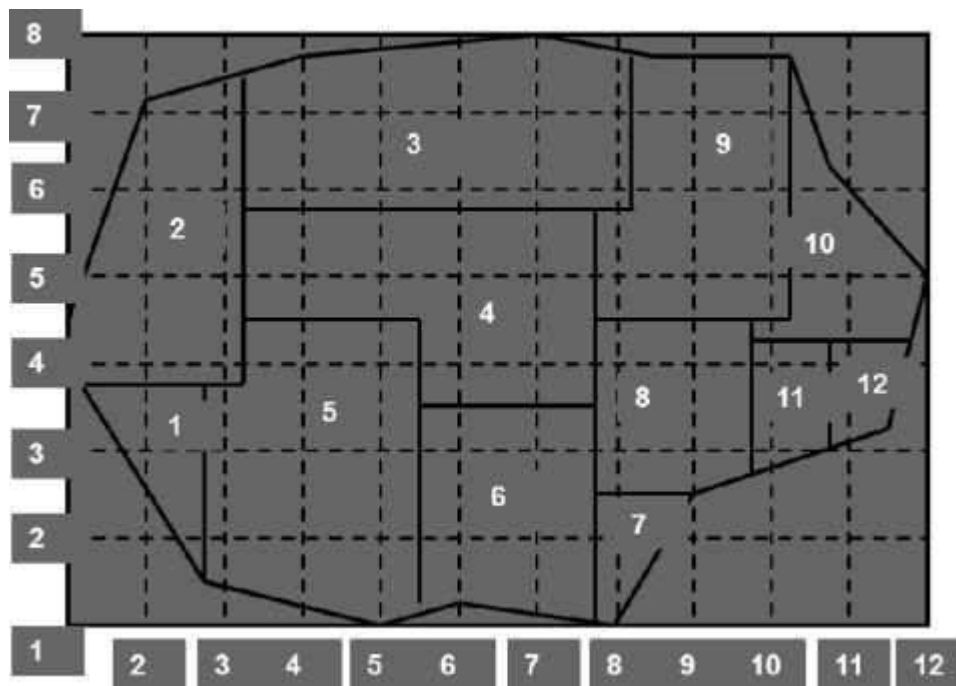
Keuntungan penggunaan rancangan ini adalah pemilihan sampel lebih mudah, identifikasi kerangka sampel yang cocok lebih mudah, serta sampel dapat tersebar secara merata. Sementara itu, kekurangan rancangan ini adalah sampel bisa menjadi bias jika terdapat “periodisasi” dalam populasi, sulit mendapatkan penduga presisi dari satu survei. Berikut ini ilustrasi pengambilan sampel secara *systematic sampling*.



Gambar 1. Contoh pengambilan sampel dengan *systematic sampling*

3) Probability Proportional to Size (PPS)

PPS merupakan teknik sampling yang menggunakan variabel tambahan yang berkorelasi kuat dengan variabel yang sedang diteliti untuk meningkatkan akurasi pemilihan sampel. Variabel tambahan menentukan besarnya peluang terpilih suatu unit dalam populasi. Kelemahan metode ini adalah dimungkinkan ada subgroup dalam populasi yang kurang terwakili atau terlalu banyak sampel. Namun demikian, kombinasi PPS dengan *systematic sampling* dapat menjadi solusi untuk mengatasi kelemahan tersebut.



Gambar 2. Ilustrasi PPS Sampling

Contoh :

Gambar 2 menunjukkan ilustrasi pemilihan petak sawah yang memiliki luas yang berbeda satu dengan yang lain dalam suatu bidang. Pemilihan petak menggunakan angka random yang menunjukkan koordinat dalam bidang tersebut. Misal terpilih koordinat (7;3), maka petak yang terpilih sebagai sampel adalah petak nomor 6. Dengan cara pemilihan sampel seperti ini, peluang terpilihnya petak-petak tersebut akan sebanding dengan luas petak tersebut. Petak yang lebih luas memiliki peluang terpilih yang lebih besar dibanding petak yang kecil. Dalam hal ini, luas petak digunakan sebagai *Measure of Size* (MOS) dalam PPS.

B. Non-Probability Sampling

Non-probability sampling merupakan metode pengambilan sampel yang menerapkan adanya elemen atau unit dalam populasi yang tidak memiliki peluang terpilih atau peluang terpilihnya tidak dapat ditentukan dengan akurat. Teknik ini memerlukan asumsi tertentu dalam penerapannya. Dengan metode ini, pemilihan sampel tidak acak (*random*), sehingga menyebabkan peluang pemilihan sampel pada rancangan ini tidak dapat dihitung. Lebih lanjut, penimbang sampling yang dibangun dari peluang sampel terpilih pun tidak dapat dihitung. Hal ini menyebabkan estimasi dan estimasi kesalahan penggunaan sampel (*sampling error*) tidak dapat dihitung pula.

Yang termasuk dalam *non-probability sampling* adalah:

- 1) *Accidental Sampling, Convenience Sampling, Haphazard Sampling, atau Opportunity Sampling.*

Teknik ini merupakan pengambilan sampel dari populasi yang terdekat atau yang bisa diwawancarai dan memudahkan tahapan wawancara. Sampel yang diambil dengan metode ini tidak dapat dilakukan generalisasi terhadap populasi, karena sampel terpilih tidak representatif terhadap populasi. Metode sampling seperti ini paling bermanfaat untuk *pilot survey* (uji coba).

- 2) *Quota Sampling*

Teknik ini merupakan pengambilan sampel dengan menentukan target sampel di setiap subgrup terlebih dahulu. Penentuan target sampel dan pengambilan sampel dilakukan dengan pertimbangan subjektif dan tidak acak. Contohnya *interviewer* mungkin akan lebih memilih responden yang terlihat lebih kooperatif. Oleh karena itu, sampel ini bisa jadi akan bias, karena tidak semua unit dalam populasi berkesempatan untuk dipilih sebagai sampel.

- 3) *Purposive Sampling*

Teknik ini merupakan pengambilan sampel berdasarkan kecenderungan atau subjektif peneliti. Teknik ini biasanya dilakukan ketika hanya terdapat sedikit orang yang ahli dalam bidang yang sedang diteliti.

BAB IV

MELAKUKAN PELATIHAN PETUGAS

Pelatihan petugas bertujuan untuk mempersiapkan petugas yang andal dalam melakukan pendataan sesuai dengan prosedur serta konsep dan definisi yang telah ditetapkan sehingga didapatkan data survei yang akurat. Petugas dilatih oleh instruktur yang sebelumnya sudah mendapatkan pelatihan. Rekrutmen petugas dilakukan secara proporsional dengan melihat beban tugas di masing-masing wilayahnya. Para petugas dapat diperoleh dari berbagai kalangan, yaitu petani, ibu rumah tangga, guru, pegawai kelurahan atau kecamatan, mahasiswa dan lain-lain.

Kegiatan pelatihan ini dapat dilaksanakan pada tempat pelatihan yang sudah ditentukan sebelumnya. Agar pelatihan para petugas lapangan dapat maksimal, beberapa kriteria tempat latihan perlu diperhatikan seperti fasilitas kelas yang cukup, ruang yang tidak bising, serta tempat istirahat yang dapat menampung petugas pengumpul data. Selanjutnya pada pelaksanaan pelatihan, selain kegiatan pemberian materi juga dilakukan kegiatan *role playing* (praktik wawancara). Kemudian, dalam rangka mengetahui penguasaan peserta terhadap materi yang diberikan, maka dilakukan tes materi atau pendalaman materi.

BAB V

MELAKUKAN PENGUMPULAN DATA

A. Metode Pengumpulan Data

Pengumpulan data memegang peranan penting dalam penyelenggaraan kegiatan statistik. Pengumpulan data yang dilakukan harus memperhatikan cara dan metode pengumpulan data. Cara pengumpulan data terdiri dari sensus, survei, dan kompilasi produk administrasi. Sensus bersumber dari data sekunder. Contoh: Registrasi Penduduk. adalah cara pengumpulan data yang dilakukan melalui pencacahan semua unit populasi untuk memperoleh karakteristik suatu populasi pada saat tertentu. Sensus bersumber dari data primer. Contoh: Sensus Penduduk 2010, Sensus Pertanian 2013, dan Sensus Ekonomi 2016.

Survei adalah cara pengumpulan data yang dilakukan melalui pencacahan sampel untuk memperkirakan karakteristik suatu populasi pada saat tertentu. Survei bersumber dari data primer. Contoh: Survei Sosial Ekonomi Nasional (Susenas) 2017. Kompilasi produk administrasi adalah cara pengumpulan data, pengolahan, penyajian, dan analisis data yang didasarkan pada catatan administrasi yang ada pada pemerintah dan/atau masyarakat. Kompilasi produk administrasi

Seperti disinggung sebelumnya, metode pengumpulan dapat dilakukan antara lain: Wawancara (*interview*), *Self Assessment* (diisi sendiri oleh responden), Pengamatan (*observation*), Pengukuran (*measurement*)

1) Wawancara

Pencacah menghubungi responden (unit penelitian) dan mengajukan pertanyaan secara langsung atau dengan menggunakan daftar pertanyaan yang telah disiapkan (terstruktur). Pilihan wawancara dapat dilaksanakan melalui telepon, wawancara perorangan secara pribadi atau kelompok.

Keunggulan dalam menggunakan wawancara adalah :

- a. Kembalinya daftar pertanyaan (kalau menggunakan daftar pertanyaan) lebih banyak, karena pencacah mengisi langsung. Tidak adanya jawaban dari responden (non response) dapat ditekan sekecil mungkin.

- b. Data lebih terjamin kebenarannya, karena pencacah mendapatkan data langsung dari responden
- c. Kelengkapan jawaban terhadap semua pertanyaan dalam kuesioner dapat dijamin.
- d. Pencacah dapat lebih memperjelas pertanyaannya, kalau responden masih ragu akan maksud pertanyaan.
- e. Pencacah dapat bertanya dengan baik.
- f. Pencacah dapat menggunakan bahasa yang dapat dimengerti oleh responden.

Kelemahan dalam menggunakan wawancara adalah, yaitu :

- a. Sifat pribadi pencacah dapat mempengaruhi jawaban
- b. Responden seringkali memberikan jawaban yang bersifat pribadi
- c. Responden kadangkala sulit dijumpai
- d. Biaya juga akan bertambah karena harus melatih pencacah
- e. Perlu organisasi lapangan yang baik

2) *Self Assessment* (diisi sendiri oleh responden)

Dalam hal ini, daftar pertanyaan dikirim kepada responden, kemudian setelah diisi oleh responden, daftar pertanyaan dikirim kembali (mailing system).

Keunggulan dari cara ini :

- a. Biaya lebih kecil dan ruang lingkup daerah bisa lebih luas
- b. Tidak memerlukan petugas pencacah, sehingga tidak ada biaya untuk latihan pencacah
- c. Responden dapat memberikan jawaban secara terbuka.

Kelemahannya yaitu:

- a. Banyak daftar isian yang tidak dikembalikan oleh responden dan juga banyak daftar yang tidak sampai kepada responden, non response menjadi sangat besar.
- b. Banyak kesalahan pengisian daftar, karena responden tidak menangkap maksud pertanyaan yang sebenarnya.
- c. Pertanyaan yang diajukan mungkin tidak dapat dijawab atau sulit dijawab secara tertulis.

3) Observasi (Pengamatan)

Cara ini dilakukan tanpa mengajukan pertanyaan-pertanyaan, sering dipergunakan dalam penelitian di dalam bidang sosial dan ekonomi serta terutama sekali dapat digunakan untuk penelitian di mana obyeknya tidak bisa diajak wawancara (komunikasi secara verbal). Misalnya pengamatan untuk mengetahui perilaku bayi, penelitian upacara adat, melihat kesibukan para karyawan yang bekerja di suatu kantor atau program TVRI apa yang digemari pirsawan dan sebagainya. Dalam pelaksanaan pengumpulan datanya dapat dilakukan oleh orang atau oleh suatu alat secara mekanis.

Kebaikan dalam cara pengamatan ini adalah, bahwa pencatatan bisa dilakukan pada waktu kejadian-kejadian itu terjadi, tanpa tergantung pada jawaban dari obyek yang diselidiki. Sedangkan kelemahannya di antaranya adalah tidak bisa dipergunakan untuk memperoleh data tentang sikap dan motivasi responden, apa yang diperbuat responden pada waktu lalu dan rencana yang akan datang.

4) Pengukuran

Dalam hal ini pencacah langsung mengadakan pengukuran terhadap obyek yang akan diteliti.

Keuntungan dari cara ini adalah:

- a. Cara ini paling teliti, karena obyek dikuasai pencacah.
- b. Data hasil pengukuran dapat diisi seutuhnya dalam daftar pertanyaan.
- c. Data tidak dipengaruhi pendapat pencacah.

B. Pelaksanaan Pengumpulan Data

Apabila cara-cara pengumpulan data di atas kita kelompokkan, maka ada dua kelompok cara pengumpulan data, yaitu:

1. Pelaksanaan pengumpulan secara langsung.

Dalam hal ini pencacah langsung mengumpulkan data di lapangan, seperti pengumpulan data yang dilakukan dengan wawancara, pengamatan dan pengukuran.

2. Pelaksanaan pengumpulan yang dilakukan secara tidak langsung. Dalam hal ini pencacah tidak langsung mengumpulkan data di lapangan, tapi hanya dengan mengirimkan daftar pertanyaan pada responden.

Dalam bagian ini pembahasan dibatasi hanya pada pelaksanaan secara langsung. Sedangkan pengumpulan yang dilakukan secara tidak langsung tidak dibicarakan karena dalam hal ini tidak diperlukan tenaga lapangan, di samping pelaksanaannya sudah langsung dari pusat pengumpulan data.

Pengumpulan data yang dilakukan secara langsung di lapangan, diperlukan petugas pencacah. Petugas pencacah ini dilatih terlebih dahulu guna menjelaskan maksud tujuan penelitian, konsep dan definisi, cara pengisian daftar pertanyaan, tata cara mencacah dan lain-lain. Petugas pencacah selain dilatih, juga dilengkapi dengan buku pedoman yang memuat segala sesuatu mengenai petunjuk pelaksanaan pengumpulan data. Buku pedoman ini berguna untuk membimbing petugas apabila telah sampai di lapangan, memecahkan berbagai kesulitan yang mungkin timbul. Petugas pencacah juga haruslah orang yang tajam pikirannya, kreatif, teliti, bebas dari kecenderungan pribadi dan prasangka, bijaksana, ramah, menyenangkan, sehat, waspada dan lincah, di samping memiliki kemampuan menulis yang rapi dan dapat dibaca.

Sub Modul Pengolahan Data

DAFTAR ISI

DAFTAR ISI	1
BAB I PENDAHULUAN.....	3
A. Deskripsi Singkat.....	3
B. Hasil Belajar (Tujuan Pembelajaran).....	3
C. Indikator Hasil Belajar	3
D. Materi Pokok.....	3
E. Manfaat	4
BAB II . DATA, STATISTIK, DAN INFORMASI	5
A. Arti dan Pengertian Data	5
B. Arti dan Pengertian Statistik	6
C. Arti dan Pengertian Informasi	7
D. Kegunaan Data dan Statistik	7
BAB III. PROSES PENGOLAHAN DATA	18
A. Melakukan Integrasi Data	18
B. Melakukan klasifikasi	18
C. Melakukan Pemeriksaan, Pengkodean, dan Penginputan	19
D. Pengolahan Data Dengan Ms. Excel.....	21

BAB I PENDAHULUAN

A. Deskripsi Singkat

Modul pengolahan data ini merupakan pengantar untuk lebih memahami proses pengolahan dari data mentah menjadi informasi dengan cakupan materi pembelajaran yaitu: arti dan konsep pengolahan data, pengolahan data secara manual, dan pengolahan data dengan komputer. Materi praktikum pengolahan membahas tentang pengolahan data statistik dengan menggunakan software Ms. Excel yang bertujuan untuk mempermudah interpretasi suatu informasi yang dijelaskan dengan perhitungan statistik.

B. Hasil Belajar (Tujuan Pembelajaran)

Setelah membaca modul ini peserta pelatihan mampu melakukan proses pengolahan data, dengan menggunakan paket program Excel.

C. Indikator Hasil Belajar

Setelah mempelajari modul ini secara tuntas, peserta diharapkan dapat:

- a. Menjelaskan pengertian data dan informasi
- b. Memahami proses pengolahan data
- c. Melakukan pengolahan data dengan menggunakan Ms. Excel

D. Materi Pokok

Materi dan submateri pokok yang dibahas dalam modul ini adalah:

1. Pengertian data, statistik, dan informasi
 - a. Pengertian data
 - b. Pengertian statistik
 - c. Pengertian informasi
 - d. Kegunaan data dan Statistik
 - e. Data statistik yang baik
 - f. Pengertian variabel
 - g. Unit observasi

2. Proses Pengolahan data

- a. Melakukan Pengintegrasian data
- b. Melakukan klasifikasi
- c. Melakukan Pemeriksaan, Pengkodean, dan Penginputan

3. Pengolahan data dengan Ms. Excel

- a. Entri data
- b. Validasi data
- c. Memeriksa duplikasi data
- d. Eksplorasi Data
- e. Tabel
- f. Grafik
- g. Pengecekan pencilan

E. Manfaat

Berbekal hasil belajar pada Modul Pengolahan Data ini, peserta diharapkan mampu menerapkan dan mereplikasi tata cara pengolahan data pada unit kerja masing-masing. Pada akhirnya, diharapkan juga dapat meningkatkan kinerja unit kerjanya

BAB II . DATA, STATISTIK, DAN INFORMASI

Sebelum pembahasan mengenai statistik secara lebih lanjut, perlu adanya pemahaman/pengertian terhadap data, informasi, dan statistik terlebih dahulu. Dalam penggunaan biasa istilah data dan informasi sering dipertukarkan, pada hal sesungguhnya secara konsep berbeda. Demikian juga data, statistik dan informasi sering dianggap sama saja. Untuk mendapatkan gambaran yang lebih pas berikut ini diberikan konsep/ pengertian dari istilah–istilah tersebut.

A. Arti dan Pengertian Data

Dalam kamus besar Bahasa Indonesia, data berarti sekumpulan keterangan yang benar atau nyata. Keterangan tersebut bisa berupa angka, huruf, simbol ataupun gambar. Demikian juga dalam kamus wikipedia (<http://id.wikipedia.org/wiki/Data>), data disebutkan sebagai catatan atas kumpulan fakta. Data merupakan bentuk jamak dari datum, berasal dari bahasa Latin yang berarti "sesuatu yang diberikan". Dalam penggunaan sehari-hari data berarti suatu pernyataan yang diterima secara apa adanya. Dari sudut pandang bisnis, data bisnis adalah deskripsi organisasi tentang sesuatu (resources) dan kejadian (transactions) yang terjadi. Pengertian yang lain menyebutkan bahwa data adalah deskripsi dari suatu kejadian yang kita hadapi .

Kata data bisa berarti juga fakta mentah dan tidak terukur, yang bila diolah akan menjadi suatu informasi. Atau dengan kata lain tujuan mendasar dari pengumpulan dan pengolahan data adalah untuk menghasilkan informasi. Meskipun data merupakan bahan utama informasi, namun perlu diingat tidak semua data relevan dan tepat waktu untuk menghasilkan informasi.

Contoh:

- a. Data umur peserta pelatihan statistisi: 28, 30, 24, 25, 25, 26, 27, 28, 31, 31, 24, 27, 29, 29, 30, 31, 28, 29, 25, 23, 26, 27, 25, 24, 23.
- b. Data nilai ujian komprehensif dari peserta diklat fungsional statistik ahli angkatan I: 89, 87, 90, 93, 94, 94, 95, 82, 80, 76, 80, 90, 100, 92, 90, 98, 80, 83, 78, 77, 90, 96, 86, 87, 80.

B. Arti dan Pengertian Statistik

Kata statistik sering menimbulkan bayangan akan angka dalam tabel atau grafik-grafik. Kata statistik berasal dari bahasa latin "status" yang dalam bahasa inggris berarti "state" yang diartikan dalam bahasa indonesia sebagai "pernyataan" (Johnson and Bhattacharya, 2007). Memang dalam pengertian awam statistik bersinonim dengan data, dimana dalam bahasa inggris disebut dengan statistic. Statistik dalam hal ini juga biasa dikenal dengan informasi, karena telah melalui proses pengolahan, analisis dan penyajian yang tepat dan menjadi bermakna.

Statistik juga berarti nama suatu ilmu, yaitu ilmu yang berkaitan dengan data, yang kalau dalam bahasa inggris dikenal dengan istilah statistics. Dalam modul ini selanjutnya kata statistik diartikan sebagai produk kegiatan statistik, khususnya pengumpulan, pengolahan data, meringkas dan menginterpretasikan sampai pada pengambilan keputusan (Johnson and Bhattacharya, 2007). Apakah suatu statistik sudah berupa informasi? Jawabannya bisa subyektif, tergantung kebutuhan khusus pengguna sendiri, karena suatu tabel statistik mungkin sudah merupakan informasi bagi A tetapi belum bagi B (sebab masih perlu diolah lagi).

Sedangkan yang dimaksud dengan data statistik adalah deskripsi numerik dari aspek kuantitatif sesuatu. Namun demikian tidak semua diskripsi numerik adalah data statistik, dikatakan data statistik jika mempunyai ciri-ciri sebagai berikut:

- a. Harus dalam bentuk agregat (seperti: jumlah, rata-rata, proporsi dan persentase).
- b. Perubahan karakteristik fenomena yang diteliti disebabkan oleh sejumlah kekuatan yang bekerja secara simultan. Contoh: perubahan yang terjadi pada tingkat penjualan, kualitas produk, jumlah penduduk, dsbnya.
- c. Diperoleh dengan cara pencacahan/pengukuran atau diestimasi berdasarkan standar akurasi yang memadai.
- d. Harus dikumpulkan secara sistematis bagi maksud/tujuan yang telah ditetapkan sebelumnya.

- e. Fakta harus dapat dibandingkan, misalnya perbandingan antar waktu bagi fakta yang sama. Fakta (data) yang sama diperoleh dengan cara dan pengukuran yang sama pula dan hanya waktu pelaksanaan pengukuran/pencacahan yang berbeda. Sebagai contoh: Jumlah impor beras Indonesia tahun 2013 dibandingkan dengan data yang sama untuk tahun 2012.

C. Arti dan Pengertian Informasi

Informasi berbeda secara mendasar dengan data. Informasi adalah hasil agregasi atau pengolahan data yang akan meningkatkan pengetahuan dan wawasan pengguna informasi. Data dapat menjadi informasi setelah dilakukan pemilihan, pengurutan, analisis dan interpretasi yang tepat dan relevan dengan masalah yang dihadapi. Fungsi utama informasi (statistik) adalah memberikan penjelasan atau keterangan, mengurangi adanya banyak pilihan/alternatif dan meningkatkan kepastian. Suatu keputusan yang berlandaskan informasi akan lebih bijaksana.

Contoh:

Pusdiklat BPS baru saja menyelenggarakan diklat statistik sektoral. Diklat diikuti oleh 25 peserta dengan umur yang relatif muda, yaitu rata-rata umur peserta diklat baru mencapai 27 tahun. Hasil ujian komprehensif menunjukkan nilai rata-rata yang cukup bagus yaitu 87,48. Namun demikian ada 3 orang peserta yang dinyatakan lulus bersyarat karena memperoleh nilai komprehensif dibawah 80.

D. Kegunaan Data dan Statistik

Setelah memahami arti dan pengertian data dan statistik, tentunya perlu diperluas wawasan tentang manfaat dan kegunaannya sehingga dapat lebih dipahami tentang konsep dasar statistik sebagai bahan pembelajaran berikutnya. Guna memahami kegunaan data dan statistik, di dalam bagian ini akan diuraikan tentang jenis dan kegunaan data dan statistik secara singkat.

D.1. Jenis Data

Seperti telah diuraikan sebelumnya, data merupakan sekumpulan angka, huruf atau simbol/gambar, Jenis-jenis data dapat dibagi berdasarkan sifatnya, sumbernya, cara memperolehnya, waktu pengumpulannya dan skala datanya sebagai berikut.

a. Menurut sifatnya

Berdasarkan sifatnya, data dapat dibedakan menjadi:

- Data Kualitatif, yaitu data yang tidak berbentuk angka, misalnya: Kuesioner Pertanyaan tentang suasana kerja, kualitas pelayanan sebuah rumah sakit atau gaya kepemimpinan, dll.
- Data Kuantitatif, data yang berbentuk angka, misalnya: harga saham, besarnya pendapatan, dll.

b. Menurut waktu pengumpulannya

Data dapat dikumpulkan pada beberapa waktu yang berbeda, maka data berdasarkan waktu pengumpulannya dapat dibedakan sebagai:

- Data *cross section*, yaitu data yang dikumpulkan pada suatu waktu tertentu (at a point of time) untuk menggambarkan keadaan dan kegiatan pada waktu tersebut. Misalnya; data penelitian yang menggunakan kuesioner.
- Data berkala (*time series data*), yaitu data yang dikumpulkan dari waktu ke waktu untuk melihat perkembangan suatu kejadian/ kegiatan selama periode tersebut. Misalnya, perkembangan uang beredar, harga 9 macam bahan pokok penduduk.

c. Menurut cara memperolehnya

Menurut cara memperolehnya, data terdiri dari:

- Data Primer (primary data): data primer adalah data yang dikumpulkan sendiri oleh perorangan/suatu organisasi secara langsung dari objek yang diteliti dan untuk kepentingan studi yang bersangkutan yang dapat berupa interview, observasi.

- Data Sekunder (secondary data): data sekunder adalah data yang diperoleh/ dikumpulkan dan disatukan oleh studi-studi sebelumnya atau yang diterbitkan oleh berbagai instansi lain. Biasanya sumber tidak langsung berupa data dokumentasi dan arsip-arsip resmi.

d. Menurut sumbernya data

Bila dilihat dari sudut pandang suatu organisasi, menurut sumbernya data dibedakan menjadi (Supranto, 1990)

- Data Internal: data internal adalah data dari dalam suatu organisasi yang menggambarkan keadaan organisasi tersebut. Contohnya: suatu perusahaan, jumlah karyawannya, jumlah modalnya, atau jumlah produksinya, dll.
- Data Eksternal: data eksternal adalah data dari luar suatu organisasi yang dapat menggambarkan faktor-faktor yang mungkin mempengaruhi hasil kerja suatu organisasi. Misalnya: daya beli masyarakat mempengaruhi hasil penjualan suatu perusahaan.

e. Menurut skala data

Bila dilihat dari skala data, maka jenis-jenis data dapat terbagi antara lain:

- Data Nominal: data yang diperoleh dengan cara katagorisasi atau klasifikasi. Data Nominal tidak dapat dilakukan operasi matematika (\times , $+$, $-$ atau $:$). Misalnya, jenis pekerjaan diklasifikasikan dalam pegawai negeri (kode 1), pegawai swasta (kode 2) dan wiraswasta (kode 3). Dari contoh itu, tidak mungkin $3-2=1$ (wiraswasta –pegawai negeri = pegawai swasta).
- Data Ordinal: data yang diperoleh dengan cara katagorissi atau klasifikasi, tetapi diantara data tersebut terdapat hubungan. Data Ordinal juga tidak bisa dilakukan operasi matematika. Misalnya, kepuasan pelanggan, diklasifikasikan sebagai: sangat puas (diberi tanda 1), Puas (diberi tanda 2), Cukup puas (diberi tanda 3), Tidak puas(diberi tanda 4), dan Sangat tidak puas (diberi tanda 5). Posisi 1 lebih tinggi dibanding

posisi 2, Tidak mungkin $1+2=3$ (yang berarti sangat puas ditambah puas = cukup puas)

- Data Interval: data yang diperoleh dengan cara pengukuran, dimana jarak antar dua titik pada skala, sudah diketahui. Berbeda dengan skala ordinal, dimana jarak dua titik tidak diperhatikan (seperti berapa jarak antara puas dan tidak puas, yang sebenarnya menyangkut perasaan orang saja)

Misalnya, temperatur ruangan. Bisa diukur dalam Celsius, atau Fahrenheit, dengan masing-masing punya skala sendiri. Untuk air membeku dan mendidih: Celcius pada 0°C sampai 100°C . Skala ini jelas jaraknya, bahwa $100-0=100$. Fahrenheit pada 32°F sampai 212°F . Skala ini jelas jaraknya, $212-32=180$

- Data Rasio: adalah data yang diperoleh dengan cara pengukuran, dimana jarak dua titik pada skala sudah diketahui, dan mempunyai titik nol yang absolut. Ini berbeda dengan skala interval, dimana tak ada titik nol mutlak/absolut. Seperti titik 0°C tentu beda dengan titik 0°F . atau pergantian tahun pada system kalender Masehi (setiap 1 Januari) berbeda dengan pergantian tahun Jawa, China dan lainnya. Sehingga tak ada tahun baru dalam artian diakui oleh semua kalender sebagai tahun baru.

Misalnya, jumlah buku di kelas: Jika 5, berarti ada 5 buku. Jika 0, berarti tak ada buku (absolut 0). Tidak ada kategorisasi atau pemberian kode. Bisa dilakukan operasi matematika. Misal: $100\text{ cm} + 35\text{ cm} = 135\text{ cm}$; $5\text{ mangga} + 2\text{ mangga} = 7\text{ mangga}$.

D.2. Kegunaan Statistik di Berbagai Sektor

Dalam perkembangan ilmu pengetahuan dan teknologi informasi, statistik banyak diterapkan dalam berbagai sektor, baik di sektor bisnis, ekonomi, pertanian, industri, maupun sosial, politik dan pertahanan/keamanan. Di sektor pemerintahan misalnya, statistik digunakan sebagai alat pengambil keputusan dalam rangka macam tujuan pembangunan, diantaranya adalah untuk mengetahui jumlah dan karakteristik penduduk yang dilakukan melalui kegiatan sensus penduduk.

Dari kegiatan statistik ini dapat diperoleh beberapa gambaran penduduk diantaranya: jumlah penduduk, tingkat pendidikan, penyebaran penduduk, ketenagakerjaan dan karakteristik sosial budaya lainnya. Perkembangan harga (dalam bidang ekonomi) juga dapat diketahui dengan data statistik yang dikumpulkan secara sampel melalui survei, yaitu Survei Harga Konsumen yang dilakukan secara harian, mingguan dan bulanan sehingga dapat menghasilkan angka inflasi.

Dalam rangka mendukung pembangunan nasional dan wilayah, maka statistik daerah sangat diperlukan sebagai bahan penyusunan perencanaan daerah dan nasional, sebagai bahan pemantauan pelaksanaan pembangunan, dan selanjutnya sebagai bahan evaluasi hasil pembangunan. Ada kata pepatah membangun itu sulit, tapi lebih sulit lagi membangun tanpa data, karena suatu perencanaan dan kebijakan tidak akan bermanfaat dan tepat sasaran kalau tidak didukung dengan adanya data yang akurat.

Dengan perkembangan ekonomi nasional dan global saat ini, data dan statistik juga dirasakan semakin penting peranannya dalam pembangunan sektor ekonomi. Sebagai contoh dengan berkembangnya bisnis asuransi dan valuta asing, maka teori probabilitas (hitung peluang) menjadi dasar pemikiran para pelaku usaha di bidang ini. Dalam sektor industri dan perdagangan, data statistik tentang besarnya kepemilikan modal dan asset perusahaan, penggunaan faktor produksi, hasil produksi dan pemasaran merupakan faktor penting untuk ditelaah demi kemajuan perkembangan bisnis atau industri dan perdagangan yang lebih luas lagi.

Soetjipto (2007) mengatakan bahwa ada empat alasan mengapa statistik itu penting dalam dunia bisnis, yaitu pertama karena statistik digunakan sebagai catatan kegiatan usaha dan keterangan lain yang berkaitan dengan kegiatan usahanya; kedua karena data statistik merupakan bagian penting dalam laporan bisnis/usahanya; ketiga karena statistik sebagai alat kontrol yang memberikan informasi kuantitatif bagi direksi dalam melakukan pengendalian operasional usaha/bisnis; dan keempat karena statistik sebagai dasar pengambilan keputusan yang telah memperhitungkan faktor resiko. Untuk pengembangan bisnis biasanya dilakukan dahulu survei pasar untuk menggali informasi tentang pangsa pasarnya, siapa target pembelinya, seberapa besarnya, kapan dan dimana cocok dipasarkannya. Survei yang dilakukan tentunya harus mengikuti prosedur yang benar

bersarkan tehnik pengumpulan data yang tepat, diolah dengan cara yang tepat dan benar serta dianalisis sesuai kebutuhan dengan menggunakan kaidah-kaidah statistik yang ada.

Penerapan ilmu statistik lainnya yang sekarang populer adalah prosedur jajak pendapat atau pooling (misalnya dilakukan sebelum pemilihan umum). Sudah sering kita dengar dan baca dalam pelaksanaan pemilihan kepala daerah, sering ada pooling tentang prediksi hasil pilkada, atau bahkan untuk melihat hasilnya secara cepat dilakukan perhitungan cepat hasil pemilu atau quick count dari para calon kepala daerah.

Dengan dasar perhitungan angka-angka indeks dapat dikembangkan berbagai indikator baru yang sedang aktual saat ini seperti Indeks Kualitas Hidup Manusia, Indeks Pembangunan Manusia, Indeks Pemberdayaan Gender, Indeks Kebahagiaan, Indeks Demokrasi Indonesia dan lain-lainnya. Semua indikator tersebut dapat memberikan gambaran perkembangan keadaan ekonomi, sosial dan politik antar wilayah maupun antar waktu.

Dengan berbagai jenis data tersebut diatas, tentunya banyak sekali kegunaan dari data yang telah dikumpulkan, karena dengan data dapat diperoleh gambaran tentang suatu keadaan yang sedang menjadi perhatian. Secara rinci kegunaan data adalah sebagai berikut:

- i. Sebagai dasar untuk penyusunan perencanaan kegiatan.
- ii. Perencanaan yang baik dan tepat harus didasari dengan data yang akurat, karena bila data yang digunakan sebagai dasar perencanaan tidak menggambarkan keadaan yang sebenarnya maka perencanaannya juga menjadi tidak tepat.
- iii. Sebagai dasar pembuatan keputusan dalam memecahkan masalah.
- iv. Setiap masalah tentu ada penyebabnya, untuk melihat penyebab terjadinya masalah maka dapat diidentifikasi faktor-faktornya melalui data-data yang berhubungan dengan permasalahan yang sedang diperhatikan.
- v. Sebagai alat kontrol/pengendalian dalam pelaksanaan suatu kegiatan yang telah direncanakan.
- vi. Dengan adanya data yang berupa pemantauan pelaksanaan suatu kegiatan, maka dapat diketahui sejauh mana pelaksanaan sesuai dengan perencanaan yang telah disusun.

- vii. Sebagai bahan evaluasi hasil suatu kegiatan.
- viii. Setelah kegiatan selesai dilaksanakan, maka diperlukan data yang dapat menggambarkan sejauh mana pelaksanaan kegiatan itu telah berhasil dilaksanakan atau gagal dalam mencapai tujuannya.

E. Syarat data dan data statistik yang baik

Dalam memanfaatkan data dan statistik dalam berbagai sektor, maka data dan statistik yang dapat digunakan adalah data dan statistik yang memiliki kualitas yang baik, dalam hal menilai apakah kualitas data dan statistik baik atau tidak, maka faktor-faktor yang dapat menjadi penilaian adalah sebagai berikut:

a. Objektif

Data yang objektif berarti bahwa data harus sesuai dengan keadaan yang sebenarnya (as it is). Contoh data yang tidak objektif adalah Rp.5.000,- yang dikatakan dengan harganya Rp.7.500,- karena dalam kuitansi tertulis Rp.7.500,- tersebut adalah data yang tidak objektif.

b. Representatif (mewakili)

Data statistik harus mewakili kelompok objek yang di amati. Misalnya data rata-rata produksi padi di daerah A yang hanya diwakili oleh rata-rata produksi dari sawah yang subur saja. Data ini tidak mewakili rata-rata (atau gambaran umum) produksi padi di daerah A yang sesungguhnya.

c. Bergalat (error) kecil

Suatu data yang baik bila mempunyai tingkat kesalahan yang kecil, yaitu beda yang kecil dari nilai yang sesungguhnya (true value). Kesalahan tersebut berupa kesalahan sampling (sampling error) (bila data tersebut berdasarkan pengamatan sebagian sampel) atau kesalahan bukan sampling (non-sampling error) (yang dapat terjadi baik data berdasarkan sampel atau berdasarkan pengamatan keseluruhan).

d. Tepat waktu (Timeliness).

Apabila data akan dipergunakan untuk melakukan perencanaan, pengendalian atau evaluasi, maka syarat data yang tepat waktu ini penting sekali agar bisa bermanfaat, misalnya dalam pengendalian kegiatan maka sempat dilakukan penyesuaian dalam

pelaksanaan kegiatan kalau ada kesalahan atau penyimpangan di dalam implementasi. Data yang tersedia tetapi tidak pada saat diperlukan tentunya kurang bermanfaat.

e. Relevan.

Data yang dikumpulkan harus ada hubungannya dengan masalah yang akan diselesaikan. Misal, pemerintah ingin menanggulangi korban banjir di suatu daerah maka yang perlu dicatat adalah data dari mereka yang terkena korban banjir di daerah tersebut, bukan data dari penduduk lain.

f. Reliabel.

Data yang reliabel adalah data yang dapat diandalkan, yaitu data yang berkualitas yang dikumpulkan dan diolah dan disajikan mengikuti kaidah-kaidah statistik.

F. Pengertian Variabel

Sebuah variabel (peubah) adalah sebuah karakteristik, angka, atau kuantitas dari unit pengamatan yang dapat diukur atau dihitung. Sebuah variabel dapat juga disebut sebagai sebuah butir data (data item). Disebut variabel atau peubah karena nilainya dapat bervariasi antar unit pengamatan atau berubah-ubah dari suatu unit pengamatan ke unit pengamatan lain, dan juga dapat berubah antar waktu. Umur, berat, tinggi, pendapatan, tempat lahir, pandangan terhadap program pemerintah, misalnya, merupakan contoh variabel yang umum ditemui dalam kehidupan sehari-hari.

Jelas bahwa variabel ini nilainya bisa berbeda-beda antar unit pengamatan dan juga bisa berubah antar waktu. Suatu unit pengamatan, misal individu, bisa diukur satu variabelnya saja, misal umur, atau bisa dihitung banyaknya teman yang dimiliki. Jenis-Jenis variabel, dapat diuraikan sebagai berikut:

a. Variabel Kontinu

Variabel yang memiliki nilai sembarang, baik berupa nilai bulat maupun pecahan, diantara dua nilai tertentu atau variabel yang mengambil seluruh nilai dalam suatu interval. Contoh: berat badan, tinggi, luas, pendapatan, dsb. Berat badan dapat ditulis 45 kg; 15 kg; atau 52,125 kg.

b. Variabel Diskrit

Konsep yang nilainya tidak dapat dinyatakan dalam bentuk pecahan atau desimal di belakang koma. Variabel Diskrit ini sering juga disebut sebagai variabel kategori. Bila dalam satu variabel tersebut mempunyai dua kategori saja maka variabel tersebut dinamakan variabel dikhotom.

Contoh dikhotom seperti: Jenis kelamin Laki-laki, Perempuan; Status perkawinan Kawin, tidak kawin. Adapun contoh Politem yaitu Tingkat pendidikan SD, SMP, SMA, PT Jumlah Anak 1, 2, 3, 4.

c. Variabel Dependen dan Independen

Variabel independen (independent variable) adalah tipe variabel yang menjelaskan atau mempengaruhi variabel yang lain. Variabel dependen (dependent variable) adalah tipe variabel yang dijelaskan atau dipengaruhi oleh variabel independen. Kedua tipe variabel ini merupakan kategori variabel penelitian yang paling sering digunakan dalam penelitian karena mempunyai kemampuan aplikasi yang luas.

Contoh, ada hubungan antara konsumsi dan pendapatan, di mana dengan bertambahnya pendapatan, konsumsi juga akan bertambah, maka Konsumsi merupakan variabel dependen (terikat dengan pendapatan), sedangkan Pendapatan adalah variabel independen (variabel bebas)

d. Variabel Moderator dan Random

Variabel moderator merupakan variabel yang mempengaruhi (memperkuat dan memperlemah) hubungan antara variabel bebas dan variabel terikat. Variabel Moderator disebut juga variabel independen kedua. Sedangkan untuk menggambarkan hasil-hasil percobaan sebagai nilai-nilai numerik secara sederhana kita menggunakan apa yang disebut sebagai variabel acak/variabel random. Jadi variabel acak dapat didefinisikan sebagai deskripsi numerik dari hasil percobaan.

Contoh, variabel yang mempengaruhi kebutuhan(demand) terhadap ikan (Y) adalah harga ikan (X1), pendapatan (X2), dan harga daging (X3). Ketiga variabel tersebut adalah variabel utama. Jika umur (X4) juga berpengaruh, tetapi bukanlah sebagai penyebab utama, maka umur (X4) ini lah yang disebut dengan variabel moderator. Disamping variabel tertentu yang nyata mempengaruhi variabel dependen, masih terdapat variabel lain yang tidak

dimasukkan dalam persamaan hubungan di atas. Variabel ini dinamakan variabel random, dan pengaruhnya dapat dilihat berdasarkan error yang timbul dalam mengadakan estimasi.

G. Unit Observasi dan Unit Analisis

Unit observasi adalah sesuatu yang dijadikan sumber untuk memperoleh data dalam rangka menggambarkan atau menjelaskan tentang satuan analisis, sesuatu yang menjadi sumber itu dapat berupa orang, dokumen, dan tempat penelitian. Unit analisis adalah unit yang kepadanya kesimpulan diberikan, unit analisis juga dipandang sebagai sesuatu yang tentangnya peneliti menghimpun informasi dan dari padanya peneliti menarik kesimpulan.

Sebagai contoh, Jika penelitian akan “menghitung rumah”, maka sesuatu yang dijadikan “satuan hitungan” itu jadilah sebagai “unit analisis.” Misal, jika yang dianalisis (dihitung) berapa orang “ibu rumah tangga” yang bekerja dan tidak bekerja dari penghuni sesuatu kompleks perumahan, maka yang menjadi unit analisisnya adalah “ibu rumah tangga”. Jika yang dianalisis tingkat pendidikan yang sedang atau sudah ditempuh oleh warga perumahan tersebut, maka yang menjadi unit analisisnya adalah orang per orang (individu) penghuni perumahan. Objek (hal; suka disebut pula sebagai “variabel penelitian”) yang diteliti adalah tingkat pendidikan (dari orang-orang tersebut).

Unit analisis dalam penelitian itu bisa berupa:

- (1) “individu” (dapat berupa orang atau “satuan-satuan” setara lainnya semisal ternak, kendaraan bermotor, alat peraga, media pendidikan, dan koleksi perpustakaan); yang menjadi objek penelitiannya misalnya mengenai umurnya (jika orang atau ternak, lama pemilikan jika kendaraan dan barang lainnya);
- (2) kelompok (keluarga atau rumah tangga, kelas/murid sekelas, dsb); yang menjadi objek penelitiannya misalnya mengenai kekerapan terjadinya “kekerasan” terhadap anggota/warganya;
- (3) organisasi (sekolah, perguruan tinggi, pemerintah daerah, organisasi profesi, dsb); yang menjadi objek penelitiannya misalnya mengenai standar layanan minimalnya, efektivitasnya dsb;
- (4) satuan geografis (kota, desa, kelurahan, kecamatan, kabupaten, propinsi, negara); yang menjadi objek penelitiannya misalnya mengenai rerata penghasilan warganya, iklim politiknya dsb;

(5) interaksi sosial (perceraian, komunikasi layanan publik, pertukaran email, pengasuhan anak, dsb); yang menjadi objek penelitiannya misalnya polanya, intensitasnya dsb.

Apa perbedaan unit analisis dan unit amatan (*unit of observation*)? *Observation* sebenarnya mengandung arti “pengumpulan data”. Jadi, unit observasi dimaksudkan satuan-satuan (seperti dalam analisis) yang menjadi “sumber data dihimpun.” Bisa saja yang menjadi unit analisis sama dengan unit observasinya. Misalnya jika yang dianalisis itu prestasi belajar (hasil UAN) siswa berupa berapa orang yang lulus dan tidak lulus, yang menjadi unit observasi (sumber data digali) adalah orang per orang siswa (siswa A lulus/tidak, siswa B lulus/tidak, siswa n lulus/tidak, dst), dan yang menjadi unit analisis (satuan hitungan) pun orang per orang siswa (dihitung berapa **orang siswa** yang lulus dan tidak lulus).

Ketika yang dianalisis rerata prestasi belajar (hasil UAN) siswa sekolah A dan Sekolah B, seperti contoh di muka, yang menjadi unit observasinya orang per orang siswa (tiap siswa “didata” skor UAN-nya berapa), tetapi yang dianalisis adalah keseluruhan siswa satu sekolah, sehingga memunculkan kesimpulan: rerata skor hasil UAN siswa (seluruh siswa) Sekolah A sekian, sedangkan siswa Sekolah B sekian. Jadi, unit observasinya orang per orang (individual) siswa, tetapi unit analisisnya “kelompok siswa” (sekolah).

BAB III. PROSES PENGOLAHAN DATA

Tahapan proses atau pengolahan data menentukan seberapa jauh tingkat akurasi dan ketepatan data statistik yang dihasilkan. Tahap ini mendeskripsikan persiapan data sebelum data tersebut dianalisis dan didiseminasikan sebagai output. Pada tahap ini, data yang digunakan merupakan data final hasil pengumpulan data.

A. Melakukan Integrasi Data

Integrasi data adalah aktivitas yang bertujuan menggabungkan data yang berasal dari dua atau lebih sumber data. Dengan kata lain, data hasil pengumpulan data yang disimpan secara terpisah akan digabungkan sebelum diolah lebih lanjut. Data dalam hal ini bisa berupa dokumen hasil pencacahan atau dalam bentuk lain sesuai dengan tools yang digunakan dalam pengumpulan data. Data yang digabungkan dapat berasal dari sumber internal dan eksternal.

Data internal adalah data yang diperoleh dari dalam organisasi penyelenggara kegiatan statistik, sedangkan data eksternal adalah data yang diperoleh dari luar organisasi penyelenggara kegiatan statistik.

Data yang telah diintegrasikan lalu dicatat dalam suatu daftar penerimaan data oleh unit kerja yang melaksanakan penerimaan data. Kegiatan penerimaan data meliputi:

1. Menerima data dari petugas lapangan/pengumpul data;
2. Memeriksa kelengkapan jumlah data;
3. Membuat laporan perkembangan penerimaan data.

B. Melakukan klasifikasi

Pengklasifikasian atau pengelompokan dokumen (*batching*) merupakan proses pengelompokan dokumen hasil pengolahan. Pengelompokan dapat dilakukan menurut jenis dokumen dan wilayah, misalnya dokumen dikelompokkan menurut desa atau kecamatan. Dokumen yang telah dikelompokkan lalu disimpan dengan menyusun dokumen pada tempat penyimpanan dan pengelolaan dokumen agar mudah diambil apabila diperlukan pada tahapan pemeriksaan, pengkodean, dan penginputan data serta mudah pula dikembalikan ke tempat penyimpanan semula.

Kegiatan penyimpanan dokumen meliputi:

- Menyusun dokumen pada tempat penyimpanan sehingga dokumen mudah dicari;
- Menyiapkan dan memberikan dokumen kepada petugas editing/coding dan petugas data entry;
- Menerima dan menyimpan kembali dokumen dari petugas editing/coding dan petugas data entry.

C. Melakukan Pemeriksaan, Pengkodean, dan Penginputan

Editing merupakan proses pemeriksaan pada isian dokumen hasil pencacahan dengan memperhatikan kaidah-kaidah editing yang telah ditetapkan. Hasil editing sangat mempengaruhi kualitas data dan proses pengolahan selanjutnya.

Berbagai informasi yang sekiranya meragukan seharusnya sudah dapat dideteksi sejak dilakukan editing sehingga akan memperlancar kegiatan pengolahan selanjutnya. Proses editing meliputi pemeriksaan:

1. Kelengkapan pengisian

Dokumen hasil wawancara harus terisi lengkap sesuai daftar pertanyaan yang ada. Setiap pertanyaan yang diajukan harus mempunyai jawaban, sekalipun jawaban itu berbunyi “TIDAK TAHU” atau “TIDAK ADA PENDAPAT” atau “TIDAK MENJAWAB”. Apabila terdapat jawaban yang kosong, kemungkinan pewawancara lupa menanyakan pertanyaan tersebut atau lupa menulis jawabannya. Jawaban yang kosong akan menimbulkan kebingungan dalam pengkodean (*coding*).

Ketidaklengkapan isian akan menyebabkan kurangnya informasi yang dibutuhkan. Oleh karena itu tahap awal pada pemeriksaan dokumen adalah memeriksa kelengkapan isian, sehingga pertanyaan yang belum terisi dapat dilengkapi oleh pengumpul data, jika memungkinkan dilakukan kunjungan ulang ke sumber data.

2. Kejelasan isian (dapat dibaca)

Kejelasan isian dokumen juga harus diperhatikan, baik isian yang berbentuk tulisan maupun yang berbentuk angka. Ketidakjelasan tulisan maupun angka dapat menyebabkan kesalahan informasi yang diperoleh. Tulisan yang buruk atau acak-acakan atau sukar dibaca

dapat mempersulit pengolahan data. Kejelasan isian termasuk juga kejelasan makna jawaban. Pewawancara perlu menuliskan jawaban dengan pola kalimat yang lengkap, jelas, logis, serta sistematis mengikuti pola subjek, predikat, obyek, dan keterangan.

3. Kesalahan isian

Kesalahan isian merupakan hal yang harus dihindari seperti dalam kesalahan kode (laki-laki berkode 1, namun ditulis kode 2 yang merupakan kode untuk perempuan). Selain itu, data harus dicatat dalam satuan-satuan sesuai yang diinginkan dalam kuesioner. Kesalahan satuan akan mengakibatkan kesalahan data yang dihasilkan termasuk kesalahan dalam analisis data. Sebagai contoh apabila data mengenai luas tanah ditetapkan untuk diukur dalam satuan hektar;

maka jangan sampai pada kuesioner ditulis lagi dalam satuan ukuran lain (meterpersegi, kubik, atau lainnya).

4. Konsistensi isian dan relevansi jawaban

Konsistensi isian merupakan kondisi isian dokumen sesuai pertanyaan yang terkait satu dengan lainnya. Petugas harus memeriksa apakah isian jawaban suatu pertanyaan konsisten dengan isian jawaban pertanyaan lain. Ini juga harus diperhatikan. Apakah jawaban-jawaban responden yang dicatat pewawancara cukup logis dan sesuai antara jawaban di pertanyaan yang satu dengan jawaban lain di pertanyaan yang lain.

Penyebabnya mungkin saja responden berusaha menutup-nutupi sesuatu atau boleh jadi pewawancara kurang kritis, kurang teliti mencatat jawaban atau bahkan malas untuk menanyakan lebih lanjut. Jawaban responden yang ada pada dokumen sering kali tidak atau kurang berkaitan dengan persoalan sebenarnya. Hal ini dapat terjadi apabila pewawancara kurang cakap merumuskan pertanyaan yang diajukan.

5. Kewajaran isian

Kewajaran isian juga harus diperiksa walaupun isian yang kurang wajar belum dapat dipastikan data tersebut salah. Pemeriksaan kewajaran ini dilakukan sebagai informasi awal bahwa isian ini harus dikonfirmasi kebenarannya. Biasanya peneliti membuat suatu batasan kewajaran data dengan membuat suatu range nilai setiap variabel. Namun demikian ada kalanya pada suatu wilayah tertentu dapat terjadi diluar range tersebut. Misalnya harga

suatu barang berkisar 5000 – 9000 per unit, tetapi di suatu wilayah dapat terjadi harga diatas 9000 yang disebabkan karena barang di wilayah tersebut langka.

Petugas editing perlu teliti dalam memeriksa kelengkapan, kewajaran, dan konsistensi antar rincian satu dengan rincian yang lainnya. Jika menemui kejanggalan isian, dapat melaporkan ke pengawas editing agar dapat diambil tindakan seperlunya atau mengembalikan dokumen tersebut ke petugas lapangan agar diperbaiki.

Setelah proses editing, maka tahap berikutnya adalah penyandian dokumen. Penyandian (coding) merupakan kegiatan mengubah data berbentuk huruf menjadi data berbentuk angka/bilangan. Pengkodean data dilakukan untuk memberikan kode yang spesifik pada respon jawaban responden untuk memudahkan proses pencatatan data. Misalnya untuk variabel pekerjaan dilakukan coding 1 = Pegawai Negeri, 2 = Wiraswasta, 3 = Pegawai Swasta dan 4 = Pensiunan. Jenis kelamin: 1 = Pria dan 2 = Wanita, dan sebagainya. Kegunaan dari coding adalah untuk mempercepat pada saat input data dan mempermudah pada saat analisis data. Setelah proses pemeriksaan dan pengkodean data secara individu untuk setiap dokumen dilakukan, maka tahap selanjutnya adalah memasukan data ke aplikasi input data yang telah dibangun pada tahap implementasi dan telah sesuai dengan kaidah validasi instrumen pengumpulan data. Dalam melakukan input data, sebaiknya dibentuk koordinator dan petugas/operator kegiatan input data. Ketersediaan komputer PC menjadi dasar penetapan jumlah operator input data.

D. Pengolahan Data Dengan Ms. Excel

D.1. Entri Data

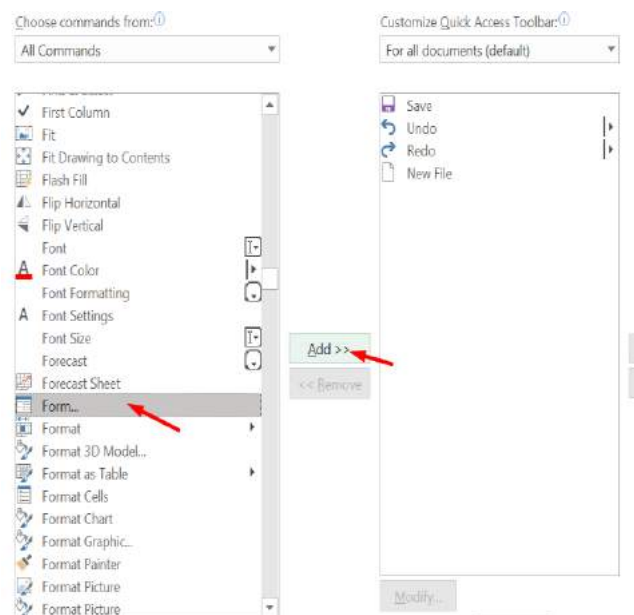
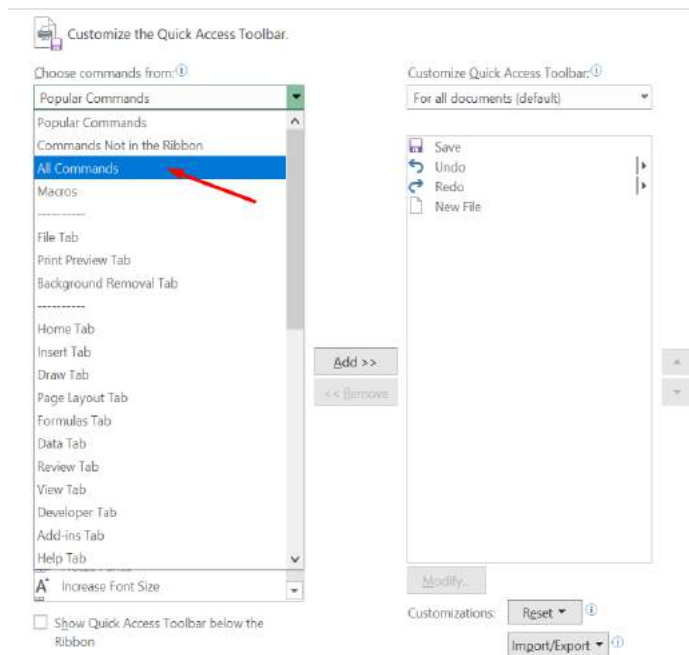
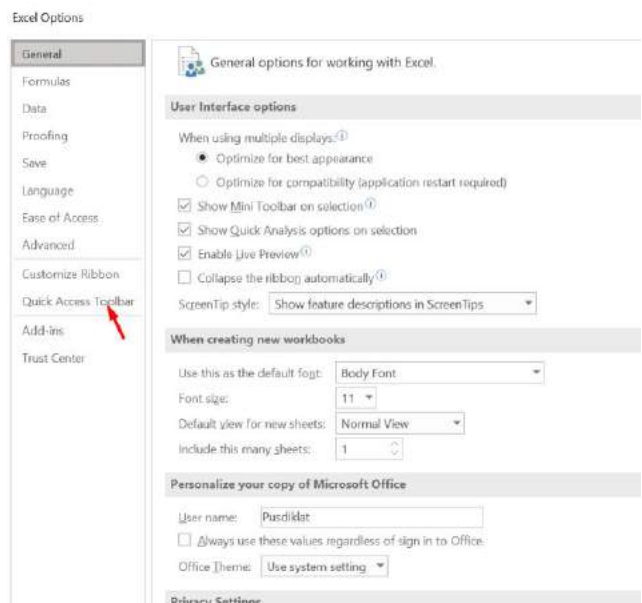
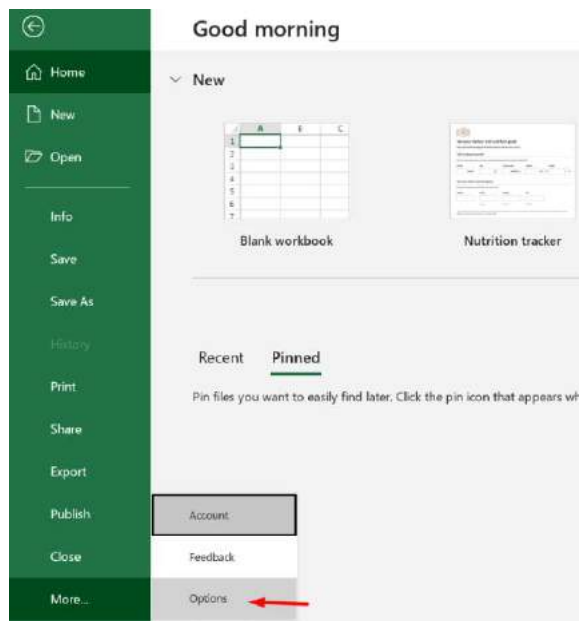
Pada Microsoft Excel kita bisa melakukan entri data secara langsung pada cell-cell nya atau kita juga bisa melakukan entri data dengan menggunakan *form*. Dengan format *form*, tampilan akan lebih menarik dan lebih mudah karena hanya muncul satu kolom yang awalnya berasal dari beberapa kolom.

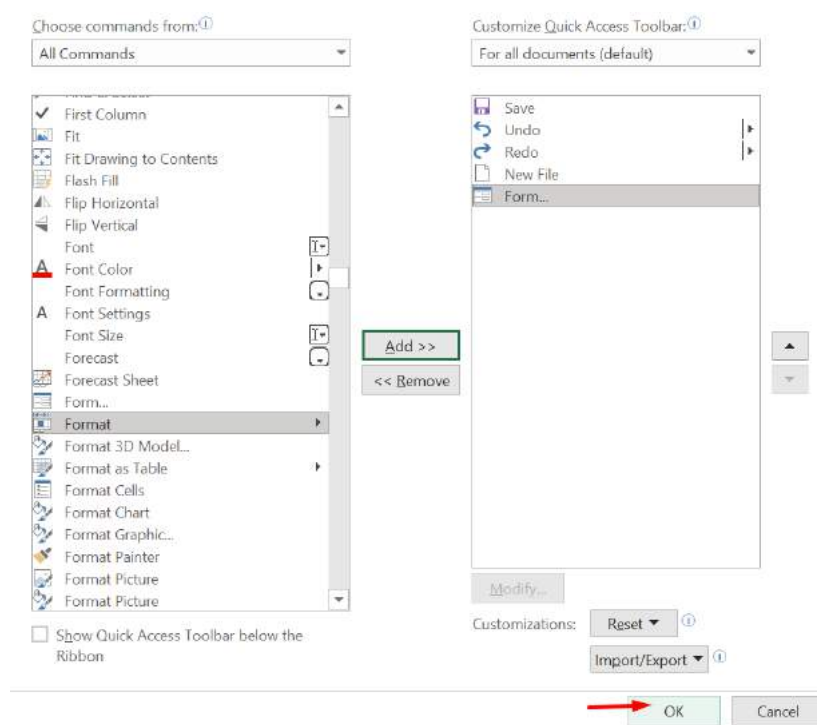
Langkah detailnya :Sebelum membuat form, tampilkan dahulu icon form lewat Customize

Quick Access Toolbar, langkahnya sebagai berikut :

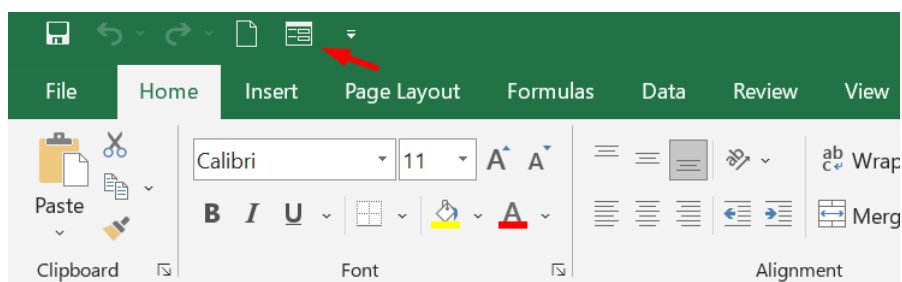
Klik: **file > options > quick access toolbar > all commands > form > add > Ok**

Secara jelas bisa dilihat pada gambar berikut ini:

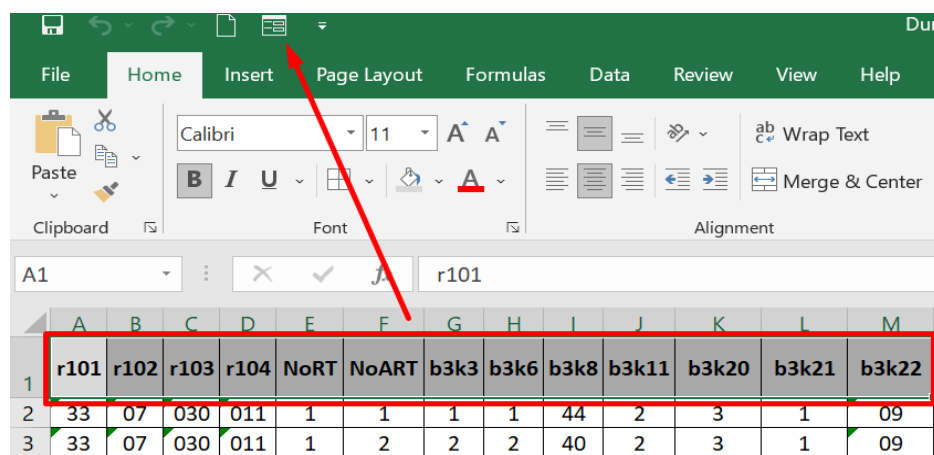




Jika kita klik Ok maka menu *form* akan tampak seperti ini



Langkah selanjutnya, buat field-field (kolom-kolom)data apa saja yang kita butuhkan jika kita belum memiliki table di excel. Jika kita sudah memiliki tabel yang memiliki judul-judul kolom, maka blok semua judul kolom kemudian klik menu *form*



individu ? X

1 of 3

r101: 33

r102: 07

r103: 030

r104: 011

NoRT: 1

NoART: 1

b3k3: 1

b3k6: 1

b3k8: 44

b3k11: 2

b3k20: 3

b3k21: 1

b3k22: 09

New

Delete

Restore

Find Prev

Find Next

Criteria

Close

Dari form yang sudah dibuat maka kita akan lebih mudah saat melakukan proses entri data.

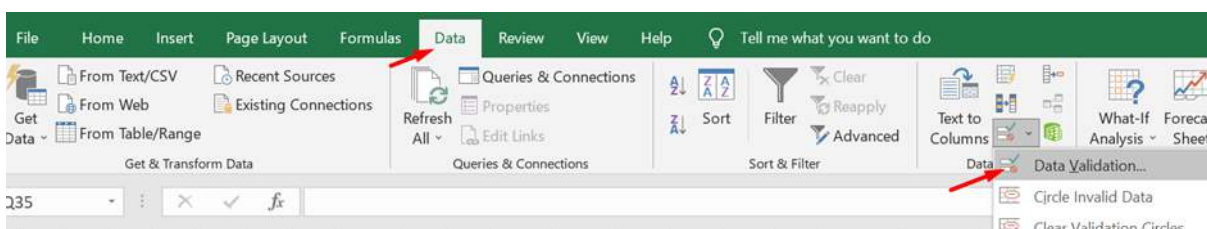
D.2. Validasi Data di Excel

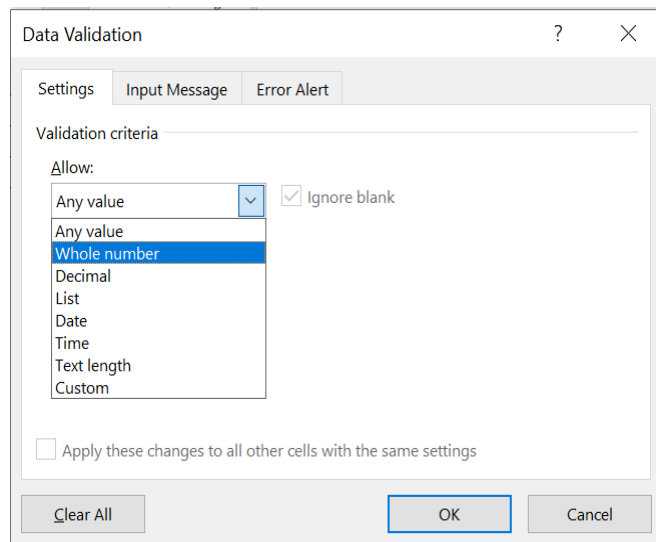
Validasi data dapat diterapkan untuk semua tipe data yang diolah Excel. Kita dapat menerapkan aturan-aturan tertentu sesuai kebutuhan untuk memberlakukan validasi data.

Misalkan kita ingin membuat isian dari variable b4r3 (luas lantai) range nya berkisar antara 5 - 200 m², hal ini bisa kita lakukan dengan memanfaatkan menu validasi di excel, Langkah-langkahnya adalah sebagai berikut :

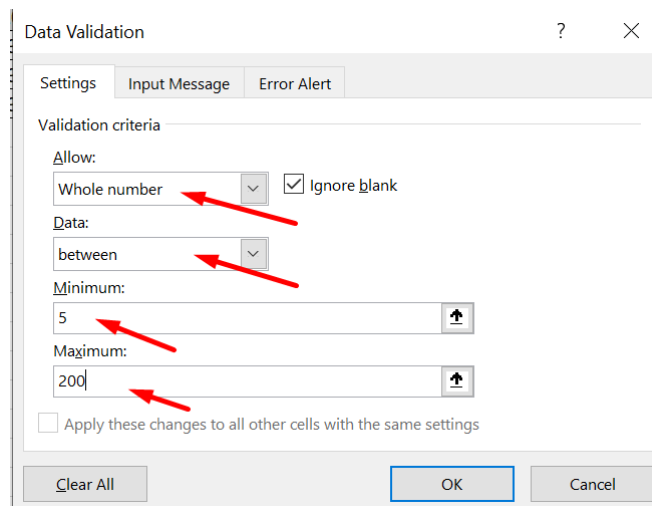
Blok range yang akan diterapkan validasi data. Pada menu **Data Tools**, klik **Data Validation**.

A	B	C	D	E	F	G	H	I	J
r101	r102	r103	r104	NoRT	JART	b4r1	b4r2	b4r3	b4r8
33	07	030	011	1	4	1	1		01
33	07	030	011	2	3	1	1		05
33	07	030	011	3	5	1	1		06

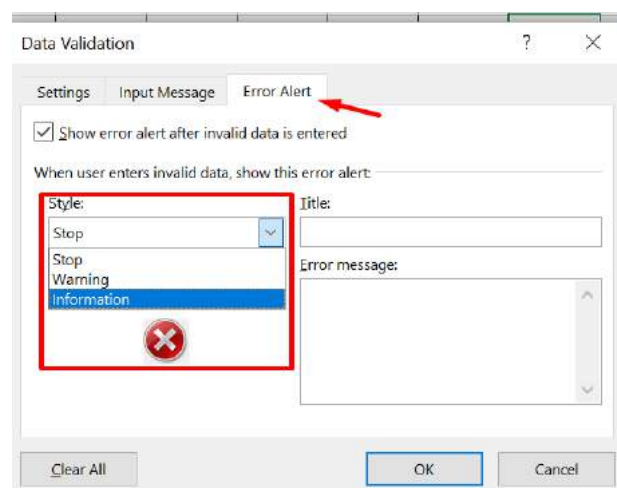




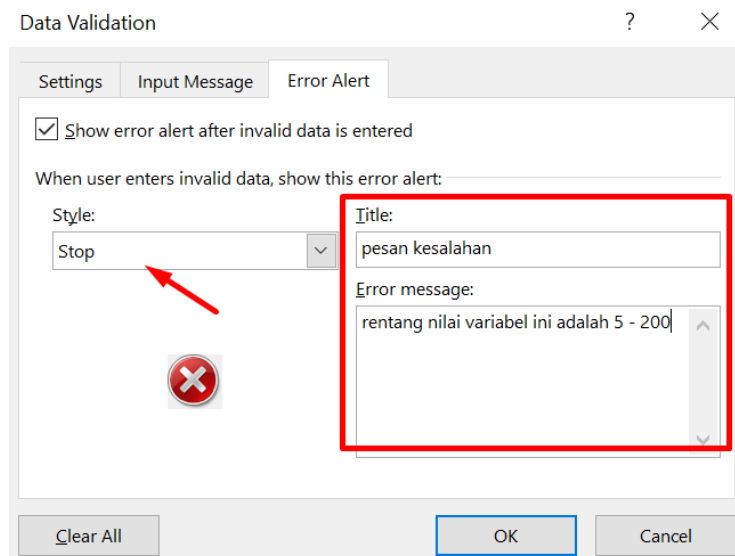
Karena kita ingin membatasi isian dalam bentuk rentang, maka kita pilih between dan isikan rentang yang kita kehendaki



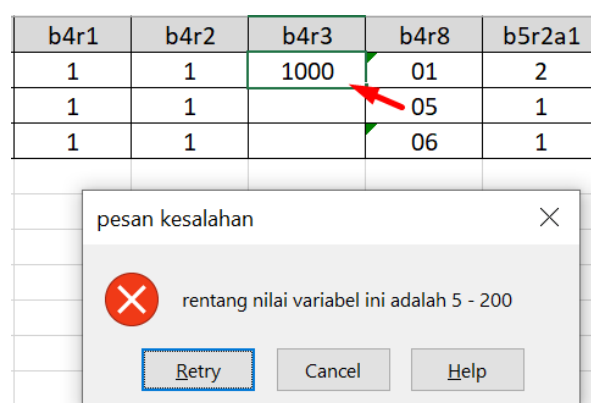
Setelah itu kita pilih *error alert*, menu ini berfungsi untuk menentukan tindakan dan pesan kesalahan jika nilai yang diisikan berada di luar rentang yang sudah ditetapkan.



kita bisa memilih beberapa style alert yaitu *Stop*, *Warning* dan *Information*. Ketiga style ini memiliki tingkat keketatan yang berbeda. Jika kita memilih style warning dan information maka jika terjadi kesalahan kita masih bisa memasukkan data tersebut, tetapi pada style Stop jika terjadi kesalahan entri kita tidak bisa melanjutkan sebelum memperbaiki isian data tersebut. Jadi kita harus memilih style terlebih dahulu. Setelah memilih style kita bisa memasukkan pesan kesalahan sesuai dengan yang dikehendaki, untuk lebih jelas bisa dilihat pada gambar berikut ini



Kemudian klik OK, maka kita sudah selesai membuat validasi. Misalkan kita isikan variable tersebut dengan angka 1000, hasilnya akan seperti ini



Terlihat dari tampilan di atas Ketika menggunakan style *Stop*, kita tidak bisa melanjutkan sampai kita memperbaiki isian yang sesuai. Untuk style *Warning* dan *Information*, bisa anda coba sendiri. Selamat mencoba !!!

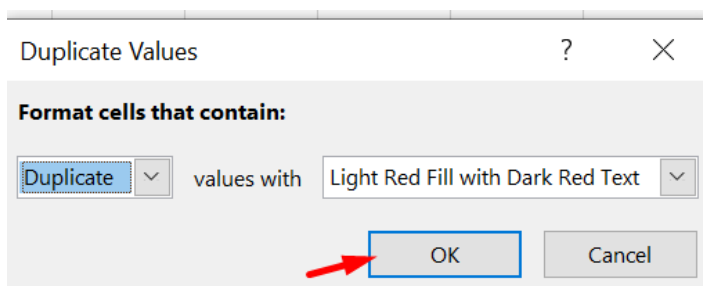
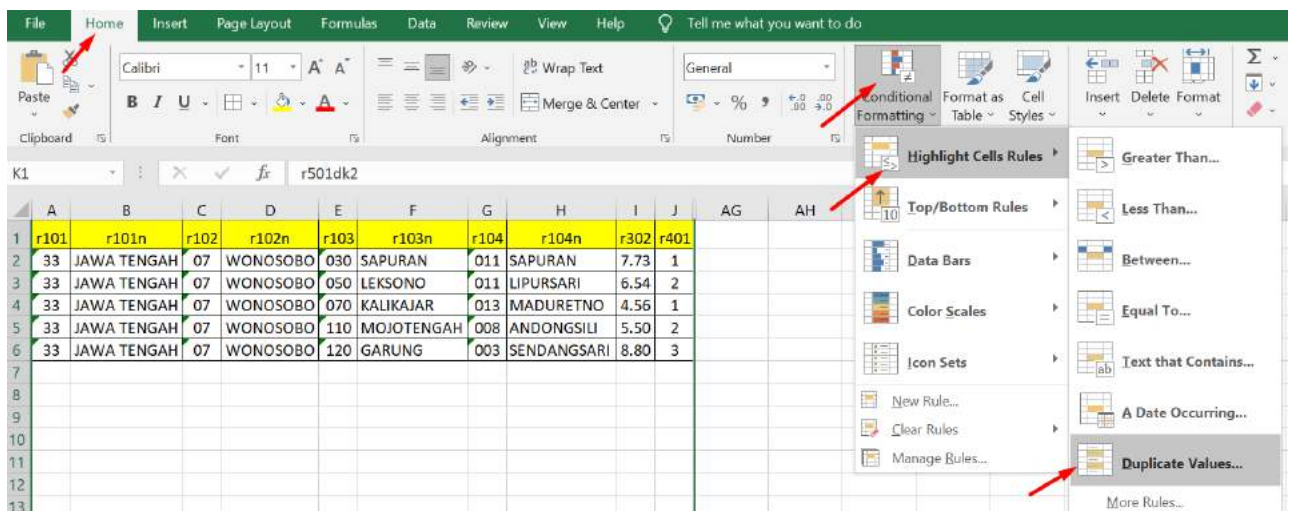
D.3. Memeriksa Duplikasi Data

Dalam menginput data, tidak jarang terjadi duplikat entri. Dengan adanya duplikat entri akan mengakibatkan hasil analisis menjadi tidak baik, sehingga pengolah data harus mengidentifikasi adanya duplikat entri, terutama untuk variable-variabel yang menjadi identitas yang harus bersifat unik, seperti NIP, NIK, No. KK, No urut ART dll. Berikut ini cara untuk mengidentifikasi adanya duplikat entri.

Blok area yang akan diidentifikasi, pada contoh ini akan dipilih variable r104 (nomor urut desa)

Pada menu Home, klik **Conditional Formatting** -> **Highlight Cell Rules** -> **Duplicate Values**

r101	r101n	r102	r102n	r103	r103n	r104	r104n	r302	r401
33	JAWA TENGAH	07	WONOSOBO	030	SAPURAN	011	SAPURAN	7.73	1
33	JAWA TENGAH	07	WONOSOBO	050	LEKSONO	011	LIPURSARI	6.54	2
33	JAWA TENGAH	07	WONOSOBO	070	KALIKAJAR	013	MADURETNO	4.56	1
33	JAWA TENGAH	07	WONOSOBO	110	MOJOTENGAH	008	ANDONGSILI	5.50	2
33	JAWA TENGAH	07	WONOSOBO	120	GARUNG	003	SENDANGSARI	8.80	3



Klik Ok maka hasilnya akan tampak sebagai berikut:

r101	r101n	r102	r102n	r103	r103n	r104	r104n	r302	r401
33	JAWA TENGAH	07	WONOSOBO	030	SAPURAN	011	SAPURAN	7.73	1
33	JAWA TENGAH	07	WONOSOBO	050	LEKSONO	011	LIPURSARI	6.54	2
33	JAWA TENGAH	07	WONOSOBO	070	KALIKAJAR	013	MADURETNO	4.56	1
33	JAWA TENGAH	07	WONOSOBO	110	MOJOTENGAH	008	ANDONGSILI	5.50	2
33	JAWA TENGAH	07	WONOSOBO	120	GARUNG	003	SENDANGSARI	8.80	3

Terlihat bahwa ada nomor urut desa yang sama, yaitu desa sapuran dan lipursari. Pemeriksaan ini berguna untuk memastikan tidak ada nomor identitas yang sama antar record. Untuk contoh di atas walaupun nomor urut desa sama, tapi tidak masalah karena berada di kecamatan yang berbeda.

E. Eksplorasi Data

Data yang kita peroleh perlu dipahami dengan baik bagaimana polanya. Hal ini dapat dilakukan melalui eksplorasi data. Tujuan dari eksplorasi data adalah sebagai berikut (Sari dan Budiarti, 2017).

1. Memperoleh pemahaman tentang karakteristik/perilaku data dengan lebih baik sehingga membantu menentukan metode analisis yang sesuai.
2. Mengenali pola yang terdapat pada data
3. Mendeteksi pencilan (*outlier*)
4. Menangkap kesalahan dalam data yang mungkin disebabkan karena terdapat kesalahan dalam pencacahan atau input data di komputer
5. Menemukan ada atau tidaknya pelanggaran asumsi dalam melakukan analisis

Data dapat terdiri angka yang banyak atau sedikit. Ketika datanya sedikit, pola data dapat diamati langsung dari angka-angka tersebut. Ketika data tersebut banyak, diperlukan teknik untuk meringkas dan mempelajari data tersebut, baik melalui tabel atau grafik.

E1. Ukuran Pemusatan dan Penyebaran Data

Data yang banyak perlu diringkas supaya lebih sederhana sehingga gambaran tentang data tersebut dapat diperoleh. Data tersebut diringkas melalui ringkasan numerik, seperti ukuran pemusatan (*central tendency*) ukuran penyebaran. Ukuran pemusatan yang banyak dipakai adalah rata-rata hitung, median dan modus. Sementara itu, ukuran penyebaran yang biasanya digunakan adalah standar deviasi, jangkauan dan jarak antar kuartil.

a. Rata-rata

Rata-rata digunakan sebagai ukuran pemusatan data dengan tipe interval dan rasio. Rata-rata lebih tepat digunakan ketika sebaran datanya simetris. Penghitungan rata-rata dilakukan dengan menjumlahkan seluruh nilai data suatu pengamatan/observasi, kemudian dibagi dengan jumlah pengamatan tersebut. Penghitungan rata-rata di Microsoft Excel dengan cara:

Pilih sel yang kosong, ketikkan `=AVERAGE()`, dalam tanda kurung diisikan kolom yang akan dihitung rata-ratanya.

Misalnya `=AVERAGE(I2:I51)` adalah perintah menghitung rata-rata data di kolom I baris 2 sampai 51.

b. Median

Median merupakan ukuran pemusatan data untuk data bertipe ordinal, interval dan rasio. Median adalah nilai yang berada tepat di tengah suatu kelompok data, antara nilai minimum dan maksimum sehingga separuh data nilainya lebih kecil daripada median dan separuh lainnya nilainya lebih besar daripada median. Nilai minimum merupakan nilai paling kecil, sedangkan nilai maksimum merupakan nilai paling besar dari sekelompok data yang sudah diurutkan. Median juga dihitung dari data yang sudah diurutkan. Untuk menghitung nilai median, jika banyak observasi ganjil maka median merupakan observasi yang berada tepat di tengah. Sedangkan jika banyak observasi genap median merupakan rata-rata dari dua nilai observasi yang di tengah.

Penghitungan median di Microsoft Excel dilakukan dengan cara

Pilih sel yang kosong, kemudian ketikkan **=MEDIAN()**. Di dalam tanda kurung diisikan kolom yang akan dihitung mediannya. Misalnya, =MEDIAN(J2:J51) adalah perintah menghitung median data di kolom J baris 2 sampai dengan 51.

c. Modus

Modus ukuran pemusatan data untuk data bertipe nominal, ordinal, interval dan rasio.

Modus dihitung berdasarkan nilai data yang paling sering muncul atau nilai data yang punya frekuensi terbesar. Penghitungan modus di Microsoft Excel dilakukan dengan cara

Pilih sel yang kosong, ketikkan =MODE(). Di dalam tanda kurung diisikan kolom yang akan dihitung modulusnya. Misalnya, =MODE(K2:K51) adalah perintah menghitung modus data di kolom K baris 2 sampai dengan 51.

Contoh penghitungan rata-rata, median dan modus dapat dilihat pada gambar berikut.

	H	I	J	K
1	b4r2	b4r3	b4r8	b5r2a1
2	1	90	1	2
3	1	35	5	1
4	1	90	6	1
49	2	85	6	1
50	1	45	5	2
51	1	45	1	2
52	Rata-rata	=AVERAGE(I2:I51)		
53	Median	=MEDIAN(I2:I51)	=MEDIAN(J2:J51)	
54	Modus	=MODE(I2:I51)	=MODE(J2:J51)	=MODE(K2:K51)

Data pada kolom I atau variabel b4r2 adalah data luas lantai (m^2), kolom J atau variabel b4r8 adalah sumber air minum, dan kolom K atau variabel b5r2a1 adalah kepemilikan aset lahan. Data luas lantai adalah data rasio sehingga rata-rata, modus, dan mediannya dapat dihitung. Namun, sumber air minum adalah data ordinal karena angka yang paling kecil menunjukkan kualitas lebih baik dibandingkan angka yang lebih besar, sehingga nilai pemusatan yang dapat dihitung adalah median dan modus.

Sementara itu, kepemilikan lahan adalah data nominal sehingga hanya bisa dihitung modusnya. Hasil penghitungan ketiga ukuran pemusatan dapat dilihat di gambar berikut

	H	I	J	K
1	b4r2	b4r3	b4r8	b5r2a1
2		1	90	1
3		1	35	5
4		1	90	6
49		2	85	6
50		1	45	5
51		1	45	1
52	Rata-rata	62,2		
53	Median	67,5	4	
54	Modus	80	5	1

d. Jarak (*Range*)

Jarak adalah jenis pengukuran penyebarang yang paling sederhana karena cara menghitungnya dari selisih antara nilai terbesar (maksimum) dan terkecil (minimum). Penghitungan jarak yang hanya berdasarkan nilai maksimum dan minimum menyebabkan jarak tidaklah stabil. Dengan kata lain, jarak tidak dapat diandalkan sebagai ukuran penyebaran. Hal ini karena jarak sangat dipengaruhi oleh nilai-nilai pencilan. Oleh karena itu, jarak tidak direkomendasikan sebagai satu-satunya ukuran untuk mengetahui penyebaran data.

Penghitungan jarak di Microsoft Excel dilakukan dengan cara

Pilih sel yang kosong, kemudian ketikkan =MAX()-MIN(). Di dalam tanda kurung diisikan kolom yang akan dihitung jaraknya. Misalnya perintah =MAX(I2:)-MIN(I2:I51) adalah jarak data di kolom I baris 2 sampai dengan 51.

e. Jarak antar kuartil (*Inter quartile range/IQR*)

Pada data yang memiliki sebaran menceng atau tidak homogen, rata-rata tidak tepat digunakan sebagai nilai yang mewakili data, sehingga IQR dapat digunakan untuk mengukur penyebaran. Berbeda dengan jarak, jarak antarkuartil dihitung berdasarkan selisih antara kuartil 3 (Q_3) dan kuartil 1 (Q_1). Q_1 merupakan suatu nilai yang membagi data menjadi 25 persen data yang lebih kecil dan 75% data yang lebih besar.

Pada Microsoft Excel, nilai Q_1 dapat diperoleh dengan perintah **=QUARTILE(array;1)**. Sementara itu, Q_3 merupakan nilai yang membagi data menjadi 75% data yang lebih kecil dan 25% data yang lebih besar. Pada Microsoft Excel, nilai Q_3 dapat diperoleh dengan perintah **=QUARTILE(array;3)**. Oleh karena itu, IQR pada Microsoft Excel dihitung dengan perintah **=QUARTILE(array;3)- QUARTILE(array;1)**

f. Simpangan baku (standart deviation)

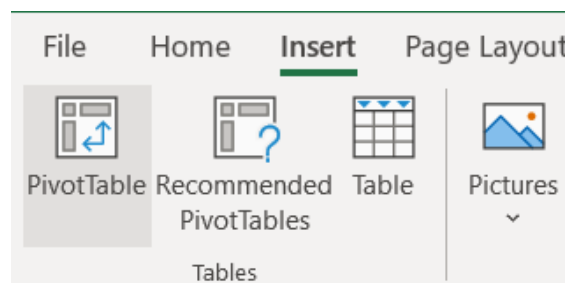
Simpangan baku menunjukkan posisi data atau penyimpangan data hasil pengamatan terhadap rata-ratanya. Simpangan baku lebih tepat digunakan ketika data menyebar secara simetris atau tidak menceng. Pada Microsoft Excel, nilai simpangan baku d dapat diperoleh dengan perintah **=STDEV()**. Di dalam tanda kurung diisikan kolom yang akan dihitung simpangan bakunya.

E2. Tabel

Tabel adalah kumpulan data yang disusun secara sistematis menurut baris dan kolom. Tabel banyak digunakan untuk menyajikan informasi statistik. Tabel mampu memberikan informasi yang lebih lengkap dan mudah dimengerti oleh pengguna data.

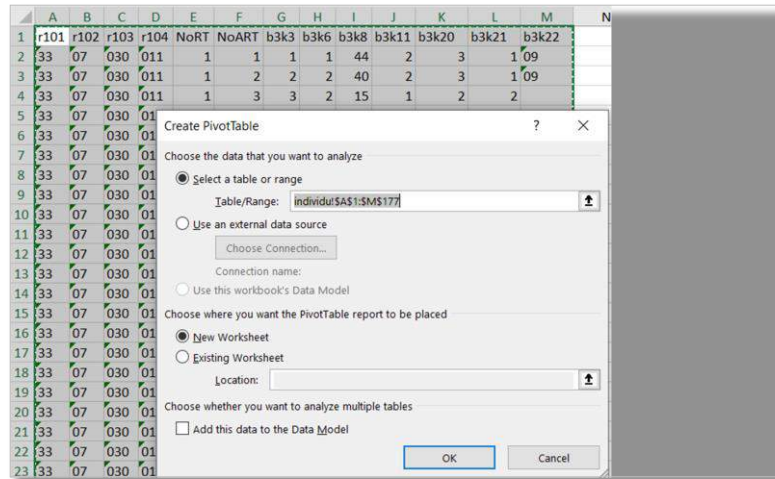
Microsoft Excel menyediakan fasilitas untuk membuat tabel dari raw data yang dikenal dengan Pivot Table. Pivot Table adalah sebuah tabel interaktif yang dalam waktu singkat mampu menampilkan ringkasan data dalam jumlah yang besar. Pivot Table bukan hanya memutar sumbu tabel, tetapi mampu melakukan kalkulasi setiap item yang dibutuhkan menggunakan cara dan perhitungan sesuai pilihan kita.

Untuk mengaktifkan pivot tabel dapat menggunakan menu Insert>Pivot Tabel.

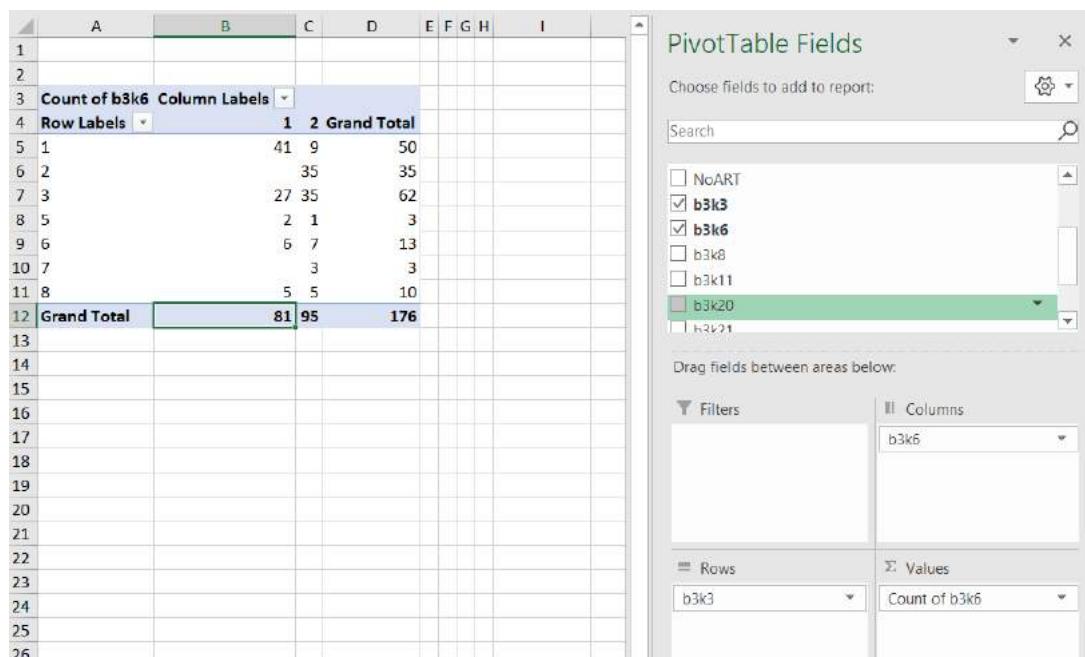


Langkah-langkah menggunakan fasilitas ini :

Blok range data yang akan dibuat tabel pivotnya. Pilih menu Insert > Pivot Table > PivotTable. Kemudian pilih hasil tabel pivot akan ditempatkan.



Kemudian pilih field yang akan ditempatkan pada Columns Labels, Row Labels dan Values.



Hasil pengaturan dari tabel di atas setelah kodediganti dengan keterangannya adalah sebagai berikut

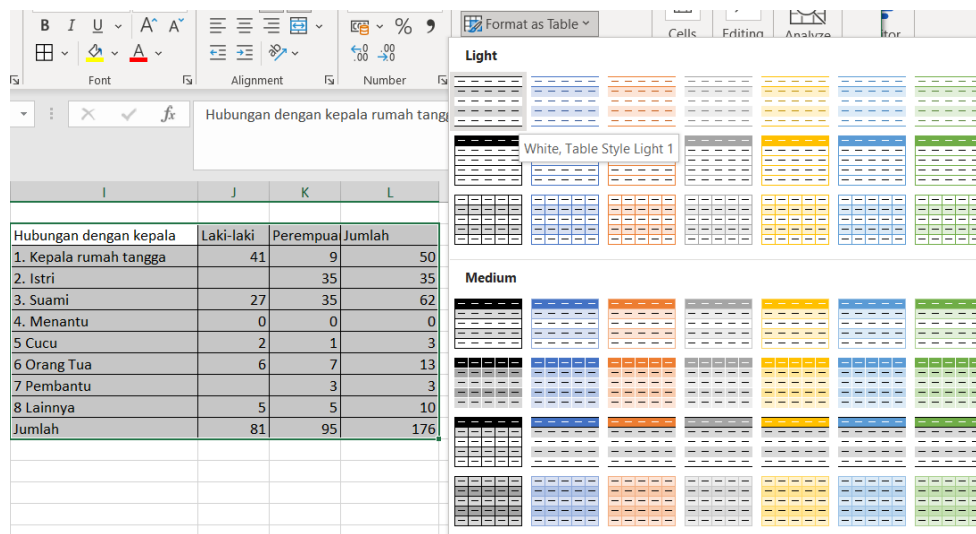
Hubungan dengan kepala rumah tangga	Laki-laki	Perempuan	Jumlah
1. Kepala rumah tangga	41	9	50
2. Istri		35	35
3. Suami	27	35	62
4. Menantu	0	0	0
5 Cucu	2	1	3
6 Orang Tua	6	7	13
7 Pembantu		3	3
8 Lainnya	5	5	10
Jumlah	81	95	176

Microsoft excel telah menyediakan berbagai format tabel yang dikenal dengan Autoformat. Dengan fasilitas ini, kita tinggal memilih dan memanfaatkannya saja. Berikut cara menggunakan fasilitas ini :

1. Blok range data yang akan dibuat tabelnya. Pilih menu Format As Table

Hubungan dengan kepala	Laki-laki	Perempuan	Jumlah
1. Kepala rumah tangga	41	9	50
2. Istri		35	35
3. Suami	27	35	62
4. Menantu	0	0	0
5 Cucu	2	1	3
6 Orang Tua	6	7	13
7 Pembantu		3	3
8 Lainnya	5	5	10
Jumlah	81	95	176

- Klik salah satu model tabel yang diinginkan



- Klik OK untuk menutup jendela ini. Maka *range* yang disorot tadi akan diubah sesuai dengan tabel yang anda pilih.

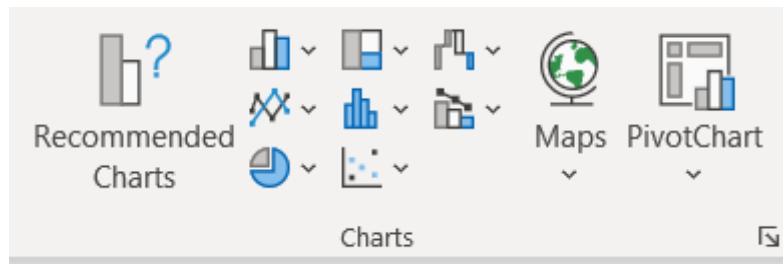
Hubungan dengan kepala rumah tangga	Laki-laki	Perempuan	Jumlah
1. Kepala rumah tangga	41	9	50
2. Istri		35	35
3. Suami	27	35	62
4. Menantu	0	0	0
5 Cucu	2	1	3
6 Orang Tua	6	7	13
7 Pembantu		3	3
8 Lainnya	5	5	10
Jumlah	81	95	176


E3. Grafik

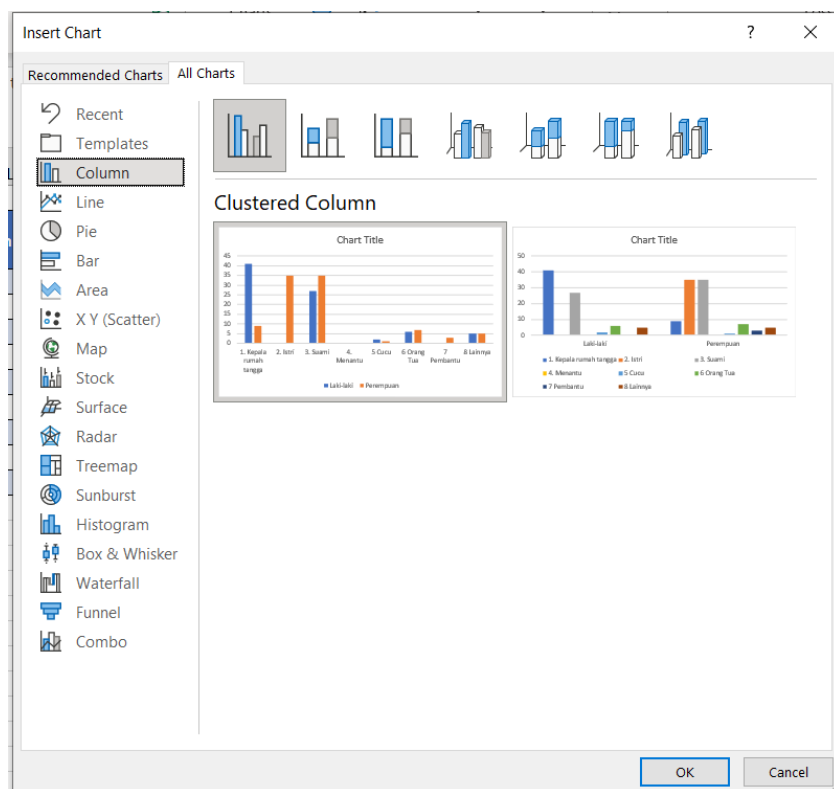
Grafik merupakan bentuk penyajian lain dari data sehingga akan mudah dimengerti ketika seseorang menganalisa suatu data. Selain dalam bentuk tabel, laporan juga dapat disajikan dalam bentuk grafik. Grafik pada Excel dapat disajikan dalam banyak bentuk dengan sudut elevasi dan perputaran yang kita dapat atur sesuai kebutuhan. Disamping itu, langkah-langkah di dalam pembuatan grafik sudah disediakan.

a. Pembuatan Grafik

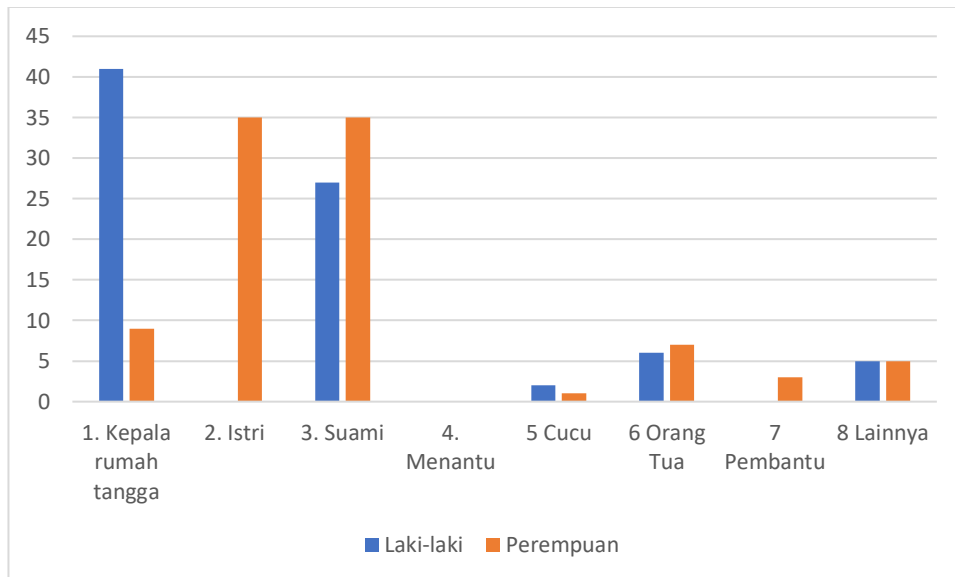
Salah satu fungsi unggul dalam Microsoft Excel adalah kita dapat melihat hasil tabel diubah menjadi ke dalam grafik dengan cepat. Microsoft Excel menyediakan berbagai macam bentuk grafik yang mencakupi Line, XY, Column, Bar, Batang, Area, Stock, dan sebagainya. Grafik dapat dilihat dalam menu INSERT sebagai berikut.



Blok tabel yang akan dibuat grafik. Setelah itu, klik tombol , maka akan muncul menu sebagai berikut:

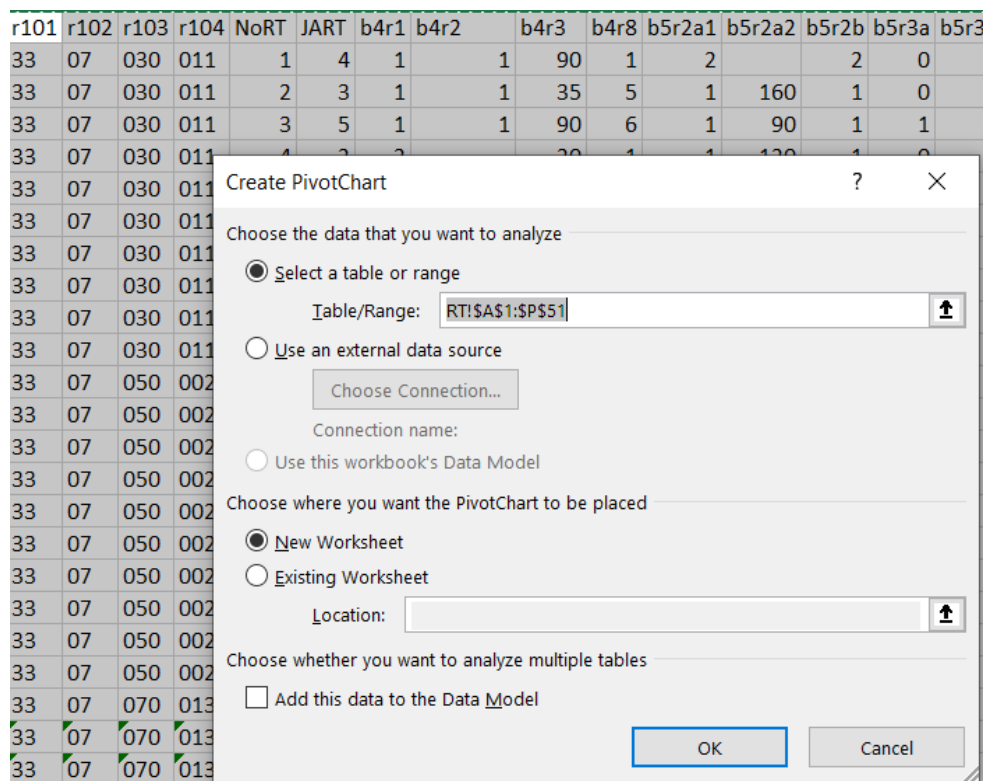


Setelah masuk ke Insert Chart, maka silakan pilih jenis grafik yang anda inginkan sesuai selera anda. Jika sudah terpilih jenis Chart yang anda inginkan, silakan klik OK sehingga didapatkan grafik sebagai berikut.

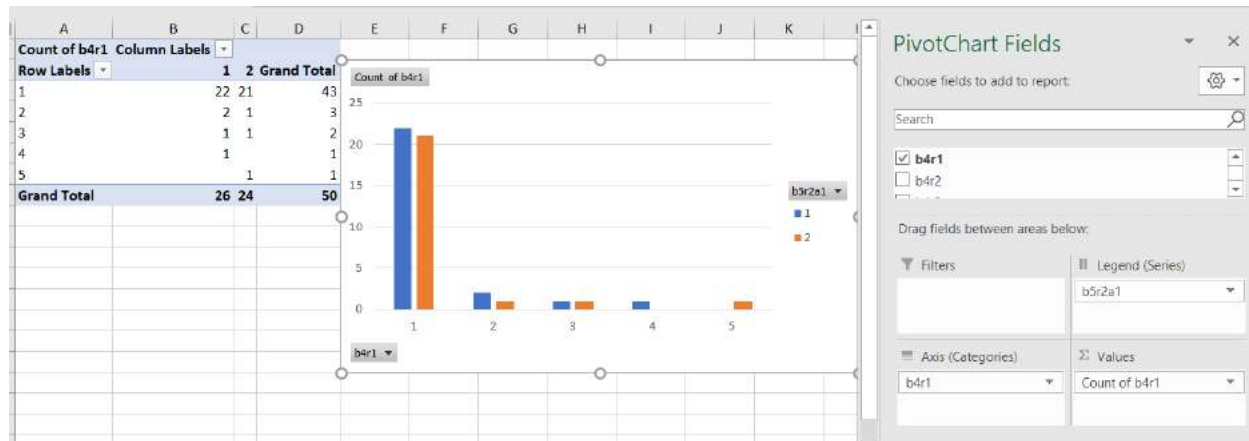


b. Pembuatan Pivot Grafik

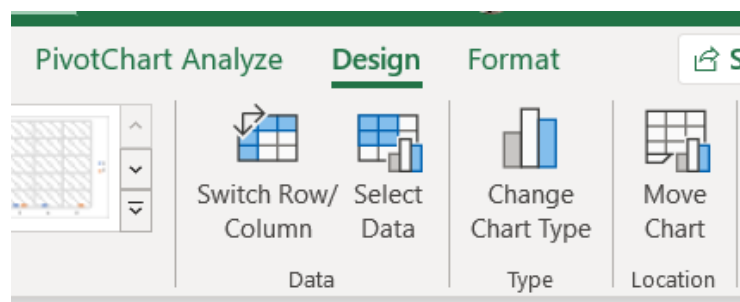
Pembuatan grafik di Microsoft Excel dapat langsung dilakukan melalui Pivot Chart. Untuk membuatnya, blok *range* data yang akan dibuat grafik pivotnya. Pilih menu Insert > Chart> PivotChart. Kemudian pilih hasil grafik pivot akan ditempatkan.



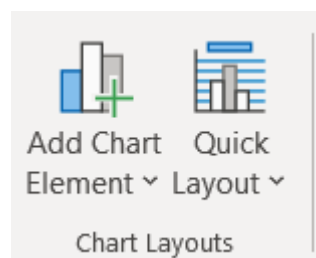
Kemudian pilih field yang akan ditempatkan pada Legends Fields (Series), Axis Fields (Categories) dan Values.



Kita bisa mengubah type grafik dengan klik bagian grafik yang mau diformat kemudian klik Change Chart Type pada tab Design.



Untuk mengatur layout untuk setiap bagian grafik misalnya Title, Kategori Axis, atau plot area, klik bagian grafik yang mau diformat, kemudian pada Design, pilih ab Add Chart Element



E5. Pengecekan Pencilan

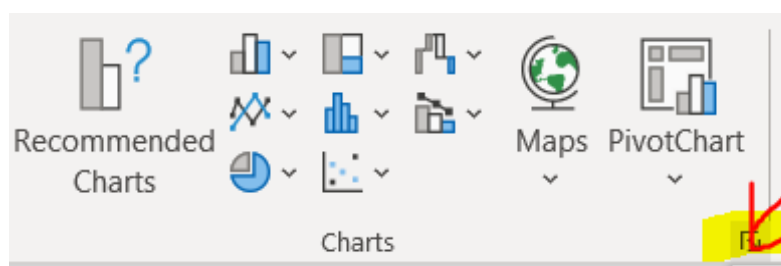
Pemeriksaan pencilan dapat dilakukan melalui *Box plot* (diagram kotak). Diagram ini dapat menggambarkan distribusi data dan memuat lima ringkasan data secara berurutan, yaitu median, Q_1 (kuartil 1), Q_3 (kuartil 3), jarak antar kuartil (IQR), nilai minimum, dan nilai maksimum. Informasi tersebut menunjukkan kemencengan data berdasarkan nilai kuartil, keruncingan data berdasarkan nilai IQR, dan adanya data pencilan (*outlier*). Pencilan dihitung berdasarkan nilai batas bawah dan batas atas. Kedua nilai tersebut digunakan sebagai penentu nilai yang berada dalam toleransi (atau tidak dikategorikan pencilan (*outlier*), yaitu $Q_1 - (1,5 \times IQR)$ untuk batas bawah, dan $Q_3 + (1,5 \times IQR)$ untuk batas atas. Nilai data pengamatan yang berada di luar rentang nilai tersebut disebut pencilan.

Pembuatan Box Plot dengan Microsoft Excel dapat dilakukan dengan langkah-langkah sebagai berikut.

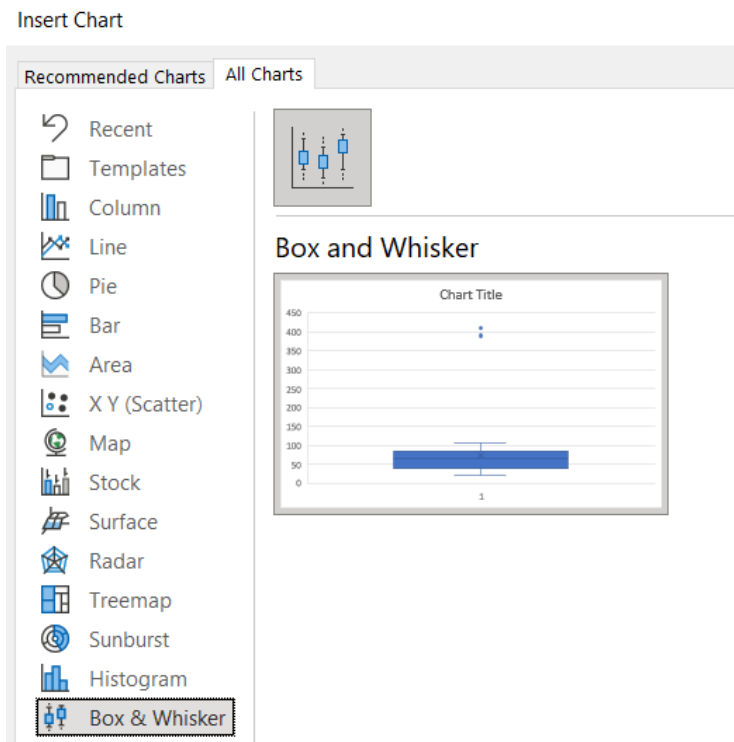
1. Blok kolom data yang akan dibuat grafik

H	I	J	K
b4r2	b4r3	b4r8	b5r2a1
1	90	01	2
1	35	05	1
1	90	06	1
	2000	01	1
1	85	02	1
1	80	04	1
1	80	06	2
1	30	03	1
2	30	01	2
1	95	06	1
1	90	04	2
	110	06	1
1	55	03	1
1	35	04	1

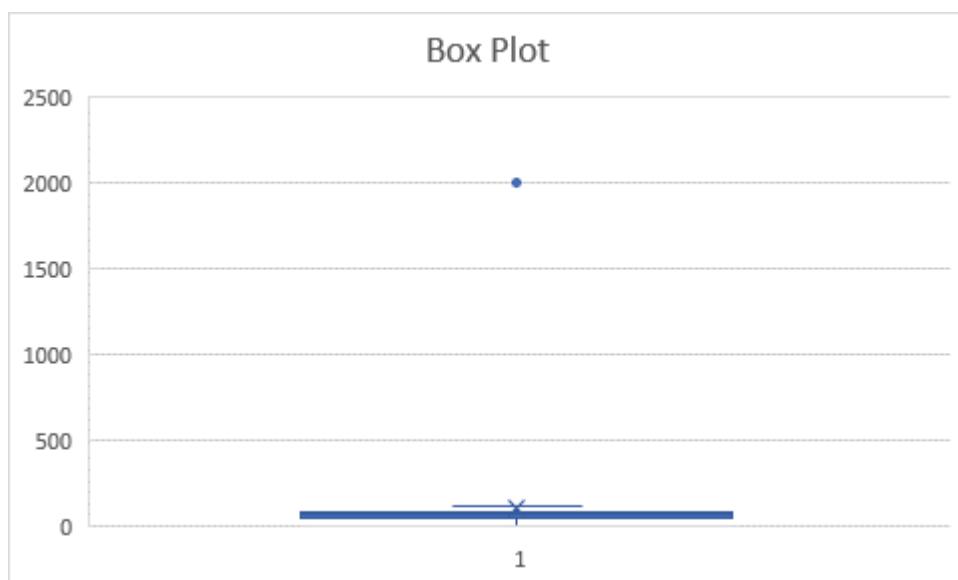
2. Klik Insert, lalu klik 



3. Pilih All Chart dan klik Box and Whisker, lalu klik OK.



Berikut adalah diagram kotak yang dihasilkan. Dari grafik ini nampak terdapat satu pengamatan yang menjadi pencilan.





Pusat Pendidikan dan Pelatihan
Badan Pusat Statistik



Jalan Raya Jagakarsa No.70, Jakarta Selatan, 12620



www.pusdiklat.bps.go.id



pusdiklat@bps.go.id



(021) 7873781 - 83



(021) 7873955

