



LABORATORIUM
PROGRAM STUDI ILMU KESEHATAN MASYARAKAT
FAKULTAS KESEHATAN MASYARAKAT
UNIVERSITAS AHMAD DAHLAN

FM-UAD-PBM-11-04/R0



Aplikasi Statistik Kehidupan

PANDUAN PRAKTIKUM

**BUKU PANDUAN PRAKTIKUM
APLIKASI STATISTIK KEHIDUPAN**

PP/FKM/ASK/VI/R3



Penyusun:

**Dedik Sulistiawan, S.KM., MPH.
Erni Gustina, S.KM., MPH.
Yuniar Wardani, S.KM., MPH.**

**LABORATORIUM FAKULTAS KESEHATAN MASYARAKAT
PROGRAM STUDI ILMU KESEHATAN MASYARAKAT
FAKULTAS KESEHATAN MASYARAKAT
UNIVERSITAS AHMAD DAHLAN**

2021

SEJARAH REVISI

Nama petunjuk praktikum : Aplikasi Statistik Kehidupan
 Semester : VI
 Program Studi : Kesehatan Masyarakat
 Fakultas : Kesehatan Masyarakat

Revisi ke	Tanggal	Uraian Revisi
1	20 Februari 2018	Terdapat perubahan materi yang dilakukan pada buku petunjuk praktikum yaitu dengan mengganti materi praktikum yang ada pada buku petunjuk praktikum edisi pertama. Perubahan materi yang dilakukan yaitu: Materi praktikum <i>Structural Equational Model</i> (SEM) digantikan Materi uji rerata 2 Mean dan uji rerata >2 Mean
2	15 Maret 2019	Pada buku praktikum edisi revisi ke-2 terdapat perubahan materi. Perubahan materi terletak pada analisis deskriptif, analisis bivariat, dan analisis multivariat data kategorikal.
3	16 Maret 2020	Buku petunjuk praktikum edisi revisi ke-3 mengalami penyesuaian materi. Terdapat pemadatan topik praktikum dan penambahan materi tentang model regresi Poisson serta reliabilitas dan validitas pengukuran yang menggantikan materi analisis survival.

KATA PENGANTAR

Aplikasi Statistik Kehidupan merupakan salah satu mata kuliah peminatan kesehatan reproduksi yang memiliki bobot 3 SKS (1 teori; 2 praktikum). Praktikum aplikasi statistik kehidupan terintegrasi dengan pemaparan teori dalam mata kuliah. Buku panduan praktikum ini disusun sebagai pedoman dalam penyelenggaraan praktikum bagi dosen dan mahasiswa.

Buku panduan praktikum ini berisi 7 acara kegiatan praktikum yaitu proses komputasi statistik menggunakan perangkat lunak STATA, analisis data deskriptif tabulasi silang dan ukuran asosiasi, analisis data kategorik multivariabel, model regresi Poisson untuk data diskrit, reliabilitas dan validitas pengukuran, dan penyajian hasil analisis data.

Tim penyusun berharap buku panduan praktikum ini dapat membantu mahasiswa dalam proses belajar, memperdalam keterampilan mahasiswa sesuai dengan kompetensi dalam mengelola data sehingga dapat meningkatkan kualitas mahasiswa dalam proses analisis.

Yogyakarta, Mei 2021

Tim Penyusun

TATA TERTIB PRAKTIKUM APLIKASI STATISTIK KEHIDUPAN

1. Praktikan wajib menyelesaikan administrasi praktikum.
2. Praktikan wajib berpakaian rapi selama praktikum.
3. Praktikan diwajibkan membawa buku petunjuk praktikum dan alat tulis secara mandiri.
4. Tidak boleh membawa makanan atau minuman ketika praktik di dalam laboratorium.
5. Praktikan wajib mengikuti *pretest* dan *posttest* sebagai salah satu komponen penilaian dalam praktikum (ketentuan *pretest* dan *posttest* diatur oleh koordinator praktikum).
6. Praktikan wajib mengikuti seluruh kegiatan praktikum, apabila tidak harus mengikuti inhal harus sesuai ketentuan yang berlaku
7. Praktikan wajib mengikuti responsi.
8. Aturan-aturan lain yang dianggap perlu, dapat disampaikan selanjutnya sebagai kesepakatan bersama.

DAFTAR ISI

HALAMAN JUDUL	i
SEJARAH REVISI	ii
KATA PENGANTAR	iii
TATA TERTIB PRAKTIKUM APLIKASI STATISTIK KEHIDUPAN	iv
DAFTAR ISI	v
PRAKTIKUM KE-1 PENGANTAR KOMPUTASI STATISTIK DENGAN STATA	1
PRAKTIKUM KE-2 ANALISIS DATA DESKRIPTIF	20
PRAKTIKUM KE-3 TABULASI SILANG DAN UKURAN ASOSIASI	24
PRAKTIKUM KE-4 ANALISIS DATA KATEGORIK MULTIVARIABEL	30
PRAKTIKUM KE-5 MODEL REGRESI POISSON UNTUK DATA DISKRIT	39
PRAKTIKUM KE-6 RELIABILITAS DAN VALIDITAS PENGUKURAN	46
PRAKTIKUM KE-7 PENYAJIAN HASIL ANALISIS DATA	56

PRAKTIKUM KE-1

PENGANTAR KOMPUTASI STATISTIK DENGAN PROGRAM STATA


A. Tujuan

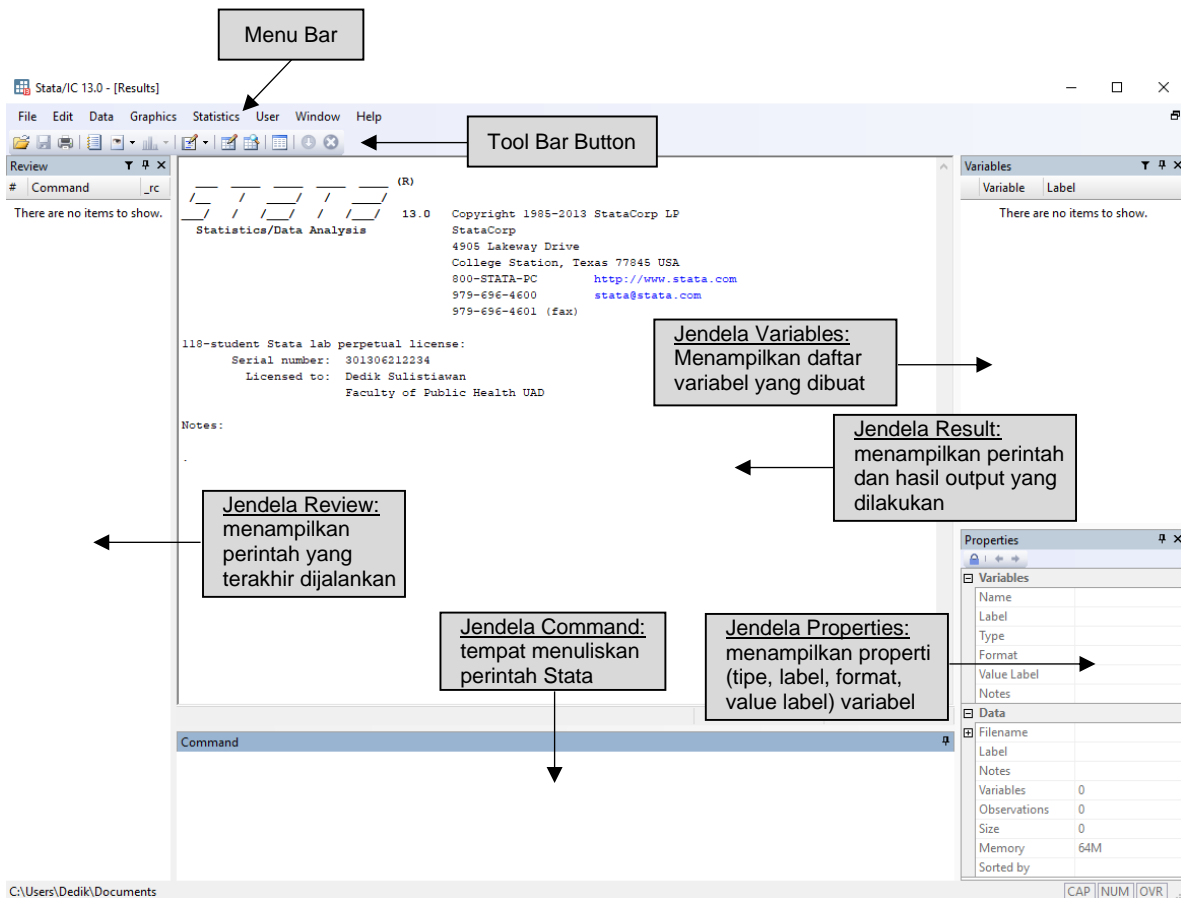
1. Mahasiswa mampu menjelaskan struktur dan fungsi komponen/ menu program Stata.
2. Mahasiswa mampu melakukan *entry* data menggunakan program Stata.
3. Mahasiswa mampu melakukan manajemen variabel menggunakan program Stata.

B. Pengantar

Stata merupakan perangkat lunak statistik yang dikembangkan oleh StataCorp, dirilis pertama kali pada tahun 1985 dengan versi 1.0-nya. Hingga saat ini Stata telah mencapai generasi ke 15.0 yang dirilis pada tahun 2017. Stata digunakan oleh banyak perusahaan dan lembaga akademis di seluruh dunia, termasuk di bidang kesehatan. Stata memiliki kemampuan yang lengkap meliputi manajemen data, analisis statistik, grafik, pemrograman, dan fitur pelaporan.

Di beberapa kalangan akademisi dan peneliti, Stata lebih digemari karena kemampuannya dalam menganalisis data yang jumlahnya sangat banyak, seperti *Indonesian Family Life Survey* (IFLS), Survei Sosial Ekonomi Nasional (SUSENAS), *Health and Demographic Surveilans System* (HDSS) Kabupaten Sleman, dan sebagainya. Pengoperasian Stata lebih direkomendasikan menggunakan bahasa pemrograman atau *syntax* (disebut juga *command*) yang dianggap lebih praktis, dibandingkan dengan menggunakan model *point and click* (*pull down menu*) seperti yang dilakukan dengan perangkat statistik lain seperti SPSS.

Hal yang diperlukan sebelum memulai praktikum ini adalah komputer telah ter-*install* program Stata versi 13.0 (atau yang lebih tinggi). Dataset yang akan digunakan dalam praktikum ini adalah data penilaian profesionalitas bidan dalam melakukan *antenatal care* (ANC) dalam format Microsoft Excel. Setelah semuanya siap, bukalah program Stata dengan melakukan klik icon Stata. 



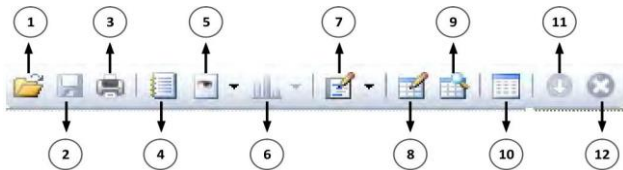
Keterangan:

Menu Bar merupakan daftar menu (*pull down menu*) untuk dapat memilih apa yang akan dilakukan di Stata.

1. **File:** membuka file baik file data, do-file maupun log-file, menyimpan data, membuka jendela penampil (viewer), menjalankan do-file, memberi/ mengubah nama file, membuat/ menghentikan/ menambah log-file, melakukan import data, melakukan export data, mencetak file (print).
2. **Edit:** melakukan copy, paste, melakukan pencarian (find), mengubah tampilan Stata (preferences).
3. **Data:** mendeskripsikan data, memunculkan dataset, mengedit dataset, memunculkan variabel manager, mengubah utilitas data.
4. **Graphics:** menu untuk membuat grafik.
5. **Statistics:** menu untuk melakukan analisis deksriptif dan analitik.
6. **User:** menampilkan data, grafik dan statistik.

7. **Window:** menampilkan jendela-jendela kerja.
8. **Help:** mencari bantuan.

Toolbar berfungsi untuk menampilkan ikon jalan pintas untuk memilih menu yang sering digunakan.



1. **Open:** membuka file, baik file data, do-file maupun log-file
2. **Save:** menyimpan file data
3. **Print:** mencetak output yang terdapat di jendela hasil
4. **Log-file:** membuat, memulai, menghentikan, dan menambah log-file
5. **Viewer:** jendela tampilan, menampilkan log-file dan bantuan (help)
6. **Graph:** untuk menampilkan grafik, akan aktif bila kita telah membuat grafik di Stata.
7. **Do-file Editor:** untuk membuat, dan melakukan editing do-file
8. **Data Editor:** untuk menampilkan data, tanpa bisa melakukan editing data
9. **Data Browser:** untuk menampilkan data, ste melakukan editing data
10. **Variable Manager:** untuk menampilkan jendela variabel manager, berisi daftar semua variabel dan sterisk dari setiap variabel yang ada (nama, tipe, format, label, dll).
11. **More:** untuk melanjutkan tampilan output di jendela hasil.
12. **Break:** untuk memotong tampilan output di jendela hasil.

C. Perintah Stata

Terdapat dua cara dalam menjalankan perintah stata, yaitu dengan **syntax** dan menu **pull down**. Cara yang akan kita bahas dalam modul ini sebagian besar menggunakan syntax.

1. Syntax

Syntax merupakan perintah/command yang dituliskan pada jendela command. Cara ini akan lebih banyak kita gunakan dalam mengolah dan menganalisis data di Stata. Karena cara ini lebih mudah digunakan jika kita sudah terbiasa menggunakan Stata. Anda tidak perlu menghafalkan syntax, secara otomatis jika Anda sering menggunakan Stata Anda akan hafal dengan sendirinya, dan bila lupa Anda bisa mencarinya melalui help, yang penting adalah paham format logika syntax.

Format syntax stata secara umum adalah:

(command) [namavariabel] [if kondisi] [in] [,] [opsi]
--

- Command adalah input perintah yang harus ada pada setiap kali memberikan perintah pada Stata, perintah pada stata selalu menggunakan huruf kecil (*lowercase*), pada perintah tertentu, perintah dapat disingkat dengan hanya memasukan 3-digit pertama dari perintah, misalnya perintah `describe` yang dapat disingkat hanya memasukan perintah `des`, tetapi perlu diingat bahwa tidak semua perintah pada Stata dapat disingkat.
- Namavariabel adalah nilai yang diberikan untuk perintah yang dapat berupa variable dari dataset yang aktif, variable lainnya yang bertipe string atau numerik.
- “If kondisi” adalah formula atau rumus yang dapat dikostumisasi sesuai kebutuhan yang memberikan batasan pada perintah yang akan dijalankan, “if kondisi” pada semua command bersifat opsional dan tidak semua command dapat ditambahkan “if kondisi”.
- “in” adalah kondisi yang mengacu pada observasi, misalnya “in 1” artinya pada observasi 1, “in 1/5” mengacu pada observasi nomor 1 sampai nomor 5.
- Opsi pada perintah stata selalu berada setelah tanda koma “,”, sama seperti command, beberapa opsi pada beberapa command juga dapat disingkat.

2. Pull Down Menu

Cara pertama adalah memilih melalui menu bar (daftar menu) atau toolbar (jika pilihan menu ada di dalam toolbar). Cara ini banyak digunakan apabila kita adalah pengguna baru Stata dan belum mengenal cara melakukan perintah dengan DOS.

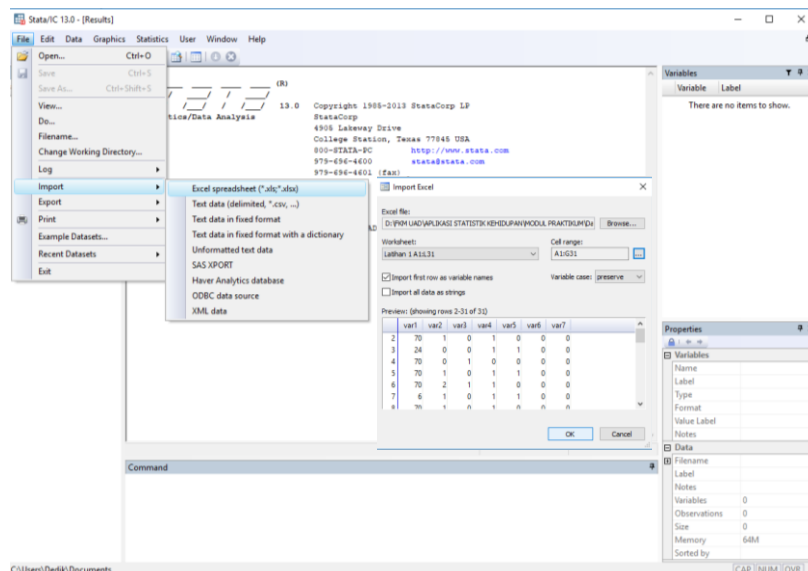
D. Memulai Stata

Pada modul ini, kita akan menggunakan data yang sudah disiapkan untuk diolah dan dianalisis. Dataset yang akan digunakan dalam praktikum ini adalah data penilaian profesionalitas bidan dalam melakukan *antenatal care* (ANC) dalam format Microsoft Excel.

1. Membuka File Data Dari Ms. Excel

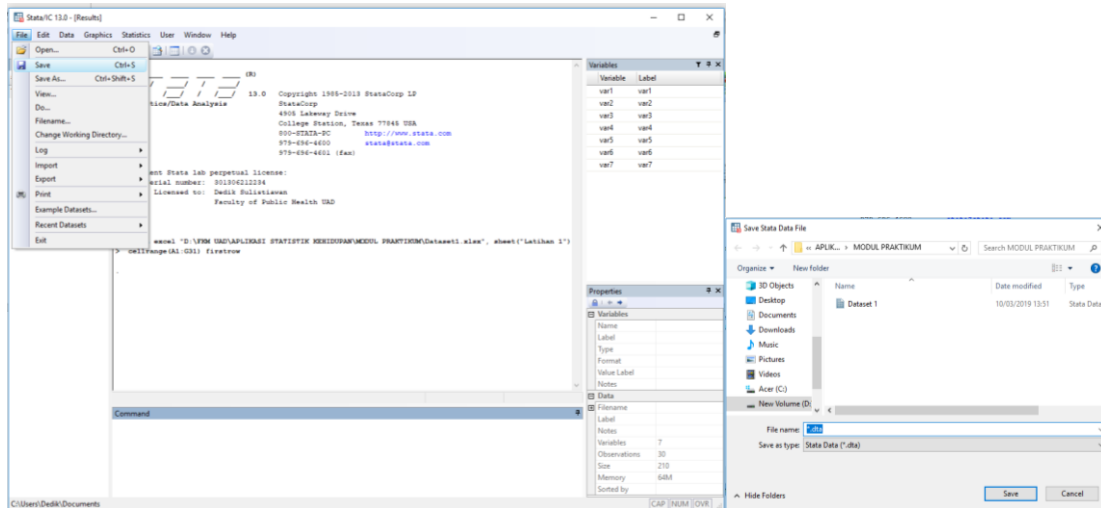
Pada modul ini akan dikenalkan cara membuka dataset yang tersimpan dalam Ms. Excel.

Caranya: File → Import → Excel Spreadsheet → Browse → Pilih direktori tempat menyimpan file → Pilih nama file → Centang Import first row as variable names → OK.



Setelah terbuka, dan ingin memastikan data telah terimport silahkan ketik `browse` pada jendela command, setelah itu tutup kembali jendelanya.

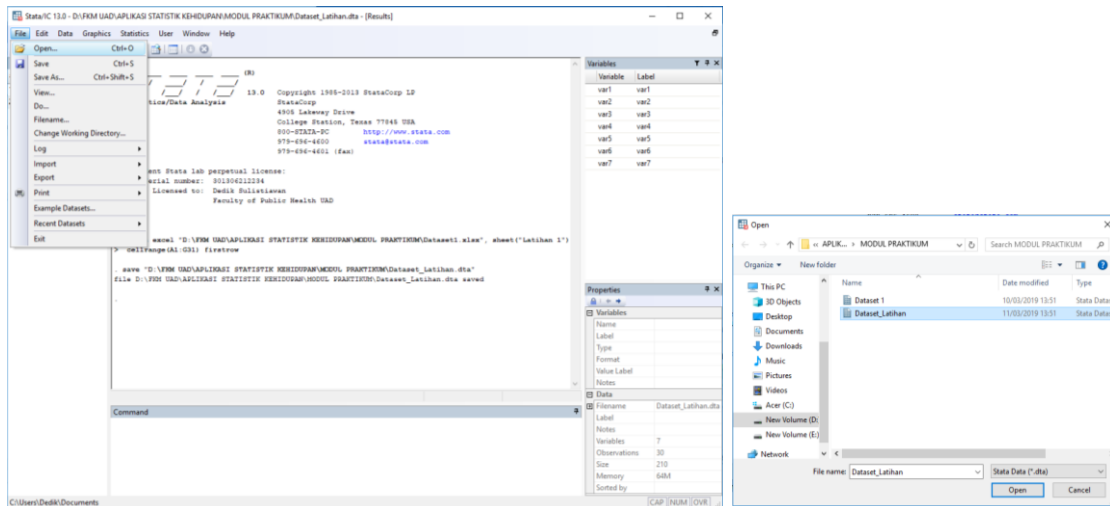
2. Cara Menyimpan Data Hasil Import ke Stata File → Save → Tulis nama file → Save



3. Cara Membuka File Stata

Cara 1 (Pull Down Menu)

File → Open → Pilih folder tempat menyimpan file → Pilih file (Contoh: Dataset_Latihan.dta) → Open



Cara 2 (Syntax)

use "alamat file dan nama file yang akan dibuka", clear

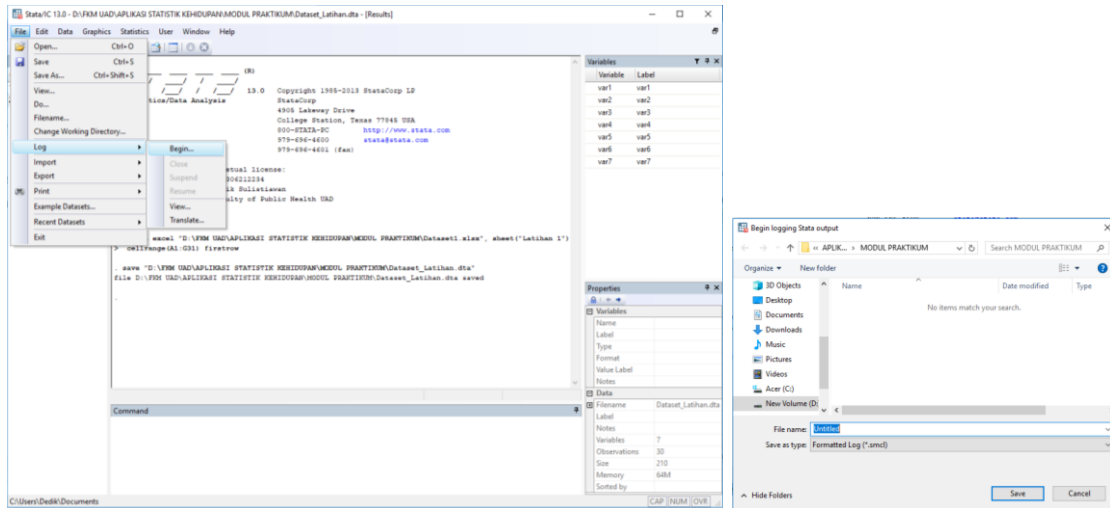
Contoh: use "D:\FKM UAD\APLIKASI STATISTIK KEHIDUPAN\MODUL PRAKTIKUM\Dataset_Latihan.dta", clear

4. Membuat Log File Stata

Sebelum memulai mengolah dan menganalisis data, kita akan memerintahkan Stata untuk merekam seluruh proses pekerjaan kita, baik output maupun perintah (command) selama kita menganalisis data.

Cara 1 (Pull Down Menu)

File → Log → Begin → Cari folder tempat menyimpan → Tulis nama file → Save



Cara 2 (Syntax)

Log using “alamat file dan nama file yang akan dibuka”

Contoh: `log using "D:\FKM UAD\APLIKASI STATISTIK KEHIDUPAN\MODUL PRAKTIKUM\Dataset_Latihan.smcl"`

Beberapa perintah log lainnya:

`log off` : memerintahkan rekaman sementara

`log on` : memulai lagi rekaman

`log close`: menghentikan rekaman keseluruhan

`log using namafile, append`: menambah rekaman baru di rekaman lama pada satu file log

`log using namafile, replace`: mengganti rekaman lama dengan rekaman pada suatu file data

E. Melihat Data

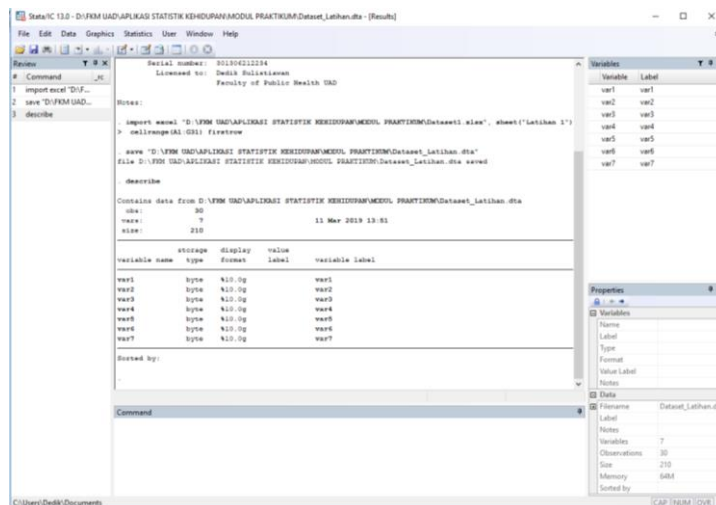
1. Melihat Deskripsi Dataset

Setelah kita siap dengan direktori tempat kerja, data telah masuk ke program Stata dan rekaman telah disiapkan, selanjutnya kita akan melihat deskripsi data set yang akan digunakan. Kegunaan dari proses ini adalah untuk mengetahui informasi lebih detail tentang data yang kita miliki (berapa jumlah observasi yang ada, berapa jumlah variabel, apakah berbeda variabel jumlah variabel yang ada di kuesioner dengan data, adakah *missing data*, dst). Perintah yang dituliskan adalah:

`describe` : deskripsi semua variabel

`describe namavariabel` : deskripsi variabel yang dipilih

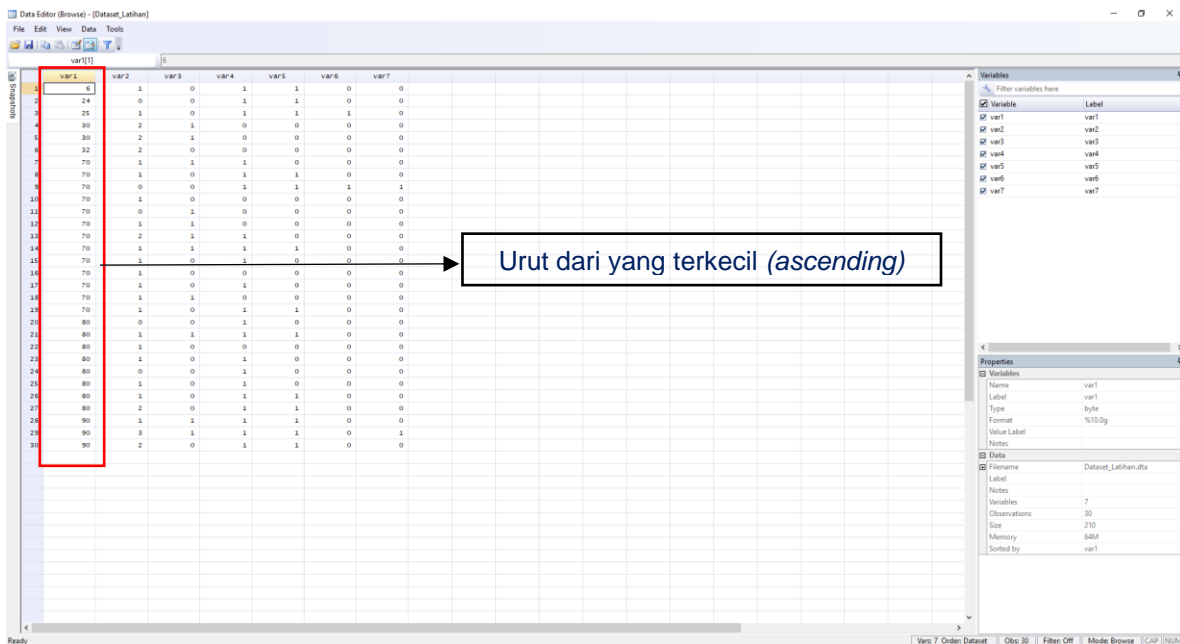
Contoh: `describe`



2. Mengurutkan Data

Apabila kita ingin data yang dimiliki tersusun secara berurutan berdasarkan nilai dari variabel, maka kita dapat menggunakan perintah: `sort`. Susunan dapat diurutkan secara alfabetis A – Z atau berdasarkan urutan angka dari terkecil – tertinggi tergantung dari jenis variabel yang dipilih.

Contoh: `sort var1`



3. Melihat Ringkasan Data per Variabel dengan Codebook

Setelah melihat deskripsi keseluruhan, lalu kita akan melihat deskripsi data per variabel sehingga akan diperoleh ringkasan data per variabel. Proses ini akan menghasilkan informasi lebih detail tentang data dalam 1 variabel. Informasi yang ada meliputi nama variabel, label variabel, tipe data, range data, nilai unik, data missing, frekuensi data dan value label.

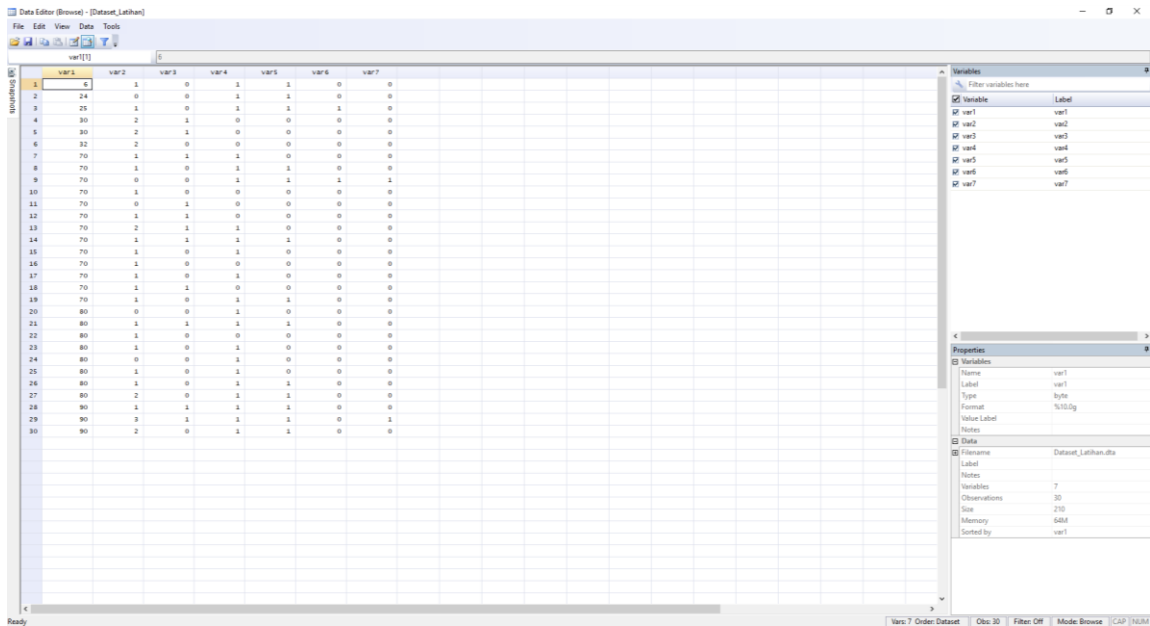
Perintah yang dituliskan:

codebook : melihat informasi semua variabel per variabel

codebook namavariabel : melihat informasi variabel yang dipilih

Contoh: codebook var1

```
var1                                                                    var1
-----                                                                    -----
                                                                    type: numeric (byte)
                                                                    range: [6,90]
                                                                    unique values: 8
                                                                    units: 1
                                                                    missing .: 0/30
                                                                    tabulation: Freq. Value
                                                                    1 6
                                                                    1 24
                                                                    1 25
                                                                    2 30
                                                                    1 32
                                                                    13 70
                                                                    8 80
                                                                    3 90
```

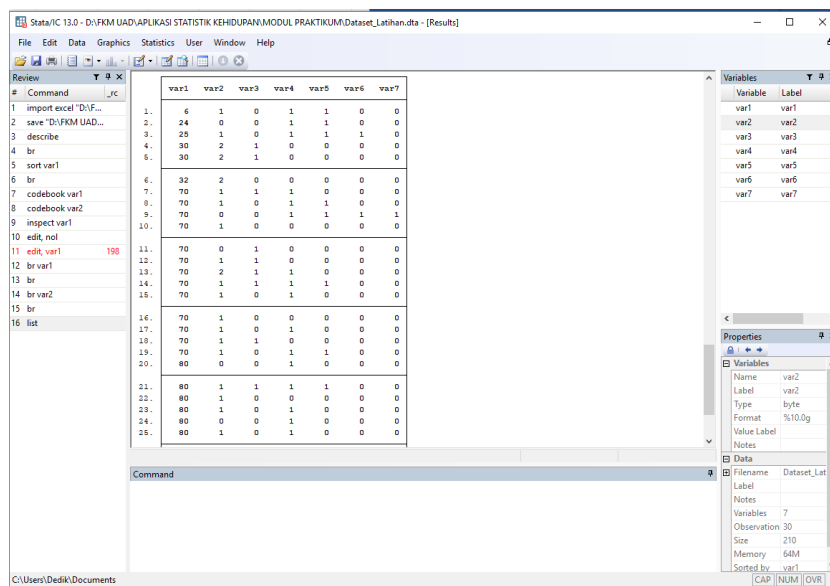
6. Melihat Data per Observasi

Jika `browse` dapat melihat data set sebagian atau keseluruhan, maka kali ini kita akan melihat data satu per satu setiap observasi dengan cara menuliskan syntax:

`list` : melihat data per observasi untuk semua observasi

`list namavariabel` : melihat daftar data per observasi untuk variabel terpilih

Contoh: `list`



7. Meringkas Data Numerik

Data numerik yang berskala rasio dan interval akan diringkas agar kita bisa mengetahui pemusatan (mean, median, modus) dan penyebaran data kita (standar deviasi, variansi).

Syntax : `sum namavariabel`
`tabstat namavariabel`

Contoh: meringkas variabel var1

```
. sum var1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
var1	30	65.56667	22.19923	6	90

Interpretasi:

Rata-rata (mean) variabel 1 (lama praktik bidan) adalah 65,5 bulan di mana rentangnya berkisar antara 6 hingga 90 bulan.

8. Meringkas Data Kategorik

Syntax:

`tab namavariabel` : meringkas satu variabel
`tab namavarA namavarB` : meringkas dua variabel menjadi tabel 2x2
`tab namavarA namavarB, row` : meringkas dua variabel menjadi tabel 2x2 dan memunculkan persen dalam row
`tab namavarA namavarB, col` : meringkas dua variabel menjadi tabel 2x2 dan memunculkan persen dalam kolom

Contoh: `tab var3`

```
. tab var3
```

var3	Freq.	Percent	Cum.
0	19	63.33	63.33
1	11	36.67	100.00
Total	30	100.00	

F. Formating Data

Pada tahap ini, kita akan membuat nama variable dan value label/ kategori variable dari data yang kita import dari Ms. Excel. Berikut adalah keterangan variabel.

Nama Variabel	Nama Baru Variabel	Label	Kode Variabel
var1	lama_praktek	Lama praktik bidan (bulan)	
var2	pendidikan	Pendidikan terakhir bidan	0 = SLTA/Sederajat 1 = D1/D2 2 = D3 3 = S1/S2/S3
var3	delima	Status bidan delima	0 = Bukan Bidan Delima 1 = Bidan Delima
var4	ukur_bb	Melakukan pengukuran berat badan ibu hamil	0 = Tidak 1 = Ya
var5	ukur_tb	Melakukan pengukuran tinggi badan ibu hamil	0 = Tidak 1 = Ya
var6	ukur_lila	Melakukan pengukuran lingkaran lengan atas ibu hamil	0 = Tidak 1 = Ya
var7	ukur_djb	Melakukan pengukuran denyut jantung bayi	0 = Tidak 1 = Ya

1. Mengubah Nama Variabel

Syntax: `rename variabel_lama variabel_baru`

Contoh: kita ingin mengubah nama var1 menjadi lama_praktek, syntax-nya:
`rename var1 lama_praktek`

Maka pada jendela variabel var1 telah berubah menjadi lama_praktek

Catatan: Nama variable tidak boleh mengandung spasi.

2. Memberi Label Nama Variabel

Syntax: `lab var namavariabel "label variabel"`

Contoh: kita akan memberi label variabel lama_praktek menjadi Lama praktik bidan (bulan), syntax-nya: `lab var lama_praktek "Lama praktik bidan (bulan) "`

3. Memberi Value Label

Syntax: `label define namavaluelabel nilai "label" nilai "label"`
`label value namavariabel namavaluelabel`

Contoh: kita akan memberikan value label variabel delima menjadi

0 = Bukan Bidan Delima

1 = Bidan Delima

Syntax:

```
lab define kode_delima 0 "Bukan Bidan Delima" 1 "Bidan Delima"
lab value delima kode_delima
```

Sebagai latihan, berilah nama variable, value label sesuai keterangan tabel diatas.

G. Transformasi Data

Transformasi data adalah teknik untuk mengubah atau mengelompokkan data, biasanya data yang ditransformasi adalah data numerik menjadi data kategori. Transformasi yang sering dilakukan adalah komputasi dan recode atau mengubah kode. Dalam modul ini kita akan mempraktikkan recode, sementara komputasi dibahas dalam pertemuan yang lain. Aturan dasar recode:

Aturan	Contoh	Arti
# = #	3 = 1	3 diubah menjadi 1
# # = #	2 . = 9	2 dan . diubah menjadi 9
##/## = #	1/5 = 4	1 sampai 5 diubah menjadi 4
nonmissing = #	nonmiss = 8	Semua yg tidak hilang diubah menjadi 8
missing = #	miss = 9	Semua yang hilang diubah menjadi 9

Contoh: Pada data latihan, terdapat varibel dengan lama_praktek yang berisi data lama praktik bidan dalam bulan yang disajikan dalam skala data rasio, kemudian kita ingin melihat berapa orang bidan yang mempunyai masa praktek kurang dari 1 tahun, 1 sampai dengan 5 tahun, dan lebih dari 5 tahun. Maka kita akan lakukan transformasi atau perubahan agar data kita terkelompok seperti yang kita inginkan.

Bila data kecil kita kelompokkan manual akan mudah, tetapi bila data cukup besar maka cara manual akan sangat membuang waktu.

Nama Variabel	Label	Kode Variabel
lama_praktek_kategori	Lama praktik bidan	0 = <1 tahun
		1 = 1-5 tahun
		2 = >5 tahun
pendidikan	Pendidikan terakhir bidan	0 = SLTA/Sederajat
		1 = D1/D2
		2 = D3
		3 = S1/S2/S3
delima	Status bidan delima	0 = Bukan Bidan Delima
		1 = Bidan Delima
ukur_bb	Melakukan pengukuran berat badan ibu hamil	0 = Tidak
		1 = Ya
ukur_tb	Melakukan pengukuran tinggi badan ibu hamil	0 = Tidak
		1 = Ya
ukur_lila	Melakukan pengukuran lingkaran lengan atas ibu hamil	0 = Tidak
		1 = Ya
ukur_djb	Melakukan pengukuran denyut jantung bayi	0 = Tidak
		1 = Ya

Syntax-nya: `recode lama_praktek min/12=0 13/60=2 61/max=2, gen(lama_praktek_kategori)`. Berilah label dan value label untuk variabel ini. Ketika melakukan analisis, kadangkala kita ingin membuang beberapa observasi. Misalnya kita hanya ingin menganalisis variable tertentu saja, maka variabel yang tidak dibutuhkan kita hapus.

Syntax-nya:

`drop namavariabel` : membuang variabel yang terpilih

`keep namavariabel` : mempertahankan variabel yang terpilih

`drop if namavariabel expression` : membuang variabel yang terpilih dengan syarat tertentu

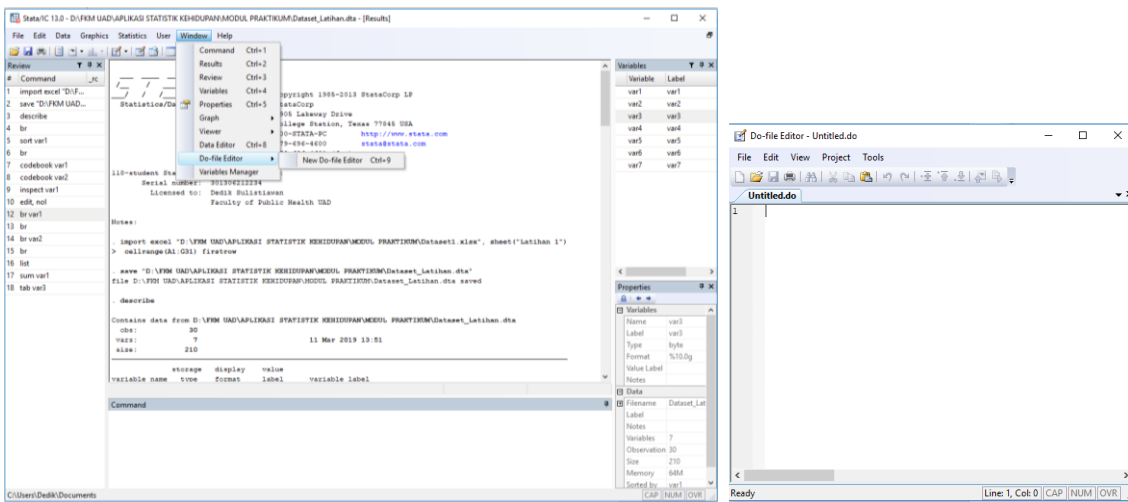
`keep if namavariabel expression` : mempertahankan variabel yang terpilih dengan syarat tertentu

H. Do-file

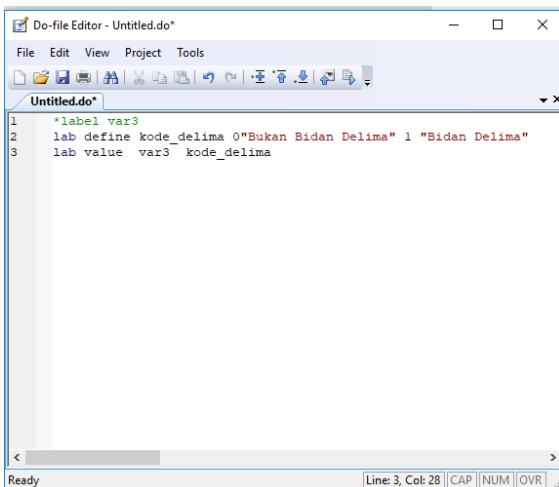
Do-file adalah suatu file teks yang berisi dokumentasi perintah-perintah yang akan diinstruksikan ke program Stata untuk dieksekusi. Do-file sangat berguna karena akan mempermudah kita untuk mengulang analisis/pekerjaan yang pernah kita lakukan di Stata.

1. Membuat Do-file

Caranya: Klik Window → Do-file Editor → New Do-file Editor atau CTRL+9 atau melalui Toolbar New Do-file Editor. Kemudian akan muncul jendela Do-file.



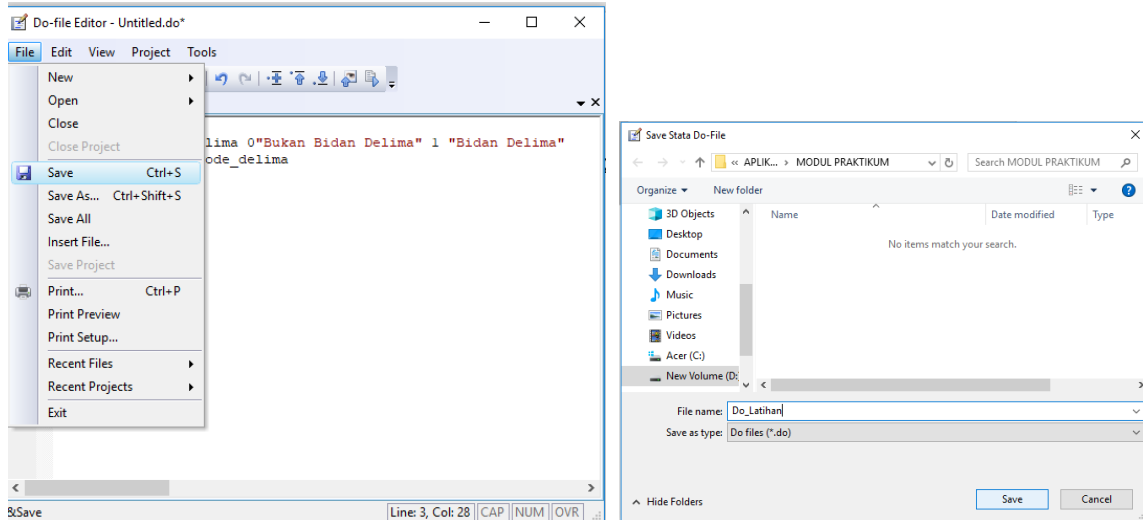
Seleksi/ pilih perintah yang ingin Anda simpan, copy (Ctrl+C) perintah yang ada di jendela review, kemudian Paste (Ctrl+V) di jendela Do-file Editor. Sebagai contoh copy perintah yang digunakan untuk melakukan labeling.



Tanda * (asterik) menginstruksikan Stata untuk mengabaikan perintah yang ada disamping kanan dari tanda bintang, sehingga memungkinkan kita melakukan catatan pada Do-file kita, seperti contoh di samping.

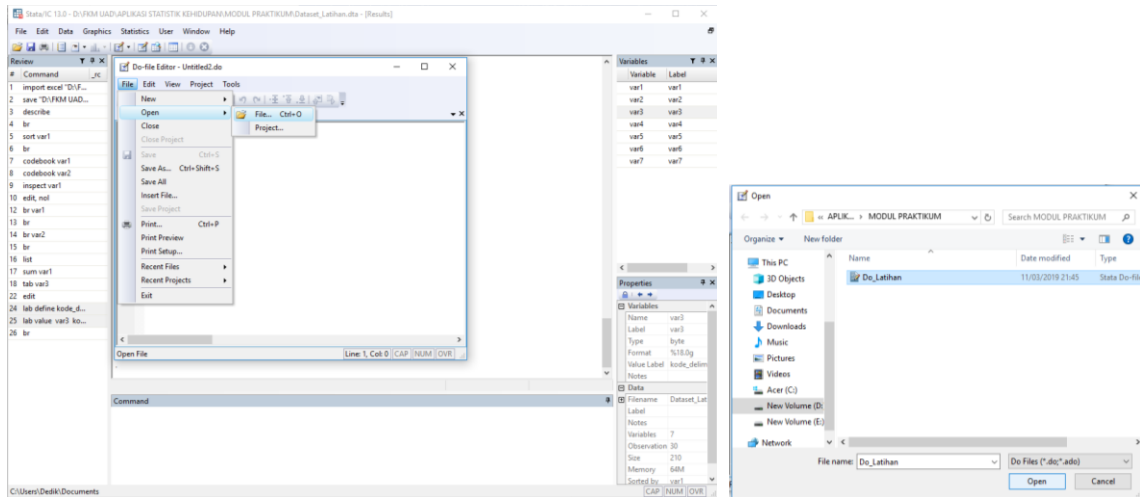
2. Menyimpan Do-File

Misalnya ingin menyimpan Do-file dengan nama Do_Latihan. Caranya: Pilih File → Save → Do_Latihan → Save



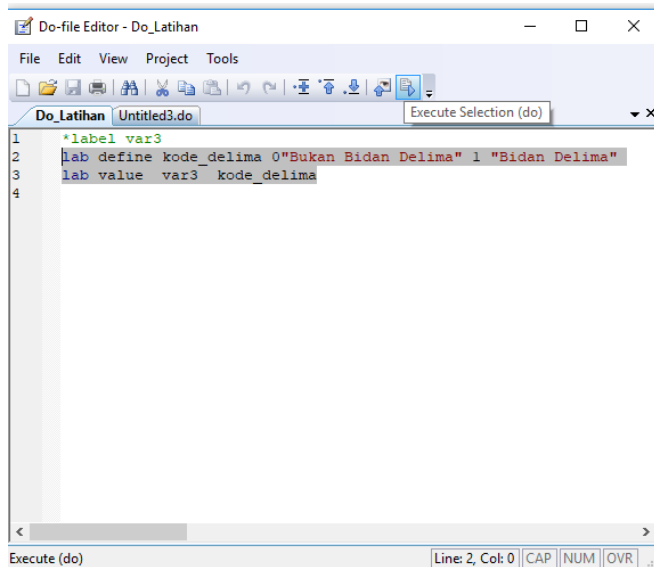
3. Membuka Do-file

Caranya: Buka Do-file editor dari toolbar kemudian klik File → Open → Pilih nama file → Open.



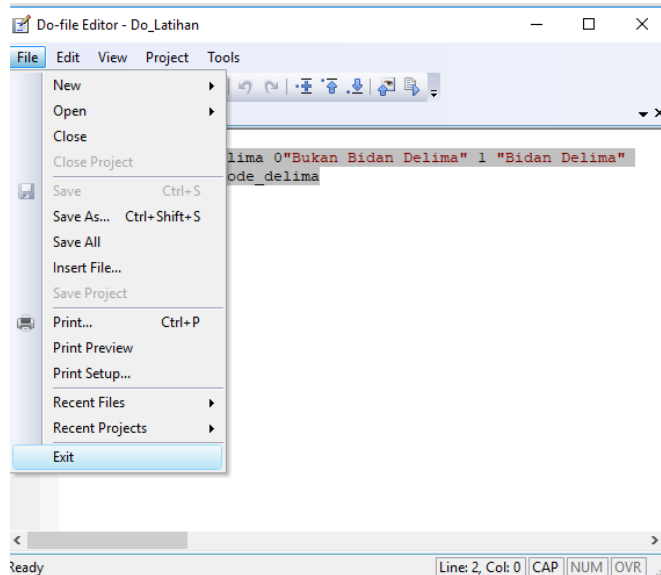
4. Menjalankan Do-file

Pilih beberapa perintah yang ada di Do-file kemudian execute selection (do). Maka secara otomatis Stata akan menjalankan perintah yang telah kita simpan dan panggil kembali ke do file.



5. Menutup Do-file

Tutup Do-file melalui menu bar File → Close atau File → Exit.



I. Latihan

1. Masukkan data berikut ke dalam program Stata.

kode_responden	status_bekerja	usia
101	1	55
102	1	46
103	0	40
104	1	41
105	1	43
106	1	40
107	1	40
108	1	48
109	0	39
110	0	45
111	0	44
112	0	43
113	0	45
114	1	49
115	0	48

2. Berikan value label untuk variabel `status_bekerja` dengan rincian:

0 = Tidak Bekerja

1 = Bekerja

Buatlah variabel baru untuk mengklasifikasikan variabel `usia` menjadi:

0 = <45 Tahun

1 = ≥45 Tahun

3. Laporkan pekerjaan Anda dalam bentuk Dataset Stata (*.dta) beserta dengan Do-file-nya (*.do).

PRAKTIKUM KE-2 ANALISIS DATA DESKRIPTIF

A. Tujuan

1. Mahasiswa mampu melakukan analisis data deskriptif (kategorik) menggunakan program Stata.
2. Mahasiswa mampu menginterpretasikan hasil analisis deskriptif (kategorik) menggunakan program Stata.
3. Mahasiswa mampu menyajikan hasil analisis ke dalam tabel dan grafik.

B. Pengantar

Modul ini merupakan ringkasan materi analisis data deskriptif (*exploratory data analysis*). Ringkasan ini tidak memberikan cakupan materi yang lengkap, hanya mencantumkan poin-poin utama. Tujuan dari *exploratory data analysis* (EDA) adalah untuk mengubah data yang tersedia dari bentuk mentah menjadi yang informatif. Saat melakukan EDA, kita harus selalu menggunakan tampilan visual (grafik atau tabel) ditambah ringkasan numerik. Jelaskan pula pola keseluruhan dan kecenderungan yang mencolok dari pola tersebut, dan interpretasikannya sesuai dengan konteks.

Ketika menjelaskan distribusi variabel tunggal, harus dibedakan antara variabel kategorik dan variabel kuantitatif (numerik). Distribusi variabel kategorik dirangkum menggunakan:

- Visual (*display*): diagram lingkaran atau diagram batang
- Ringkasan numerik: persentase kategori (grup) dalam bentuk tabel.

Sedangkan untuk distribusi variabel kuantitatif (numerik) dirangkum menggunakan:

- Visual (*display*): histogram (atau stem plot, terutama untuk set data kecil)
- Ringkasan numerik: statistik deskriptif (ukuran pusat dan ukuran penyebaran).

Jika distribusi data simetris tanpa outlier, digunakan mean dan standar deviasi. Jika tidak, digunakan median dan rentang antar-kuartil (IQR).

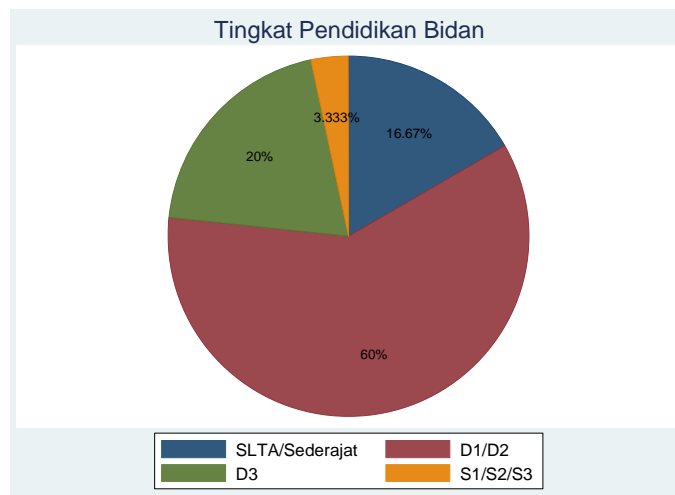
Modul ini terbatas pada proses analisis data deskriptif untuk variabel yang memiliki skala data nominal dan ordinal. Sedangkan untuk skala data interval dan rasio (numerik) dapat dipelajari secara mandiri di luar aktivitas praktikum menggunakan literatur yang direkomendasikan. Skala nominal dan ordinal disebut sebagai variabel kategorik karena memiliki kategori. Sebagai contoh variabel jenis kelamin memiliki kategori laki-laki dan perempuan. Skala data nominal memiliki kategori yang sederajat. Berbeda dengan skala data ordinal yang memiliki kategori bertingkat, misalnya status pendidikan dengan kategori rendah, sedang, dan tinggi.

C. Analisis Deskriptif Data Kategorik

Data kategorik dideskripsikan dalam bentuk jumlah (n) dan persentase (%). Adapun dataset latihan secara deskriptif disajikan dalam tabel berikut.

Variabel	n	%
Lama praktik bidan		
Kurang dari 1 tahun	1	3,3
Lebih dari 5 tahun	29	96,7
Pendidikan terakhir bidan		
SLTA/Sederajat	5	16,7
D1/D2	18	60,0
D3	6	20,0
S1/S2/S3	1	3,3
Status bidan delima		
Bukan Bidan Delima	19	63,3
Bidan Delima	11	36,7
Melakukan pengukuran berat badan ibu hamil		
Tidak	9	30,0
Ya	21	70,0
Melakukan pengukuran tinggi badan ibu hamil		
Tidak	17	56,7
Ya	13	43,3
Melakukan pengukuran lingkaran lengan atas ibu hamil		
Tidak	28	93,3
Ya	2	6,7
Melakukan pengukuran denyut jantung bayi		
Tidak	28	93,3
Ya	2	6,7
Total	30	100,0

Selain itu, data tersebut dapat juga disajikan dalam diagram lingkaran. Sebagai contoh adalah grafik tingkat pendidikan bidan sebagai berikut.



D. Mengeksekusi dengan Program Stata

Perintah umum untuk memperoleh deskripsi beberapa variabel kategorik secara bersamaan adalah menggunakan syntax: `tab1 var1 var2 var3 var4`. Sebagai contoh, syntax yang dijalankan untuk menghasilkan tabel di atas adalah: `tab1 lama_praktek_kategori pendidikan delima ukur_bb ukur_tb ukur_lila ukur_djb`. Output yang dihasilkan oleh Stata untuk syntax tersebut adalah sebagai berikut.

```
-> tabulation of lama_praktek_kategori
```

RECODE of lama_praktek (var1)	Freq.	Percent	Cum.
<1 tahun	1	3.33	3.33
>5 tahun	29	96.67	100.00
Total	30	100.00	

```
-> tabulation of pendidikan
```

var2	Freq.	Percent	Cum.
SLTA/Sederajat	5	16.67	16.67
D1/D2	18	60.00	76.67
D3	6	20.00	96.67
S1/S2/S3	1	3.33	100.00
Total	30	100.00	

-> tabulation of delima

var3	Freq.	Percent	Cum.
Bukan Bidan Delima	19	63.33	63.33
Bidan Delima	11	36.67	100.00
Total	30	100.00	

-> tabulation of ukur_bb

var4	Freq.	Percent	Cum.
Tidak	9	30.00	30.00
Ya	21	70.00	100.00
Total	30	100.00	

-> tabulation of ukur_tb

var5	Freq.	Percent	Cum.
Tidak	17	56.67	56.67
Ya	13	43.33	100.00
Total	30	100.00	

-> tabulation of ukur_lila

var6	Freq.	Percent	Cum.
Tidak	28	93.33	93.33
Ya	2	6.67	100.00
Total	30	100.00	

-> tabulation of ukur_djb

var7	Freq.	Percent	Cum.
Tidak	28	93.33	93.33
Ya	2	6.67	100.00
Total	30	100.00	

Adapun syntax untuk menyajikan data dalam bentuk diagram lingkaran secara umum adalah `graph pie, over(namavariabel)`. Sebagai contoh syntax yang dijalankan untuk menghasilkan diagram lingkaran seperti di atas adalah:

`graph pie, over(pendidikan) plabel(_all percent) title(Tingkat Pendidikan Bidan) legend(on)` yang akan menghasilkan output sama dengan diagram lingkaran di atas.

E. Latihan

Sajikan data pada latihan sebelumnya yang sudah dikategorikan ke dalam bentuk tabel dan diagram. Sertakan dataset Stata-nya (*.dta) dan Do-file-nya (*.do).

PRAKTIKUM KE-3 TABULASI SILANG DAN UKURAN ASOSIASI

A. Tujuan

1. Mahasiswa mampu melakukan analisis data kategorik menggunakan pendekatan tabulasi silang dengan program Stata.
2. Mahasiswa mampu menginterpretasikan hasil analisis data kategorik menggunakan pendekatan tabulasi silang dengan menggunakan program Stata.
3. Mahasiswa mampu menyajikan hasil analisis ke dalam tabel.

B. Pengantar

Modul ini merupakan ringkasan materi analisis data kategorik bivariable menggunakan pendekatan tabulasi silang. Ringkasan ini tidak memberikan cakupan materi yang lengkap, hanya mencantumkan poin-poin utama. Modul ini terbatas pada proses analisis untuk variabel dengan jenis data kategorik. Sedangkan untuk skala numerik dapat dipelajari secara mandiri di luar aktivitas praktikum menggunakan literatur yang direkomendasikan.

Selain itu, modul ini juga hanya membahas analisis data kategorik bivariabel sampel bebas (independen) saja. Untuk sampel berpasangan, dapat dipelajari secara mandiri di luar aktivitas praktikum menggunakan literatur yang direkomendasikan. Oleh karena itu, jenis uji statistik yang akan digunakan yaitu uji Chi-square, Fisher Exact Test.

Sebelum memulai analisis, dilakukan identifikasi jumlah kategori dari masing-masing variabel. Berdasarkan jumlah kategori dari masing-masing variabel, dapat diperoleh berbagai macam tabel baris x kolom ($b \times k$) yang dapat diklasifikasikan menjadi tabel 2×2 , $>2 \times 2$, $2 \times >2$, atau $>2 \times >2$. Uji statistik yang digunakan untuk semua jenis tabel tersebut adalah uji Chi-square, jika syaratnya terpenuhi. Syarat uji Chi-square adalah tidak ada *expected value* tiap sel yang bernilai kurang dari 5 lebih dari 20% dari keseluruhan sel. Jika syarat tersebut tidak terpenuhi, maka harus

digunakan uji alternatifnya, yaitu Fisher Exact Test untuk tabel 2x2. Modul ini menyajikan simulasi analisis (tabulasi silang) untuk tabel 2x2 dan >2x2. Oleh karena itu, uji yang akan digunakan yaitu Uji Chi-square (untuk tabel 2x2 yang memenuhi syarat dan >2x2) dan Fisher Exact Test sebagai alternatif uji Chi-square untuk tabel 2x2 yang tidak memenuhi syarat.

Selain itu, dalam studi epidemiologis biasanya peneliti juga menganalisis kekuatan hubungan (ukuran asosiasi) antara variabel independent dengan independennya. Ukuran ini dinyatakan dalam bentuk *Odds Ratio* (OR) untuk desain *case control* (retrospektif), *Relative Risk* (RR) untuk desain *cohort* (prospektif), dan *Prevalence Ratio* (PR) untuk desain *cross sectional*. Modul ini akan membahas pula cara untuk menampilkan hubungan asosiatif tersebut menggunakan program Stata. Khusus untuk desain studi *cross sectional*, hubungan asosiatif yang akan di-generate adalah *Odds Ratio* (OR) sebagai fungsi peluang.

C. Mengeksekusi dengan Program Stata

1. Tabulasi Silang 2x2

Syntax-nya: `tabulate var1 var2, expected`

Contoh: `tabulate delima ukur_tb, expected`

Key
<i>frequency</i>
<i>expected frequency</i>

var3	var5		Total
	Tidak	Ya	
Bukan Bidan Delima	10 10.8	9 8.2	19 19.0
Bidan Delima	7 6.2	4 4.8	11 11.0
Total	17 17.0	13 13.0	30 30.0

Interpretasi: Besarnya *expected value* untuk sel a, b, c, dan d berturut-turut adalah 10,8; 8,2; 6,2; dan 4,8. Terdapat satu sel (25%) yang memiliki *expected value* kurang dari 5, sehingga tidak memenuhi syarat untuk dianalisis menggunakan uji Chi-square. Alternatifnya, digunakan Fisher Exact Test.

Syntax-nya: `tabulate delima ukur_tb, exact column row`

Key
<i>frequency</i>
<i>row percentage</i>
<i>column percentage</i>

var3	var5		Total
	Tidak	Ya	
Bukan Bidan Delima	10	9	19
	52.63	47.37	100.00
	58.82	69.23	63.33
Bidan Delima	7	4	11
	63.64	36.36	100.00
	41.18	30.77	36.67
Total	17	13	30
	56.67	43.33	100.00
	100.00	100.00	100.00

Fisher's exact = 0.708
 1-sided Fisher's exact = 0.421

Interpretasi: Tidak ada perbedaan dalam melakukan pengukuran tinggi badan ibu hamil antara bidan delima dan bukan bidan delima.

2. Tabulasi Silang >2x2

Syntaxnya: `tabulate var1 var2, chi2 column row`

Contoh: `tabulate pendidikan ukur_djb, chi2 column row`

var2	var7		Total
	Tidak	Ya	
SLTA/Sederajat	4	1	5
	80.00	20.00	100.00
	14.29	50.00	16.67
D1/D2	18	0	18
	100.00	0.00	100.00
	64.29	0.00	60.00
D3	6	0	6
	100.00	0.00	100.00
	21.43	0.00	20.00
S1/S2/S3	0	1	1
	0.00	100.00	100.00
	0.00	50.00	3.33
Total	28	2	30
	93.33	6.67	100.00
	100.00	100.00	100.00

Pearson chi2(3) = 17.1429 Pr = 0.001

Key
<i>frequency</i>
<i>row percentage</i>
<i>column percentage</i>

Interpretasi: Ada perbedaan dalam melakukan pengukuran denyut jantung bayi antara bidan dengan tingkat pendidikan yang berbeda.

3. Ukuran Asosiasi *Relative Risk* (RR)

Relative Risk (RR) tepat dipakai sebagai ukuran risiko relatif antara kelompok terpapar (ekspos) dan tidak terpapar (nonekspos) pada penelitian prospektif (kohort) dan eksperimen. Nilai RR = 1, berarti risiko sakit/ mengalami kejadian antara kelompok yang memiliki risiko sama dengan risiko sakit/ mengalami kejadian pada kelompok yang tidak memiliki risiko. Artinya, jika RR = 1 menunjukkan bahwa faktor tersebut tidak meningkatkan risiko terjadinya penyakit/ kejadian. Nilai RR > 1 mengindikasikan bahwa keberadaan faktor risiko tersebut meningkatkan risiko terjadinya penyakit pada kelompok terpapar. Nilai RR < 1 mengindikasikan bahwa keberadaan faktor yang diteliti tersebut memberikan efek pencegahan atau dapat menurunkan risiko terjadinya penyakit/ kejadian pada kelompok yang memiliki faktor tersebut. Untuk mengetahui apakah RR di populasi berbeda dengan 1, maka nilai RR yang didapatkan dari sampel dihitung nilai *confidence interval*-nya atau diuji kemaknaannya dengan uji Chi-square.

Syntaxnya: `cs variabeldependen variabelindependen`

Contoh: `cs ukur_djb delima`

	var3		Total
	Exposed	Unexposed	
Cases	1	1	2
Noncases	10	18	28
Total	11	19	30
Risk	.0909091	.0526316	.0666667
	Point estimate		[95% Conf. Interval]
Risk difference	.0382775		-.1590612 .2356162
Risk ratio	1.727273		.119559 24.95397
Attr. frac. ex.	.4210526		-7.364072 .9599262
Attr. frac. pop	.2105263		

chi2(1) = 0.16 Pr>chi2 = 0.6855

Interpretasi: Tidak ada hubungan antara status bidan delima dengan perilaku memeriksa denyut jantung bayi.

4. Ukuran Asosiasi *Odds Ratio* (OR)

Odds Ratio (OR) tepat dipakai untuk menentukan perbedaan risiko pada penelitian *case control* (retrospektif). Nilai OR sama dengan 1 menunjukkan bahwa odds terjadi sakit/ mengalami kejadian antara kelompok ekspos dan non-ekspos tidak berbeda. Nilai OR > 1 menunjukkan bahwa odds terjadinya sakit/ mengalami kejadian pada kelompok ekspos lebih tinggi dari nonekspos. Nilai OR < 1 mengindikasikan terjadinya penurunan odds sakit/ mengalami kejadian pada kelompok terekspos atau juga disebut keberadaan faktor tersebut memberikan efek pencegahan. Untuk mengetahui apakah OR di populasi berbeda dengan 1, maka nilai OR yang didapatkan dari sampel dihitung nilai *confidence interval*-nya atau diuji kemaknaannya dengan uji Chi-square.

Syntaxnya: `cc variabeldependen variabelindependen`

Contoh: `cc ukur_djb delima`

`. cc ukur_djb delima`

	Exposed	Unexposed	Total	Proportion Exposed
Cases	1	1	2	0.5000
Controls	10	18	28	0.3571
Total	11	19	30	0.3667
	Point estimate		[95% Conf. Interval]	
Odds ratio	1.8		.0208645	149.0103 (exact)
Attr. frac. ex.	.4444444		-46.92839	.9932891 (exact)
Attr. frac. pop	.2222222			

chi2(1) = 0.16 Pr>chi2 = 0.6855

Interpretasi: Tidak ada hubungan antara status bidan delima dengan perilaku memeriksa denyut jantung bayi.

Ukuran asosiasi yang sebenarnya tepat digunakan untuk desain *cross sectional* adalah *Prevalence Ratio* (PR). Meskipun demikian, banyak peneliti menggunakan OR untuk menyatakan ukuran asosiasi untuk desain *cross-sectional* dengan sedikit modifikasi pada saat melakukan interpretasi (menganalogikan OR sebagai peluang/ probabilitas). Meskipun demikian, beberapa studi menyebutkan bahwa mengestimasi PR menggunakan OR rawan sekali terjadi *overestimate*.

D. Latihan

Kasus: Peneliti ingin menganalisis hubungan antara jenis kelamin, usia, status merokok, dan kategori *body mass index* (BMI) dengan status penyakit jantung koroner (PJK). Data hasil penelitian ini disajikan dalam dokumen `Dataset_Latihan3.dta`

1. Lakukan tabulasi silang untuk seluruh variabel independen dengan variabel dependen. Laporkan dalam bentuk tabel ringkasan yang disertai dengan signifikansi statistik berdasarkan hasil analisis yang dilakukan. Berikan narasi/interpretasi.
2. Asumsikan desain penelitian ini adalah *case control*. Laporkan ukuran asosiasi (OR) beserta interpretasinya untuk variabel independen jenis kelamin dan status merokok.

Sertakan Do-file-nya (*.do).

PRAKTIKUM KE-4 ANALISIS DATA KATEGORIK MULTIVARIABEL

A. Tujuan

1. Mahasiswa mampu melakukan analisis data kategorik multivariabel menggunakan program Stata.
2. Mahasiswa mampu menginterpretasikan hasil analisis data kategorik multivariabel menggunakan program Stata.
3. Mahasiswa mampu menyajikan hasil analisis ke dalam tabel.

B. Pengantar

Modul ini merupakan ringkasan materi analisis data kategorik multivariabel. Ringkasan ini tidak memberikan cakupan materi yang lengkap, hanya mencantumkan poin-poin utama. Modul ini terbatas pada proses analisis data untuk variabel yang memiliki skala data kategorik. Sedangkan untuk skala data numerik dapat dipelajari secara mandiri di luar aktivitas praktikum menggunakan literatur yang direkomendasikan.

Dalam penelitian kesehatan atau kedokteran, variabel *outcome* yang sering diteliti merupakan variabel dengan dua kategori atau disebut *binary (binomial) outcome*. Misalnya pada penelitian faktor risiko penyakit jantung koroner (PJK), yang menjadi variabel *outcome* (tergantung) adalah PJK dengan dua kategori, yaitu menderita PJK dan bukan PJK. Selain itu, tak jarang pula variabel *outcome* terdiri atas lebih dari 2 kategori atau disebut multinomial. Selain itu, pada umumnya masalah kesehatan masyarakat sifatnya kompleks dan jarang terjadinya satu kejadian yang disebabkan oleh satu faktor tunggal (multifaktor). Sebagai contoh, misalnya penyakit jantung koroner (PJK) disebabkan oleh banyak faktor usia, jenis kelamin, kebiasaan merokok, indeks massa tubuh, dan sebagainya.

Untuk mempelajari hubungan satu masalah dengan berbagai faktor yang terkait tidak bias lagi dianalisis secara bivariat. Metode analisis multivariat atau multivariabel merupakan metode analisis yang memungkinkan kita mempelajari

hubungan beberapa variabel bebas dengan satu variabel tergantung. Dalam praktikum ini, akan dibahas tentang analisis data kategorik multivariabel dengan variabel dependen biner/ binomial menggunakan regresi logistik berganda (*multiple logistic regression*) dan variabel dependen poli/ multinomial menggunakan regresi logistik berganda multinomial (*multinomial multiple logistic regression*).

C. Mengeksekusi dengan Program Stata

1. Regresi Logistik Berganda

Kriteria pemilihan dan seleksi variabel independen untuk model regresi logistik berbeda antar disiplin ilmu. Prinsip dasar utama yaitu mencari model yang paling parsimoni (*parsimonious*), yang masih mampu “menjelaskan” data. Model demikian secara numerik akan lebih stabil dalam arti kata memiliki estimasi interval koefisien regresi dan *standard error* yang relatif sempit, serta lebih mudah digeneralisasikan.

Dalam epidemiologi misalnya, dianjurkan untuk menginklusi semua variabel yang relevan secara klinik ataupun intuitif dalam model, tanpa tergantung pada kemaknaan statistiknya. Praktik ini diharapkan akan mampu mengendalikan semua kemungkinan *confounding* pada dataset. Sebaliknya, inklusi terlalu banyak variabel dalam model, relatif terhadap jumlah anggota sampelnya dapat menyebabkan “overfitting”, yaitu estimasi yang tak stabil dengan estimasi koefisien regresi dan *standard error* yang terlalu besar.

Beberapa ahli menyepakati bahwa kandidat variabel yang akan digunakan sebagai prediktor dalam analisis regresi logistik adalah variabel yang ketika dianalisis secara bivariat menunjukkan hasil *p-value* $< 0,25$. Oleh karena itu, apabila terdapat kandidat variabel yang secara bivariat memiliki *p-value* $> 0,25$ maka tidak dimasukkan sebagai kandidat prediktor dalam analisis regresi logistik berganda. Secara umum, urutan langkah pada seleksi variabel yaitu:

- a. Pilih calon variabel independen dengan teliti, lalu lakukan analisis univariabel tiap calon variabel independen ini dengan variabel dependen. Tiap calon variabel independen yang pada analisis univariabel ini menghasilkan nilai $p <$

- 0.25 dan secara substantive memiliki kemaknaan substantif menurut ranah bidang ilmu yang diteliti dapat dipertahankan menuju langkah berikutnya. Perhatikan bahwa calon prediktor yang pada analisis univariabel hanya menunjukkan asosiasi lemah dengan respons, mungkin menjadi prediktor penting pada analisis multivariabel.
- b. Lakukan analisis multivariabel dengan variabel dependen, menggunakan himpunan variabel independen yang lolos dari seleksi tahap 1. Beberapa alternatifnya:
- Inklusikan seluruh variabel independen yang dianggap “relevan secara ilmiah” dalam analisis multivariabel tanpa tergantung pada hasil analisis univariabel pada langkah a. Seandainya jumlah variabel independen tersebut terlalu banyak (dalam perbandingan dengan ukuran sampel), dapat dilakukan seleksi ulang dengan mendefinisikan kembali kriteria “relevan secara ilmiah”.
 - Lakukan seleksi variabel dengan metode “*stepwise*”. Inklusi dan eksklusi variabel dengan metode ini sepenuhnya didasarkan pada kriteria statistik, yaitu nilai p -nya. Dua versi metode *stepwise* ialah “seleksi ke depan” (*forward selection*) dan “eliminasi ke belakang” (*backward elimination*).
Pada *forward selection*, analisis dimulai dengan 1 variabel independen yang nilai p -nya terkecil pada analisis univariabel, lalu setiap kali ditambah 1 variabel independen yang nilai p -nya terkecil berikutnya pada analisis univariabel. Lihat hasil uji Wald pada tiap penambahan variabel baru, tiap variabel yang tak bermakna secara statistik dikeluarkan kembali dari model. Penambahan ke depan berlanjut sampai semua calon variabel independen dicoba untuk dimasukkan dalam model.
Pada *backward elimination*, analisis multivariabel dimulai dengan memasukkan semua variabel independen yang menunjukkan kemaknaan statistik pada analisis univariabel, lalu dicoba mengeliminasi 1 variabel independen yang pada uji Wald memiliki nilai p terbesar dan tak bermakna secara statistik. Lakukan kembali analisis multivariabel setelah

mengeliminasi 1 variabel ini. Proses percobaan eliminasi ini berlanjut sampai yang tersisa seluruhnya adalah variabel independen yang bermakna secara statistik.

Prinsip dasar yang terpenting untuk diingat pada tahap ini, yaitu bahwa yang terutama seharusnya bertanggung jawab menelaah-ulang dan mengevaluasi model adalah peneliti, bukan komputer.

- c. Verifikasikan kepentingan tiap variabel independen yang diinkluskikan dalam model sebagai hasil langkah b.
 - Periksa kembali statistik uji Wald untuk tiap variabel.
 - Bandingkan hasil estimasi tiap koefisien regresi dengan hasil estimasi pada analisis univariabel pada langkah a. Perbedaan yang terlalu besar mungkin mengindikasikan bahwa 1 atau lebih variabel yang telah dieksklusikan sebenar masih perlu dipertahankan demi “penyesuaian” efek bersama seluruh variabel independen.
- d. Periksa semua kemungkinan interaksi antar variabel utama yang dihasilkan dari langkah-langkah di atas. Keputusan untuk menginkluskikan suatu interaksi harus didasarkan atas pertimbangan statistik maupun praktik. Inklusi interaksi yang tak bermakna secara statistik hanya akan memperbesar rentang estimasi standard error tanpa mengubah nilai estimasi koefisiennya sendiri. Interaksi yang diinkluskikan juga harus memiliki arti sesuai ranah bidang ilmu yang diteliti.

Syntax: `logistic variabeldependen i.variabelindependen1
i.variabelindependen2 i.variabelindependen3`

Contoh: `logistic pjk i.jenis_kelamin i.usia i.rokok i.bmi`

Output Stata untuk analisis ini adalah sebagai berikut.


```

Logistic regression              Number of obs =      500
                                LR chi2(7) =      126.06
                                Prob > chi2 =      0.0000
Log likelihood = -71.416669      Pseudo R2 =      0.4688
    
```

pkj	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
jenis_kelamin					
Laki-laki	6.375352	3.574941	3.30	0.001	2.124212 19.13421
usia					
36-45 tahun	1.101717	.7794643	0.14	0.891	.275325 4.408539
46-55 tahun	17.52746	13.41672	3.74	0.000	3.909767 78.57544
>55 tahun	9.980274	6.213241	3.70	0.000	2.945948 33.81114
rokok					
Merokok	6.589764	3.186573	3.90	0.000	2.554227 17.00123
bmi					
Kurus	.7019434	.5995908	-0.41	0.679	.1315882 3.744443
Gemuk	2.981477	2.072352	1.57	0.116	.7634523 11.64343
_cons	.0024426	.0019957	-7.36	0.000	.0004925 .0121144

Keluarkan variable BMI yang tidak signifikan pada saat uji parsial (Uji Wald).

```
. logistic pkj i.jenis_kelamin i.usia i.rokok
```

```

Logistic regression              Number of obs =      500
                                LR chi2(5) =      119.50
                                Prob > chi2 =      0.0000
Log likelihood = -74.69514      Pseudo R2 =      0.4444
    
```

pkj	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
jenis_kelamin					
Laki-laki	9.44338	5.077842	4.18	0.000	3.291729 27.09136
usia					
36-45 tahun	1.186135	.8306387	0.24	0.807	.3006395 4.679745
46-55 tahun	15.53641	10.71831	3.98	0.000	4.019028 60.05933
>55 tahun	16.50904	9.568278	4.84	0.000	5.301318 51.41142
rokok					
Merokok	5.741759	2.57721	3.89	0.000	2.382209 13.83917
_cons	.0027028	.0019065	-8.38	0.000	.0006782 .0107705

Bandingkan dengan hasil analisis univariatnya. Identifikasi *unadjusted* OR dan *adjusted* OR nya.

2. Regresi Logistik Berganda Multinomial

Regresi logistik multinomial (regresi logistik polinomial) adalah pemodelan regresi logistik untuk data prediktor-respons dengan respons kategorik nominal non-biner. Misalkan dimiliki data prediktor-respons dengan respons berskala nominal non-biner yang memiliki M kategori, maka akan dipilih 1 kategori sebagai kategori

dasar (*reference*), dan tiap kategori lainnya masing-masing akan dibandingkan dengan *reference category* ini, sehingga diperoleh $(M - 1)$ model regresi logistik. Jika tidak dispesifikasikan, umumnya yang diambil untuk *reference category* secara default adalah kategori dengan nilai respons terendah. Jika respons memiliki 4 kategori, I, II, III, dan IV, maka secara default kategori I akan menjadi *reference category*, lalu dilakukan 3 kali pemodelan regresi logistik, II vs I, III vs I, dan IV vs I. Karena tiap pemodelan menggunakan himpunan nilai prediktor yang berbeda, akan diperoleh 3 model regresi logistik yang berbeda nilai-nilai estimasi koefisien regresinya maupun estimasi konstantanya.

Contoh: File data `Dataset_Latihan3&4_2` yang digunakan memuat data tentang 500 orang responden. Variabel dependen berupa Status PJK yang dikategorikan menjadi 0 = Tidak PJK; 1 = PJK; 2 = PJK+Hipertensi. Prediktornya adalah jenis kelamin, usia, status merokok, dan BMI.

Syntax: `mlogit pjk i.jenis_kelamin i.usia i.rokok i.bmi`

```
Iteration 0: log likelihood = -253.71191
Iteration 1: log likelihood = -180.41236
Iteration 2: log likelihood = -153.64139
Iteration 3: log likelihood = -147.5588
Iteration 4: log likelihood = -147.28063
Iteration 5: log likelihood = -147.27955
Iteration 6: log likelihood = -147.27955
```

```
Multinomial logistic regression          Number of obs   =       500
                                         LR chi2(14)     =       212.86
                                         Prob > chi2     =       0.0000
                                         Pseudo R2      =       0.4195

Log likelihood = -147.27955
```

pjk	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Tidak_PJK	(base outcome)					
PJK						
jenis_kelamin						
Laki-laki	.9006148	.6257734	1.44	0.150	-.3258785	2.127108
usia						
36-45 tahun	-.0976718	.7109107	-0.14	0.891	-1.491031	1.295688
46-55 tahun	1.590543	.7007009	2.27	0.023	.2171947	2.963892
>55 tahun	2.427595	.8121482	2.99	0.003	.8358135	4.019376
rokok						
Merokok	2.756383	.7192678	3.83	0.000	1.346644	4.166122
bmi						
Kurus	.5129741	.7949195	0.65	0.519	-1.045039	2.070988
Gemuk	1.078176	.7411765	1.45	0.146	-.3745035	2.530855
_cons	-6.020761	.9035716	-6.66	0.000	-7.791729	-4.249793

PJK_Hipertensi						
jenis_kelamin						
Laki-laki	1.081039	.5034061	2.15	0.032	.0943814	2.067697
usia						
36-45 tahun	.689357	.6586094	1.05	0.295	-.6014936	1.980208
46-55 tahun	2.596203	.6762208	3.84	0.000	1.270835	3.921572
>55 tahun	4.272251	.7733371	5.52	0.000	2.756538	5.787964
rokok						
Merokok	1.894636	.5242101	3.61	0.000	.8672036	2.922069
bmi						
Kurus	1.123369	.688516	1.63	0.103	-.2260976	2.472836
Gemuk	.516805	.6733114	0.77	0.443	-.802861	1.836471
_cons	-5.885016	.8209642	-7.17	0.000	-7.494076	-4.275956

Keluarkan variabel BMI yang tidak signifikan pada saat uji parsial (Uji Wald).

```
. mlogit pjk i.jenis_kelamin i.usia i.rokok
```

```
Iteration 0: log likelihood = -253.71191
Iteration 1: log likelihood = -181.82914
Iteration 2: log likelihood = -155.83769
Iteration 3: log likelihood = -150.19796
Iteration 4: log likelihood = -149.96013
Iteration 5: log likelihood = -149.95924
Iteration 6: log likelihood = -149.95924
```

```
Multinomial logistic regression          Number of obs   =       500
                                         LR chi2(10)     =       207.51
                                         Prob > chi2     =       0.0000
Log likelihood = -149.95924             Pseudo R2      =       0.4089
```

pjk	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Tidak_PJK	(base outcome)					
PJK						
jenis_kelamin						
Laki-laki	1.074265	.5766498	1.86	0.062	-.0559484	2.204477
usia						
36-45 tahun	.0210472	.7036315	0.03	0.976	-1.358045	1.40014
46-55 tahun	1.880276	.6706035	2.80	0.005	.5659172	3.194635
>55 tahun	2.858009	.7483183	3.82	0.000	1.391332	4.324685
rokok						
Merokok	2.612252	.6742351	3.87	0.000	1.290776	3.933729
_cons	-5.515559	.7474765	-7.38	0.000	-6.980586	-4.050532
PJK_Hipertensi						
jenis_kelamin						
Laki-laki	.852207	.4755371	1.79	0.073	-.0798286	1.784243
usia						
36-45 tahun	.7832897	.6509816	1.20	0.229	-.4926108	2.05919
46-55 tahun	2.668498	.6739953	3.96	0.000	1.347491	3.989504
>55 tahun	4.118064	.7009165	5.88	0.000	2.744293	5.491835
rokok						
Merokok	2.053822	.5148627	3.99	0.000	1.04471	3.062934
_cons	-5.208913	.6392268	-8.15	0.000	-6.461775	-3.956052

Keluarkan variable jenis kelamin yang tidak signifikan pada saat uji parsial (Uji Wald).

```
. mlogit pjk i.usia i.rokok
```

```
Iteration 0: log likelihood = -253.71191
Iteration 1: log likelihood = -185.0983
Iteration 2: log likelihood = -159.68067
Iteration 3: log likelihood = -153.22929
Iteration 4: log likelihood = -152.89987
Iteration 5: log likelihood = -152.8987
Iteration 6: log likelihood = -152.8987
```

```
Multinomial logistic regression      Number of obs   =      500
LR chi2(8)                          =      201.63
Prob > chi2                          =      0.0000
Log likelihood = -152.8987           Pseudo R2      =      0.3974
```

pjk	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Tidak_PJK (base outcome)						
PJK						
usia						
36-45 tahun	-.1327015	.6958245	-0.19	0.849	-1.496492	1.231089
46-55 tahun	2.040703	.657843	3.10	0.002	.7513543	3.330051
>55 tahun	3.068133	.7095793	4.32	0.000	1.677383	4.458883
rokok						
Merokok	2.802317	.6613157	4.24	0.000	1.506162	4.098472
_cons	-5.043095	.6743728	-7.48	0.000	-6.364842	-3.721349
PJK_Hipertensi						
usia						
36-45 tahun	.6716947	.6471923	1.04	0.299	-.5967789	1.940168
46-55 tahun	2.778618	.6659081	4.17	0.000	1.473462	4.083774
>55 tahun	4.24328	.6795694	6.24	0.000	2.911349	5.575212
rokok						
Merokok	2.249438	.4995492	4.50	0.000	1.270339	3.228536
_cons	-4.877635	.6000004	-8.13	0.000	-6.053614	-3.701655

Setelah mendapatkan model yang dirasa fit, untuk memperoleh estimasi nilai-nilai rasio risiko, digunakan opsi `rrr`.

```
. mlogit, rrr
```

```
Multinomial logistic regression      Number of obs   =      500
LR chi2(8)                          =      201.63
Prob > chi2                          =      0.0000
Log likelihood = -152.8987           Pseudo R2      =      0.3974
```

```
. mlogit, rrr

Multinomial logistic regression      Number of obs   =      500
                                      LR chi2(8)      =     201.63
                                      Prob > chi2     =      0.0000
Log likelihood = -152.8987           Pseudo R2      =      0.3974
```

pjk	RRR	Std. Err.	z	P> z	[95% Conf. Interval]	
Tidak_PJK	(base outcome)					
PJK						
usia						
36-45 tahun	.8757265	.6093519	-0.19	0.849	.2239142	3.424959
46-55 tahun	7.696017	5.06277	3.10	0.002	2.119869	27.93978
>55 tahun	21.50172	15.25717	4.32	0.000	5.351532	86.39093
rokok						
Merokok	16.48279	10.90033	4.24	0.000	4.509391	60.24816
_cons	.0064537	.0043522	-7.48	0.000	.001721	.0242013
PJK_Hipertensi						
usia						
36-45 tahun	1.957552	1.266913	1.04	0.299	.5505823	6.959922
46-55 tahun	16.09676	10.71896	4.17	0.000	4.364318	59.36909
>55 tahun	69.63589	47.32241	6.24	0.000	18.38157	263.8053
rokok						
Merokok	9.482403	4.736927	4.50	0.000	3.562061	25.24268
_cons	.007615	.004569	-8.13	0.000	.0023494	.0246826

Interpretasi: Prediktor kejadian PJK adalah usia dan status merokok. Usia >55 tahun berpeluang mengalami kejadian PJK 21,5 kali dibandingkan usia 25-35 tahun. Usia 46-55 tahun berpeluang mengalami kejadian PJK 7,7 kali dibandingkan usia 25-35 tahun. Perilaku merokok berpeluang mengalami kejadian PJK 16,5 kali dibandingkan tidak merokok.

Prediktor kejadian PJK+Hipertensi adalah usia dan status merokok. Usia >55 tahun berpeluang mengalami kejadian PJK+Hipertensi 69,6 kali dibandingkan usia 25-35 tahun. Usia 46-55 tahun berpeluang mengalami kejadian PJK+Hipertensi 16,1 kali dibandingkan usia 25-35 tahun. Perilaku merokok berpeluang mengalami kejadian PJK+Hipertensi 9,5 kali dibandingkan tidak merokok.

D. Latihan

Lakukan analisis regresi logistik berganda menggunakan Dataset_Latihan5.dta dengan variabel dependen berupa kesertaan KB. Interpretasikan hasilnya. Sertakan pula dataset Stata-nya (*.dta) dan Do-file-nya (*.do).

PRAKTIKUM KE-5 MODEL REGRESI POISSON UNTUK DATA DISKRIT

A. Tujuan Praktikum

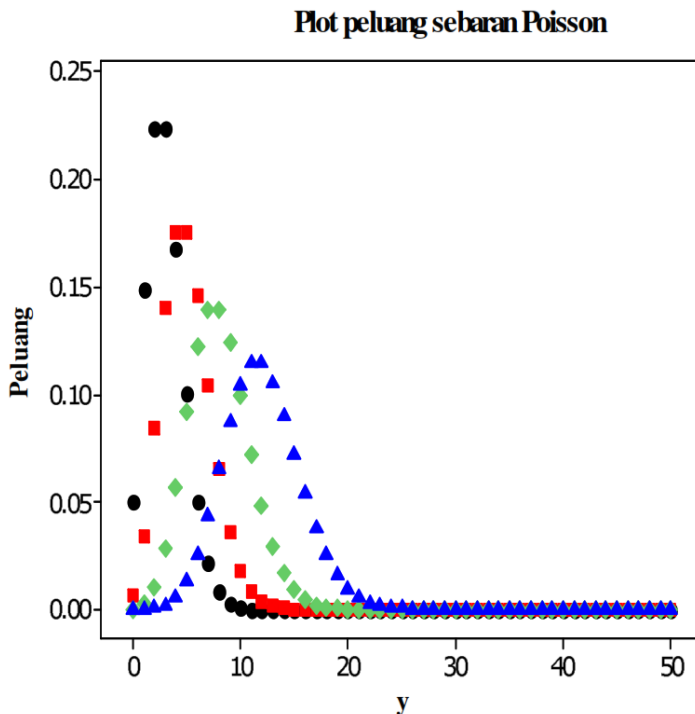
1. Mahasiswa mampu mengidentifikasi tujuan pemodelan dengan regresi Poisson.
2. Mahasiswa mampu melakukan analisis regresi Poisson menggunakan program Stata.
3. Mahasiswa mampu menginterpretasikan hasil regresi Poisson menggunakan program Stata.

B. Pengantar

Analisis regresi adalah suatu metode yang digunakan untuk menganalisis hubungan antara variabel respon dan variabel prediktor. Pada umumnya analisis regresi digunakan untuk menganalisis variabel respon yang merupakan data kontinu dan mengikuti distribusi normal atau data ketegorik. Namun dalam beberapa aplikasinya, variabel respon yang akan dianalisis dapat berupa data diskrit. Salah satu contoh dimana variabel responnya diskrit adalah banyaknya kejadian yang jarang terjadi (*rare event*). Misalkan banyaknya kematian ibu akibat komplikasi kehamilan dan persalinan setiap tahun, penderita kanker paru-paru yang meninggal setiap tahun, jumlah kecelakaan lalu lintas dalam sebulan, dan sebagainya. Salah satu model regresi yang dapat digunakan untuk menjelaskan hubungan antara variabel respon Y yang berupa data diskrit dengan variabel prediktor X berupa data diskrit, kontinu, kategorik atau campuran adalah model regresi Poisson.

Regresi Poisson didapatkan dari distribusi Poisson, yaitu suatu distribusi untuk peristiwa yang probabilitas kejadiannya kecil, dimana kejadiannya tergantung pada interval waktu tertentu atau di suatu daerah tertentu dengan hasil pengamatan berupa variabel diskrit dan antar variabel saling independen. Interval waktu tersebut dapat berapa saja panjangnya, misalnya semenit, sehari, seminggu, sebulan atau

bahkan setahun. Distribusi Poisson sering digunakan untuk kejadian-kejadian yang jarang terjadi dengan data berupa cacahan (data hitung/ *count*) yang bernilai non-negatif.



Gambar: Grafik Sebaran Distribusi Poisson

C. Mengeksekusi dengan Program STATA

Prosedur analisis regresi Poisson adalah sebagai berikut.

1. Buka dataset yang telah disediakan.
2. Lakukan pengujian distribusi data untuk memastikan bahwa variabel dependen yang akan dianalisis mengikuti distribusi Poisson.

```
. swilk jumlah_anc
```

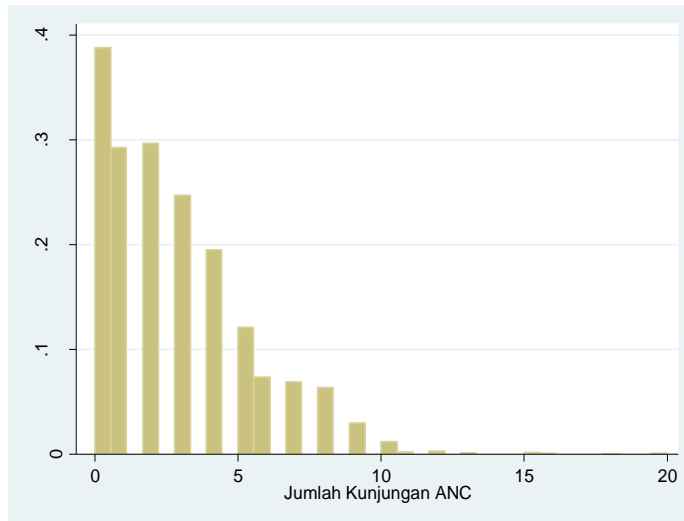
Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
jumlah_anc	4,493	0.95001	123.133	12.587	0.00000

Ho ditolak → variabel dependen tidak mengikuti distribusi normal.

3. Lakukan pengecekan plot variabel dependen dengan membuat histogram.

hist jumlah_anc



Data yang mengikuti distribusi Poisson selalu memiliki bentuk yang menceng ke kanan, dan umum dijumpai pada data hitung (count data) yang bernilai selalu positif. Oleh karena itu, analisis poisson mungkin sesuai untuk dilakukan.

4. Lakukan analisis regresi Poisson.

poisson jumlah_anc i.kategori_wilayah i.tingkat_pendidikan
i.indeks_kesejahteraan i.status_kehamilan

```
. poisson jumlah_anc i.kategori_wilayah i.tingkat_pendidikan i.indeks_kesejahteraan i.status_kehamilan

Iteration 0: log likelihood = -9856.7293
Iteration 1: log likelihood = -9856.6926
Iteration 2: log likelihood = -9856.6926
```

```
Poisson regression              Number of obs   =      4,493
                               LR chi2(9)         =     1858.48
                               Prob > chi2         =      0.0000
                               Pseudo R2          =      0.0862

Log likelihood = -9856.6926
```

	jumlah_anc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
kategori_wilayah						
	Rural	-.1727083	.0206284	-8.37	0.000	-.2131393 -.1322773
tingkat_pendidikan						
	Tamat SD	.1228632	.0298246	4.12	0.000	.064408 .1813184
	Tamat SMP	.2775914	.029831	9.31	0.000	.2191237 .3360592
	Tamat SMA	.4617219	.0336348	13.73	0.000	.395799 .5276448
indeks_kesejahteraan						
	Menengah	.1981508	.0352444	5.62	0.000	.1290731 .2672285
	Kaya	.3044342	.0346498	8.79	0.000	.2365219 .3723466
	Terkaya	.4741512	.0340182	13.94	0.000	.4074768 .5408256
	5	.6498982	.036384	17.86	0.000	.578587 .7212095
status_kehamilan						
	Diinginkan	.24587	.0352495	6.98	0.000	.1767823 .3149577
	_cons	.3170494	.045577	6.96	0.000	.22772 .4063788

5. Transformasikan koefisien regresi Poisson ke dalam bentuk *Incidence Rate Ratio (IRR)*. Caranya, setelah mengeksekusi command di atas, ketik `poisson, irr`.

```
. poisson, irr
```

```
Poisson regression              Number of obs   =      4,493
                               LR chi2(9)        =     1858.48
                               Prob > chi2         =      0.0000
Log likelihood = -9856.6926     Pseudo R2       =      0.0862
```

jumlah_anc	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
kategori_wilayah						
Rural	.841383	.0173564	-8.37	0.000	.8080436	.876098
tingkat_pendidikan						
Tamat SD	1.13073	.0337236	4.12	0.000	1.066527	1.198797
Tamat SMP	1.319947	.0393754	9.31	0.000	1.244985	1.399422
Tamat SMA	1.586804	.0533718	13.73	0.000	1.485571	1.694936
indeks_kesejahteraan						
Menengah	1.219146	.042968	5.62	0.000	1.137773	1.306339
Kaya	1.355858	.0469802	8.79	0.000	1.266835	1.451136
Terkaya	1.60665	.0546553	13.94	0.000	1.503021	1.717424
5	1.915346	.0696879	17.86	0.000	1.783516	2.056919
status_kehamilan						
Diinginkan	1.278733	.0450747	6.98	0.000	1.193371	1.370201
_cons	1.37307	.0625805	6.96	0.000	1.255734	1.501371

6. Lakukan evaluasi atas model yang telah terbentuk (*goodness of fit test*), apakah regresi Poisson tepat untuk digunakan pada data tersebut. Caranya, setelah mengeksekusi command di atas, ketik `estat gof`.

```
. estat gof
```

```
Deviance goodness-of-fit = 9330.297
Prob > chi2(4483)        = 0.0000
Pearson goodness-of-fit  = 9234.933
Prob > chi2(4483)        = 0.0000
```

Nilai chi-square di *goodness of fit test* yang besar adalah indikator bahwa distribusi poisson bukan pilihan yang baik untuk digunakan. Hasil uji statistik *goodness of fit test* juga menunjukkan hasil yang signifikan ($p < 0,05$). Artinya, model poisson tidak sesuai (tidak fit) untuk digunakan. Oleh karena itu, dilakukan prosedur analisis alternatifnya, yaitu menggunakan regresi binomial negatif.

Regresi binomial negatif lebih tepat digunakan dalam kasus overdispersi. Overdispersi adalah suatu kondisi dimana varians data lebih besar daripada mean-nya. Untuk membuktikan apakah data tersebut mengalami overdispersi, kita dapat melihat *summary statistics*-nya.

```
. sum jumlah_anc, detail
```

Jumlah Kunjungan ANC				
	Percentiles	Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	4,493
25%	1	0	Sum of Wgt.	4,493
50%	2		Mean	2.763632
		Largest	Std. Dev.	2.554263
75%	4	16		
90%	7	18	Variance	6.52426
95%	8	20	Skewness	1.183376
99%	10	20	Kurtosis	4.980571

Berdasarkan output tersebut, terbukti bahwa variabel dependen mengalami overdispersi yang terlihat dari nilai varians yang lebih besar dari pada mean. Oleh karena itu regresi Poisson kurang tepat digunakan, karena syarat regresi Poisson adalah data harus equidispersi (variens sama dengan mean-nya).

7. Karena data mengalami overdispersi, maka dilakukan analisis regresi binomial negatif sebagai alternatif regresi Poisson untuk data diskrit yang mengalami overdispersi. Adapun jika overdispersi disebabkan karena banyaknya pengamatan (variabel dependen) yang bernilai nol, maka alternatif yang digunakan adalah regresi *Zero-Inflated Poisson* (ZIP). Dalam kasus ini, Ketika melihat *summary statistics*, overdispersi bukan disebabkan karena banyaknya nilai nol. Oleh karena itu, regresi binomial negatif tepat untuk digunakan.

```
nbreg jumlah_anc i.kategori_wilayah i.tingkat_pendidikan
i.indeks_kesejahteraan i.status_kehamilan
```

```
Negative binomial regression      Number of obs   =    4,493
LR chi2(9)                       =    795.96
Dispersion   = mean              Prob > chi2     =    0.0000
Log likelihood = -9276.7067      Pseudo R2      =    0.0411
```

jumlah_anc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
kategori_wilayah Rural	-.1759234	.0305863	-5.75	0.000	-.2358714	-.1159754
tingkat_pendidikan Tamat SD	.1221176	.0398349	3.07	0.002	.0440426	.2001926
Tamat SMP	.2778605	.0408878	6.80	0.000	.1977219	.3579991
Tamat SMA	.4590862	.0486583	9.43	0.000	.3637177	.5544546
indeks_kesejahteraan Menengah	.1991268	.0460012	4.33	0.000	.1089661	.2892874
Kaya	.3014946	.0460732	6.54	0.000	.2111929	.3917964
Terkaya	.4720556	.0461367	10.23	0.000	.3816293	.562482
5	.6470757	.0509545	12.70	0.000	.5472067	.7469446
status_kehamilan Diinginkan	.2419496	.047498	5.09	0.000	.1488553	.3350438
_cons	.3248446	.0602474	5.39	0.000	.2067619	.4429274
/lnalpha	-.974219	.0482845			-1.068855	-.8795831
alpha	.3774871	.0182268			.3434015	.4149559

LR test of alpha=0: chibar2(01) = 1159.97 Prob >= chibar2 = 0.000

8. Transformasikan koefisien regresi binomial negatif ke dalam bentuk *Incidence Rate Ratio (IRR)*. Caranya, setelah mengeksekusi command di atas, ketik `nbreg, irr`.

```
Negative binomial regression      Number of obs   =    4,493
LR chi2(9)                       =    795.96
Dispersion   = mean              Prob > chi2     =    0.0000
Log likelihood = -9276.7067      Pseudo R2      =    0.0411
```

jumlah_anc	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
kategori_wilayah Rural	.8386823	.0256522	-5.75	0.000	.7898823	.8904972
tingkat_pendidikan Tamat SD	1.129887	.045009	3.07	0.002	1.045027	1.221638
Tamat SMP	1.320302	.0539842	6.80	0.000	1.218623	1.430464
Tamat SMA	1.582627	.0770079	9.43	0.000	1.438668	1.740991
indeks_kesejahteraan Menengah	1.220337	.0561369	4.33	0.000	1.115125	1.335476
Kaya	1.351878	.0622853	6.54	0.000	1.235151	1.479636
Terkaya	1.603287	.0739704	10.23	0.000	1.464669	1.755023
5	1.909947	.0973204	12.70	0.000	1.728418	2.110542
status_kehamilan Diinginkan	1.27373	.0604996	5.09	0.000	1.160505	1.398002
_cons	1.383816	.0833713	5.39	0.000	1.22969	1.557259
/lnalpha	-.974219	.0482845			-1.068855	-.8795831
alpha	.3774871	.0182268			.3434015	.4149559

LR test of alpha=0: chibar2(01) = 1159.97 Prob >= chibar2 = 0.000

Interpretasi:

1. Terdapat hubungan yang signifikan antara kategori wilayah tempat tinggal ibu hamil dengan jumlah ANC yang dilakukan. Ibu hamil yang tinggal di daerah perdesaan berpeluang lebih jarang melakukan ANC dibandingkan ibu hamil yang tinggal di perkotaan.
2. Terdapat hubungan yang signifikan antara tingkat pendidikan ibu hamil dengan jumlah ANC yang dilakukan. Semakin tinggi tingkat pendidikan ibu hamil, peluangnya untuk melakukan ANC lebih sering dibandingkan ibu hamil dengan pendidikan rendah.
3. Terdapat hubungan yang signifikan antara indeks kesejahteraan ibu hamil dengan jumlah ANC yang dilakukan. Semakin tinggi indeks kesejahteraan ibu hamil, peluangnya untuk melakukan ANC lebih sering dibandingkan ibu hamil dengan indeks kesejahteraan rendah.
4. Terdapat hubungan yang signifikan antara status kehamilan dengan jumlah ANC yang dilakukan ibu hamil. Ibu hamil yang kehamilannya diinginkan berpeluang lebih sering melakukan ANC dibandingkan ibu hamil yang kehamilannya tidak diinginkan.

PRAKTIKUM KE-6 RELIABILITAS DAN VALIDITAS PENGUKURAN

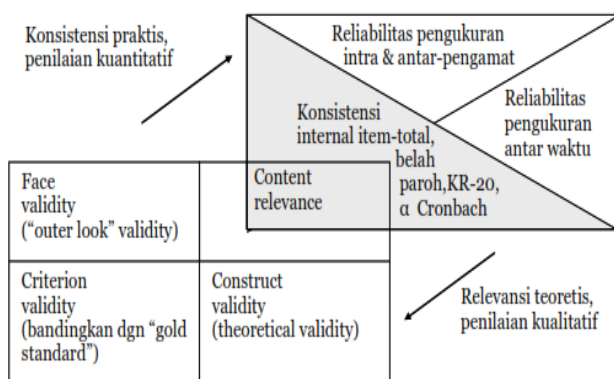
A. Tujuan Praktikum

1. Mahasiswa mampu mengidentifikasi tujuan penilaian reliabilitas dan validitas pengukuran.
2. Mahasiswa mampu melakukan penilaian reliabilitas dan validitas pengukuran menggunakan program Stata.
3. Mahasiswa mampu menginterpretasikan hasil penilaian reliabilitas dan validitas pengukuran menggunakan program Stata.

B. Pengantar

Pengukuran yang valid adalah pengukuran dari alat ukur yang dibuat) dengan metodologi yang benar dan implementasi pengukuran yang benar pula. Jika implementasi pengukuran benar, tetapi alat ukur tidak benar, maka hasil pengukuran juga tidak benar, menghasilkan kesalahan pengukuran yang disebut *measurement bias (measurement error)*. Demikian juga jika metodologi alat ukur benar, tetapi pelaksanaan pengukuran tidak benar (misalnya, asal-asalan), maka hasil pengukuran juga tidak benar.

Alat ukur (instrumen) yang baik harus mengukur dengan benar (valid) dan konsisten (andal, reliabel). Aspek validitas yang berhubungan dengan reliabilitas adalah validitas isi (*content validity*). Sedangkan aspek reliabilitas yang berkaitan dengan validitas adalah konsistensi internal antar item, konsistensi item-total, dan konsistensi belah paroh.



validitas pengukuran, di samping reliabilitas. Jika item pertanyaan

memberikan lebih dari dua alternative respons (jawaban), maka korelasi item-total dinilai dengan menggunakan korelasi Pearson. Jika item-item memberikan alternatif jawaban dikotomi, maka korelasi item-total dinilai dengan korelasi titik-biserial (*point-biserial correlation*).

C. Mengeksekusi dengan Program STATA

Untuk simulasi praktikum kali ini, pada kasus pertama, diasumsikan bahwa peneliti melakukan pengukuran reliabilitas dan validitas instrumen kepuasan pengguna layanan konseling remaja di suatu Puskesmas PKPR menggunakan 10 item pertanyaan dengan skala Likert (1-5 poin). Berikut adalah rekapitulasi pengukuran reliabilitas dan validitas instrumen yang dari 25 orang responden.

id	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	total
1	5	5	3	4	4	5	4	4	4	4	42
2	4	4	4	3	3	5	5	3	5	5	41
3	4	4	3	4	3	4	3	3	3	3	34
4	3	4	3	3	3	4	3	3	4	4	34
5	5	4	5	4	4	3	3	4	3	5	40
6	4	4	5	4	4	4	4	4	5	5	43
7	4	4	4	4	4	4	3	4	4	4	39
8	4	4	4	4	4	4	4	4	4	4	40
9	5	4	4	4	3	4	3	3	3	3	36
10	4	4	4	4	4	4	4	4	4	4	40
11	4	4	4	3	4	4	4	4	4	4	39
12	4	4	4	4	4	4	4	4	4	4	40
13	4	4	5	4	4	4	4	5	4	5	43
14	5	4	3	4	4	4	3	3	4	5	39
15	4	3	4	2	3	4	3	3	3	2	31
16	4	4	4	4	4	4	4	4	4	4	40
17	5	5	5	5	5	5	4	5	5	5	49
18	5	5	5	5	4	5	5	4	5	4	47
19	4	3	3	3	3	4	3	3	3	4	33
20	3	3	4	4	3	4	4	3	4	4	36
21	3	3	3	4	4	4	3	3	4	3	34
22	4	4	3	4	5	4	3	4	3	4	38
23	4	4	4	4	3	3	3	4	4	4	37
24	4	4	4	4	4	3	4	3	4	4	38
25	3	3	3	3	2	3	3	3	3	2	28

Prosedur:

1. Salin dataset tersebut ke Stata dengan cara *copy-paste* atau menggunakan dataset yang telah disediakan.
2. Lakukan penilaian konsistensi internal item-total sekaligus konsistensi belah paroh (menggunakan *Cronbach's alpha*) dengan command: `alpha q1-q10, item` atau menggunakan drop-down menu: **Statistics** → **Multivariate Analysis** → **Cronbach's Alpha** sehingga muncul *output* sebagai berikut.

Test scale = mean(unstandardized items)

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
q1	25	+	0.6132	0.5164	.2028241	0.8806
q2	25	+	0.8233	0.7773	.1903704	0.8642
q3	25	+	0.6783	0.5802	.1931019	0.8768
q4	25	+	0.7058	0.6256	.1948148	0.8731
q5	25	+	0.7285	0.6468	.1903241	0.8715
q6	25	+	0.5549	0.4605	.2100926	0.8837
q7	25	+	0.6810	0.5958	.1968981	0.8752
q8	25	+	0.7530	0.6839	.1913426	0.8691
q9	25	+	0.7394	0.6635	.190787	0.8703
q10	25	+	0.7563	0.6612	.1794444	0.8716
Test scale					.194	0.8849

3. Penilaian konsistensi internal item-total juga dapat dilakukan dengan mengkorelasikan masing-masing item dengan total dengan menggunakan command: `pwcorr q1-total, sig` sehingga muncul *output* sebagai berikut.

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	total
q1	1.000										
q2	0.701	1.000									
q3	0.380	0.378	1.000								
q4	0.444	0.633	0.391	1.000							
q5	0.437	0.566	0.336	0.598	1.000						
q6	0.338	0.505	0.099	0.224	0.314	1.000					
q7	0.098	0.010	0.636	0.283	0.127		1.000				

q7	0.182	0.474	0.516	0.300	0.262	0.559	1.000				
	0.385	0.017	0.008	0.145	0.206	0.004					
q8	0.380	0.604	0.623	0.526	0.674	0.226	0.344	1.000			
	0.061	0.001	0.001	0.007	0.000	0.277	0.092				
q9	0.121	0.521	0.486	0.427	0.366	0.542	0.756	0.385	1.000		
	0.564	0.008	0.014	0.034	0.072	0.005	0.000	0.058			
q10	0.393	0.513	0.470	0.445	0.551	0.258	0.430	0.516	0.587	1.000	
	0.052	0.009	0.018	0.026	0.004	0.214	0.032	0.008	0.002		
total	0.613	0.823	0.678	0.706	0.729	0.555	0.681	0.753	0.739	0.756	1.000
	0.001	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000

4. Lakukan rekapitulasi nilai *item-total (item-test) correlation* serta nilai Cronbach's Alpha ke dalam tabel khusus untuk memudahkan menginterpretasikan hasilnya. Berikut adalah tabel rekapitulasi hasil penilaian reliabilitas dan validitas untuk data latihan di atas.

No	Item Pertanyaan	<i>r</i> hitung (Hasil Item-Total/ Item-Test/ Pearson Correlation)	<i>r</i> hitung (Hasil Corrected Item- Total/ Item-Rest Correlation)	<i>r</i> tabel*	<i>p</i> -value	Ket.**
1	q1	0.613	0.516	0,396	0.001	Valid
2	q2	0.823	0.777	0,396	<0.001	Valid
3	q3	0.678	0.580	0,396	<0.001	Valid
4	q4	0.706	0.626	0,396	<0.001	Valid
5	q5	0.729	0.647	0,396	<0.001	Valid
6	q6	0.555	0.461	0,396	0.004	Valid
7	q7	0.681	0.596	0,396	<0.001	Valid
8	q8	0.753	0.684	0,396	<0.001	Valid
9	q9	0.739	0.664	0,396	<0.001	Valid
10	q10	0.756	0.661	0,396	<0.001	Valid
Nilai Cronbach's Alpha			0.885	0,396		Reliabel

* *r* tabel dilihat dari tabel nilai kritis *r* Pearson pada $df=n-2$; $\alpha=0.05$

** dinyatakan valid jika *r* hitung > *r* tabel atau $p < 0.05$

** dinyatakan reliabel jika nilai Cronbach's Alpha > *r* hitung atau > 0.60

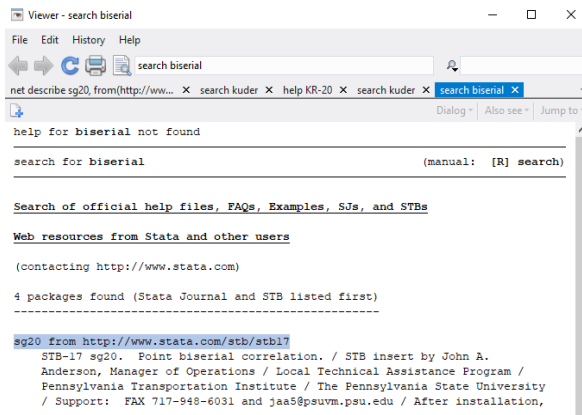
Pada simulasi kasus kedua, diasumsikan bahwa peneliti melakukan pengukuran reliabilitas dan validitas instrumen untuk mengukur pengetahuan wanita usia subur (WUS) menggunakan 10 item pertanyaan tentang fakta jenis metode

kontrasepsi modern yang diketahui, dengan pilihan jawaban dikotomi/ nominal (1=benar dan 0=salah). Berikut adalah rekapitulasi pengukuran reliabilitas dan validitas instrumen yang dari 20 orang responden.

id	k1	k2	k3	k4	k5	k6	k7	k8	k9	k10	total
1	1	1	1	1	1	1	0	1	1	1	9
2	1	1	1	1	1	1	1	1	1	1	10
3	1	1	1	1	1	1	1	1	1	1	10
4	1	1	1	1	0	1	1	1	1	1	9
5	1	1	1	1	0	1	1	1	1	0	8
6	1	1	1	1	1	1	1	1	1	1	10
7	1	1	1	0	1	1	0	1	1	1	8
8	1	1	1	1	0	1	1	1	1	1	9
9	1	1	1	1	0	1	1	0	1	1	8
10	1	1	0	1	1	1	0	1	1	1	8
11	1	1	0	1	1	1	1	0	1	1	8
12	1	1	1	1	1	1	1	0	1	0	8
13	1	1	1	1	0	1	0	1	1	0	7
14	1	1	1	0	1	0	1	1	1	1	8
15	1	1	0	1	0	0	0	1	1	1	6
16	0	1	1	1	0	1	0	0	1	0	5
17	0	1	1	0	0	0	0	0	1	0	3
18	1	0	1	0	0	0	0	1	1	0	4
19	0	1	0	0	0	1	0	0	0	1	3
20	1	0	1	0	0	0	0	0	0	0	2

Prosedur:

1. Salin dataset tersebut ke Stata dengan cara *copy-paste* atau menggunakan dataset yang telah disediakan.
2. Lakukan penilaian konsistensi internal item-total menggunakan korelasi titik-biserial (*point-biserial correlation*). Namun, untuk menjalankan *command*-nya, terlebih dahulu harus dilakukan instalasi karena *by default*, jenis analisis ini tidak disediakan langsung di STATA. Caranya, jalankan *command* `help biserial`.



Maka akan muncul jendela bantuan (*help*) seperti di samping. Instalasi *package* korelasi titik-biserial dapat dilakukan dengan cara klik tautan/*link* pada hasil pencarian teratas (lihat gambar di samping dengan *highlight* biru), lalu klik petunjuk "*click here to install*".

3. Setelah *package point-biserial correlation* ter-install, jalankan *command* berikut: `pbis k1 total` sejumlah item yang ada (k_1 sampai dengan k_{10}) dengan cara menyalin *command* tersebut dan mengganti k_1 dengan k_2, k_3, \dots, k_{10} . Berikut adalah *output* apabila *command* tersebut telah dijalankan.

```
. pbis k1 total

(obs= 20)
Np= 17  p= 0.85
Nq= 3   q= 0.15
-----+-----+-----+-----+
Coef.= 0.5908          t= 3.1068          P>|t| = 0.0061          df= 18

. pbis k2 total

(obs= 20)
Np= 18  p= 0.90
Nq= 2   q= 0.10
-----+-----+-----+-----+
Coef.= 0.5585          t= 2.8568          P>|t| = 0.0105          df= 18

. pbis k3 total

(obs= 20)
Np= 16  p= 0.80
Nq= 4   q= 0.20
-----+-----+-----+-----+
Coef.= 0.1817          t= 0.7839          P>|t| = 0.4433          df= 18

. pbis k4 total

(obs= 20)
Np= 14  p= 0.70
Nq= 6   q= 0.30
-----+-----+-----+-----+
Coef.= 0.6564          t= 3.6914          P>|t| = 0.0017          df= 18

. pbis k5 total

(obs= 20)
Np= 9   p= 0.45
Nq= 11  q= 0.55
-----+-----+-----+-----+
Coef.= 0.5945          t= 3.1366          P>|t| = 0.0057          df= 18

. pbis k6 total
```

```
(obs= 20)
Np= 15  p= 0.75
Nq= 5  q= 0.25
-----+-----+-----+-----+
Coef.= 0.5944          t= 3.1362          P>|t| = 0.0057          df= 18

. pbis k7 total

(obs= 20)
Np= 10  p= 0.50
Nq= 10  q= 0.50
-----+-----+-----+-----+
Coef.= 0.6662          t= 3.7900          P>|t| = 0.0013          df= 18

. pbis k8 total

(obs= 20)
Np= 13  p= 0.65
Nq= 7  q= 0.35
-----+-----+-----+-----+
Coef.= 0.5523          t= 2.8111          P>|t| = 0.0116          df= 18

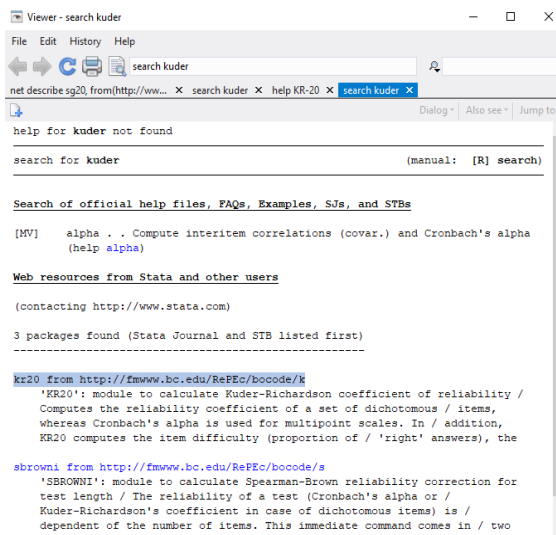
. pbis k9 total

(obs= 20)
Np= 18  p= 0.90
Nq= 2  q= 0.10
-----+-----+-----+-----+
Coef.= 0.6258          t= 3.4042          P>|t| = 0.0032          df= 18

. pbis k10 total

(obs= 20)
Np= 13  p= 0.65
Nq= 7  q= 0.35
-----+-----+-----+-----+
Coef.= 0.5523          t= 2.8111          P>|t| = 0.0116          df= 18
```

4. Berikutnya, lakukan prosedur penilaian konsistensi belah paroh menggunakan pendekatan Kuder-Richardson. Namun, untuk menjalankan *command*-nya,



terlebih dahulu harus dilakukan instalasi karena *by default*, jenis analisis ini tidak disediakan langsung di STATA. Caranya, jalankan *command* `help kuder`. Maka akan muncul jendela bantuan (*help*) seperti di samping. Instalasi *package* penilaian konsistensi belah paroh menggunakan pendekatan Kuder-Richardson dapat dilakukan dengan

cara klik tautan/ link pada hasil pencarian teratas (lihat gambar di samping dengan highlight biru), lalu klik petunjuk “*click here to install*”.

5. Setelah *package* Kuder-Richardson (KR-20) ter-*install*, jalankan *command* berikut: `kr20 k1-k10`. Berikut adalah output apabila *command* tersebut telah dijalankan.

```
Kuder-Richarson coefficient of reliability (KR-20)
```

```
Number of items in the scale = 10
```

```
Number of complete observations = 20
```

Item	Obs	Item difficulty	Item variance	Item-rest correlation
k1	20	0.8500	0.1275	0.4866
k2	20	0.9000	0.0900	0.4681
k3	20	0.8000	0.1600	0.0205
k4	20	0.7000	0.2100	0.5336
k5	20	0.4500	0.2475	0.4426
k6	20	0.7500	0.1875	0.4652
k7	20	0.5000	0.2500	0.5327
k8	20	0.6500	0.2275	0.3985
k9	20	0.9000	0.0900	0.5455
k10	20	0.6500	0.2275	0.3985
Test		0.7150		0.4292

```
KR20 coefficient is 0.7819
```

6. Prosedur penilaian konsistensi belah paroh juga dapat dilakukan menggunakan pendekatan *Cronbach's Alpha*. *Cronbach's Alpha* merupakan perluasan KR-20, dapat digunakan untuk item-item dengan respons dikotomi atau lebih. Jika *Cronbach's Alpha* digunakan untuk item-item dikotomi, maka hasilnya identik dengan KR-20. Jalankan *command* `alpha k1-k10, item,` maka akan muncul output sebagai berikut.

Test scale = mean(unstandardized items)

Item	Obs	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
k1	20	+	0.6062	0.4992	.047076	0.7378
k2	20	+	0.5730	0.4803	.0491228	0.7430
k3	20	+	0.1864	0.0211	.0580409	0.7936
k4	20	+	0.6735	0.5475	.0429825	0.7273
k5	20	+	0.6099	0.4540	.0444444	0.7419
k6	20	+	0.6099	0.4773	.0454678	0.7381
k7	20	+	0.6835	0.5465	.0418129	0.7267
k8	20	+	0.5667	0.4088	.0461988	0.7484
k9	20	+	0.6421	0.5597	.0476608	0.7353
k10	20	+	0.5667	0.4088	.0461988	0.7484
Test scale					.0469006	0.7646

7. Lakukan rekapitulasi nilai *item-total (item-test) correlation* dari hasil *point-biserial correlation* serta nilai Kuder-Richardson (KR-20) dan Cronbach's Alpha ke dalam tabel khusus untuk memudahkan menginterpretasikan hasilnya. Berikut adalah tabel rekapitulasi hasil penilaian reliabilitas dan validitas untuk data latihan di atas.

No	Item Pertanyaan	<i>r</i> hitung (Hasil Item-Total/ Point-Biserial)	<i>r</i> tabel*	<i>p</i> -value	Keterangan**
1	k1	0.591	0.444	0.006	Valid
2	k2	0.559	0.444	0.011	Valid
3	k3	0.182	0.444	0.443	Tidak Valid
4	k4	0.656	0.444	0.002	Valid
5	k5	0.595	0.444	0.006	Valid
6	k6	0.594	0.444	0.006	Valid
7	k7	0.666	0.444	0.001	Valid
8	k8	0.552	0.444	0.012	Valid
9	k9	0.626	0.444	0.003	Valid
10	k10	0.552	0.444	0.012	Valid
Nilai Kuder-Richardson		0.782	0.444		Reliabel
Nilai Cronbach's Alpha		0.765	0.444		Reliabel

* *r* tabel dilihat dari tabel nilai kritis *r* Pearson pada $df=n-2$; $\alpha=0.05$ ** dinyatakan valid jika *r* hitung > *r* tabel atau *p*-value < 0.05** dinyatakan reliabel jika nilai KR-20 dan Cronbach's Alpha > *r* hitung atau > 0.60

Lampiran. Tabel Nilai Kritis r Pearson

df = n - 2	Tingkat signifikansi (two-tailed)			
	0,05	0,025	0,01	0,005
	Tingkat signifikansi (one-tailed)			
	0,10	0,05	0,02	0,01
1	0,988	0,997	0,9995	0,9999
2	0,900	0,950	0,980	0,990
3	0,805	0,878	0,934	0,959
4	0,729	0,811	0,882	0,917
5	0,669	0,754	0,833	0,874
6	0,622	0,707	0,789	0,834
7	0,582	0,666	0,750	0,798
8	0,549	0,632	0,716	0,765
9	0,521	0,602	0,685	0,735
10	0,497	0,576	0,658	0,708
11	0,476	0,553	0,634	0,684
12	0,458	0,532	0,612	0,661
13	0,441	0,514	0,592	0,641
14	0,426	0,497	0,574	0,623
15	0,412	0,482	0,558	0,606
16	0,400	0,468	0,542	0,590
17	0,389	0,456	0,528	0,575
18	0,378	0,444	0,516	0,561
19	0,369	0,433	0,503	0,549
20	0,360	0,423	0,492	0,537
21	0,352	0,413	0,482	0,526
22	0,344	0,404	0,472	0,515
23	0,337	0,396	0,462	0,505
24	0,330	0,388	0,453	0,496
25	0,323	0,381	0,445	0,487
26	0,317	0,374	0,437	0,479
27	0,311	0,367	0,430	0,471
28	0,306	0,361	0,423	0,463
29	0,301	0,355	0,416	0,456
30	0,296	0,349	0,409	0,449
35	0,275	0,325	0,381	0,418
40	0,257	0,304	0,358	0,393
45	0,243	0,288	0,338	0,372
50	0,231	0,273	0,322	0,354
60	0,211	0,250	0,295	0,325
70	0,195	0,232	0,274	0,302
80	0,183	0,217	0,256	0,283
90	0,173	0,205	0,242	0,267
100	0,164	0,195	0,230	0,254

PRAKTIKUM KE-7 PENYAJIAN HASIL ANALISIS DATA

A. Tujuan Praktikum

1. Mahasiswa mampu menjelaskan prinsip penyajian hasil analisis data yang efektif dan efisien.
2. Mahasiswa mampu menyajikan hasil analisis dalam bentuk tabel/ grafik secara efektif dan efisien.

B. Pengantar

Kita semua pernah mendengar pepatah lama: "sebuah gambar bernilai seribu kata". Salah satu teknik terbaik untuk memahami data adalah menyajikan angka ke dalam bentuk visual. Hal ini dapat memudahkan melihat suatu pola atau dapat mengekspos pola yang mungkin tersembunyi. Data dapat divisualisasikan ke dalam berbagai cara, mulai dari diagram batang sederhana hingga hingga diagram pencar yang kompleks, tabel, peta, dan sebagainya. Penyajian data ke dalam bentuk visual cukup penting karena pembaca akan lebih mudah memahami presentasi visual. Namun, perlu diingat bahwa penyajian statistik secara visual yang buruk dapat memberikan informasi yang *misleading*, *misunderstanding*, dan *misinterpretations*.

Meskipun saat ini telah banyak revolusi dalam metode visualisasi data melalui bantuan beberapa perangkat lunak dan teknologi, penting bagi analis data untuk menghindari presentasi data yang terlalu berlebihan yang justru merusak konten. Ingatlah bahwa teknologi hanyalah pelayan. Anda tidak boleh menambahkan sesuatu yang tidak berguna dan elemen yang tidak jelas (cenderung berlebihan) hanya karena Anda bisa. Buatlah pesan sederhana untuk pembaca melalui visualisasi data.

1. Visualisasi Data dalam Bentuk Tabel

Tabel yang baik adalah yang melengkapi teks. Tabel harus menyajikan angka secara ringkas dan terorganisir dengan baik untuk mendukung analisis. Tabel

berfungsi untuk membantu meminimalkan angka dalam cerita statistik (narasi). Satu tempat desimal akan cukup untuk sebagian besar data. Namun, dalam kasus tertentu, dua atau lebih tempat desimal mungkin diperlukan untuk menggambarkan perbedaan dalam distribusi yang lebih detail

Visualisasi data dalam tabel diurutkan menurut kriteria tertentu sehingga informasi lebih mudah dicerna, misalnya dengan cara mengurutkan dari yang tertinggi sampai ke yang terendah. Tampilkan data angka dengan rata kanan. Tabel harus mampu berdiri sendiri, baik ketika diterbitkan dalam laporan, artikel, atau halaman web. Setiap tabel harus mengandung metadata yang cukup, seperti judul dan sumbernya.

Untuk memastikan bahwa tabel Anda mudah dimengerti, Anda harus mempertimbangkan pedoman berikut.

- Hindari teks yang tidak perlu.
- Tampilkan data Anda baik dengan urutan kronologis atau dengan menggunakan beberapa klasifikasi standar. Untuk rangkaian waktu, lebih tepat untuk menggunakan urutan kronologis terbalik (misalnya dimulai dengan periode terbaru dan mundur).
- Gunakan tempat desimal minimum (misal satu angka di belakang koma).
- Gunakan pemisah ribuan, misal titik atau koma atau spasi (menghindari masalah perbedaan antarbahasa).
- Sejajarkan angka pada titik desimal (atau rata/ di sebelah kanan jika tidak ada angka desimal) sehingga jelas. Hindari menuliskan angka di tengah sebuah kolom, kecuali semua besarnya sama.
- Jangan biarkan sel dengan data kosong. Nilai yang hilang harus diidentifikasi sebagai "tidak tersedia" atau "tidak berlaku". Singkatan "NA" dapat berlaku untuk keduanya, sehingga perlu didefinisikan.

GOOD EXAMPLE**Share of total energy consumption, by sector (in percent)**

Ireland, 1980-2003

	1980	1985	1990	2000	1995	2002	2003
Transport	27.8	27.9	28.2	31.1	36.8	39.5	39.1
Residential	31.1	33.9	30.4	27.6	24.3	23.7	24.0
Industry	31.5	27.2	23.9	22.1	21.4	19.5	18.8
Agriculture	n/a ¹	n/a ¹	3.5	3.7	3.1	2.9	2.8
Services	9.6	11.0	14.0	15.5	14.4	14.4	15.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

¹ Data on energy consumption for the agricultural sector was not collected until 1990.

Source: Department of Public Enterprise, Ireland

2. Visualisasi Data dalam Bentuk Grafik

Statistik seringkali dapat lebih dipahami ketika disajikan dalam grafik daripada dalam bentuk tabel. Grafik merupakan representasi visual dari data statistik, di mana data diwakili oleh simbol seperti batang (bar) atau garis. Grafik merupakan alat visualisasi yang sangat efektif untuk menampilkan data dengan cepat dan mudah, memfasilitasi perbandingan dan dapat mengungkapkan tren serta hubungan dalam data.

Grafik umumnya berbentuk figur satu atau dua dimensi, seperti grafik batang atau grafik garis. Meskipun terdapat grafik tiga dimensi, jenis ini biasanya dianggap terlalu rumit untuk dipahami. Bagan dapat digunakan untuk menggambarkan pola dalam sejumlah besar data atau untuk mengomunikasikan temuan atau pesan utama. Grafik dipertimbangkan untuk digunakan jika ingin menunjukkan:

- Perbandingan: Berapa banyak? Item mana yang lebih besar atau lebih kecil?
- Perubahan dari waktu ke waktu: Bagaimana suatu variabel berkembang?
- Distribusi frekuensi: Bagaimana distribusi suatu item? Apa perbedaannya?
- Korelasi: Apakah dua variabel berhubungan?
- Bagian dari total keseluruhan: Bagaimana satu item dibandingkan dengan total?

Jenis grafik yang akan dibuat tergantung dari jenis datanya. Misalnya, grafik garis digunakan untuk data *time series* (perkembangan dari waktu ke waktu), atau grafik batang untuk data kategorikal. Untuk memastikan Anda tidak memuat terlalu banyak hal ke dalam grafik, tuliskan kalimat topik untuk grafik tersebut. Grafik dapat diperjelas dengan cara:

- Menggunakan padatan (*solid*) daripada pola (*pattern*) untuk gaya garis dan isi;
- Menghindari penanda titik data pada grafik garis;
- Menggunakan nilai data pada grafik jika tidak mengganggu kemampuan pembaca untuk melihat gambaran utama;
- Memulai skala sumbu Y pada nol;
- Menggunakan hanya satu unit pengukuran per grafik;
- Menggunakan desain dua dimensi untuk data dua dimensi;
- Membuat semua teks pada grafik mudah dimengerti;
- Tidak menggunakan singkatan;
- Menghindari akronim;
- Menulis label dari kiri ke kanan;
- Menggunakan tata bahasa yang tepat;
- Menghindari legenda kecuali di peta.

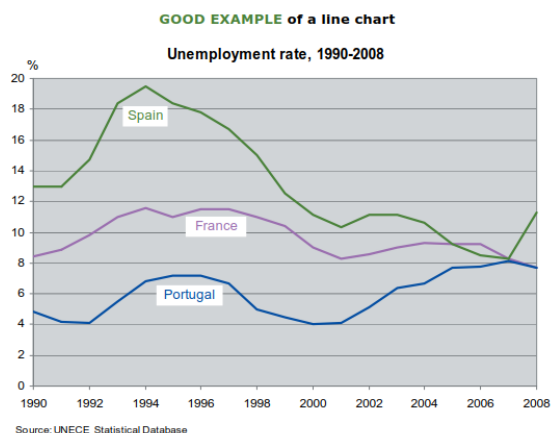
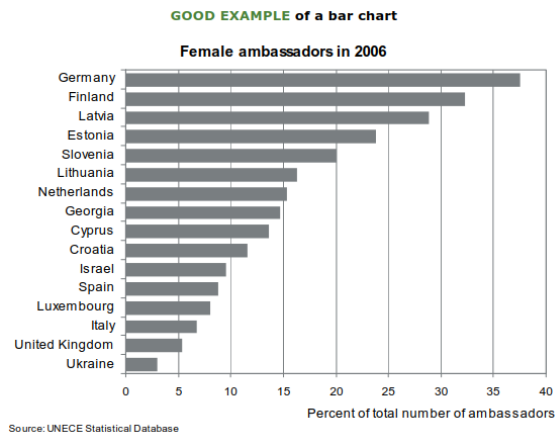
Jenis Grafik:

Grafik batang : membandingkan frekuensi atau nilai untuk kategori atau grup.

Grafik garis : memvisualisasikan tren dalam data dari waktu ke waktu

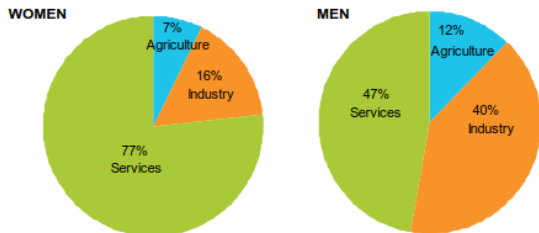
Grafik lingkaran : untuk menunjukkan distribusi persentase satu variable

Scatter plot : untuk menunjukkan hubungan antara dua variable



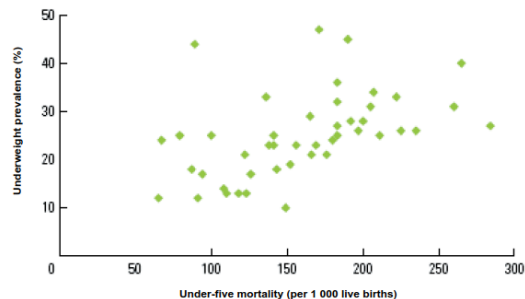
GOOD EXAMPLE of a pie chart

Employment by major sectors in Latvia, 2007



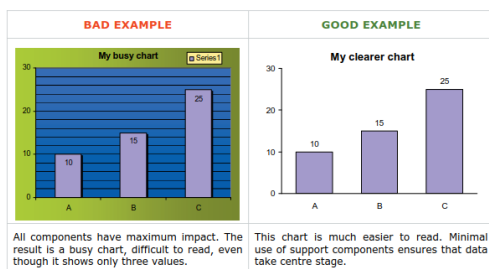
Source: UNECE Statistical Database

GOOD EXAMPLE of a scatter plot
Under-five mortality and underweight prevalence
In Sub-Saharan African countries, 2003

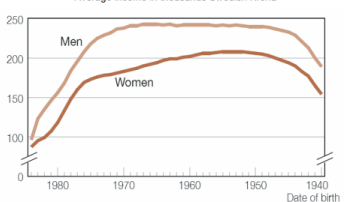


Source: Jamison et al. (2006) *Disease and Mortality in Sub-Saharan Africa*, 2nd edition, Washington D.C., The World Bank.

Beberapa kesalahan dalam menyajikan data dalam bentuk visual sehingga rawan menyebabkan *misleading*, *misunderstanding*, atau *misinterpretations*:

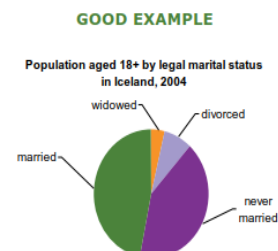
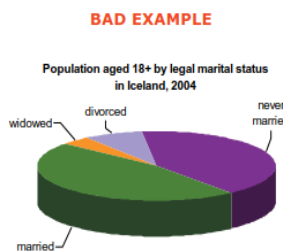
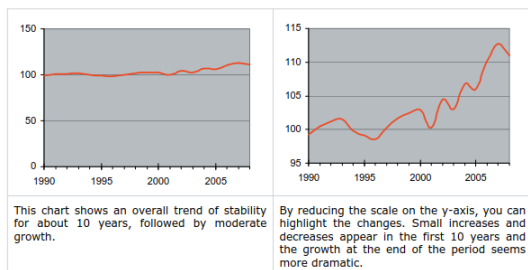
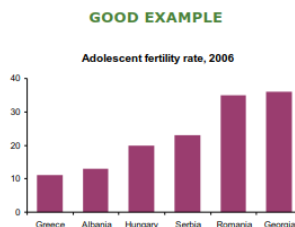
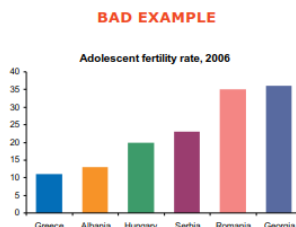
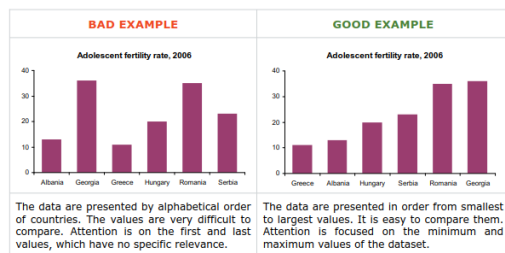
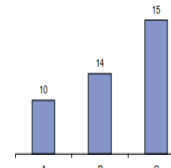


GOOD EXAMPLE of a chart with a y-axis not starting at zero
Pensionable income for persons aged 20-64 in Sweden, 2004
Average income in thousands Swedish Krona



Source: Statistics Sweden (2006), *Women and Men in Sweden: Facts and figures 2006*.

BAD EXAMPLE of relative size between chart objects



C. Mengeksekusi dengan Program Stata

1. Menyajikan Hasil Analisis dengan Grafik

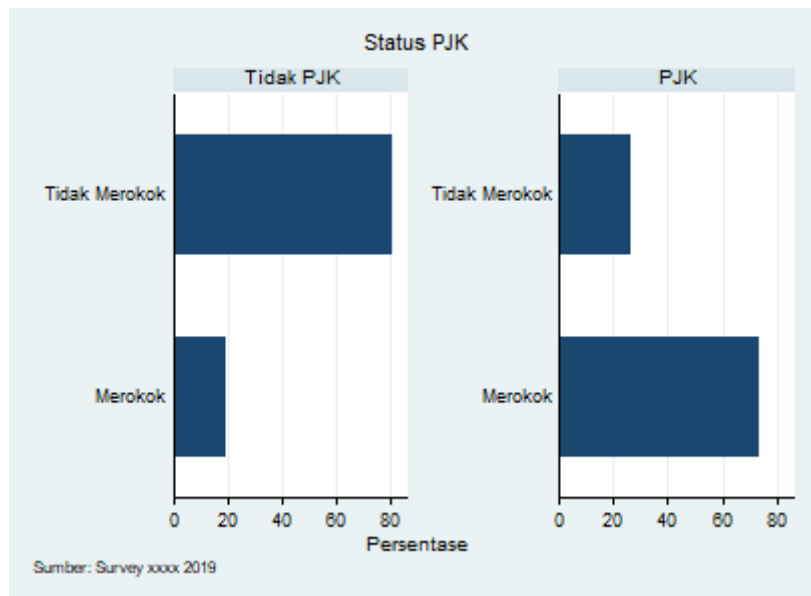
Selain menggunakan tabel, grafik dapat dipilih untuk dijadikan sebagai alat visualisasi hasil analisis data. Grafik berfungsi untuk menampilkan suatu informasi yang ingin ditonjolkan (*highlight*) dalam suatu dokumen. Salah satu jenis grafik yang paling sederhana adalah grafik batang, tetapi menjadi tidak sederhana lagi ketika data yang dianalisis adalah data kategorik, apalagi untuk frekuensi dari satu atau lebih variabel kategorik. Jika kita menggunakan apa yang tersedia di Stata resmi, pendekatan yang dapat digunakan adalah menggunakan `histogram` (yang sebenarnya lebih tepat untuk data numerik) atau `graph bar` atau `graph hbar`. Namun, pendekatan ini seringkali menghasilkan output yang terkesan “aneh” atau dipaksakan. Selain itu, pendekatan tersebut juga memerlukan *adjustment* atau persiapan tertentu.

Untuk mengatasi hal tersebut dapat digunakan *command* `catplot`. Tujuannya adalah agar grafik frekuensi tersedia melalui satu perintah, tanpa perlu persiapan atau restrukturisasi dataset. Sebelumnya, kita dapat menginstal `catplot` dari *Statistical Software Components* (SSC) menggunakan perintah `ssc install catplot`. Secara lebih detil dapat dipelajari dalam: Cox, Nicholas J. 2004. *Speaking Stata: Graphing categorical and compositional data*. *The Stata Journal* (2004) 4, Number 2, pp. 190–215 (<https://journals.sagepub.com/doi/pdf/10.1177/1536867X0400400209>).

Misalnya kita ingin melaporkan distribusi status PJK menurut status merokok menggunakan diagram batang. Masih menggunakan dataset sebelumnya, syntax yang dieksekusi menggunakan program Stata adalah sebagai berikut.

```
catplot rokok, ylabel(, angle(h)) percent(pjk) by(pjk,
note("Sumber: Survey xxxx 2019") subtitle(Status PJK))
```

Output syntax tersebut adalah sebagai berikut.



2. Menyajikan Hasil Analisis dengan Tabel

Menyajikan data/ hasil analisis statistik ke dalam bentuk tabel juga harus mengikuti prosedur, khususnya apabila hasil analisis akan dipublikasikan ke dalam suatu penerbit ilmiah (jurnal). Beberapa rambu-rambu umum yang harus diperhatikan untuk melaporkan hasil analisis statistik ke dalam tabel adalah sebagai berikut.

- Untuk variabel numerik, jika data berdistribusi normal, maka ukuran pemusatan yang dilaporkan adalah mean dan standar deviasi. Sebaliknya, jika data tidak mengikuti distribusi normal, maka yang dilaporkan adalah nilai median dan *inter-quartile ranges* (IQR).
- Untuk variabel kategorik, yang dilaporkan adalah jumlah (n) dan persentase (%). Selain itu, juga harus dilaporkan jumlah/ total sampelnya. Gunakan “n” untuk menyatakan jumlah sampel. “N” adalah symbol untuk populasi.
- Laporkan juga *effect size* dan *confidence interval*. Hindari hanya melaporkan signifikansi statistik (*p-value*). Misalnya, untuk uji beda, *effect size* yang dilaporkan adalah *mean differences*. Untuk asosiasi (data kategorik), yang dilaporkan adalah OR/ RR atau χ^2 . Laporkan dengan dua angka decimal.

- d. *Confodence interval* dilaporkan dengan dua angka decimal, sedangkan *p-value* disajikan dengan tiga angka decimal. Untuk *-p-value* yang bernilai “0.000”, dilaporkan menjadi “<0.001”.
- e. Sajikan seluruh angka statistic dengan rata kanan (*align right*).

Contoh melaporkan hasil analisis deskriptif:

Table 1. Sociodemographic Characteristics of Respondents (n=1,237)

Variable	n	%
Sex		
Male	141	11.4
Female	1096	88.6
Age Category		
15-34	341	27.6
35-49	896	72.4
Age of First Marriage		
Under 21	346	28.0
21+	891	72.0
Education Attainment		
Primary/ Never Attended	161	13.0
Secondary-1	248	20.0
Secondary-2	619	50.0
University	209	16.9
Parity		
0-1 Child	378	30.6
2 Children	595	48.1
3+ Children	264	21.3
Working Status		
Housewife	590	47.7
Formal Sectors	246	19.9
Informal Sectors	401	32.4

Table 2. Percentage Distribution of Respondents by Contraception Types, Health Insurance, and Payment for Contraceptive Services (n=1,237)

Variable	n	%
Modern Method Used		
LARC-Sterilization	479	38.7
Non-LARC	758	61.3
Type of Health Insurance Owned		
Public-Subsidized	763	61.7
Public-Non-Subsidized	419	33.9
Private Insurance	55	4.4
Payment Category		
Free	329	26.6
OOP Payment	908	73.4

Contoh melaporkan hasil tabulasi silang:

Table 3. Distribution of Respondents by Out-of-Pocket Expenditure of Contraception Services (n=1,237)

Variable	Out-of-Pocket Expenditure					χ^2	p-value
	No		Yes		Total		
	n	%	n	%	n		
Sex							
Male	17	12.1	124	87.9	141	17.23	<0.001
Female	312	28.5	784	71.5	1,096		
Age Category							
15-34	92	27.0	249	73.0	341	0.04	0.851
35-49	237	26.5	659	73.5	896		
Age of First Marriage							
Under 21	106	30.6	240	69.4	346	4.01	0.045
21+	223	25.0	668	75.0	891		
Education Attainment							
Primary/Never Attended	55	34.2	106	65.8	161	11.54	0.009
Secondary-1	78	31.5	170	68.5	248		
Secondary-2	149	24.1	470	75.9	619		
University	47	22.5	162	77.5	209		
Parity							
0-1 Child	71	18.8	307	81.2	378	26.83	<0.001
2 Children	160	26.9	435	73.1	595		
3+ Children	98	37.1	166	62.9	264		
Working Status							
Housewife	168	28.5	422	71.5	590	3.79	0.151
Formal Sectors	54	22.0	192	78.0	246		
Informal Sectors	107	26.7	294	73.3	401		
Classification of Modern Method							
LARC-Sterilization	223	46.6	256	53.4	479	159.50	<0.001
Non-LARC	106	14.0	652	86.0	758		
Type of Insurance							
Public-Subsidized	248	32.5	515	67.5	763	35.60	<0.001
Public-Non-Subsidized	72	17.2	347	82.8	419		
Private Insurance	9	16.4	46	83.6	55		

Contoh melaporkan ukuran asosiasi (Odds Ratio/ Relative Risk):

Variable	Professional ANC Status			Odds Ratio (95% CI)	
	Incomplete %	Complete %	Total n	Unadjusted	Adjusted
Pregnancy status					
Intended pregnancy	16.9	83.1	9303	1.00	1.00
Unintended pregnancy	27.9	72.1	1812	1.89** [1.69-2.13]	1.85** [1.63-2.10]
Age category					
20-34	18.1	81.9	7273	1.00	1.00
15-19	37.4	62.6	174	2.70** [1.98-3.69]	1.87** [1.35-2.60]
35-49	19.1	80.9	3668	1.07 [0.97-1.18]	0.86* [0.77-0.98]
Educational attainment					
Higher	14.0	86.0	5317	1.00	1.00
Lower	23.1	76.9	5798	1.85** [1.68-2.04]	1.24** [1.11-1.38]
Respondent currently working					
Yes	17.3	82.7	5672	1.00	1.00
No	20.2	79.8	5434	1.22** [1.10-1.34]	1.14** [1.03-1.26]
Wealth index					
Higher	11.8	88.2	3883	1.00	1.00
Middle-Lower	22.4	77.6	7232	2.16** [1.93-2.42]	1.93** [1.71-2.17]
Number of living children					
Two or less	16.2	83.8	7412	1.00	1.00
More than two	23.7	76.3	3703	1.60** [1.46-1.77]	1.43** [1.27-1.62]
Age at first birth					
Ideal age of pregnancy	15.9	84.1	7451	1.00	1.00
Risky age of pregnancy	24.4	75.6	3664	1.71** [1.55-1.88]	1.33** [1.19-1.48]
Type of place of residence					
Urban	15.0	85.0	5624	1.00	1.00
Rural	22.5	77.5	5491	1.65** [1.50-1.81]	1.57** [1.41-1.73]

Exponentiated coefficients; 95% confidence intervals in brackets

** $p < 0.05$, ** $p < 0.01$*

Contoh melaporkan ukuran asosiasi dengan banyak model (regresi):

Table 4. Structural Model and Odds Ratio of Factors Related to Out-of-Pocket Expenditures for Contraception Services in Sleman Regency, 2016

Predictors	Model 1	Model 2	Full Model
	AOR [95% CI]	AOR [95% CI]	AOR [95% CI]
Sex			
Male	1		1
Female	0.51* [0.28 - 0.96]		1.01 [0.52 - 1.98]
Age Category			
15-34	1		1
35-49	1.20 [0.87 - 1.65]		1.18 [0.84 - 1.67]
Age of First Marriage			
Under 21	1		1
21+	0.97 [0.71 - 1.31]		0.99 [0.71 - 1.38]
Education Attainment			
Primary/Never Attended	1		1
Secondary-1	1.10 [0.72 - 1.69]		1.15 [0.72 - 1.85]
Secondary-2	1.46 [0.98 - 2.17]		1.38 [0.89 - 2.16]
University	1.66 [0.98 - 2.80]		1.75 [0.95 - 3.21]
Parity			
0-1 Child	1		1
2 Children	0.76 [0.52 - 1.10]		0.85 [0.57 - 1.27]
3+ Children	0.47** [0.31 - 0.72]		0.66 [0.41 - 1.05]
Working Status			
Housewife	1		1
Formal Sectors	0.93 [0.62 - 1.39]		0.59* [0.37 - 0.94]
Informal Sectors	0.95 [0.71 - 1.28]		0.90 [0.65 - 1.25]
Classification of Modern Method			
LARC-Sterilization		1	1
Non-LARC		6.38** [4.79 - 8.50]	6.34** [4.66 - 8.62]
Type of Insurance			
Public-Subsidized		1	1
Public-Non-Subsidized		3.12** [2.25 - 4.30]	3.10** [2.12 - 4.53]
Private Insurance		3.47** [1.60 - 7.52]	3.42** [1.51 - 7.71]
Observations	1237	1237	1237
Pseudo R-squared	0.029	0.150	0.158
AIC	1413.5	1225.3	1234.4
BIC	1469.9	1245.8	1306.1

AOR: Adjusted Odds Ratio; AIC: Akaike's Information Criterion; BIC: Bayesian Information Criterion
Exponentiated coefficients; 95% confidence intervals in brackets; * $p < 0.05$, ** $p < 0.01$

Catatan Tambahan:**Melaporkan Hasil Regresi dengan Banyak Model.**

Misalnya kita ingin melaporkan hasil analisis regresi logistik berganda yang telah kita lakukan dalam pertemuan sebelumnya (analisis multivariabel). Namun, pelaporan hasil analisis regresi ke dalam tabel membutuhkan usaha yang tidak mudah mengingat *output* yang dihasilkan dari analisis regresi banyak dan mengakomodir beberapa model yang dibentuk. Di sisi lain, kita diharuskan melaporkan hasil analisis yang informatif serta efektif dan efisien sesuai dengan kaidah pelaporan.

Terkait dengan hal tersebut, Stata mengakomodir penyajian data yang informatif tanpa melakukan proses editing secara manual dengan bantuan *Statistical Software Components* (SSC) khusus yang belum terpasang by default di Stata. Salah satu contoh yang akan dipraktikkan di pertemuan ini adalah membuat tabel pelaporan analisis regresi menggunakan `ssc estout`. Secara lebih detil dapat dipelajari dalam: Jann, Ben. 2007. Making regression tables simplified. *The Stata Journal* (2007) 7, Number 2, pp. 227–244 (https://www.stata-journal.com/sjpdf.html?articlenum=st0085_1).

Misalnya kita gunakan dataset pertemuan sebelumnya tentang analisis regresi logistik berganda (Dataset_Latihan3&4.dta).

Syntax:

```
ssc install estout → instalasi estout package (online)
logistic pjk i.jenis_kelamin
estimates store est1
logistic pjk i.usia
estimates store est2
logistic pjk i.rokok
estimates store est3
logistic pjk i.bmi
estimates store est4
eststo est5: logistic pjk i.jenis_kelamin i.usia i.rokok
```

```
eststo est6: logistic pjk i.jenis_kelamin i.usia i.rokok i.bmi
esttab est1 est2 est3 est4 est5 est6, eform ci star(* 0.05 **
0.01) noobs noconst label pr2 aic bic nonnumbers
mtitles("Unadjusted OR" "Unadjusted OR" "Unadjusted OR"
"Unadjusted OR" "Adjusted OR Model 1" "Adjusted OR Model 2")
title(Tabel 1. Odds Ratio Faktor yang Berhubungan dengan Status
Penyakit Jantung Koroner (PJK))
```

Maka *output*-nya adalah sebagai berikut.

Tabel 1. Odds Ratio Faktor yang Berhubungan dengan Status Penyakit Jantung Koroner (PJK)

	Unadjusted OR	Unadjusted OR	Unadjusted OR	Unadjusted OR	Adjusted OR Model 1	Adjusted OR Model 2
Status PJK						
Perempuan	1 [1,1]				1 [1,1]	1 [1,1]
Laki-laki	11.99** [4.593,31.31]				9.443** [3.292,27.09]	6.375** [2.124,19.13]
25-35 tahun		1 [1,1]			1 [1,1]	1 [1,1]
36-45 tahun		0.986 [0.261,3.724]			1.186 [0.301,4.680]	1.102 [0.275,4.409]
46-55 tahun		26.76** [7.965,89.90]			15.54** [4.019,60.06]	17.53** [3.910,78.58]
>55 tahun		20.74** [7.384,58.28]			16.51** [5.301,51.41]	9.980** [2.946,33.81]
Tidak Merokok			1 [1,1]		1 [1,1]	1 [1,1]
Merokok			11.57** [5.424,24.69]		5.742** [2.382,13.84]	6.590** [2.554,17.00]
Normal				1 [1,1]		1 [1,1]
Kurus				2.030 [0.497,8.286]		0.702 [0.132,3.744]
Gemuk				6.727** [2.008,22.53]		2.981 [0.763,11.64]
Pseudo R-squared	0.148	0.267	0.175	0.068	0.444	0.469
AIC	233.0	205.2	225.9	256.7	161.4	158.8
BIC	241.4	222.0	234.3	269.4	186.7	192.6

Exponentiated coefficients; 95% confidence intervals in brackets
 * p<0.05, ** p<0.01

Output tersebut dapat disimpan dalam Ms. Office (Word, Excel, dsb) dengan syntax: `esttab using regresilogistik.rtf` atau diedit dalam Ms. Office agar lebih rapi. Hasil akhir tabel yang sudah dirapikan adalah sebagai berikut.

Tabel 1. Odds Ratio Faktor yang Berhubungan dengan Status Penyakit Jantung Koroner (PJK)

Prediktor	Unadjusted OR		Adjusted OR	
			Model 1	Model 2
Status PJK				
Jenis Kelamin				
Perempuan	[1,1]		[1,1]	[1,1]
Laki-laki	11.99** [4.593,31.31]		9.443** [3.292,27.09]	6.375** [2.124,19.13]
Usia				
25-35 tahun	[1,1]		[1,1]	[1,1]
36-45 tahun	0.986 [0.261,3.724]		1.186 [0.301,4.680]	1.102 [0.275,4.409]
46-55 tahun	26.76** [7.965,89.90]		15.54** [4.019,60.06]	17.53** [3.910,78.58]
>55 tahun	20.74** [7.384,58.28]		16.51** [5.301,51.41]	9.980** [2.946,33.81]
Status Merokok				
Tidak Merokok	[1,1]		[1,1]	[1,1]
Merokok	11.57** [5.424,24.69]		5.742** [2.382,13.84]	6.590** [2.554,17.00]
BMI				
Normal				[1,1]
Kurus	2.030 [0.497,8.286]		2.030 [0.497,8.286]	0.702 [0.132,3.744]
Gemuk	6.727** [2.008,22.53]		6.727** [2.008,22.53]	2.981 [0.763,11.64]
Pseudo R-squared	0.148	0.267	0.444	0.469
AIC	233.0	205.2	161.4	158.8
BIC	241.4	222.0	186.7	192.6

Exponentiated coefficients (OR); 95% confidence intervals dalam kurung

* p<0.05, ** p<0.01

Terima Kasih