

MODULE BOOK
BLOCK OF STATISTIC
(ELECTIVE)



Person in Charge :

Dr. TITIH HURIAH, M.Kep.,Sp.Kep.K

MASTER OF NURSING
POST GRADUATE PROGRAM
UNIVERSITAS MUHAMMADIYAH YOGYAKARTA
2016/2017



HALAMAN PENGESAHAN

Nama Mata Kuliah	: Blok Elektif
Nomor Kode/ SKS	: MN 102 / 3 SKS (4 Minggu)
Bidang Ilmu	: Statistik
Status Mata Kuliah	: Pilihan
Nama Penanggungjawab (Koordinator)	: Dr. Titih Huriah, M.Kep.,Ns.,Sp.Kep.K
NIP	: 173.045
Pangkat/ Golongan	: Lektor / IIC
Jabatan	: Lektor
Fakultas/Program Studi	: Magister Keperawatan / Program Pasca Sarjana
Universitas	: Universitas Muhammadiyah Yogyakarta
Jumlah Tim Pengajar	: 5 Orang
	- Dr. Titih Huriah, M.Kep.,Ns.,Sp.Kep.K
	- Shanti Wardaningsih, M.Kep.,Ns.,Sp.Kep.J.,Ph.D
	- Fitri Arofiati, S.Kep.,Ns.,MAN.,Ph.D
	- Erna Rochmawati, SKp.,MNS. M.Med.Ed.,DNSc
	- Dr. Elsy Maria Rosa, S.Kp.,M.Kep

Yogyakarta, Februari 2017

Menyetujui
Ka Prodi Magister Keperawatan,

Mengetahui
PJ Blok,

Fitri Arofiati, S.Kep.,Ns.,MAN.,Ph.D Dr. Titih Huriah, M.Kep.,Ns.,Sp.Kep.Kom

PENDAHULUAN

A. Visi, Misi Dan Tujuan Pendidikan Prodi

Visi

Menjadi program magister keperawatan yang unggul dalam pengembangan ilmu pengetahuan dan teknologi melalui penelitian keperawatan berlandaskan nilai-nilai ke-Islaman untuk kemaslahatan umat di Indonesia pada 2020 dan di Asia pada 2025.

Misi

1. Menyelenggarakan pendidikan keperawatan di tingkat magister dengan standar nasional dan internasional.
2. Mengembangkan penelitian yang menghasilkan teknologi keperawatan.
3. Menerapkan ilmu keperawatan sebagai bagian dari pengabdian kepada masyarakat untuk kemaslahatan umat.
4. Menghasilkan magister keperawatan yang berakhlak mulia dan mampu mengintegrasikan nilai

Tujuan

1. Tujuan Umum

Mampu mewujudkan magister keperawatan yang unggul dalam pengembangan ilmu pengetahuan dan teknologi, berakhlak mulia sehingga mampu bersaing di tingkat nasional dan internasional

2. Tujuan Khusus

- a. Meningkatkan kualitas hasil pendidikan yang berstandar nasional dan Internasional
- b. Meningkatkan budaya meneliti yang menghasilkan penelitian yang berkualitas dan tepat guna
- c. Meningkatkan pelayanan keperawatan profesional berbasis bukti dalam upaya peningkatan derajat kesehatan masyarakat
- d. Meningkatkan pelayanan keperawatan professional yang dilandasi nilai-nilai ke-Islaman.

B. Capaian Pembelajaran (*Learning Outcome*)

Capaian Pembelajaran Prodi Magister Keperawatan berdasarkan Profil Lulusan sebagai berikut :

Tabel 1. Kaitan Profil Lulusan Prodi M.Kep dan Capaian Pembelajaran

NO	PROFIL LULUSAN & DES	UNSUR SN PT & KKNI	CAPAIAN PEMBELAJARAN (CP)	
1	<p><i>Care provider</i> : Seorang perawat yang memiliki jiwa kepemimpinan dalam pengembangan ilmu keperawatan secara interdisiplin berdasarkan <i>evidence based nursing</i>, nilai-nilai ke-Islaman, dan berwawasan global</p>	Sikap	S1	Bertakwa kepada Tuhan Yang Maha Esa dan mampu menunjukkan sikap religius;
			S2	Menjunjung tinggi nilai kemanusiaan dalam menjalankan tugas berdasarkan agama, moral, dan etika;
			S3	Menginternalisasi nilai, norma, dan etika akademik;
			S5	Menghargai keanekaragaman budaya, pandangan, agama, dan kepercayaan, serta pendapat atau temuan orisinal orang lain;
			S6	Berkontribusi dalam peningkatan mutu kehidupan bermasyarakat, berbangsa, bernegara dan kemajuan peradaban berdasarkan Pancasila
			S8	Taat hukum dan disiplin dalam kehidupan bermasyarakat dan bernegara
			S10	Menunjukkan sikap bertanggungjawab atas pekerjaan di bidang keahliannya secara mandiri.
			S11	Mampu bertanggung gugat terhadap praktik profesional meliputi kemampuan menerima tanggung gugat terhadap keputusan dan tindakan profesional sesuai dengan lingkup praktik di bawah tanggungjawabnya, dan hukum/peraturan perundangan;
			S12	Mampu melaksanakan praktik keperawatan dengan prinsip etis dan peka budaya sesuai dengan Kode Etik Perawat Indonesia;
		S13	Memiliki sikap menghormati hak privasi, nilai budaya yang dianut dan martabat klien, menghormati hak klien untuk memilih dan menentukan sendiri asuhan keperawatan dan kesehatan yang diberikan, serta bertanggung jawab atas kerahasiaan dan keamanan informasi tertulis, verbal dan elektronik yang diperoleh dalam kapasitas sesuai dengan lingkup tanggungjawabnya;	
Penguasaan Pengetahuan	PP1	Menguasai berbagai jenis dan jenjang teori keperawatan dan teori lain yang diperlukan sebagai dasar seorang pemimpin dan manajer keperawatan		

		PP2	Menguasai teori aplikasi terkait metoda dan perangkat analisis fungsi manajemen (perencanaan, pelaksanaan, pengarahan, pemantauan, evaluasi, dan pengendalian) serta fungsi organisasi (SDM, operasional, sarana prasarana, pemasaran dan keuangan) pada berbagai jenis organisasi;
		PP3	Menguasai konsep dan teknik menyusun rencana strategis dan menjabarkannya dalam rencana operasional;
		PP4	Menguasai prinsip kepemimpinan dan kewirausahaan dalam berbagai tipe organisasi;
		PP12	Menguasai berbagai teori etika yang dapat dijadikan acuan dalam menyelesaikan masalah pelayanan dan asuhan keperawatan
		PP13	Menguasai teori transkultural yang dapat dijadikan dasar dalam menyelesaikan masalah pelayanan dan asuhan keperawatan
		PP14	Menguasai prinsip-prinsip <i>hospitality</i> dalam berperan sebagai pemimpin dan manajer keperawatan
		PP15	Menguasai teori sistem pemberian asuhan keperawatan, mutu asuhan, audit, dan penjaminan mutu
		PP16	Menguasai pengetahuan Islam yang berkembang sesuai Al Quran dan As Sunah.
	Keterampilan Umum	KU2	Mampu melakukan validasi akademik atau kajian sesuai bidang keahliannya dalam menyelesaikan masalah di masyarakat atau industri yang relevan melalui pengembangan pengetahuan dan keahliannya
		KU3	Mampu menyusun ide, hasil pemikiran dan argumen saintifik secara bertanggung jawab dan berdasarkan etika akademik, serta mengkomunikasikan melalui media kepada masyarakat akademik dan masyarakat luas
		KU5	Mampu mengambil keputusan dalam konteks menyelesaikan masalah pengembangan ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora berdasarkan kajian ,analisis atau eksperimental terhadap informasi dan data.

			KU6	Mampu mengelola, mengembangkan dan memelihara jaringan kerja dengan kolega, sejawat di dalam lembaga dan komunitas penelitian yang lebih luas.
			KU7	Mampu meningkatkan kapasitas pembelajaran secara mandiri.
		Keterampilan Khusus	KK1	Mampu mengembangkan inovasi sistem pemberian asuhan keperawatan
			KK2	Menerjemahkan standar patient care sesuai standar akreditasi, JCI, CMS, dan profesional nursing literature
			KK3	Membantu proses perubahan ketika asuhan keperawatan didesain ulang
			KK4	Membantu pengembangan dan implementasi program keselamatan pasien
			KK5	Mendesain sistem keselamatan klinis proses kebijakan dan prosedur
2	Educator : Seorang perawat yang mampu mengelola dan mengembangkan manajemen pendidikan keperawatan baik di tingkat pendidikan akademik maupun klinik dengan mengikuti perkembangan teknologi pendidikan keperawatan secara global berdasarkan nilai-nilai ke-Islaman		Sikap	S1
			S2	Menjunjung tinggi nilai kemanusiaan dalam menjalankan tugas berdasarkan agama, moral dan etika
			S5	Menghargai keanekaragaman budaya, pandangan, agama, dan kepercayaan serta pendapat atau temuan orisinal orang lain
			S6	Berkontribusi dalam peningkatan mutu kehidupan bermasyarakat, berbangsa, bernegara dan kemajuan peradaban berdasarkan pancasila
			S7	Menunjukkan sikap kepemimpinan, bekerja sama dan memiliki kepekaan sosial serta kepedulian terhadap masyarakat dan lingkungan
			S10	Menunjukkan sikap jujur, percaya diri, empati, dan bertanggungjawab atas pekerjaan di bidang keahliannya secara mandiri.
			S11	Sebagai role model yang menunjukkan nilai-nilai ke-Islaman dan berwawasan global dalam pendidikan dan pelayanan keperawatan
			S13	Memiliki sikap menghormati hak asasi manusia, privasi, nilai budaya yang dianut, martabat, serta bertanggung jawab atas kerahasiaan, dan keamanan informasi tertulis, verbal, dan elektronik yang diperoleh dalam

				kapasitas yang sesuai dengan lingkup tanggung jawabnya
		Penguasaan Pengetahuan	PP8	Menguasai etika bisnis dan nilai-nilai kemanusiaan (<i>humanity values</i>);
			PP9	Menguasai pengetahuan tentang jenis dan regulasi organisasi lokal, nasional, regional, dan global;
			PP10	Menguasai kaidah, prinsip dan teknik komunikasi lintas fungsi, level organisasi, dan budaya;
			PP11	Mampu menggunakan bahasa lisan dan tulisan dalam Bahasa Indonesia dan Bahasa Inggris dengan baik untuk kegiatan akademik maupun non akademik.
			PP13	Menguasai teori transkultural yang dapat dijadikan dasar dalam menyelesaikan masalah pelayanan dan asuhan keperawatan
			PP16	Menguasai pengetahuan Islam yang berkembang sesuai Al Quran dan As Sunah.
		Keterampilan Umum	KU3	Mampu menyusun ide, hasil pemikiran dan argumen saintifik secara bertanggung jawab dan berdasarkan etika akademik, serta mengkomunikasikan melalui media kepada masyarakat akademik dan masyarakat luas
			KU4	Mampu mengidentifikasi bidang keilmuan yang menjadi obyek penelitiannya dan memosisikan ke dalam suatu peta penelitian yang dikembangkan melalui pendekatan inter atau multi disipliner.
			KU6	Mampu mengelola, mengembangkan dan memelihara jaringan kerja dengan kolega, sejawat di dalam lembaga dan komunitas penelitian yang lebih luas.
			KU7	Mampu meningkatkan kapasitas pembelajaran secara mandiri.
		Keterampilan Khusus	KK6	Mampu merancang program edukasi/pembelajaran yang kreatif dan inovatif pada bidang kepemimpinan dan manajemen keperawatan
3	Researcher : Seorang perawat yang memiliki kemampuan	Sikap	S5	Menghargai keanekaragaman budaya, pandangan, agama, dan kepercayaan serta pendapat atau temuan orisinal orang lain
			S6	Berkontribusi dalam peningkatan mutu kehidupan bermasyarakat, berbangsa,

<p>menyelesaikan masalah keperawatan melalui penelitian keperawatan di tatanan pendidikan dan pelayanan keperawatan berdasarkan nilai-nilai ke-Islaman dan berwawasan global</p>			bernegara dan kemajuan peradaban berdasarkan pancasila
	S7		Menunjukkan sikap kepemimpinan, bekerja sama dan memiliki kepekaan sosial serta kepedulian terhadap masyarakat dan lingkungan
	S10		Menunjukkan sikap jujur, percaya diri, empati, dan bertanggungjawab atas pekerjaan di bidang keahliannya secara mandiri.
	S11		Sebagai role model yang menunjukkan nilai-nilai ke-Islaman dan berwawasan global dalam pendidikan dan pelayanan keperawatan
	Penguasaan Pengetahuan	PP5	Menguasai konsep tentang metode penelitian yang mencakup studi kasus, kesejarahan, survei, simulasi, dan eksperimen pada lingkup kualitatif dan kuantitatif, secara eksploratif, deskriptif, dan verifikatif;
		PP6	Menguasai minimal satu metode penelitian (studi kasus, kesejarahan, survei, simulasi, dan eksperimen pada lingkup kualitatif dan kuantitatif, secara eksploratif, deskriptif, dan verifikatif);
		PP7	Menguasai konsep-konsep tentang statistik deskriptif dan inferensial.
		PP16	Menguasai pengetahuan Islam yang berkembang sesuai Al Quran dan As Sunah.
	Keterampilan Umum	KU1	Mampu mengembangkan pemikiran logis, kritis, sistematis, dan kreatif melalui penelitian ilmiah, penciptaan desain atau karya seni dalam bidang ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora sesuai dengan bidang keahliannya, menyusun konsepsi ilmiah dan hasil kajiannya berdasarkan kaidah, tata cara, dan etika ilmiah dalam bentuk tesis yang dipublikasikan tulisan dalam jurnal ilmiah yang terakreditasi.
		KU2	Mampu melakukan validasi akademik atau kajian sesuai bidang keahliannya dalam menyelesaikan masalah di masyarakat atau industri yang relevan melalui pengembangan pengetahuan dan keahliannya
		KU3	Mampu menyusun ide, hasil pemikiran dan argumen saintifik secara bertanggung jawab dan berdasarkan etika akademik, serta

				mengkomunikasikan melalui media kepada masyarakat akademik dan masyarakat luas
			KU4	Mampu mengidentifikasi bidang keilmuan yang menjadi obyek penelitiannya dan memosisikan ke dalam suatu peta penelitian yang dikembangkan melalui pendekatan inter atau multi disipliner.
			KU8	Mampu mendokumentasikan, menyimpan, mengamankan, dan menemukan kembali data hasil penelitian dalam rangka menjamin kesahihan dan mencegah plagiasi.
		Keterampilan Khusus	KK7	Mampu melakukan kajian empirik dan pemodelan dengan menggunakan metode ilmiah pada pengelolaan asuhan dan pelayanan di berbagai tipe organisasi berdasarkan fungsi organisasi;
			KK8	Mengembangkan peta penelitian dalam bidang kepemimpinan dan manajemen keperawatan
			KK9	Menggunakan hasil riset untuk mengembangkan, melaksanakan standar praktik dan model patient care di organisasi
			KK10	Mampu menganalisis data, menyajikan, dan mendesiminasikan hasil riset untuk peningkatan, pengelolaan asuhan dan pelayanan

Dari hasil Capaian Pembelajaran (*Learning Outcome*) profil di atas maka dapat di susun *Learning Outcome* program studi secara keseluruhan berdasarkan pada empat unsur utama yaitu sikap, penguasaan pengetahuan, ketrampilan umum dan ketrampilan khusus. Adapun Capaian Pembelajaran Prodi Magister Keperawatan sebagai berikut :

NO	UNSUR SN PT & KKNI		CAPAIAN PEMBELAJARAN (CP)
1	Sikap	S1	Bertakwa kepada Tuhan Yang Maha Esa dan mampu menunjukkan sikap religius.
		S2	Menjunjung tinggi nilai kemanusiaan dalam menjalankan tugas berdasarkan agama, moral, dan etika.
		S3	Menginternalisasi nilai, norma, dan etika akademik.
		S4	Berperan sebagai warga negara yang bangga dan cinta tanah air, memiliki nasionalisme serta rasa tanggungjawab pada negara dan bangsa.
		S5	Menghargai keanekaragaman budaya, pandangan, agama, dan kepercayaan, serta pendapat atau temuan orisinal orang lain.
		S6	Berkontribusi dalam peningkatan mutu kehidupan bermasyarakat, berbangsa, bernegara, dan kemajuan peradaban berdasarkan pancasila.
		S7	Bekerja sama dan memiliki kepekaan sosial serta kepedulian terhadap masyarakat dan lingkungan
		S8	Taat hukum dan disiplin dalam kehidupan bermasyarakat dan bernegara.
		S9	Menginternalisasi semangat kemandirian, kejuangan, dan kewirausahaan.
		S10	Menunjukkan sikap bertanggungjawab atas pekerjaan di bidang keahliannya secara mandiri.
		S11	Mampu bertanggung gugat terhadap praktik profesional meliputi kemampuan menerima tanggung gugat terhadap keputusan dan tindakan profesional sesuai dengan lingkup praktik di bawah tanggungjawabnya, dan hukum/peraturan perundangan.
		S12	Mampu melaksanakan praktik keperawatan dengan prinsip etis dan peka budaya sesuai dengan Kode Etik Perawat Indonesia.
		S13	Memiliki sikap menghormati hak privasi, nilai budaya yang dianut dan martabat klien, menghormati hak klien untuk memilih dan menentukan sendiri asuhan keperawatan dan kesehatan yang diberikan, serta bertanggung jawab atas kerahasiaan dan keamanan informasi tertulis, verbal dan elektronik yang diperoleh dalam kapasitas sesuai dengan lingkup tanggungjawabnya.
2	Penguasaan Pengetahuan	PP1	Menguasai berbagai jenis dan jenjang teori keperawatan dan teori lain yang diperlukan sebagai dasar seorang pemimpin dan manajer keperawatan

PP2	Menguasai teori aplikasi terkait metoda dan perangkat analisis fungsi manajemen (perencanaan, pelaksanaan, pengarahannya, pemantauan, evaluasi, dan pengendalian) serta fungsi organisasi (SDM, operasional, sarana prasarana, pemasaran dan keuangan) pada berbagai jenis organisasi;
PP3	Menguasai konsep dan teknik menyusun rencana strategis dan menjabarkannya dalam rencana operasional;
PP4	Menguasai prinsip kepemimpinan dan kewirausahaan dalam berbagai tipe organisasi;
PP5	Menguasai konsep tentang metode penelitian yang mencakup studi kasus, kesejarahan, survei, simulasi, dan eksperimen pada lingkup kualitatif dan kuantitatif, secara eksploratif, deskriptif, dan verifikatif;
PP6	Menguasai minimal satu metode penelitian (studi kasus, kesejarahan, survei, simulasi, dan eksperimen pada lingkup kualitatif dan kuantitatif, secara eksploratif, deskriptif, dan verifikatif);
PP7	Menguasai konsep-konsep tentang statistik deskriptif dan inferensial.
PP8	Menguasai etika bisnis dan nilai-nilai kemanusiaan (<i>humanity values</i>);
PP9	Menguasai pengetahuan tentang jenis dan regulasi organisasi lokal, nasional, regional, dan global;
PP10	Menguasai kaidah, prinsip dan teknik komunikasi lintas fungsi, level organisasi, dan budaya;
PP11	Mampu menggunakan bahasa lisan dan tulisan dalam Bahasa Indonesia dan Bahasa Inggris dengan baik untuk kegiatan akademik maupun non akademik.
PP12	Menguasai berbagai teori etika yang dapat dijadikan acuan dalam menyelesaikan masalah pelayanan dan asuhan keperawatan
PP13	Menguasai teori transkultural yang dapat dijadikan dasar dalam menyelesaikan masalah pelayanan dan asuhan keperawatan
PP14	Menguasai prinsip-prinsip <i>hospitality</i> dalam berperan sebagai pemimpin dan manajer keperawatan
PP15	Menguasai teori sistem pemberian asuhan keperawatan, mutu asuhan, audit, dan penjaminan mutu
PP16	Menguasai pengetahuan Islam yang berkembang sesuai Al Quran dan As Sunah.

3	Ketrampilan Umum	KU1	Mampu mengembangkan pemikiran logis, kritis, sistematis, dan kreatif melalui penelitian ilmiah, penciptaan desain atau karya seni dalam bidang ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora sesuai dengan bidang keahliannya, menyusun konsepsi ilmiah dan hasil kajiannya berdasarkan kaidah, tata cara, dan etika ilmiah dalam bentuk tesis yang dipublikasikan tulisan dalam jurnal ilmiah yang terakreditasi.
		KU2	Mampu melakukan validasi akademik atau kajian sesuai bidang keahliannya dalam menyelesaikan masalah di masyarakat atau industri yang relevan melalui pengembangan pengetahuan dan keahliannya
		KU3	Mampu menyusun ide, hasil pemikiran dan argumen saintifik secara bertanggung jawab dan berdasarkan etika akademik, serta mengkomunikasikan melalui media kepada masyarakat akademik dan masyarakat luas
		KU4	Mampu mengidentifikasi bidang keilmuan yang menjadi obyek penelitiannya dan memosisikan ke dalam suatu peta penelitian yang dikembangkan melalui pendekatan inter atau multi disiplin.
		KU5	Mampu mengambil keputusan dalam konteks menyelesaikan masalah pengembangan ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora berdasarkan kajian ,analisis atau eksperimental terhadap informasi dan data.
		KU6	Mampu mengelola, mengembangkan dan memelihara jaringan kerja dengan kolega, sejawat di dalam lembaga dan komunitas penelitian yang lebih luas.
		KU7	Mampu meningkatkan kapasitas pembelajaran secara mandiri.
		KU8	Mampu mendokumentasikan, menyimpan, mengamankan, dan menemukan kembali data hasil penelitian dalam rangka menjamin kesahihan dan mencegah plagiasi.
4	Ketrampilan Khusus	KK1	Mampu mengembangkan inovasi sistem pemberian asuhan keperawatan
		KK2	Menerjemahkan standar patient care sesuai standar akreditasi, JCI, CMS, dan profesional nursing literature
		KK3	Membantu proses perubahan ketika asuhan keperawatan didesain ulang
		KK4	Membantu pengembangan dan implementasi program keselamatan pasien
		KK5	Mendesain sistem keselamatan klinis proses kebijakan dan prosedur

	KK6	Mampu merancang program edukasi/pembelajaran yang kreatif dan inovatif pada bidang kepemimpinan dan manajemen keperawatan
	KK7	Mampu melakukan kajian empirik dan pemodelan dengan menggunakan metode ilmiah pada pengelolaan asuhan dan pelayanan di berbagai tipe organisasi berdasarkan fungsi organisasi;
	KK8	Mengembangkan peta penelitian dalam bidang kepemimpinan dan manajemen keperawatan
	KK9	Menggunakan hasil riset untuk mengembangkan, melaksanakan standar praktik dan model patient care di organisasi
	KK10	Mampu menganalisis data, menyajikan, dan mendesiminasikan hasil riset untuk peningkatan, pengelolaan asuhan dan pelayanan

INFORMASI BLOK

A. Nama dan bobot SKS, Kode Blok dan Semester Penawaran

Nama Blok	:	Statistik
Bobot SKS	:	3 SKS (4 Minggu)
Kode Blok	:	MN 102
Semester	:	II (Genap) 2016/2017
Jumlah Pertemuan	:	

B. Ketercapaian Pembelajaran berdasarkan Sikap, Penguasaan Pengetahuan, Ketrampilan Umum & Ketrampilan Khusus melalui Blok yang bersangkutan

Capaian Pembelajaran yang dimiliki oleh Mahasiswa setelah mengikuti blok Statistik adalah :

SOFTSKILL

SIKAP	1.	Mampu bertanggung gugat terhadap praktik profesional meliputi kemampuan menerima tanggung gugat terhadap keputusan dan tindakan profesional sesuai dengan lingkup praktik di bawah tanggungjawabnya, dan hukum/peraturan perundangan
	2.	Mampu melaksanakan praktik keperawatan dengan prinsip etis dan peka budaya sesuai dengan Kode Etik Perawat Indonesia
	3.	Menginternalisasi nilai, norma, dan etika akademik

HARDSKILL

PENGUASAAN PENGETAHUAN	1.	Menguasai konsep tentang metode penelitian yang mencakup studi kasus, kesejarahan, survei, simulasi, dan eksperimen pada lingkup kualitatif dan kuantitatif, secara eksploratif, deskriptif, dan verifikatif;
	2.	Menguasai minimal satu metode penelitian (studi kasus, kesejarahan, survei, simulasi, dan eksperimen pada lingkup kualitatif dan kuantitatif, secara eksploratif, deskriptif, dan verifikatif);
	3.	Menguasai konsep-konsep tentang statistik deskriptif dan inferensial
	4.	Menguasai pengetahuan Islam yang berkemajuan sesuai Al-Qur'an dan As Sunah
KETRAMPILAN UMUM	1.	Mampu mendokumentasikan, menyimpan, mengamankan, dan menemukan kembali data hasil penelitian dalam rangka menjamin kesahihan dan mencegah plagiasi.

**KETRAMPILAN
KHUSUS**

1. Mampu melakukan kajian empirik dan pemodelan dengan menggunakan metode ilmiah pada pengelolaan asuhan dan pelayanan di berbagai tipe organisasi berdasarkan fungsi organisasi;
2. Mengembangkan peta penelitian dalam bidang kepemimpinan dan manajemen keperawatan

DAFTAR ISI

- I Ringkasan Blok
- II Kegiatan Belajar Mengajar
 - A. Tujuan
 - B. Karakteristik Mahasiswa
 - C. Learning Outcomes
- III Modul
 - 1. Bekerja dengan SPSS
 - A. Tujuan Pembelajaran
 - B. Topik
 - C. Pre-Assessment
 - 2. Bekerja dengan STATA
 - A. Tujuan Pembelajaran
 - B. Topik
 - C. Pre-Assessment
 - 3. Bekerja dengan EPI DATA
 - A. Tujuan Pembelajaran
 - B. Topik
 - C. Pre-Assessment
- IV Course Time Table
- V Expertise

I. OVERVIEW BLOCK

Blok statistik merupakan salah satu blok elektif yang bisa dipilih oleh mahasiswa pada tahun ke dua kurikulum PBL pada Magister Keperawatan Program Pasca Sarjana UMY. Blok ini terdiri dari tiga modul yaitu Epi Info, SPSS, dan Stata. Blok elektif ini terdiri dari 3 SKS dengan pembagian 1 SKS kuliah dan 2 SKS praktek.

Sebagian besar pembelajaran di blok statistik ini adalah pembelajaran dengan menggunakan program SPSS, yang merupakan program yang paling sering digunakan oleh praktisi kesehatan. Pada modul Bekerja dengan SPSS, mahasiswa akan dipaparkan cara menganalisis data dengan pendekatan statistik univariat, bivariat dan multivariat. Modul ini terdiri dari kegiatan perkuliahan di kelas, tutorial dan praktek di laboratorium IT. Modul bekerja dengan STATA dan Epi Info hanya bersifat pengenalan kepada mahasiswa. Modul ini terdiri dari kuliah di kelas dan praktek.

II. KEGIATAN BELAJAR MENGAJAR

A. Tujuan

Pada akhir pembelajaran, mahasiswa mengetahui, mempunyai kemampuan dalam menganalisis data penelitian, melakukan olah data dengan SPSS, STATA dan Epi Info, serta mampu melakukan interpretasi atau membuat kesimpulan dari hasil analisis data penelitian yang dilakukan.

B. Karakteristik Mahasiswa

Blok statistik ini akan dilaksanakan pada tahun kedua bersamaan dengan kegiatan blok 4 pada program Magister Keperawatan Program Pasca Sarjana UMY.

C. Learning outcome

Kompetensi blok :

1. Memahami konsep basis data dan teknik transformasi data
2. Memahami dan terampil membuat statistik dengan menggunakan SPSS, STATA dan Epi Info
3. Memahami dan terampil memilih dan melakukan uji statistik yang sesuai
4. Mampu menginterpretasikan hasil-hasil uji statistik

D. Perhitungan SKS

$$3 \text{ SKS} = 150 \text{ menit} \times 14 = 2100 \text{ menit}$$

$$6 \text{ minggu} = 2100/6 = 350 \text{mnt/mg}$$

$$1 \text{ minggu} = 350/6 = 5,8 \text{ jam (dibulatkan 6 jam)}$$

III. MODULE

1. Working with SPSS

A. Learning Outcomes

Mahasiswa mampu memahami, terampil membuat statistik dengan menggunakan SPSS dan mampu menginterpretasi hasil analisis data dengan SPSS.

B. Topics

Profil	Core Competency	Strategy	Topics
Researcher	1. Mengetahui prinsip-prinsip umum pengolahan dan analisis data penelitian	Lecture, Discussion	<ul style="list-style-type: none">• Penjelasan silabus• Penjelasan prinsip dan tahapan pengolahan data penelitian• Penjelasan struktur file data
	2. Mengembangkan struktur data untuk keperluan analisis data	Kuliah Praktikum	<ul style="list-style-type: none">• Latihan membuka data SPSS• Latihan memberikan variabel label dan value label• Latihan merge file, transpose data, dan select cases
	3. Mampu mengembangkan variabel baru untuk keperluan analisis data 4. Mampu mengeluarkan ukuran-ukuran pemusatan,	Kuliah Praktikum	<ul style="list-style-type: none">• Latihan pengelompokan nilai variabel• Latihan membuat variabel baru berdasarkan perhitungan matematis

	penyebaran data dan grafik dengan SPSS		<ul style="list-style-type: none"> • Menampilkan distribusi frekuensi • Menampilkan dan interpretasi nilai pemusatan dan penyebaran data • Menampilkan kurva batang, pie, normal, box plot, stem&leaf, histogram dan diagram sebar
	5. Mampu melakukan uji validitas dan reliabilitas	Kuliah Praktikum	<ul style="list-style-type: none"> • Uji validitas • Uji reliabilitas
	6. Mampu mengeluarkan dan menginterpretasikan uji statistik beda proporsi : uji kai kuadrat	Kuliah Praktikum	<ul style="list-style-type: none"> • Uji kai kuadrat
	7. Mampu mengeluarkan dan menginterpretasikan uji statistik beda dua mean dan lebih dari dua mean	Kuliah Praktikum	<ul style="list-style-type: none"> • Uji beda dua mean : t paired dan t independent • Uji beda lebih dari dua mean : ANOVA
	8. Mampu mengeluarkan dan menginterpretasikan uji statistik korelasi dan regresi linier sederhana	Kuliah Praktikum	<ul style="list-style-type: none"> • Uji statistik korelasi dan regresi linier sederhana
	9. Mampu mengeluarkan dan menginterpretasikan uji statistik regresi linier berganda	Kuliah Praktikum	<ul style="list-style-type: none"> • Uji statistik regresi linier berganda
	10. Mampu mengeluarkan dan menginterpretasikan uji statistik regresi logistik berganda	Kuliah Praktikum	<ul style="list-style-type: none"> • Uji statistik regresi logistik berganda

C. Pre-Assessment

Learning activities must be followed by the students as the requirements to do final examination. **Minimal attendance of the learning activity :**

1. Lecture : 75%
2. Practice : 100%

2. Bekerja dengan STATA

A. Learning Outcomes

Mahasiswa mampu mengenal mengetahui langkah-langkah membuat statistik dengan menggunakan STATA.

B. Topics

Topik pertemuan tentang stata hanya dua kali yaitu kuliah dengan topik mengenal STATA dan praktikum dengan topik Sample session with STATA.

C. Pre-Assessment

Learning activities must be followed by the students as the requirements to do final examination. **Minimal attendance of the learning activity :**

1. Lecture : 75%
2. Practice : 100%

3. EPI DATA

A. Learning Outcomes

Mahasiswa mampu memahami, terampil membuat statistik dengan menggunakan Epi Data dan mampu menginterpretasi hasil analisis data dengan Epi Data.

B. Topik

Profil	Kompetensi Utama	Strategi	Topik
Researcher	<ol style="list-style-type: none"> 1. Mengetahui prinsip-prinsip umum pengolahan dan analisis data penelitian dengan menggunakan Epi Info 2. Mengetahui variabel dan struktur data 3. Mampu mengoperasikan dasar-dasar Epi-Info khususnya pengembangan template data entry 	Lecture, Discussion	<ul style="list-style-type: none"> • Penjelasan program Epi data • Pembuatan struktur file data dengan Epi data (file_QES) • Pembuatan file data dengan Epi data (file_REC) • Pembuatan program check dengan Epi Data (file_CHECK)
	<ol style="list-style-type: none"> 4. Mampu mengembangkan data penelitian dengan Epi-data 	Praktik	<ul style="list-style-type: none"> • Latihan pembuatan struktur file data dengan Epi data (file_QES) • Latihan pembuatan file data dengan Epi data (file_REC) • Latihan pembuatan program check dengan Epi Data (file_CHECK)

C. Pre-Assessment

Learning activities must be followed by the students as the requirements to do final examination. **Minimal attendance of the learning activity :**

1. Lecture : 75%
2. Practice : 100%

IV. COURSE TIME TABLE

Bulan	Hari/tanggal	Hari/tanggal	Hari/tanggal	Hari/tanggal
1 08.00 - 10.00	1. Penjelasan silabus 2. Prinsip umum pengolahan dan analisis data	Mengembangkan struktur data untuk keperluan analisis data	1. Mengembangkan variabel baru untuk keperluan analisis data 2. Mengeluarkan ukuran-ukuran pemusatan, penyebaran data dan grafik dengan SPSS	Uji validitas dan reliabilitas
	Dosen : Titih Huriyah, M.Kep.,Sp.Kom	Dosen : Titih Huriyah, M.Kep.,Sp.Kom	Dosen : Agus Karmayana R, M.Kes	Dosen : Bondan Palestin, M.Kep.,Sp.Kom
2 10.00 - 11.00	Belajar mandiri : mempersiapkan data untuk olah data selasa, 13/3	Praktikum : Latihan membuka data SPSS	Praktikum : Latihan membuat variabel baru, dll	Praktikum : Uji validitas dan reliabilitas
		Fasilitator : Titih	Fasilitator : Agus, Titih	Fasilitator : Bondan, Titih

Bulan	Hari/tanggal	Hari/tanggal	Hari/tanggal	Hari/tanggal
1 08.00 - 10.00	Uji statistik korelasi (parametrik dan nonparametrik) : dan interpretasinya	Uji statistik beda dua mean dan lebih dari dua mean dan interpretasinya	Uji statistik korelasi dan regresi linier sederhana dan interpretasinya	Uji statistik regresi linier berganda dan interpretasinya
	Dosen : Agus Karmayana R, M.Kes	Dosen : Drs.Zulaela,Dipl.M ed.Stats.,Msi	Dosen : Drs.Zulaela,Dipl.Me d.Stats.,Msi	Dosen : Drs.Zulaela,Dipl. Med.Stats.,Msi
2 10.00 - 11.00	Praktikum : Uji korelasi	Praktikum : <ul style="list-style-type: none"> • Uji beda dua mean : t paired dan t independent • Uji beda lebih dari dua mean : ANOVA 	Praktikum : Uji statistik korelasi dan regresi linier sederhana	Praktikum : Uji statistik regresi linier berganda

	Fasilitator : Agus, Titih	Fasilitator : Drs.Zulaela, Ambar	Fasilitator : Drs.Zulaela, Ambar	Fasilitator : Drs.Zulaela, Titih
--	------------------------------	--	-------------------------------------	-------------------------------------

Bulan	Hari/tanggal	Hari/tanggal	Hari/tanggal	Hari/tanggal
1 08.00 - 10.00	Uji statistik regresi logistik berganda dan interpretasinya	Ujian SPSS Fasilitator : Prima, Titih	Ujian Ulang SPSS Fasilitator : Prima, Titih	1. prinsip-prinsip umum pengolahan dan analisis data penelitian dengan menggunakan Epi Info 2. Mengoperasikan dasar-dasar Epi- Info khususnya pengembangan template data entry
	Dosen : Drs.Zulaela,Dipl.M ed.Stats.,Msi			Dosen : Arif Kurniawan, M.Kes
2 10.00 – 11.00	Praktikum : Uji statistik regresi logistik berganda			Praktikum : • Pembuatan struktur file data dengan Epi data (file_QES) • Pembuatan file data dengan Epi data (file_REC) • Pembuatan program check dengan Epi Data (file_CHECK)
	Fasilitator : Drs.Zulaela, Titih			Fasilitator : Arif Kurniawan, Titih

Bulan	Hari/tanggal
1 08.00 -10.00	Pengenalan STATA
	Dosen : Arif Kurniawan, M.Kes
2 10.00 - 11.00	Praktikum : Pengenalan program STATA
	Fasilitator : Arif, Titih

No	Name of experts	Expertise	Contact Number
1.	Drs.Zulaela,Dipl.Med.Stats.,Msi	Statistic	email : zulaela@ugm.ac.id
2.	Bondan Palestin, M.Kep.,Sp.Kom	Nursing statistic	08121525754
3.	Agus Karmayana Rubaya, M.Kes	Statistic	08174118767
4.	Arif Kurniawan, M.Kes	Statistic	08112515890
5.	Titih Huriah, M.Kep.,Sp.Kom	Nursing statistic	081392405406

APPENDIX 1

PRACTICAL GUIDANCE WITH SPSS

BLOCK OF STATISTIC (ELECTIVE)



Person in Charge :

Dr. TITIH HURIAH, M.Kep.,Sp.Kep.Kom

**MASTER OF NURSING
POST GRADUATE PROGRAM
UNIVERSITAS MUHAMMADIYAH YOGYAKARTA
2016/2017**

FIRST MEETING

WORKING WITH SPSS

Pengantar

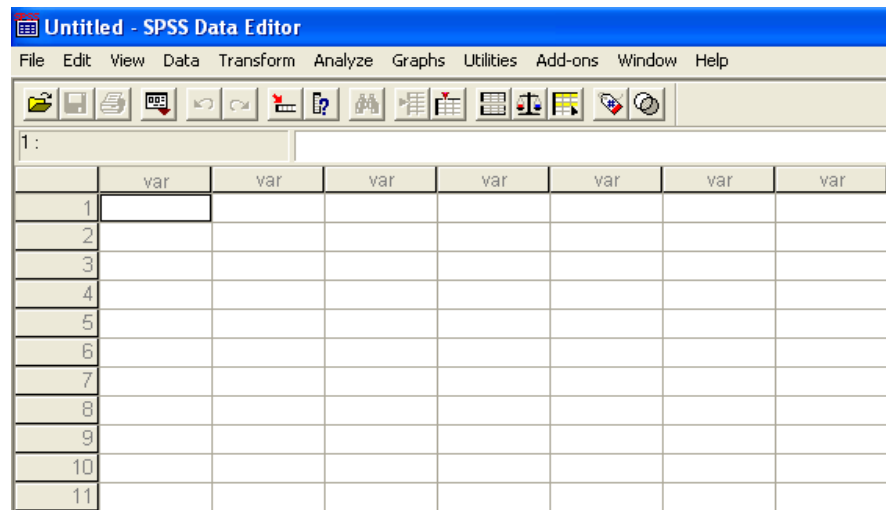
Dalam melakukan penelitian, kita perlu menguasai berbagai komponen metodologis, antara lain desain penelitian, skala pengukuran, penentuan subyek, perhitungan besar sample, pengolahan data dan pemilihan uji hipotesis yang tepat. Panduan ini dibuat khususnya untuk pengolahan data dengan program computer. Ada beberapa program yang dapat digunakan sebagai alat pengolahan data (statistic), bahkan di dalam MS Excel yang lengkap pun ada fasilitas untuk pengolahan data. Pada praktikum IT blok ini, yang akan dipelajari dan dipraktikkan adalah SPSS. Berikut ini dibahas cara penggunaan SPSS secara umum, meskipun kenyataannya ada banyak versi, namun pada dasarnya cara penggunaannya adalah hampir sama. Hal penting dalam menggunakan SPSS adalah bahwa pengguna harus mempunyai bekal penguasaan statistic yang memadai. SPSS hanyalah alat untuk mengolah data. Selebihnya pengguna harus dapat menginterpretasikan hasil pengolahan data dengan tepat, agar dapat menarik kesimpulan yang tepat pula.

MATERI I

BEKERJA DENGAN SPSS

A. Memulai SPSS

Aktifkan program SPSS di computer dengan meng-klik icon program tersebut. Pada saat pertama kali membuka SPSS, selalu nampak tampilan pertama sebagai berikut :



Gambar 1 Menu Utama SPSS

Windows (tampilan) di atas disebut **SPSS DATA EDITOR** yang selalu muncul setiap kali SPSS dibuka dan merupakan windows utama pada SPSS. Perhatikan kalimat "SPSS Data Editor" yang menunjukkan anda sedang berada pada area tersebut; hal ini untuk membedakan dengan output SPSS yang disebut VIEWER.

B. Mengenal menu-menu pada data editor

1. Menu **File** :

Merupakan menu pertama dari DATA EDITOR yang dibuka, dan berfungsi mengatur operasional file-file SPSS seperti membuat sebuah file baru (NEW), membuka file yang sudah ada (OPEN), mencetak file tertentu (PRINT), seperti pada windows MS Office.

2. Menu **Edit**

Digunakan untuk melakukan perbaikan atau perubahan berkenaan dengan data yang telah dibuat ataupun berbagai option lainnya. Berbagai perubahan tersebut meliputi menghapus data/kasus, mengganti isi data, mencari data, dll. Pada dasarnya langkah-langkah penggunaan pada menu ini serupa dengan penggunaan pada program MS Office yang telah kita pelajari di awal praktikum IT.

3. Menu **View**

Pada dasarnya berfungsi menyajikan penampilan data, toolbars dan output SPSS pada layer monitor; pengerjaan pada menu ini tidak mengubah isi variable atau data, juga tidak berpengaruh pada perhitungan statistic yang dilakukan

4. Menu **Data**

Berfungsi membuat perubahan data SPSS secara keseluruhan

5. Menu **Transform**

Berfungsi membuat perubahan data pada variable yang telah dipilih

6. Menu **Analyze**

Menu Ini merupakan jantung dari program SPSS karena pada menu inilah seluruh perhitungan statistic dilakukan.

7. Menu **Graphs**

Menu ini berfungsi menampilkan grafik/chart yang merupakan hasil perhitungan statistic data yang ada pada DATA EDITOR. Menu ini cukup penting dan kompleks yang berguna dalam penyajian data serta laporan yang berupa grafik.

8. Menu **Utilities**

Berfungsi sebagai tambahan pengerjaan data statistik dengan SPSS

9. **Add-ons**

Berisi berbagai macam prosedur statistic lanjutan yang bias dilakukan oleh SPSS, seperti CONJOINT, CATEGORIES, ADVANCED MODEL, dll.

10. **Help**

Berfungsi untuk menyediakan bantuan informasi program SPSS

C. **Membuat file data**

Dimulai dengan membuka view DATA EDITOR. Pada area DATA VIEW terdapat 2 bagian utama, yaitu :

1. kolom : terdapat kata var dalam setiap sel kolomnya. Kolom ini adalah tempat untuk menuliskan macam variable yang akan dianalisis (seperti : tinggi badan, berat badan, dll)
2. baris : dicirikan dengan adanya angka 1, 2, 3 dst. Baris ini adalah tempat mengisikan data/kasus (seperti : nama responden, angka berat badan, angka tinggi badan, dll)

Setelah membuka Data Editor, anda akan mulai dengan lembar kerja baru. Hal penting yang harus dikuasai sebelum memasukkan data adalah mendefinisikan variable Misalnya jika kolom 1 akan dibuat nama responden,

maka pertama kali, klik mouse pada sheet tab VARIABLE VIEW, atau dengan mengklik menu VIEW lalu submenu VARIABLE. Oleh karena ini variable pertama, tempatkan pointer pada baris 1. Ada beberapa point penting yang harus dibuat untuk mendefinisikan suatu variable pada variable view, yaitu :

Name : double klik di bawah kolom Name, ketik nama, lalu tekan tombol enter

Type : definisikan variable jika tidak diubah, maka secara default SPSS memberi tipe numeric. Untuk variable nama termasuk gabungan huruf (non angka) maka klik kotak kecil di kanan sel tersebut, pilih String, tekan ok

Ada kalanya, kita memasukkan data yang sebenarnya bukan data numeric (misalnya data tipe string), namun ingin kita analisis dengan SPSS. Untuk itu data tersebut dapat dinumerikkan dengan kode/kategorisasi, yaitu dengan memilih tipe data numeric pada saat mendefinisikan variable pada variable view .

Width : menentukan banyaknya karakter yang bisa dimasukkan ke dalam kolom tersebut. Pada kolom ini disediakan 1 s/d 255 karakter yang bisa (untuk data string, dan 40 untuk data numeric) dimuat. Anda bisa menentukan besar/lebar kolom sesuai kebutuhan, dengan langsung mengetik

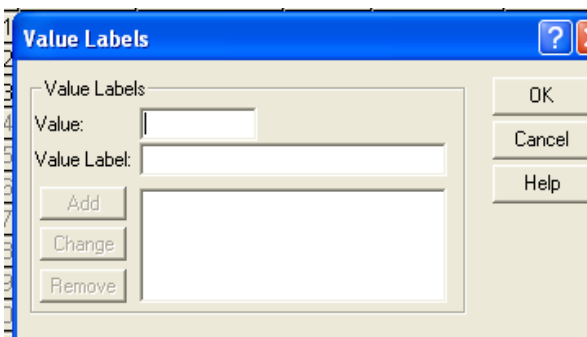
Atau menggunakan scroll number, untuk angka naik atau turun

Decimals : fasilitas ini hanya aktif jika tipe variable berupa angka. Fasilitas ini menentukan berapa angka di belakang koma, data yang akan dimasukkan. Jika variable bertipe string, maka otomatis tidak ada decimal, sehingga fasilitas ini tidak aktif (warna buram)

Label : adalah keterangan untuk nama variable yang bisa disertakan ataupun tidak. Fasilitas ini perlu diisi apabila ingin dibuat label (kategorisasi) untuk variable tertentu. Label diisi dengan nama kategori misalnya ; jenis kelamin. Label yang tidak diisi tidak mempengaruhi proses data. Sebagai contoh double klik pada kolom label diketik : jenis kelamin.

Values : untuk data kuantitatif dan tanpa kategorisasi, fasilitas ini dapat diabaikan. Untuk melanjutkan contoh, ketik angka 1. Pilihan values berguna pada proses pemberian kode. Misalnya untuk jenis kelamin, double klik pada kotak values, muncul kotak dialog

:



Gambar 2 : Kotak dialog Values

Untuk memberi kode, isikan angka 1 pada kotak value, kemudian ketik Wanita pada kotak value Label. Otomatis tombol Add aktif. Tekan tombol tersebut, otomatis keterangan 1="wanita" tampak sebagai kodifikasi gender pertama. Lanjutkan dengan mengetik angka 2 pada kotak value dan pria pada kotak value Label, lalu tekan Add, lanjutkan OK.

- Missing** : adalah data yang hilang atau tidak ada isinya. Jika semuanya lengkap, maka bisa diabaikan.
- Column** : hampir sama dengan Wdth, berfungsi menyediakan lebar kolom yang diperlukan untuk pemasukan data.
- Align** : adalah posisi data (di kiri, tengah atau kanan)
- Measure** : adalah hal yang penting di SPSS karena menyangkut tipe variable (skala pengukuran) yang nantinya menentukan jenis analisis yang digunakan. Untuk data string (karakter) ada 2 pilihan, yaitu data adalah nominal atau ordinal.

Pemahaman skala pengukuran variabel menggambarkan pemahaman terhadap data yang anda miliki. Skala pengukuran dibagi menjadi kategorik (nominal-ordinal) dan numerik (rasio-interval).

Variabel nominal mempunyai kategori yang sederajat atau tidak bertingkat (contoh : variabel jenis kelamin, yaitu laki-laki dan perempuan), sedangkan variabel ordinal mempunyai kategori yang tidak sederajat atau kategori yang bertingkat (contoh: variabel kadar kolesterol baik, kadar kolesterol sedang, dan kadar kolesterol buruk).

Variabel numerik dapat dibedakan menjadi variabel rasio dan interval. Variabel yang mempunyai nilai nol alami, tergolong variabel rasio (contoh : tinggi badan, berat badan), sedangkan variabel yang tidak memiliki nilai nol alami termasuk variabel interval (contoh: suhu, suhu nol derajat Celcius berbeda dengan nol derajat pada skala Fahrenheit)

Dalam SPSS, variabel rasio dan interval disebut sebagai variabel scale, yang dilambangkan dengan gambar penggaris).

Jika pengisian pada variable view benar, maka diperoleh hasil seperti Pada gambar di bawah ini :

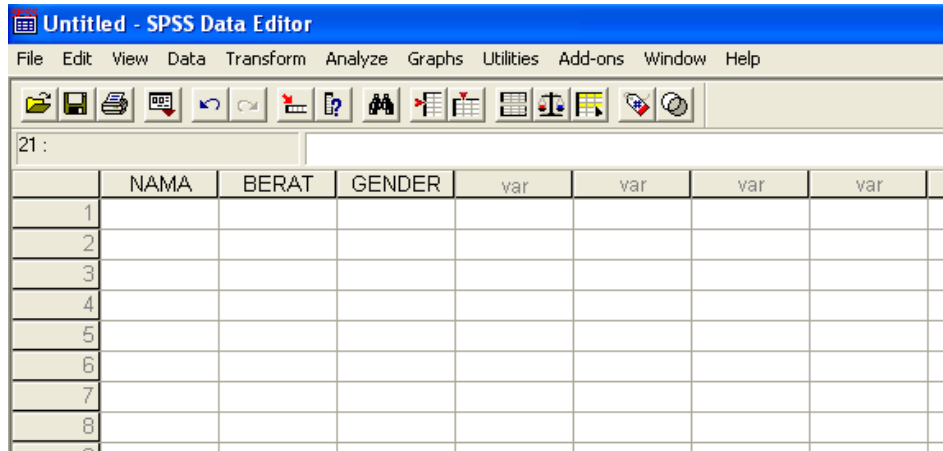
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	NAMA	String	20	0		None	None	8	Left	Nominal
2	BERAT	Numenc	8	2		None	None	8	Right	Scale
3	GENDER	String	8	0	jenis kelamin	[1, WANITA]	None	8	Left	Nominal
4										
5										

Gambar 3 : Variable View

Mengisi data

Untuk menginput data harus dilakukan di area DATA VIEW. Area ini dapat dibuka dengan mngklik tab sheet DATA VIEW yang ada di bagian

bawah layer, atau dengan memilih menu ViIEW, lalu klik submenu DATA. Cara yang lain adalah dengan langsung menekan Ctrl+T. Dari pengisian di area variable view yang telah dibuat perhatikanlah ada 3 nama variable yang tampak di data view ini.



Gambar 4 : Data View

Untuk mengisi variable nama, letakkan pointer pada baris 1 kolom variable Nama, lalu ketik menurun ke bawah sesuai data yang ada. Pengisian seperti mengisi data Microsoft Excel, atau mengetik table pada MS Word. Begitu pula dengan pengisian variable berat

Untuk pengisian variable Gender:

Sebelum mengisi data ini, arahkan pointer ke menu utama SPSS, pilihlah menu View, lalu klik mouse pda sub menu Value Label.

Menyimpan Data

Data yang telah kita buat dapat disimpan dengan prosedur berikut :

Dari menu utama SPSS, pilih menu **File**, kemudian pilih submenu **Save As** ; beri nama file, dan tempatkan pada directory yang dikehendaki. Prosedurnya pada dasarnya seperti pada program MS Office. File yang disimpan pada umumnya akan mempunyai nama ekstensi (tipe) **sav**.

D. Memformat data/mengelola data

Setelah sebuah variable didefinisikan dan data yang ada dimasukkan ke dalam SPSS DATA EDITOR, pada data tersebut dapat dilakukan berbagai pengelolaan menggunakan Menu Data. Menu ini digunakan untuk melakukan berbagai pengerjaan pada data SPSS yang bukan berupa prosedur statistic. Dalam beberapa hal, menu ini mempunyai fungsi yang berkaitan dengan menu Edit

1. Menyisipkan variabel baru

Kita dapat menyisipkan variable baru pada file data yang telah dibuat. Misalnya pada file berat yang terdiri dari 3 variabel akan ditambah dengan variable baru : tinggi. Dapat dilakukan dengan cara berikut :

- 1) buka file berat, letakkan pointer padasembarang tempat di kolom (variable) gender (pada DATA VIEW)

- 2) dari menu utama SPSS dipilih menu Data, klik mouse pada pilihan Insert Variabel. Maka akan muncul kolom baru dengan nama var00004 (atau nama baru lainnya) yang merupakan variabel baru. Untuk pendefinisian/pengisian variable (tipe, nama dan lainnya) dapat digunakan perintah sama seperti pada pemasukan variable baru yang telah dijelaskan sebelum ini.

2. Menyisipkan case

Untuk menyisipkan kasus pada file data kita dapat dilakukan dengan cara berikut :

- 1) letakkan pointer pada sel paling kiri pada posisi baris yang sesuai, misal akan disisipkan kasus pada baris ke 3, maka pointer diletakkan pada sel paling kiri baris ke 3
- 2) pilih submenu Insert Case pada menu DATA atau klik kanan mouse pada posisi tadi, lalu klik Insert Case
- 3) dengan demikian akan tersedia 1 baris baru yang siap untuk diisi dengan kasus yang akan disisipkan.

3. Mengganti isi data

Untuk mengganti isi data dapat dilakukan dengan cara sederhana yaitu dengan melatukkan pointer pada sel yang akan diganti isinya sehingga tampak terblok, langsung ketik dengan data yang baru.

4. Menghapus data

Misal dari file berat akan dihapus data atau kasus nomor 5 (lengkap dari nama, gender, berat), maka penggantian dapat dilakukan dengan cara :

- 1) letakkan pointer pada angka 5 (kotak paling kiri sehingga tampak seluruh kotak di baris tersebut terblok
- 2) dari menu utama, pilih menu Edit, pilih Cut, atau klik kanan pada sel tersebut pilih cut, atau dengan menekan Ctrl+X yang akan berefek sama yaitu hilangnya seluruh kasus.
- 3) Jika hanya akan dihapus namanya saja, cukup dihapus dengan cara menekan tombol Del pada sel tersebut atau perintah Clear pada menu Edit.
- 4) Untuk membatalkan perintah, tekan undo pada toolbar atau pada menu Edit

5. Mengcopy data

Sebuah data (variable) dapat dilakukan proses copy dan paste secara biasa, seperti perlakuan pada sebuah data pada umumnya. yaitu dengan cara mengCopy . Paste seperti biasa. Namun demikian , dalam praktek sering dijumpai proses penggantian variable yang lebih dari satu, untuk itu digunakan submenu PASTE VARIABEL yang dijumpai dengan kita membuka pada area Variabel View.

Jika kita telah memiliki data yang ditulis di MS Excel, kita dapat langsung memindah data tersebut ke Data View pada Data Editor SPSS, dengan cara biasa, namun untuk dapat melanjutkan ke pengolahan data,

maka kita terlebih dahulu harus mengidentifikasi variable di Variable View.

6. Mengurutkan data

Untuk beberapa keperluan khusus, pada data yang banyak terkadang diperlukan pengurutan data berdasarkan variable tertentu. Hal ini dapat dilakukan dengan memanfaatkan perintah SORT CASES dengan cara :

- 1) sebagai contoh, akan diurutkan data dari file Berat (yang telah anda buat) berdasarkan variable nama, maka untuk memulai, letakkan pointer pada sembarang tempat pada file Berat. Klik menu Data, kemudian Sort Cases, akan tampak kotak dialog
- 2) karena akan diurutkan berdasarkan Nama, maka klik pada variable nama, dan masukkan dalam kolom Sort by
- 3) jika pengurutan dilakukan naik ke atas dari abjad A sampai Z, maka klik pada pilihan ascending, lalu tekan ok.
- 4) Dengan demikian data sudah diurutkan berdasarkan inisial nama, dan data berat dan gender akan mengikuti variable nama yang bersangkutan.

E. Keluar dari SPSS

Setelah seluruh pengerjaan di SPSS dianggap selesai, untuk keluar dari program SPSS adalah sama caranya seperti biasa, yaitu dengan menekan submenu Exit yang terdapat pada menu File; atau tekan tanda silang di pojok kanan atas.

LATIHAN/TUGAS:

Latihan pertama pada praktikum ini adalah latihan menginput data dan mengidentifikasi data, sekaligus mempraktekkan mengelola/memformat data seperti yang telah dijelaskan.. Untuk itu soal akan diberikan pada saat praktikum.

SECOND MEETING

MATERI II STATISTIK DESKRIPTIF

Statistik deskriptif lebih berhubungan dengan pengumpulan dan peringkasan data serta penyajian hasil peringkasan tersebut. Data-data yang dapat diperoleh dari hasil sensus, percobaan, survey, dll umumnya masih acak, mentah dan tidak terorganisir dengan baik (raw data). Dari data tersebut harus disusun, diringkas dengan baik dan teratur, baik dalam bentuk table atau grafik, sebagai dasar untuk berbagai pengambilan keputusan (statistic Inferensi)

Untuk mengetahui deskripsi data diperlukan ukuran yang lebih eksak, ukuran statistic tersebut disebut *summary statistics* (ringkasan statistic). Dua kelompok ukuran statistic yang sering digunakan dalam pengambilan keputusan adalah :

1. Mencari Central Tendency (kecenderungan terpusat) seperti : Mean, Median dan Modus
2. Mencari ukuran Dispersi, seperti Standart Deviasi, Varians

Untuk menggambarkan data digunakan menu **Descriptive Statistic**. Menu-menu yang berhubungan dengan statistic deskriptif adalah : Frequencies, Descriptive, Explore, Crosstab, dan Ratio pada submenu Descriptif Statistics yang terdapat pada menu Analyze, sedangkan Case Summaries pada submenu Report.

1. Frequencies

Membahas beberapa ukuran statistic dasar, seperti Mean, Median, Kuartil, Persentil, Standard Deviasi, dan lainnya. Fungsi utama dari FREQUENCIES adalah memberi gambaran sekilas dan ringkas (first look) dari sekelompok data.

2. Descriptive

Lebih kompleks daripada FREQUENCIES, menu ini dapat menyajikan ukuran statistic beberapa variable dalam satu table, serta mengetahui skor z dari satu distribusi data. Skor z biasa digunakan untuk pengujian kenormalan distribusi data.

3. Explore

Menu ini lebih lengkap daripada menu DESCRIPTIVES, berfungsi untuk memeriksa lebih teliti sekelompok data, antarlain *data screening*, menguji ada tidaknya *outlier*, uji asumsi kenormalan data dan kesamaan varians, serta penganganan data per subgroup atau per kasus.

4. Crosstab

Digunakan untuk menyajikan deskripsi data dalam bentuk tabel silang (crosstab) yang terdiri atas baris dan kolom. Selain itu, menu ini juga dilengkapi dengan analisis hubungan di antara baris dan kolom, seperti independensi di antara mereka, besar hubungannya dan lainnya (hal inisebenarnya termasuk pada statistic induktif atau inferensi, dan merupakan perluasan dari statistic deskriptif.

5. Ratio

Menu ini menyediakan ringkasan statistic untuk data hasil perbandingan dua data tertentu. Berbeda dengan menu lainnya, menu ini hanya menangani perbandingan dua data, bukan satu data yang bersifat sendiri.

6. Case Summaries

Menu ini digunakan untuk melihat lebih jauh isi statistic deskripsi yang meliputi subgroup dari sebuah kasus, seperti kasus dengan grup Pria dan Wanita; pada grup ini bias dibuat subgroup untuk pria berusia remaja dan dewasa, dan dapat pula dibagi lagi pada remaja yang tinggal di kota atau kombinasi lainnya.

FREQUENCIES

Untuk latihan, cobalah untuk membuat file data seperti di bawah ini. Data di bawah adalah data tinggi dan jenis kelamin mahasiswa Fakultas Kedokteran Angkatan 2009..

	tinggi	gender
1	170.2	1.00
2	172.5	1.00
3	180.3	1.00
4	172.5	1.00
5	159.6	2.00
6	168.5	2.00
7	168.5	1.00
8	172.5	1.00
9	174.5	1.00
10	159.6	2.00
11	170.4	2.00
12	161.3	2.00
13	172.5	1.00
14	170.4	2.00
15	168.9	2.00
16	168.9	2.00
17	177.5	2.00
18	174.5	1.00
19	168.6	2.00
20	164.8	2.00
21	170.4	1.00
22	168.9	1.00
23	164.8	2.00
24	167.2	2.00
25	167.2	2.00

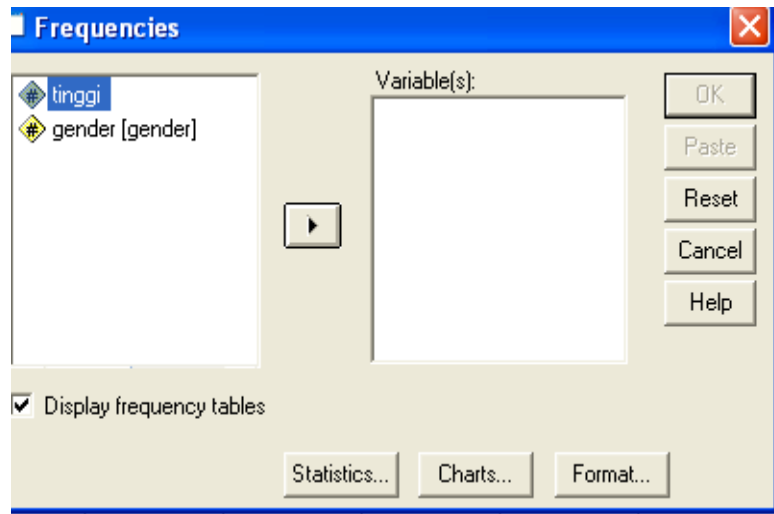
NB : untuk variabel gender, angka 1 berarti responden berjenis kelamin pria, sedangkan angka 2 berarti responden wanita (gunakan Value Label)

Cobalah untuk memasukkan data seperti yang telah dijelaskan di bab 1. Mulailah dengan mendefinisikan variabel pada Variabel View. Setelah itu, lanjutkan dengan mengisikan data pada Data View, seperti pada data di atas. Silakan simpan data ini dengan nama file DESKRIPTIF. Lanjutkan dengan melakukan analisis statistic berikut ini :

Tabel frekuensi dan statistic deskriptif untuk variabel tinggi

Variabel tinggi termasuk data kuantitatif sehingga dapat dibuat table frekuensi serta statistic yang meliputi Mean, SD, Skewness, dll. Selain itu, akan dilengkapi dengan Chart yang sesuai untuk data tersebut, yaitu histogram atau Bar Chart. Coba lakukan analisis statistic dengan langkah seperti berikut ini (masih pada file Deskriptif)

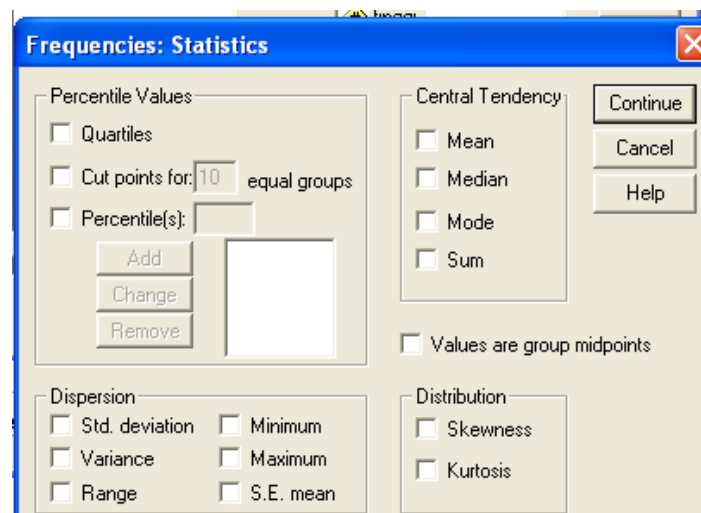
Klik menu **Analyze** → **Descriptive Statistics** → **Frequencies**, sehingga tampak kotak dialog seperti di bawah ini



Gambar 5 : kotak dialog Frequencies

Pengisian :

1. pada kotak variabel isi dengan tinggi, dengan cara sorot tulisan tinggi pada kotak kiri, lalu klik pada kotak panah di tengah sehingga tulisan tinggi masuk kedalam kotak variabel
2. klik **Statistics**, sehingga tampak kotak dialog Statistics berikut :



Gambar 6: Kotak Dialog Statistics

Pilihan Statistics meliputi berbagai ukuran statistic untuk menggambarkan data (statistic Deskriptif)

Pengisian

Pada petunjuk ini hanya akan dijelaskan beberapa ukuran statistic yang sering dipergunakan untuk menggambarkan data.

Dispersion:

Atau penyebaran data, pilihlah keenam jenis pengukuran disperse yang ada, yaitu SD, Variance, Range, Minimum, maximum, SE Mean.

Central Tendency

Atau pusat pengukuran data, yang biasa digunakan adalah **Mean** dan **Median**. Tekan continue setelah selesai input untuk melanjutkan proses berikutnya. Lalu tekan ok. Namun jika masih ingin melakukan proses yang lain, misalnya membuat grafik (chart). Sebelum tekan ok, tekan dulu Charts sehingga tampak kotak dialog Charts.

Cara pengisian **Charts** :

1. tekan charts, lalu pada pilihan Charts type.pilih charts yang akan dibuat, untuk latihan ini, pilihlah **Histogram** dan untuk menampilkan kurva normal, pilih juga **With normal curve**. Lanjutkan dengan menekan **Continue**.
2. klik pilihan **Format**, hal ini berhubungan dengan susunan format data. Pada **Order by** kita bias tentukan susunan data akan dimulai dari nilai terkecil ataukah sebaliknya. Untuk keseragaman pada latihan ini silakan tekan **Ascendeing values**. Abaikan bagian lain, lalu tekan **Continue**, dilanjutkan dengan **ok**

Output SPSS dan analisis

Output

Statistics

tinggi

N	Valid	25
	Missing	0
Mean		169.400
Std. Error of Mean		.9927
Median		168.900
Std. Deviation		4.9633
Variance		24.634
Range		20.7
Minimum		159.6
Maximum		180.3

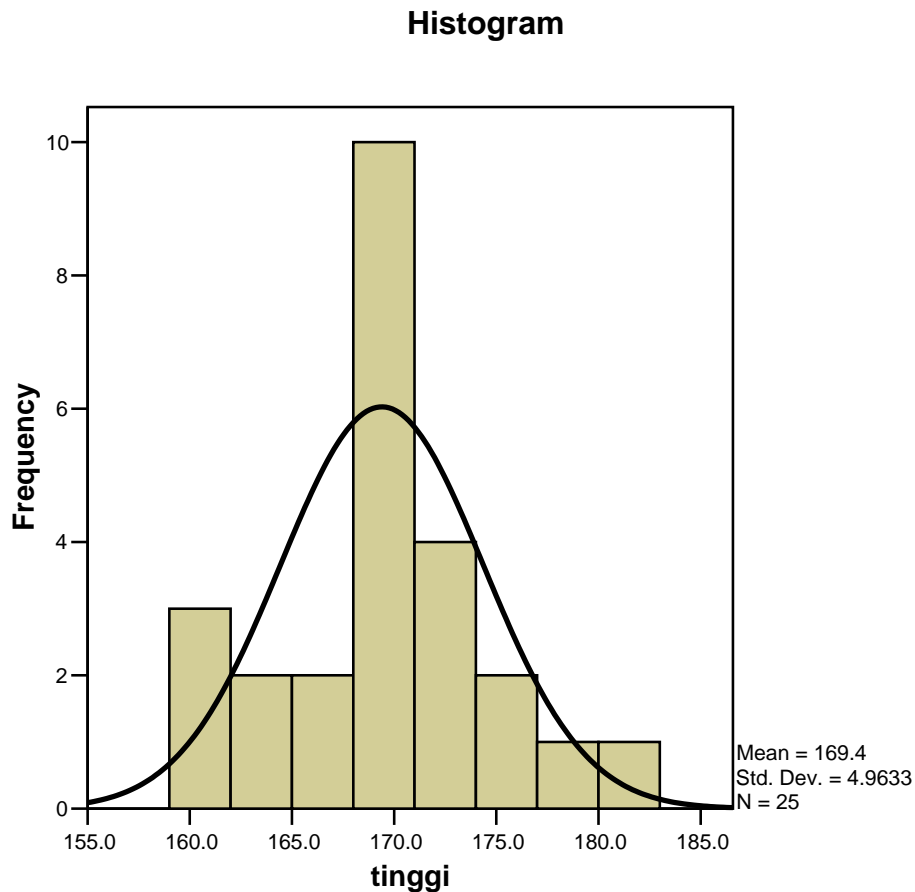
tinggi

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	159.6	2	8.0	8.0	8.0
	161.3	1	4.0	4.0	12.0
	164.8	2	8.0	8.0	20.0
	167.2	2	8.0	8.0	28.0
	168.5	2	8.0	8.0	36.0
	168.6	1	4.0	4.0	40.0
	168.9	3	12.0	12.0	52.0
	170.2	1	4.0	4.0	56.0
	170.4	3	12.0	12.0	68.0
	172.5	4	16.0	16.0	84.0
	174.5	2	8.0	8.0	92.0
	177.5	1	4.0	4.0	96.0
	180.3	1	4.0	4.0	100.0
Total		25	100.0	100.0	

Analisis

- N adalah jumlah data yang valid = 25 buah
 - Missing (data yang hilang) = 0
 - Mean (rata-rata) tinggi adalah = 169.400 cm
 - Std.Error of mean = 0.993 cm. penggunaan Standard error of mean untuk memperkirkn besar rata-rata populasi yang diperkirakan dari sample. Untuk itu, dengan standard error of mean tertentu dan pada tingkat kepercayaan 95% (SPSS sebagian besar menggunakan angka ini sebagai standar), rata-rata populasi menjadi :
 - Rata-rata \pm 2 standard error of mean
 - NB : angka 2 digunakan karena tingkat kepercayaan 95%
- Maka :
- $$169,4 \text{ cm} \pm (2 \times 0,993 \text{ cm}) = 167,414 \text{ sampai } 171,386 \text{ cm}$$
- Median atau titik tengah data jika semua data diurutkan dan dibagi dua sama besar. Angka median 168,9 cm menunjukkan bahwa 50% tinggi badan adalah 168,9 cm ke atas dan 50%-nya adalah 168,9 cm ke bawah
 - Standard Deviasi adalah 4,963 cm dan varians yang merupakan kelipatan standard deviasi adalah 24,634 cm
- Penggunaan SD untuk menilai dispersi rata-rata dari sample. Untuk itu dengan Sd tertentu dan pada tingkat kepercayaan 95% (SPSS biasa menggunakan angka ini sebagai standar) rata-rata tinggi badan menjadi :
- Rata-rata \pm 2 SD
- NB : angka 2 digunakan karena tingkat kepercayaan 95%
- Maka :
- $$169,4 \text{ cm} \pm (2 \times 4,963 \text{ cm}) = 159,474 \text{ sampai } 179,326 \text{ cm}$$
- Perhatikan kedua batas angka yang berbeda tipis dengan nilai minimum dan maksimum. Ini membuktikan sebaran data adalah baik.

- Data minimum adalah 159,6 cm sedangkan data maksimum adalah 180,3 cm
- Range adalah data maksimum – data minimum, atau dalam kasus ini :
 $180,3 \text{ cm} - 159,6 \text{ cm} = 20,7$



Analisis histogram :

Grafik histogram yang telah dibuat frekuensinya, terlihat bahwa batang histogram mempunyai kemiripan dengan kurva normal (berbentuk seperti lonceng) yang disertakan. Ini membuktikan bahwa distribusi tersebut sudah dapat dikatakan normal atau mendekati normal. Namun pengujian normalitas sendiri akan dibahas pada materi berikut ini.

Menguji Normalitas Data dan Varians

Kita harus terampil melakukan dan menginterpretasikan apakah suatu data memiliki distribusi normal atau tidak, karena pemilihan penyajian data dan uji hipotesis yang dipakai tergantung dari normal tidaknya distribusi data. Jika distribusi data normal, maka dianjurkan untuk menyajikan data dengan menggunakan mean dan standar deviasi, sedangkan jika distribusi data tidak normal dianjurkan menggunakan median dan minimum-maksimum sebagai pasangan ukuran pemusatan dan penyebaran.

Untuk pemilihan uji hipotesis, jika distribusi data normal, maka digunakan uji parametrik, namun jika distribusi data tidak normal, dipilih uji nonparametrik.

Untuk menilai normal tidaknya distribusi data, ada 2 metode, yaitu dengan :

A. metode deskriptif

Dengan melihat output SPSS pada bagian Deskriptif, berdasarkan koefisien varians, rasio skewness dan kurtosis, histogram, Q-Q plot, Detrendended normal Q-Q plot (langkah lengkap, lihat di bawah)

B. metode analitik

Dengan menggunakan uji Kolmogorov-Smirnov (untuk sampel banyak:>50) atau Shapiro-Wilk. (untuk sampel sedikit:≤ 50)

Tabel 1 : Metode untuk mengetahui suatu set data berdistribusi normal atau tidak

Metode	Parameter	Kriteria distribusi data dikatakan normal	Keterangan
Deskriptif	Koefisien varian	Nilai koefisien varians , 30%	$\frac{SD}{Mean} \times 100\%$
	Rasio Skewness	Nilai rasio skewness - $2s/d^2$	$\frac{Skewness}{SE\ Skewness}$
	Rasio Kurtosis	Nilai rasio Kurtosis - $2s/d^2$	$\frac{Kurtosis}{SE\ Kurtosis}$
	Histogram	Simetris tidak miring kiri atau kanan, tidak terlalu tinggi, tidak terlalu rendah	
	Box plot	Simetris, median tepat di tengah, tidak ada outlier atau nilai ekstrim	
	Normal Q-Q plots	Data menyebar sekitar garis	
	Detrended Q-Q plots	Data menyebar sekitar garis pada nilai 0	
Analitik	Kolmogorov-Smirnov	Nilai kemaknaan (p)>0,05	Untuk sampel besar (>50)
	Shapiro-Wilk	Nilai kemaknaan (p)>0,05	Untuk sampel kecil (≤50)

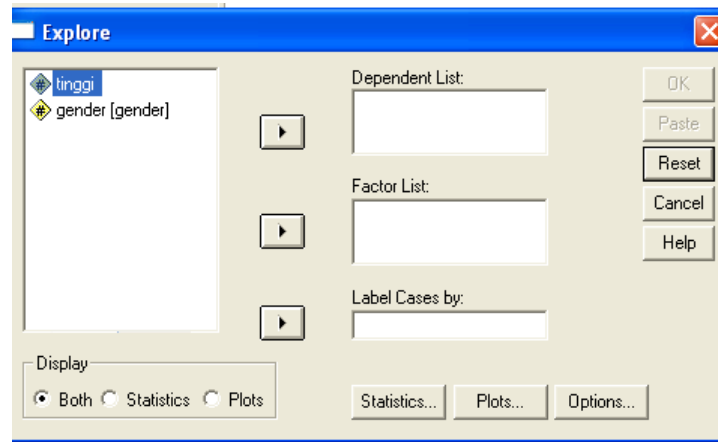
Salah satu konsep penting dalam statistic inferensi adalah :

- Apakah beberapa sample yang telah diambil berasal dari populasi yang sama (populasi data berdistribusi normal) ?
- Apakah sampel-sampel tersebut mempunyai varians yang sama?

- Pengujian kenormalan data dapat dilakukan dengan gambar ataupun uji skewness dan kurtosis. Dengan kasus yang sama yaitu tinggi badan akan diuji kenormalan dan kesamaan varians dari sample dengan bantuan uji SHAPIRO-WILK dan LILLIEFORS serta gambar NORMAL PROBABILITY PLOTS.

Langkah-langkah pengujian

- Kerjakan untuk file deskriptif, ikuti prosedur berikut :
- Dari menu utama SPSS pilih menu Analyze → Descriptive Statistics
- Pilih submenu Explore, sehingga tampak kotak dialog seperti di bawah ini :



Gambar 7 : Kotak dialog Explore

Pengisian :

- **Dependent List**, klik variable tinggi kemudian klik tanda '>' (yang sebelah atas), maka variable tinggi berpindah ke Dependent List
- **Factor List**, klik variable gender, kemudian klik tanda '>' yang sebelah atas), maka variable gender akan berpindah ke Factor List.
- **List cases by** atau kasus akan diurutkan menurut variable mana? Dalam kasus ini karena variable hanya dua dan semua sudah masuk ke dependent dan factor, maka isian tersebut bisa diabaikan.
- Lanjutkan dengan klik pilihan **Statistics** : terlihat default SPSS yang memilih Description. Untuk keseragaman, biarkan saja pilihan tersebut
NB : bagaimanapun kotak pilihan ini harus terisi karena SPSS akan menolak jika semua pilihan ditiadakan.
- Tekan **Continue** setelah selesai input, untuk melanjutkan proses berikutnya.
- Lanjutkan dengan klik pilihan **Plots**, pilihlah :
 - Pada **Box plot**, pilihlah None karena tidak akan dibuat Boxplot
 - Pada **Descriptive**, tidak usah memilih apapun, kecuali akan dibuat Stem dan Leaf atau histogram. Jika default menunjuk salah satu pilihan tersebut, maka lakukanlah deselect (klik untuk menghapus tanda √)
 - Klik pilihan **Normality Plots with test**. **Pilihan ini untuk membuat gambar uji normalitas**
 - Pada pilihan **Spread vs Level with Levene Test**, pilih **Power estimation** untuk menguji kesamaan varians
 - Tekan **Continue**

- Pada Display awal tampak pilihan Both, Statistic & Plots, pilihlah **Both**
- Tekan **Ok** jika semua pengisian sudah selesai

Output SPSS dan Analisis

Explore gender

Case Processing Summary

	gender	Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
tinggi	pria	11	100.0%	0	.0%	11	100.0%
	wanita	14	100.0%	0	.0%	14	100.0%

Descriptives

	gender		Statistic	Std. Error
tinggi	pria	Mean	172.482	.9886
		95% Lower Bound Confidence Interval for Mean	170.279	
		5% Upper Bound Confidence Interval for Mean	174.684	
		5% Trimmed Mean	172.269	
		Median	172.500	
		Variance	10.750	
		Std. Deviation	3.2787	
		Minimum	168.5	
		Maximum	180.3	
		Range	11.8	
	Interquartile Range	4.3		
	Skewness	1.263	.661	
	Kurtosis	2.546	1.279	
	wanita	Mean	166.979	1.2778
		95% Lower Bound Confidence Interval for Mean	164.218	
		5% Upper Bound Confidence Interval for Mean	169.739	
		5% Trimmed Mean	166.804	
		Median	167.850	
		Variance	22.859	
		Std. Deviation	4.7811	
Minimum		159.6		
Maximum		177.5		
Range		17.9		
Interquartile Range	5.3			
Skewness	.213	.597		
Kurtosis	.745	1.154		

Tests of Normality

	gender	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
tinggi	pria	.225	11	.125	.882	11	.110
	wanita	.166	14	.200(*)	.927	14	.279

* This is a lower bound of the true significance.
a Lilliefors Significance Correction

Analisis dari output test of normality

Pedoman pengambilan keputusan :

- Nilai Sig. atau signifikansi atau probabilitas < 0,05, distribusi adalah tidak normal (simetris)
- Nilai Sig. atau signifikansi atau nilai probabilitas > 0,05 distribusi adalah normal (simetris)

Ada 2 macam alat uji kenormalan distribusi data yang biasa digunakan yaitu :

- Komogorov Smirnov dengan keterangan adalah sama dengan uji Lilliefors (lihat tanda "a" di bawah table). Diperoleh baik untuk gender pria maupun wanita, tingkat signifikansi atau nilai probabilitas di atas 0,05 (0,125 dan 0,200 lebih besar dari 0,05); maka dapat dikatakan distribusi kedua sample adalah normal.
- Shapiro Wilk baik untuk gender pria maupun wanita, tingkat signifikansi atau nilai probabilitas di atas 0,05 (0,125 dan 0,344 lebih besar dari 0,05; maka dapat dikatakan distribusi tinggi badan pria dan wanita adalah normal

Test of Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
tinggi	Based on Mean	1.507	1	23	.232
	Based on Median	1.260	1	23	.273
	Based on Median and with adjusted df	1.260	1	21.148	.274
	Based on trimmed mean	1.549	1	23	.226

Analisis dari output test of Homogeneity

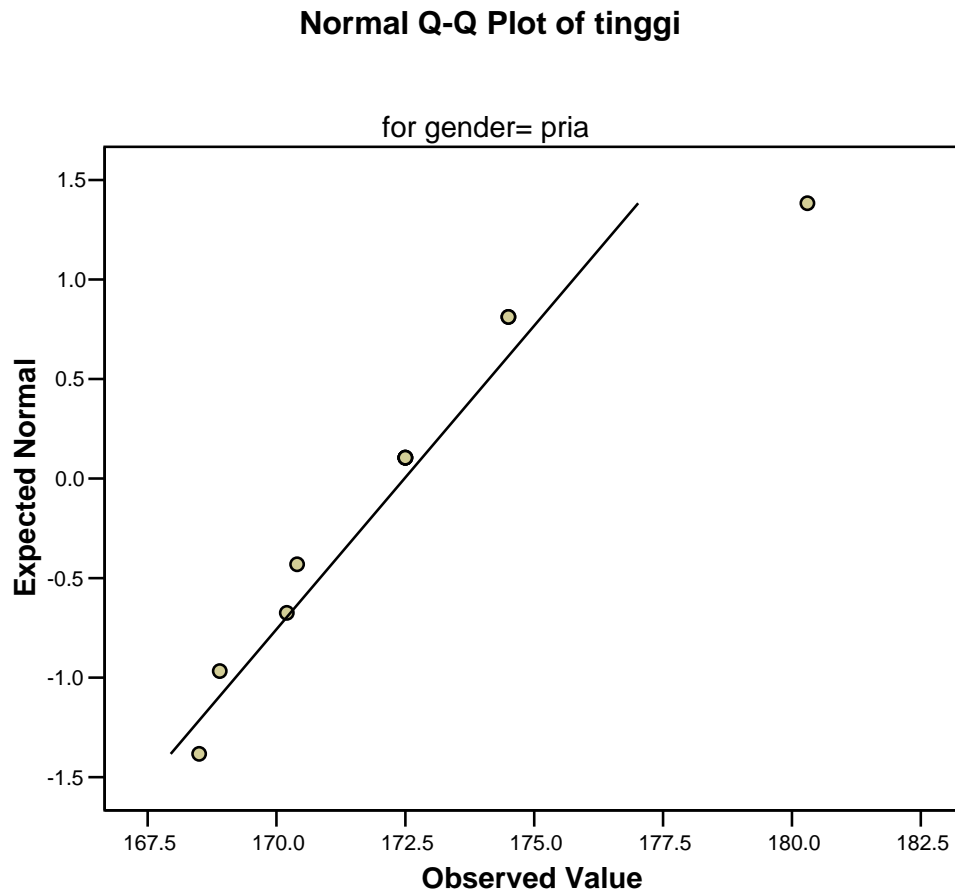
Output ini untuk menguji apakah dua sample yang diambil mempunyai varians yang sama. Pedoman pengambilan keputusan :

- Nilai Sig, atau signifikansi atau nilai probabilitas < 0,05, data berasal dari populasi-populasi yang mempunyai varians tidak sama

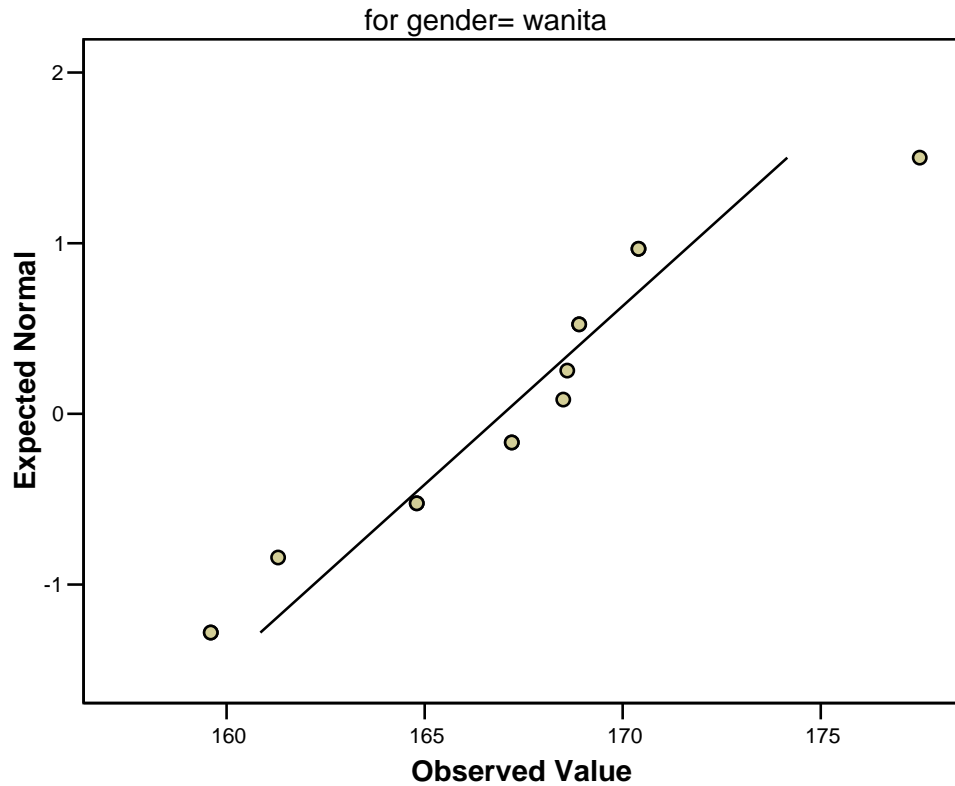
- Nilai Sig, atau signifikansi atau nilai probabilitas $> 0,05$, data berasal dari populasi-populasi yang mempunyai varians sama

Pada output di atas, ada satu alat uji yang digunakan, yaitu Levene test. Hasil pengujian menunjukkan tingkat signifikansi atau nilai probabilitas mean (rata-rata) yang berada di atas $0,05$ ($0,232$ lebih besar dari $0,05$). Demikian pula jika dasar pengukuran adalah median data, angka Sig. adalah $0,273$ yang tetap di atas $0,05$. Maka bisa dikatakan data berasal dari populasi-populasi yang mempunyai varians sama, atau sample pria dan wanita tersebut di atas di ambil dari populasi pria dan wanita yang mempunyai varians tinggi badan sama, dalam arti varians populasi tinggi badan pria sama dengan populasi tinggi badan wanita.

Normal Q-Q Plots



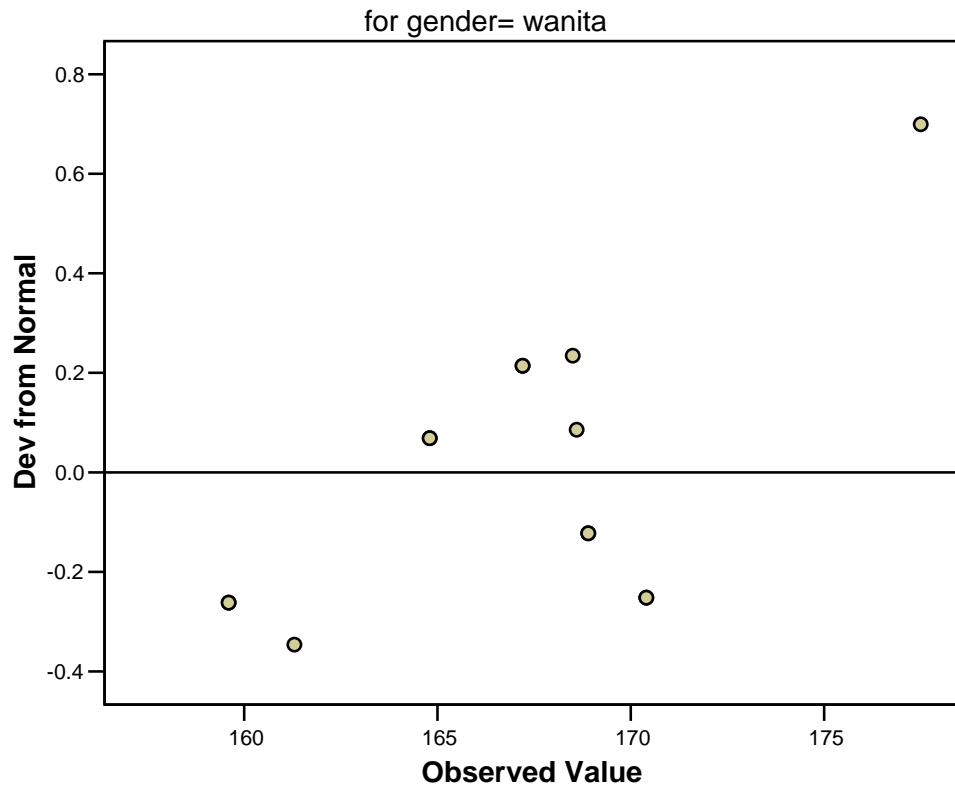
Normal Q-Q Plot of tinggi



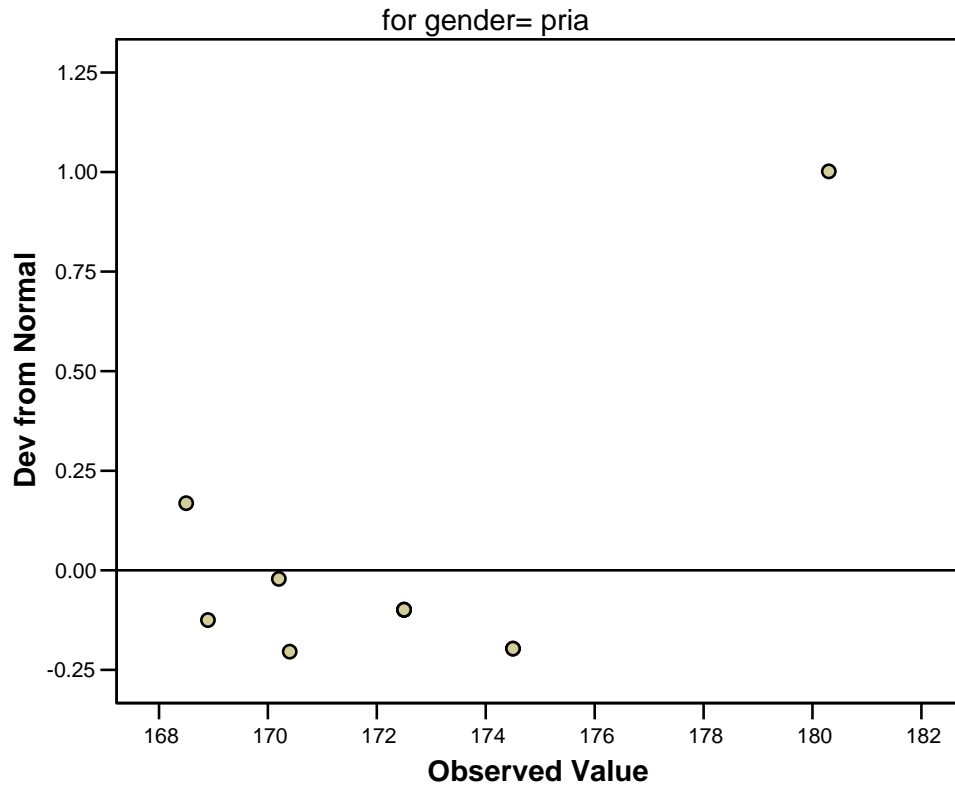
Pada gambar Q-Q plot untuk variable tinggi, baik untuk gender pria maupun wanita, terlihat ada garis lurus dari kiri bawah ke kanan atas. Garis ini berasal dari nilai z (lihat pembahasan z pada materi kuliah). Jika suatu distribusi data normal, maka data akan tersebar di sekeliling garis. Pada kasus ini terlihat bahwa memang data tersebar di sekeliling garis (kecuali ada 1 data pria dan 1 data wanita yang “outlier”).

Detrended Normal Q-Q Plots

Detrended Normal Q-Q Plot of tinggi



Detrended Normal Q-Q Plot of tinggi



Kedua output ini mendeteksi pola-pola dari titik-titik yang bukan dari kurva normal. Terlihat pada kedua gender sebagian besar data terpola di sekitar garis garis , kecuali 2 buah data yang masing-masing ada di pojok kanan atas. Hal ini membuktikan bahwa distribusi data adalah normal.

Latihan/tugas:

Mahasiswa diberi contoh set data (diberikan pada saat praktikum) untuk diinput dan selanjutnya mempraktekkan melakukan uji normalitas dengan cara-cara yang telah diberikan sekaligus menginterpretasi hasil uji tersebut.

THIRD MEETING

MATERI III STATISTIK INFERENSI

Pada pembahasan mengenai statistic deskriptif telah diketahui bahwa untuk mengetahui karakteristik sebuah populasi dilakukan prosedur statistic deskriptif. Dari prosedur itu dapat diperoleh hasil berupa parameter yang menggambarkan ciri-ciri populasi tersebut. Jika populasi tidak begitu besar, hal ini tidak menimbulkan kesulitan. Namun jika populasi begitu besar, misalnya mencakup balita di seluruh Indonesia, tentunya akan menimbulkan kesulitan untuk melakukan penggambaran yang jelas tentang populasi dan berbagai pengambilan keputusan sehubungan dengan ciri-ciri populasi. Untuk itu dapat dilakukan pengambilan sample sejumlah tertentu populasi tersebut kemudian dengan sample tersebut dilakukan berbagai keputusan (inferensi) terhadap populasi, yaitu :

- Melakukan perkiraan (estimasi) terhadap populasi,
Misalnya : berapa rata-rata populasi? Berapa deviasi standarnya?
- Melakukan uji hipotesis terhadap parameter populasi
Misal : apakah rata-rata populasi yang diperkitrakan sudah benar?

Dengan kata lain, dari informasi sample yang telah ada akan dilakukan berbagai penggambaran dan kesimpulan terhadap isi populasi. Kegiatan itu disebut statistic inferensi

Dalam prakteknya, metode statistic inferensi cukup beragam, dan salah satu criteria penting dalam pemilihan metode statistic yang akan digunakan adalah melihat distribusi sebuah data. Jika data yang diuji berdistribusi normal atau mendekati normal, maka selanjutnya dengan data-data tersebut dapat dilakukan berbagai inferensi atau pengambilan keputusan dengan metode statistic parametric.

Tetapi jika terbukti data tidak berdistribusi normal atau jauh dari criteria distribusi normal, maka tidak dapat digunakan metode parametric, sehingga untuk kegiatan inferensi digunakan statistic nonparametric.

Kegiatan inferensi dapat dibedakan menjadi :

- Pengujian beda rata-rata, yang meliputi uji t dan uji F (Anova)
- Pengujian asosiasi (hubungan) dua variable atau lebih, alat uji yang digunakan seperti Chi-square, korelasi dan regresi.

Macam-macam uji hipotesis sesuai dengan skala pengukuran dapat dilihat pada tabel berikut ini :

Tabel 2 : skala pengukuran dan uji hipotesis yang digunakan

Skala pengukuran	Jenis Hipotesis				
	Komparatif				Korelatif
Numerik	Tidak berpasangan		Berpasangan		Pearson
	2 kelompok	>2 kelompok	2 kelompok	> 2 kelompok	
	Uji t tidak berpasangan ↓	One way ANOVA ↓	Uji t berpasangan ↓	Repeated ANOVA ↓	
Kategorik (Ordinal)	Mann Whitney	Kruskal Wallis	Wilcoxon	Friedman	Spearman Sommers Gamma
Kategorik (Nominal/Ordinal)	Chi-Square Fisher Kolmogorov-Smirnov		Mc Necmar, Cochran Marginal Homogeneity Wilcoxon, Friedman		Koefisien Kontngensi Lambda

Statistic Inferensi dalam SPSS

SPSS menyediakan berbagai metode parametrik untuk melakukan inferensi terhadap data statistic. Beberapa menu statistic parametric yang tersedia dalam SPSS adalah ; COMPARE MEANS, General Linear Model (GLM), CORRELATE dan REGRESSION.

COMPARE MEANS

Menu ini meliputi :

- **MEANS**
- **Uji t**, yang meliputi :
 - Uji t satu sample (**One Sample t Test**)
 - Uji t untuk 2 sampel independent (**Independent Sample t Test**)
 - Uji t untuk 2 sampel berpasangan (**Paired Sampel t Test**)
- **One Way Anova** : jika uji t untuk 2 sampel, maka Anova digunakan untuk menguji lebih dari 2 sampel

General Linear Model

GLM merupakan kelanjutan dari ANOVA, yaitu pada GLM dibahas satu variable dependen namun mempunyai satu atau lebih factor.

CORRELATE

Membahas uji hubungan antara dua variable

REGRESSION

Membahas pembuatan model regresi untuk menggambarkan hubungan dua variable atau lebih

Paired Sample t Test (Uji t berpasangan)

Fasilitas ini digunakan untuk menguji 2 sampel yang berpasangan (paired). Paired dalam hal ini diartikan sebagai sebuah sample dengan subjek yang sama, namun mengalami 2 perlakuan atau pengukuran yang berbeda, seperti subjek A mendapat perlakuan I kemudian perlakuan II. Contoh berikut akan menjelaskan hal ini.

Contoh :

Produsen obat penurun berat badan ingin mengetahui apakah obat yang diproduksinya benar-banar mempunyai efek terhadap penurunan berat badan konsumennya. Untuk itu diambil sample yang terdiri dari 10 orang yang masing-masing diukur berat badannya sebelum menggunakan obat tersebut. setelah sebulan mengkonsumsi obat tersebut secara rutin dan teratur, kembali sample diukur berat badannya. Berikut adalah hasilnya (dalam kilogram)

	Sebelum	Sesudah		Sebelum	Sesudah
1	76.85	76.22	6	88.15	82.53
2	77.95	77.89	7	92.54	92.56
3	78.65	79.02	8	96.25	92.33
4	79.25	80.21	9	84.56	85.12
5	82.65	82.65	10	88.25	84.56

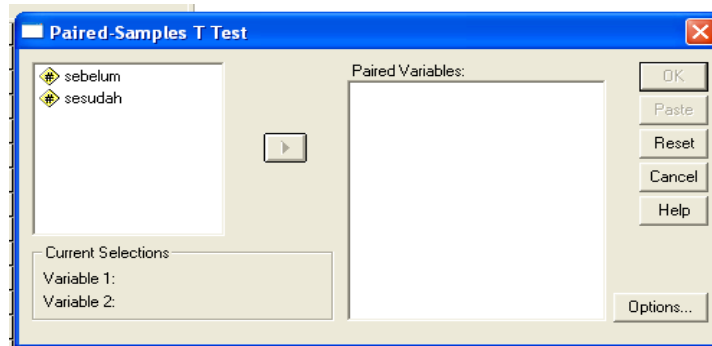
Penyelesaian :

Perhatikan ciri dari sample berpasangan, yakni subjeknya tetap sepuluh orang; kepada setiap orang tersebut diberikan 2 kali perlakuan, yang dalam kasus ini adalah efektivitas sebuah obat. Kasus di atas terdiri atas 2 sampel yang berhubungan atau berpasangan satu dengan yang lain, yaitu sample sebelum makan obat dan sample sesudah makan obat. Di sini populasi berdistribusi **normal** dan karena **sample sedikit**, dipakai uji t untuk 2 sampel yang berpasangan (paired)

Tugas :

Cobalah untuk memasukkan data di atas pada data editor SPSS. Lanjutkan dengan pengolahan data dengan langkah-langkah berikut :

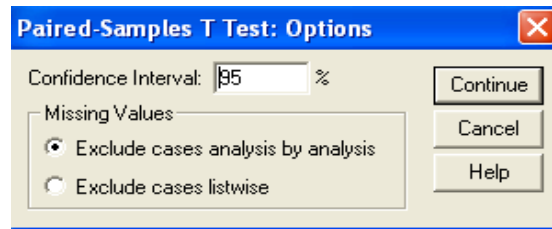
- ✚ Simpanlah data yang telah anda buat dengan nama uji t paired dan tempatkan pada direktori yang dikehendaki
- ✚ Bukalah file tadi, pilih menu ANALYZE → Compare-Means
- ✚ Pilih Paired-Sample T Test sehingga tampak kotak dialog seperti ini :



Gambar 3.1 : kotak dialog Paired t Test

Pengisian :

- ✚ Karena akan diuji data sebelum dan sesudah, maka klik variable sebelum, lanjutkan dengan klik variable sesudah, lanjutkan dengan meng-klik kotak kecil bertanda ">" maka pada kotak Paired Variable terlihat : sebelum-sesudah.
NB : variable sebelum dan sesudah harus dipilih bersamaan, jika tidak, SPSS tidak dapat menginput dalam kolom Paired Variable.



Gambar 3.2 : Kotak dialog Option

Pengisian :

- ✚ **Confidence Interval** : atau tingkat kepercayaan, sebagai default, SPSS menggunakan tingkat kepercayaan 95% atau tingkat signifikansi 100%-95% = 5%.
- ✚ **Missing Value** atau data yang hilang, karena data komplit, maka abaikan bagian ini (tetap pada default SPSS, yaitu **Exclude case analysis by analysis**)
- ✚ Tekan **Continue** jika pengisian dianggap selesai, lanjutkan dengan klik **Ok** untuk mengakhiri pengisian prosedur analisis.

**Output
T test**

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	sebelum	84.5100	10	6.63931	2.09953
	sesudah	83.3090	10	5.58235	1.76530

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	sebelum & sesudah	10	.943	.000

Paired Samples Test

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	sebelum - sesudah	1.20100	2.30738	.72966	-.44960	2.85160	1.646	9	.134

Analisis

Output Bagian 1

Output ini menunjukkan ringkasan statistic dari kedua sample. Berat badan sebelum minum obat rata-rata 84,51 kg, sedangkan setelah minum obat, konsumen mempunyai rata-rata berat badan 83,3090 kg

Output bagian 2

Output menunjukkan korelasi antara kedua variable yang menghasilkan angka 0,943 dengan nilai probabilitas jauh di bawah 0,05 (lihat nilai signifikansi output yang 0,000) ini menyatakan bahwa korelasi antara berat sebelum dan sesudah minum obat adalah sangat erat dan benar-benar berhubungan secara nyata.

Output bagian 3

Hipotesis

H_0 = kedua rata-rata populasi adalah identik (rata-rata populasi berat sebelum minum obat dan sesudah minum obat adalah sama/tidak berbeda secara nyata)

H_1 = kedua rata-rata populasi adalah tidak identik (rata-rata populasi berat sebelum minum obat dan sesudah minum obat adalah tidak sama/ berbeda secara nyata)

Pengambilan keputusan

a. Berdasarkan perbandingan t hitung dengan t table (dasar pengambilan keputusan sama dengan uji t) :

Jika statistic hitung (angka t output) > statistic table (table t), maka H ditolak

Jika statistic hitung (angka t output) < statistic table (table t), maka H diterima

T hitung dari output adalah 1,646

Sedang statistic table dapat dilihat pada table t :

- Tingkat signifikansi (α) adalah 5% (lihat input data pada bagian OPTION yang memilih tingkat kepercayaan 95%)
- Df atau derajat kebebasan adalah n (jumlah data) - 1 atau $10 - 1 = 9$
- Uji dilakukan dua sisi karena akan diketahui apakah rata-rata sebelum sama dengan sesudah atau tidak. Jadi dapat lebih besar atau lebih kecil, karenanya dipakai uji dua sisi. Perlunya uji dua sisi dapat diketahui pula dari output SPSS yang menyebutkan adanya Two tailed test.

Dari table t, diperoleh angka 2,2622

Oleh karena t hitung terletak pada daerah H_0 diterima, maka berarti obat tersebut tidak efektif dalam upaya menurunkan berat badan

b. Berdasarkan nilai probabilitas :

- Jika probabilitas > 0,05, maka H_0 diterima
- Jika probabilitas < 0,05, maka H_0 ditolak

Keputusan :

Terlihat bahwa t hitung adalah 1,646 dengan probabilitas 0,134 (> 0,05) maka H_0 diterima atau dapat diartikan bahwa berat badan (BB) sebelum dan sesudah minum obat relative sama. Atau dengan kata lain, obat penurun BB tersebut tidak efektif dalam menurunkan BB secara nyata.

Pada prinsipnya pengambilan keputusan berdasar t hitung dan t table ataupun berdasar angka probabilitas adalah sama hasilnya. Tetapi untuk kemudahan dan kepraktisan, penggunaan angka probabilitas lebih sering digunakan sebagai dasar pengambilan keputusan inferensi.

Catatan :

Dalam kasus ini ,dapat pula dinyatakan bahwa terdapat perbedaan Mean sebesar 1,2010. angka ini berasal dari :

$$\begin{aligned} & \text{BB sebelum minum obat} - \text{BB sesudah minum obat} \\ & = 84,5100 \text{ kg} - 83,3090 \text{ kg} = 1,2010 \text{ kg} \end{aligned}$$

Perbedaan ini memiliki range antara lower (batas bawah) sebesar -0,4496 (tanda negative berarti BB sebelum minum obat < BB sesudah minum obat) sampai upper (batas atas) 2,8516 kg. Walau demikian, dari uji t terbukti bahwa perbedaan 1,2010 kg dengan range > 0 kg - -2.8616 kg tersebut tidak cukup berarti untuk menyatakan bahwa obat tersebut efektif untuk menurunkan BB.

One Sample t Test

Pengujian satu sample pada prinsipnya bertujuan untuk menguji apakah suatu nilai tertentu (yang diberikan sebagai pembanding) berbeda secara nyata atau tidak dengan rata-rata sebuah sample.

Sebagai contoh, diduga rata-rata konsumsi rokok pada karyawan PT X adalah sebanyak 3 batang perhari. Jika seluruh karyawan PT X dianggap populasi, maka angka tersebut adalah suatu parameter. Selanjutnya akan dibuktikan secara statistic apakah konsumsi tersebut memang benar demikian. Untuk itu diambil sejumlah sample dan pada sample itu dilakukan perhitungan rata-rata konsumsi rokok perhari. Kemudian dilakukan proses perbandingan yang disebut sebagai uji satu sample (one sample test). Digunakan uji t karena jumlah sample yang diambil pada uji semacam itu kurang dari 30 buah.

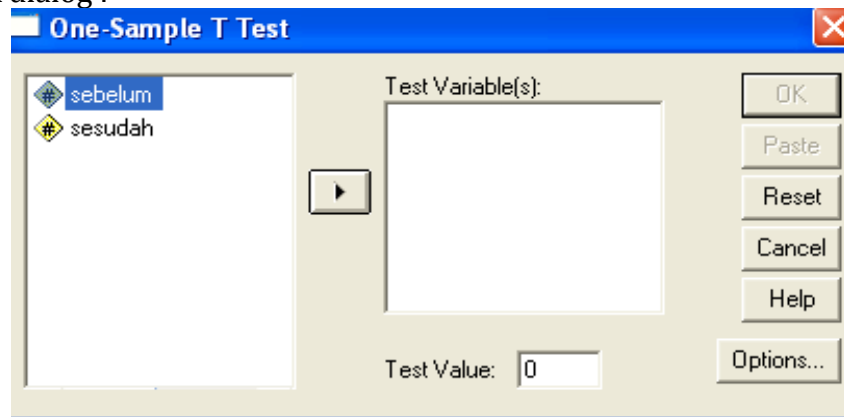
Untuk mudah dan cepatnya, kita gunakan data BB pada kasus Paired t test, hanya saja kita gunakan data yBB sebelum minum obat.

Misalnya diduga populasi rata-rata berat sebelum minum obat = 84,51 kg (lihat output terdahulu), untuk membuktikan hal tersebut, sekelompok anak muda ditimbang, dan mereka mempunyai BB rata-rata 90 kg. dengan data di atas, apakah dapat disimpulkan bahwa berat rata-rata populasi memang 84,51 kg ?

Penyelesaian :

Kasus di atas terdiri atas 1 sampel yang akan dipakai dengan nilai populasi hipotesis yaitu 90 kg. Di sini populasi diketahui berdistribusi normal, dan karena sample sedikit, digunakan uji t.

- Data diinput seperti cara yang sudah dipraktekkan pada latihan sebelum ini. Atau pada kasus ini digunakan data file yang sudah ada.
- Pengolahan data, ikuti langkah-langkah berikut :
 - Buka ;embar kerja file uji t paired yang telah ada
 - Klik menu Analyze → Compare Means → One Sampel t Test, sehingga tampak kotak dialog :



Gambar 3.3: kotak dialog One Sample t Test

Pengisian :

- ❖ Masukkan variable sebelum ke dalam kotak Test Variable, dengan cara : sorot variable sebelum → klik kotak panah kecil
- ❖ Ketik 90 pada kotak Test Value (karena akan diuji nilai hipotesis 90 kg)
- ❖ Karena tidak ada data missing dan tingkat kepercayaan tetap 95%, abaikan pilihan OPTIONS
- ❖ Tekan OK untuk proses data

Output dan analisis T test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
sebelum	10	84.5100	6.63931	2.09953

Output di atas menunjukkan ringkasan statistic dari variable sebelum. Rata-rata BB sebelum minum obat adalah 84,51 kg.

One-Sample Test

	Test Value = 90					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
	Lower	Upper				
sebelum	-2.615	9	.028	-5.49000	-10.2395	-.7405

Hipotesis :

H_0 = berat kelompok anak muda tidak berbeda dengan rata-rata berat populasi sebelum minum obat

H_0 = berat kelompok anak muda berbeda dengan rata-rata berat populasi sebelum minum obat

Pengambilan keputusan

a. Berdasarkan perbandingan t hitung dan t Tabel

❖ Jika statistic hitung (angka t output) > statistic table (table t), maka H_0 ditolak

❖ Jika statistic hitung (angka t output) < statistic table (table t), maka H_0 diterima

T hitung dari output adalah -2,615

Sedang statistic table dapat dihitung pada table t :

- ❖ Tingkat signifikansi (α) adalah 5% (lihat input data pada bagian OPTIONS yang memilih tingkat kepercayaan 95%)
- ❖ Df atau derajat kebebasan adalah n(jumlah data) - 1 atau 10 - 9 = 9
- ❖ Uji dilakukan 2 sisi karena akan dicari apakah rata-rata sebelum sama dengan berat anak muda ataukah tidak. Jadi dapat lebih kecil atau lebih besar. Perlunya Uji 2 sisi dapat diketahui pula dari output SPSS yang menyebut adanya Two tailed test.
- ❖ Karena t hitung terletak pada daerah H_0 , maka dapat disimpulkan bahwa berat kelompok anak muda tersebut memang berbeda dengan berat rata-rata populasi sebelum minum obat.

b. Berdasarkan probabilitas

- ❖ Jika probabilitas > 0,05, maka H₀ diterima
- ❖ Jika probabilitas < 0,05, maka H₀ ditolak

Keputusan :

T hitung = -2,615 dengan probabilitas 0,028 (<0,05) maka H₀ ditolak atau berarti berat kelompok anak muda tersebut memang berbeda dengan berat rata-rata populasi sebelum minum obat.

Independent Sample t Test (Uji t tidak berpasangan)

= uji t untuk 2 sampel independent (bebas)

Contoh kasus :

Seorang peneliti ingin mengetahui ada tidaknya perbedaan tinggi dan berat badan seorang pria dan seorang wanita. Untuk itu , 7 pria dan 7 wanita masing-masing diukur tinggi dan berat badannya. Berikut adalah hasil pengukurannya :

	tinggi	berat	gender
1	174.5	65.8	pria
2	178.6	62.7	pria
3	170.8	66.4	pria
4	168.2	68.9	pria
5	159.7	67.8	pria
6	167,8	67.8	pria
7	165.5	65.8	pria
8	154.7	48.7	wanita
9	152.7	45.7	wanita
10	155.8	46.2	wanita
11	154.8	43.8	wanita
12	157.8	58.1	wanita
13	156.7	54.7	wanita
14	154.7	49.7	wanita

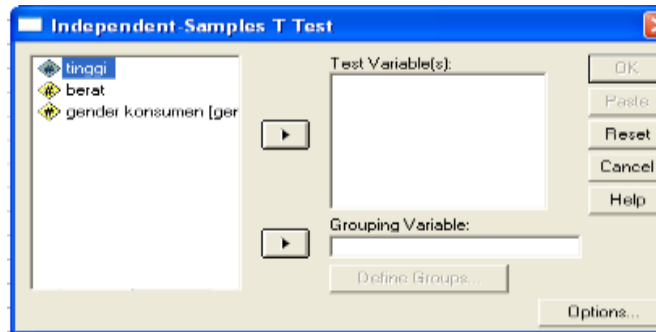
Penyelesaian :

Pada kasus ini terdapat 2 sampel yang bebas satu dengan yang lain, yaitu sample bergender pria tentu berbeda dengan sample bergender wanita. Di sini populasi diketahui berdistribusi normal dan karena sample sedikit, maka digunakan uji t untuk 2 sampel.

Untuk mempelajari uji statistic ini, inputlah data di atas ke dalam lembar kerja SPSS, dengan cara seperti yang telah biasa dilakukan pada SPSS.

Ikuti langkah-langkah berikut ini

- Setelah data selesai diinput, pilih menu **Analyze** → **Compare Means** → **Independent Sample t Test**



Gambar 3.4 : Kotak dialog Independent t Test

Pengisian ;

- Sorot variable tinggi, klik panah dalam kotak kecil bagian atas sehingga variable tinggi masuk ke dalam kotak **Test Variable**.
- Ulangi langkah ini untuk variable berat
- Sorot variable gender, klik pada panah dalam kotak kecil bagian bawah, sehingga pada kotak **Grouping Variable** terisi gender. Dengan demikian kotak **Define Groups** menjadi aktif.
- Klik **Define Groups**, sehingga muncul kotak dialog Define Groups
- Untuk group 1, isi dengan angka 1 yang berarti grup 1 berisi tanda 1 atau “pria”
- Untuk group 2, isi dengan angka 2 yang berarti grup 2 berisi tanda 2 atau “wanita”
- Tekan Continue untuk kembali ke menu sebelumnya.
- Tekan OK untuk mengakhiri pengisian prosedur analisis dan memulai proses data.

Output dan Analisis

Group Statistics

	gender sampel	N	Mean	Std. Deviation	Std. Error Mean
tinggi	pria	7	169.300	6.1351	2.3189
	wanita	7	155.314	1.6426	.6208
berat	pria	7	66.457	2.0231	.7647
	wanita	7	49.557	5.1555	1.9486

Output di atas menunjukkan ringkasan statistic dari kedua sample. Dari data tersebut ada tidaknya perbedaan secara signifikan (jelas dan nyata) antara BB dan tinggi antara pria dan wanita belum dapat diketahui. Untuk itu analisis dilanjutkan pada bagian kedua output.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
tinggi	Equal variances assumed	5.475	.037	5.826	12	.000	13.9857	2.4005	8.7554	19.2160
	Equal variances not assumed			5.826	6.856	.001	13.9857	2.4005	8.2850	19.6864
berat	Equal variances assumed	4.345	.059	8.074	12	.000	16.9000	2.0933	12.3392	21.4608
	Equal variances not assumed			8.074	7.805	.000	16.9000	2.0933	12.0518	21.7482

Uji t 2 sampel dilakukan dalam 2 tahap, yaitu pertama menguji apakah varians dari dua populasi dapat dianggap sama? Tahap selanjutnya dilakukan pengujian untuk melihat ada tidaknya perbedaan rata-rata populasi. Pada dasarnya uji t mensyaratkan adanya kesamaan varians dari dua populasi yang diuji. Jika asumsi tersebut tidak terpenuhi, maka SPSS akan menyediakan alternative jawaban uji t yang lain.

Tinggi badan

Pertama-tama dilakukan uji F untuk menguji asumsi kesamaan varians

Hipotesis :

H_0 = kedua varians populasi adalah identik (varians populasi tinggi badan pria dan wanita adalah sama)

H_1 = kedua varians populasi adalah tidak identik (varians populasi tinggi badan pria dan wanita adalah berbeda)

Pengambilan Keputusan

- Jika probabilitas > 0,05, maka H_0 diterima
- Jika probabilitas < 0,05, maka H_0 ditolak

Keputusan :

- Terlihat bahwa F hitung untuk tinggi badan dengan equal variance assumed (diasumsi kedua varians sama atau menggunakan pooled variance t test) adalah

5,475 dengan probabilitas 0,037. karena probabilitas $< 0,05$ maka H_0 ditolak, artinya kedua varians benar-benar berbeda.

- Oleh karena perbedaan yang nyata dari kedua varians, maka penggunaan varians untuk membandingkan rata-rata populasi dengan t test sebaiknya menggunakan dasar *Equal variance not assumed* (diasumsi kedua varians tidak sama)
- Setelah uji asumsi kesamaan varians selesai, kemudian dilakukan analisis dengan menggunakan t test untuk mengetahui apakah rata-rata tinggi badan pria dan wanita adalah berbeda secara signifikan?

Hipotesis :

Hipotesis untuk kasus ini

H_0 = kedua rata-rata populasi adalah identik (rata-rata populasi tinggi badan pria dan wanita adalah sama)

H_0 = kedua rata-rata populasi adalah tidak identik (rata-rata populasi tinggi badan pria dan wanita adalah berbeda)

Catatan : tidak seperti asumsi sebelumnya yang menggunakan varians, sekarang dipakai mean atau rata-rata hitung

Karena tidak ada kalimat lebih tinggi atau kurang tinggi, maka dilakukan uji dua sisi.

Pengambilan keputusan

- Jika probabilitas $> 0,05$, maka H_0 diterima
- Jika probabilitas $< 0,05$, maka H_0 ditolak

Keputusan :

Dari table output dapat diketahui bahwa t hitung untuk Tinggi badan dengan Equal Variance not assumed (diasumsi kedua varians tidak sama atau menggunakan separate variance test) adalah 5,826 dengan probabilitas 0,001. Oleh karena probabilitas $< 0,05$, maka H_0 ditolak, berarti kedua rata-rata tinggi badan pria dan wanita benar-benar berbeda, Jika dilihat dari rata-ratanya pria mempunyai rata-rata tinggi lebih besar dari wanita.

Perhatikan bahwa perbedaan dari penggunaan Equal variance assumed ke Equal variance not assumed mengakibatkan menurunnya degree of freedom (derajat kebebasan) dari 12 menjadi 6,856 atau kegagalan mengasumsikan kesamaan varians berakibat keefektivan ukuran sample menjadi berkurang sekitar 40% lebih.

Tugas :

Berdasar table output yang sama, coba jelaskan (analisis) arti statistic dari nilai-nilai yang ada pada table tersebut, untuk variable berat badan. Perhatikan nilai F hitung, probabilitas. Ada tidaknya beda nyata antara kedua varians, perbandingan rata-rata populasi menggunakan Equal variance assumed atukah not assumed.

Ringkasan dari test di atas :

- Diuji dengan F test dahulu (Levene test) apakah hipotesis varians sama , ditolak atukah tidak

- Jika hipotesis ditolak, atau varians berbeda, maka untuk membandingkan Means digunakan t test dengan asumsi varians tidak sama
- Jika hipotesis diterima, atau varians sama, maka terlihat otomatis pada output SPSS tidak ada angka untuk t test Equal variance not assumed. Oleh karena itu test dengan uji t untuk membandingkan means langsung dilakukan dengan equal variance assumed.

One way ANOVA

Untuk pengujian lebih dari 2 sampel uji statistic yang dapat digunakan adalah uji ANOVA . asumsi yang digunakan pada pengujian menggunakan ANOVA adalah :

- a. populasi-populasi yang akan diuji berdistribusi normal
- b. varians dari populasi-populasi tersebut adalah sama
- c. sample tidak berhubungan satu dengan yang lain

Contoh :

Fulan melakukan penelitian KTI mengenai pengaruh berbagai dosis ekstrak pare terhadap kecepatan gerakan sperma. Untuk itu ia menggunakan hewan coba mencit jantan sebanyak 12 ekor. Dari ke 12 mencit dibagi menjadi 4 kelompok (masing-masing terdiri 3 ekor mencit) dengan perlakuan pemberian ekstrak pare yang berbeda. Kelompok A diberi perlakuan dosis ekstrak 100mg/hari, kelompok B 200 mg/hari, C 300 mg/hari, D 400 mg/hari. Pemberian ekstrak pare diberikan selama 30 hari berurut-turut. Dari percobaan tersebut diperoleh hasil sbb :

Mencit	A	B	C	D
1	22,4	20,6	15,3	13,3
2	21,5	19,3	16,1	12,7
3	22,1	18,9	15,8	13,5

Penyelesaian :

Kasus di atas terdiri atas sample yang bebas satu sama lain, yaitu dosis A, B, C, D yang berbeda dosisnya. Di sini populasi keempat kelompok dosis diketahui berdistribusi normal, karena jumlah sample lebih dari 2 kelompok, maka digunakan uji ANOVA

Input data :

Supaya data dapat diolah dengan SPSS, maka format data harus disesuaikan dengan formatnya SPSS. Cobalah menginput data tersebut sehingga diperoleh hasil seperti di bawah ini :

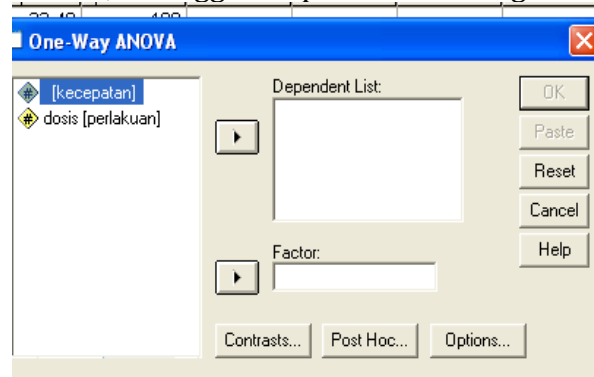
	kecepatan	perlakuan	var
1	22.40	100	
2	21.50	100	
3	22.10	100	
4	20.60	200	
5	19.30	200	
6	18.90	200	
7	15.30	300	
8	16.10	300	
9	15.80	300	
10	13.30	400	
11	12.70	400	
12	13.50	400	
13			
14			
15			

Gambar 3.5 : Data View ANOVA

Pengolahan data :

Langkah-langkah:

- ❖ (masih di area data view seperti di atas) pilih menu **Analyze** → **Compare Means** → **One-Way ANOVA**, sehingga tampak kotak dialog berikut ini



Gambar 3.6 : Kotak Dialog ANOVA

Pengisian :

- ❖ **Dependent List** (variable yang akan diuji), masukkan variable kecepatan
- ❖ Factor atau grup, masukkan dosis(perlakuan)
- ❖ Klik **Options**
- ❖ Untuk keseragaman, pilih **Deskriptive & homogeneity of variance**, hingga kotak kecil terisi tanda \checkmark
- ❖ Untuk **Missing Value**, karena data kita komplit tidak ada yang hilang, maka pilihan ini dapat diabaikan
- ❖ Tekan **Continue** untuk melanjutkan proses pengolahan data, kembali ke kotak dialog **ANOVA**, lanjutkan dengan memilih Post Hoc atau analisis lanjutan. Untuk keseragaman, pilih **Tukey**

- ❖ Tekan **Continue** jika pengisian dianggap selesai
- ❖ Kemudian tekan **OK** untuk mengakhiri pengisian prosedur analisis.

Output SPSS dan Analisis :
Oneway

Descriptives

Perla kuan	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
100	3	22.0000	.45826	.26458	20.8616	23.1384	21.50	22.40
200	3	19.6000	.88882	.51316	17.3921	21.8079	18.90	20.60
300	3	15.7333	.40415	.23333	14.7294	16.7373	15.30	16.10
400	3	13.1667	.41633	.24037	12.1324	14.2009	12.70	13.50
Total	12	17.6250	3.59498	1.03778	15.3409	19.9091	12.70	22.40

Output ini menggambarkan ringkasan statistic dari keempat sample. Sebagai contoh, pada perlakuan 100 mg. rata-rata kecepatan gerak sperma adalah 22,00 mm/mnt. Kecepatan minimum 21,5 mm/mnt dan maksimum 22,4 mm/mnt. Dengan tingkat kepercayaan 95% atau signifikansi 5%, rata-rata kecepatan ada pada range 20,8816 sampai 23,1384.

Untuk latihan, coba baca dan jelaskan ringkasan table statistic untuk 3 sample yang lain.

Test of Homogeneity of Variances

Levene Statistic	df1	df2	Sig.
1.645	3	8	.255

Analisis ini bertujuan untuk menguji berlaku tidaknya asumsi untuk ANOVA, yaitu apakah keempat sample mempunyai varians yang sama

Hipotesis :

H_0 = keempat varians populasi adalah identik

H_1 = keempat varians populasi adalah tidak identik

Pengambilan keputusan

- Jika probabilitas > 0,05, maka H_0 diterima
- Jika probabilitas < 0,05, maka H_0 ditolak

Keputusan :

Terlihat bahwa Levene T hitung adalah 1,645 dengan nilai probabilitas 0,255. Oleh karena probabilitas > 0,05, maka H_0 diterima atau berarti keempat varians adalah sama. Dengan demikian, asumsi kesamaan varians untuk uji ANOVA sudah terpenuhi.

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	139.489	3	46.496	139.141	.000
Within Groups	2.673	8	.334		
Total	142.163	11			

Uji ini digunakan untuk menguji apakah dari keempat sample mempunyai rata-rata yang sama. Analisis menggunakan ANOVA :

Hipotesis :

H_0 = keempat rata-rata populasi adalah identik

H_1 = keempat rata-rata populasi adalah tidak identik

Catatan: pada uji ini yang digunakan untuk asumsi adalah mean, bukan lagi varians.

Pengambilan Keputusan

Dapat didasarkan pada :

1. perbandingan F hitung dan F table
dasar pengambilan keputusan sama dengan uji F
2. nilai probabilitas → lebih praktis
jika probabilitas > 0,05, maka H_0 diterima
jika probabilitas < 0,05, maka H_0 ditolak
keputusan : melihat bahwa F hitung adalah 139,141 dengan probabilitas 0,000 < 0,05, maka H_0 ditolak → berarti rata-rata kecepatan gerak sperma dari keempat perlakuan dosis tersebut memang berbeda. Untuk mengetahui di antara keempat kelompok, mana saja kelompok yang berbeda dan mana saja yang tidak berbeda, hal ini akan dibahas pada analisis Tukey dalam post hoc test berikut.

Post Hoc Tets

Multiple Comparisons

Dependent Variable:

Tukey HSD

(I) dosis	(J) dosis	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
100	200	2.40000(*)	.47199	.004	.8885	3.9115
	300	6.26667(*)	.47199	.000	4.7552	7.7782
	400	8.83333(*)	.47199	.000	7.3218	10.3448
200	100	-2.40000(*)	.47199	.004	-3.9115	-.8885
	300	3.86667(*)	.47199	.000	2.3552	5.3782
	400	6.43333(*)	.47199	.000	4.9218	7.9448
300	100	-6.26667(*)	.47199	.000	-7.7782	-4.7552
	200	-3.86667(*)	.47199	.000	-5.3782	-2.3552
	400	2.56667(*)	.47199	.003	1.0552	4.0782
400	100	-8.83333(*)	.47199	.000	-10.3448	-7.3218
	200	-6.43333(*)	.47199	.000	-7.9448	-4.9218

300 | -2.56667(*) | .47199 | .003 | -4.0782 | -1.0552 |

* The mean difference is significant at the .05 level.

Uji signifikansi perbedaan, berdasarkan nilai probabilitas

- jika probabilitas > 0,05, maka H_0 diterima
- jika probabilitas < 0,05, maka H_0 ditolak

Post hoc di atas menunjukkan probabilitas kesemuanya < 0,05, maka H_0 ditolak, berarti perbedaan mean diantara keempat kelompok tersebut benar-benar nyata (hubungan antar variable). Hal tersebut juga dapat dilihat dengan adanya tanda (*) dibelakang angka Mean Difference.

Homogeneous Subsets

Tukey HSD

dosis	N	Subset for alpha = .05			
		1	2	3	4
400	3	13.1667			
300	3		15.7333		
200	3			19.6000	
100	3				22.0000
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

a Uses Harmonic Mean Sample Size = 3.000.

Homogeneous Subsets bertujuan untuk mencari grup/subset mana saja yang mempunyai perbedaan rata-rata yang tidak berbeda secara signifikan. Cara membaca yang paling mudah dari output ini adalah, jika nilai rata-rata terletak dalam satu kolom subsets yang sama, maka menunjukkan tidak ada perbedaan yang nyata. Output di atas menunjukkan bahwa keempat kelompok saling berbeda nyata satu dengan yang lain → ada 4 subsets yang masing-masing terisi satu nilai rata-rata.

Uji-uji komparatif di atas adalah beberapa contoh uji statistic untuk data-data yang memenuhi syarat untuk diuji parametric. Data-data yang tidak memenuhi syarat untuk diuji parametric, maka harus digunakan uji statistic alternative, yaitu yang nonparametric. Beberapa uji statistic non parametric akan dijelaskan pada saat praktikum, berikut dengan contoh-contohnya.

Panduan praktis interpretasi uji hipotesis

No.	Nama Uji	Makna jika $p < 0,05$ (Hipotesis nol ditolak)
1.	Uji normalitas Kolmogorov Smirnov, Shapiro-Wilk	Distribusi data tidak normal
2.	Uji Varians Levene's	Distribusi data yang dibandingkan mempunyai varian yang berbeda
3.	Uji One Way Anova	

4.	Uji Repeated ANOVA	Paling tidak, terdapat dua kelompok data yang mempunyai perbedaan yang bermakna (untuk mengetahui kelompok mana yang berbeda secara bermakna, harus dilakukan analisis Post Hoc
5.	Uji Kruskal Wallis	
6.	Uji Friedman	
7.	Uji Mc Nemar	Terdapat perbedaan yang bermakna antara dua pengukuran
8.	Uji Marginal Homogeneity	Terdapat perbedaan yang bermakna antara dua pengukuran
9.	Uji Cochran	Paling tidak, terdapat perbedaan pada dua pengukuran

Latihan/tugas :

Pada saat praktikum, mahasiswa diberi set data dari suatu kasus (contoh penelitian). Mahasiswa diberi tugas untuk melakukan praktek input data, menentukan uji statistik yang tepat sesuai prosedur yang telah diajarkan, menginterpretasikan hasil uji statistik tersebut untuk menarik kesimpulan yang tepat.

REGRESI dan KORELASI

Telah diketahui bahwa uji statistic pada dasarnya meliputi dua kegiatan, yakni uji beda dan uji asosiasi. Uji beda (difference) digunakan untuk mengetahui apakah ada perbedaan yang jelas antara rerata beberapa sample. Sedangkan uji asosiasi digunakan untuk mengetahui apakah di antara dua variable terdapat hubungan yang signifikan. Uji asosiasi meliputi korelasi dan regresi.

KORELASI

Pada bab ini akan ditinjau 2 aspek untuk analisis korelasi, yaitu apakah data sample yang ada menyediakan bukti cukup bahwa ada kaitan antara variable-variabel dalam populasi asal sample. Dan yang kedua adalah jika ada hubungan, seberapa kuat hubungan antar variable tersebut. Keeratan hubungan itu dinyatakan dengan nama koefisien korelasi (kadang disebut korelasi saja). Untuk uji korelasi yang dapat digunakan, sesuai dengan jenis variabelnya, dapat dilihat di dalam table berikut ini :

Tabel 3 : Pemilihan Hipotesis Korelatif

Variabel 1	Variabel 2	Uji Korelasi
Nominal	Nominal	Koefisien Kontengensi, Lambda
Nominal	Ordinal	Koefisien Kontengensi, Lambda
Ordinal	Ordinal	Spearman, Gamma, Somers'd
Ordinal	Numeric	Spearman
Numeric	Numeric	Pearson

Catatan : jika syarat untuk uji Pearson tidak memenuhi, maka digunakan uji alternative, yaitu uji korelasi Spearman (uji nonparametric)

Bagaimana Interpretasi hasil uji korelasi?

Interpretasi hasil uji korelasi didasarkan pada nilai p, kekuatan korelasi serta arah korelasinya.

Tabel 4 : Interpretasi Hasil Uji Korelasi

No.	Parameter	Nilai	Interpretasi
1.	Kekuatan Korelasi (r)	0,00-0,199	Sangat Lemah
		0,20-0,399	Lemah
		0,40-0,599	Sedang
		0,60-0,799	Kuat
		0,80-1,00	Sangat Kuat
2.	Nilai p (dalam SPSS, ditunjukkan dengan nilai Sig.)	$p < 0,05$	Terdapat korelasi yang bermakna antara dua variabel yang diuji
		$p > 0,05$	Tidak terdapat korelasi yang bermakna antara dua variabel yang diuji

3.	Arah korelasi	+ (positif) - (negatif)	Searah, semakin besar nilai satu variabel semakin besar pula nilai variabel lainnya Berlawanan arah, semakin besar nilai satu variabel, semakin kecil nilai variabel lainnya
----	---------------	--------------------------------	---

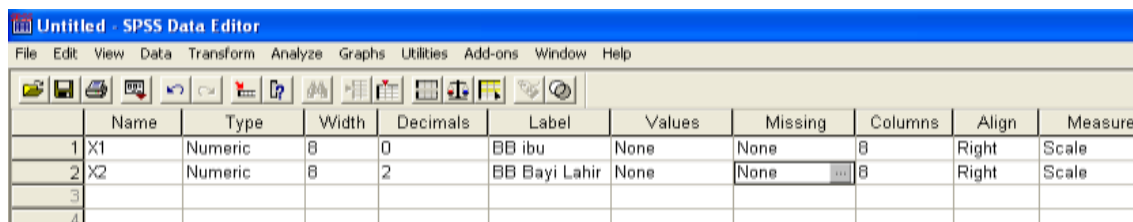
Contoh :

Untuk korelasi sederhana dengan data x1 dan x2 menggunakan angka adalah sebagai berikut :

Ingin diketahui apakah ada korelasi antara berat badan ibu dengan berat badan bayi yang dilahirkan di Sleman, ada 9 sampel yang digunakan. Berikut datanya :

No.	BB ibu (X1)	BB bayi lahir (X2)
1	80	3
2	86	3,10
3	87	3,14
4	90	3,30
5	78	2,60
6	70	2,50
7	65	2,51
8	60	1,80
9	62	1,90

Coba input data tersebut dengan mengisi variable view, kemudian lanjutkan dengan memasukkan data pada data view. Diumpamakan distribusi data adalah normal, maka uji dapat dilanjutkan dengan uji korelasi Pearson.



Gambar 4.1 : Variable View untuk contoh kasus Korelasi

Coba lakukan analisis statistic dengan langkah seperti berikut ini

- Klik menu **Analyze** → **Correlate** → **Bivariate**
- Masukkan variable BB ibu dan BB bayi lahir ke dalam kotak Variables
- Pilih (klik) Pearson dan Two-tailed
- Tekan OK



Gambar 4.2 : kotak dialog Bivariate Correlations

Output dan Analisis

Correlations

		BB ibu	BB bayi lahir
BB ibu	Pearson	1	.953(**)
	Correlation		
	Sig. (2-tailed)		
	N	9	9
BB bayi lahir	Pearson	.953(**)	1
	Correlation		
	Sig. (2-tailed)		
	N	9	9

** Correlation is significant at the 0.01 level (2-tailed).

Masalah : Apakah terdapat hubungan antara berat badan ibu dengan berat badan bayi lahir di Sleman?

Hipotesis :

H_0 : tidak terdapat hubungan antara BB Ibu dengan BB bayi lahir di Sleman

H_1 : terdapat hubungan antara BB Ibu dengan BB bayi lahir di Sleman

Keputusan :

Jika Sig > 0,05 maka H_0 diterima

Jika Sig < 0,05 maka H_0 ditolak

Pengambilan Keputusan :

Nilai sig = 0,000, maka H_0 ditolak berarti terdapat hubungan antara BB ibu dengan BB bayi lahir. Koefisien korelasi = 0,953, berarti hubungannya sangat kuat sekali.

REGRESI

Regresi bertujuan untuk menguji pengaruh antara satu variable dengan variable lain. Variabel yang dipengaruhi disebut variable tergantung atau dependen, sedangkan variable yang mempengaruhi disebut variable bebas atau variable independent. Uji regresi ada 2, yaitu :

1. Regresi linier sederhana (= Simple Regression)

- yaitu jika hanya ada satu variable dependent dan satu variable independent
2. Regresi linier berganda (Multiple Regression)
yaitu jika ada satu variable dependen dan lebih dari satu variable independent

REGRESI LINIER SEDERHANA

Model persamaan regresi linier sederhana adalah sebagai berikut :

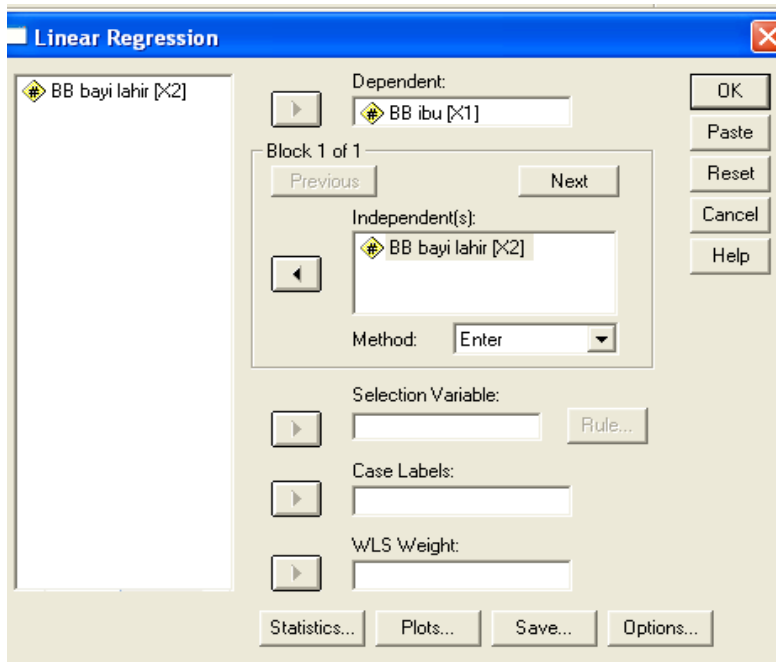
$$Y = a + b_1X_1 + e$$

Contoh soal :

Ingin diketahui apakah ada pengaruh antara BB ibu dengan BB bayi yang dilahirkan di Sleman, dengan sample 9. Dengan menggunakan data untuk korelasi pada bab sebelum ini, dilakukan analisis regresi dengan langkah-langkah berikut :

1. Pengolahan Data

- Buka lembar kerja (file Korelasi)
- Klik **Analyze** → **Regression** → **Linear**
- Masukkan BB ibu ke kotak Dependent dan BB bayi pada kotak Independent sehingga tampak kotak dialog seperti di bawah ini :



Gambar 4.3 : kotak dialog Linear Regression

- Langkah selanjutnya, untuk keseragaman, biarkan sesuai default yang ada.

2. Output dan analisis

Regression

Variables Entered/Removed(b)

Model	Variables Entered	Variables Removed	Method

1	BB bayi lahir(a)	.	Enter
---	------------------	---	-------

a All requested variables entered.
b Dependent Variable: BB ibu

Model Summary(b)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.953(a)	.908	.895	3.695

a Predictors: (Constant), BB bayi lahir
b Dependent Variable: BB ibu

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	946.422	1	946.422	69.315	.000(a)
	Residual	95.578	7	13.654		
	Total	1042.000	8			

a Predictors: (Constant), BB bayi lahir
b Dependent Variable: BB ibu

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients		Sig.
		B	Std. Error	Beta	t	
1	(Constant)	21.778	6.550		3.325	.013
	BB bayi lahir	20.210	2.427	.953	8.326	.000

a Dependent Variable: BB ibu

Perumusan masalah :

Apakah terdapat pengaruh antara BB ibu dengan BB bayi lahir ?

Hipotesis :

H₀ : tidak ada pengaruh antara BB ibu dengan BB bayi lahir

H_a : ada pengaruh antara BB ibu dengan BB bayi lahir

Pengambilan Keputusan

Y = BB ibu

X = BB bayi lahir

Cara 1 :

Jika Sig > 0,05 maka H₀ diterima

Jika Sig < 0,05 maka H₀ ditolak

Cara 2

Jika $t_{table} < t_{hitung}$, t_{table} maka H_0 diterima

Jika $t_{hitung} > t_{table}$ dan $-t_{hitung} > t_{table}$ maka H_0 ditolak

Cara 1 dari penelitian di atas, diperoleh $Sig = 0,000 < 0,05$ maka H_0 ditolak

Cara 2 untuk t_{table} kita melihat di table ($df = n - 1$; dua sisi) = 2,3060

$t_{hitung} = 8,326 \rightarrow$ berada pada daerah H_0 ditolak.

Setelah diketahui bahwa ada pengaruh antara BB ibu dengan BB bayi, besar pengaruhnya adalah 20,210 jadi persamaan regresinya adalah :

$$Y = 21,778 + 20,210X + e$$

Jika BB bayi lahir naik 1 satuan, maka BB ibu akan meningkat 20,210. Nilai R square adalah 0,908 (adalah pengkuadratan dari koefisien korelasi $(0,953)^2$). R square dapat disebut koefisien determinasi yang dalam hal ini berarti 90,8% BB ibu dipengaruhi oleh BB bayi lahir.

UJI CHI-SQUARE

Uji Chi Square untuk satu sampel

Uji ini dapat dipakai untuk menguji apakah data sebuah sample yang diambil menunjang hipotesis yang menyatakan bahwa populasi asal sample tersebut mengikuti suatu distribusi yang telah ditetapkan. Uji ini juga dapat disebut uji keselarasan (goodness of fit test) karena untuk menguji apakah sebuah sample selaras dengan salah satu distribusi teoritis (seperti distribusi normal, uniform, binomial dan lainnya) Namun pada penerapannya, uji ini tetap mengikuti prinsip dasar pengujian Chi-Square, yaitu membandingkan antara frekuensi-frekuensi harapan dengan frekuensi-frekuensi teramati.

Contoh kasus :

Manager PT Vita Indo yang menjual produk vitamin C dosis tinggi dengan empat macam rasa ingin mengetahui apakah konsumen menyukai keempat macam rasa vitamin tersebut. Untuk keperluan ini, dia dalam waktu satu minggu diamati pembelian permen di suatu outlet dan diperoleh hasil sebagai berikut :

Rasa	Jumlah
Orange	35
Lemon	28
Grape	10
Mango	27

Penyelesaian :

Oleh karena akan menguji apakah sebuah sample mengikuti distribusi tertentu, maka digunakan uji Chi-Square

Langkah-langkah :

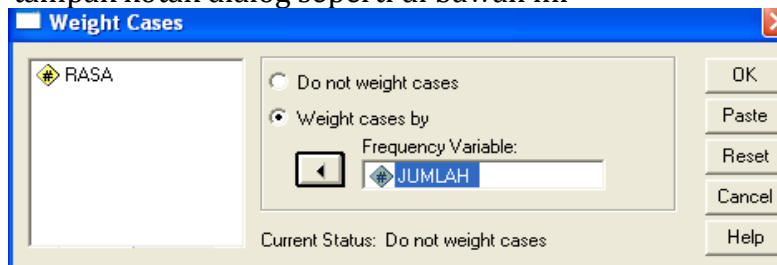
1. buka lembar kerja baru

2. definisikan Variabel dan property yang diperlukan, dengan cara seperti biasanya dalam pendefinisian data dalam SPSS, sehingga tampak di layar seperti di bawah ini. Lakukan kodifikasi untuk variable RASA

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	RASA	Numeric	8	0	{1, Orange}...	None	None	8	Right	Scale
2	JUMLAH	Numeric	8	0	None	None	None	8	Right	Scale

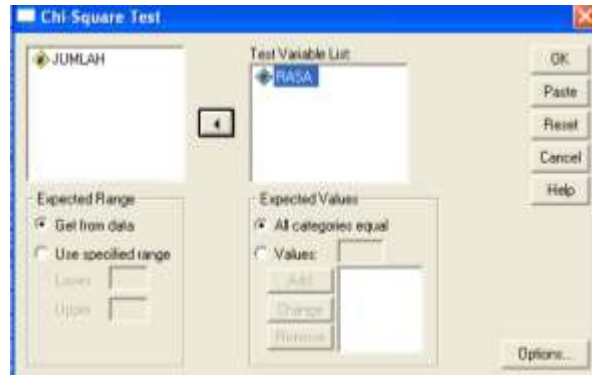
Gambar 4.4 : Variable View untuk contoh kasus Chi-Square Test

3. Mengisi data : sebelum mengisi, klik menu VIEW → VALUE LABEL
Lanjutkan dengan mengisi data sesuai kolomnya
4. Melakukan Process Weight Cases
Variabel RASA yang telah dikodifikasi, kemudian dilakukan proses Weight Cases untuk menghubungkan dengan variable jumlah. Adapun caranya adalah sbb :
 - Letakkan pointer pada kolom variable RASA
 - Dari menu utama SPSS, pilih menu Data → Weight Cases
 - Karena variable yang akan dihubungkan adalah JUMLAH, maka pindahkan variable JUMLAH ke pilihan Frequency Variable, sehingga tampak kotak dialog seperti di bawah ini



Gambar 4.5 : Kotak dialog Weight Cases

5. Pengolahan data ;
 - Masih di file Chi-Square, lanjutkan dengan memilih Menu **Analyze → Nonparametrik Test → Chi-Square...**,
 - Variabel yang akan diuji yaitu RASA dipindahkan ke kotak **Test Variable List**
 - Kolom **Expected Range** pilih Get from data, karena akan dihitung dari data kasus
 - Kolom Expected Values, pilih **All categories equal**, karena distribusinya adalah uniform (semua kemungkinan adalah sama), seperti pada kotak dialog di bawah ini.



Gambar 4.6 : kotak dialog Chi-Square test

- Klik OK

Output dan Analisis

NPar Tests

Chi-Square Test

Frequencies

RASA

	Observed N	Expected N	Residual
Orange	35	25.0	10.0
Lemon	28	25.0	3.0
Grape	10	25.0	-15.0
Mango	27	25.0	2.0
Total	100		

Expected N, artinya jumlah Rasa vitamin yang diharapkan terbeli. Oleh karena dipakai distribusi yang seragam, maka diharapkan sama rata, yaitu 25% (100% dibagi 4 rasa), kolom Residual adalah selisih antara jumlah yang dibeli dengan jumlah yang diharapkan.

Hipotesis

H₀ : sample diambil dari populasi yang mengikuti distribusi seragam, atau rasa vitamin yang ada disukai konsumen secara merata.

H₁ : sample bukan berasal dari populasi yang mengikuti distribusi seragam, atau setidaknya sebuah rasa vitamin lebih disukai daripada setidaknya sebuah rasa lain.

Test Statistics

	RASA
Chi-Square(a)	13.520
df	3
Asymp. Sig.	.004

a 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 25.0.

Pengambilan Keputusan :

- a. Berdasarkan perbandingan Chi-Square Uji dan Tabel
Jika Chi-Square hitung < Chi-Square table, maka H_0 diterima
Jika Chi-Square hitung > Chi-Square table, maka H_0 ditolak
 - Chi-Square hitung → lihat pada output, adalah = 13,520
 - Chi-Square table → lihat table Chi-Square dengan $\alpha = 5\%$, dan $df = 3$ (lihat output atau dari rumus $k - 1$) → Chi-Square table untuk kasus ini adalah 7,814
 - Jadi Chi-Square hitung > Chi Square table ($13,520 > 7,814$) → maka H_0 ditolak
- b. Berdasarkan Probabilitas
 - Jika probabilitas > 0,05, maka H_0 diterima
 - Jika probabilitas < 0,05, maka H_0 ditolak

Keputusan :

- Dapat dibaca pada kolom **Asymp.Sig/Asiymptotic significance** adalah 0,04, atau probabilitas di bawah 0,05 → H_0 ditolak

Berdasar 2 analisis diatas, dapat diambil kesimpulan yang sama, yaitu H_0 ditolak, artinya populasi tidak seragam (konsumen tidak mempunyai kesukaan yang sama terhadap keempat macam rasa vitamin tersebut.

Uji Chi-Square (Hipotesis Komparatif Karegorik tidak berpasangan)

Contoh kasus (sekaligus untuk **latihan/tugas**)

Jika ingin mengetahui hubungan antara perilaku merokok merokok dan tidak merokok) dengan status fertilitas pria (infertil dan fertil). Dirumuskan pertanyaan sebagai berikut : " Apakah terdapat hubungan antara perilaku merokok dengan status fertilitas seorang pria?

Uji apakah yang mungkin digunakan untuk menjawab pertanyaan tersebut?

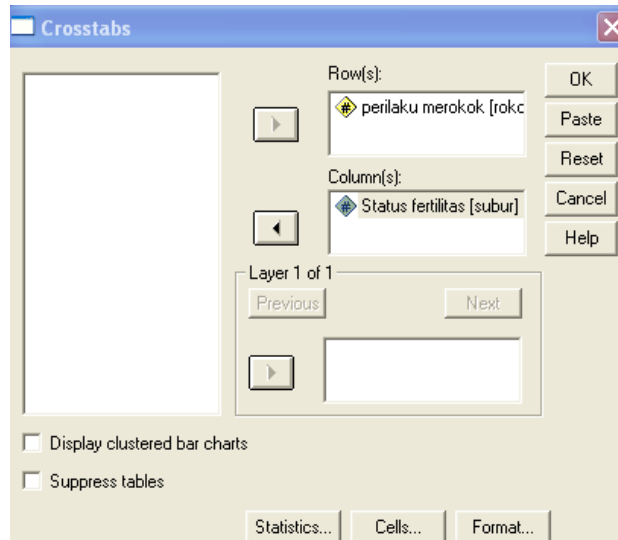
Langkah-langkah yang digunakan untuk menjawab pertanyaan tersebut adalah sebagai berikut :

1. Tentukan variabel yang akan dihubungkan
2. Tentukan jenis hipotesis
3. Tentukan masalah skala variabel
4. Tentukan berpasangan atau tidak
5. Tentukan jenis tabel B x K (Baris x Kolom)
6. Menarik kesimpulan uji statistik yang tepat

Coba praktekkkan langkah-langkah tersebut, buatlah jawabannya (alasan), mengapa digunakan uji Chi Square? Uji alternatif apakah yang digunakan jika syarat-syarat untuk uji ini tidak terpenuhi?

Langkah-langkah pengujian:

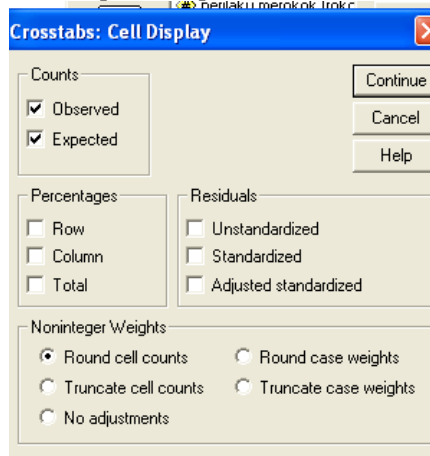
1. Lakukan input data (data diberikan pada saat praktikum) seperti yang sudah kalian lakukan pada materi sebelumnya.
Lakukan langkah berikut (klik):
2. Analyze → Descriptive statistics → Croostabs
3. masukkan variabel rokok ke dalam Rows (sebagai variabel bebas)
4. masukkan variabel subur ke dalam Columns (sebagai variabel terikat)



5. Klik kotak Statistics, lalu pilih Chi-Square pada kiri atas kotak, lalu klik Continue



6. Klik Cell, pilih Observed untuk menampilkan nilai observed dan pilih Expected untuk menampilkan nilai expected pada kotak Count, lalu klik Continue



7. Proses telah selesai, Klik Continue, Klik OK

Coba lihat outputnya!

perilaku merokok * Status fertilitas Crosstabulation

			Status fertilitas		Total
			subur	tidak subur	
perilaku merokok	tidak merokok	Count	35	15	50
		Expected Count	27.5	22.5	50.0
	merokok	Count	20	30	50
		Expected Count	27.5	22.5	50.0
Total		Count	55	45	100
		Expected Count	55.0	45.0	100.0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	9.091(b)	1	.003		
Continuity Correction(a)	7.919	1	.005		
Likelihood Ratio	9.240	1	.002		
Fisher's Exact Test				.005	.002
Linear-by-Linear Association	9.000	1	.003		
N of Valid Cases	100				

a Computed only for a 2x2 table

b 0 cells (.0%) have expected count less than 5. The minimum expected count is 22.50.

Interpretasi Hasil

1. Tabel atas mendeskripsikan masing-masing sel untuk nilai observed dan expected. Nilai observed untuk sel a, b, c & d masing-masing 35, 15, 20 dan 30; sedangkan expectednya masing-masing 27.5, 22.5, 27.5 dan 22.5.
2. Tabel 2x2 ini layak diuji dengan Chi Square karena tidak ada nilai expected yang kurang dari 5
3. Tabel di bawahnya menunjukkan hasil uji Chi Square. Nilai yang dipakai adalah pada nilai Pearson Chi-Square, dengan nilai Significancy nya adalah 0.003, berarti terdapat hubungan antara perilaku merokok dengan status fertilitas.

Interpretasi lengkap nilai p

Jika tidak ada hubungan antara perilaku merokok dengan status kesuburan, maka faktor peluang saja menerangkan 0.003 hasil yang diperoleh. Karena faktor peluang kurang dari 5%, maka hasil tersebut bermakna.

APPENDIX 2 :
PRACTICAL GUIDANCE

STATA

A SAMPLE SESSION

The dataset that we will use for this session is a set of data about vintage 1978 automobiles sold in the United States.

To follow along using point-and-click, note that the menu items are given by **Menu > Menu Item > Submenu Item > etc.** To follow along using the Command window, type the commands that follow a dot (.) in the boxed listing into the small window labeled **Command**. When there is something of note about the structure of a command, it will be pointed out as a **Syntax note**.

Start by loading the **auto** dataset, which is included with Stata. To use the menus,

1. Select **File > Example Datasets...**
2. Click on **Example datasets installed with Stata.**
3. Click on **use on auto.dta.**

The result of this command is threefold:

- Some output appears in the large Results window.
- **. sysuse auto**
- (1978 Automobile Data)

The output consists of a command and its result. The command is bold and follows the period (.): **sysuse auto**. The output is in the standard face here and is a brief description of the dataset.


Note: If a command intrigues you, you can type `help commandname` in the Command window to find help. If you want to explore at any time, **Help > Search...** can be informative.

- The same command, **sysuse auto** appears in the small Review window to the upper left. The Review window keeps track of commands Stata has run, successful and unsuccessful. The commands can then easily be rerun. See chapter 4 for more information.
- A series of variables appears in the small Variables window to the lower left.

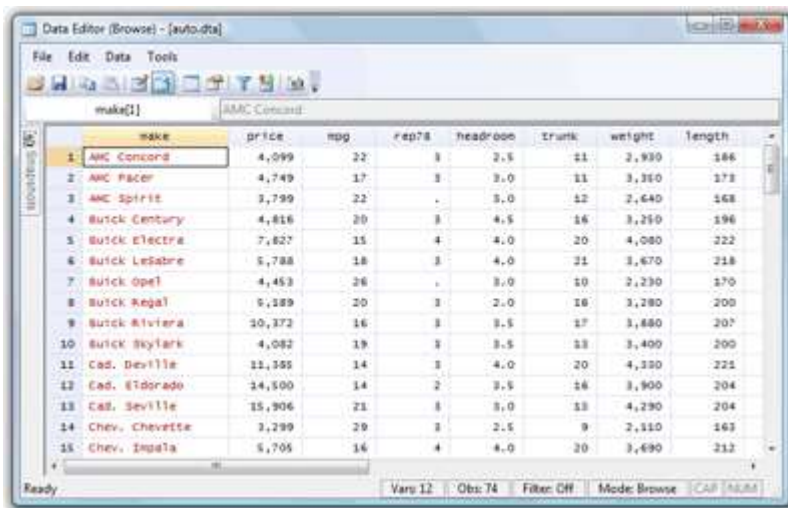
You could have opened the dataset by typing **sysuse auto** followed by *Enter* into the Command window. Try this now. **sysuse** is a command that loads (uses) example (system) datasets. As you will see during this session, Stata commands are often simple enough that it is faster to use them directly. This will be especially true once you become familiar with the commands you use the most in your daily use of Stata.

Syntax note: in the above example **sysuse** is the Stata command, whereas **auto** is the name of a Stata data file.

Simple data management

We can get a quick glimpse at the data by looking at it in the **Data Browser**. This can be done by clicking on the Data Editor (Browse) button, , or selecting **Data > Data Editor (Browse)** from the menus, or typing the command **browse**.

When the Data Browser window opens, you can see that Stata regards the data as one rectangular table. This is true for all Stata datasets. The columns represent *variables*, whereas the rows represent *observations*. The variables have somewhat descriptive names, whereas the observations are numbered.



The screenshot shows the Stata Data Editor (Browse) window for a dataset named 'auto.dta'. The window displays a table with 15 rows and 8 columns. The columns are labeled: make, price, mpg, rep78, headroom, trunk, weight, and length. The rows are numbered 1 through 15. The 'make' column contains car makes such as AMC Concord, AMC Pacer, AMC Spirit, Buick Century, Buick Electra, Buick LeSabre, Buick Opel, Buick Regal, Buick Riviera, Buick Skylark, Cad. Deville, Cad. Eldorado, Cad. Seville, Chev. Chevette, and Chev. Impala. The 'price' column contains numerical values ranging from 4,099 to 15,906. The 'mpg' column contains numerical values ranging from 14 to 29. The 'rep78' column contains numerical values ranging from 1 to 5, with some cells containing a period (.) representing missing values. The 'headroom', 'trunk', 'weight', and 'length' columns contain numerical values ranging from 2.5 to 222.

	make	price	mpg	rep78	headroom	trunk	weight	length
1	AMC Concord	4,099	22	1	2.5	11	2,930	186
2	AMC Pacer	4,749	17	1	3.0	11	3,360	173
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168
4	Buick Century	4,816	20	1	4.5	16	3,250	196
5	Buick Electra	7,827	15	4	4.0	20	4,080	222
6	Buick LeSabre	5,788	18	1	4.0	21	3,670	218
7	Buick Opel	4,453	26	.	3.0	10	2,230	170
8	Buick Regal	5,189	20	1	2.0	16	3,280	200
9	Buick Riviera	10,372	16	1	3.5	17	3,680	207
10	Buick Skylark	4,082	19	1	3.5	13	3,400	200
11	Cad. Deville	11,355	14	1	4.0	20	4,330	221
12	Cad. Eldorado	14,500	14	2	3.5	18	3,900	204
13	Cad. Seville	15,906	21	1	3.0	13	4,290	204
14	Chev. Chevette	3,299	29	1	2.5	9	2,110	163
15	Chev. Impala	5,705	16	4	4.0	20	3,690	212

The data are displayed in multiple colors—at first glance it appears that the variables listed in black are numeric, whereas those that are in colors are text. This is worth investigating. Click on a cell under the **make** variable: the input box at the top displays the make of the car. Scroll to the right until you see the **foreign** variable. Click on one of its cells. Although the cell may display “Domestic”, the input box displays a 0. This shows that Stata can store categorical data as numbers but display human-readable text. This is done by what Stata calls *value labels*. Finally, under the **rep78** variable, which looks to be numeric, there are some cells containing just a period (.). As we will see, these correspond to missing values.

Syntax note: here the command is **browse** and there are no other arguments.

Looking at the data in this fashion, though comfortable, lends little information about the dataset. It would be useful for us to get more details about what the data are and how the data are stored. Close the Data Browser by clicking its close button—while it is open, we cannot give any commands to Stata.

We can see the structure of the dataset by *describing* its contents. This can be done either by going to **Data > Describe data > Describe data in memory** in the menus

and clicking **OK** or by typing **describe** in the Command window and pressing *Enter*. Regardless of which method you choose, you will get the same result:

. describe

```
Contains data from C:\Stata11\ado\base\a\auto.dta
obs:      74                1978 Automobile Data
vars:     12                13 Apr 2009 17:45
size:     3,478 (99.9% of memory free) (_dta has notes)
-----
      storage display  value
variable name type format  label  variable label
-----
make      str18 %-18s          Make and Model
price     int  %8.0gc              Price
mpg       int  %8.0g             Mileage (mpg)
rep78     int  %8.0g             Repair Record 1978
headroom  float %6.1f              Headroom (in.)
trunk     int  %8.0g              Trunk space (cu. ft.)
weight    int  %8.0gc              Weight (lbs.)
length    int  %8.0g              Length (in.)
turn      int  %8.0g              Turn Circle (ft.)
displacement int %8.0g          Displacement (cu. in.)
gear_ratio float %6.2f          Gear Ratio
foreign   byte %8.0g  origin  Car type
-----
Sorted by: foreign
```

If your listing stops short, and you see a blue **--more--** at the base of the Results window, pressing the space bar or clicking on the blue **--more--** itself will allow the command to be completed.

At the top of the listing, some information is given about the dataset, such as where it is stored on disk, how much memory it occupies, and when the data were last saved. The bold **1978 Automobile Data** is the short description that appeared when the dataset was opened and is referred to as a *data label* by Stata. The phrase **_dta has notes** informs us that there are notes attached to the dataset. We can see what notes there are by typing **notes** in the Command window:

. notes

_dta:

1. from Consumer Reports with permission

Here we see a short note about the source of the data.

Looking back at the listing from **describe**, we can see that Stata keeps track of more than just the raw data. Each variable has the following:

1. A *variable name*, which is what you call the variable when communicating with Stata.
2. A *storage type*, which is the way in which Stata stores its data. For our purposes, it is enough to know that types beginning with **str** are string, or text, variables, whereas all others are numeric. See [U] **12 Data**.
3. A *display format*, which controls how Stata displays the data in tables. See [U] **12.5 Formats: controlling how data are displayed**.
4. A *value label* (possibly). This is the mechanism that allows Stata to store numerical data while displaying text. See chapter 9 and [U] **12.6.3 Value labels**.
5. A *variable label*, which is what you call the variable when communicating with other people. Stata uses the variable label when making tables, as we will see.

A dataset is far more than simply the data it contains. It is also information that makes the data usable by someone other than the original creator.

Although describing the data says something about the structure of the data, it says little about the data themselves. The data can be summarized by clicking **Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Summary statistics**, and clicking the **OK** button. You could also type **summarize** in the Command window and press *Enter*. The result is a table containing summary statistics about all the variables in the dataset:

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk	74	13.75676	4.277404	5	23
weight	74	3019.459	777.1936	1760	4840
length	74	187.9324	22.26634	142	233
turn	74	39.64865	4.399354	31	51
displacement	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1

From this simple summary, we can learn a bit about the data. First of all, the prices are nothing like today's car prices—of course, these cars are now antiques. We can see that the gas mileages are not particularly good. Automobile aficionados can gain some feel for other characteristics.

There are two other important items here:

The variable **make** is listed as having no observations. It really has no numerical observations, because it is a string (text) variable.

The variable **rep78** has 5 fewer observations than the other numerical variables. This implies that **rep78** has five missing values.

Although we could use the **summarize** and **describe** commands to get a bird's eye view of the dataset, Stata has a command that gives a good in-depth description of the structure, contents, and values of the variables: the **codebook** command. Either type **codebook** in the Command window and press *Enter* or navigate the menus to **Data > Describe data > Describe data contents (codebook)**, and click **OK**. We get a large amount of output that is worth investigating. Look it over to see that much can be learned from this simple command. You can scroll back in the Results window to see earlier results, if need be. We'll focus on the output for **make**, **rep78**, and **foreign**.

To start our investigation, we would like to run the **codebook** command on just one variable, say, **make**. We can do this via menus or the command line, as usual. To get the **codebook** output for **make** via the menus, start by navigating as before, to **Data > Describe data > Describe data contents (codebook)**. When the dialog appears, there are multiple ways to tell Stata to consider only the **make** variable:

- We could type **make** into the *Variables* field.
- The *Variables* field is actually a combobox control that accepts variable names. Clicking the button to the right of the *Variables* field displays a list of the variables from the current dataset. Selecting a variable from the list will, in this case, enter the variable name into the edit field.

A much easier solution is to type **codebook make** in the Command window and then press *Enter*. The result is informative:

```
. codebook make
-----
make                               Make and Model
type: string (str18), but longest is str17

unique values: 74                   missing "": 0/74

examples: "Cad. Deville"
           "Dodge Magnum"
           "Merc. XR-7"
           "Pont. Catalina"

warning: variable has embedded blanks
```

The first line of the output tells us the variable name (**make**) and the variable label (**Make and Model**). The variable is stored as a string (which is another way of saying “text”) with a maximum length of 18 characters, though a size of only 17 characters

would be enough. All the values are unique, so if need be, **make** could be used as an identifier for the observations—something that is often useful when putting together data from multiple sources or when trying to weed out errors from the dataset. There are no missing values, but there are blanks within the makes. This latter fact could be useful if we were expecting **make** to be a single-word string variable.

Syntax note: telling the **codebook** command to run on the **make** variable is an example of using a *varlist* in Stata's syntax.

Looking at the **foreign** variable can teach us about value labels. We would like to look at the codebook output for this variable, and on the basis of our latest experience, it would be easy to type **codebook foreign** into the Command window to get the following output:

```
. codebook foreign
-----
foreign                               Car type
      type: numeric (byte)
      label: origin

      range: [0,1]                    units: 1
unique values: 2                      missing.: 0/74

      tabulation: Freq. Numeric Label
                   52    0 Domestic
                   22    1 Foreign
```

We can glean that **foreign** is an indicator variable, because its only values are 0 and 1. The variable has a value label that displays “Domestic” instead of 0 and “Foreign” instead of 1. There are two advantages of storing the data in this form:

- Storing the variable as a byte takes less memory, because each observation uses 1 byte instead of the 8 bytes needed to store “Domestic”. This is important in large datasets. See [U] **12.2.2 Numeric storage types**.
- As an indicator variable, it is easy to incorporate into statistical models. See [U] **25.2 Using indicator variables in estimation**.

Finally, we can learn a little about a poorly labeled variable with missing values by looking at the **rep78** variable. Typing **codebook rep78** into the Command window yields

```
. codebook rep78
```

rep78

Repair Record 1978

type: numeric (int)

range: [1,5]

units: 1

unique values: 5

missing.: 5/74

tabulation: Freq. Value

2 1

8 2

30 3

18 4

11 5

5 .

rep78 appears to be a categorical variable, but because of lack of documentation, we do not know what the numbers mean. (To see how we would label the values, see chap. 10.) This variable has five missing values, meaning that there are 5 observations for which the repair record is not recorded. We can use the Data Browser to investigate these 5 observations—and we will do this using the Command window only. (Doing so is much simpler.) If you recall from earlier, the command brought up by clicking the Data Browser button was **browse**. We would like to browse only those observations for which **rep78** is missing, so we could type

. browse if missing(rep78)

Obs	make	price	mpg	rep78	headroom	trunk	weight	length
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168
7	Buick Opel	4,453	26	.	3.0	10	2,230	170
45	Plym. Sapporo	6,486	26	.	1.5	8	2,520	182
51	Pont. Phoenix	4,424	19	.	3.5	13	3,420	203
64	Peugeot 604	22,990	14	.	3.5	14	3,420	192

From this, we see that the **.** entries are indeed missing values—though other missing values are allowable. See [U] **12.2.1 Missing values**. Close the Data Browser after you are satisfied with this statement.

Syntax note: using the *if qualifier* above is what allowed us to look at a subset of the observations.

Looking through the data lends no clues about why these particular data are missing. We decide to check the source of the data to see if the missing values were originally missing or if they were omitted in error. Listing the makes of the cars whose repair records are missing will be all we need, since we saw earlier that the values of **make** are unique. This can be done via the menus and a dialog.

1. Select **Data > Describe data > List data**.
2. Click the button to the right of the *Variables* field to show the variable names.
3. Click **make** to enter it into the *Variables* field.
4. Click the **by/if/in** tab in the dialog.
5. Type **missing(rep78)** into the *If: (expression)* box.
6. Click **Submit**. Stata executes the proper command but the dialog box remains open. **Submit** is useful when experimenting, exploring, or building complex commands. We will use primarily **Submit** in the examples. You may click **OK** in its place if you like.

The same ends could be achieved by typing **list make if missing(rep78)**. The latter is easier, once you know that the command **list** is used for listing observations. In any case, here is the output:

. list make if missing(rep78)

	make
3.	AMC Spirit
7.	Buick Opel
45.	Plym. Sapporo
51.	Pont. Phoenix
64.	Peugeot 604

We go to the original reference and find that the data were truly missing and cannot be resurrected. See chapter 11 for more information about all that can be done with the **list** command.

Syntax note: this command uses two new concepts for Stata commands—the **if** qualifier and the **missing()** function. The **if** qualifier restricts the observations on which the command runs to only those observations for which the expression is true. See [U] **11.1.3 if exp**. The **missing()** function tests each observation to see if it is missing. See [U] **13.3 Functions**.

Now that we have a good idea about the underlying dataset, we can investigate the data themselves.

Descriptive statistics

We saw above that the **summarize** command gave brief summary statistics about all the variables. Suppose now that we became interested in the prices while summarizing the data, because they seemed fantastically low (it was 1978, after all). To get an in-depth look at the **price** variable, we can use the menus and a dialog:

1. Select **Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Summary statistics**.
2. Enter or select **price** in the *Variables* field.

3. Select **Display additional statistics**.
4. Click **Submit**.

Syntax note: as can be seen from the Results window, typing **summarize price, detail** will get the same result. The portion after the comma contains *options* for Stata commands; hence, **detail** is an example of an option.

. summarize price, detail

Price					
Percentiles		Smallest			
1%	3291	3291			
5%	3748	3299			
10%	3895	3667	Obs	74	
25%	4195	3748	Sum of Wgt.	74	
50%	5006.5		Mean	6165.257	
		Largest	Std. Dev.	2949.496	
75%	6342	13466			
90%	11385	13594	Variance	8699526	
95%	13466	14500	Skewness	1.653434	
99%	15906	15906	Kurtosis	4.819188	

From the output, we can see that the median price of the cars in the dataset is only about \$5,006! We could also see that the four most expensive cars are all priced between \$13,400 and \$16,000. If we wished to browse the cars that are the most expensive (and gain some experience with Stata's command syntax), we could use an **if** qualifier,

. browse if price > 13000

and see from the Data Browser that the expensive cars are two Cadillacs and two Lincolns, which have low gas mileage and are fairly heavy:



We now decide to turn our attention to foreign cars and repairs, because as we glanced through the data, it appeared that the foreign cars had better repair records. Let's start by looking at the proportion of foreign cars in the dataset along with the

proportion of cars with each type of repair record. We can do this with one-way tables. The table for **foreign** cars can be done via menus and a dialog starting with **Statistics > Summaries, tables, and tests > Tables > One-way tables** and then choosing the variable **foreign** in the *Categorical variable* field. Clicking **Submit** yields

```
. tabulate foreign
```

Car type	Freq.	Percent	Cum.
Domestic	52	70.27	70.27
Foreign	22	29.73	100.00
Total	74	100.00	

We see that roughly 70% of the cars in the dataset are domestic, whereas 30% are foreign made. The value labels are used to make the table, so that the output is nicely readable.

Syntax note: we also see that this one-way table could be made by using the **tabulate** command together with a one variable: **foreign**.

Making a one-way table for the repair records is simple—it will be simpler if done via the Command window. Typing **tabulate rep78** yields

```
. tabulate rep78
```

Repair Record 1978	Freq.	Percent	Cum.
1	2	2.90	2.90
2	8	11.59	14.49
3	30	43.48	57.97
4	18	26.09	84.06
5	11	15.94	100.00
Total	69	100.00	

We can see that most cars have repair records of 3 and above, though the lack of value labels makes us unsure what a “3” means. The five missing values are indirectly evident, because the total number of observations listed is 69 rather than 74.

These two one-way tables do not help us compare the repair records of foreign and domestic cars. A two-way table would help greatly, which we can get by using the menus and a dialog:

1. Select **Statistics > Summaries, tables, and tests > Tables > Two-way tables with measures of association**.
2. Choose **rep78** as the row variable.
3. Choose **foreign** as the column variable.
4. It would be nice to have the percentages within the **foreign** variable, so check the **Within-row relative frequencies** checkbox.

5. Click **Submit**.

Here is the resulting output.

. tabulate rep78 foreign, row

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

Repair Record 1978	Car type		Total
	Domestic	Foreign	
1	2 100.00	0 0.00	2 100.00
2	8 100.00	0 0.00	8 100.00
3	27 90.00	3 10.00	30 100.00
4	9 50.00	9 50.00	18 100.00
5	2 18.18	9 81.82	11 100.00
Total	48 69.57	21 30.43	69 100.00

The output indicates that foreign cars are generally much better than domestic cars when it comes to repairs. If you like, you could repeat the previous dialog and try some of the hypothesis tests available from the dialog. We will abstain.

Syntax note: we see that typing the command **tabulate rep78 foreign, row** would have given us the same table. Thus, using **tabulate** with two variables yielded a two-way table. It makes sense that **row** is an option, because we went out of our way to check it in the dialog, so we changed the behavior of the command from its default.

Continuing our exploratory tour of the data, we would like to compare gas mileages between foreign and domestic cars, starting by looking at the summary statistics for each group by itself. We know that a direct way to do this would be to summarize **mpg** for each of the two values of **foreign**, by using **if** qualifiers:

. summarize mpg if foreign==0

Variable	Obs	Mean	Std. Dev.	Min	Max
mpg	52	19.82692	4.743297	12	34

. summarize mpg if foreign==1

Variable	Obs	Mean	Std. Dev.	Min	Max
mpg	22	24.77273	6.611187	14	41

It appears that foreign cars get somewhat better gas mileage—we will test this soon.

Syntax note: we needed to use a double equal sign (==) for testing equality. This could be familiar to you if you have programmed before. If it is unfamiliar, it is a common source of errors when initially using Stata. Thinking of equality as “really equal” can cut down on typing errors.

There are two other methods that we could have used to produce these summary statistics. These methods are worth knowing, because they are less error prone. The first method duplicates the concept of what we just did by exploiting Stata's ability run a command on each of a series of nonoverlapping subsets of the dataset. To use the menus and a dialog, do the following:

1. Select **Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Summary statistics** and click the **Reset** button.
2. Select **mpg** in the *Variables* field.
3. Select *Standard display* option (if it is not already selected).
4. Select the **by/if/in** tab.
5. Check the *Repeat command by groups* checkbox.
6. Select or type **foreign** for the grouping variable.
7. **Submit** the command.

You can see that the results match those from above. They have a better appearance because the value labels are used rather than the numerical values. The method is more appealing because the results were produced without knowing the possible values of the grouping variable ahead of time.

```
. by foreign, sort: summarize mpg
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mpg	52	19.82692	4.743297	12	34

```
-> foreign = Foreign
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mpg	22	24.77273	6.611187	14	41

There is something different about the equivalent command that appears above: it contains a *prefix command* called a **by** prefix. The **by** prefix has its own option, namely, **sort**, to ensure that like members are adjacent to each other before being summarized. The **by** prefix command is important for understanding data manipulation and working with subpopulations within Stata. Store this example away, and consult [U] **11.1.2 by varlist:** and [U] **27.2 The by construct** for more information. Stata has other prefix commands for specialized treatment of commands, as explained in [U] **11.1.10 Prefix commands**.

The third method for tabulating the differences in gas mileage across the cars' origins involves thinking about the structure of desired output. We need a one-way table of automobile types (foreign versus domestic) within which we see information about gas mileages. Looking through the menus yields the menu item **Statistics > Summaries, tables, and tests > Tables > One/two-way table of summary statistics**. Selecting this, entering **foreign** for *Variable 1* and **mpg** for the *Summarize variable*, and submitting the command yields a nice table:

```
. tabulate foreign, summarize(mpg)
```

Car type	Summary of Mileage (mpg)		
	Mean	Std. Dev.	Freq.
Domestic	19.826923	4.7432972	52
Foreign	24.772727	6.6111869	22
Total	21.297297	5.7855032	74

The equivalent command is evidently **tabulate foreign, summarize(mpg)**.

Syntax note: this is a one-way table, so **tabulate** uses one variable. The variable being summarized is passed to the **tabulate** command via an option.

A simple hypothesis test

We would like to run a hypothesis test for the difference in the mean gas mileages. Under the menus, **Statistics > Summaries, tables, and tests > Classical tests of hypotheses > Two-group mean-comparison test** leads to the proper dialog. Enter **mpg** for the *Variable name* and **foreign** for the *Group variable name*, and **Submit** the dialog. The results are

```
. ttest mpg, by(foreign)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
Domestic	52	19.82692	.657777	4.743297	18.50638	21.14747
Foreign	22	24.77273	1.40951	6.611187	21.84149	27.70396
combined	74	21.2973	.6725511	5.785503	19.9569	22.63769
diff		-4.945804	1.362162		-7.661225	-2.230384

diff = mean(Domestic) - mean(Foreign) t = -3.6308
Ho: diff = 0 degrees of freedom = 72

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0003 Pr(|T| > |t|) = 0.0005 Pr(T > t) = 0.9997

From this, we could conclude that the mean gas mileage for foreign cars is different from that of domestic cars (though we really ought to have wanted to test this before snooping through the data). We can also conclude that the command, **ttest mpg**,

by(foreign) is easy enough to remember. Feel free to experiment with unequal variances, various approximations to the number of degrees of freedom, and the like.

Syntax note: the **by()** option used here is not the same as the **by** prefix command used earlier. Although it has a similar conceptual meaning, its usage is different because it is a particular option for the **ttest** command.

Descriptive statistics, correlation matrices

We now change our focus from exploring categorical relationships to exploring numerical relationships: we would like to know if there is a correlation between miles per gallon and weight. We select **Statistics > Summaries, tables, and tests > Summary and descriptive statistics > Correlations and covariances** in the menus. Entering **mpg** and **weight**, either by clicking or typing, and then submitting the command yields

```
. correlate mpg weight
(obs=74)
      |      mpg  weight
-----|-----
mpg   |  1.0000
weight| -0.8072  1.0000
```

The equivalent command for this is natural: **correlate mpg weight**. There is a negative correlation, which is not surprising, because heavier cars should be harder to push about.

We could see how the correlation compares for foreign and domestic cars by using our knowledge of how the **by** prefix works. We can reuse the *correlate* dialog or use the menus as before if the dialog is closed. Click the **by/if/in** tab, check the *Repeat command by groups* checkbox, and enter the **foreign** variable to define the groups. As done above on page 19, a simple **by foreign, sort:** prefix in front of our previous command would work, too:

```
. by foreign, sort: correlate mpg weight
-> foreign = Domestic
(obs=52)
      |      mpg  weight
-----|-----
mpg   |  1.0000
weight| -0.8759  1.0000
-----|-----
-> foreign = Foreign
(obs=22)
      |      mpg  weight
-----|-----
mpg   |  1.0000
weight| -0.6829  1.0000
```

We see from this that the correlation is not as strong among the foreign cars.

Syntax note: although we used the **correlate** command to look at the correlation of two variables, Stata can make correlation matrices for an arbitrary number of variables:

```
. correlate mpg weight length turn displacement
(obs=74)
```

	mpg	weight	length	turn	displa~t
mpg	1.0000				
weight	-0.8072	1.0000			
length	-0.7958	0.9460	1.0000		
turn	-0.7192	0.8574	0.8643	1.0000	
displacement	-0.7056	0.8949	0.8351	0.7768	1.0000

This can be useful, for example, when investigating collinearity among predictor variables. In fact, simply typing **correlate** will yield the complete correlation matrix.

Graphing data

We have found several things in our investigations so far: We know that the average MPG of domestic and foreign cars differs. We have learned that domestic and foreign cars differ in other ways as well, such as in frequency-of-repair record. We found a negative correlation of MPG and weight—as we would expect—but the correlation appears stronger for domestic cars.

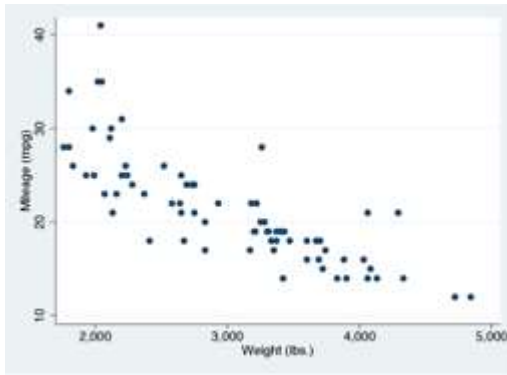
We would now like to examine, with an eye toward modeling, the relationship between MPG and weight, starting with a graph. We can start with a scatterplot of **mpg** against **weight**. The command for this is simple: **scatter mpg weight**. Using the menus requires a few steps because the graphs in Stata can be customized heavily.

1. Select **Graphics > Twoway graph (scatter, line, etc.)**.
2. Click the **Create...** button.
3. Select the *Basic plots* radio button (if it is not already selected).
4. Select *Scatter* as the basic plot type (if it is not already selected).
5. Select **mpg** as the *Y variable* and **weight** as the *X variable*.
6. Click the **Submit** button.

The Results window shows the command that was issued from the menu:

```
. twoway (scatter mpg weight)
```

The command issued when the dialog was submitted is a bit more complex than the command suggested above. There is good reason for this: the more complex structure allows combining and overlaying graphs, as we will soon see. In any case, the graph that appears is



We see the negative correlation in the graph, though the relationship appears to be nonlinear.

Note: when you draw a graph, the Graph window appears, probably covering up your Results window. Click on the main Stata window to get the Results window back on top. Want to see the graph again? Click the **Graph** button. See chapter 14 for more information about the **Graph** button.

We would now like to see how the different correlations for foreign and domestic cars are manifested in scatterplots. It would be nice to see a scatterplot for each type of car, along with a scatterplot for all the data.

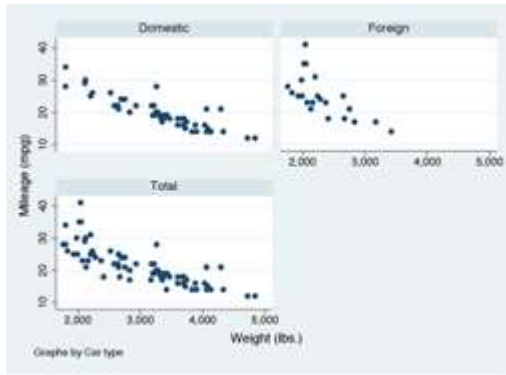
Syntax note: since we are looking at subgroups, this looks like it is a job for **by**. We want one graph—think about whether it should be a **by()** option or a **by** prefix.

Start as before:

1. Select **Graphics > Twoway graph (scatter, line, etc.)** from the menus.
2. If *Plot 1* is displayed under the *Plot definitions*, and **(scatter mpg weight)** appears below the *Plot definitions* box, select it and skip to step 4.
3. Go through the process to create the graph on the previous page.
4. Click the **By** tab.
5. Check the *Draw subgraphs for unique values of variables* checkbox.
6. Enter **foreign** as the *Variable*.
7. Check the *Add a graph with totals* checkbox.
8. Click the **Submit** button.

The command and the associated graph are

```
. twoway (scatter mpg weight), by(foreign, total)
```



The graphs show that the relationship is nonlinear for both types of cars.

Syntax note: this is another use of a **by()** option. If you had used a **by** prefix, two separate graphs would have been generated.

Model fitting: Linear regression

After looking at the graphs, we would like to fit a regression model that predicts MPG from the weight and type of the car. From the graphs, the relationship is nonlinear and so we will try modeling MPG as a quadratic in weight. Also from the graphs, we judge the relationship to be different for domestic and foreign cars. We will include an indicator (dummy) variable for foreign and evaluate afterward whether this adequately describes the difference. Thus, we will fit the model

$$\text{mpg} = \beta_0 + \beta_1 \text{weight} + \beta_2 \text{weight}^2 + \beta_3 \text{foreign} + \hat{\epsilon}$$

foreign is already an indicator (0/1) variable, but we need to create the weight-squared variable. This can be done via the menus, but here using the command line is simpler:

```
. gen wtsq = weight^2
```

Now that we have all the variables we need, we can run a linear regression. We'll use the menus and see that the command is also simple. To use the menus, select **Statistics > Linear models and related > Linear regression**. In the resulting dialog, choose **mpg** as the dependent variable and **weight**, **wtsq**, and **foreign** as the independent variables. **Submit** the command. Here is the equivalent simple **regress** command and the resulting analysis-of-variance table.

```
. regress mpg weight wtsq foreign
```

Source	SS	df	MS
Model	1689.15372	3	563.05124
Residual	754.30574	70	10.7757963
Total	2443.45946	73	33.4720474

```
Number of obs = 74
F( 3, 70) = 52.25
Prob > F = 0.0000
R-squared = 0.6913
Adj R-squared = 0.6781
```

Root MSE = 3.2827

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
weight	-.0165729	.0039692	-4.18	0.000	-.0244892	-.0086567
wtsq	1.59e-06	6.25e-07	2.55	0.013	3.45e-07	2.84e-06
foreign	-2.2035	1.059246	-2.08	0.041	-4.3161	-.0909002
_cons	56.53884	6.197383	9.12	0.000	44.17855	68.89913

The results look encouraging, so we will plot the predicted values on top of the scatterplots for each of the types of cars. To do this, we need the predicted, or fitted, values. This can be done via the menus, but doing it by hand is simple enough. We will create a new variable **mpghat**:

```
. predict mpghat  
(option xb assumed; fitted values)
```

The output from this command is simply a notification. Go over to the Variables window and scroll to the bottom to confirm that there is now an **mpghat** variable. If you were to try this command when **mpghat** already exists, Stata will refuse to overwrite your data:


```
. predict mpghat  
mpghat already defined  
r(110);
```

The **predict** command, when used after a regression, is called a *postestimation command*. As specified, it creates a new variable called **mpghat** equal to

$$-.0165729\text{weight} + 1.59 * 10^{-6}\text{wtsq} - 2.2035\text{foreign} + 56.53884$$

For careful model fitting, there are several features available to you after estimation—one is calculating predicted values. Be sure to read [U] **20 Estimation and postestimation commands**.

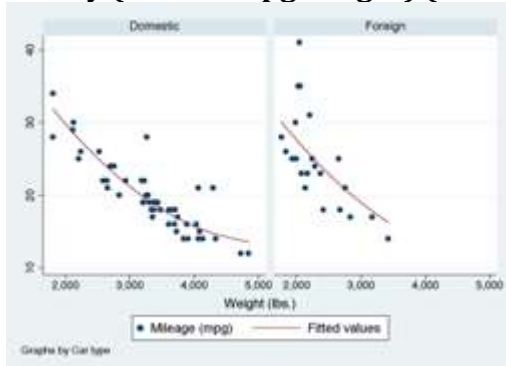
We can now graph the data and the predicted curve to evaluate the fit on the foreign and domestic data separately to determine if our shift parameter is adequate. We can draw both graphs together. Using the menus and a dialog, do the following:

1. Select **Graphics > Twoway graph (scatter, line, etc.)** from the menus.
2. If there are any plots listed, click the **Reset** button, .
3. Create the graph for **mpg** versus **weight**.
 - a. Click the **Create...** button.
 - b. Be sure that *Basic plots* and *Scatter* are selected.
 - c. Select **mpg** as the *Y variable* and **weight** as the *variable*.
 - d. Click **Accept**.

4. Create the graph showing **mpg** versus **weight**.
 - a. Click the **Create...** button.
 - b. Be sure that *Basic plots* and *Line* are selected.
 - c. Select **mpg** as the *Y variable* and **weight** as the *X variable*.
 - d. Check the *Sort on x variable* box. Doing so ensures that the lines connect from smallest to largest **weight** values, instead of the order in which the data happen to be.
 - e. Click **Accept**.
5. Show two plots, one each for domestic and foreign cars, on the same graph.
 - a. Click the **By** tab.
 - b. Check the *Draw subgraphs for unique values of variables*.
 - c. Enter **foreign** in the *Variables* field.
6. Click the **Submit** button.

Here are the resulting command and graph.

. twoway (scatter mpg weight) (line mpghat weight, sort), by(foreign)

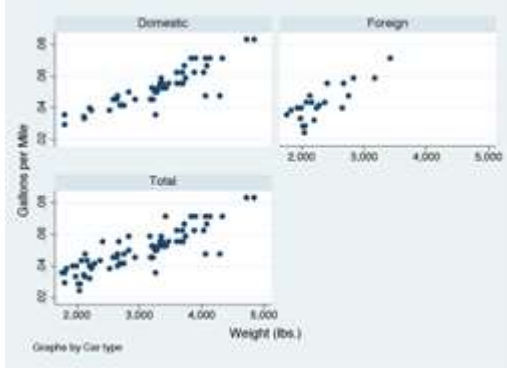


Here we can see the reason for enclosing the separate **scatter** and **line** commands in parentheses: they can then be overlaid by submitting them together. The fit of the plots looks good and is cause for initial excitement. So much excitement, in fact, that we decide to print the graph and show it to an engineering friend. We print the graph, being careful to print the graph and not all our results: **File > Print** from the Graph window menu bar.

When we show our graph to our engineering friend, she is concerned. “No,” she says. “It should take twice as much energy to move 2,000 pounds 1 mile compared with moving 1,000 pounds the same distance, and therefore it should consume twice as much gasoline. Miles per gallon is not a quadratic in weight; gallons per mile is a linear function of weight. Don't you remember any physics?”

We try out what she says. We need to generate a gallons-per-mile variable and make a scatterplot. Here are the commands that we would need—note their similarity to commands issued earlier in the session. There is one new command: the **label variable** command, which allows us to give the **gpm** variable a *variable label* so that the graph is labeled nicely.

- . generate gpm = 1/mpg
- . label variable gpm "Gallons per mile"
- . twoway (scatter gpm weight), by(foreign, total)



Sadly satisfied that the engineer is indeed correct, we rerun the regression:

- . regress gpm weight foreign

Source	SS	df	MS
Model	.009117618	2	.004558809
Residual	.00284001	71	.00004
Total	.011957628	73	.000163803

Number of obs = 74
 F(2, 71) = 113.97
 Prob > F = 0.0000
 R-squared = 0.7625
 Adj R-squared = 0.7558
 Root MSE = .00632

gpm	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
weight	.0000163	1.18e-06	13.74	0.000	.0000139 .0000186
foreign	.0062205	.0019974	3.11	0.003	.0022379 .0102032
_cons	-.0007348	.0040199	-0.18	0.855	-.0087504 .0072807

We find that although foreign cars had better gas mileage than domestic cars in 1978, it was because they were so light. In fact, according to our model a foreign car with the same weight as a domestic car would use an additional 1/160 gallon per mile driven. With this, we are satisfied with our analysis.

Commands versus menus

In this chapter you have seen that Stata can operate either via menu choices and dialog boxes or via the Command window. As you become more familiar with Stata, you will find that the Command window is typically much faster for often-used commands, whereas the menus and dialogs are faster when building up complex commands such as graphs.

One of Stata's great strengths is the consistency of its *command syntax*. Most of Stata's commands share the following syntax, where square brackets mean that something is optional and a *varlist* is a list of variables.

`[prefix:] command [varlist] [if] [in] [weight] [, options]`

Some general rules:

- Most commands accept prefix commands that modify their behavior; see [U] **11.1.10 Prefix commands** for details. One of the more common prefix commands is **by**.
- If an optional *varlist* is not specified, all the variables are used.
- *if* and *in* restrict the observations on which the command is run.
- *options* modify what the command does.
- Each command's syntax is found in the online help and the reference manuals.
- Stata's command syntax includes more than we have shown you here, but this should get you started. For more information, see [U] **11 Language syntax** and **help language**.

We saw examples using all the pieces of this except for the **in** qualifier and the **weight** clause. The syntax for all commands can be found in the online help along with examples—see chapter 6 for more information. The explanation of the syntax can be found online using **help language**. The consistent syntax makes it straightforward to learn new commands and to read others' commands when examining an analysis.

Here is an example of reading the syntax diagram by using the **summarize** command from earlier in this chapter. The syntax diagram for **summarize** is typical:

summarize `[varlist] [if] [in] [weight] [, options]`

This means that

command by itself is **summarize**
valid:

command followed by a **summarize mpg**
varlist (variable list) is
valid:

summarize mpg
weight

command with *if* (with **summarize if**
or without a *varlist*) is **mpg>20**
valid:


```
summarize mpg  
weight if mpg>20
```

and so on.

You can learn about **summarize** in [R] **summarize**, or select **Help > Stata Command...** and enter **summarize**, or type **help summarize** in the Command window.

Keeping track of your work

It would have been useful if we had made a log of what we did so that we could conveniently look back at interesting results or track any changes that were made. Your logs will contain commands and their output—another reason to learn command syntax, so that you can remember what you've done.

To make a log file that keeps track of everything appearing in the Results window, click the button that looks like a lab notebook, . Choose a place to store your log file, and give it a name, just as you would any other document. The log file will save everything that appears in the Results window from the time you start a log file to the time that you close it.

STATA LEARNING MODULES dapat diunduh di :

<http://www.ats.ucla.edu/stat/stata/modules/default.htm>