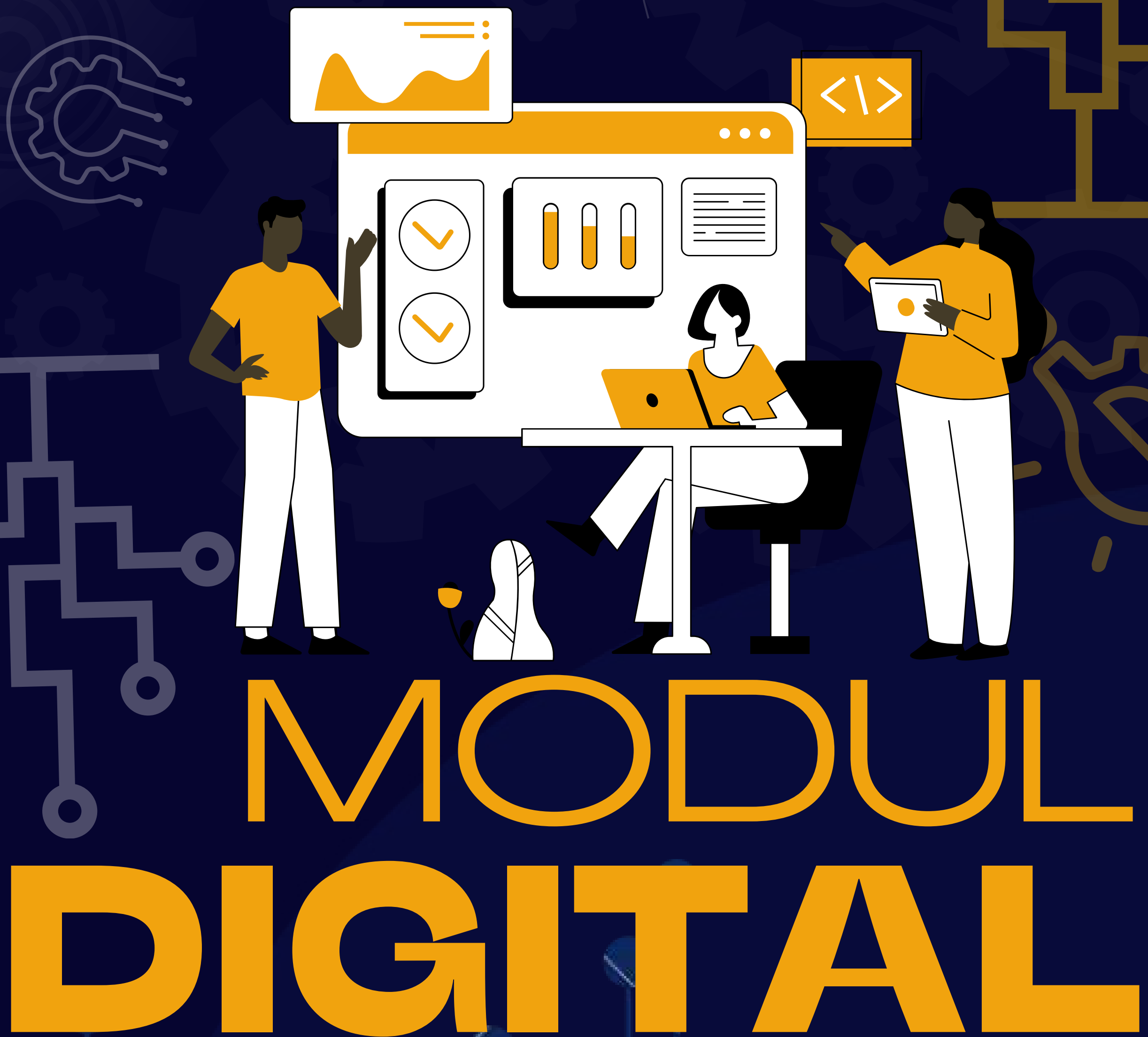


Dr. Arif Sabta Aji S.Gz., MQM. | Bunga Astria Paramashanti, MPH, PhD.  
| Hastrin Hositanisita, M.Sc. | Herwinda Kusuma Rahayu, MPH. | Sintha  
Dewi Purnamasari, MS. | Herni Dwi Herawati, MPH.



# MODUL DIGITAL

Biostatistik untuk *Applied Research Method  
in Nutrition 2*

# MODUL DIGITAL

Biostatistik untuk *Applied Research Method in Nutrition 2*

Penulis:

Dr. Arif Sabta Aji, S.Gz., MQM |

Bunga Astria Paramashanti, MPH, PhD. |

Hastrin Hositanisita, M.Sc. | Herwinda Kusuma Rahayu, MPH. |

Sintha Dewi Purnamasari, MS. | Herni Dwi Herawati, MPH.



**UNDANG-UNDANG REPUBLIK INDONESIA  
NOMOR 28 TAHUN 2014 TENTANG HAK CIPTA**

**LINGKUP HAK CIPTA**

Pasal 1

1. Hak Cipta adalah hak eksklusif pencipta yang timbul secara otomatis berdasarkan prinsip deklaratif setelah suatu ciptaan diwujudkan dalam bentuk nyata tanpa mengurangi pembatasan sesuai dengan ketentuan peraturan perundang-undangan.

**KETENTUAN PIDANA**

Pasal 113

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp 100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).
3. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/ atau pidana denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah).
4. Setiap Orang yang memenuhi unsur sebagaimana dimaksud pada ayat (3) yang dilakukan dalam bentuk pembajakan, dipidana dengan pidana penjara paling lama 10 (sepuluh) tahun dan/atau pidana denda paling banyak Rp4.000.000.000,00 (empat miliar rupiah).

# MODUL DIGITAL

## Biostatistik untuk *Applied Research Method in Nutrition 2*

Penulis :

Dr. Arif Sabta Aji, S.Gz., MQM | Bunga Astria Paramashanti, MPH, PhD. |  
Hastrin Hositanisita, M.Sc. | Herwinda Kusuma Rahayu, MPH. |  
Sintha Dewi Purnamasari, MS. | Herni Dwi Herawati, MPH.

*All rights reserved*

Hak Cipta Dilindungi Undang-undang  
Hak Penerbitan pada Jejak Pustaka  
Isi di Luar Tanggung Jawab Penerbit

ISBN:

**Editor:**

Pramitha Sari, S.Gz., MHKes

**Tata Letak Isi:**

Dr. Arif Sabta Aji

**Desain Cover:**

Dr. Arif Sabta Aji

iii + 118 hlm: 14 x 21 cm

Cetakan Pertama, Juli 2023

Penerbit:

**Jejak Pustaka**

Anggota IKAPI No. 141/DIY/2021  
Wirokerten RT.002 Desa Wirokerten  
Banguntapan Bantul Yogyakarta  
[jejakpustaka@gmail.com](mailto:jejakpustaka@gmail.com)  
081320748380



## **KATA PENGANTAR**

### ***Assalamualaikum Wr. Wb***

Puji syukur kita panjatkan kehadirat Allah SWT, karena hanya dengan Ridha-Nya kita selalu mendapatkan kebahagiaan, kedamaian, serta ketenteraman yang senantiasa menyertai kita. Shalawat serta salam kita haturkan kepada junjungan kita Nabi Muhammad SAW, beserta sahabat dan keluarganya.

Modul biostatistik untuk applied research method in nutrition 2 merupakan salah satu penunjang pada kurikulum berbasis kompetensi dan merupakan kelanjutan dari rencana pembelajaran semester applied research Method in nutrition 1. Buku ini mempunyai beberapa kajian yaitu biostatistik lanjut, analisis data hubungan, penyajian data, hasil dan pembahasan penelitian, dan penelitian kualitatif.

Buku modul digital biostatistik untuk applied research method in nutrition 2 ini disusun agar menjadi buku yang menarik, edukatif, dan membantu mahasiswa sehingga dapat menjadi pemicu bagi mahasiswa untuk terus belajar secara mandiri dengan sumber-sumber referensi yang tidak terbatas di dunia ini. Taklupa kami ucapkan terima kasih yang sebesar-besarnya kepada semua pihak dan kontributor yang telah memberikan saran yang begitu berharga bagi pengembangan kurikulum dan blok applied research

method in nutrition 2 ini khususnya. kami menyadari bahwa rpp blok applied research method in nutrition 2 ini masih jauh dari sempurna, oleh karena itu kami akan terus menerima dan menampung masukan, kritik dan saran yang membangun dari berbagai pihak, untuk dapat kami kembangkan demi kemajuan kurikulum.

Akhir kata, kami harap mahasiswa mendapatkan pengalaman yang menarik dan berharga selama mengikuti blok ini. Keep your learning passion, youth! *Be a truly successful Nutritionist!*

***Wassalammualaikum Wr. Wb.***

Yogyakarta, Juli 2023  
Penulis,

# DAFTAR ISI

<b>Kata Pengantar .....</b>	<b>i</b>
<b>Daftar Isi .....</b>	<b>iii</b>

## **BAB I**

<b>PENDAHULUAN .....</b>	<b>1</b>
A. Pengantar Biostatistik .....	1
B. Aplikasi Biostatistik dalam Bidang Gizi dan Kesehatan.....	1
A. Pengantar Analisis Data .....	4
B. Uji Hipotesis .....	9
C. Konsep Uji Parametrik dan Non-Parametrik .....	14

## **BAB II ANALISIS UNIVARIAT**

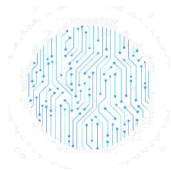
A. Analisis Univariat.....	17
B. Analisis Deskriptif .....	17
C. Bentuk Penyajian Data Analisis Univariat .....	23
D. Uji Normalitas Data.....	25
E. Kasus Analisis Deskriptif.....	25

## **BAB III**

<b>ANALISIS BIVARIAT .....</b>	<b>37</b>
A. Analisis Bivariat.....	37
B. Uji t Independen.....	38
C. Uji t Dependen.....	43



D. Uji ANOVA.....	48
E. Uji Chi Square .....	62
F. Uji Korelasi .....	78
G. Uji Regresi Linier Sederhana.....	91
<b>DAFTAR PUSTAKA .....</b>	<b>115</b>
<b>PROFIL PENULIS.....</b>	<b>117</b>



# **BAB I**

## **PENDAHULUAN**

### **A. Pengantar Biostatistik**

Statistik adalah salah satu ilmu yang bertujuan untuk menentukan sebuah keputusan dan menemukan hal yang baru didasarkan dari data yang didapatkan setelah melakukan rancangan penelitian atau eksperimen, pengambilan data dan menganalisisnya. Ilmu statistik digunakan di berbagai bidang keilmuan baik sosial, ekonomi, biologi, teknik, kedokteran dan keilmuan lainnya. Biostatistik sendiri menjadi salah satu cabang dari ilmu statistik terapan yang fokus pada rancangan penelitian dan analisis data yang berhubungan dengan ilmu kedokteran, biologi dan masalah kesehatan. Cakupan keilmuan biostatistik meliputi proses pengambilan data, pengorganisasian data, interpretasi data numerik, merancang penelitian dan analisis data penelitian dalam bidang biologi dan kesehatan.

### **B. Aplikasi Biostatistik dalam Bidang Gizi dan Kesehatan**

Data yang telah didapatkan baik dari hasil penelitian observasional maupun eksperimental. Beberapa contoh

aplikasi atau manfaat penggunaan biostatistik adalah bisa digunakan untuk menentukan:

1. Mengetahui faktor-faktor risiko yang berhubungan dengan permasalahan gizi atau penyakit pada populasi

Adanya permasalahan gizi atau penyakit di sebuah populasi membutuhkan identifikasi tentang faktor-faktor yang berhubungan dan menjadi faktor risiko. Identifikasi faktor risiko yang berhubungan ini bisa diaplikasikan untuk melakukan langkah preventif atau pencegahan. Contohnya berdasarkan hasil penelitian oleh RD Semba et.al (2007), menggunakan uji *multivariate logistic regression*, didapatkan hasil ayah perokok berhubungan dengan meningkatnya risiko anak stunting dan wasting keluarga miskin di wilayah urban di Indonesia.

2. Membandingkan efektivitas beberapa jenis intervensi gizi

Berdasarkan penelitian ditemukan bahwa hipertensi dipengaruhi oleh faktor gaya hidup dan pola makan. Peneliti kemudian ingin membandingkan beberapa jenis terapi diet untuk hipertensi untuk mengetahui jenis intervensi diet yang bisa menurunkan tekanan darah secara signifikan. Contohnya dalam penelitian LJ Appel (2005) *‘Effects of Protein, Monounsaturated Fat, and Carbohydrate Intake on Blood Pressure and Serum Lipids*

*Results of the OmniHeart Randomized Trial* bertujuan untuk membandingkan pengaruh 3 jenis diet terhadap tekanan darah (tujuan utama) dan terhadap perubahan berat badan (tujuan khusus). Beberapa jenis uji statistik seperti *paired t-test* (uji t-test berpasangan), *Wilcoxon rank-based*, dan *Analysis of Covariance* (Ancova) untuk menjawab tujuan penelitian tersebut.

3. Mengetahui ada tidaknya hubungan kausalitas atau sebab akibat antar variable

Penggunaan uji statistik yang tidak tepat bisa menimbulkan kesalahan dalam pembuatan kesimpulan. Tidak setiap variabel independen menjadi variabel yang berhubungan sebab akibat dengan variabel dependen. Uji statistik bisa digunakan untuk mengontrol dan menemukan ada atau tidaknya hubungan ini. Contohnya, dalam penelitian yang menyebutkan bahwa ayah perokok meningkatkan risiko anak terkena *wasting* dan *stunting*. Namun, apakah rokok menjadi penyebab *wasting* dan *stunting* pada anak? Adakah hubungan kausalitas atau sebab akibat dari faktor ayah perokok pada kejadian *stunting* dan *wasting*? Untuk menjawab pertanyaan tersebut dibutuhkan metode penelitian, jumlah sampel, dan uji statistik yang sesuai untuk mengontrol berbagai macam variabel yang kemungkinan menjadi faktor perancu.

### **C. Pengantar Analisis Data**

Analisis berasal dari Bahasa Yunani yang terdiri dari kata “ana” yang berarti atas dan “lysis” yang berarti memecahkan, sehingga Ian Dey (1995) mendefinisikan analisis sebagai suatu proses pemecahan data menjadi komponen-komponen penyusunnya untuk mengungkapkan elemen dan struktur karakteristiknya. Adapun data adalah sekumpulan informasi yang berisi kumpulan fakta berupa angka-angka atau kata-kata hasil dari observasi, pengukuran, perhitungan ataupun pendeskripsian suatu benda. Maka, untuk menganalisis data, data tersebut harus dipecah terlebih dahulu, kemudian digabungkan kembali bersama untuk memperoleh pemahaman yang baru.

Analisis data adalah bagian penting dalam membantu menjawab pertanyaan penelitian bagi peneliti, dosen dan mahasiswa maupun memberikan informasi-informasi penting bagi praktisi di berbagai bidang, seperti kesehatan, bisnis, sosial dan sebagainya. Keterampilan dalam menganalisis data membutuhkan pemikiran kritis di sepanjang prosesnya, termasuk dalam pemilihan metode analisis data yang tepat, interpretasi data dan penyajian datanya. Dalam modul ini, cakupan analisis data dan contoh-contohnya dikhususkan untuk tujuan penelitian bagi para peneliti, dosen dan mahasiswa di bidang gizi dan kesehatan. Dalam sebuah penelitian, analisis data merupakan proses yang diawali dengan perencanaan yang hati-hati, diikuti oleh pengumpulan data, analisis data dan interpretasi hasil

analisis data. Analisis data tidak hanya soal statistik dan angka-angka, melainkan data dalam sebuah konteks. Data adalah kumpulan informasi dimana konteks data membuatnya menjadi berarti. Contohnya, seorang mahasiswa mendapatkan nilai rerata skor Z tinggi badan menurut umur sebesar -2.50 standar deviasi dibandingkan dengan populasi rujukan. Angka ini akan lebih memberikan informasi penting jika data tersebut dikaitkan dengan informasi lainnya, misalnya, konteks lokasi dimana data tersebut diambil, prevalensi saat ini, tren prevalensi dari tahun-tahun sebelumnya, faktor-faktor yang berkaitan dengan nilai rerata tersebut dan sebagainya. Dalam hal ini, konteks data akan membantu dalam interpretasi data dan proses analisis secara keseluruhan. Oleh karena itu, modul ini mencakup contoh-contoh yang relevan dalam mengantar peneliti untuk melakukan analisis data sesuai konteks di bidang gizi dan kesehatan.

Secara umum, tujuan dari analisis data adalah untuk memperoleh informasi yang dapat digunakan sebagai dasar pengambilan keputusan dalam memecahkan suatu masalah. Baik data kualitatif maupun kuantitatif, tujuan khusus dari analisis data meliputi: 1) untuk mendeskripsikan dan meringkas data, 2) untuk membandingkan dan menguji teori/konsep berdasarkan data, 3) menemukan adanya teori/konsep baru berdasarkan data dan 4) mencari penjelasan apakah teori/konsep baru yang diuji berlaku secara umum atau hanya pada kondisi tertentu.

Secara garis besar, tahapan dalam analisis data meliputi persiapan, analisis dan interpretasi hasil analisis sebagaimana dijelaskan sebagai berikut:

### 1. Persiapan

Pada tahap awal ini, peneliti mempersiapkan data yang selanjutnya dapat dianalisis untuk menjawab pertanyaan penelitiannya.

#### a. Pengecekan atau validasi data

Validasi data dilakukan untuk memahami apakah data yang dikumpulkan sesuai standar. Lebih spesifik, tujuannya adalah 1) untuk memastikan bahwa subjek atau responden penelitian lah yang memberikan data atau tanggapan pada kuesioner atau instrument penelitian, 2) untuk memastikan setiap subjek atau responden dipilih sesuai kriteria penelitian, 3) untuk memastikan standar etike penelitian dipertahankan saat pengumpulan data, dan 4) untuk memastikan bahwa subjek atau responden penelitian telah menjawab semua pertanyaan atau sebaliknya, pewawancara atau pengumpul data telah menanyakan semua pertanyaan yang telah dirancang dalam kuesioner atau instrumen penelitian.

#### b. Editing

Seringnya sampel data penelitian sarat dengan kesalahan. Baik responden maupun pewawan-

cara terkadang salah dalam mengisi pertanyaan kuesioner atau instrumen penelitian secara tidak sengaja. Fase pengeditan ini adalah proses dimana peneliti memastikan bahwa data yang terkumpul bebas dari kesalahan tersebut. Oleh karena itu, peneliti perlu memeriksa, memastikan kembali dan mengedit data tersebut terlebih dahulu.

c. Koding

Koding terkait dengan pengelompokan dan pemberian nilai atas variabel atau respon dari subjek atau responden. Misalnya, peneliti membuat kelompok usia untuk membedakan subjek berdasarkan kategori usia mereka sehingga memudahkan dalam menganalisis kumpulan data dalam kelompok-kelompok kecil daripada sekumpulan data yang besar.

2. Analisis data

Detil analisis data akan dijelaskan pada bab-bab berikutnya. Pada prinsipnya, analisis data akan disesuaikan dengan pertanyaan dan tujuan penelitian, serta jenis dan karakteristik dari data yang dikumpulkan.

Dalam analisis data kuantitatif, teknik analisis data dibagi menjadi dua, yaitu statistik deskriptif dan statistik inferensial. Statistik inferensial meliputi uji parametrik dan uji non-parametrik.



- a. Statistik deskriptif
- b. Uji statistik deskriptif digunakan untuk menganalisis data dengan mendeskripsikan atau menggambarkan data yang terkumpul sebagaimana adanya tanpa bermaksud untuk membuat kesimpulan atau generalisasi hasil temuan. Dalam statistik deskriptif, peneliti perlu memahami jenis datanya. Misalnya, pada data numerik, peneliti bisa mendapatkan frekuensi, persentase dan ukuran tendensi sentral (seperti modus, median, rerata). Hasil pengolahan data dengan statistik deskriptif memberikan kesan data yang lebih mudah dipahami maknanya. Selain itu, peneliti dapat mengelompokkan data berdasarkan variabel kelompoknya sehingga mudah diinterpretasikan maksudnya.
- c. Statistik inferensial  
Statistik inferensial digunakan untuk menghasilkan temuan sehingga dapat disimpulkan dan digeneralisasikan dalam konteks yang lebih luas lagi, seperti wilayah atau populasi yang lebih besar. Dalam uji ini, peneliti akan menguji hipotesisnya. Oleh karena itu, dalam berbagai literatur, analisis statistik inferensial juga sering disebut dengan analisis uji hipotesis. Jenis analisis ini juga terkadang disebut dengan statistik induktif atau statistik probabilitas, yaitu

jenis analisis statistik yang digunakan untuk menganalisis data sampel dan hasilnya diberlakukan untuk populasi. Mengingat pentingnya statistik inferensial dalam menyimpulkan temuannya, maka diperlukan persyaratan yang lebih ketat. Misalnya, dalam teknik pengambilan sampel, sampel harus representatif terhadap populasi dimana sampel memiliki karakteristik sebagaimana yang dimiliki oleh populasi yang diwakilinya.

### 3. Interpretasi hasil analisis

Detail analisis data akan dijelaskan pada bab-bab berikutnya. Pada prinsipnya, data yang telah dianalisis akan diinterpretasikan sesuai dengan hasil uji statistik dan didiskusikan sesuai dengan konteks pertanyaan penelitian.

## **D. Uji Hipotesis**

Hipotesis berasal dari kata “hupo” dan “thesis”. Hupo artinya sementara atau lemah kebenarannya, sedangkan thesis artinya pernyataan atau teori. Maka, hipotesis adalah pernyataan yang perlu diuji kebenarannya. Untuk menguji kebenaran suatu hipotesis, maka diperlukan uji hipotesis. Uji hipotesis memegang peranan penting dalam pengambilan keputusan berbasis data. Penelitian tanpa adanya uji hipotesis kemungkinan dapat mengarah ke penarikan

kesimpulan dan pengambilan keputusan salah. Berikut adalah langkah-langkah dalam uji hipotesis:

1. Membuat asumsi awal, yaitu dengan menetapkan hipotesis nol dan hipotesis alternative

Dalam uji hipotesis terdapat dua jenis hipotesis, yaitu hipotesis nol ( $H_0$ ) dan hipotesis alternatif ( $H_a$ ) sebagaimana penjelasan berikut:

- a. Hipotesis nol ( $H_0$ )

Hipotesis nol adalah pernyataan matematika yang digunakan untuk menunjukkan bahwa tidak ada perbedaan antara dua kemungkinan. Dengan kata lain, tidak ada perbedaan antara karakteristik data tertentu.

Contoh:

- ✓ Tidak ada perbedaan rerata asupan zat besi antara remaja putri yang menerima edukasi audiovisual dan remaja putri yang tidak menerima edukasi audiovisual.
- ✓ Tidak ada hubungan antara edukasi audiovisual dengan asupan zat besi pada remaja putri.

- b. Hipotesis alternatif ( $H_a$ )

Sebagaimana sebutannya, hipotesis alternatif adalah alternatif dari hipotesis nol. Hipotesis alternatif kebalikan logis dari hipotesis nol, sehingga penerimaan hipotesis alternatif akan mengikuti penolakan hipotesis nol.

Contoh:

- ✓ Ada perbedaan rerata asupan zat besi antara remaja putri yang menerima edukasi audiovisual dan remaja putri yang tidak menerima edukasi audiovisual.
  - ✓ Ada hubungan antara edukasi audiovisual dengan asupan zat besi pada remaja putri.
2. Menetapkan kriteria untuk pengambilan keputusan, yaitu tingkat signifikansi

Hasil uji statistik akan menentukan apakah hipotesis nol ditolak atau gagal ditolak. Kriteria atau batasan yang digunakan untuk memutuskan apakah hipotesis nol ditolak atau gagal ditolak disebut dengan tingkat signifikansi. Tingkat signifikansi juga biasa disebut dengan tingkat kemaknaan. Tingkat signifikansi merupakan kesalahan tipe I dari suatu uji yang diberi notasi  $\alpha$  atau alfa. Nilai ini menunjukkan besarnya peluang salah dalam menolak hipotesis nol. Dengan kata lain nilai alfa merupakan batas toleransi peluang salah dalam menolak hipotesis nol atau batas maksimal menolak hipotesis nol. Penentuan nilai alfa sendiri bergantung pada tujuan dan kondisi penelitian. Nilai alfa yang biasa digunakan adalah 1%, 5% dan 10%. Dalam bidang kesehatan masyarakat, nilai alfa biasanya adalah 5%, sedangkan uji obat-

obatan atau uji dengan menggunakan hewan coba biasanya menggunakan nilai alfa 1%.

Interval kepercayaan adalah salah satu bagian penting dalam uji hipotesis. Hal ini karena tingkat signifikansi atau alfa dapat ditentukan dari interval kepercayaan yang diberikan. Misalnya, interval kepercayaan 95%, maka nilai alfa adalah  $100\% - 95\% = 5\%$  atau 0,05 untuk pengujian hipotesis satu sisi. Untuk uji hipotesis dua sisi, maka nilai 0,05 dibagi dengan 2, yaitu 0,025.

Biasanya peneliti akan berpegang pada nilai-p dari hasil uji statistik. Nilai p adalah probabilitas yang mengasumsikan hipotesis nol benar. Nilainya selalu berupa angka antara 0 dan 1. Lalu, nilai-p akan dibandingkan dengan tingkat signifikansi atau alfa, yaitu risiko yang dapat diterima dari penolakan hipotesis nol yang salah. Jika nilai p lebih kecil daripada tingkat signifikansi yang diharapkan, maka hipotesis nol ditolak. Namun jika nilai p lebih besar dari tingkat signifikansi yang diharapkan, maka hipotesis nol gagal ditolak.

### 3. Menentukan uji statistik yang sesuai

Ada berbagai jenis uji statistik yang dapat digunakan untuk menguji hipotesis penelitian. Setiap uji statistik memiliki persyaratan atau asumsi tertentu yang harus dipenuhi. Jenis uji statistik bergantung pada variabel yang akan dianalisis dan distribusi

populasinya, normal atau tidak. Detil penjelasan mengenai jenis-jenis uji statistik akan dipaparkan di bab-bab berikutnya.

4. Berdasarkan hasil uji statistic

Menentukan apakah asumsi awal ditolak atau gagal ditolak (diterima). Digarisbawahi bahwa istilah hipotesis “diterima” sebetulnya kurang tepat. Tepatnya adalah hipotesis “gagal ditolak”. Gagal menolak hipotesis berbeda dengan mengakui kebenaran (menerima hipotesis), melainkan peneliti tidak memiliki cukup bukti untuk menolak hipotesis.

Hasil uji yang signifikan atau tidak akan menentukan pengambilan keputusan yang dibuat terkait nilai yang dinyatakan dalam hipotesis nol. Ketika hipotesis nol ditolak, maka hasil uji mencapai signifikansi dan ketika hipotesis nol dipertahankan, maka hasil uji gagal mencapai signifikansi. Sebagai contoh, ketika nilai  $p$  kurang dari 5% ( $p < 0,05$ ), maka peneliti menolak hipotesis nol. Peneliti akan merujuk  $p < 0,05$  sebagai kriteria untuk memutuskan untuk menolak hipotesis nol. Ketika nilai  $p$  lebih besar dari 5% ( $p > 0,05$ ), peneliti gagal menolak hipotesis nol. Keputusan untuk menolak atau mempertahankan hipotesis nol disebut signifikansi. Ketika nilai  $p$  kurang dari 0,05, maka hasil penelitian mencapai signifikansi dan keputusannya adalah menolak hipotesis nol. Ketika nilai  $p$  lebih besar dari 0,05, maka hasil penelitian gagal mencapai

signifikansi dan keputusannya adalah untuk mempertahankan hipotesis nol.

### **E. Konsep Uji Parametrik dan Non-Parametrik**

Uji statistik parametrik didasarkan pada asumsi tentang distribusi populasi dari mana sampel diambil. Asumsi-asumsi uji parametrik meliputi;

1. Data terdistribusi normal
2. Melibatkan parameter statistik, seperti rerata dan standar deviasi
3. Sampel diambil secara acak
4. Sampel memiliki varians yang sama (*equal variance*)
5. Berskala interval dan rasio
6. Ukuran pemusatan berupa nilai rerata

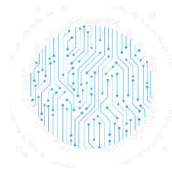
Adapun uji statistik nonparametrik tidak didasarkan pada asumsi, yaitu data dapat dikumpulkan dari sampel yang tidak mengikuti distribusi tertentu. Pada umumnya, uji nonparametric menggunakan data yang berskala nominal dan ordinal, serta ukuran pemusatan berupa nilai median. Tabel berikut menunjukkan jenis-jenis uji statistik parametrik dan nonparametrik.

**Tabel, 1.** Jenis Uji Parametrik dan Nonparametrik

<b>Uji Parametrik</b>	<b>Uji Nonparametrik</b>
One sample t-test	Sign test/Wilcoxon signed rank test
Two (independent) sample t test	Mann whitney u test/ Wilcoxon rank sum test
Paired t-test	Sign test/Wilcoxon signed rank test
One-way ANOVA	Kruskall wallis test
Repeated ANOVA	Friedman test
Pearson correlation	Spearman







## **BAB II**

# **ANALISIS UNIVARIAT**

### **A. Analisis Univariat**

Biostatistika memberikan beragam cara untuk mengolah dan menganalisis data. Analisis univariat sendiri adalah jenis analisis statistic yang menggunakan satu variabel saja. Pembahasan analisis satu variabel ini dilakukan dengan menjelaskan berbagai macam informasi baik untuk data dalam bentuk kategorik atau numerik. Variable terpilih akan dihitung dan dianalisis secara terpisah sehingga menghasilkan gambaran secara keseluruhan pada satu variabel tersebut secara jelas. Karakteristik dari variabel akan dieksplorasi dan dipresentasikan. Metode statistic seperti distribusi frekuensi, diagram batang, histogram, diagram lingkaran, dan ukuran pemusatan seperti mean, median, dan modus. Bentuk data seperti akan sangat menentukan bentuk peringkasan datanya.

### **B. Analisis Deskriptif**

Analisis deskriptif memiliki bentuk yang bergantung dari jenis datanya. Untuk data numerik yang digunakan adalah nilai mean (rata-rata), median, standard deviasi dan

inter kuartil range, minimal dan maksimal. Sedangkan untuk data kategorik, presentasi data tersebut menggunakan distribusi frekuensi dengan ukuran persentase atau proporsi.

### 1. Data Numerik

Peringkasan untuk data jenis numerik diantaranya adalah dengan menentukan ukuran tengah dimana cerminan dari konsentrasi nilai-nilai hasil pengukuran seperti mean, median, dan modus. Selain itu, ada juga pengelompokkan berdasarkan ukuran variasi seperti range, jarak inter kuartil, dan standar deviasi.

### 2. Mean

Mean adalah rata-rata hasil dari jumlah semua nilai pengukuran yang dibagi oleh banyaknya nilai pengukuran tersebut. Keuntungan nilai mean adalah sudah melibatkan seluruh hasil pengukuran. Kelemahan dari mean adalah sangat dipengaruhi oleh nilai ekstrim (ekstrim tinggi atau rendah) pada suatu pengukuran. Jika pada suatu data hasil pengukuran memiliki nilai ekstrim maka akan menghasilkan kurva yang menceng/miring dan tidak bisa mewakili rata-rata pengukuran hasil pengamatan. Contohnya adalah nilai ekstrim dari tinggi badan. Berdasarkan data dari tinggi badan anak remaja laki-laki Indonesia usia 18 tahun didapatkan rata-rata 165 cm, sebenarnya Sebagian besar tinggi badannya adalah di bawah 165 cm.

Mean sebesar 165 cm diperoleh karena tarikan sekelompok kecil orang yang memiliki tinggi badan sangat tinggi. Dengan demikian penggunaan mean untuk data yang ada nilai ekstrimnya (data yang hasil kurva distribusinya menceng/miring) kurang tepat.

Contoh dalam perhitungan rata-rata lama pasien di rawat di RS adalah 1 hari, 3 hari, 4 hari, 2 hari, 90 hari. Perhitungan meannya adalah

$$\text{Mean} = (1+3+4+2+90)/5 = 20 \text{ hari}$$

Dari hasil perhitungan didapatkan rata-rata lama hari rawat 20 hari, hasil ini tentunya tidak dapat mewakili karena secara visual datanya Sebagian besar kurang dari 5 hari. Keadaan ini bisa terjadi karena kumpulan data di atas ada nilai ekstrimnya.

### 3. Median

Median adalah nilai dimana setengah banyaknya pengamatan mempunyai nilai di bawahnya dan setengahnya lagi mempunyai nilai di atasnya. Berbeda dengan nilai mean, penghitungan median hanya mempertimbangkan urutan nilai hasil pengukuran, besar beda antar nilai di abaikan. Karena mengabaikan besar beda, maka median tidak dipengaruhi oleh nilai ekstrim.

Prosedur penghitungan median melalui langkah

- a. Data diurutkan/di-array dari nilai kecil ke besar
- b. Hitung posisi median dengan rumus  $(n+1)/2$

c. Hitung nilai mediannya

Contoh ada usia 6 mahasiswa 20 th, 26 th, 24 th, 30 th, 40 th, 36 th

Data diurutkan: 20, 24, 26, 30, 36, 40

Posisi =  $(6+1)/2 = 3,5$

Mediannya adalah data yang urutannya ke 3,5 yaitu  $(26 + 30)/2 = 28$

Jadi 50% mahasiswa berumur dibawah 28 tahun dan 50% mahasiswa berumur di atas 28 tahun.

4. Modus/Mode

Mode adalah nilai pengamatan yang mempunyai frekuensi/jumlah terbanyak.

Contoh mode data umur mahasiswa: 18 th, 22 th, 21 th, 20 th, 23th, 20 th. Dari data tersebut berarti mode-nya adalah 20 tahun.

a. Bentuk distribusi data

Hubungan nilai mean, median dan mode akan menentukan bentuk distribusi data:

- 1) Bila nilai mean, median dan mode sama, maka bentuk distribusi datanya normal
- 2) Bila nilai mean  $>$  median  $>$  mode, maka bentuk distribusi datanya menceng/ miring ke kanan
- 3) Bila nilai mean  $<$  median  $<$  mode, maka bentuk distribusi datanya menceng/miring ke kiri

b. Range

Range merupakan ukuran variasi yang paling dasar, dihitung dari selisih nilai terbesar dengan nilai terkecil. Kelemahan range adalah dipengaruhi nilai ekstrim. Keuntungan penghitungan dapat dilakukan dengan cepat.

c. Jarak Inter Quartil

Nilai observasi disusun berurutan dari nilai ke cil ke besar, kemudian ditentukan kuartil bawah dan atas. Kuartil merupakan pembagian data menjadi 4 bagian yang dibatasi oleh tiga ukuran kuartil, yaitu kuartil I, kuartil II dan kuartil III.

Kuartil I mencakup 25% data berada di bawahnya dan 75% data berada di atasnya.

Kuartil II (median) mencakup 50% data berada di bawahnya dan 50% data berada di atasnya.

Kuartil III mencakup 75% data berada di bawahnya dan 25% data berada di atasnya.

Jarak inter kuartil adalah selisih anatar kuaril III dan kuaril I. Ukuran ini lebih baik dari range, terutama kalau frekuensi pengamatan banyak dan distribusi sangat menyebar.

d. Standar Deviasi

Variasi data yang diukur melalui penyimpangan /deviasi dari nilai-nilai pengamatan terhadap nilai

meannya. Rata-rata hitung dari kuadrat deviasi terhadap mean disebut varian. Semakin besar nilai varian akan semakin bervariasi, karena satuan varian (kuadrat) yang tidak sama dengan satuan nilai pengamatan, maka dikembangkan suatu ukuran variasi yang mempunyai satuan yang sama dengan satuan pengamatan, yaitu Standard Deviasi. Seperti halnya varian, semakin besar SD semakin besar variasinya. Apabila tidak ada variasi, maka  $SD=0$ .

Dari uraian tersebut dapat disimpulkan, untuk data numerik digunakan nilai mean (rata-rata), median, standard deviasi dan inter quartil range, minimal dan maksimal. Bila data yang terkumpul tidak menunjukkan adanya nilai ekstrim (distribusi normal), maka perhitungan nilai mean dan standard deviasi merupakan cara analisis univariat yang tepat. Sedangkan bila dijumpai nilai ekstrim (distribusi data tidak normal), maka nilai median dan inter quartil range (IQR) yang lebih tepat dibandingkan nilai mean.

#### e. Data Kategorik

Pada data kategorik memiliki peringkasan data hanya menggunakan distribusi frekuensi dalam ukuran presentasi atau proporsi. Bila data kategorik, tentunya informasi/peringkasan disampaikan

kan adalah berdasarkan kelompoknya, bukan menggunakan ukuran mean atau median atau modus. Data yang disajikan dalam jenis data kategorik adalah informasi jumlah dan presentase. Untuk ukuran variasi, pada data kategorik variasi maksimal adalah apabila jumlah antar kategori sama.

Pada kelas A, jenis kelamin mahasiswa bervariasi (heterogen) karena 50% pria dan 50% wanita.

Pada kelas B, jenis kelamin mahasiswa tidak bervariasi (homogen pada pria) karena pria 90% dan wanita hanya 10%.

### C. Bentuk Penyajian Data Analisis Univariat

Bentuk penyajian data dari hasil analisis univariat adalah menggunakan tabel atau grafik. Tetapi, perlu diingat bahwa dianjurkan kita memilih salah satu saja dan tidak dianjurkan untuk memilih keduanya. Contoh penyajian analisis univariat adalah sebagai berikut;

#### 1. Data Numerik

**Tabel 1.** Distribusi Usia dan Tinggi Badan Anak Kelas IX

Variabel	Mean Median	SD	Minimal- Maksimal
Usia	17,18	4,5	16-19
Tinggi badan	165,50 162,12	5,89	155,24- 181,23



## 2. Data Kategorik

**Tabel 2.** Distribusi Kategori Tingkat Pendidikan Orang Tua Siswa

<b>Pendidikan</b>	<b>Jumlah</b>	<b>Presentase</b>
SD	12	8,6
SMP	11	7,8
SMA	45	32,1
Diploma	33	23,6
Sarjana	27	19,3
Pascasarjana	12	8,6
Total	140	100.0

Bagaimana menginterpretasi tabel di atas? “dilihat konsentrasi/jumlah yang terbesar data pada kelompok mana?”

Selain untuk mendeskripsikan masing-masing variabel, analisis univariat dapat juga sekaligus untuk mengeksplorasi variabel yang dapat berguna dalam mendiagnosis asumsi statistik lanjut (terutama untuk variabel jenis numerik), misalnya apakah variannya homogen atau heterogen, apakah distribusinya normal atau tidak. Eksplorasi data juga dapat untuk mendeteksi adanya nilai ekstrim/outlier, bila ada nilai ekstrim sangat menentukan analisis selanjutnya (bivariat) apakah nilainya akan berkurang.

#### **D. Uji Normalitas Data**

Uji kenormalan data:

Untuk mengetahui suatu data berdistribusi normal, ada 3 cara untuk mengetahuinya yaitu:

- a. Dilihat dari grafik histogram dan kurve normal, bila bentuknya menyerupai bel shape, berarti distribusi normal
- b. Menggunakan nilai Skewness dan standar error-nya, bila nilai Skewness dibagi standar error-nya menghasilkan angka  $\leq 2$ , maka distribusinya normal
- c. Uji kolmogorov smirnov, bila hasil uji signifikan ( $p$  value  $> 0,05$ ) maka distribusi normal. Namun uji kolmogorov sangat sensitif dengan jumlah sampel, maksudnya: untuk jumlah sampel yang besar uji kolmogorov cenderung menghasilkan uji yang signifikan (yang artinya bentuk distribusinya tidak normal). Atas dasar kelemahan ini dianjurkan untuk mengetahui kenormalan data lebih baik menggunakan angka skewness atau melihat grafik histogram dan kurve normal

#### **E. Kasus Analisis Deskriptif**

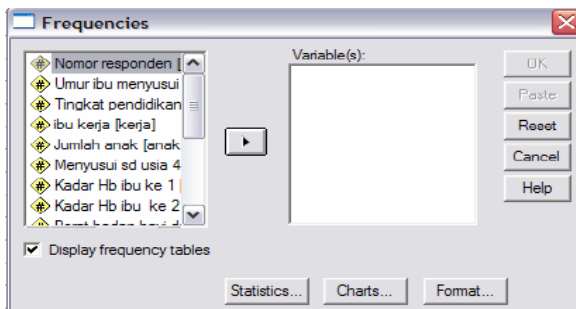
Tujuan analisis ini adalah untuk mendeskripsikan karakteristik masing-masing variabel yang diteliti. Bentuknya tergantung dari jenis datanya. Untuk data numerik digunakan nilai mean (rata-rata), median, standard

deviasi dll. Sedangkan untuk data katagorik tentunya hanya dapat menjelaskan angka/nilai jumlah dan persentase masing-masing kelompok. Berikut akan dipelajari cara mengeluarkan analisis deskriptif di SPAA, dimulai untuk variabel katagorik (sebagai latihan digunakan variabel “pendidikan”) dan kemudian dilanjutkan variabel numerik (variabel umur).

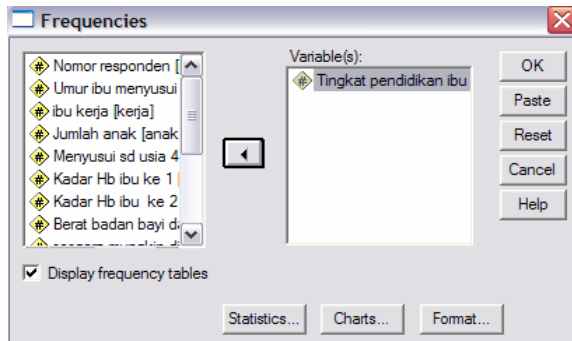
### 1. Data Kategorik

Untuk menampilkan tabulasi data katagorik digunakan tampilan frekuensi. Sebagai contoh kita akan menampilkan tabel distribusi frekuensi untuk variabel pendidikan dari file ‘ASI.SAV’.

- a. Dari menu utama SPSS pilih ‘Analyze’, kemudian ‘Descriptive Statistic’ dan pilih ‘Frequencies’, sehingga muncul tampilan:



- b. Sorot variabel ‘didik’. Klik tanda panah dan masukkan ke kotak “Variable (s)”



- c. Klik 'OK', hasil dapat dilihat di jendela output, seperti sbb:

**Frequencies**

Statistics

pendidikan formal ibu menyusui

N	Valid	50
	Missing	0

pendidikan formal ibu menyusui

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	10	20.0	20.0	20.0
	2	11	22.0	22.0	42.0
	3	16	32.0	32.0	74.0
	4	13	26.0	26.0	100.0
	Total	50	100.0	100.0	

Kolom 'Frequency' menunjukkan jumlah kasus dengan nilai yang sesuai. Pada contoh di atas, total responden 50 orang, dari jumlah tersebut 10 ibu yang berpendidikan SD, proporsi dapat dilihat pada kolom 'Percent', pada contoh di atas ada 20% ibu yang berpendidikan SD. Kolom 'Valid Percent' memberi hasil yang

sama karena pada data ini tidak ada 'missing cases'. 'Cumulative Percent'.

Penyajian dan Interpretasi di Laporan Penelitian adalah dari angka-angka tersebut kemudian kita masukkan ke tabel penyajian di laporan penelitian/laporan tesis. Adapun penyajian dan interpretasinya sbb:

Tabel ...  
Distribusi Responden Menurut Tingkat Pendidikan  
Di ..... X tahun ....

Pendidikan	Jumlah	Persentase
SD	10	20,0
SMP	11	22,0
SMU	16	32,0
PT	13	26,0
Total	50	100,0

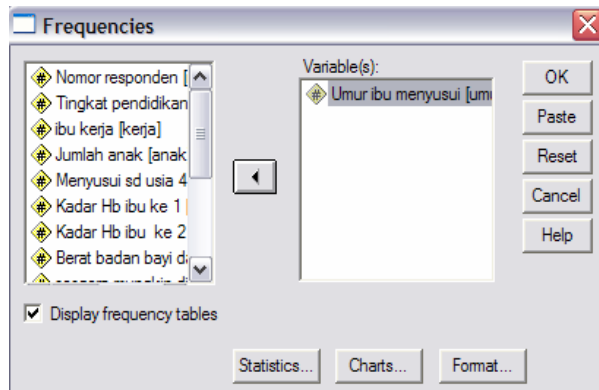
Distribusi tingkat pendidikan responden hampir merata untuk masing-masing tingkat pendidikan. Paling banyak responden berpendidikan SMU yaitu 16 orang (32,0%) sedangkan untuk pendidikan SD, SMP dan PT masing-masing 20,0%, 22,0% dan 26,0%.

## 2. Data Numerik

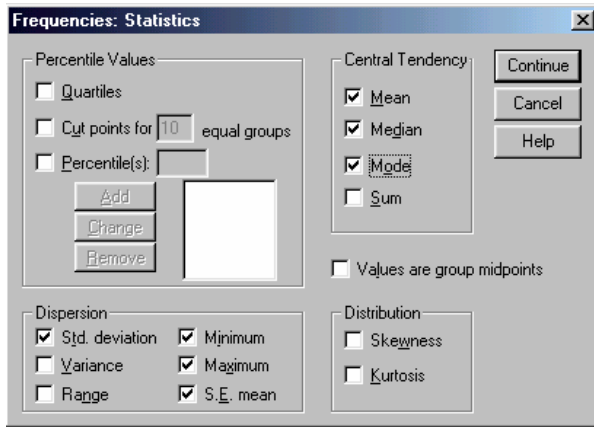
Pada data numerik, peringkasan data dapat dilakukan dengan melaporkan ukuran tengah dan sebarannya. Ukuran yang digunakan adalah rata-rata, median dan modus. Sedangkan ukuran sebarannya (variasi) yang digunakan adalah range, standard

deviasi, minimal dan maksimal. Pada SPSS ada dua cara untuk mengeluarkan analisis deskriptif yaitu dapat melalui perintah '*Frequencies*' atau perintah '*Explore*'. Biasanya digunakan adalah *Frequencies* oleh karena ukuran statistik yang dapat dihasilkan pada menu '*Frequencies*' sangat lengkap (seperti mean, median, varian dll), selain itu pada perintah ini juga dapat ditampilkan grafik histogram dan kurve normalnya. Berikut akan dicoba mengeluarkan analisis deskriptif variabel umur dengan menggunakan perintah *frequencies*.

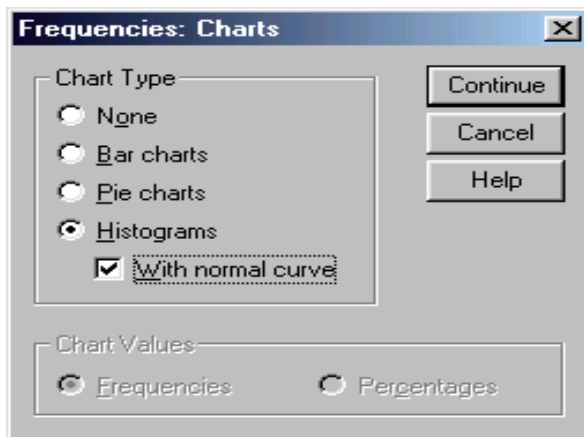
- a. Aktifkan data "susu.sav"
- b. Pilih '*Analyze*'
- c. Pilih '*Descriptive Statistic*'
- d. Pilih '*Frequencies*', terlihat kotak *frequencies*:
- e. Sorot variabel yang akan dianalisis, sorot **umur**, dan klik tanda panah sehingga umur masuk ke kotak variable (s).



- f. Klik tombol option ‘Statistics...’, pilih ukuran yang anda minta misalnya mean, median, standard sevasi, minimum, maximum, SE.



- g. Klik continue
- h. Klik tombol option ‘Charts’ lalu muncul menu baru dan klik ‘Histogram’, lalu klik ‘With Normal Curve’



- i. Klik continue
- j. Klik 'OK', dan pada layar terlihat distribusi frekuensi disertai ukuran statistic yang diminta dan dibawahnya tampak grafik histogram beserta curve normalnya.

**Frequencies**

Statistics

Umur ibu menyusui

N	Valid	50
	Missing	0

Statistics

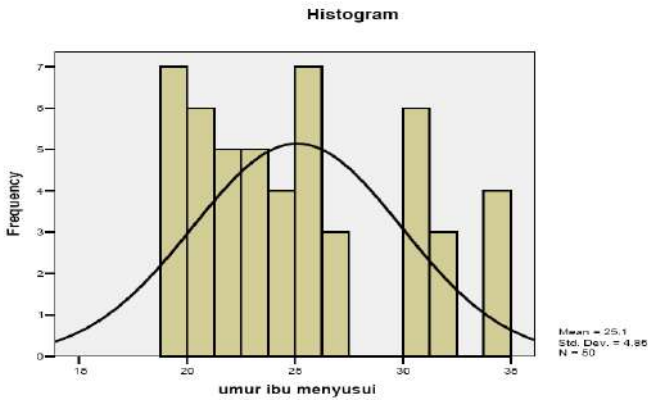
Umur ibu menyusui

N	Valid	50
	Missing	0
Mean		25.10
Std. Error of Mean		.686
Median		24.00
Mode		19
Std. Deviation		4.850
Minimum		19
Maximum		35

Umur ibu menyusui

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 19	7	14.0	14.0	14.0
20	3	6.0	6.0	20.0
21	3	6.0	6.0	26.0
22	5	10.0	10.0	36.0
23	5	10.0	10.0	46.0
24	4	8.0	8.0	54.0
25	2	4.0	4.0	58.0
26	5	10.0	10.0	68.0
27	3	6.0	6.0	74.0
30	3	6.0	6.0	80.0
31	3	6.0	6.0	86.0
32	3	6.0	6.0	92.0
34	2	4.0	4.0	96.0
35	2	4.0	4.0	100.0
Total	30	100.0	100.0	

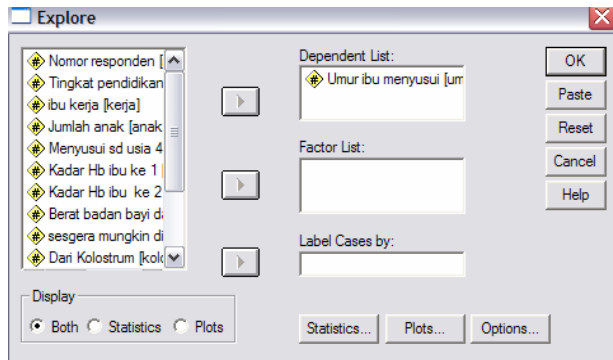




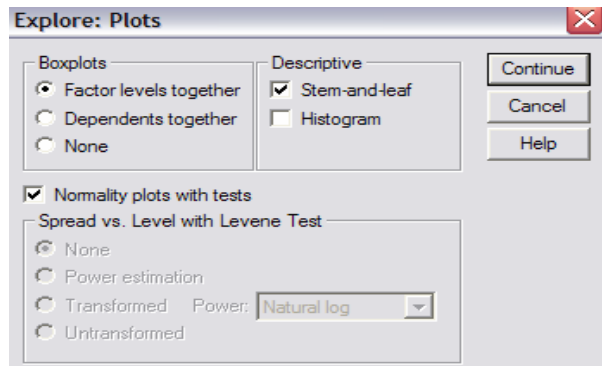
Dari hasil di atas, nilai rata-rata dapat dilihat pada baris mean, sedangkan nilai standard deviasi dapat dilihat pada baris std. Seviation. Pada contoh di atas, rata-rata umur ibu adalah 25,10 tahun, median 24,0 tahun dan standard deviasi 4,85 tahun dengan umur termuda 19 tahun dan yang tertua 35 tahun. Distribusi frekuensi ditampilkan menurut umur termuda sampai dengan umur tertua dengan informasi tentang jumlah dan persentasenya. Bentuk distribusi data dapat diketahui dari grafik histogram dan kurve normalnya. Dari tampilan grafik dapat dilihat bahwa distribusi variabel umur berbentuk normal.

Dari hasil di atas belum diperoleh informasi estimasi interval yang penting untuk melakukan estimasi parameter populasi. Bila anda ingin memperoleh estimasi interval lakukan analisis eksplorasi data dengan perintah 'Explore'. Adapun caranya sbb:

- 1) Dari menu utama SPSS, pilih menu 'Analyze', kemudian pilih submenu 'descriptive Statistics', lalu pilih 'Explore'.
- 2) Isikan kotak 'Dependent List' dengan variabel 'umur', kotak 'Factor List' dan 'Label Cases By' biarkan kosong, sehingga tampilannya sbb:



- 3) Klik tombol 'Plots', dan pilih 'Normality Plots With Test'



- 4) Klik continue

- 5) Klik OK, hasilnya dapat dilihat di layer adalah sebagai berikut:

Explore

Descriptives

		Statistic	Std. Error	
umur ibu menyusui	Mean	25.10	.686	
	95% Confidence Interval for Mean	Lower Bound	23.72	
		Upper Bound	26.48	
	5% Trimmed Mean	24.90		
	Median	24.00		
	Variance	23.520		
	Std. Deviation	4.850		
	Minimum	19		
	Maximum	35		
	Range	16		
	Interquartile Range	9		
	Skewness	.547	.337	
	Kurtosis	-.812	.662	

Tests of Normality

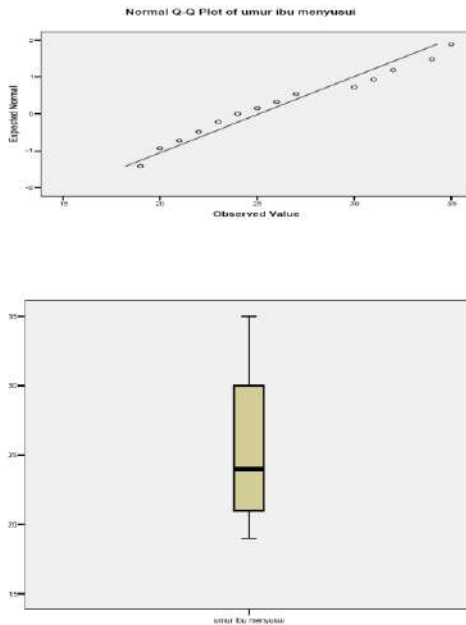
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
umur ibu menyusui	.130	50	.035	.920	50	.002

<sup>a</sup>. Lilliefors Significance Correction

umur ibu menyusui

umur ibu menyusui Stem-and-Leaf Plot

Frequency	Stem & Leaf
7.00	1 .9999999
20.00	2 .0001112222233334444
10.00	2 .5566666777
11.00	3 .00011122244
2.00	3 .55
Stem width:	10
Each leaf:	1 case(s)



Dari hasil analisis 'Explore' terlihat juga nilai mean, median dan mode. Namun yang paling penting dari tampilan explore munculnya angka estimasi interval. Dari hasil tersebut kita dapat melakukan estimasi interval dari umur ibu. Kita dapat menghitung 95% confidence interval umur yaitu 23,72 s.d. 26,48. Jadi kita 95% yakin bahwa rata-rata umur ibu di populasi berada pada selang 23,72 sampai 26,48 tahun. Untuk variabel umur diatas, dilihat dari histogram dan kurve normal terlihat bentuk yang normal, selain itu hasil dari perbandingan skwness dan standar error didapatkan:  $0,547/0,337 = 1,62$  , hasilnya masih

dibawah 2, berarti distribusi normal. Dari hasil tersebut diatas dengan demikian variabel umur disimpulkan berdistribusi normal.

### 3. Penyajian dan Interpretasi di Bagian Hasil Penelitian

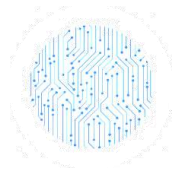
Dari angka-angka tersebut kemudian kita masukkan ke tabel penyajian di laporan penelitian/laporan tesis. Adapun penyajian dan interpretasinya adalah sbb:

Tabel 1

Distribusi Umur dan Lama Hari Rawat pasien Rumah sakit X Tahun x

Variabel	Mean	SD	Minimal-Maksimal	95% CI
Umur	25,10	4,85	19 - 35	23,72 – 26,48

Hasil analisis didapatkan rata-rata umur ibu adalah 25,10 tahun (95% CI: 23,72 – 26,48), dengan standar deviasi 4,85 tahun. Umujr termuda 19 tahun dan umur tertua 35 tahun. Dari hasil estimasi interval dapat disimpulkan bahwa 95% diyakini bahwa rata-rata umur ibu adalah diantara 23,72 sampai dengan 26,48 tahun.



## **BAB III**

### **ANALISIS BIVARIAT**

#### **A. Analisis Bivariat**

Setelah diketahui karakteristik masing-masing variabel dapat diteruskan analisis lebih lanjut. Pada analisis univariat, misalnya ada dua variabel: jenis pembayaran berobat dan kepuasan pasien, kita hanya melakukan pendeskripsian sendiri-sendiri untuk variabel jenis pembayaran dan kepuasan pasien. Untuk variabel jenis pembayaran akan diketahui berapa persen yang berobat dengan biaya sendiri dan berapa persen yang dibiayai askes. Begitu juga untuk variabel kepuasan pasien, akan diketahui berapa persen yang puas dan berapa persen yang tidak puas.

Apabila diinginkan analisis hubungan antara dua variabel, dalam contoh diatas berarti kita ingin mengetahui hubungan jenis pembayaran dengan kepuasan pasien, maka analisis dilanjutkan pada tingkat bivariat. Pada analisis bivariat kita dapat mengetahui apakah ada perbedaan kepuasan pasien antara pasien dengan membayar sendiri dengan pasien dengan biaya askes. Kegunaan analisis bivariat bisa untuk mengetahui apakah ada hubungan yang

signifikan antara dua variabel, atau bisa juga digunakan untuk mengetahui apakah ada perbedaan yang signifikan antara dua atau lebih kelompok (sampel).

Perbedaan Substansi/Klinis dan perbedaan Statistik perlu dipahami/diadari bagi peneliti bahwa berbeda bermakna/signifikan secara statistik tidak berarti (belum tentu) bahwa perbedaan tersebut juga bermakna dipandang dari segi substansi/klinis. Seperti diketahui bahwa semakin besar sampel yang dianalisis akan semakin besar menghasilkan kemungkinan berbeda bermakna. Dengan sampel besar perbedaan-perbedaan sangat kecil, yang sedikit atau bahkan tidak mempunyai manfaat secara substansi/klinis dapat berubah menjadi bermakna secara statistik. Oleh karena itu arti kegunaan dari setiap penemuan jangan hanya dilihat dari aspek statistik semata, namun harus juga dinilai/dilihat kegunaannya dari segi klinis/substansi. Sebagai contoh ada studi eksperimen yang akan menguji dua obat (katakanlah obat A dan Obat B) untuk mengetahui pengaruhnya terhadap penurunan tekanan darah. Kemudian obat A dan B diujicobakan pada dua kelompok relawan penderita hipertensi. Hasil eksperimen didapatkan bahwa rata-rata penurunan tekanan darah setelah minum obat A adalah 40 mmHg dan pada kelompok yang minum Obat B rata-rata penurunannya 39 mmHg. Kemudian dilakukan uji statistik dan hasilnya signifikan/bermakna ( $p$  value  $<$   $\alpha$ ), apa yang dapat disimpulkan dari temuan ini? Secara statistik memang terjadi perbedaan bermakna,

namun secara substansi tidaklah mempunyai perbedaan yang berarti, oleh karena perbedaan mean penurunan tekanan darah antara obat A dan B hanya 1 mmHg. Dengan hasil ini dapat disimpulkan bahwa sebenarnya antara obat A dan B tidak ada perbedaan (sama saja) kasiatnya.

## **B. Uji t Independen**

Pada prinsipnya tujuan uji 2 sampel adalah ingin mengetahui apakah ada perbedaan rata-rata (mean) antara dua populasi, dengan melihat rata-rata dua sampelnya. Sebagai contoh ada perbedaan antara rata-rata Asupan Protein orang yang Sarapan dengan yang Tidak Sarapam. Untuk menguji hal tersebut, tentu tidak bisa semua orang obesitas dan tidak diobservasi karena memakan biaya dan waktu. Alternatif terbaik adalah dengan mengambil sampel di kedua kelompok tersebut, dari data dua kelompok sampel akan ditarik kesimpulan untuk populasi.

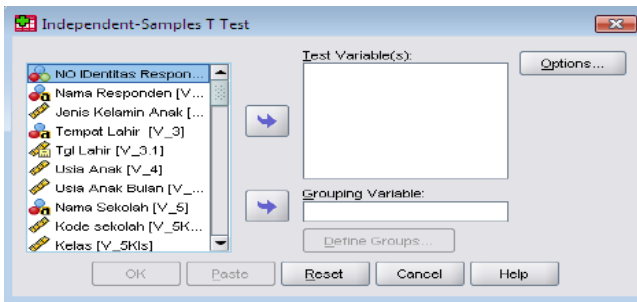
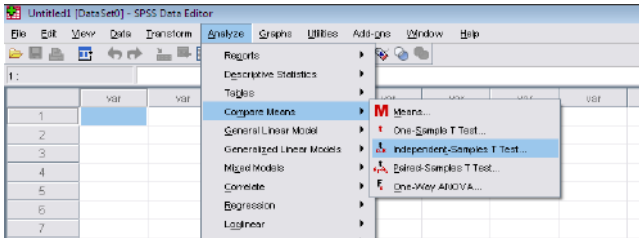
Perhatikan kata 'independen' atau 'bebas' yang berarti tidak ada hubungan antara 2 sampel yang diuji. Dalam contoh, jelas antara orang yang Obesitas dengan yang Tidak Obesitas bukanlah orang yang sama. Hal ini berbeda dengan uji t berpasangan, dimana orang atau sampel yang sama akan diuji 2 kali.

### **Langkah-Langkah Uji Dua Sampel Independen**

- 1. Klik Menu Analyse --> Compare means --> independent samples t test seperti pada layar berikut ini;**

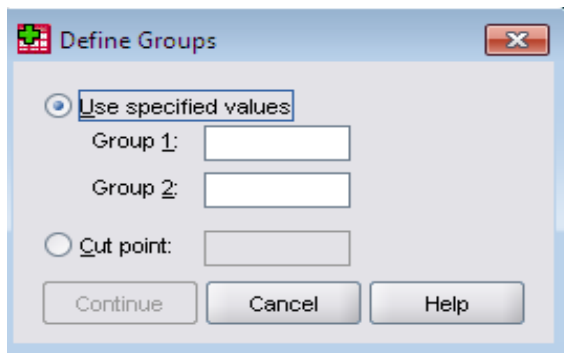


MODUL DIGITAL: Biostatistik untuk *Applied Research Method in Nutrition 2*



Entry;

- Pada Kolom Test Variable (s): masukkan variabel yang akan diuji. Misal pada contoh adalah variabel Asupan Lemak.
- Pada Kolom Grouping Variable (s): klik Define Groups, masukkan angka satu pada Group 1, dan angka 2 pada Group 2.



Tekan Continue untuk kembali ke menu sebelumnya.

Kemudian tekan OK untuk mengakhiri pengisian prosedur analisis.

## HASIL ANALISIS

### Output Bagian 1

Group Statistics

	Apakah...	N	Mean	Std. Deviation	Std. Error Mean
Asupan Protein Sehari	Ya	111	41.0896	15.62305	1.48287
	Tidak	15	35.1733	11.31322	2.92106

Pada bagian pertama terlihat ringkasan statistik dari kedua sampel. Untuk Asupan Protein, Orang yang Sarapan mempunyai asupan rata-rata 41,08 gram yang jauh di atas rata-rata Asupan Protein Orang yang Tidak Sarapan, yaitu 35,17 gram. Dari data tersebut, apakah ada perbedaan yang signifikan (jelas dan nyata) antara asupan protein orang sarapan dan tidak sarapan? 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.	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference		
								Lower	Upper	
Asupan Protein Sehari	Equal variances assumed	1.801	.183	111	.180	5.91631	4.19070	-2.35595	14.19176	
	Equal variances not assumed			1.808	21.980	.085	5.91631	3.27590	-.87621	12.71082

Uji t dua sampel dilakukan dalam dua tahapan; tahapan pertama adalah menguji apakah varians dari dua populasi bisa dianggap sama? Setelah itu baru 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 alternatif jawaban uji t yang lain.

Pertama dilakukan pengujian apakah ada kesamaan varians pada data orang yang sarapan dengan yang tidak; pengujian asumsi kesamaan varians dilakukan lewat uji F (*Levene's Test*).

Hipotesis untuk pengujian varians.

$H_0$  = 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:

Dasar Pengambilan Keputusan (uji varians menggunakan uji satu sisi):

Jika probabilitas  $> 0,05$ , maka  $H_0$  diterima.

Jika probabilitas  $< 0,05$ , maka  $H_0$  ditolak.

Keputusan:

Terlihat bahwa F hitung untuk Asupan Protein dengan Equal variance assumed (diasumsi kedua varians sama atau menggunakan pooled variance t test) adalah 1,601 dengan probabilitas 0,208. Oleh karena probabilitas  $> 0,05$ , maka  $H_0$  diterima, atau kedua varians tidak berbeda.

Perbedaan yang nyata dari kedua varians membuat penggunaan varians untuk membandingkan rata-rata

populasi dengan t test sebaiknya menggunakan dasar Equal variance assumed (diasumsi kedua varians sama). Setelah uji asumsi kesamaan varians selesai, selanjutnya dilakukan analisis dengan memakai 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).

$H_a$  = Kedua rata-rata Populasi adalah tidak identik (rata-rata populasi tinggi badan pria dan wanita adalah berbeda).

Nb: berbeda dengan asumsi sebelumnya yang menggunakan varians, sekarang dipakai mean atau rata-rata hitung.

Oleh karena tidak ada kalimat “lebih tinggi” atau “kurang tinggi”, maka dilakukan uji dua sisi.

Keputusan:

Terlihat bahwa t hitung untuk Asupan Protein dengan Equal variance assumed adalah 1,415 dengan p *value* 0,16, atau  $p > 0,05$ , sehingga tidak ada perbedaan yang signifikan ( $H_0$  diterima).

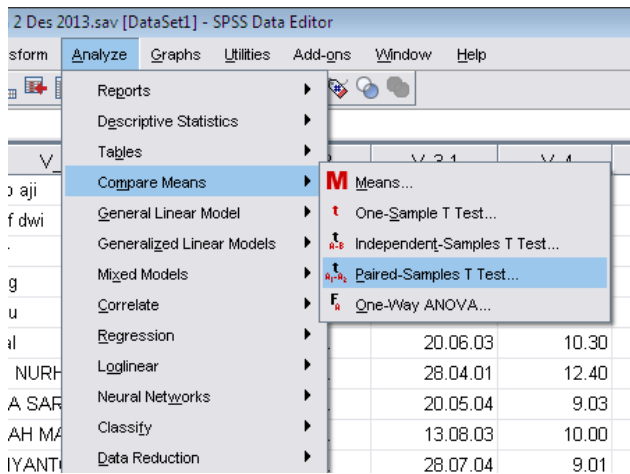
### C. Uji t Dependen

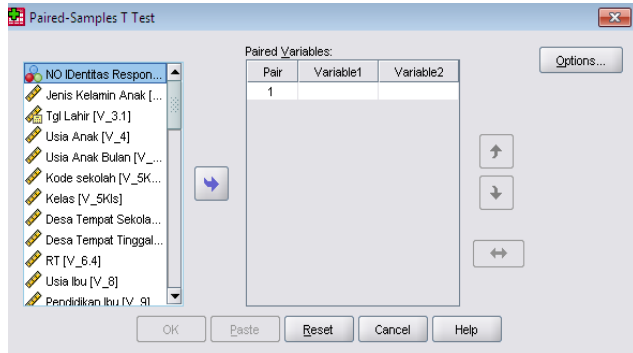
Uji ini dilakukan terhadap dua sampel yang berpasangan (paired); Sampel yang berpasangan diartikan sebagai sebuah sampel dengan subjek yang sama, namun mengalami dua perlakuan atau pengukuran yang berbeda, seperti subjek A akan mendapat perlakuan I kemudian

perlakuan II. Contoh berikut akan menjelaskan hal di atas. Anak sekolah diukur status gizinya pada bulan Mei, kemudian karena banyaknya anak sekolah yang mengalami gizi kurang, maka pemerintah memberikan bantuan pemberian makanan tambahan PMT pada bulan Juni. Untuk melihat efek dari PMT tersebut, pada bulan Agustus anak sekolah tersebut diukur lagi status gizinya. Status gizi anak sekolah sebelum dan sesudah pemberian PMT inilah yang akan dibandingkan dengan menggunakan uji sampel berpasangan.

## Langkah-Langkah Uji Dua Sampel Berpasangan

### 1. Menu Analyze --> Compare Means --> Paired Samples T test





Pengisian:

⇒ Paired Variable(s) atau Variabel yang akan diuji. Oleh karena di sini akan diuji data “sebelum” dan “sesudah”, maka klik mouse pada variabel sebelum, kemudian klik mouse sekali lagi pada variabel sesudah, maka terlihat pada kolom Current

Selection di bawah, terdapat keterangan untuk variable 1 dan 2.

Kemudian klik mouse pada tanda ‘>’ (yang sebelah atas), maka pada Paired variables terlihat tanda sebelum .. sesudah.

Nb: variabel sebelum dan sesudah harus dipilih berbarengan

Jika tidak, SPSS tidak bisa menginput dalam kolom Paired Variables, dengan tanda tidak aktifnya ikon.

## HASIL ANALISIS

**Paired Samples Statistics**

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Status Gizi sebelum	-.8739	106	1.28381	.12469
	Status Gizi sesudah	-.7764	106	.94127	.09142

Pada bagian pertama terlihat ringkasan statistik dari kedua sampel. Untuk sebelum diberi PMT, anak sekolah dasar mempunyai rata-rata Z score -0,87, dan sedangkan setelah diberi PMT responden mempunyai rata-rata Z score -0,77.

**Paired Samples Correlations**

		N	Correlation	Sig.
Pair 1	Status Gizi sebelum & Status Gizi sesudah	106	.723	.000

**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	Status Gizi sebelum - Status Gizi sesudah	-.09745	.80840	.09809	-.28616	.07326	-1.132	105	.260

Bagian kedua output adalah hasil korelasi antara kedua variabel, yang menghasilkan angka 0,723 dengan nilai probabilitas jauh di bawah 0,05 (lihat nilai signifikansi output yang 0,000). Hal ini menyatakan bahwa korelasi antara status gizi sebelum dan sesudah PMT adalah sangat erat dan benar-benar berhubungan secara nyata.

Hipotesis:

Hipotesis untuk kasus ini.

$H_0$  = Kedua rata-rata Populasi adalah identik (rata-rata status gizi sebelum dan setelah pemberian PMT adalah tidak berbeda secara nyata).

$H_1$  = Kedua rata-rata Populasi adalah tidak identik (rata-rata status gizi sebelum dan sesudah pemberian PMT adalah memang berbeda secara nyata).

Melihat nilai signifikansi yang lebih dari 0,05, sehingga pada contoh ini, kedua populasi adalah identik, artinya tidak ada perbedaan status gizi anak sekolah dasar sebelum dengan sesudah diberi PMT.

### LATIHAN UJI T.

Berikut adalah data primer untuk melihat asupan responden wanita dewasa muda di Kompleks Perumahan Tamantirto Indah.

NoID	Nama	Status Pekerjaan	Asupan Energi (kcal)	Asupan Energi setelah Penyuluhan (kcal)
(1)	(2)	(3)	(4)	(5)
1	Lalita	1	1700	1900
2	Nana	2	2300	2000
3	Amira	2	2400	1850
4	Santi	1	1750	2100
5	Lia	1	1500	1950
6	Tamara	2	2300	2000
7	Hanifa	1	1400	1800



8	Nurisa	2	2250	2000
9	Anin	1	1600	1800
10	Kaisa	1	1800	1800
11	Mimi	2	2400	2000
12	Jingga	1	1800	2150
13	Opie	2	2150	2100
14	Restia	2	2300	2050
15	Ghaita	2	2450	2050

1: Bekerja

2: Tidak bekerja

- a) Apakah terdapat perbedaan rata-rata asupan energi antara subjek yang bekerja dengan yang tidak?
- b) Kemudian di kompleks perumahan tersebut diadakan penyuluhan terkait gizi seimbang karyawan. Apakah terdapat perbedaan rata-rata asupan energi antara subyek sebelum diberi penyuluhan dengan setelah diberi penyuluhan?

#### D. Uji ANOVA

Uji Anova adalah uji statistik inferensial parametrik yang digunakan oleh peneliti untuk membandingkan dua atau lebih *mean* dari kelompok. Laporan hasil meliputi skor F dan tingkat probabilitas. Anova dan Uji F memiliki tujuan yang sama yakni menguji rata-rata populasi, tetapi pada Anova, diuji lebih dari dua rata-rata populasi, sedangkan uji F bertujuan untuk menguji sama atau tidaknya varians. Pengujian rata-rata pada lebih dari dua kelompok sampel digunakan One-Way Anova. Pada uji Anova me-

miliki syarat yang harus dipenuhi yakni normalitas data antar kelompok dan homogenitas varian. Uji normalitas data antarkelompok yang sering digunakan adalah uji Shapiro Wilk, sedangkan pada uji homogenitas varian One-Way Anova digunakan adalah uji Barlett's. Jika pada uji Anova hasilnya menolak  $H_0$  (nilai  $p$  value  $< 0,05$ ), maka kita lanjutkan ke uji perbandingan berganda. Ada beberapa uji berganda salah satunya adalah uji Bonferroni.

Asumsi yang digunakan pada uji Anova antara lain:

- Populasi yang akan diuji memiliki distribusi normal
- Varians dari populasi tersebut sama
- Sampel tidak berhubungan satu sama lain
- Jenis data yang dihubungkan adalah numerik dengan kategori (untuk kategori yang lebih dari 2 kelompok)

Rumus Anova sebagai berikut:

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

$df = k - 1 \rightarrow$  untuk pembilang

$N - k \rightarrow$  untuk penyebut

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

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

$$\bar{X} = \frac{n1 \cdot X1 + n2 \cdot X2 + \dots + nk \cdot Xk}{N}$$

Keterangan : N = jumlah seluruh data

$$(n1 + n2 + \dots + nk)$$

**Kriteria Pengujian:**

- Ho diterima, jika nilai F hitung  $\leq$  F table
- Ho ditolak, jika nilai F hitung  $>$  F table

**Contoh kasus**

Sebuah penelitian ingin mengetahui perbedaan skor kognitif siswa kelas 6 SD pada tiga kelompok yang mendapatkan suplementasi zat besi yang berbeda. Data yang dikumpulkan sebagai berikut:

Fe Folat (kelompok 1)	:	16	22	18	20	19
Fe Folat + Zink (kelompok 2)	:	26	22	30	24	27
Fe Folat + Vitamin A (kelompok 3)	:	20	23	18	16	20

Apakah ada perbedaan skor kognitif pada ketiga jenis suplementasi zat besi tersebut dengan alpha 5%?

Jawab:

Hipotesis:

Ho:  $\mu1 = \mu2 = \mu3$

Tidak ada perbedaan mean skor kognitif pada ketiga jenis suplementasi zat besi.

$H_a: \mu_1 \neq \mu_2 = \mu_3$ , atau

$\mu_1 = \mu_2 \neq \mu_3$ , atau

$\mu_1 \neq \mu_2 \neq \mu_3$

Ada perbedaan mean skor kognitif pada ketiga jenis suplementasi zat besi.

### Perhitungan Uji Anova (Uji F)

Fe Folat (kelompok 1)	:	mean = 19,0	standar deviasi = 2,23
Fe Folat + Zink (kelompok 2)	:	mean = 25,8	standar deviasi = 3,03
Fe Folat + Vitamin A (kelompok 3)	:	mean = 19,4	standar deviasi = 2,61

$$\bar{X} = \frac{(5)(19) + (5)(25,8) + (5)(19,4)}{15} = 21,4$$

$$s_b^2 = \frac{(5)(19 - 21,4)^2 + (5)(25,8 - 21,4)^2 + (5)(19,4 - 21,4)^2}{3 - 1} = 72,8 \text{ (73)}$$

$$s_w^2 = \frac{(5 - 1)(2,23)^2 + (5 - 1)(3,03)^2 + (5 - 1)(2,61)^2}{15 - 3} = 7$$

$$F = \frac{73}{7} = 10,4$$

df pembilang =  $3 - 1 = 2$

df penyebut =  $15 - 3 = 12$

Nilai kritis F pada table F adalah = 3,89

Kesimpulan: Karena nilai F hitung (10,4) lebih besar dari nilai F tabel (3,89) dapat disimpulkan bahwa ada

perbedaan mean skor kognitif pada ketiga jenis suplementasi zat besi.

Cuplikan Tabel F

Titik Persentase Distribusi F untuk Probabilita = 0,05															
df untuk penyebut (N2)	df untuk pembilang (N1)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161	199	216	225	230	234	237	239	241	242	243	244	245	245	246
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86
5	6.01	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85
11	4.84	3.98	3.59	3.36	3.21	3.10	3.02	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46

**Analisis Multiple Comparison (Posthoc Test)**

Analisis ini bertujuan untuk mengetahui lebih lanjut kelompok mana saja yang berbeda meannya bilamana pada pengujian Anova dihasilkan ada perbedaan yang bermakna (Ho ditolak). Salah satu uji analisis multiple comparison adalah uji Bonferroni.

Rumus perhitungan Bonferroni sebagai berikut.

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{VSw^2 \left[ \left( \frac{1}{n_i} \right) + \left( \frac{1}{n_j} \right) \right]}}$$

Dengan level of significance (α) sebagai berikut.

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

Kriteria Pengujian:

- $H_0$  ditolak, jika  $P \text{ value} < \alpha$

**Contoh kasus**

Berdasarkan soal di atas kita akan telusuri lebih lanjut kelompok mana saja yang skor kognitifnya berbeda dengan menggunakan alpha 5%.

$$(3_2) = \frac{3!}{2!(3-2)!} = 3$$

$$\alpha^* = \frac{0,05}{3} = 0,0167$$

Uji kelompok 1 dan 2:

$$t_{12} = \frac{19 - 25,8}{\sqrt{7 \left[ \left( \frac{1}{5} \right) + \left( \frac{1}{5} \right) \right]}} = -4,06$$

Langkah selanjutnya mencari nilai t tabel dengan menggunakan table t dengan df (n-k) atau (15-3) yaitu 12.

df	Pr	0.25	0.10	0.05	0.025	0.01	0.005	0.001
		0.50	0.20	0.10	0.050	0.02	0.010	0.002
1		1.00000	3.07768	6.31375	12.70620	31.82052	63.65674	318.30884
2		0.81650	1.88562	2.91999	4.30265	6.96456	9.92484	22.32712
3		0.76489	1.63774	2.35336	3.18245	4.54070	5.84091	10.21453
4		0.74070	1.53321	2.13185	2.77645	3.74695	4.60409	7.17318
5		0.72669	1.47588	2.01505	2.57058	3.36493	4.03214	5.89343
6		0.71756	1.43976	1.94318	2.44691	3.14267	3.70743	5.20763
7		0.71114	1.41492	1.89458	2.36462	2.99795	3.49948	4.78529
8		0.70639	1.39682	1.85955	2.30600	2.89646	3.35539	4.50079
9		0.70272	1.38303	1.83311	2.26216	2.82144	3.24984	4.29681
10		0.69981	1.37218	1.81246	2.22814	2.76377	3.16927	4.14370
11		0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.02470
12		0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	3.92963
13		0.69383	1.35017	1.77093	2.16037	2.65031	3.01228	3.85198
14		0.69242	1.34503	1.76131	2.14479	2.62449	2.97684	3.78739

**Kesimpulan:**

Dengan nilai  $t = -4,06$  dan  $df = 12$ , maka nilai  $p < 0,001$  karena nilai  $4,06$  diasumsikan berada pada sebelah kanan nilai  $3,92$ . Maka  $p < 0,001$  lebih kecil dari nilai  $\alpha^* = 0,0167$ , dengan demikian  $H_0$  ditolak. Secara statistik ada perbedaan rata-rata skor kognitif antara kelompok 1 (Fe Folat) dan kelompok 2 (Fe Folat dan Zink).

Uji kelompok 1 dan 3:

$$t_{13} = \frac{19 - 19,4}{\sqrt{7 \left[ \left(\frac{1}{5}\right) + \left(\frac{1}{5}\right) \right]}} = -0,23$$

Cari nilai  $t$  tersebut pada table distribusi  $t$  dengan  $df = N-k \rightarrow df = 15-3 = 12$

Pr	0.25	0.10	0.05	0.025	0.01	0.005	0.001
df	0.50	0.20	0.10	0.050	0.02	0.010	0.002
1	1.00000	3.07768	6.31375	12.70620	31.82052	63.65674	318.30884
2	0.81650	1.88562	2.91999	4.30265	6.96456	9.92484	22.32712
3	0.76489	1.63774	2.35336	3.18245	4.54070	5.84091	10.21453
4	0.74070	1.53321	2.13185	2.77645	3.74695	4.60409	7.17318
5	0.72669	1.47588	2.01505	2.57058	3.36493	4.03214	5.89343
6	0.71756	1.43976	1.94318	2.44691	3.14267	3.70743	5.20763
7	0.71114	1.41492	1.89458	2.36462	2.99795	3.49948	4.78529
8	0.70639	1.39682	1.85955	2.30600	2.89646	3.35539	4.50079
9	0.70272	1.38303	1.83311	2.26216	2.82144	3.24964	4.29681
10	0.69981	1.37218	1.81246	2.22814	2.76377	3.16927	4.14370
11	0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.02470
12	0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	3.92963
13	0.69383	1.35017	1.77093	2.16037	2.65031	3.01228	3.85198
14	0.69242	1.34503	1.76131	2.14479	2.62449	2.97684	3.78739

**Kesimpulan:**

Dengan nilai  $t = -0,23$  dan  $df = 12$ , maka nilai  $p > 0,25$  karena nilai  $0,23$  diasumsikan berada pada sebelah kiri nilai  $0,69$ . Maka  $p > 0,25$  lebih besar dari nilai  $\alpha^* =$

0,0167, dengan demikian  $H_0$  gagal ditolak. Secara statistik tidak ada perbedaan rata-rata skor kognitif antar kelompok 1 (Fe Folat) dan kelompok 3 (Fe Folat dan Vitamin A).

Uji kelompok 2 dan 3:

$$t_{23} = \frac{25,8 - 19,4}{\sqrt{7 \left[ \left( \frac{1}{5} \right) + \left( \frac{1}{5} \right) \right]}} = 3,82$$

df	Pr	0.25 0.50	0.10 0.20	0.05 0.10	0.025 0.050	0.01 0.02	0.005 0.010	0.001 0.002
1		1.00000	3.07768	6.31375	12.70620	31.82052	63.65674	318.30884
2		0.81650	1.88562	2.91999	4.30265	6.96456	9.92484	22.32712
3		0.76489	1.63774	2.35336	3.18245	4.54070	5.84091	10.21453
4		0.74070	1.53321	2.13185	2.77645	3.74695	4.60409	7.17318
5		0.72669	1.47588	2.01505	2.57058	3.36493	4.03214	5.89343
6		0.71756	1.43976	1.94318	2.44691	3.14267	3.70743	5.20763
7		0.71114	1.41492	1.89458	2.36462	2.99795	3.49948	4.78529
8		0.70639	1.39682	1.85955	2.30600	2.89646	3.35539	4.50079
9		0.70272	1.38303	1.83311	2.26216	2.82144	3.24984	4.29681
10		0.69981	1.37218	1.81246	2.22814	2.76377	3.16927	4.14370
11		0.69745	1.36343	1.79588	2.20099	2.71808	3.10581	4.02470
12		0.69548	1.35622	1.78229	2.17881	2.68100	3.05454	3.92963
13		0.69383	1.35017	1.77093	2.16037	2.65031	3.01228	3.85198
14		0.69242	1.34503	1.76131	2.14479	2.62449	2.97684	3.78739

Cari nilai t tersebut pada table distribusi t dengan  $df = N - k$   
 $\rightarrow df = 15 - 3 = 12$

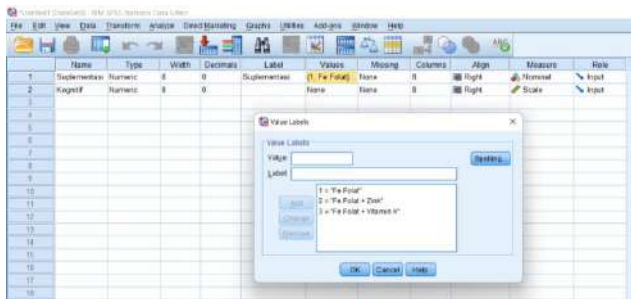
### Kesimpulan:

Dengan nilai  $t = 3,82$  dan  $df = 12$ , maka nilai  $p$  terletak diantara  $3,054 - 3,929$  (antara  $0,005 - 0,001$ ) sehingga nilai  $0,001 < p < 0,005$  lebih kecil dari nilai  $\alpha^* = 0,0167$ , dengan demikian  $H_0$  ditolak. Secara statistik ada perbedaan rata-rata skor kognitif antar kelompok 2 (Fe Folat dan Zink) dan 3 (Fe Folat dan Vitamin A).



### Langkah-Langkah Uji Anova Dengan Spss

1. Lakukan uji normalitas terhadap variable numerik dan varians. Kita asumsikan data berdistribusi normal dan varians data sama.
2. Buka **Tab Variabel View**, buat 2 variabel: Suplementasi dan Kognitif
3. Ubah Type Suplementasi ke “Numeric”, Decimals “0”, beri label “Supelementasi”, ubah measure menjadi “Nominal” dan isi value dengan kategori; 1 = Fe Folat, 2 = Fe Folat dan Zink, 3 = Fe Folat dan Vitamin A.



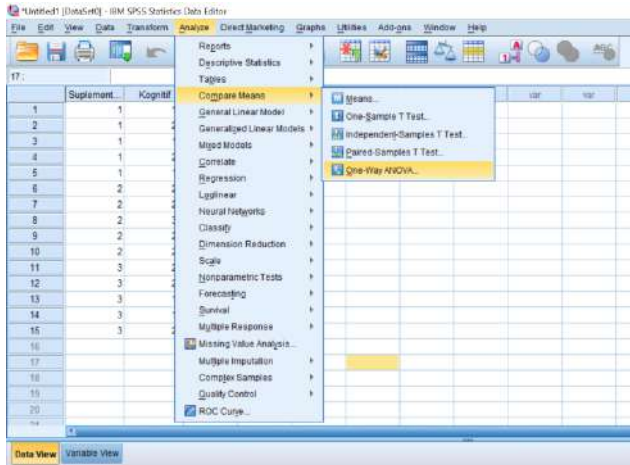
4. Buka Data View dan isikan data sebanyak 15 responden sebagai berikut;

A screenshot of the SPSS Data View showing 15 rows of data. The first two columns are 'Suplement' and 'Kognitif'. The data is as follows:

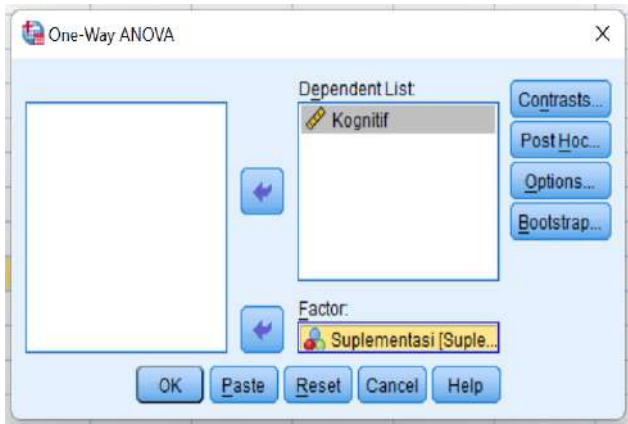
	Suplement	Kognitif
1	1	15
2	1	22
3	1	18
4	1	20
5	1	19
6	2	26
7	2	22
8	2	30
9	2	24
10	2	27
11	3	20
12	3	23
13	3	18
14	3	16
15	3	20

5. Lakukan uji anova

*Analyze* → *Compare Means* → *One Way Anova*



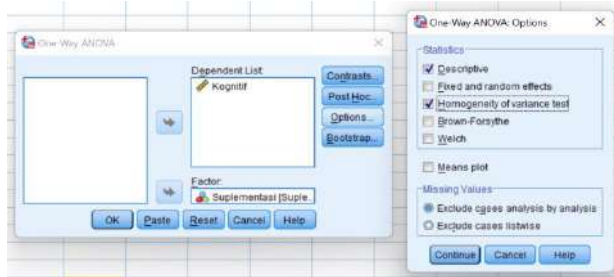
6. Kotak *Dependent List* diisi variable numerik Kognitif dan kotak *Factor* diisi variable kategori Suplementasi.



7. Klik **Post Hoc** → **Bonferroni** → **Continue**



8. Klik **Options** → **Descriptive** dan **Homogeneity of variance test** → **Continue** → **OK**



9. Output SPSS

**Descriptives**

Kognitif									
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	
Fe Folat	5	19.00	2.236	1.000	16.22	21.78	16	22	
Fe Folat + Zink	5	25.00	3.033	1.356	22.03	29.57	22	30	
Fe Folat + Vitamin A	5	19.40	2.608	1.166	16.16	22.64	16	23	
Total	15	21.40	4.050	1.046	19.16	23.64	16	30	

**Test of Homogeneity of Variances**

Kognitif			
Levene Statistic	df1	df2	Sig.
.222	2	12	.804

**ANOVA**

Kognitif					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	145.800	2	72.900	10.400	.002
Within Groups	84.000	12	7.000		
Total	229.800	14			

**Post Hoc Tests**

**Multiple Comparisons**

Dependent Variable: Kognitif  
Bonferroni

(i) Suplementasi	(j) Suplementasi	Mean Difference (i-j)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Fe Folat	Fe Folat + Zink	-6.800 <sup>*</sup>	1.673	.005	-11.45	-2.15
	Fe Folat + Vitamin A	-.400	1.673	1.000	-5.05	4.25
Fe Folat + Zink	Fe Folat	6.800 <sup>*</sup>	1.673	.005	2.15	11.45
	Fe Folat + Vitamin A	6.400 <sup>*</sup>	1.673	.007	1.75	11.05
Fe Folat + Vitamin A	Fe Folat	.400	1.673	1.000	-4.25	5.05
	Fe Folat + Zink	-6.400 <sup>*</sup>	1.673	.007	-11.05	-1.75

\*. The mean difference is significant at the 0.05 level.

**10. Interpretasi**

- Nilai mean dan standar deviasi dari setiap kelompok. Rata-rata skor kognitif anak kelas 6 SD dengan Suplementasi Fe Folat adalah 19 poin dengan standar deviasi 2,23 poin, pada anak dengan Suplementasi Fe Folat dan Zink adalah 25,8 poin dengan standar deviasi 3,03 poin, dan Suplementasi Fe Folat dan Vitamin A adalah 19,4 poin dengan standar deviasi 2,6 poin.
- Perlu diingat bahwa salah satu asumsi uji Anova adalah variansnya sama. Berdasarkan *Test of Homogeneity of Variances* menunjukkan bahwa varians ketiga kelompok tersebut sama (p value = 0,804), sehingga uji Anova valid untuk menguji hubungan antar kelompok.
- Dari table Anova diperoleh nilai Sig = 0,002 < alpha 0,05 sehingga menolak Ho, dapat disimpulkan bahwa ada perbedaan mean skor

kognitif pada ketiga jenis suplementasi zat besi.

- Dari table *Post Hoc Test* di atas memperlihatkan bahwa kelompok yang menunjukkan adanya perbedaan mean skor kognitif ditandai dengan tanda bintang “\*”. Dengan kesimpulan sebagai berikut:
  - ✓ Mean Fe Folat (19) = mean Fe Folat + Vitamin A(19,4)  $\neq$  mean Fe Folat + Zink (25,8).
  - ✓ P value (Fe Folat – Fe Folat dan Vitamin A) = 1.000, Ho diterima, kesimpulan kedua mean sama atau tidak ada perbedaan rata-rata skor kognitif antar kelompok 1 (Fe Folat) dan kelompok 3 (Fe Folat dan Vitamin A).
  - ✓ P value (Fe Folat dan Vitamin A – Fe Folat dan Zink)= 0.007, Ho ditolak, kesimpulan kedua mean berbeda atau ada perbedaan rata-rata skor kognitif antar kelompok 2 (Fe Folat dan Zink) dan 3 (Fe Folat dan Vitamin A).
  - ✓ Dari hasil di atas mean skor kognitif pada pemberian Fe Folat dan Fe Folat plus Vitamin A adalah sama, sedangkan Fe Folat plus Zink mempunyai rata-rata yang berbeda (lebih besar) di antara

ketiganya. Jadi dapat ditarik kesimpulan: *Pemberian Fe Folat+Zink lebih optimal untuk meningkatkan kognitif pada siswa dari pada Fe Folat atau Fe Folat+Vitamin A.*

### SOAL LATIHAN

Suatu penelitian dilakukan untuk mengetahui hubungan kondisi ekonomi sosial ekonomi keluarga dengan berat badan bayi yang dilahirkan. Penelitian dilakukan dengan menimbang berat badan bayi (kg) pada 30 ibu yang baru melahirkan yang terbagi dalam kelompok sosial ekonomi rendah, sedang, dan tinggi. Adapun hasilnya adalah sebagai berikut;

No	Sosek rendah	Sosek sedang	Sosek tinggi
1	2.4	3.0	3.1
2	3.0	3.1	2.4
3	2.1	2.7	3.5
4	3.0	2.6	2.9
5	3.4	3.1	3.4
6	2.3	2.9	4.0
7	2.3	2.9	3.4
8	2.3	2.6	3.9
9	2.5	2.6	3.7
10	2.5	2.6	3.7

Pertanyaan:

1. Ujilah dengan alpha 5% apakah ada perbedaan berat badan bayi antara tingkat sosial ekonomi tersebut!
2. Bila ada perbedaan, kelompok mana saja yang berbeda?

### E. Uji Chi Square

Seringkali dalam suatu penelitian, kita menemui data yang tidak dapat dinyatakan dalam bentuk angka-angka pengukuran (data numerik). Sebaliknya justru yang kita jumpai adalah data hasil dari menghitung jumlah pengamatan yang diklasifikasikan atas beberapa katagori. Data seperti ini disebut data katagorik (kualitatif), misalnya jenis kelamin yang mempunyai katagori: laki-laki dan perempuan; status merokok yang mempunyai katagori; perokok berat, perokok ringan dan tidak merokok. Dalam penelitian kesehatan seringkali peneliti perlu melakukan analisis hubungan variabel katagorik dengan variabel katagorik. Analisis ini bertujuan untuk menguji perbedaan proporsi dua atau lebih kelompok sampel. Uji statistik yang digunakan untuk menjawab kasus tersebut adalah **UJI KAI KUADRAT (CHI SQUARE)**.

Misalnya ingin diketahui hubungan jenis pekerjaan dengan perilaku menyusui ibu, apakah ada perbedaan proporsi kejadian menyusui eksklusif antara ibu yang bekerja dengan ibu yang tidak bekerja. Dari contoh terlihat

bahwa variabel jenis pekerjaan (bekerja/tidak bekerja) merupakan variabel katagorik, dan variabel perilaku menyusui (eksklusif/non eksklusif) juga merupakan variabel katagorik.

Sebelum berlanjut lebih dalam tentang kai kuadrat terlebih dahulu kita pahami dengan benar apa itu variabel katagorik. Suatu variabel disebut katagorik bila isi variabel tersebut terbentuk dari hasil klasifikasi/penggolongan, misalnya variabel sex, jenis pekerjaan, golongan darah, pendidikan. Di lain pihak variabel numerik (misalnya berat badan, umur dll) dapat masuk/dapat menjadi variabel katagorik bila variabel tersebut sudah mengalami pengelompokan. Misalkan kita ambil satu contoh variabel berat badan, berat badan bila nilainya masih riil (50 kg, 63 kg dst) maka masih termasuk variabel numerik, namun bila sudah dilakukan pengelompokan menjadi (<50 kg (kurus), 50-60 kg (sedang) dan > 60 (gemuk) maka variabel tersebut sudah berjenis katagorik.

#### 1. Tujuan Uji kai Kuadrat

Tujuan dari digunakannya uji kai kuadrat adalah untuk untuk menguji perbedaan proporsi/persentase antara beberapa kelompok data. Dilihat dari segi data-nya uji kai kuadrat dapat digunakan untuk mengetahui hubungan antara variabel katagorik dengan variabel katagorik. Contoh pertanyaan penelitian untuk kasus yang dapat dipecahkan oleh uji kai kuadrat misalnya;



- a) Apakah ada perbedaan kejadian hipertensi antara wanita dan pria. Kasus ini berarti akan menguji hubungan variabel hipertensi (kategori dengan klasifikasi ya dan tidak) dengan variabel jenis kelamin (kategori dengan klasifikasi wanita dan pria)
  - b) Apakah ada perbedaan kejadian anemia antara ibu yang kondisinya tinggi, sedang dan rendah. Pada kasus ini akan menguji hubungan variabel anemia kategori dengan klasifikasi ya dan tidak) dengan variabel Sosek (kategori dengan klasifikasi rendah, sedang dan tinggi).
2. Prinsip dasar Uji Kai Kuadrat

Proses pengujian kai kuadrat adalah membandingkan frekuensi yang terjadi (observasi) dengan frekuensi harapan (ekspektasi). Bila nilai frekuensi observasi dengan nilai frekuensi harapan sama, maka dikatakan tidak ada perbedaan yang bermakna (signifikan). Sebaliknya, bila nilai frekuensi observasi dan nilai frekuensi harapan berbeda, maka dikatakan ada perbedaan yang bermakna (signifikan). Pembuktian dengan uji kai kuadrat dengan menggunakan formula:

$$X^2 = \sum \frac{(O - E)^2}{E}$$

$$df = (k-1)(n-1)$$

ket :

O = nilai observasi

E = nilai ekspektasi (harapan)

k = jumlah kolom

b = jumlah baris

Untuk mempermudah analisis kai kuadrat, nilai data kedua variabel disajikan dalam bentuk tabel silang:

Variabel 1	Variabel 2		Jumlah
	Tinggi	Rendah	
Ya	a	b	a+b
Tidak	c	d	c+d
Jumlah	a+c	b+d	n

a, b, c, d merupakan nilai observasi, sedangkan nilai ekspektasi (harapan) masing-masing sel dicari dengan rumus:

$$E = \frac{\text{Total barisnya} \times \text{total kolomnya}}{\text{Jumlah keseluruhan data}}$$

Misalkan untuk mencari nilai ekspektasi (E) untuk sel a adalah:

$$E_a = (a+b) \times (a+c) / n$$

Untuk  $E_b$ ,  $E_c$  dan  $E_d$  dapat dicari dengan cara yang sama. Khususnya untuk tabel 2x2, dapat mencari nilai  $X^2$  dengan menggunakan rumus:

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

Uji kai kuadrat sangat baik untuk tabel dengan derajat kebebasan (df) yang besar. Sedangkan khusus untuk tabel 2 x 2 (df-nya adalah 1) sebaiknya digunakan uji kai kuadrat yang sudah dikoreksi (*Yate Corrected* atau *Yate's Correction*). Formula kai kuadrat *Yate's Correction* adalah sbb:

$$X^2 = \frac{(|O - E| - 0,5)^2}{E}$$

Atau

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

### 3. Keterbatasan Kai Kuadrat

Seperti kita ketahui, uji kai kuadrat menuntut frekuensi harapan/ekspektasi (E) dalam masing-masing sel tidak boleh terlampau kecil. Jika frekuensi sangat kecil, penggunaan uji ini mungkin kurang tepat. Oleh karena itu dalam penggunaan kai kuadrat harus memperhatikan keterbatasan-keterbatasan uji ini. Adapun keterbatasan uji kai kuadrat adalah sbb;

- a) Tidak boleh ada sel yang mempunyai nilai harapan (nilai E) kurang dari 1.
- b) Tidak boleh ada sel yang mempunyai nilai harapan (nilai E) kurang dari 5, lebih dari 20% dari jumlah sel.

Jika keterbatasan tersebut terjadi pada saat uji kai kuadrat, peneliti harus menggabungkan katagori-katagori yang berdekatan dalam rangka memperbesar frekuensi harapan dari sel-sel tersebut (penggabungan ini dapat dilakukan untuk analisis tabel silang lebih dari  $2 \times 2$ , misalnya  $3 \times 2$ ,  $3 \times 4$  dsb). Penggabungan ini tentunya diharapkan tidak sampai membuat datanya kehilangan makna. Andai saja keterbatasan tersebut terjadi pada tabel  $2 \times 2$  (ini berarti tidak bisa menggabung katagori-katagorinya lagi), maka dianjurkan menggunakan uji **Fisher's Exact**.

### **ODDS RATIO (OR) dan RISIKO RELATIF (RR)**

Hasil uji Chi Square hanya dapat menyimpulkan ada tidaknya perbedaan proporsi antar kelompok atau dengan kata lain kita hanya dapat menyimpulkan ada/tidaknya hubungan du variabel katagorik. Dengan demikian uji Chi Square tidak dapat menjelaskan derajat hubungan, dalam hal ini uji Chi Square tidak dapat mengetahui kelompok mana yang memiliki risiko lebih besar disbanding kelompok lain.

Dalam bidang kesehatan untuk mengetahui derajat hubungan, dikenal ukuran Risiko Relatif (RR) dan Odds Rasio (OR). Risiko relatif membandingkan risiko pada kelompok terekspose dengan kelompok tidak terekspose. Sedangkan Odds Rasio membandingkan Odds pada kelompok ter-ekspose dengan Odds kelompok tidak ter-eksp[ose. Ukuuran RR pada umumnya digunakan pada disain Kohort, sedangkan ukuran OR biasanya digunakan pada desain kasus kontrol atau ptong lintang (*Cross Sectional*).

***Pengkodean Variabel:***

Perlu diketahui bahwa dalam mengeluarkan nilai OR dan RR harus hati hati jangan sampai terjadi kesalahan pengkodean. Pemberian kode harus ada konsistensi antara variabel independen dengan variabel dependen. Untuk variabel independen, kelompok yang berisiko/expose diberi kode tinggi (kode 1) dan kode rendah (kode 0) untuk kelompok yang tidak berisiko/non expose. Pada variabel dependennya, kode tinggi (kode 1) untuk kelompok kasus atau kelompok yang menjadi fokus pembahasan penelitian dan kode rendah (kode 0) untuk kelompok non kasus atau yang bukan menjadi fokus penelitian. Sebagai contoh data di atas pengkodeannya adalah sbb: Ibu tidak bekerja diberi kode 1 dan bekerja kode 0 dan ibu yang menyusui secara eksklusif diberi kode 1 dan non

eksklusif diberi kode 0. Sebetulnya bisa juga kodenya dibalik, tapi harus konsisten, misalnya kodenya: tidak bekerja =0, bekerja =1 dan eksklusif =0, tdk eksklusif =1.

Tabel ...  
Distribusi Responden menurut Tingkat Pendidikan dan Pengetahuan

Pendidikan	Pengetahuan				Total	
	Rendah		Tinggi			
	N	%	n	%	n	%
SD	25	50,0	25	50,0	50	34,4
SMP	16	40,0	24	60,0	40	27,6
SMU	10	33,3	20	66,7	30	20,7
PT	5	20,0	20	80,0	25	17,3
Jumlah	56	38,7	89	61,3	145	100,0

Pembuatan persentase pada analisis tabel silang harus diperhatikan agar tidak salah dalam menginterpretasi. Pada jenis penelitian survei/Cross sectional atau Kohort, pembuatan persentasenya berdasarkan nilai variabel independen. Contoh di atas jenis penelitiannya Cross Sectional, variabel pendidikan sebagai variabel independen dan pengetahuan sebagai variabel dependen. Dapat dilihat di tabel persentasenya berdasarkan masing-masing kelompok tingkat pendidikan (persentase baris). Contoh di atas dapat diinterpretasikan sbb: Dari 50 pasien yang berpendidikan SD, ada sebanyak 25 (50,0%) pasien mempunyai pengetahuan tinggi. Dari 40 pasien yang berpendidikan SMP, ada sebanyak 24 (60,0%) yang berpengetahuan tinggi. Dari 30 pasien yang berpendi-

dikan SMU ada sebanyak 20 (66,7%) yang berpengetahuan tinggi. Dan dari 25 pasien yang berpendidikan PT, ada sebanyak 20 (80,0%) yang berpengetahuan tinggi. Dari data ini terlihat ada kecenderungan bahwa semakin tinggi tingkat pendidikan akan semakin tinggi tingkat pengetahuannya.

Pada penelitian yang berjenis kasus kontrol (*Case Control*) pembuatan persentasenya berdasarkan variabel dependennya, misalkan terlihat pada tabel berikut:

Tabel ...  
Distribusi Responden Menurut Kasus kanker paru dan Jenis Kelamin

Jenis Kelamin	Kanker Paru				Total	
	Kasus		Kontrol			
	N	%	n	%	n	%
Laki-laki	75	75,0	30	30,0	105	52,5
Perempuan	25	25,0	70	70,0	95	47,5
Jumlah	100	50,0	100	50,0	200	100,0

Interpretasinya:

Dari mereka yang menderita kanker paru, ada sebanyak 75 (75%) responden berjenis kelamin laki-laki. Sedangkan pada kelompok yang tidak menderita kanker paru, ada sebanyak (30%) responden yang berjenis kelamin laki-laki.

**KASUS:**

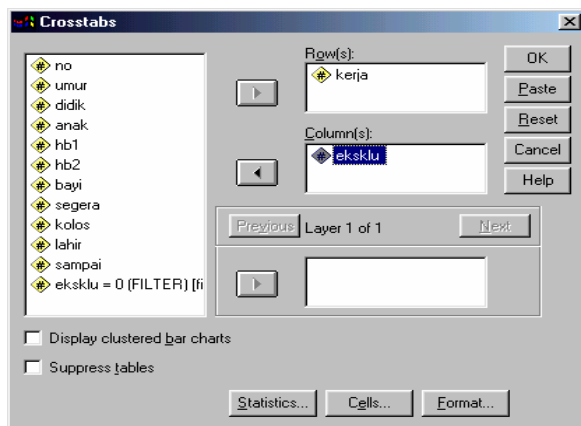
**UJI KAI KUADRAT**

Suatu penelitian ingin mengetahui hubungan pekerjaan dengan perilaku menyusui. Variabel pekerjaan

berisi dua nilai yaitu tidak bekerja dan bekerja, dan variabel menyusui berisi dua nilai yaitu eksklusif dan non eksklusif. Untuk mengerjakan soal ini gunakan data “Susu. SAV”.

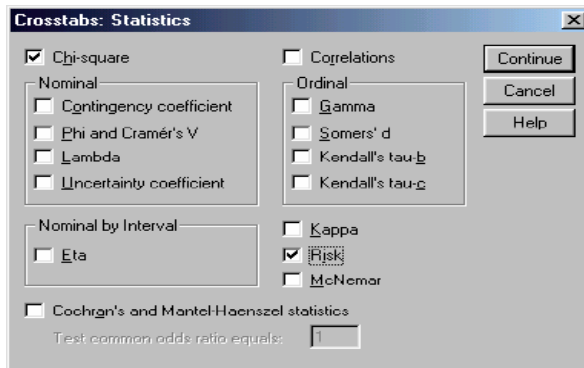
Adapun prosedur di SPSS sbb:

- a) Pastikan anda berada pada data editor ASI.SAV
- b) Dari menu SPSS, klik “*Analyze*”, kemudian pilih “*Descriptive statistic*”, lalu pilih “*Crosstab*”, sesaat akan muncul menu *Crosstabs*
- c) Dari menu *crosstab*, ada dua kotak yang harus diisi, pada kotak “*Row(s)*” diisi variabel independen (variabel bebas), dalam contoh ini variabel pekerjaan masuk ke kotak “*Row(s)*”.
- d) Pada kotak “*Column(s)*” diisi variabel dependennya, dalam contoh ini variabel perilaku menyusui masuk ke kotak “*Column(s)*”.

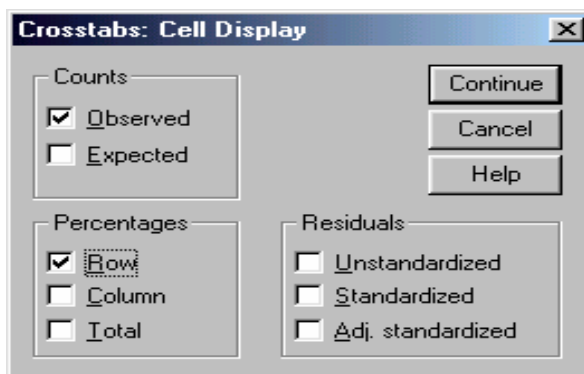




- e) Klik option “*Statistics..*”, klik pilihan “*Chi Square*” dan klik pilihan “*Risk*”



- f) Klik “*Continue*”  
g) Klik option “*Cells*”, bawa bagian “*Percentages*” dan klik “*Row*”



- h) Klik “*Continue*”  
i) Klik “*OK*” hasilnya tampak sbb:

### Crosstabs

status pekerjaan ibu \* status menyusui asi Crosstabulation

			status menyusui asi		Total
			tdk EKSKLUSIVE	EKSKLUSIVE	
status pekerjaan ibu	KERJA	Count % within status pekerjaan ibu	17 68.0%	8 32.0%	25 100.0%
	tidak kerja	Count % within status pekerjaan ibu	7 28.0%	18 72.0%	25 100.0%
Total		Count % within status pekerjaan ibu	24 48.0%	26 52.0%	50 100.0%

### Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	8.013 <sup>b</sup>	1	.005		
Continuity Correction <sup>a</sup>	6.490	1	.011		
Likelihood Ratio	8.244	1	.004		
Fisher's Exact Test				.010	.005
Linear-by-Linear Association	7.853	1	.005		
N of Valid Cases	50				

- a. Computed only for a 2x2 table
- b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.00.

### Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for status pekerjaan ibu (TIDAK KERJA / KERJA)	5.464	1.627	18.357
For cohort status menyusui asi = YA EKSKLUSIVE	2.250	1.209	4.189
For cohort status menyusui asi = TIDAK EKSKLUSIVE	.412	.208	.816
N of Valid Cases	50		

Pada hasil di atas tertampil tabel silang antara pekerjaan dengan pola menyusui, dengan angka di masing-masing selnya. Angka yang paling atas adalah jumlah kasus masing-masing sel, angka kedua adalah persentase menurut baris (data yang kita analisis “ASI.SAV, berasal dari penelitian Cross Sectional sehingga persen yang ditampilkan adalah persentase baris, namun bila jenis penelitiannya *Case Control* angka persentase yang digunakan adalah persentase kolom.

Dari analisis di atas maka interpretasinya:

Ada sebanyak 18 (72,0%) ibu yang tidak bekerja menyusui bayi secara eksklusif. Sedangkan diantara ibu yang bekerja, ada 8 (32,0%) yang menyusui secara eksklusif. Hasil uji *Chi Square* dapat dilihat pada kotak “*Chi Square Test*”. Dari print out muncul dengan beberapa bentuk/angka sehingga menimbulkan pertanyaan, “Angka yang mana yang kita pakai?” apakah *Pearson*, *Continuity Correction*, *Likelihood* atau *Fisher*?”

Aturan berlaku pada *Chi Square* adalah sbb:

- a) Bila pada  $2 \times 2$  dijumpai nilai Expected (harapan) kurang dari 5, maka yang digunakan adalah “*Fisher’s Exact Test*”

- b) Bila tabel  $2 \times 2$ , dan tidak ada nilai  $E < 5$ , maka uji yang dipakai sebaiknya “***Continuity Correction (a)***”
- c) Bila tabelnya lebih dari  $2 \times 2$ , misalnya  $3 \times 2$ ,  $3 \times 3$  dsb, maka digunakan uji “***Pearson Chi Square***”
- d) Uji “***Likelihood Ratio***” dan “***Linear-by-Linear Assciation***”, biasanya digunakan untuk keperluan lebih spesifik, misalnya analisis stratifikasi pada bidang epidemiologi dan juga untuk mengetahui hubungan linier dua variable katagorik, sehingga kedua jenis ini jarang digunakan.

Untuk mengetahui adanya nilai E kurang dari 5, dapat dilihat pada footnote b dibawah kotak Chi-Square Test, dan tertulis diatas nilainya 0 cell (0 %) berarti pada tabel silang diatas tidak ditemukan ada nilai  $E < 5$ . Dengan demikian kita menggunakan uji Chi Square yang sudah dilakukan koreksi (***Continuity Correction***) dengan p value dapat dilihat pada kolom “***Asymp. Sig***” dan terlihat p valuenya = 0,011. berarti kesimpulannya ada perbedaan perilaku menyusui eksklusif antara ibu yang bekerja dengan ibu yang tidak bekerja. Dengan kata lain dapat disimpulkan bahwa ada hubungan status

pekerjaan dengan perilaku menyusui eksklusif. Uji Chi square hanya dapat digunakan untuk mengetahui ada/tidaknya hubungan dua variabel, sehingga uji ini tidak dapat untuk mengetahui derajat/kekuatan hubungan dua variabel. Untuk mengetahui besar/kekuatan hubungan banyak metodenya tergantung latar belakang disiplin keilmuannya, misal untuk ilmu sosial dengan melihat koefisien Phi, koefisien Contingency dan cramer's V. sedangkan untuk bidang kesehatan terutama kesehatan masyarakat digunakan nilai OR atau RR. Nilai OR digunakan untuk jenis penelitian Cross Sectional dan Case Control, sedangkan nilai RR digunakan bila jenis penelitiannya Kohort.

Pada hasil di atas nilai OR terdapat pada baris Odds ratio yaitu 5,464 (95% CI: 1,627 – 18,357). Sedangkan nilai RR terlihat dari baris **For Cohort** yaitu besarnya 2,250 (95% CI: 1,209 – 4,189). Pada data ini berasal dari penelitian Cross Sectional maka kita dapat menginterpretasikan nilai OR=5,464 sbb: Ibu yang tidak bekerja mempunyai peluang 5,46 kali untuk menyusui eksklusif dibandingkan ibu yang bekerja.. Pada perintah Crosstab nilai OR akan keluar bila tabel silang 2 x 2, bila tabel silang lebih dari 2 x 2, misalnya 3 x 2, 4 x 2

dsb, maka nilai OR dapat diperoleh dengan analisis regresi logistik sederhana dengan cara membuat “Dummy variable”

*Penyajian dan Interpretasi di Laporan Penelitian:*

Tabel ...

Distribusi Responden Menurut Jenis Pekerjaan dan Perilaku menyusui

Jenis Pekerjaan	Menyusui				Total		OR (95% CI)	P value
	Tidak Eksklusif		Eksklusif		n	%		
	n	%	n	%				
Bekerja	17	68,0	8	32,0	25	100	5,464	0,011
Tdk Bekerja	7	28,0	18	72,0	25	100	1,6 – 18,3	
Jumlah	26	52,0	24	48,0	50	100		

Hasil analisis hubungan antara status pekerjaan dengan perilaku menyusui eksklusif diperoleh bahwa ada sebanyak 8 (32%) ibu yang bekerja menyusui bayi secara eksklusif. Sedangkan diantara ibu yang tidak bekerja, ada 18 (72,0%) yang menyusui secara eksklusif. Hasil uji statistik diperoleh nilai  $p=0,011$  maka dapat disimpulkan ada perbedaan proporsi kejadian menyusui eksklusif antara ibu tidak bekerja dengan ibu yang bekerja (ada hubungan yang signifikan antara pekerjaan dengan perilaku menyusui). Dari hasil analisis diperoleh pula nilai  $OR=5,464$ , artinya ibu tidak bekerja mempunyai peluang 5,46 kali untuk menyusui eksklusif dibanding ibu yang bekerja.

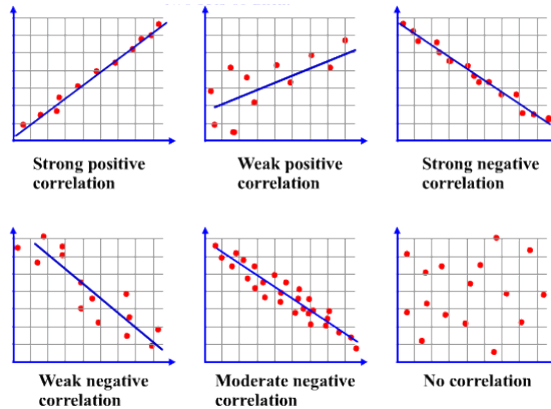
## F. Uji Korelasi

Uji korelasi digunakan untuk mengetahui derajat atau besarnya hubungan. Selain itu, uji korelasi digunakan untuk melihat arah hubungan dua variabel yang bersifat numerik. Korelasi dapat dilihat melalui interpretasi data kualitatif dan kuantitatif. Contoh uji korelasi:

- Korelasi antara asupan lemak (dalam gram) dengan skor IMT?
- Korelasi antara berat badan (dalam gram) dengan kadar hemoglobin?
- Korelasi antara skor pengetahuan tentang gizi dengan frekuensi konsumsi sayur dan buah?

### 1. Interpretasi korelasi secara kualitatif

Secara kualitatif hubungan dua variabel dapat dilihat dari diagram scatter. Diagram scatter menunjukkan titik-titik perpotongan nilai data dari dua variabel (variabel X= variabel bebas dan variabel Y=variabel terikat). Grafik variabel bebas (X) diletakkan pada garis horizontal, sedangkan variabel terikat (Y) diletakkan pada garis vertikal. Berdasarkan diagram scatter dapat diperoleh data terkait pola hubungan atau arah kedua variabel serta kekuatan hubungannya. Apabila titik-titik perpotongan data variabel menyempit maka menunjukkan kekuatan korelasi yang kuat, sedangkan apabila titik-titik perpotongan data variabel melebar maka menunjukkan kekuatan korelasi yang lemah.



## 2. Interpretasi korelasi secara kuantitatif

Interpretasi hasil korelasi adalah dengan melihat beberapa parameter, meliputi:

### a. Kekuatan korelasi secara statistic

Secara kuantitatif atau derajat hubungan dua variable dapat diketahui melalui *koefisien korelasi pearson product momen*, yang dinotasikan dengan huruf r kecil.

Koefisien korelasi (r) dapat diperoleh dari formula berikut:

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

Keterangan:

- r = koefisien korelasi
- n = jumlah subjek/ sampel
- X = nilai variabel bebas
- Y = nilai variabel terikat



Nilai koefisien korelasi ( $r$ ) berkisar pada angka 0-1. Jika nilai  $r=0$ , maka diketahui bahwa tidak terdapat hubungan linear. Jika nilai  $r=1$ , maka diketahui bahwa terdapat hubungan yang sempurna atau kuat. Kekuatan hubungan dua variabel secara kuantitatif dibagi menjadi beberapa kategori:

**Tabel, 3.** Kekuatan Hubungan Nilai R

<b>Nilai r</b>	<b>Kekuatan hubungan</b>
0.0 - <0.2	Sangat lemah
0.2 - <0.4	Lemah
0.4 - <0.6	Sedang
0.6 - <0.8	Kuat
0.8 – 1.00	Sangat kuat

### **b. Arah korelasi**

Tanda +/- (positif atau negatif) menunjukkan arah hubungan kedua variabel. Apabila tanda positif maka menunjukkan arah hubungan yang positif, artinya setiap terjadi peningkatan variabel independen (X) akan diikuti oleh peningkatan variabel dependen (Y). Sedangkan, apabila tanda negatif menunjukkan arah hubungan yang negatif, artinya setiap terjadi peningkatan nilai variabel independent (X) akan diikuti oleh penurunan variabel dependen (Y).

### **c. Nilai p atau kemaknaan secara statistic**

Koefisien korelasi ( $r$ ) hanya menunjukkan kekuatan dan arah hubungan linear dua variabel, namun

untuk mengetahui apakah hubungan kedua variabel bermakna secara statistik atau signifikan, maka uji hipotesis perlu dilakukan yaitu dengan menentukan nilai p value. Apabila p value menunjukkan  $< 0.05$  maka korelasi bermakna, sedangkan apabila p value  $> 0.05$  maka menunjukkan korelasi tidak bermakna. Uji hipotesis dapat dilakukan dengan dua cara, yaitu pertama: membandingkan nilai r hitung dengan r tabel, kedua: menggunakan pengujian dengan pendekatan distribusi t. Pada modul ini kita gunakan pendekatan distribusi t, dengan

Formula:

$$t = r \frac{n - 2}{\sqrt{1 - r^2}}$$

Keterangan:

n = jumlah subjek/ sampel

df = derajat kebebasan untuk mencari nilai tabel t

df = n-2

Untuk menentukan signifikansi dapat dilakukan dengan uji statistik pearson atau spearman, dengan beberapa syarat yang harus terpenuhi:

	<b>Pearson</b>	<b>Spearman</b>
Syarat	- Paling tidak <b>salah satu variabel terdistribusi normal</b>	- Bila <b>kedua variabel terdistribusi tidak normal</b>
	- Syarat <b>linearitas terpenuhi</b>	- Syarat <b>linearitas terpenuhi</b>

- Apabila syarat linearitas tidak terpenuhi, jangan melakukan uji korelasi, namun pertimbangkan untuk melakukan uji komparatif.
- Apabila syarat linearitas tidak terpenuhi, jangan melakukan uji korelasi, namun pertimbangkan untuk melakukan uji komparatif.

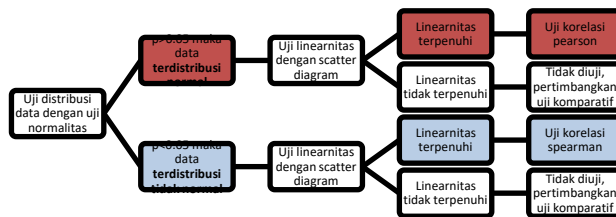
Contoh Apakah ada korelasi antara asupan Fe (numerik) dengan skor Hb (numerik) pada wanita usia subur?

Apakah ada korelasi antara asupan Fe (numerik) dengan skor Hb (numerik) pada wanita usia subur?

- Data asupan Fe terdistribusi normal
- Data asupan Fe dan skor Hb tidak terdistribusi normal

#### d. Uji korelasi dengan aplikasi SPSS

Langkah untuk melakukan uji korelasi:



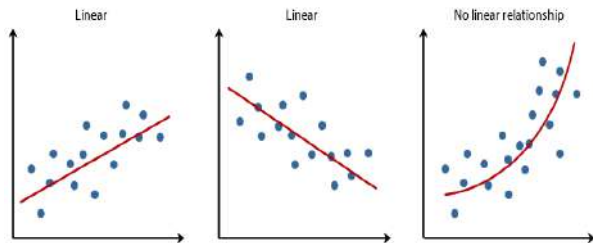
##### 1) Uji normalitas

- ✓ Digunakan untuk mengetahui distribusi data kontinyu/ rasio/ numerik

- ✓ Contoh data numerik: berat badan, tinggi badan, skor IMT, nilai pre-test, nilai posttest, dll
- ✓ Uji normalitas yaitu **shapiro wilk** (jumlah sampel kecil) dan **Kolmogorov-Smirnov** (jumlah sampel besar)
- ✓ Interpretasi : apabila p value  $<0.05$  maka data terdistribusi tidak normal dan apabila p value  $>0.05$  maka data terdistribusi normal

## 2) Uji linieritas

Menggunakan grafik scatterplot



## 3) Uji kemaknaan/ signifikansi

- ✓ Ditentukan berdasarkan p value
- ✓ Interpretasi : apabila p value  $<0.05$  maka data secara statistik signifikan, sedangkan apabila p value  $>0.05$  maka data secara statistik tidak signifikan

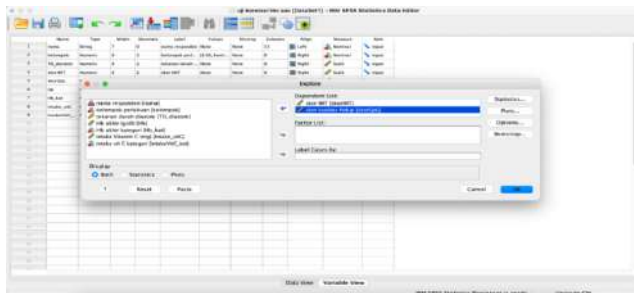
**Uji Korelasi Pearson (Uji hipotesis korelatif berdistribusi normal)**

Contoh kasus:

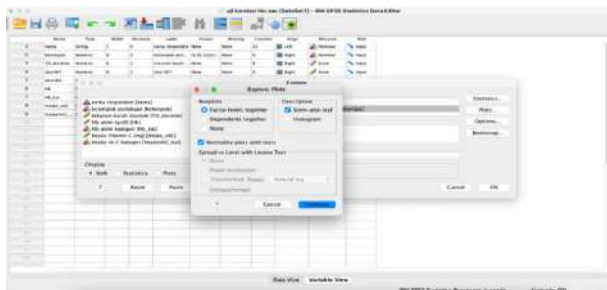
Seorang peneliti ingin mengetahui korelasi antara skor index massa tubuh (skor IMT) dengan skor kualitas hidup (skor QoL). Dengan kesalahan tipe satu 5% dan power penelitian 80%. Jumlah sampel adalah 30 orang dewasa.

1. Melakukan **uji normalitas** pada variabel skor IMT dan skor kualitas hidup (QoL)

Analyze → Descriptive statistics → Explore → masukkan variabel skor IMT dan skor kualitas hidup pada box Dependent list



Pada explore → klik Plots → aktifkan Normality plots with tests → Continue



Klik → OK hingga muncul output SPSS

**Tests of Normality**

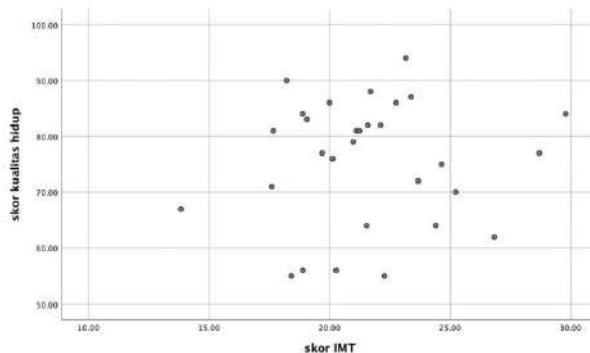
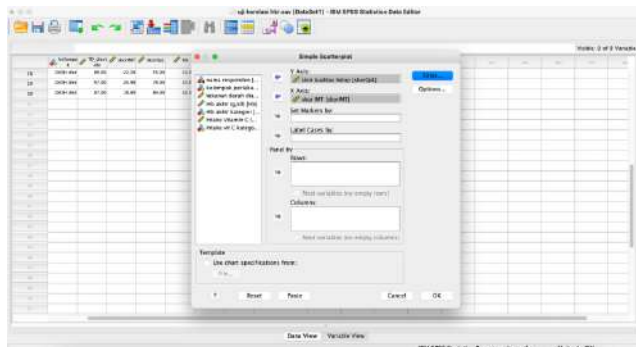
	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
skor IMT	.088	30	.200 <sup>*</sup>	.973	30	.619
skor kualitas hidup	.155	30	.064	.932	30	.055

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

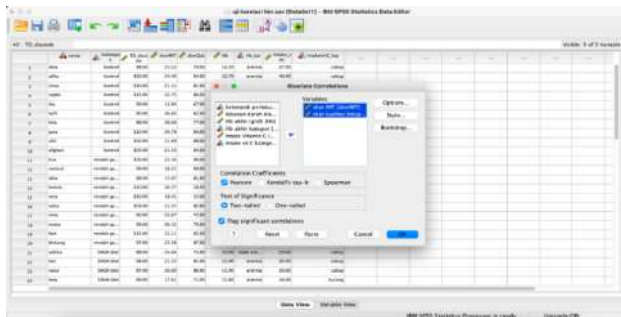
2. Melakukan uji linearitas dengan membuat grafik scatter

Graph → Legacy Dialogs → Scatter/Dots → Simple scatter → Define



3. Melakukan uji Pearson

Analyze → Correlate → Bivariate → masukkan variabel skor IMT dan skor kualitas hidup pada box Variables



Pada Correlation Coefficients → aktifkan Pearson  
 Klik → OK hingga muncul output SPSS

		skor IMT	skor kualitas hidup
skor IMT	Pearson Correlation	1	.081
	Sig. (2-tailed)		.671
	N	30	30
skor kualitas hidup	Pearson Correlation	.081	1
	Sig. (2-tailed)	.671	
	N	30	30

Interpretasi

No	Contoh Kasus	Skala (numerik /kategori kal)	Uji normalita (Normal/ Tidak)	Linearitas (Linear/tidak linear)	Uji
1	Korelasi antara skor	- Skor IMT: numerik	- Skor IMT= 0,619	- Skor IMT → linear - Skor	Uji pearso

IMT dengan skor QoL	- Skor QoL: numerik	(>0.05): distribusi data normal	QoL → linear	n
		- Skor QoL= 0.055 (>0.05): distribusi data normal		
p value	<b>Kekuatan korelasi (r)</b>	<b>Arah korelasi</b>		
p= 0,671	r= 0,081	Positif		
Interpretasi: secara statistik tidak bermakna	Interpretasi= kekuatan hubungannya sangat lemah	Interpretasi: semakin tinggi skor IMT maka semakin tinggi pula skor QoL		

### Uji Korelasi Spearman (Uji hipotesis korelatif berdistribusi tidak normal)

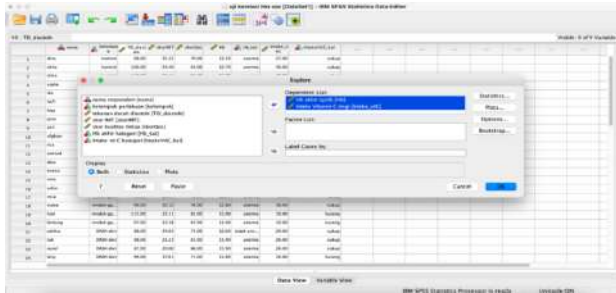
Contoh kasus:

Seorang peneliti ingin mengetahui korelasi antara asupan vitamin C dengan nilai Hb. Dengan kesalahan tipe satu 5% dan power penelitian 80%. Jumlah sampel adalah 30 orang dewasa.

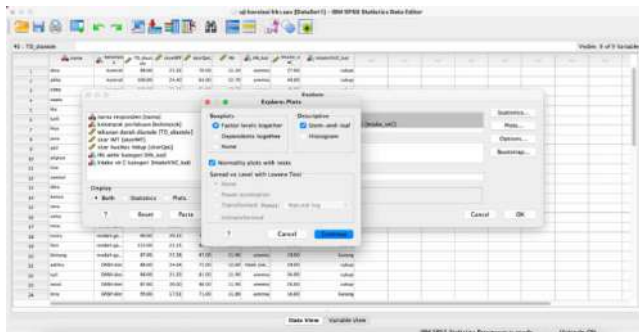
1. Melakukan uji normalitas pada variabel asupan vitamin C dan nilai Hb  
Analyze → Descriptive statistics → Explore → masukkan variabel asupan vitamin C dan nilai Hb pada box Dependent list



MODUL DIGITAL: Biostatistik untuk *Applied Research Method in Nutrition 2*



Pada explore → klik Plots → aktifkan Normality plots with tests → Continue



Klik → OK hingga muncul output SPSS

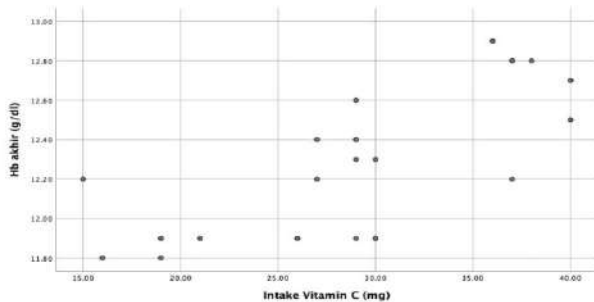
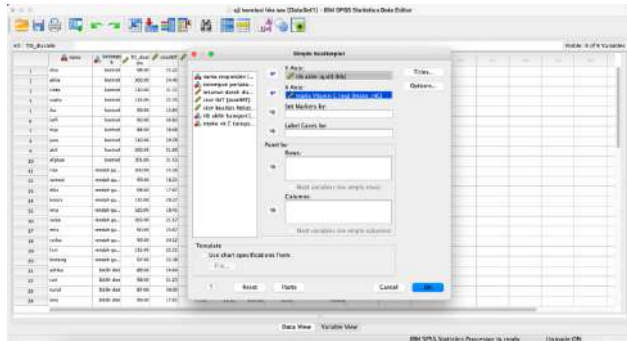
**Tests of Normality**

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Hb akhir (g/dl)	.217	30	.001	.894	30	.006
Intake Vitamin C (mg)	.145	30	.109	.924	30	.035

a. Lilliefors Significance Correction

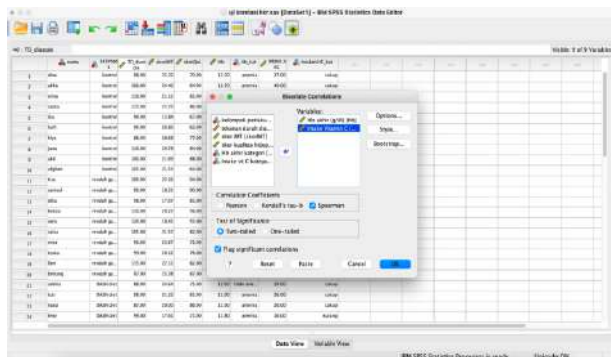
- Melakukan uji linearitas dengan membuat grafik scatter  
Graph → Legacy Dialogs → Scatter/Dots → Simple scatter → Define

# MODUL DIGITAL: Biostatistik untuk *Applied Research Method in Nutrition 2*



### 3. Melakukan uji Spearman

Analyze → Correlate → Bivariate → masukkan variabel skor IMT dan skor kualitas hidup pada box Variables



Pada correlation coefficients → aktifkan Spearman  
 Klik → OK hingga muncul output SPSS

**Correlations**

		Hb akhir (g/dl)		Intake Vitamin C (mg)
Spearman's rho	Hb akhir (g/dl)	Correlation Coefficient	1.000	.721**
		Sig. (2-tailed)	.	.000
		N	30	30
	Intake Vitamin C (mg)	Correlation Coefficient	.721**	1.000
		Sig. (2-tailed)	.000	.
		N	30	30

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

### Interpretasi

Kasus	Skala (numerik/kategorikal)	Uji normalita (Normal/Tidak)	Linearitas (Linear/tidak linear)	Uji
Korelasi antara asupan vitamin C dengan kadar Hb	- Asupan vitamin C: - Kadar Hb: numerik	- Asupan vitamin C= 0,006 (<0.05): distribusi data tidak normal - Kadar Hb= 0.035 (<0.05): distribusi data normal	- Asupan vitamin C → linear - Kadar Hb → linear	Uji Spearman
p value		Kekuatan korelasi (r)	Arah korelasi	
p= 0,000		r= 0,721	Positif	

Interpretasi:  
secara statistik  
bermakna

Interpretasi=  
kekuatan  
hubungannya  
kuat

Interpretasi:  
semakin tinggi  
asupan vitamin C  
semakin tinggi pula  
kadar Hb

## G. Uji Regresi Linier Sederhana

### 1. Pengertian

Uji regresi linier sederhana adalah jenis uji statistik yang bertujuan untuk mengetahui adanya hubungan antar dua data numerik atau *continuous*. Berbeda dengan uji korelasi, uji regresi bertujuan untuk melihat seberapa besar pengaruh perubahan nilai variabel independen terhadap variabel dependen. Contohnya peneliti ingin mengetahui seberapa besar pengaruh konsumsi gula terhadap kenaikan IMT. Uji korelasi akan memberikan informasi berupa ada atau tidaknya korelasi atau hubungan antar variabel, apakah korelasi antar variabel kuat atau lemah, dan apakah korelasi bersifat positif atau negatif. Namun, uji korelasi tidak bisa menjawab seberapa besar pengaruh konsumsi gula dalam jumlah tertentu mampu menaikkan IMT. Maka dibutuhkan uji regresi untuk menjawabnya.

Dasar dari pengujian uji regresi adalah untuk mendapatkan bagaimana model hubungan antara variabel dependen (atau variabel response) terhadap satu atau lebih variabel independen (atau variabel prediktor). Uji regresi linear sederhana menganalisis satu variabel independen, jika akan menganalisis beberapa variabel independen

digunakan uji *multiple linear regression*. Model uji regresi memiliki beberapa manfaat penggunaan diantaranya;

- a. Memprediksi faktor-faktor yang memberikan efek pada variabel dependen atau variabel hasil. Contohnya, berdasarkan hasil uji regresi dari beberapa penelitian disebutkan bahwa faktor konsumsi telur tidak berhubungan dengan kadar kolesterol total. Maka peneliti lain tidak akan memasukkan atau mengontrol variabel konsumsi telur ketika akan melakukan penelitian sejenis di penelitian lainnya. Hal ini bermanfaat untuk menghemat waktu, sumber daya dan dana pada penelitian atau ujicoba lainnya.
- b. Rumus yang terbentuk dari uji regresi bisa memberikan nilai prediksi pada variabel dependen atau variabel respon meskipun tidak ada nilai atau angka yang muncul pada saat diteliti. Contohnya peneliti ingin menganalisis respon pemberian dosis suplemen zat besi pada kadar hemoglobin remaja perempuan. Karena keterbatasan dana, peneliti hanya melakukan penelitian pada beberapa level dosis saja. formula regresi yang terbentuk akan memberikan estimasi kadar hemoglobin pada remaja meskipun dosis tersebut tidak diteliti secara langsung oleh peneliti.
- c. Data yang didapatkan juga bisa digunakan untuk melihat derajat hubungan antara variabel dependen

dan variabel independen, sehingga bisa melihat seberapa kuat hubungan antar variabel tersebut.

## 2. Konsep Uji Regresi Linear Sederhana

Hasil dari uji regresi akan diterjemahkan ke dalam rumus matematika regresi linear yaitu

$$Y = a + bx$$

- **x** disebut juga variabel independen atau predictor
  - Pada uji regresi linear sederhana hanya memiliki satu variabel prediktor 'x', dan pada uji multivariabel analisis akan digunakan beberapa pengujian variabel independen.
- **Y** disebut juga variabel dependent atau variabel respon, atau variabel hasil
- **a** merupakan *intercept*; yaitu nilai Y ketika  $x = 0$
- **b** disebut *slope* atau *gradien* dari estimasi garis yang didapatkan slope merepresentasikan kenaikan nilai rata-rata Y jika terdapat kenaikan satu unit nilai x
  - Nilai slope positif artinya y akan meningkat jika x meningkat
  - Nilai slope negatif artinya y menurun ketika x meningkat
- **a dan b** disebut juga dengan **koefisien regresi** dari garis estimasi regresi linier (*estimated linear regression line*)

Berdasarkan rumus tersebut bisa dilihat bahwa variabel terikat (Y) (variabel dependent, variabel respon atau variabel hasil) dipengaruhi oleh variabel independen (x) □

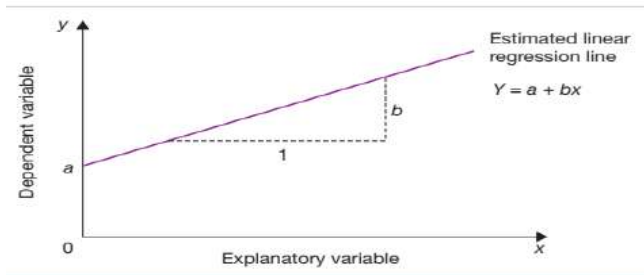


Figure SEQ Figure\\* ARABIC 1. Garis estimasi regresi linear yang ditunjukkan dengan intercept (a), slope (b) rata-rata kenaikan Y untuk setiap kenaikan 1 unit x

### 3. Teknis Uji Statistik

#### a. Dasar Uji Regresi Linier Sederhana

Gambaran untuk memahami uji regresi linear sederhana misalnya seorang peneliti ingin mengetahui hubungan antara tinggi badan (cm) terhadap tekanan darah sistolik (TDS) (mmHg) dan mendapatkan sampel 100 anak. Peneliti kemudian menggunakan uji regresi linier sederhana dengan tekanan darah sebagai variabel terikat/dependen dan tinggi badan sebagai variabel independen.

Berdasarkan hasil analisis didapatkan hasil output sebagai berikut:

**Tabel, 1.** Output hasil uji regresi

Variable	Parameter estimate	Standard error	Test statistic	P-value
Intercept	46.2817	16.7845	2.7574	0.0070
Height	0.4842	0.1396	3.4684	0.0008

Maka berdasarkan rumusan matematika regresi:  $Y = a + bx$ ; dengan ‘intercept’ sebagai nilai a, tinggi badan (TB) direpresentasikan sebagai ‘b’ (slope dari garis regresi). Maka rumus dari garis estimasi regresinya adalah: **TDS = 46,28 + 0,48 x TB**

Rumus regresi tersebut bisa digunakan untuk **memprediksi TDS** subjek lain pada populasi berdasarkan TB-nya. Misalnya jika seorang anak memiliki TB 115 cm, maka kita bisa memprediksi TDS nya adalah:  $46,28 + (0,48 \times 115) = 101,48$  mmHg; dan jika terdapat anak dengan TB 130 cm, maka prediksi TDS-nya adalah :  $46,28 + (0,48 \times 130) = 108,68$  mmHg.

b. Model Regresi Linear Sederhana

Regresi linear sederhana diperoleh dari data yang didapatkan dari subjek atau sampel. Perhitungan dari sampel ini harus bisa merepresentasikan kondisi yang ada di populasi.

Bagaimana menghitung rumus regresi dari sampel yang didapatkan untuk bisa memprediksi nilai pada populasi?

Metode yang sering digunakan adalah metode *sum least square* yaitu metode untuk menentukan



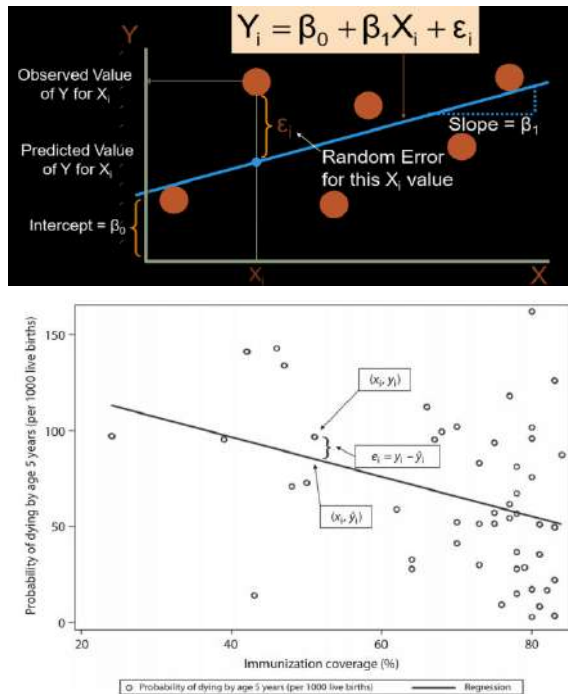
persamaan regresi dengan meminimalkan perbedaan jumlah nilai kuadrat antara nilai  $y$  dari hasil observasi dan nilai  $Y$  estimasi.

**Contohnya pada gambar 2:**

- $b_0$  atau intercept adalah garis estimasi dari nilai  $Y$  ketika  $X=0$
- $b_1$  adalah *slope* atau *gradient* atau kemiringan dari garis estimasi
  - o Merepresentasikan estimasi perubahan nilai  $Y$  jika terdapat peningkatan nilai 1 unit dari  $x$
- $E$  atau random error menggambarkan faktor lain baik yang diketahui maupun yang tidak diketahui
  - o Perbedaan angka observasi  $Y$  dari  $x_i$  dan nilai prediksi  $y$  dari  $x_i$  disebut dengan nilai residual. perbedaan ini bisa disebabkan karena faktor yang diketahui maupun faktor yang tidak diketahui

Figure SEQ Figure 1\* ARABIC 2 Model garis regresi linear

The diagram shows the linear regression equation  $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$  enclosed in a light orange box. Labels with arrows point to each part of the equation: 'Dependent Variable' points to  $Y_i$ ; 'Population Y intercept' points to  $\beta_0$ ; 'Population Slope Coefficient' points to  $\beta_1$ ; 'Independent Variable' points to  $X_i$ ; and 'Random Error term' points to  $\epsilon_i$ . Below the box, a bracket under  $\beta_0 + \beta_1 X_i$  is labeled 'Linear component', and a bracket under  $\epsilon_i$  is labeled 'Random Error component'.



*Figure 3.* Gambaran dari hasil angka observasi, garis least square regression, dan residual. Gambar kurung terbuka menunjukkan angka observasi dari data (x adalah cakupan dari imunisasi pada anak usia 1 tahun, y adalah probabilitas kematian anak pada usia 5 tahun). Garis least square regression direpresentasikan dengan garis lurus tebal. Angka prediksi ( $\hat{y}_i$ ) dari setiap  $x_i$  berada di sepanjang garis regresi. Residual ( $\hat{e}_i$ ), adalah perbedaan angka dari

nilai yang diobservasi ( $y_i$ ) dan nilai yang diprediksi ( $\hat{y}_i$ ).

c. Asumsi Uji Regresi Linear Sederhana  
Seperti uji statistic lainnya, uji regresi linear sederhana juga memiliki beberapa asumsi yang harus di cek terlebih dahulu;

- 1) Linearitas; Terdapat hubungan linear antara variabel x dan y
- 2) Normalitas; Asumsi normalitas mengacu pada data residual terdistribusi normal.
- 3) Independensi; Bahwa tidak ada hubungan antara residual dengan variabel dependen (y)
- 4) Homoscedasticity; Homoscedasticity menyatakan bahwa kesalahan prediksi (error) harus konstan

d. Dasar hitungan dari *sum least square* dijadikan dasar perhitungan atau penentuan nilai intercept dan slope

Rumus hitungan manual dari *sum least square* adalah:

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \hat{Y} = a + bX$$

n = jumlah sampel

x = angka observasi x (variabel independent/predictor)

$y$  = angka observasi yang didapatkan dari variabel  $x$  (variabel dependent atau terikat)

$b$  = slope atau estimasi perubahan nilai  $Y$  jika terdapat peningkatan nilai 1 unit dari  $x$

e. Contoh Soal dan Pembahasan

Peneliti ingin melihat hubungan antara pendapatan dengan pengeluaran konsumsi bahan makanan sumber konsumsi jumlah protein pada keluarga. Selebihnya, peneliti ingin memprediksi berapa konsumsi protein sebuah keluarga berdasarkan variabel pendapatan sebuah keluarga. Didapatkan 10 sampel keluarga dengan jumlah pendapatan (dalam ribuan) seperti pada table 2.

**Tabel, 2.** Jumlah konsumsi makanan sumber protein (gram) berdasarkan jumlah pendapatan (ribuan)

Variabel	1	2	3	4	5	6
Pendapatan	80	100	120	140	160	
		180	200	220	240	
		260				
Konsumsi	70	65	90	95	110	
		115	120	140	155	
		150				

**Pertanyaan :**

- 1) Dugalah bentuk (persamaan) linear hubungan tersebut secara manual!

- 2) Tentukan konsumsi protein keluarga Adi jika pendapatannya Rp 175.000
- 3) Interpretasikan koefisien regresi yang diperoleh !
- 4) Hitunglah *Standard error of estimate*-nya !
- 5) Ujilah apakah pendapatan berpengaruh dalam peningkatan konsumsi protein!
- 6) Hitunglah korelasi antara kedua variabel tersebut !
- 7) Tentukan hasil uji statistik menggunakan SPSS dan cek hasil hitungan manual dengan hasil output SPSS!

**Jawaban;**

**1) Bentuk persamaan linear hubungan tersebut!**

Untuk menentukan persamaan linear perlu dihitung nilai ‘b’ atau ‘*slope*’ dengan rumus sbb:

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

Sampel	Pendapatan (x)	Konsumsi Protein (Y)	X <sup>2</sup>	XY	Y <sup>2</sup> (untuk uji r)
1	80	70	6400	5600	4900
2	100	65	10000	6500	4225
3	120	90	14400	10800	8100
4	140	95	19600	13300	9025
5	160	110	25600	17600	12100
6	180	115	32400	20400	13225
7	200	120	40000	24000	14400
8	220	140	48400	30800	19600
9	240	155	57600	37200	24025
10	260	150	67600	39000	22500
<b>Total</b>	<b>1700</b>	<b>1110</b>	<b>322000</b>	<b>205500</b>	<b>132100</b>

$$b = \frac{(10 \times 205500) - (1110 \times 1700)}{10 (322000) - (1700)^2} = 0,5091$$

Kemudian menggunakan rumus awal dapat kita tentukan  $a = \bar{Y} - b\bar{X}$

Maka intercept (a);

$$a = 1110 - (0,5091) (1700)$$

$$a = 24,454$$

Maka persamaan garis prediksi regresi linear sederhana adalah:  $\hat{Y} = 24,454 + 0,51 x$

**2) Tentukan jumlah konsumsi keluarga Adi jika pendapatannya Rp 175.000**

Berdasarkan perhitungan didapatkan rumus untuk prediksi konsumsi protein adalah;

$$Y = 24,454 + 0,51 x$$

$$Y = 24,454 + (0,51) (175)$$

$$Y = 24,454 + 88, 534$$

$$Y = 122,98 \text{ gram}$$

Jadi prediksi konsumsi protein keluarga Adi dengan pendapatan Rp 175.000 adalah 122,98 gram.

**3) Interpretasikan koefisien regresi yang diperoleh**

- Koefisien regresi adalah nilai a dan b.
- Interpretasi dari  $b = 0,51$  adalah setiap kenaikan Rp 1000 pendapatan sebuah keluarga akan

meningkatkan konsumsi protein sebanyak 0,51 gram.

- Interpretasi dari  $a = 24,454$  adalah jika variabel independen atau income atau  $x = 0$  maka nilai prediksi  $y$  adalah 24,454. Meskipun kadang tidak memungkinkan angka  $x=0$ , misalnya pada hubungan tinggi badan dengan tekanan darah, bahwa tidak mungkin tinggi badan = 0.

#### 4) **Hitunglah *Standard error of estimate*-nya !**

*Standard error of estimation* atau kesalahan baku pendugaan merupakan ukuran tentang variasi titik-titik di sekitar garis regresi, sehingga dapat digunakan sebagai ukuran ketepatan pendugaan atau formula dugaan nilai regresi.

Seperti pada gambar 3 bisa dilihat bahwa nilai  $e$  didapatkan dari selisih angka nilai  $y$  observasi dengan nilai  $y$  prediksi. Tingginya angka standard error mengindikasikan rendahnya ketepatan nilai prediksi dari hasil regresi.

$$Se = \sqrt{\frac{\sum e^2}{n-2}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}}$$

Menggunakan hitungan manual seperti tertuang dalam tabel berikut:

No	Pendapatan (x)	Konsumsi Protein (y observasi)	Y (prediksi) = 24,45 + 0,51 x	e (y observasi - y prediksi)	e <sup>2</sup>
1	80	70	65,25	4,75	22,5625
2	100	65	75,45	-10,45	109,2025
3	120	90	85,65	4,35	18,9225
4	140	95	95,85	-0,85	0,7225
5	160	110	106,05	3,95	15,6025
6	180	115	116,25	-1,25	1,5625
7	200	120	126,45	-6,45	41,6025
8	220	140	136,65	3,35	11,2225
9	240	155	146,85	8,15	66,4225
10	260	150	157,05	-7,05	49,7025
<b>Total</b>	<b>1700</b>	<b>1110</b>	<b>1111,5</b>	<b>-1,5</b>	<b>337,525</b>

Maka standard error =  $\sqrt{\frac{337,525}{10-2}} = \sqrt{42,19} = 6,49$

**5) Hitunglah korelasi antara kedua variabel tersebut**

Uji korelasi bisa digunakan untuk melihat kekuatan korelasi antar variabel yang direpresentasikan dengan hitungan nilai  $r^2$ .

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

Dengan menggunakan hitungan tabel perhitungan sebelumnya maka didapatkan:



$$r^2 = \frac{10(205500) - (1700)(1110)}{\sqrt{[10(322000) - (1700)^2][10(132100) - (1110)^2]}} = 0,98$$

Nilai  $r = \sqrt{0,98} = 0,99$

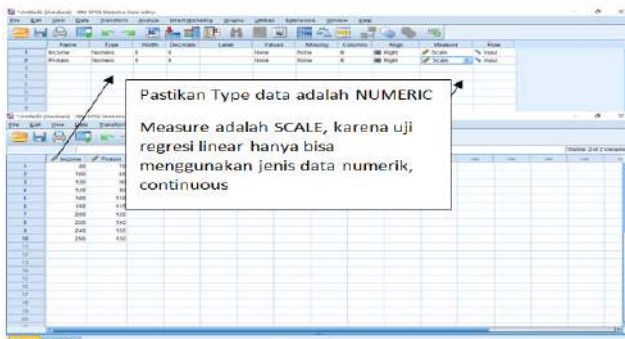
**Tabel, 2.** Interpretasi Nilai R

<i>Nilai r</i>	<i>Interpretasi</i>
0	Tidak ada korelasi
$r < 0,5$	Korelasi lemah
$(0,5 < r < 0,75)$	Korelasi sedang
$0,75 < r < 0,9$	Korelasi kuat
<b><math>0,9 &lt; r &lt; 1</math></b>	<b>Korelasi sangat kuat</b>
1	Korelasi sempurna

## Penggunaan Software SPSS untuk Uji Regresi Linear Sederhana

### a) Input data

Masukkan nama variabel di sheet **variabel view**  
 Karena uji regresi menggunakan data numerik, continuous maka pastikan pilih opsi **NUMERIC** pada kolom Type dan opsi **SCALE** pada kolom Measure.

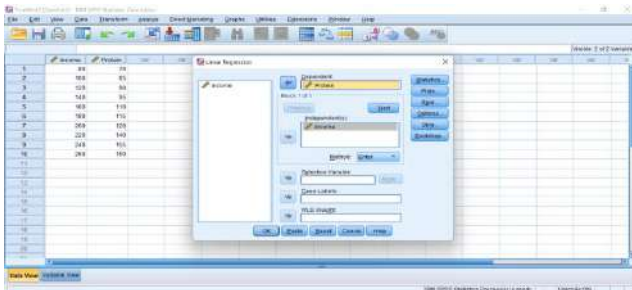
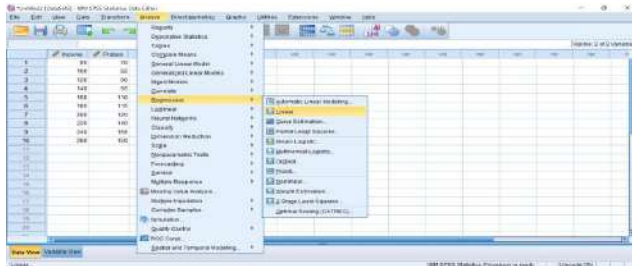


## b) Uji Regresi Pengecekan Asumsi Uji Regresi

Untuk melakukan tes uji regresi sederhana

Klik **Analyze**  **Regression**  **Linear**

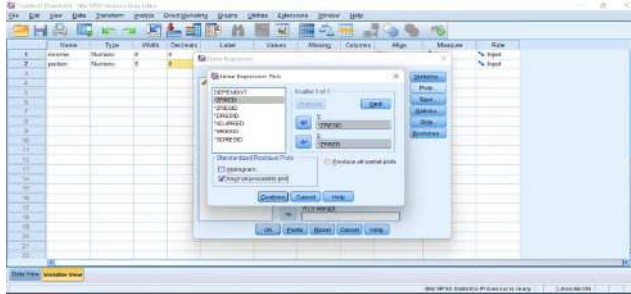
Masukkan Variabel **Income** ke form Variabel **Independent** dan **Protein** ke Varibel **Dependent**



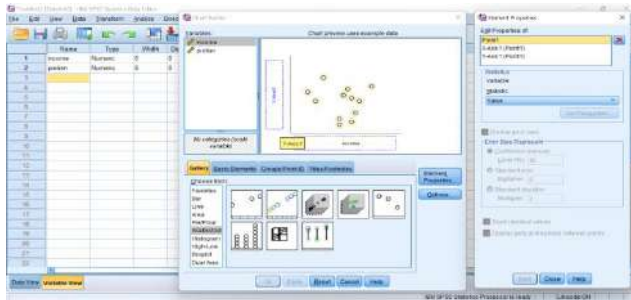
Klik **Statistic** → dan pilih Opsi Model Fit dan



Klik Plots → Masukkan **ZRESID** di form **Y** dan **ZPRED** di kolom **X** → Klik Normal probability Plot dan Histogram → Continue



Untuk Melihat Grafik Regresi  
Graph → Chart Builder → Scatter/Dot  
Tarik variabel Income di garis X  
Tarik variabel Protein di garis Y  
OK

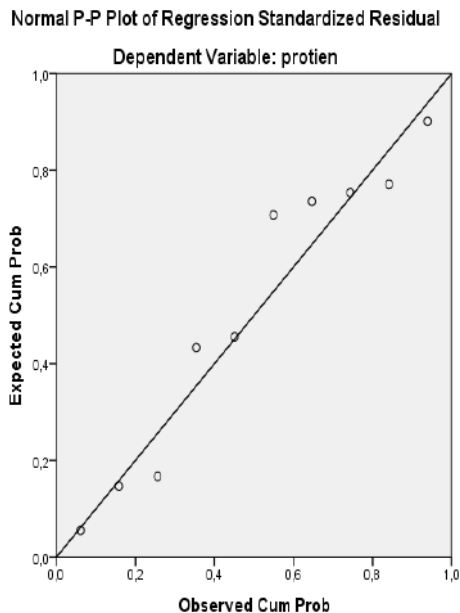


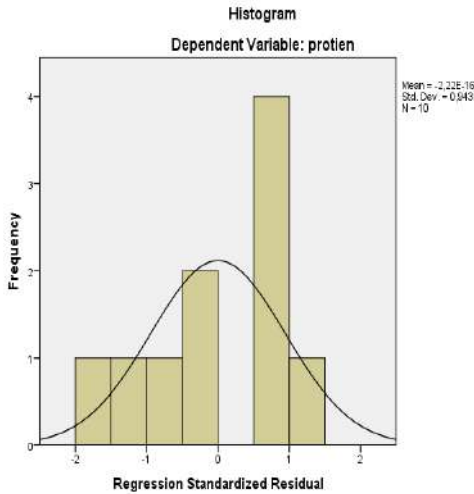
## Hasil Output

### 1) Hasil Asumsi

#### a) Normalitas

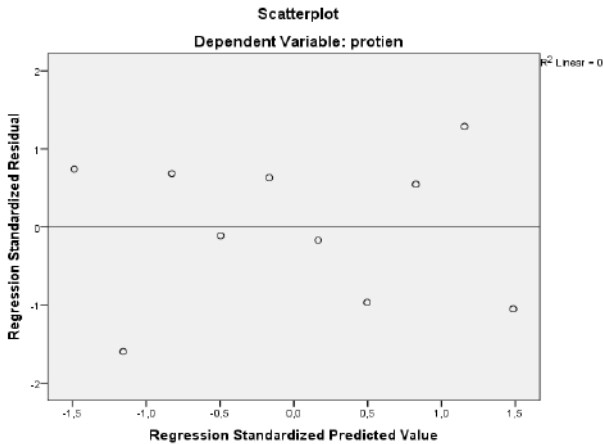
Untuk mengecek normalitas bisa dilihat dari hasil Output P-P plot dan histogram. Dari gambar bisa dilihat bahwa persebaran data tidak terlalu mengikuti arah garis diagonal atau ada sedikit penyimpangan data. Hal ini bisa terjadi karena jumlah data yang sedikit (10 sampel). Maka bisa dilakukan pengecekan hasil histogram, dimana arah grafik menunjukkan grafik yang mengarah pada distribusi normal.





**b) Independensi dan Homoscedasticity**

Grafik ini menunjukkan residual dari variabel dependen atau variabel hasil (y). Menunjukkan tidak ada pola khusus dari variabel hasil.



### c) Linearitas

Linearitas menunjukkan bahwa variabel independen dari regresi memiliki hubungan dengan variabel dependen. Jika nilai residual terdistribusi normal, dan memenuhi asumsi homoscedastic, maka bisa dipastikan bahwa terdapat linearitas antar variabel.

## 2) Hasil Pengujian Regresi

**Variables Entered/Removed<sup>a</sup>**

Model	Variables Entered	Variables Removed	Method
1	Income <sup>b</sup>	.	Enter

a. Dependent Variable: Protein

b. All requested variables entered.

Tabel ini menunjukkan bahwa variabel dependen adalah protein, dengan memasukkan satu variabel predictor yaitu income atau pendapatan.

**Model Summary<sup>b</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	Change Statistics			Sig. F Change
						F Change	df1	df2	
1	.981 <sup>a</sup>	.962	.957	6.493	.962	202.868	1	8	.000

a. Predictors: (Constant), Income

b. Dependent Variable: Protein

Tabel ini menunjukkan kekuatan model berdasarkan variabel yang dimasukkan

- **R square** menunjukkan proporsi varians variabel 'konsumsi protein' yang diperhitungkan dari variabel 'pendapatan'. Angka R Square

0,962 bisa diartikan bahwa pada kasus ini, 96,2% varian pada ‘konsumsi protein’ bisa dijelaskan dengan variabel ‘pendapatan’.

- **R** menunjukkan seberapa besar kekuatan hubungan variabel independen bisa memprediksi nilai variabel dependen.
- Bisa dilihat bahwa hasil menggunakan analisis statistic menunjukkan hasil yang mendekati angka perhitungan manual dengan nilai  $r = 0,98$ .
- Cek juga nilai *standard error estimate* yang didapatkan pada tabel dengan hasil perhitungan (error = 6,49) □ menunjukkan nilai residual atau perbedaan angka ‘y’ prediksi dan ‘y’ hasil observasi.

**ANOVA<sup>a</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8552,727	1	8552,727	202,868	,000 <sup>b</sup>
	Residual	337,273	8	42,159		
	Total	8890,000	9			

a. Dependent Variable: Protein

b. Predictors: (Constant), Income

- Tabel ANOVA pada output analisis regresi menunjukkan adanya hubungan yang linear secara signifikan antar variabel, dilihat dari nilai signifikansi F yaitu 0,000 ( $p < 0,05$ ).

- Kekuatan korelasi  $r^2$  juga bisa dihitung secara manual berdasarkan hasil output tabel ANOVA =  $100 \times (8552,727:8890) = 0,962$  (**cek kolom R square Tabel Model Summary**)

Model	Unstandardized Coefficients			Standardized Coefficients			95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Tolerance	VIF	
1	(Constant)	24,455	6,414		3,813	,005	9,864	39,245		
	Income	,509	,036	,991	14,243	,000	,427	,592	1,000	1,000

a. Dependent Variable: Protein

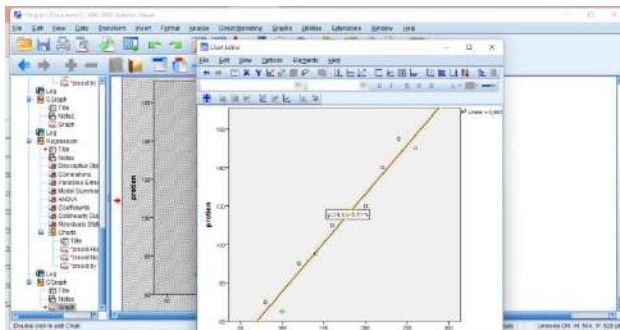
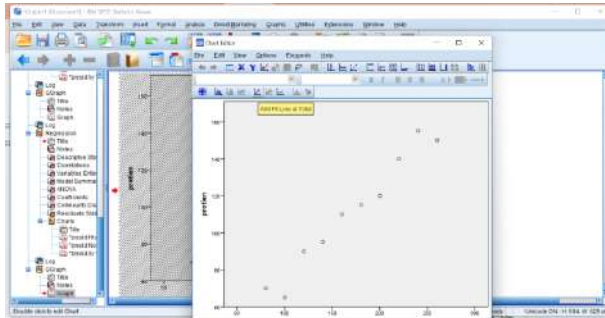
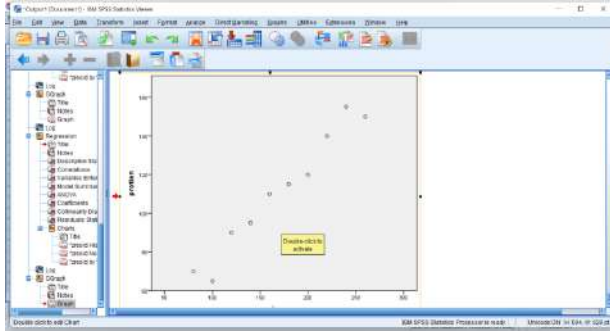
### Tabel ini menunjukkan koefisien regresi

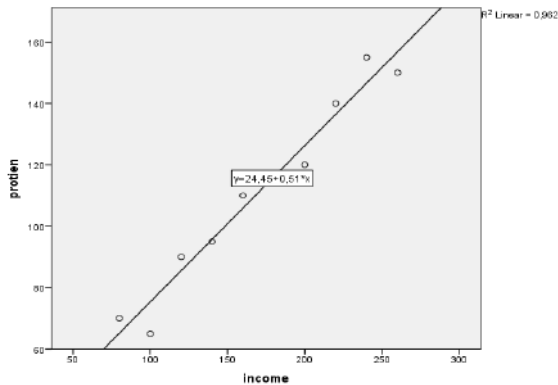
- Dengan nilai intercept ditunjukkan pada baris constant = 24,455; dan nilai slope untuk predictor (x) income adalah 0,509 - 0,51
- Maka berdasarkan hasil output uji regresi linier sederhana SPSS didapatkan prediksi persamaan regresi:  $Y = 24,455 + (0,51) x$  **cek hasil hitung manual**
- Nilai p-value untuk variabel independen yaitu income atau pendapatan adalah 0,000 ( $p < 0,005$ ) menunjukkan variabel pendapatan sebagai faktor predictor yang signifikan secara statistik mempengaruhi hasil variabel outcome atau dependent yaitu konsumsi protein.



## Grafik Regresi

Pada menu output → Double click gambar/grafik → Klik Add Fit Line to Total → OK





### Latihan Soal Mandiri

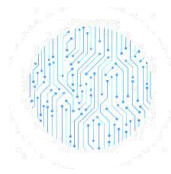
Peneliti ingin mengetahui hubungan antara BMI dengan tekanan darah. Sampel yang didapatkan adalah 20 responden dengan rincian data berikut:

No	BMI	Tekanan Darah
1	19,72	121
2	23,62	124
3	21,22	130
4	20,90	133,5
5	21,60	138,5
6	23,26	143,4
7	24,46	147,0
8	21,80	145,7
9	20,96	154,0
10	25,65	162,3
11	23,56	159,9
12	28,38	166,8
13	25,07	169,0
14	23,95	170,2

15	25,56	168,0
16	31,18	172,5
17	29,30	184,3
18	27,46	174,0
19	26,48	178,3
20	31,20	188,0

**Pertanyaan:**

- Dugalah bentuk (persamaan) linear hubungan tersebut secara manual!
- Interpretasikan koefisien regresi yang diperoleh!
- Hitunglah korelasi dan interpretasi hasil korelasi uji regresi!
- Tentukan hasil uji statistik menggunakan SPSS dan cek hasil hitungan manual dengan hasil output SPSS!
- Tentukan nilai prediksi tekanan darah dari Bapak Yudi jika BMI-nya adalah 23,9



## DAFTAR PUSTAKA

- Appel LJ, Sacks FM, Carey VJ, Obarzanek E, Swain JF, Miller ER et al. Effects of protein, monounsaturated fat, and carbohydrate intake on blood pressure and serum lipids: Results of the OmniHeart randomized trial. *JAMA*. 2005 Nov 16;294(19):2455-2464. doi: 10.1001/jama.294.19.2455
- Najmah. 2016. *Statistik Kesehatan Aplikasi STATA dan SPSS*. Jakarta: Penerbit Salemba Medika.
- Ott, R. L. & Longnecker, M. An introduction to statistical methods and data analysis (7th ed.). 2015. Boston :Brooks/Cole, Cengage Learning
- Petrie, Aviva, and Caroline Sabin. 2020. *Medical Statistics at a Glance*. 4th ed. Hoboken, NJ: Wiley-Blackwell.
- R. Lyman Ott, Michael Longnecker. 2016. *An Introduction to Statistical Methods and Data Analysis, Seventh Edition*.
- Rosner, Bernard. 2016. *Fundamentals of Biostatistics* (8th ed.). Boston :Brooks/Cole, Cengage Learning
- Sabri, Luknis dan Hastanto, Sutanto Priyo. 2014. *Statistik Kesehatan*. Jakarta: PT Raja Grafindo Persada.

Semba RD, Kalm LM, Pee Sde, Ricks MO, Sari M, Bloem MW. Paternal smoking is associated with increased risk of child malnutrition among poor urban families in Indonesia. *Public Health Nutrition*. Cambridge University Press; 2007;10(1):7–15.

# MODUL DIGITAL

Biostatistik untuk *Applied Research Method in Nutrition 2*