



Modul 1

SFS 510 - Metodologi Penelitian Fisioterapi 2

Materi 1

Manajemen Data

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

2020

Pendahuluan

Bagian penting dari proses perencanaan penelitian adalah pengembangan rencana manajemen data yang menentukan bagaimana data akan direkam, diorganisir, dikurangi dan dianalisis. Rencana ini dimulai dengan proposal penelitian, menentukan pertanyaan penelitian, hipotesis dan desain. Sebelum data dikumpulkan, peneliti harus dapat mengidentifikasi variabel apa yang akan diukur, dengan menggunakan instrumen dan satuan pengukuran tertentu.

Mereka yang akan mengumpulkan data mungkin perlu dilatih dan penilaian keandalan yang dilakukan. Tidak diragukan lagi, beberapa rencana ini akan berubah setelah proyek dimulai, tapi tidak ada yang harus dimulai tanpa rencana di awal. Perencanaan ini membutuhkan pengetahuan tentang pengkodean data dan persyaratan format, statistik dan komputer. Tujuan materi ini adalah untuk menjelaskan prosedur pengaturan data yang akan dimasukkan ke dalam komputer dan dianalisis dengan program statistik.

Kerahasiaan dan Keamanan Data

Proposal penelitian akan mencakup rencana untuk menangani data, termasuk menjaga kerahasiaan informasi peserta. Semua subjek harus diberi nomor ID unik yang tidak terkait dengan nama, nomor unit medis, nomor jaminan sosial, atau pengidentifikasi pribadi lainnya. Dokumen untuk pengumpulan data harus menyertakan ID subjek saja. Daftar nama subjek, alamat atau nomor telepon dan kode ID yang sesuai dapat disimpan terpisah dan diamankan dari file lain jika peserta harus dihubungi.

Sebagai bagian dari *inform consent*, subyek harus yakin bahwa informasi pribadi mereka, data dari catatan medis dan data yang dikumpulkan sebagai bagian dari proyek hanya akan diakses jika diperlukan untuk penelitian. *Institutional review board* (IRB) yang menyetujui proyek ingin mengetahui jenis data yang akan dikumpulkan, tujuan yang digunakan untuk menggunakan data, akses, dan perlindungan untuk keamanan dan kerahasiaan.

Banyak negara memiliki peraturan di tempat yang menentukan standar ini. Di Amerika Serikat, ini adalah bagian dari *Privacy Rule of the Health Insurance Portability*

and Accountability Act (HIPAA). Di Kanada, hal ini digambarkan dalam the *Privacy Rule of the Health Insurance Portability and Accountability Act*.

Monitor Partisipasi Subjek

Selama proyek berlangsung, peneliti harus memiliki prosedur untuk menjaga catatan yang akurat dan lengkap dari keterlibatan subyek. Catatan harus menunjukkan berapa banyak subjek yang direkrut dan mengapa ada yang tidak memenuhi syarat, berapa banyak yang setuju untuk berpartisipasi, dan berapa banyak akhirnya berpartisipasi. Perubahan protokol penelitian harus dijelaskan. Pembagian kelompok dan penyimpangan yang terjadi harus didokumentasikan. Informasi ini relevan dengan validitas proyek dan akan menjadi penting pada saat peneliti menyelesaikan proyek penelitian.

Program Statistik

Sejumlah paket statistik untuk digunakan telah berkembang dengan pesat, Dua program yang paling umum digunakan adalah *Statistical Package for the Social Sciences* (SPSS), yang sekarang disebut sebagai SPSS (hak cipta IBM) dan *Statistical Analysis System* (SAS).

SPSS secara tradisional telah digunakan lebih banyak dalam penelitian ilmu sosial dan perilaku, meskipun penggunaannya telah meningkat dalam penelitian bidang kesehatan. Dalam bidang penelitian fisioterapi, mayoritas pengolahan data menggunakan program SPSS. SAS lebih banyak digunakan sebagai program biostatistik dalam penelitian kedokteran dan epidemiologi.

Meskipun paket ini sedikit berbeda, mereka mematuhi standar tertentu yang penting untuk manajemen data. Sebagian besar program menyediakan format untuk *entri* data yang mirip dengan *spreadsheet*. Data juga dapat diimpor ke program statistik dari *spreadsheet* seperti *Microsoft Excel*®.

Formulir Pengumpulan Data

Sistem perekaman data harus dikembangkan dengan cermat. Biasanya, data dikumpulkan dari setiap subjek dan direkam pada lembar terpisah atau langsung ke

program komputer. Kode identifikasi subjek dicantumkan, serta informasi relevan lainnya seperti tanggal, individu yang mengumpulkan data (jika ada lebih dari satu investigator), tugas kelompok subjek dan informasi demografis seperti usia, jenis kelamin, dan diagnosis. Jika memungkinkan, semua data harus dicantumkan dalam urutan yang akan dimasukkan dalam file data, untuk memfasilitasi *entri* data. Gambar 1 mengilustrasikan bentuk pengumpulan data untuk studi dua jenis diet pada pasien dengan diabetes.

Subject ID #	<u>4</u>	Date of enrollment	<u>9/23/04</u>
Group Assignment:	<input checked="" type="radio"/> Diet 1	<input type="radio"/> Diet 2	
Subject's Age (in years)	<u>58</u>		
Gender	<input checked="" type="radio"/> Male	<input type="radio"/> Female	
Year of Onset	<u>1998</u>		
BASELINE	Date	<u>11/04/04</u>	
Blood Sugar	Trial 1	<u>152</u>	Trial 2 <u>155</u>
Weight	<u>198</u>	(lbs)	
FOLLOW-UP	Date	<u>10/18/05</u>	
Blood Sugar	Trial 1	<u>120</u>	Trial 2 <u>125</u>
Weight	<u>178</u>	(lbs)	
Questions at Follow-up			
How many times per week (on average) do you exercise?			
0	Less than once a week		
<input checked="" type="radio"/> 1	Once a week		
2	2-3 times per week		
3	4 or more times per week		
-98	Refused to answer		
-99	Don't know		
	Other _____		
COMMENTS			

Gambar 1. Contoh Form Pengumpulan Data

Peneliti harus membuat keputusan tentang bagaimana data akan direkam pada setiap subjek. Apakah ada tingkat presisi dalam pengukuran yang harus digunakan, seperti mengukur ke milimeter terdekat atau setengah inci? Format untuk merekam respon terbuka atau data kualitatif harus ditentukan. Jika data hilang, alasannya harus disertakan. Pentingnya skema pengumpulan data terorganisasi dengan baik menjadi

jelas ketika peneliti mulai memasukkan data ke dalam komputer. Jika data tidak direkam dengan jelas dan dalam format yang konsisten, entri data akan menjadi proses yang sulit dan berpotensi terjadi kesalahan.

Data Coding

Bagian penting dari rencana pengumpulan data adalah pengembangan skema untuk merekam data. Beberapa pengukuran menghasilkan data kuantitatif, seperti rentang gerak dan tekanan darah. Variabel seperti jenis kelamin, kelompok dan ras menghasilkan data kategoris. Survei dan studi kualitatif dapat menghasilkan tanggapan terbuka yang harus dikodekan.

Jenis Variabel

Data dapat dimasukkan sebagai angka atau karakter. Data kuantitatif adalah numerik, memiliki nilai satu atau beberapa digit, terkadang termasuk titik desimal, dan terdiri dari hanya angka. Nilai numerik dapat didahului dengan tanda tambah atau minus, meskipun tanda plus diasumsikan dan tidak dimasukkan. Variabel karakter, disebut *alphanumeric* atau *string variable*, terdiri dari huruf atau karakter dan mungkin termasuk digit. *String variable* mungkin berupa huruf atau kata yang merepresentasikan nilai variabel, seperti pria/wanita atau nama negara bagian atau kota. Nilai uang dapat dikodekan untuk unit moneter yang berbeda, dengan atau tanpa tempat desimal. Variabel juga dapat dimasukkan sebagai tanggal, menggunakan salah satu dari banyak bentuk yang dapat diterima, seperti MM-DD-YYYY. Kolom tanggal dapat ditambahkan atau dikurangi untuk menentukan lamanya waktu dalam hari, minggu, bulan, atau tahun.

Kode untuk Kategori Variabel

Data untuk variabel kategori dimasukkan sebagai label. Misalnya, jika gender adalah variabel, kita masukkan baik pria atau wanita sebagai nilai data. Meskipun kita bisa memasukkan label lengkap sebagai data, jauh lebih mudah untuk mengkodennya. Menggunakan kode karakter, misalnya, gender dapat dikodekan F untuk perempuan dan M untuk pria. Hal ini umumnya direkomendasikan bahwa kode yang dimasukkan

sebagai data numerik untuk memfasilitasi analisis statistik, seperti kode 1 untuk perempuan dan 0 untuk laki-laki.

Untuk variabel *dichotomous* secara konvensional menggunakan 1 dan 0 sebagai kode, biasanya menandakan tidak adanya sifat sebagai nol. Sebagai label murni tidak peduli apakah kita kode gender sebagai 1 dan 0, sebagai 1 dan 2, atau nomor lainnya; Namun, banyak prosedur statistik hanya akan memanipulasi kategori data dengan 1 dan 0 sebagai kategori kode. Ketika desain penelitian mencakup perbandingan grup, penetapan grup masing-masing subjek harus diidentifikasi dengan kode untuk variabel pengelompokan. Keputusan pengkodean kategori variabel harus dibuat sebelum data dikumpulkan. Kode harus digunakan pada formulir pengumpulan data untuk mempercepat transfer data ke komputer.

Kesalahan Data

Hal yang tidak biasa untuk beberapa data yang akan hilang dari catatan subjek karena kesalahan dalam perekaman, tidak tersedianya informasi, tidak ada respon pada survei, atau masalah dalam pengumpulan data. Untuk mengidentifikasi nilai yang hilang, pada sebagian besar program computer akan dikosongkan. Yang lain memiliki aturan khusus untuk mengidentifikasi nilai yang hilang, seperti penggunaan periode di tempat data yang hilang. Hal ini tidak dianjurkan untuk menggunakan nol untuk mewakili nilai yang hilang, karena angka nol akan dibaca sebagai nomor dan mungkin benar ada nol dalam data.

Hal ini sering berguna untuk menetapkan kode khusus untuk nilai yang hilang, untuk mengidentifikasi alasan informasi yang hilang. Misalnya, kode terpisah dapat digunakan untuk membedakan penolakan untuk menjawab pertanyaan, tanggapan dari "tidak tahu," pertanyaan yang tidak diminta, penyidikan kesalahan dan sebagainya. Perbedaan tersebut dapat membantu untuk interpretasi hasil, terutama ketika ada banyak data yang hilang poin. Data yang hilang harus dikodekan menggunakan nilai numerik yang berada di luar jangkauan nilai data aktual apa pun. Sebagai contoh, kode-99 umumnya digunakan.

Data Entry

Standar struktur untuk entri data mengharuskan setiap variabel dimasukkan dalam kolom terpisah, dan setiap baris mewakili subjek individu. Data dapat diketik langsung ke dalam program statistik, atau dapat dimasukkan dalam *spreadsheet* pertama dan kemudian diimpor ke program statistik. Tidak peduli bagaimana informasi yang dimasukkan, peneliti yang bijak akan menyimpan data sesering mungkin dan membuat cadangan file data secara teratur.

Ketika data berasal sebagai *spreadsheet*, baris pertama di setiap kolom harus berisi nama variabel, yang kemudian akan terbaca oleh program statistik. Untuk memudahkan transfer ini, nama variabel harus sesuai dengan pembatasan dari program statistik. Selain baris pertama ini, Semua baris lain dalam *spreadsheet* harus berisi hanya data. Tidak ada formula atau diagram yang harus disertakan. Jika formula digunakan dalam sel tertentu, mereka harus dikonversi ke nilai data aktual sebelum mentransfer ke program statistik.

Nama Variabel

Nama variabel diidentifikasi pada setiap poin data dalam berkas. Setiap variabel dalam file harus memiliki nama yang unik. Ketika nama variabel panjang, singkatan dapat digunakan. Sebanyak mungkin, nama variabel harus mudah diidentifikasi. Aturan tertentu berlaku untuk nama variabel, tergantung pada paket statistik yang digunakan. Banyak program mengharuskan nama variabel tidak lebih dari delapan karakter (angka atau huruf), meskipun versi yang lebih baru dari beberapa paket memungkinkan nama variabel yang lebih panjang.

Nama variabel biasanya harus diawali dengan huruf dan tidak memiliki spasi. Beberapa program memungkinkan tanda hubung, garis bawah, simbol dolar atau tanda nomor dalam nama variabel. Umumnya karakter khusus seperti !, ? dan / tidak dapat digunakan dalam nama variabel. Sebagai contoh, nilai *pretest* dan *posttest* untuk nyeri dapat dikodekan *pain1* untuk *pretest* dan *pain2* untuk *posttest*. Para peneliti harus terbiasa dengan persyaratan untuk paket statistik yang mereka gunakan.

Bidang Variabel

Setiap baris data, yang mewakili skor satu subjek disebut rekaman atau kasus. Setiap skor individu atau nilai variabel diidentifikasi sebagai bidang. Sebuah kasus terdiri dari beberapa bidang data. Bidang dijelaskan sesuai dengan lebarnya, yaitu jumlah digit atau ruang yang diperlukan untuk nilai maksimum yang mungkin. Banyak program menetapkan lebar bidang *default* yang dapat diubah oleh peneliti.

Label

Karena nama variabel pendek dan kategori dikodekan, terkadang membingungkan untuk membaca cetakan analisis dengan banyak singkatan. Untuk memudahkan membaca output, sebagian besar program akan memungkinkan peneliti untuk menentukan label untuk nama variabel dan kode nilai kategori.

Label biasanya dapat diperluas menjadi 40-60 karakter atau lebih. Hal tersebut memungkinkan peneliti untuk menyesuaikan catatan dengan cara yang lebih mudah untuk melakukan interpretasi. Untuk membuat ini terjadi peneliti harus meluangkan waktu untuk mengetikkan semua label. Label dapat dicantumkan dalam formulir pengumpulan data.

Pembersihan Data

Ketika data diinput ke komputer, dan sebelum analisis dijalankan, data harus dicek terhadap data mentah untuk memastikan tidak ada perbedaan atau kesalahan pengkodean. Proses ini disebut pembersihan data, dan meskipun mungkin memakan waktu dan membosankan, sangat penting untuk memastikan validitas analisis data. File data dapat dicetak atau ditampilkan pada layar komputer dan diperiksa secara visual untuk akurasi terhadap data asli. Menjalankan statistik deskriptif pada data akan memungkinkan peneliti untuk melihat apakah ada perbedaan yang jelas. Jumlah frekuensi harus diperiksa untuk semua kategori variabel.

Output akan mencantumkan semua kode untuk setiap variabel dan berapa kali kode muncul di data. Ini juga akan menunjukkan berapa banyak subjek yang dihitung, dan jika ada data yang hilang untuk variabel tersebut. Hal ini memungkinkan peneliti untuk menentukan apakah ada kesalahan dalam kode, atau jika variabel memiliki terlalu

sedikit entri yang akan berguna. Untuk variabel kontinu, statistik deskriptif dan grafik, seperti histogram atau plot, harus dijalankan untuk menganalisis nilai minimum dan maksimal untuk memastikan bahwa kisaran skor sesuai. Dengan cara ini, peneliti dapat memastikan jika nilai dari kisaran yang mungkin telah dimasukkan. Misalnya, jika nilai gula darah maksimum dicetak sebagai 560, peneliti tahu ada kesalahan dan dapat kembali dan memperbaiki entri itu. Terkadang berguna untuk menyortir data, mengurutkan kembali subjek sesuai dengan nilai variabel tertentu, untuk menentukan apakah nomor yang sesuai telah dimasukkan.

Modifikasi Data

Semua program statistik termasuk proses untuk modifikasi data atau transformasi untuk membuat variabel baru atau untuk menetapkan kode baru untuk variabel yang ada. Sebagai contoh, kita mungkin ingin menghitung *mean* beberapa percobaan untuk digunakan untuk analisis data. Atau kita mungkin memiliki nilai untuk beberapa item pada skala dan ingin mendapatkan jumlah. Mungkin variabel kontinu akan dikonversi ke kategori. Ketika jenis transformasi dilakukan, variabel baru dibuat, dan harus diberi nama variabel baru dan unik.

Komputasi variabel baru

Komputasi variabel baru mensyaratkan bahwa beberapa operasi aritmatika dilakukan pada data yang ada. Semua program menggunakan simbol yang sama untuk mewakili operasi logis. Simbol ini, yang dikenal sebagai operator, digunakan untuk membuat ekspresi terkait instruksi komputer.

Pengkodean Ulang Variabel

Kami juga dapat menggunakan perbandingan untuk pengkodean ulang variabel dengan menentukan hubungan antara mereka. Perbandingan dapat ditentukan sebagai simbol atau kombinasi huruf. Ketika menetapkan nilai ke variabel baru, peneliti harus berhati-hati untuk tidak tumpang tindih kategori apapun, atau komputer tidak akan mampu melakukan fungsi yang diinginkan. Selain itu, pengelompokan harus mencerminkan berbagai nilai yang ada di data.

Prosedur Statistik

Banyak prosedur statistik juga menyediakan mekanisme untuk membuat dan menyimpan variabel baru. Misalnya, saat menjalankan analisis faktor, nilai faktor dibuat untuk setiap subjek pada setiap faktor. Nilai ini dapat disimpan dan digunakan sebagai variabel dalam analisis mendatang. Ketika prosedur regresi dijalankan, nilai residu dapat dihitung dan disimpan. Sebagian besar program memerlukan petunjuk khusus untuk opsi ini.

Analisis Data

Ketika pengumpulan data telah selesai, semua data yang dimasukkan dan disimpan (dan di-*backup*), dan kita menetapkan untuk memulai analisis data. Jika proposal penelitian dilakukan dengan baik, kita siap untuk mendekati tahap proses penelitian ini dengan cara yang terorganisir. Ini adalah ide yang baik untuk memulai dengan menjadi akrab dengan data dengan melihat statistik deskriptif-frekuensi untuk variabel kategori dan sarana untuk variabel kontinyu. Histogram, garis plot, *stem-and-leaf plots* atau *box plots* sangat membantu untuk menilai secara visual bentuk distribusi, dan untuk mengidentifikasi kesenjangan atau *outliers*. Untuk data korelasi, scatterplots harus dibuat untuk mendapatkan linearitas dan tingkat hubungan dalam data.

Langkah awal ini diperlukan untuk memahami lingkup data, dan mungkin menyarankan pendekatan statistik. Sebagai contoh, transformasi mungkin diperlukan untuk variabel nonlinier langkah berikutnya adalah menerapkan prosedur statistik untuk menjawab pertanyaan penelitian. Beberapa petunjuk bermanfaat untuk membuat proses ini efisien, Siapkan daftar hipotesa tertentu, variabel dan prosedur statistik yang sesuai untuk memandu waktu kita di komputer. Misalnya, jika kita bermaksud untuk membandingkan dua kelompok, tentukan *t-test*, dipasangkan atau tidak berpasangan, dan variabel mana yang akan digunakan. Jika kita menjalankan beberapa regresi, cantumkan variabel independen dan dependen untuk masing-masing kategori yang akan dianalisis.

Lihatlah output dan periksa temuan. Sering kali, pertanyaan tambahan muncul dan kita dapat memilih untuk menjalankan tes lebih lanjut. Misalnya, kita mungkin menemukan hubungan antara beberapa variabel yang tidak terantisipasi. Mungkin

menarik untuk melakukan analisis tertentu pada subkelompok dalam data. Program statistik memberikan pilihan penyaringan yang berbeda untuk memilih subyek sesuai dengan kriteria yang ditentukan.

Akhirnya, sebagian besar program statistik termasuk pilihan untuk membuat tabel atau grafik langsung dari data. Banyak dari program ini menyediakan pilihan yang cukup canggih, dengan berbagai huruf dan warna untuk menyesuaikan presentasi data. Grafik dan tabel ini dapat diimpor ke dalam program pengolah kata atau presentasi. Banyak jenis grafik yang berbeda biasanya tersedia, dan sering kali membantu untuk mencoba format yang berbeda untuk melihat yang menyajikan data yang terbaik.

Referensi

Keppel G. Design and Analysis: A Researcher's Handbook (4th ed.). Englewood Cliffs, NJ: Prentice Hall, 2004.

Green SB, Salkind NJ, Akey TM. Using SPSS for Windows and Macintosh: Analyzing and Understanding Data (4th ed.). Upper Saddle River, NJ: Prentice Hall, 2004.

Dennis Howitt, Duncan Cramer. Introduction to SPSS Statistics in Psychology For Version 19 and earlier Fifth edition. Pearson Education Limited Edinburgh Gate Harlow Essex England 2011

National Institutes of Health. HIPAA Privacy Rule: Information for researchers. Available at: <<http://privacyruleandresearch.nih.gov/>> Accessed August 17, 2006.

Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada, TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans. 1998 (with 2000, 2002 and 2005 amendments). Available at: <<http://www.pre.ethics.gc.ca/english/policystatement/policystatement.cfm>> Accessed August 17, 2006.