

بنام خدا

فرمت داده های مورد نیاز برای تحلیل آماری

ارائه دهنده:

زهرا شایان

دکترای آمار زیستی

استادیار گروه پزشکی اجتماعی

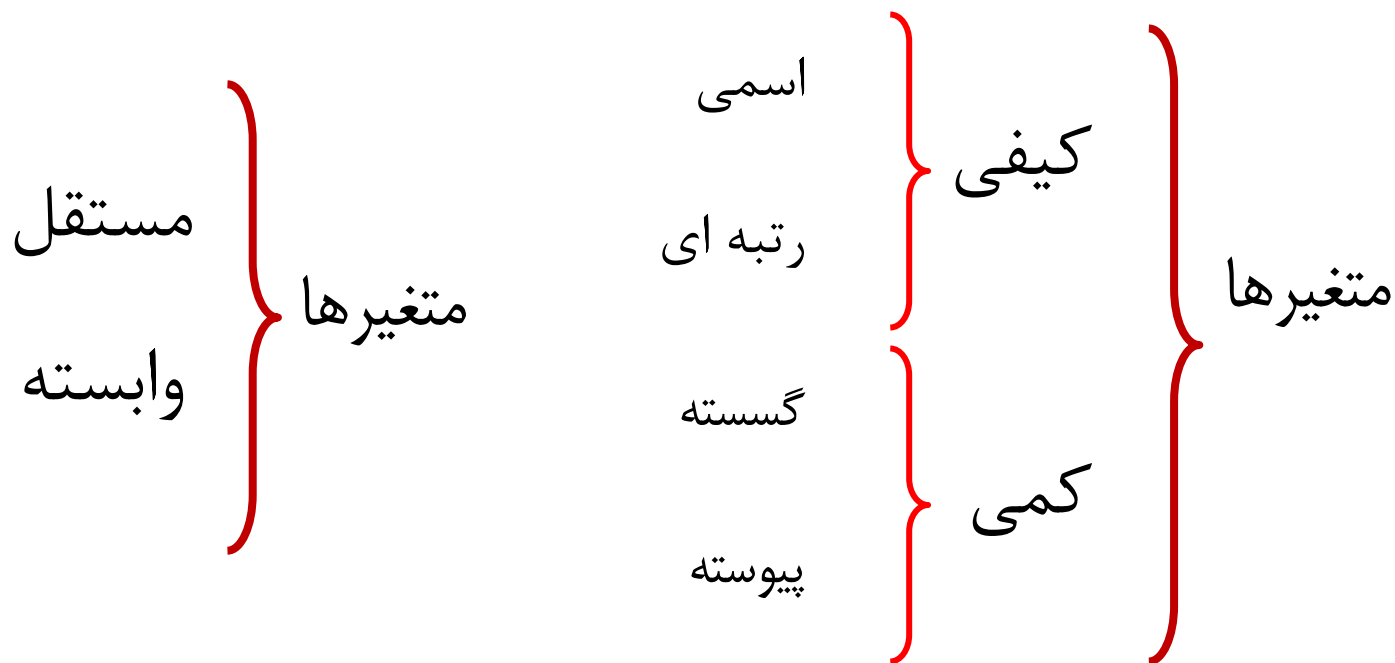
بهمن ۱۳۹۵

جدول متغیرها

یکی از قسمتهای مهم پروپوزال، تکمیل دقیق جدول متغیرها می باشد. به این منظور باید تعریف انواع متغیرها را بدانیم.

ردیف	عنوان متغیر		نوع متغیر		کمی		کیفی	
	مستقل	وابسته	پیوسته	گسسته	اسمی	رتبه‌ای		
۱								
۲								

متغیر ویژگی است که از یک فرد به فرد دیگر تغییر می کند.
دو نوع تقسیم بندی برای متغیرها وجود دارد:



کمی		کیفی	
متغیرهایی که از راه شمارش و اندازه‌گیری به دست می‌آید و با عدد و رقم بیان می‌شوند.		متغیرهایی که براساس آن جمعیت را گروه‌بندی می‌کنند و با عدد و رقم بیان نمی‌شوند	
پیوسته	گسسته (شمارشی)	رتبه‌ای	اسمی
با اندازه‌گیری بدست می‌آید. مقادیر اعشاری هم می‌پذیرد.	مقادیر از راه شمارش بدست می‌آید.	متغیر به چند گروه مجزا افراز شده و یک نوع ترتیب در اندازه متغیر وجود دارد.	طبقه‌بندی متغیر به رده‌های مجزا از هم را مشخص می‌کند. اندازه‌ها هیچگونه ترتیبی ندارند. به هر طبقه یک کد عددی داده می‌شود. عملیات ریاضی روی این اعداد انجام نمی‌شود.
قد وزن فشارخون	تعداد روزهای بستری در ICU تعداد دفعات عود بیماری	شدت بیماری (ضعیف=۱، متوسط=۲، شد ید=۳) سطح تحصیلات	جنس (زن=۱، مرد=۲) نوع درمان

متغیرهای کمی دارای **دو مقیاس** هستند:
فاصله ای ، نسبتی

مقیاس اندازه گیری

نسبتی	فاصله ای
صفر واقعی است علاوه بر اختلاف دو عدد، نسبت دو عدد هم دارای مفهوم است با تغییر واحد اندازه گیری نسبت دو عدد تغییر نمی کند	صفر قرار دادی است اختلاف بین اعداد را می توان محاسبه کرد ولی نسبت دو عدد دارای مفهوم نیست. با تغییر واحد اندازه گیری نسبت دو عدد تغییر می کند
قد وزن	درجه حرارت نمره هوش

تقسیم بندی دیگری از متغیرها:

مستقل ، وابسته

متغیر	
وابسته	مستقل
متغیری که تغییرات آن در نتیجه تغییر یا تغییرات چند متغیر مستقل است.	متغیری که تحت دستکاری قرار می گیرد و یا رفتار آن تحت مطالعه قرار می گیرد تا اثر و یا اثرات آن مشخص شود.
بررسی رابطه سیگار کشیدن با بیماریهای قلبی - عروقی	
بیماریهای قلبی - عروقی: متغیر وابسته	سیگار: متغیر مستقل

فراخوانی اطلاعات به نرم افزار جهت تحلیل آماری

برای تحلیل داده ها در SPSS باید اطلاعات هر فرد سطری وارد شود و متغیرها در ستونهای مجزا.

اگر یک متغیر در چند نوبت اندازه گیری شده باشد (مثلا مقدار قند خون هر ماه ثبت شود)، در SPSS باید در ستونهای کنار هم قرار گیرد. موقع استخراج داده ها به نرم افزار SPSS باید فرمت مناسب رعایت شود.

اطلاعات یک نمونه پرسشنامه

کد شناسایی

شهرستان

جنس

سال تولد

گروه خونی

قد بیمار

نتایج آزمایش: Na, K, PH, ...

فراخوانی داده ها در Excel

نمونه ای از اطلاعات فراخوانی شده در Excel

H	G	F	E	D	C	A
قد بیمار	گروه خونی	تاریخ تولد	سن	جنسیت	شهرستان	شناسه بیمار
۱۸۰,۰۰	-O	۱۳۳۸	۵۷	مرد	ارسنجان	۱
۱۶۵,۰۰	+B	۱۳۳۷	۵۸	مرد	شیراز	۲
	+A	۱۳۵۵	۴۰	مرد	صدرا	۳
۱۷۰,۰۰		۱۳۴۵	۵۰	زن	شیراز	۴
۱۵۷,۰۰	+B	۱۳۷۰	۲۵	مرد	اقلید	۵
		۱۳۱۰	۸۵	مرد	شیراز	۶
۱۶۴,۰۰	+A	۱۳۲۶	۶۹	مرد	شیراز	۷
	+O	۱۳۵۰	۴۵	مرد	کازرون	۸
۱۷۰,۰۰	+B	۱۳۲۰	۷۵	زن	مرودشت	۹

فراخوانی داده ها از Excel به SPSS

	شماره بیمار	شهرستان	سن	جنسیت	تاریخ تولد	گروه خونی	قد بیمار
1	1	ارمنجان	57.0	مرد	1338	O-	180.00
2	2	شیراز	58.0	مرد	1337	B+	165.00
3	3	صنرا	40.0	مرد	1355	A+	
4	4	شیراز	50.0	زن	1345		170.00
5	5	اقلید	25.0	مرد	1370	B+	157.00
6	6	شیراز	85.0	مرد	1310		
7	7	شیراز	69.0	مرد	1326	A+	164.00
8	8	کازرون	45.0	مرد	1350	O+	
9	9	مرودشت	75.0	زن	1320	B+	170.00

Name	Type	Width	Decimals	Label
شماره بیمار	String	36	0	شماره بیمار
شهرستان	String	20	0	
سن	Numeric	12	1	
جنسیت	String	6	0	
تاریخ تولد	String	4	0	تاریخ تولد
گروه خونی	String	3	0	گروه خونی
قد بیمار	String	6	0	قد بیمار

اندازه گیری آزمایشات در چند نوبت

داد های زیر یک نمونه از استخراج داده از نرم افزار است. اطلاعات همه بیماران و تمام آزمایشات در چند نوبت بصورت سطری استخراج شده است. این نوع داده برای تحلیل مناسب نیست.

D	C	B	A
مقدار آزمایش	تاریخ آزمایش	نام آزمایش	شناسه بیمار
۱۴۰	۱۳۹۱/۰۴/۱۸	NA	۱
۱۴۲	۱۳۹۱/۰۵/۰۸	NA	۱
۱۴۹	۱۳۹۱/۰۶/۱۲	NA	۱
۱۴۰	۱۳۹۱/۰۷/۰۹	NA	۱
۱۴۶	۱۳۹۱/۰۸/۰۶	NA	۱
۶	۱۳۹۱/۰۴/۱۸	K	۱
۵,۷	۱۳۹۱/۰۵/۰۸	K	۱
۷,۳	۱۳۹۱/۰۶/۱۲	K	۱
۷,۱	۱۳۹۱/۰۷/۰۹	K	۱
۷,۱	۱۳۹۱/۰۸/۰۶	K	۱

روش مناسب فراخوانی داده های تکراری در Excel

The screenshot displays the Microsoft Excel interface. The title bar at the top right reads "Book1 - Microsoft Excel". The ribbon is set to the "Home" tab, which is divided into three groups: "Clipboard", "Font", and "Alignment". The "Clipboard" group includes icons for Paste, Cut, Copy, and Format Painter. The "Font" group includes options for font face (Calibri), size (11), bold (B), italic (I), underline (U), text color, and background color. The "Alignment" group includes options for text alignment (left, center, right, justified), wrap text, and merge & center. Below the ribbon, the active cell is K8, and the formula bar is empty. The main workspace shows a table with the following data:

	A	B	C	D	E	F	G	H
1	Na1	date1	Na2	date2	Na3	date3	Na4	date4
2	140	1391/04/18	142	1391/05/08	149	1391/06/12	140	1391/07/09
3								
4								

- پس برای استخراج اطلاعات باید به چند نکته توجه شود:
- ❖ اگر امکان داشته باشد نام متغیرها بصورت انگلیسی وارد شود.
 - ❖ موقع فراخوانی، متغیرهای کیفی به کد عددی تبدیل شوند.
 - ❖ برای استخراج داده های تکراری از همان ابتدا به برنامه نویس در مورد نحوه استخراج آن توضیحات لازم داده شود.
 - ❖ برای ورود تاریخ باید فرمت Date انتخاب شود نه حالت رشته ای (string)، اگر فرمتی غیر از این باشد در فراخوانی داده ها مشکل ایجاد می شود، بخصوص برای کم کردن تاریخ ها (مثلا محاسبه سن). یا به فرمت تاریخ شمسی و میلادی باید توجه شود. چون نرم افزارها براساس تاریخ میلادی طراحی شده اند.

بررسی صحت داده های موجود در پرونده

عدم درستی و دقت در جمع آوری و ثبت داده ها منجر به تصمیم گیری نادرست و به تبع آن سیاست گذاری غلط خواهد شد.

پس باید به این نکات توجه شود:

- ❖ داده ها به دقت جمع آوری و ثبت شود (از افراد آموزش دیده استفاده شود)
- ❖ اگر تعدادی از پرسش ها بدون پاسخ مانده است باید تکمیل شود
- ❖ اگر پاسخ سؤال ها با یکدیگر همخوانی ندارند می بایست علت روشن شده و پرسشنامه اصلاح گردد
- ❖ داده های گردآوری شده بازبینی شود. حتی اگر موارد ذکر شده رعایت شود نمی توان مطمئن شد داده ها دقیق هستند. برای این منظور بهتر است قسمتی از داده ها (حدود ۱۰ درصد) انتخاب شود و دوباره همه موارد چک شد اگر در صد خطا پایین بود می توان به صحت داده ها مطمئن شد.

با تشکر از
توجه شما