۱۲۰ نکته برای تحلیل آماری با

SPSS

تألیف

دکتر علی شهنوازی

محتوی آموزشی در کانال تلگرام

@drshahnavazi

در دسترس می‌باشد.

مقدمه

تحلیل آماری بخشی از روش تحقیق است که برای انجام آن علاوه بر دانش تخصصی نیاز به نرم‌افزارهای آماری می‌باشد. از میان انواع مختلف نرم‌افزارهای موجود، SPSS به دلایل متعدد از قبیل بروز بودن، در دسترس بودن، جامع و شناخته شده بودن از جایگاه متفاوتی برخوردار هست. هر اندازه که نرم‌افزارهای دیگر تلاش در تخصص گرایی داشته‌اند، SPSS کوشیده به نیاز عمومی پاسخ دهد. به همین جهت حجم و تعداد مطالبی که در خصوص این نرم‌افزار تهیه گردیده از هر نرم‌افزار آماری دیگر بیشتر می‌باشد. بسیاری از کتب مربوط بیشتر به خود نرم‌افزار پرداخته‌اند و کمتر به تحلیل نتایج کوشیده‌اند. در مجموعه «**۱۲۰ نکته برای تحلیل آماری با SPSS**» تلاش می‌شود به هر دو جنبه توجه گردد.

□ نکته ۱

معرفی نرم‌افزار SPSS

SPSS مخفف Statistical Package for the Social Sciences می‌باشد. با خرید SPSS توسط IBM از نسخه ۱۹ به بعد نام آن به‌طور رسمی به IBM SPSS تغییر یافت.

□ نکته ۲

انواع فایل در SPSS

- فایل داده یا data با پسوند sav.
- فایل خروجی یا output با پسوند spv.
- فایل سینتکس یا syntax برای کد نویسی با پسوند sps.

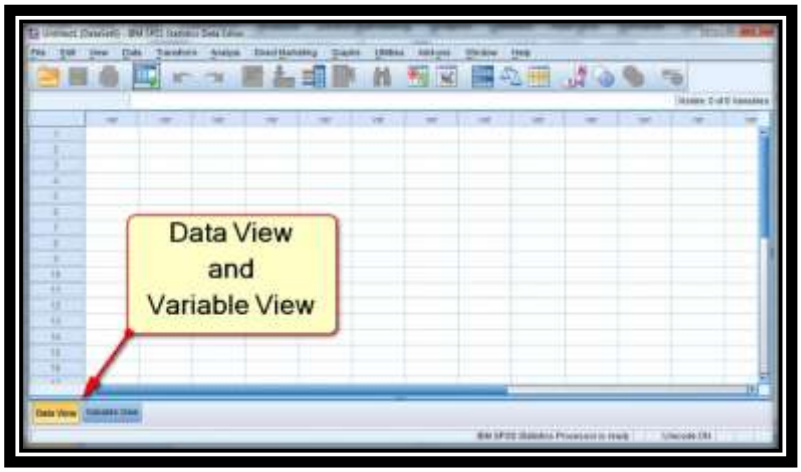
□ نکته ۳

نمایش پسوند فایل‌ها

در ویندوز ۷ مسیر Control Panel→Folder Options→View Tab و سپس گزینه Hide known file types را بررسی نمایید.

□ نکته ٤

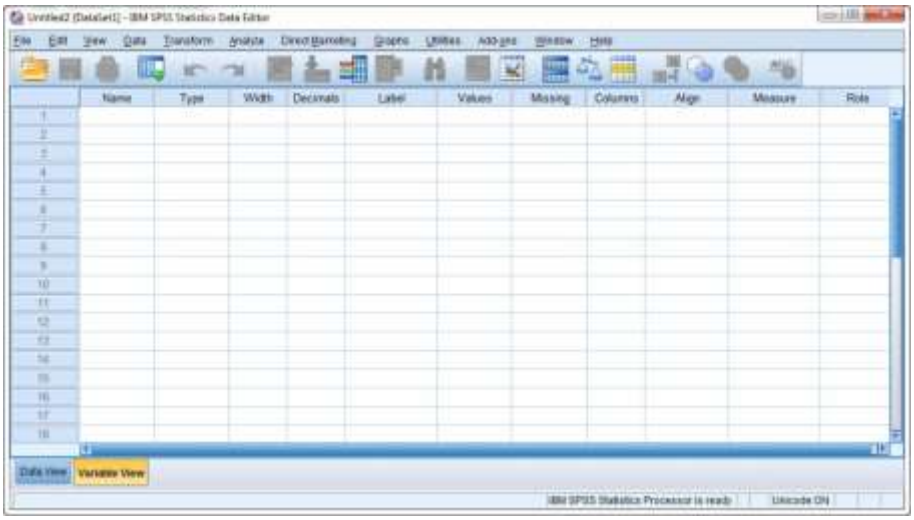
پنجره معرفی داده



شکل ۱: پنجره معرفی داده

□ نکته ۵

پنجره variable view



شکل ۲: پنجره variable View

□ نکته ٦

اجزای پنجره variable view

این پنجره برای تعیین ویژگی‌های متغیرها در فایل داده بوده و شامل ۱۱ قسمت Name، Type، Width، Decimals، Label، Values، Missing، Columns، Align، Measure و Role می‌باشد.

□ نکته ۷

Name

- توضیح مختصر ولی قابل تشخیص از متغیر می‌باشد.
- حداکثر تا ۶۴ کاراکتر مجاز بوده و رها نمودن فضای خالی (spaces) و استفاده از کاراکترهای ویژه (special characters)مجاز نمی‌باشد.
- تنها استفاده از حروف و اعداد توصیه شده است.

□ نکته ۸

Type

- نوع متغیر بوده و می‌تواند عددی، حرفی یا تاریخ باشد.
- پیش‌فرض نوع متغیر در SPSS عددی (Numeric) بوده و تغییر آن توصیه نمی‌شود.
- به‌عنوان‌مثال بجای استفاده از حروف f و m برای تفکیک جنسیت از اعداد ۱ و ۲ استفاده گردد.

□ نکته ۹

Width

- هر چه عدد بزرگ‌تر باشد به پهنای بیشتری نیاز خواهد بود.
- پهنای پیش‌فرض، ۸ می‌باشد.
- چنانچه پهنای انتخاب شده کمتر از میزان مورد نیاز باشد، SPSS اعداد را گرد خواهد کرد.
- تغییر پهنای متغیر تأثیری بر دقت محاسبات نمی‌گذارد.

□ نکته ۱۰

Decimals

- این گزینه تعداد اعشار را در پنجره Data View مشخص می‌سازد.
- تعداد اعشار به‌طور پیش‌فرض، ۲ می‌باشد.
- SPSS محاسبات را تا ۱۶ رقم اعشار بدون توجه به تعداد اعشار مشخص شده انجام می‌دهد.

□ نکته ۱۱

Label

- برچسب عبارتی (phrase) برای توضیح بیشتر متغیر می‌باشد.
- معمولاً چنانچه نام متغیر به‌اندازه کافی گویا (descriptive) باشد، از این گزینه صرف نظر می‌شود.
- SPSS برچسب را تفسیر نمی‌کند و صرفاً در برونداد، کنار نام متغیر گزارش می‌دهد.

□ نکته ۱۲

Values

- برچسب‌های مقداری در متغیرهای اسمی و ترتیبی کاربرد دارند.

- در تفکیک جنسیت، منطقه جغرافیایی یا دین ضروری است از برچسب مقداری استفاده نماییم.
- استفاده از این نوع برچسب برای متغیرهای فاصله‌ای و اسمی ضرورتی ندارد.

□ نکته ۱۳

Missing

- مقادیر ازدست‌رفته هنگامی کاربرد دارد که علت مفقودی مهم می‌باشد.
- به‌صورت گسسته به‌تنهایی یا در ترکیب با گستره‌ای از مقادیر ازدست‌رفته می‌شود استفاده نمود.
- عدد استفاده شده نباید در محدوده داده‌های معتبر باشد.

□ نکته ١٤

Columns

- به‌منظور تنظیم عرض ستون در پنجره View Data کاربرد دارد.
- با کاهش عرض ستون تعداد بیشتری از متغیرها بدون تغییر در نوار پیمایش مشاهده می‌شود.
- توصیه می‌شود عرض ستون کمتر از تعداد کاراکترهای نام متغیر نباشد.

□ نکته ۱۵

Align

- برای تنظیم متن در خانه جدول مشاهده داده (Data View) کاربرد دارد.
- گزینه‌های راست، چپ و وسط‌چین قابل انتخاب هستند.
- پیش‌فرض راست‌چین می‌باشد.

□ نکته ١٦

Measure

- سطح اندازه‌گیری متغیرها مشخص می‌شود .
- گزینه‌های موجود اسمی(Nominal) ، ترتیبی (Ordinal) و نسبی و فاصله‌ای (Scale) هستند.
- تنظیم سطح اندازه‌گیری هنگام ایجاد نمودارهای تعاملی کاربرد دارد.

□ نکته ۱۷

Role

- نقش‌های مختلفی (Input, Target, Both, None, Partion & Split) به متغیرها می‌توان داد.
- پیش‌فرض Input می‌باشد که اجازه می‌دهد متغیر به‌طور کامل در تحلیل وارد شود.
- تغییری در پیش‌فرض داده نمی‌شود.

□ نکته ۱۸

مفاهیم آماری: داده‌ها، متغیرها و موارد

- مشاهدات کمّی یا کیفی ثبت شده که بر اساس آن‌ها نتایج حاصل می‌شوند را داده‌های پژوهش می‌نامند.
- متغیر مقداری است که اندازه آن متفاوت و درعین‌حال قابل‌اندازه‌گیری می‌باشد .

□ نکته ۱۹

مفاهیم آماری: سطوح اندازه‌گیری

- مقیاس یا روش ثبت داده‌ها، سطوح اندازه‌گیری را مشخص می‌سازند .
- سطوح اندازه‌گیری منعکس کننده قابلیت‌های آماری داده‌ها هستند.
- سطوح در SPSS شامل اسمی، ترتیبی و مقیاسی می‌باشد.

□ نکته ۲۰

سطوح اندازه‌گیری: داده‌های اسمی

- داده‌های اسمی تنها برچسب بوده و معنی خاصی ندارند.
- برای تبدیل داده‌های کیفی به کمّی مورد استفاده قرار می‌گیرند.
- در این نوع داده‌ها نمی‌شود میانگین‌گیری نمود ولی فراوانی قابل گزارش می‌باشد .
- جنسیت، تأهل، شاغل بودن و سابقه بیماری داده اسمی هستند.

□ نکته ۲۱

سطوح اندازه‌گیری: داده‌های ترتیبی

- این داده‌ها برای رتبه‌بندی ویژگی خاص استفاده می‌شوند.
- تفاوت مطلق میان رتبه‌ها مشخص نمی‌باشد .
- طیف لیکرت یکی از روش‌های سنجش داده‌های ترتیبی هست .
- احساس درد، میزان رضایت، خشکی یا سفتی جزء داده‌های ترتیبی می‌باشند.

□ نکته ۲۲

سطوح اندازه‌گیری: داده‌های فاصله‌ای و نسبی

- داده‌های فاصله‌ای دارای صفر قراردادی می‌باشند (نمره درسی)
- صفر در داده‌های نسبی مطلق بوده و بیانگر میزان صفت مورد سنجش هست (درآمد)
- SPSS با هر دو داده به یک شیوه (مقیاسی یا Scale) رفتار می‌کند.

□ نکته ۲۳

مفاهیم آماری: طرح‌های وابسته و غیر وابسته

- در طرح‌های وابسته یا همبسته چند ویژگی گروهی خاص (مرد یا زن) در یک مقطع زمانی یا در طول زمان اندازه گرفته می‌شود.
- در طرح‌های غیر وابسته یا غیرهمبسته ویژگی‌های گروه‌های خاص (مرد و زن یا گروه‌های خونی مختلف) بررسی می‌شوند.
- SPSS به شیوه متفاوت داده‌های این طرح‌ها را بررسی می‌کند.

□ نکته ۲٤

مفاهیم آماری: آمار توصیفی

- آمار توصیفی به شرح وضعیت وجود پرداخته و به دلایل تفاوت‌ها توجهی نمی‌کند.
- از این آمار برای خلاصه نمودن داده‌ها در چند شاخص استفاده می‌شود (میانگین، نما، میانه، فراوانی، دامنه، حداقل و حداکثر).

- این آمار به شفاف شدن چارچوب مطالعه یاری می‌رساند.

□ نکته ۲۵

مفاهیم آماری: آمار استنباطی

- روش‌های آماری با توجه به اهداف پژوهش انتخاب می‌شوند.
- پاسخ‌هایی تجربی و متکی به شواهد برای فرضیه‌ها تهیه می‌شود.
- تعیین میزان تفاوت، کیفیت ارتباط و بررسی روابط علّی و معلولی از خروجی‌های آمار استنباطی هستند.

□ نکته ۲۶

مفاهیم آماری: فرضیه‌ها و فرض صفر

- پیش‌بینی نتایج پژوهش بر اساس اطلاعات موجود می‌باشد.
- فرضیه‌ها با نام‌های فرضیه‌های پژوهش (Research Hypothesis)، تجربی (Experimental Hypothesis) و جایگزین (Alternative Hypothesis) شناخته می‌شوند.
- فرضیه صفر برخلاف فرضیه پژوهش طراحی می‌شود.

□ نکته ۲۷

مفاهیم آماری: فرضیه‌های یک و دوطرفه

- در فرضیه‌های یک‌طرفه همزمان معنی‌داری و جهت ارتباط آزمون می‌شود.
- در فرضیه‌های دوطرفه توجه بر وجود ارتباط است.
- آزمون‌های یک‌طرفه به‌ندرت استفاده می‌شوند.

□ نکته ۲۸

مفاهیم آماری: جامعه و نمونه

- جامعه (Population) همه مشاهدات محتمل برای متغیر مورد بررسی را در بر می‌گیرد.
- نمونه (Sample) بخشی از جامعه بوده و نتایج حاصل از آن به جامعه تعمیم داده می‌شود.
- چنانچه عناصر جامعه محدود باشد، تمام شماری انجام می‌شود.

□ نکته ۲۹

مفاهیم آماری: پارامترها و آماره‌ها

- پارامترها مقادیر واقعی ویژگی‌های جامعه می‌باشند.
- آماره‌ها روش‌های آماری برای برآورد یا آزمون پارامترها با استفاده از اطلاعات نمونه هستند.
- پارامترها ناشناخته بوده، در نتیجه دقت برآورد اهمیت می‌یابد.

□ نکته ۳۰

مفاهیم آماری: احتمالات

- نمونه تا چه اندازه نماینده‌ای واقعی از جامعه مورد بررسی می‌باشد.
- تفاوت مشاهده شده تا چه اندازه به شانسی بودن فرآیند نمونه‌گیری مربوط است.
- در آزمون‌های آماری میزان احتمال یا مقدار p گزارش می‌شود.

□ نکته ۳۱

مفاهیم آماری: معنی‌داری

- p میزان احتمال اینکه نتیجه به چه میزان می‌تواند شانسی باشد را مشخص می‌سازد (صحت فرض صفر).
- چنانچه p محاسباتی کمتر از پنج درصد باشد، شانسی بودن نتایج را نمی‌توان قبول کرد.
- با رد فرضیه صفر، نتایج معنی‌دار (Significant) گزارش می‌شوند.

□ نکته ۳۲

مفاهیم آماری: معنی‌داری مجانبی

- معنی‌داری مجانبی (Asymptotic) در مقابل یک توزیع مجانبی آزمون می‌شود.
- توزیع مجانبی توزیعی است که احتمال مقادیر بسیار بالا یا بسیار پائین هیچ‌گاه صفر نمی‌شود.
- توزیع نرمال یک توزیع مجانبی است.

□ نکته ۳۳

مفاهیم آماری: معنی‌داری واقعی

- استفاده از توزیع‌های مجانبی برای نمونه‌های بزرگ پیشنهاد شده است.
- برای نمونه‌های کوچک از معنی‌داری واقعی استفاده می‌شود (مانند آزمون دقیق فیشر یا Fisher's exact test).
- در تحلیل نتایج به استفاده از آزمون واقعی اشاره می‌گردد.

□ نکته ۳٤

مفاهیم آماری: سطوح اطمینان

- استفاده از سطوح اطمینان به‌جای برآوردهای نقطه‌ای توصیه شده است.
- سطوح اطمینان دارای حد بالا و حد پائین هستند.
- حدود بالا و پائین دامنه‌ای که انتظار می‌رود مقدار واقعی در آن قرار گیرد را تعیین می‌کنند.

□ نکته ۳۵

مفاهیم آماری: مثال سطوح اطمینان

- چنانچه در سطح اطمینان ۹۵/۰، حد پائین و بالای متغیری به ترتیب ۲ و ۳ باشد، در آن صورت احتمال قرارگیری پارامتر مورد بررسی در دامنه مذکور ۹۵ درصد می‌باشد.
- یعنی چنانچه ۱۰۰ بار نمونه‌گیری انجام داده شود، در ۹۵ مورد پارامتر در محدوده ۲ و ۳ قرار خواهد گرفت.

□ نکته ۳٦

مفاهیم آماری: مزایای کاربرد سطوح اطمینان

- دامنه برآورد از مقدار واقعی پارامتر مشخص می‌گردد.
- با استفاده از سطوح اطمینان می‌توان میزان همپوشانی متغیرهای مختلف را هنگام رسم میانگین‌ها بررسی نمود.
- امکان آزمون فرضیه به‌سادگی فراهم می‌شود.

□ نکته ۳۷

مفاهیم آماری: اندازه اثر (effect size)

- اندازه اثر، میزان تأثیر متغیر مستقل بر متغیر وابسته را اندازه‌گیری می‌کند (اندازه تفاوت میانگین‌ها).
- معنی‌داری همواره به معنی اثرگذاری قابل‌توجه نمی‌باشد. این مطلب با اندازه اثر سنجیده می‌شود.
- d کوهن (Cohen's d) برای محاسبه اندازه اثر بکار می‌رود.

□ نکته ۳۸

مفاهیم آماری: d کوهن

۰/۲، ۰/۵ و ۰/۸ به ترتیب اندازه اثر کم، متوسط و زیاد شناخته می‌شوند. چنانچه انحراف معیار دو گروه ۱۲ و ۱۳ تایی به ترتیب ۵/۴۸ و ۵/۹۱ و تفاوت میانگین آن‌ها۷/۸۸- باشد، در آن صورت d کوهن به شکل زیر محاسبه می‌شود:

الف) محاسبه میانگین وزنی انحراف معیارها:

$$\frac{(5.484 * 12) + (5.911 * 13)}{12 + 13} = 5.706$$

ب) تعیین d کوهن:

$$\text{Cohen's d} = \frac{-7.878}{5.706} = -1.38$$

میانگین مشاهدات گروه اول بیش از یک برابر انحراف معیار از گروه دوم کمتر بوده و اندازه اثر به‌طور قابل‌توجهی بزرگ است.

□ نکته ۳۹

مفاهیم آماری: اندازه اثر (نکات)

- در مداخلات حیاتی اندازه اثر کم، بسیار با ارزش می‌تواند باشد.
- اندازه اثر با هزینه انجام مداخله مقایسه می‌شود.
- گزارش اندازه اثر در نتایج تحلیلی آمار استنباطی هنگام معنی‌داری متغیرها توصیه شده است.

□ نکته ٤٠

مفاهیم آماری: توان آماری

- توان آماری (Statistical Power) میزان قدرت آزمون در تشخیص بین صحت یا اشتباه بودن فرض صفر را تعیین می‌کند.
- توان آماری تحت تأثیر اندازه اثر و حجم نمونه می‌باشد.
- چنانچه فرض صفر رد نشود می‌توان آن را به صحت فرض یا پائین بودن توان آماری نسبت داد.

□ نکته ٤١

مفاهیم آماری: تساوی عملی

- چنانچه اجرا پذیری میدانی نتیجه اهمیت داشته باشد، در آن صورت موضوع تساوی عملی (Equivalence) مطرح می‌گردد.
- با عدم تأمین حداقل میزان اندازه اثر مورد نیاز، روش جدید با وجود معنی‌داری، پذیرفته‌نشده و گزارش می‌شود تساوی عملی وجود دارد.

□ نکته ٤٢

مفاهیم آماری: تساوی عملی (نکات)

- چنانچه توان آماری برای اندازه اثر مطلوب حدود ۰/۸ باشد، گزارش می‌شود میان روش‌ها تفاوت عملی وجود ندارد.
- اگر توان آماری مورد نیاز برای اندازه اثر درخواستی کمتر از ۵/۰ هست، گزارش می‌شود ارزشمندی روش جدید هنوز نامشخص می‌باشد.

□ نکته ٤٣

مفاهیم آماری: آزمون‌های پارامتری و نا پارامتری

- آزمون‌های پارامتری (Parametric) و نا پارامتری (Non Parametric) جزء آمار استنباطی هستند.
- آزمون‌های پارامتری توان آماری بیشتری نسبت به آزمون‌های نا پارامتری دارند.
- آزمون‌های نا پارامتری به پیش‌فرض‌های کمتری نیاز دارند.

□ نکته ٤٤

مفاهیم آماری: آزمون‌های پارامتری (شرایط لازم)

- داده‌ها در مقیاس فاصله‌ای یا نسبی اندازه‌گیری شده‌اند.
- توزیع احتمالی داده‌ها تصادفی، نرمال یا طبیعی هست.
- گروه‌ها یا نمونه‌های مورد مقایسه دارای واریانس برابر می‌باشند (واریانس همسان).
- دقت پیش‌بینی، وابسته به سطح متغیر مستقل نیست.

□ نکته ٤٥

وارد کردن داده در IBM SPSS

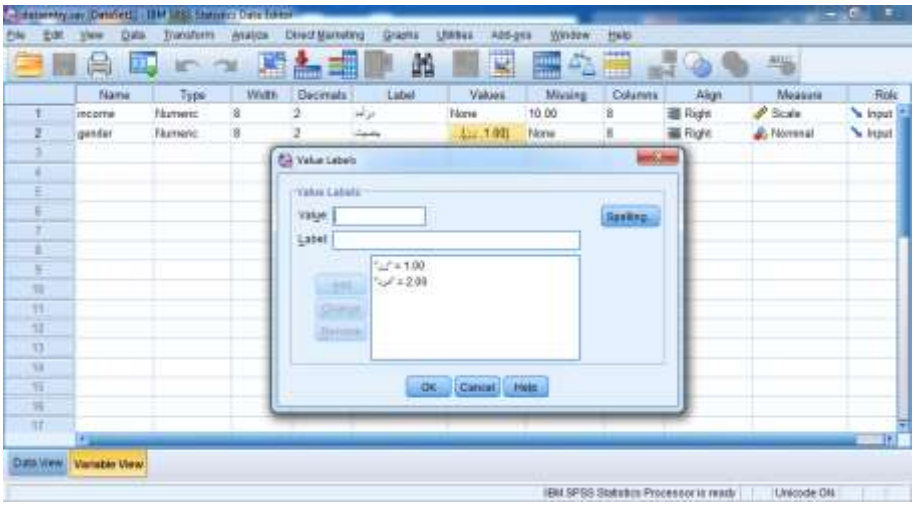
- با فعال کردن SPSS در رایانه ابتدا شکل ۳، ظاهر می‌شود.
- در این مرحله امکان انتخاب فایل‌های قبلی از قسمت Open another file در Recent Files یا وارد کردن مستقیم داده با انتخاب New Dataset در New Files وجود دارد.
- با انتخاب New Dataset پنجره Data Editor ظاهر می‌شود.

شکل ۳: نحوه انتخاب فایل‌های قبلی یا وارد کردن مستقیم داده

□ نکته ٤٦

وارد کردن داده در IBM SPSS (مثال)

- درآمد ماهانه ۴۰ نفر زن و مرد جمع‌آوری‌شده است.
- درآمد با income و جنسیت با gender مشخص شده است.
- داده‌های درآمد، نسبی و جنسیت، اسمی (۱ و ۲) می‌باشد.
- داده‌های مورد نظر در فایل dataentry.sav، آمده است.
- شکل ۴، نمایی از Variable View می‌باشد.



شکل ۴: معرفی ویژگی‌های داده‌های مورد بررسی

□ نکته ٤٧

آمار توصیفی (Descriptive Statistics)

- معیارهای وضعیت نسبی: فراوانی، رتبه درصدی، صدک ها و چارک‌ها
- معیارهای گرایش مرکزی: میانگین، میانه و نما
- معیارهای پراکندگی: دامنه تغییرات، واریانس و انحراف معیار
- کجی: در توزیع متقارن کجی صفر هست. چنانچه دامنه توزیع به سمت راست پخش شده باشد، توزیع دارای کجی مثبت و در حالت عکس، کجی منفی خواهد بود.
- کشیدگی: بیانگر میزان بلندا و پهنای توزیع می‌باشد.

□ نکته ٤٨

فراوانی (Frequencies)

- با دستور فراوانی مقادیر متغیر، طبقه‌بندی می‌شود.
- در جدول فراوانی، فراوانی نسبی (Percent)، درصد معتبر (Valid Percent) و درصد تجمعی (Cumulative Percent) ارائه می‌شود.
- به‌منظور اجرای دستور مراحل زیر انجام داده می‌شود:

Analyze → Descriptive Statistics → Frequencies

□ نکته ٤٩

فراوانی: تعاریف

- فراوانی: تعداد مشاهدات هر مورد یا گروه را مشخص می‌سازد.
- درصد معتبر: نسبت تعداد مشاهدات هر گروه به کل داده‌های معتبر (کل مشاهدات به‌جز داده‌های مفقود) می‌باشد.
- درصد تجمعی: تجمیع پیوسته‌ای از درصدهای معتبر می‌باشد و برای داده‌های اسمی کاربرد ندارد.

□ نکته ٥٠

فراوانی: مثال

الف) برای متغیر طبقه‌ای یا CATEGORICAL VARIABLES جنسیت که با gender مشخص شده است.

ب) برای متغیر کمّی یا QUANTITATIVE VARIABLES درآمد که با income تعریف شده است.

شکل ۵: تعریف متغیر طبقه‌ای

جدول ۱: نتایج

جنسیت

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	زن	20	50.0	50.0	50.0
	مرد	20	50.0	50.0	100.0
	Total	40	100.0	100.0	

□ نکته ۵۱

فراوانی: متغیر طبقه‌ای (گزارش دهی)

تعداد نمونه مورد مطالعه ۴۰ نفر بود که از میان آن‌ها ۲۰ نفر زن و ۲۰ نفر مرد می‌باشند. داده مفقود مشاهده نشده، لذا درصد معتبر با فراوانی نسبی تفاوتی ندارد. در مطالعه حاضر سهم زنان و مردان از نمونه مساوی و هر یک برابر با ۵۰ درصد می‌باشد.

□ نکته ۵۲

فراوانی: متغیر کمّی (QUANTITATIVE VARIABLES)

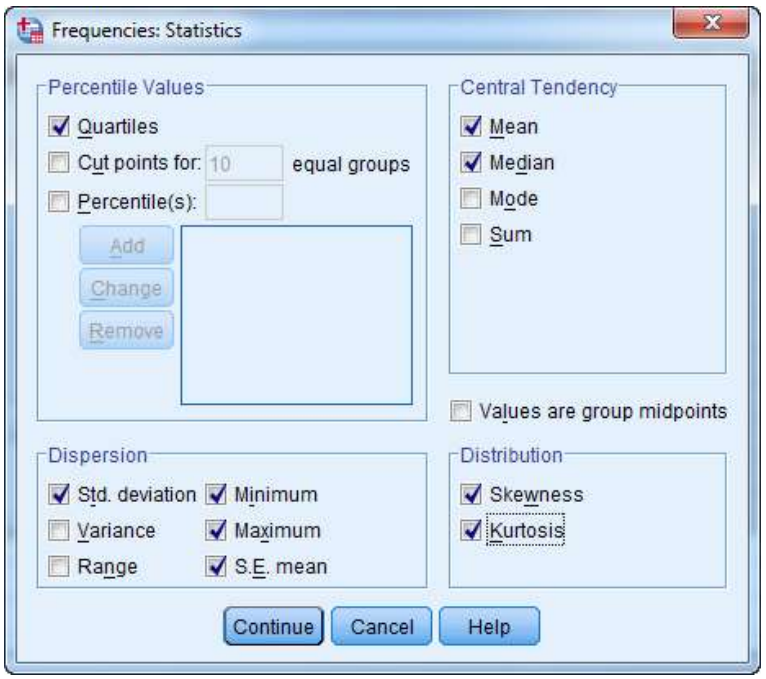
- فایل frequencies.sav را باز کنید.
- از منوی اصلی (main menu) مسیر زیر را انتخاب نمایید.

Analyze ➔Descriptive Statistics ➔ Frequencies

- متغیر income را به پانل متغیر(ها) یا Variable(s) panel به‌صورت شکل زیر انتقال دهید.

شکل ۶: فراوانی متغیر کمّی

- Display frequency tables را انتخاب نمایید.
- دکمه (pushbutton) آماره‌ها یا Statistics را انتخاب کنید.
- کادر گفتگوی آماره‌ها (Statistics screen) به‌صورت شکل زیر ظاهر می‌شود.



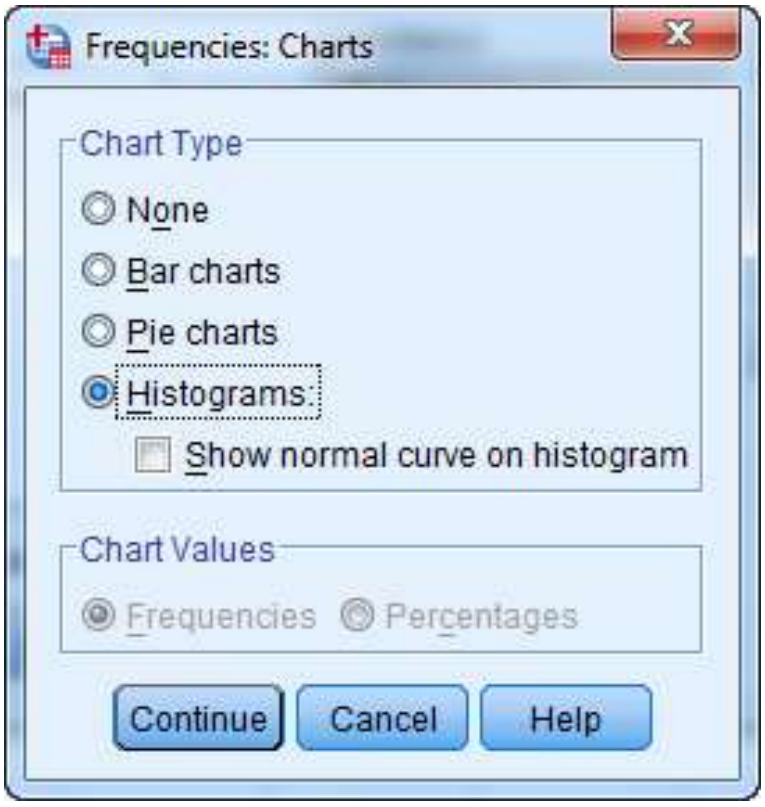
شکل ۷: کادر گفتگوی آماره‌ها

- در سمت چپ و بالا زیر عنوان Percentile Values، چارک‌ها (Quartiles) را انتخاب نمایید.
- سمت راست و بالا زیر عنوان گرایش مرکزی یا Central Tendency، میانگین (Mean) و میانه (Median) را انتخاب کنید.
- زیر عنوان پراکندگی یا Dispersion، انحراف معیار (Std. deviation)، حداقل (Minimum)، حداکثر (Maximum) و انحراف معیار میانگین (S.E. mean) را انتخاب نمایید.
- زیر عنوان توزیع یا Distribution، کجی یا چولگی (Skewness) و کشیدگی (Kurtosis) را انتخاب کنید.

□ نکته ۵۳

فراوانی: نمودارها (charts)

- در کادر گفتگوی آماره‌ها، دکمه ادامه یا Continue را انتخاب نمایید تا به کادر گفتگوی اصلی (main dialog window) بازگردید.
- دکمه نمودارها (Charts) را انتخاب و در پنجره نمودارها (Charts window)، هیستوگرام (Histograms) نمایید.



شکل ۸: پنجره نمودارها

- با انتخاب دکمه ادامه یا Continue ابتدا به پنجره اصلی گفتگو برگشته و سپس با تائید OK تحلیل نتایج آغاز می‌شود.
- آمار توصیفی متغیر درآمد به‌صورت زیر در پنجره مشاهده گر (Viewer Window) نمایش داده می‌شود.

جدول ۲: آمار توصیفی متغیر کمّی

Statistics

درآمد

N	Valid	38
	Missing	2
Mean		6.4150
Std. Error of Mean		.21210
Median		6.6300
Std. Deviation		1.30745
Skewness		-.018
Std. Error of Skewness		.383
Kurtosis		-.961
Std. Error of Kurtosis		.750
Minimum		4.43
Maximum		9.00
Percentiles	25	5.2700
	50	6.6300
	75	7.2050

□ نکته ٥٤

آمار توصیفی متغیر کمّی: خطای معیار میانگین

خطای معیار میانگین (Std. Error of Mean) از تقسیم انحراف معیار داده‌ها (Std. Deviation) به ریشه دوم (Square Root) تعداد مشاهدات معتبر (N) محاسبه می‌شود:

$$\text{Std. Error of Mean} = \frac{1.30745}{\sqrt{38}} = 0.21210$$

□ نکته ٥٥

خطای معیار میانگین: کاربرد

- خطای معیار میانگین، برآوردی از انحراف معیار میانگین‌های نمونه‌های فرضی تصادفی N تایی از جامعه می‌باشد.
- توزیع احتمالی میانگین، نرمال فرض می‌شود.
- خطای معیار میانگین در محاسبه فاصله اطمینان (Confidence Interval) کاربرد دارد.

□ نکته ٥٦

خطای معیار میانگین: مثال

فاصله اطمینان ۹۵ درصدی برای میانگین درآمد به‌صورت زیر محاسبه می‌شود:

Lower confidence value=6.415- (1.96*0.2121)=5.9993

Upper confidence value= 6.415+ (1.96*0.2121)=6.8307

۱/۹۶ مقدار z برای ۹۵ درصد بوده و ۶/۴۱ میانگین درآمد و ۰/۲۱ خطای معیار میانگین است.

□ نکته ٥٧

خطای معیار میانگین: تحلیل

- سطح پائین اطمینان (Lower confidence value) و سطح بالای اطمینان (Upper confidence value) برای میانگین درآمد به ترتیب ۵/۹۹ و ۶/۸۳ می‌باشد.
- چنانچه نمونه گیری به‌دفعات انجام شود در ۹۵ درصد موارد میانگین در فاصله ۵/۹۹ الی ۶/۸۳ قرار خواهد گرفت.

□ نکته ٥٨

کجی: توضیح

- کجی یا Skewness میزان عدم تقارن توزیع را نشان می‌دهد.
- منحنی نرمال دارای کجی صفر می‌باشد.
- چنانچه بیشتر اعداد نمونه کوچک باشند، کجی مثبت (Positive Skewness) و دنباله توزیع (The "tail" of the distribution) به سمت مقادیر مثبت محور X خواهد بود.

□ نکته ٥٩

کجی: مثال

- اگر بیشتر اعداد نمونه ارقام بزرگی هستند، کجی منفی (Negative Skewness) و دنباله توزیع به سمت مقادیر منفی محور X خواهد بود.
- کجی متغیر درآمد، ۰۱۸/- می‌باشد که بیانگر کجی منفی توزیع احتمالی است.

□ نکته ۶۰

کشیدگی: توضیح

- کشیدگی (Kurtosis) میزان فشردگی (Compression) یا پهنی (Flattening) توزیع را نسبت به منحنی نرمال (Normal Curve) مشخص می‌سازد.
- کشیدگی منحنی نرمال صفر بوده و mesokurtosis نامیده می‌شود.

□ نکته ۶۱

کشیدگی مثبت

- کشیدگی مثبت (Positive Kurtosis) هنگامی است که توزیع نسبت به منحنی نرمال حول مرکز فشردگی (more compressed toward the center) بیشتری دارد.
- کشیدگی مثبت، leptokurtosis نامیده می‌شود.

□ نکته ۶۲

کشیدگی منفی

- زمانی که توزیع نسبت به منحنی نرمال پهن‌تر باشد، گفته می‌شود کشیدگی منفی (Negative Kurtosis) وجود دارد.
- کشیدگی منفی به نام platykurtosis شناخته می‌شود.
- کشیدگی مثال فوق ۰/۹۶- محاسبه شده است.

□ نکته ۶۳

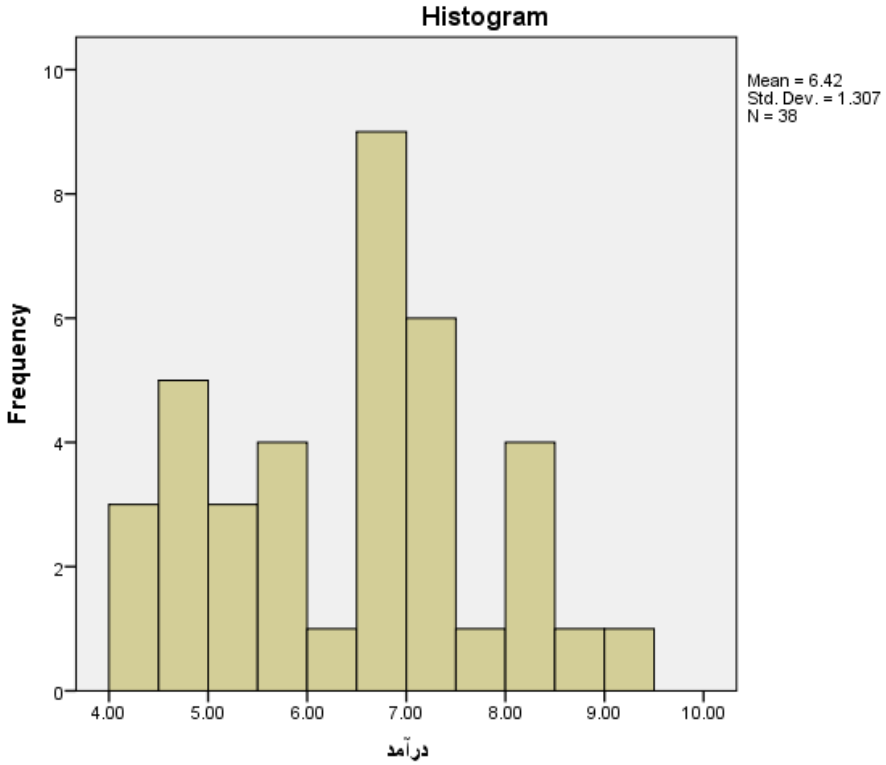
کشیدگی: تحلیل

- چنانچه کشیدگی در فاصله ۱- تا ۱+ باشد، گفته می‌شود توزیع بیش از اندازه کشیدگی ندارد.
- اگر کشیدگی در فاصله ۰/۵- تا ۰/۵+ باشد، محققین از عبارت a bit of compression برای توضیح وضعیت استفاده می‌کنند.
- برای کشیدگی ۱- یا ۱+ واژه mild compression کاربرد دارد.

□ نکته ٦٤

کشیدگی: مثال

- کشیدگی متغیر درآمد ۰/۹۶- است، لذا کشیدگی منفی داشته و نسبت به محنی نرمال پهن‌تر می‌باشد.
- خطای معیار کشیدگی، ۰/۷۵ است که در فاصله اعتماد ۹۵ درصدی، دامنه پائین و بالای آن (۰/۹۶۱ ± × ۰/۷۵۰ ×۱/۹۶)، ۲/۴۳- و ۰/۵۱ خواهد بود.



شکل ۹: هیستوگرام

جدول ۳: جدول فراوانی

درآمد

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	4.43	3	7.5	7.9	7.9
	4.54	2	5.0	5.3	13.2
	4.65	1	2.5	2.6	15.8
	4.76	2	5.0	5.3	21.1
	5.12	1	2.5	2.6	23.7
	5.32	1	2.5	2.6	26.3
	5.41	1	2.5	2.6	28.9
	5.87	4	10.0	10.5	39.5
	6.00	1	2.5	2.6	42.1
	6.54	2	5.0	5.3	47.4
	6.62	1	2.5	2.6	50.0
	6.64	1	2.5	2.6	52.6

ویدیوی ۱: فراوانی متغیر کمی

قابل دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ٦٥

آمار توصیفی متغیر کمّی: گزارش دهی

از ۴۰ نمونه جمع‌آوری‌شده، ۳۸ مورد معتبر (Valid) و ۲ مورد مفقوده (Missing) هستند. میانگین مشاهدات معتبر ۶/۴۱ می‌باشد که با توجه به خطای معیار میانگین ۰/۰۲۱ حد پائین و حد بالای فاصله اعتماد ۹۵ درصدی به ترتیب ۵/۹۹ و ۶/۸۳ خواهد بود. میانه مشاهدات ۶/۶۳ بوده و انحراف معیار آن‌ها ۱/۳۱ می‌باشد. کجی متغیر درآمد، ۰/۰۱۸- بوده که بیانگر تقارن نسبی توزیع احتمالی می‌باشد (در محدوده ۱± قرار گرفته است) ولی با کشیدگی ۰/۹۶-، کشیدگی منفی، محسوس هست. حداقل و حداکثر داده‌ها نیز به ترتیب ۴/۴۳ و ۹ می‌باشد.

□ نکته ٦٦

آماره‌های توصیفی (Descriptives)

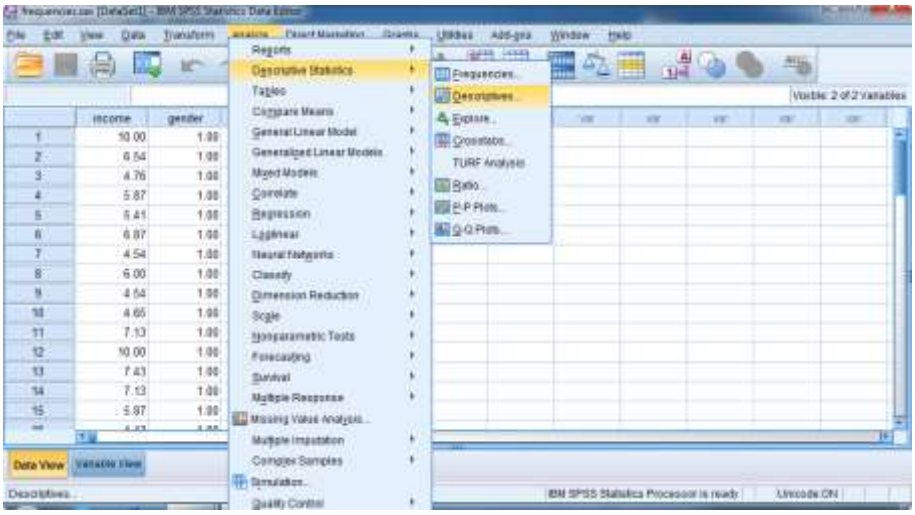
- آماره‌های توصیفی برای متغیرهای کمّی (Quantitative Variables) کاربرد دارد.

- آماره‌های این قسمت به‌جز درصدها (Percentiles) و گراف‌ها (Graphic Output) مشابه قسمت فراوانی (Frequencies) است.
- امکان تبدیل متغیرها به مقادیر استاندارد وجود دارد.

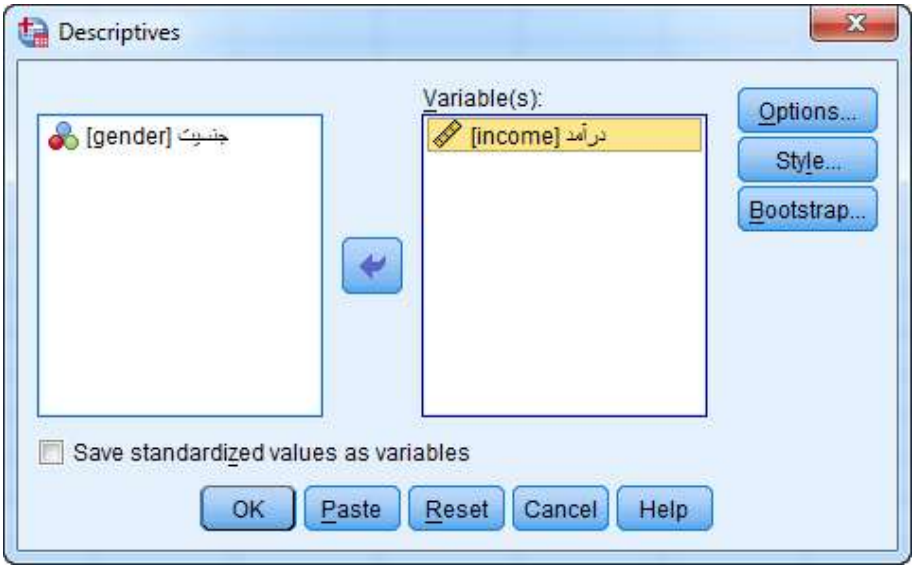
□ نکته ٦٧

آماره‌های توصیفی: مثال

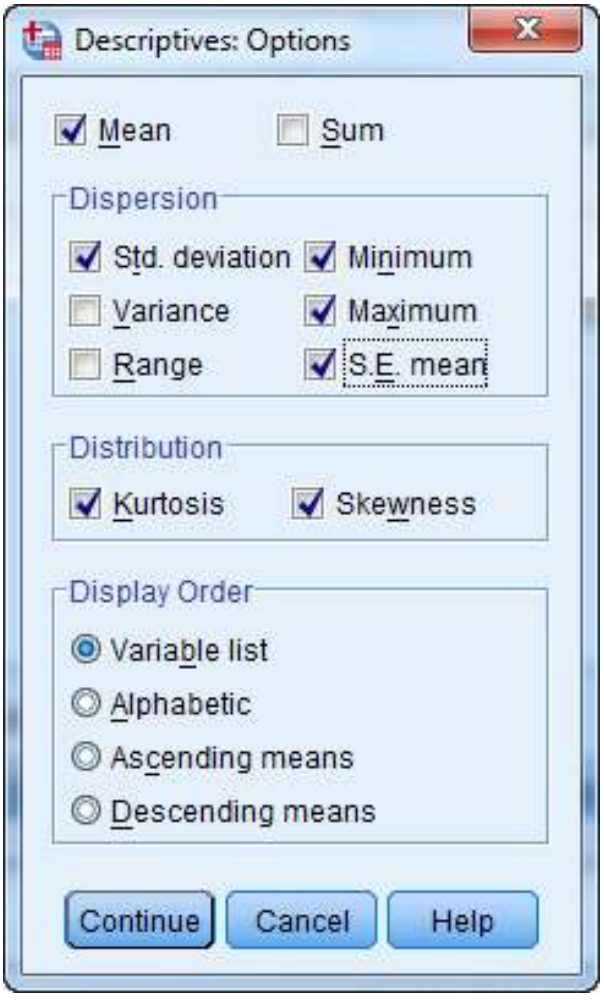
- فایل frequencies.sav را باز کنید.
- از مسیر زیر متغیر درآمد (income) را انتخاب نمایید.
- Analyze ➔Descriptive Statistics ➔Descriptives
- دکمه گزینه‌ها (Options pushbutton) را برگزینید.
- آماره‌های میانگین، انحراف معیار، حداقل، حداکثر، خطای معیار میانگین، کشیدگی و کجی را انتخاب نمایید.



شکل ۱۰: مسیر دستور آمار توصیفی



شکل ۱۱: انتخاب متغیر آمار توصیفی



شکل ۱۲: انتخاب آماره‌های توصیفی

جدول ۴: نتایج توصیفی

Descriptive Statistics

	N	Minimum	Maximum	Mean		Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error	Statistic	Std. Error
درآمد income	38	4.43	9.00	6.4150	.21210	1.30745	-.018	.383	-.961	.750
Valid N (listwise)	38									

☐ نکته ٦٨

آماره‌های توصیفی: تحلیل

بر اساس اطلاعات جدول می‌توان گفت که از ۴۰ داده موجود ۳۸ مورد معتبر بوده و حداقل و حداکثر آن به ترتیب ۴/۴۳و ۹ هست. میانگین متغیر ۶/۴۱ با خطای معیار ۰/۲۱ می‌باشد. انحراف معیار متغیر ۱/۳۱ بوده و دارای توزیعی به نسبت متقارن با کشیدگی منفی است.

ویدیوی ۲: آماره‌های توصیفی

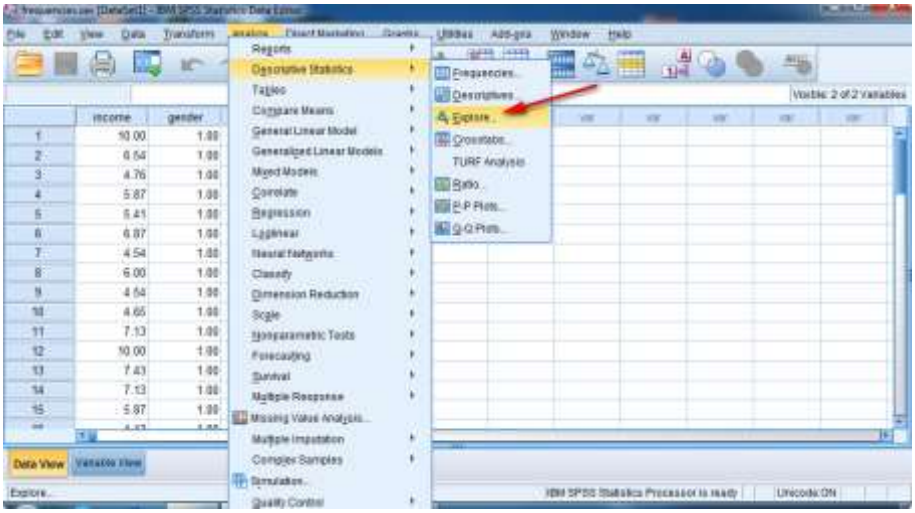
قابل دسترس در سی دی و کانال تلگرام @drshahnavazi

☐ نکته ٦٩

اکتشاف

- فرمان اکتشاف یا Explore برای داده‌های کمّی که از طرح‌های مستقل جمع‌آوری‌شده‌اند، کاربرد دارد.
- امکان توصیف داده‌ها به تفکیک گروه‌های مختلف را فراهم می‌سازد.
- برای استفاده از امکانات این فرمان مسیر زیر را انتخاب نمایید:

Analyze ➔Descriptive Statistics ➔ Explore



شکل ۱۳: مسیر دستور اکتشاف

□ نکته ۷۰

اکتشاف: Statistics

- در این قسمت حد پائین و بالای میانگین با انتخاب Descriptives و تعیین فاصله اطمینان برای میانگین (Confidence Interval for Mean) محاسبه می‌شود.
- با انتخاب Percentiles مجموعه‌ای از صدک ها به همراه چارک‌ها یا Quartiles گزارش می‌شود.

□ نکته ۷۱

اکتشاف: Plots

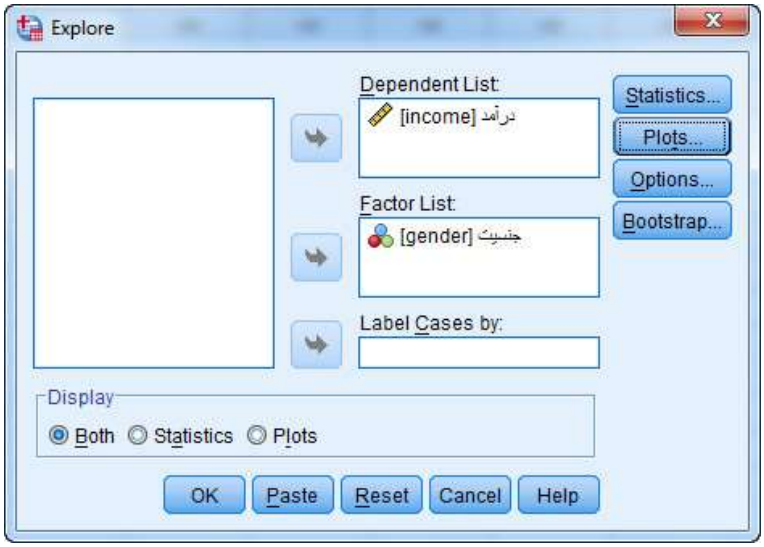
- در این قسمت امکان ترسیم نمودارهای جعبه‌ای و ویسکرز (box and whiskers plot) وجود دارد.
- نمودارهای جعبه‌ای برای بررسی نحوه توزیع و شناسایی داده‌های پرت و افراطی کاربرد دارد.
- این نمودارها توسط توکی در سال ۱۹۷۷ معرفی شدند.

□ نکته ۷۲

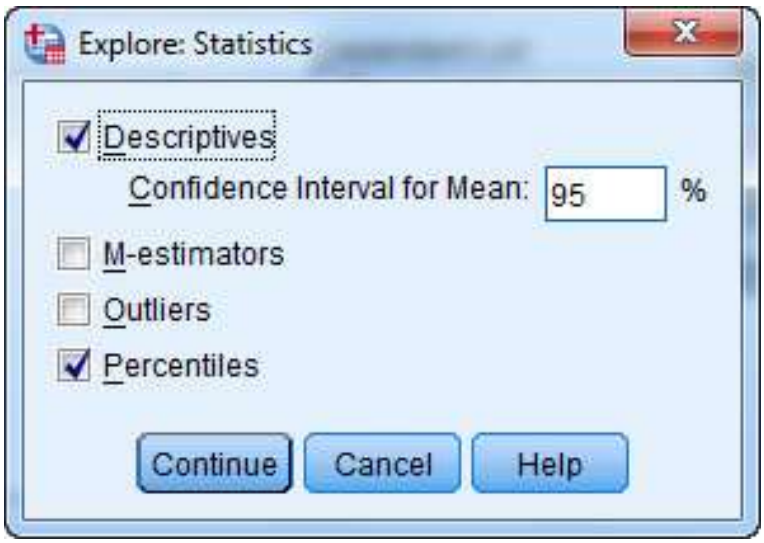
اکتشاف: اصطلاحات

- فاصله چارکی یا Interquartile Range (IQR)، اختلاف مطلق چارک اول (درصد ۲۵ ام) با چارک سوم (درصد ۷۵ ام) است.

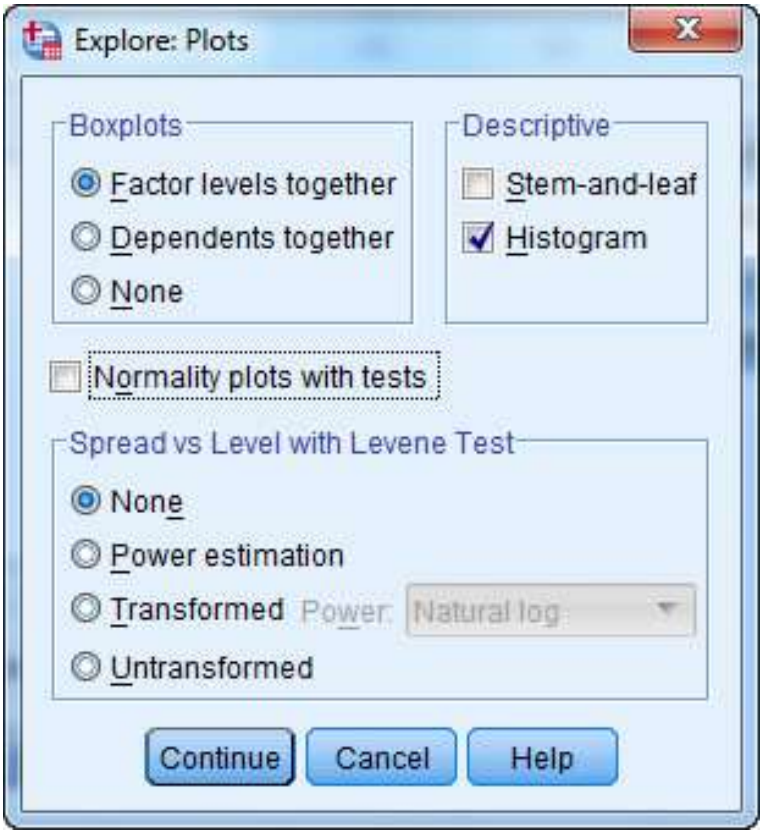
- میانگین اصلاح شده ۵ درصدی (5% Trimmed Mean) با حذف پنج درصد داده‌ها از هر طرف محاسبه می‌شود.
- Exclude cases pairwise امکان انجام محاسبه با حداکثر داده معتبر را فراهم می‌سازد.



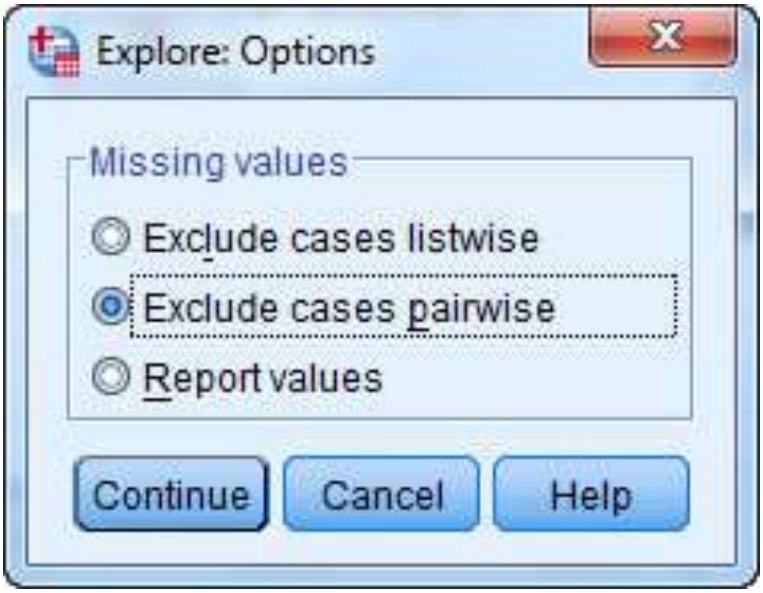
شکل ۱۴: انتخاب متغیرهای اکتشافی



شکل ۱۵: انتخاب آماره‌های اکتشافی



شکل ۱۶: انتخاب گراف‌های اکتشافی



شکل ۱۷: گزینه‌های اکتشافی

ویدیوی ۳: اکتشاف

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

جدول ۵: نتایج اکتشافی

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	جنسیت	N	Percent	N	Percent	N	Percent
درآمد	زن	18	90.0%	2	10.0%	20	100.0%
	مرد	20	100.0%	0	0.0%	20	100.0%

□ نکته ۷۳

نتایج دستور اکتشاف: خلاصه (تحلیل)

از ۴۰ نمونه مورد مطالعه، ۳۸ داده معتبر (Valid) و ۲ داده مفقوده (Missing) مشاهده شده است. موارد مفقود متعلق به گروه زنان بوده در نتیجه درصد معتبر این گروه ۹۰ درصد محاسبه شده است. درصد داده‌های مفقود گروه زنان و مردان به ترتیب ۱۰ و ۰ درصد می‌باشد.

بررسی داده‌های پژوهش به تفکیک گروه‌های مختلف نشان می‌دهد که میانگین درآمد زنان ۵/۶۸ بوده و کمتر از میانگین درآمد مردان با ۷/۰۸ می‌باشد. میانگین اصلاح شده ۵ درصدی زنان و مردان به ترتیب ۵/۶۵ و ۷/۱۲ است که تفاوت چندانی با میانگین اصلی ندارد. عدم وجود تفاوت قابل‌توجه، دلیلی بر عدم وجود داده‌های پرت (Outlier) می‌باشد.

دامنه (Range) که بیانگر فاصله حداقل و حداکثر داده‌هاست در زنان ۳ و در مردان ۴/۵۷ می‌باشد، یعنی پراکنش درآمد در مردان بیشتر از زنان است. میانه در گروه زنان ۵/۶۴ و در گروه مردان ۷/۰۲ بوده و داده‌های گروه زنان کجی مثبت و گروه مردان کجی منفی نشان می‌دهند. فاصله چارکی در زنان و مردان نیز به ترتیب ۲ و ۱/۴۸ شده است.

جدول ۵: نتایج آمار توصیفی اکتشافی

Descriptives

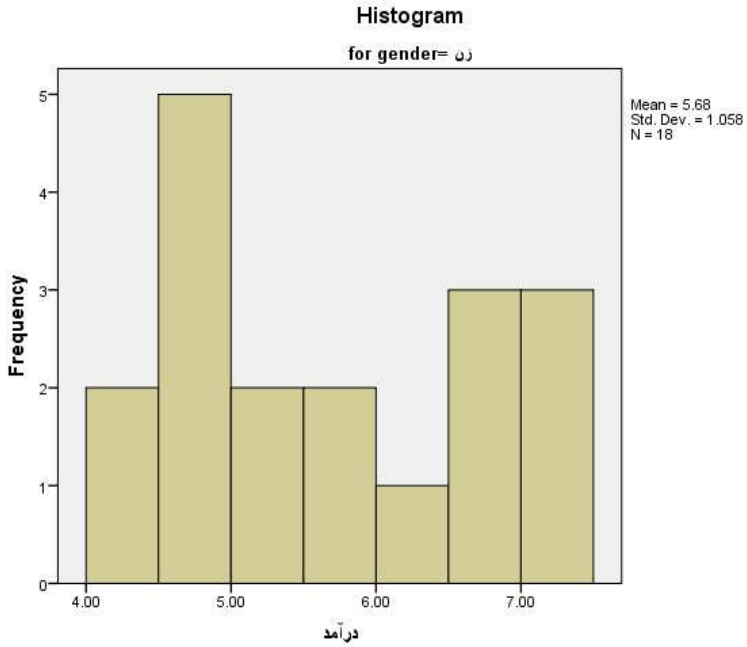
Dependent Variables	gender	Statistics		Stat Type: Statistic	Stat Type: Std. Error
income	زن	Mean		5.6789	.24930
		95% Confidence Interval for Mean	Lower Bound	5.1529	
			Upper Bound	6.2049	
		5% Trimmed Mean		5.6510	
		Median		5.6400	
		Variance		1.119	
		Std. Deviation		1.05768	
		Minimum		4.43	
		Maximum		7.43	
		Range		3.00	
		Interquartile Range		2.00	
		Skewness		.276	.536
		Kurtosis		-1.456	1.038
	مرد	Mean		7.0775	.26006
		95% Confidence Interval for Mean	Lower Bound	6.5332	
			Upper Bound	7.6218	
		5% Trimmed Mean		7.1178	
		Median		7.0200	
		Variance		1.353	
		Std. Deviation		1.16301	
		Minimum		4.43	
		Maximum		9.00	
		Range		4.57	
		Interquartile Range		1.48	
		Skewness		-.523	.512
		Kurtosis		.103	.992

جدول ۶: صدک ها

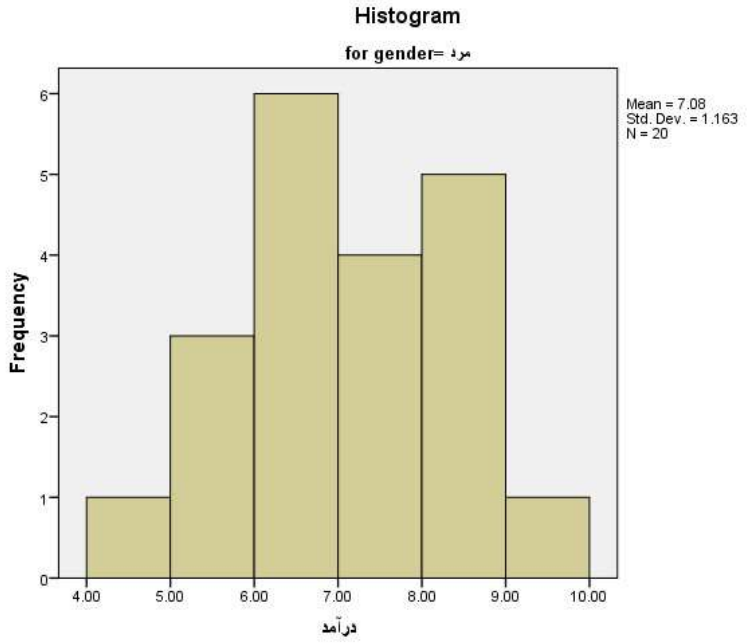
Percentiles

		جنسیت	Percentiles: 5	10	25	50	75	90	95
Weighted Average (Definition 1)	درآمد	زن	4.4300	4.4300	4.6225	5.6400	6.6225	7.1600	.
		مرد	4.4645	5.1950	6.6250	7.0200	8.1050	8.5230	8.9780
Tukey's Hinges	درآمد	زن			4.6500	5.6400	6.5400		
		مرد			6.6300	7.0200	8.1000		

در ردیف نخست جدول صدک های ۵، ۱۰، ۲۵، ۵۰، ۹۰ و ۹۵ گروه‌های زنان و مردان به تفکیک محاسبه شده است. صدک ۲۵ ام مردان نشان می‌دهد که ۲۵ درصد افراد، درآمدی کمتر از ۶/۶۲ دارند. ردیف دوم چارک‌های اول، دوم (میانه) و سوم توکی هستند (Tukey’s Hinges) که در بررسی داده‌های پرت کاربرد دارند.



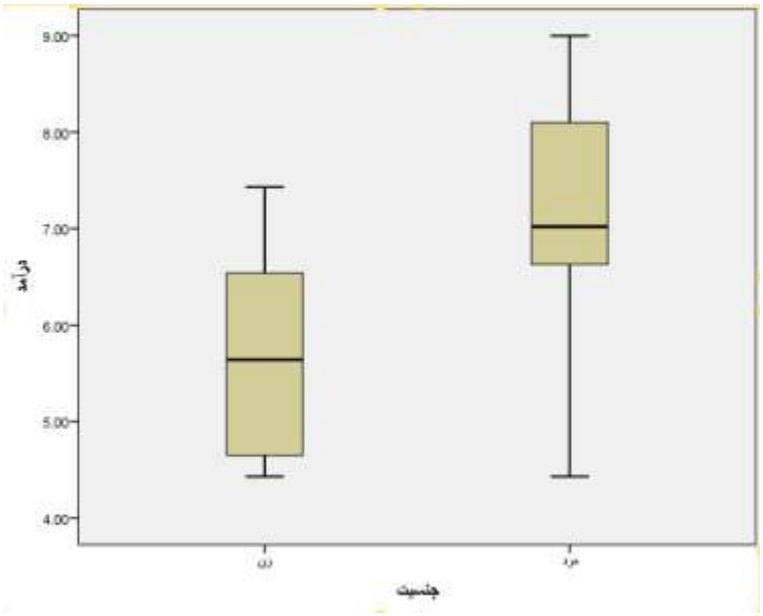
شکل ۱۸: هیستوگرام زنان



شکل ۱۹: هیستوگرام مردان

هیستوگرام گروه زنان نشان می‌دهد که فراوانی داده‌ها بیشتر در مقادیر پائین بوده، درنتیجه کجی مثبت مشاهده می‌شود و دامنه توزیع به سمت مقادیر مثبت است. کشیدگی توزیع نیز منفی بوده و خوابیده‌تر از منحنی نرمال عادی می‌باشد.

هیستوگرام گروه مردان رفتاری متفاوت داشته و فراوانی مقادیر بالا، قابل‌توجه می‌باشد، لذا نوع کجی منفی و دامنه توزیع به سمت مقادیر چپ است. توزیع همچنین دارای کشیدگی مثبت بوده و ارتفاع آن بیش‌ازحد نرمال می‌باشد. با احتیاط می‌شود گفت توزیع داده‌های زنان و مردان چندان نامتقارن نیست.



شکل ۲۰: نمودار جعبه‌ای

□ نکته ۷٤

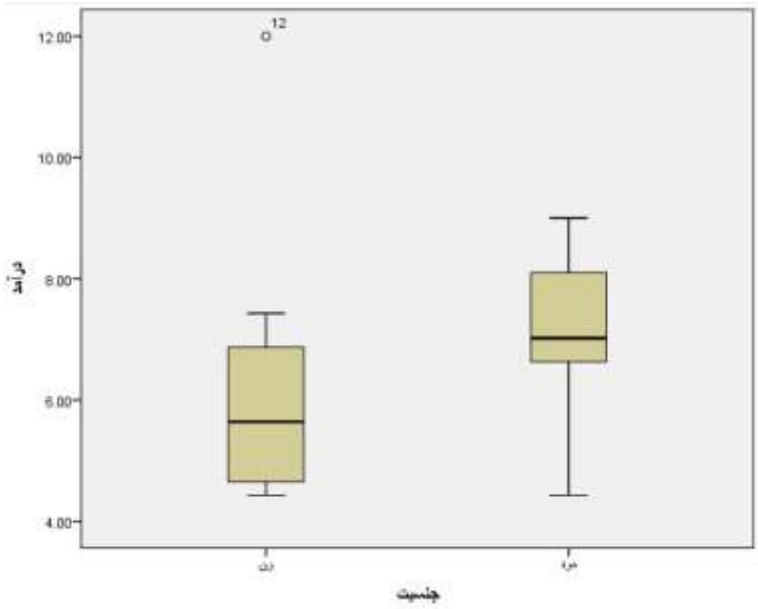
نمودار جعبه‌ای (توضیح)

مستطیل وسط، بخش میانه توزیع (The middle portion of the distribution) را مشخص می‌سازد. قسمت پائین و بالای آن به‌طور تقریبی با چارک‌های اول و سوم برابر بوده و خط پررنگ درون مستطیل، بیانگر میانه توزیع هست. خطوطی که از دو طرف مستطیل خارج شده‌اند ویسکرز (Whiskers) نامیده می‌شوند.

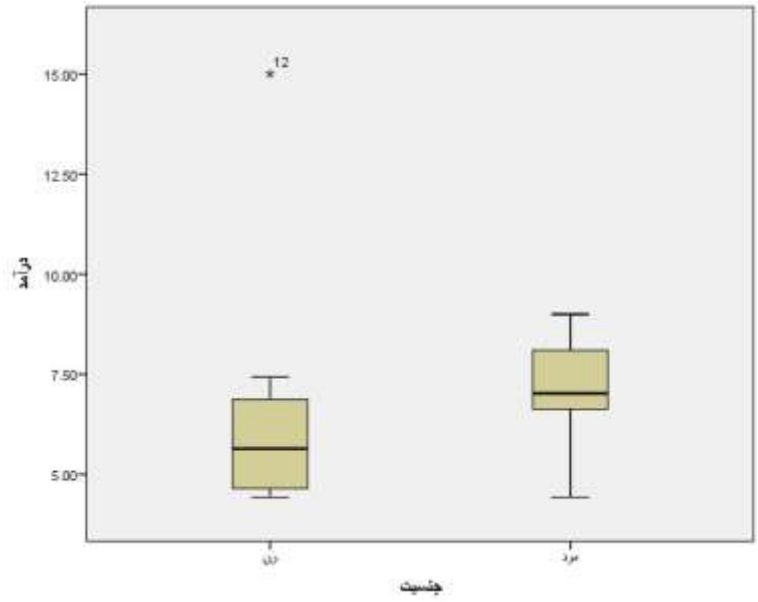
□ نکته ۷۵

نمودار جعبه‌ای (کاربرد)

- نمودار جعبه‌ای در شناسایی داده‌های پرت (Outliers) و انتهایی (Extreme) کاربرد دارد.
- چنانچه داده‌ای در فاصله ۱/۵ تا ۳ برابری فاصله چارکی قرار گیرد با دایره (O) مشخص شده و بیانگر پرت بودن داده است.
- اگر داده‌ای دورتر از ۳ برابر فاصله چارکی باشد، داده انتهایی بوده و با ستاره مشخص می‌شوند.



شکل ۲۱: شناسایی داده پرت



شکل ۲۲: شناسایی داده انتهایی

□ نکته ۷۶

نمودار جعبه‌ای (گزارش دهی)

بررسی نمودار جعبه‌ای گروه‌های زنان و مردان نشان می‌دهد که پراکنش داده‌ها در گروه مردان بیشتر از گروه زنان است. هر دو گروه باآنکه حداقل مقدار یکسانی دارند ولی حداکثر درآمد در گروه مردان بیشتر است. میانگین مشاهدات در گروه مردان بیشتر از گروه زنان بوده و در هیچ‌یک از گروه‌ها داده پرت مشاهده نمی‌شود.

□ نکته ۷۷

تبدیل داده‌ها: مقادیر استاندارد z

- در قسمت Descriptives می‌توان با انتخاب standardized values of variables Save مقادیر استاندارد داده‌های خام (Raw Data) را محاسبه نمود.
- مقادیر z، میزان تفاوت داده برحسب انحراف معیار از میانگین نمونه را مشخص می‌سازند.
- مقادیر استاندارد z از رابطه زیر محاسبه می‌شوند:

$$z = (X - M)/SD$$

که در آن X داده اصلی، M میانگین و SD انحراف معیار می‌باشد.

- مقدار مثبت نشان می‌دهد که داده از میانگین بزرگ‌تر و مقدار منفی بیانگر کوچک بودن داده از میانگین هست.

□ نکته ۷۸

تبدیل داده‌ها: مقادیر استاندارد z (مثال)

- فایل frequencies.sav را باز کرده و از منوی اصلی (main menu) مسیر زیر انتخاب نمایید:

Analyze ➔Descriptive Statistics ➔ Descriptives

- سپس متغیر درآمد را به قسمت Variable(s) منتقل و standardized values of variables Save را انتخاب نمایید.

□ نکته ۷۹

تبدیل داده‌ها: مقادیر استاندارد z (خروجی)

خروجی اجرای دستور از دو قسمت تشکیل یافته است:

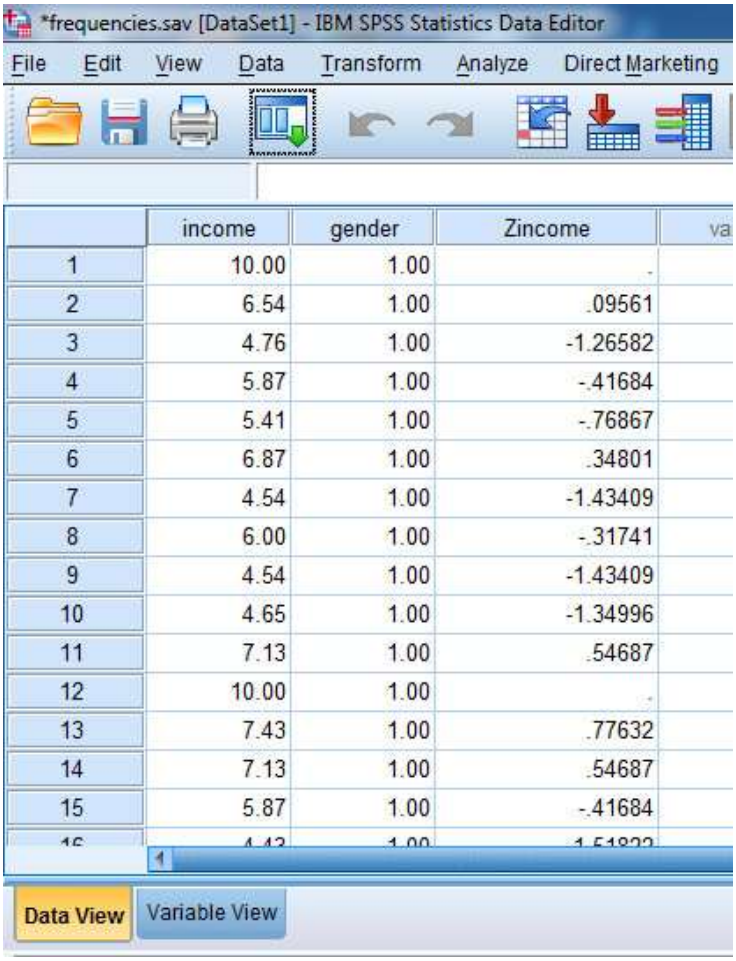
نخست؛ جدول آماره‌های توصیفی شامل حداقل، حداکثر، میانگین و انحراف معیار که داده‌ها بر اساس آن‌ها، به مقادیر استاندارد تبدیل می‌شوند،

دوم؛ متغیر Zincome به فهرست متغیرها اضافه می‌شود.

جدول ۷: آماره‌های توصیفی

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
درآمد	38	4.43	9.00	6.4150	1.30745
Valid N (listwise)	38				



	income	gender	Zincome
1	10.00	1.00	.
2	6.54	1.00	.09561
3	4.76	1.00	-1.26582
4	5.87	1.00	-.41684
5	5.41	1.00	-.76867
6	6.87	1.00	.34801
7	4.54	1.00	-1.43409
8	6.00	1.00	-.31741
9	4.54	1.00	-1.43409
10	4.65	1.00	-1.34996
11	7.13	1.00	.54687
12	10.00	1.00	.
13	7.43	1.00	.77632
14	7.13	1.00	.54687
15	5.87	1.00	-.41684

شکل ۲۳: متغیر استاندارد

ویدیوی ٤: متغیر استاندارد

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۸۰

تبدیل داده‌ها: رمزگذاری مجدد

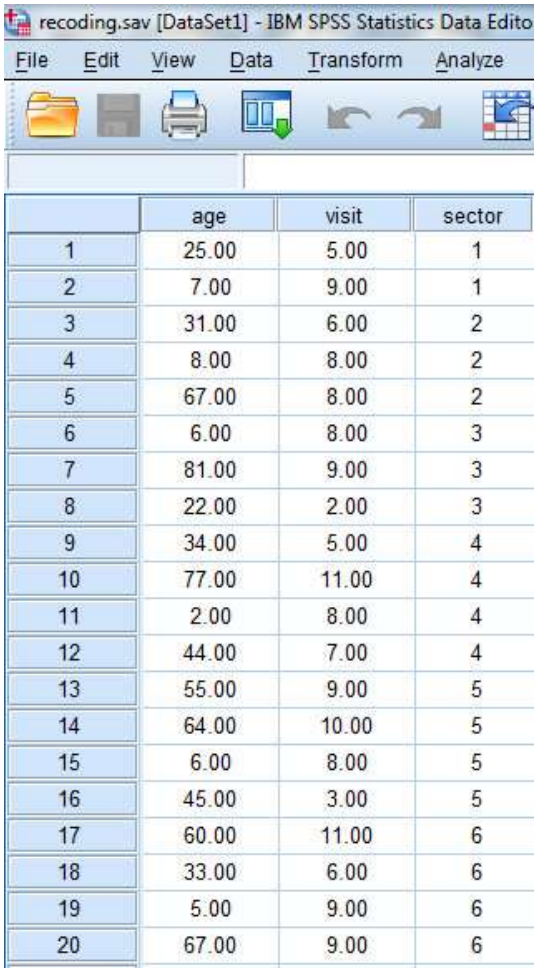
- متغیرهای اسمی و کمّی را با توجه به اهداف پژوهش می‌توان رمزگذاری مجدد (Recoding) کرد.
- متغیرهای اسمی بیشتر رمزگذاری مجدد می‌شوند.
- هنگامی‌که تعداد اعضای گروه‌ها برای انجام محاسبات کافی نباشند، از این روش برای تجمیع داده‌ها استفاده می‌شود.

- در داده‌های کمّی برای مدیریت مشاهدات پرت ولی واقعی می‌توان از روش Recoding استفاده نمود و به آن‌ها اعداد قابل قبول اختصاص داد.
- دو روش کدگذاری recode a variable into the same variable و recode a variable into a different variable وجود دارد.
- استفاده از روش recode a variable into the same variable به دلیل حذف داده‌های اولیه، توصیه نمی‌شود.
- رمزگذاری مجدد برای داده‌های کدگذاری شده نیز کاربرد دارد، برای مثال داده‌های که بر حسب شهرستان کدگذاری شده‌اند را می‌توان به تفکیک منطقه جغرافیایی، رمزگذاری مجدد نمود.

□ نکته ۸۱

تبدیل داده‌ها: رمزگذاری مجدد (مثال: داده کمّی)

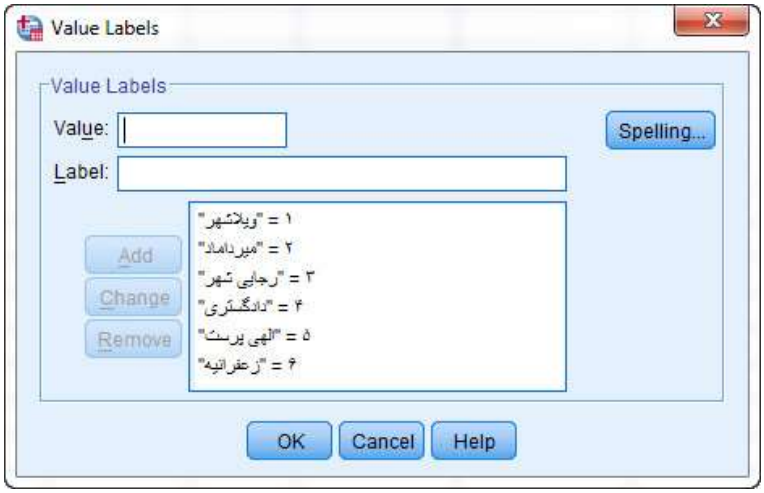
- فایل Recoding شامل متغیر سن (age)، تعداد مراجعه به درمانگاه (visit) و محله سکونت (sector) را باز نمایید.
- از مسیر زیر متغیر سن (age) را به پنجره گفتگوی رمز مجدد (Recode dialog window) بیاورید.
- متغیر age را به کادر Numeric Variable ➔ Output Variable منتقل نمایید.


recoding.sav [DataSet1] - IBM SPSS Statistics Data Edito

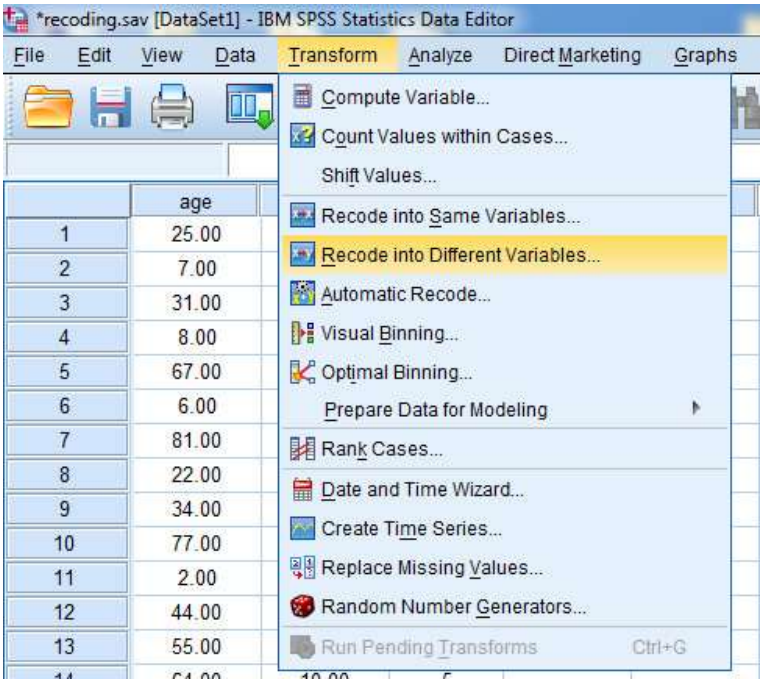
File Edit View Data Transform Analyze

	age	visit	sector
1	25.00	5.00	1
2	7.00	9.00	1
3	31.00	6.00	2
4	8.00	8.00	2
5	67.00	8.00	2
6	6.00	8.00	3
7	81.00	9.00	3
8	22.00	2.00	3
9	34.00	5.00	4
10	77.00	11.00	4
11	2.00	8.00	4
12	44.00	7.00	4
13	55.00	9.00	5
14	64.00	10.00	5
15	6.00	8.00	5
16	45.00	3.00	5
17	60.00	11.00	6
18	33.00	6.00	6
19	5.00	9.00	6
20	67.00	9.00	6

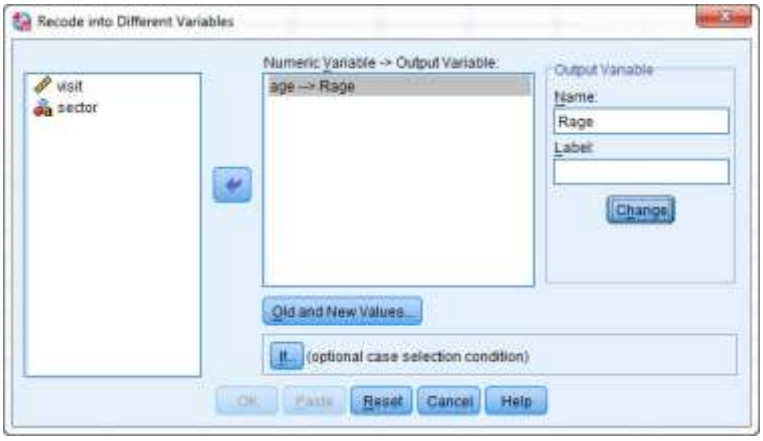
شکل ۲۴: داده‌های رمزگذاری مجدد



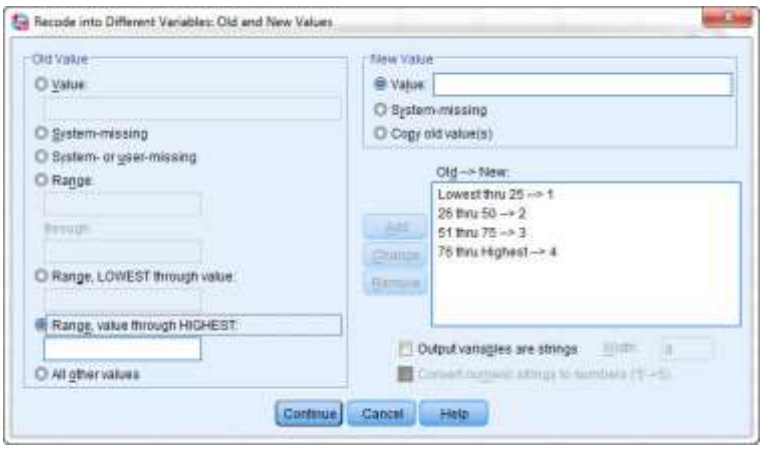
شکل ۲۵: تعریف داده‌ها



شکل ۲۶: مسیر رمزگذاری مجدد



شکل ۲۷: تعریف نام متغیر برای رمزگذاری مجدد



شکل ۲۸: رمزگذاری مجدد

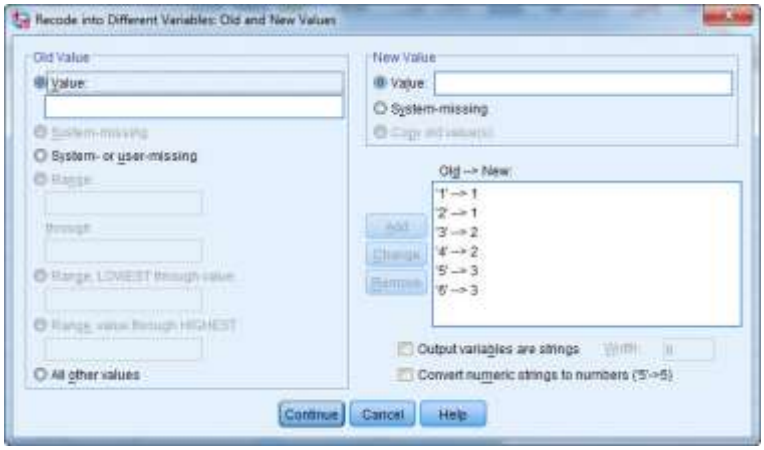
ویدیوی ۵: رمزگذاری مجدد داده کمی

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۸۲

تبدیل داده‌ها: رمزگذاری مجدد (مثال: داده طبقه‌ای)

- فایل recoding.sav را باز کرده و از مسیر زیر متغیر محله سکونت (sector) را به پنجره گفتگوی رمزگذاری مجدد (Recode dialog window) بیاورید.
- متغیر sector را به کادر Output Variable ➔ Numeric Variable منتقل کنید.
- محله‌های ویلاشهر (۱) و میرداماد (۲) را در هم ادغام و طبقه ۱ را شکل دهید.
- محله‌های رجائی شهر (۳) و دادگستری (۴) را در هم ادغام و طبقه ۲ را ایجاد نمایید.
- محله‌های الهی پرست (۵) و زعفرانیه (۶) را ادغام و طبقه ۳ را ایجاد کنید.



شکل ۲۹: رمزگذاری مجدد داده‌های طبقه‌ای

ویدیوی ٦: رمزگذاری مجدد داده طبقه‌ای

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۸۳

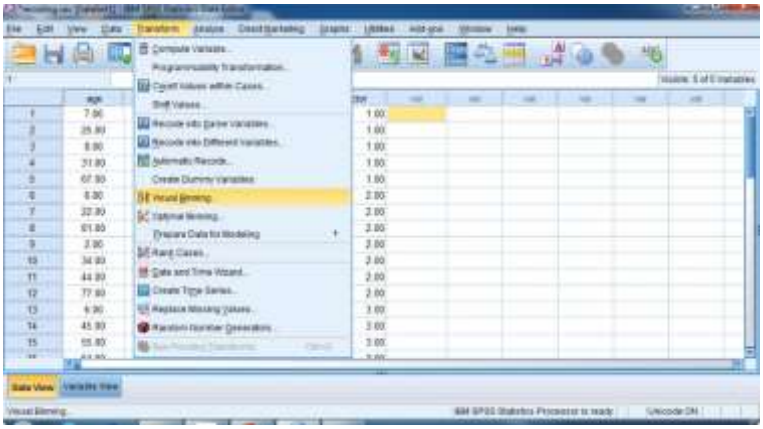
تبدیل داده‌ها: Visual binning

- Visual Binning، روش دیگری برای رمزگذاری مجدد داده‌های کمّی به گروه‌های مختلف (Bins) می‌باشد.
- این روش، دیداری (Visual) نامیده می‌شود، زیرا SPSS با نمایش هیستوگرام متغیر، امکان مشاهده نحوه تقسیم‌بندی را فراهم می‌سازد.
- Visual Binning، باعث از بین رفتن اطلاعات موجود در داده‌ها می‌شود و غیر از مواقع ضروری توصیه نمی‌شود.
- در این روش امکان گروه‌بندی بر اساس انحراف معیار (standard deviation) وجود دارد.
- −1 SD, 0 SD, and +1 SD به ترتیب معادل با صدک های 15.9, 50, and 84.1 می‌باشد.
- رمزگذاری بر اساس انحراف معیار باعث تشکیل گروه‌هایی با اندازه‌های مختلف می‌شود. در این نوع گروه‌بندی بیشتر داده‌ها در گروه‌های میانی متمرکز می‌شوند.
- برای ایجاد گروه‌های با تعداد مشاهده مساوی، تقسیم‌بندی بر اساس صدک ها انجام می‌پذیرد.

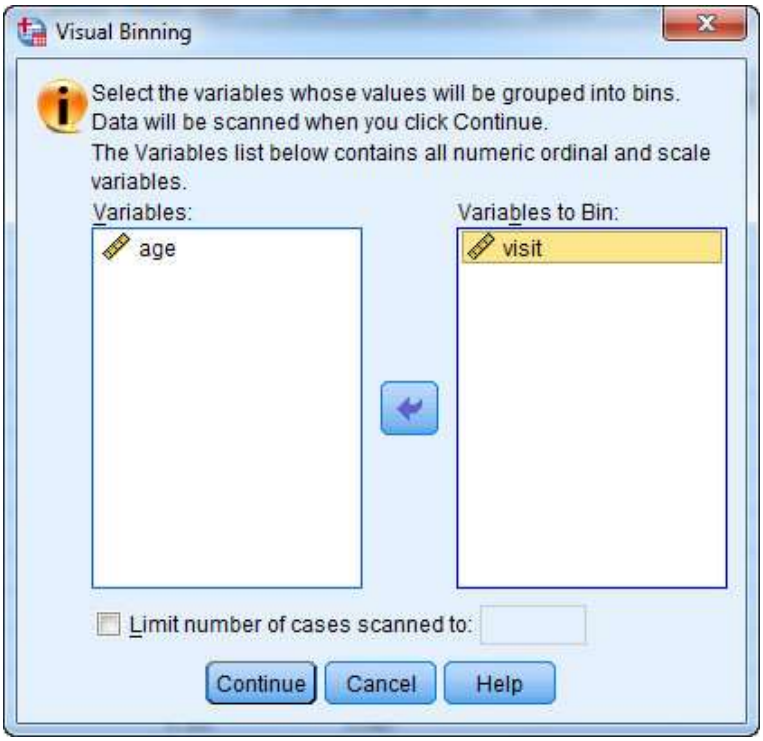
□ نکته ۸٤

مثال: Visual binning

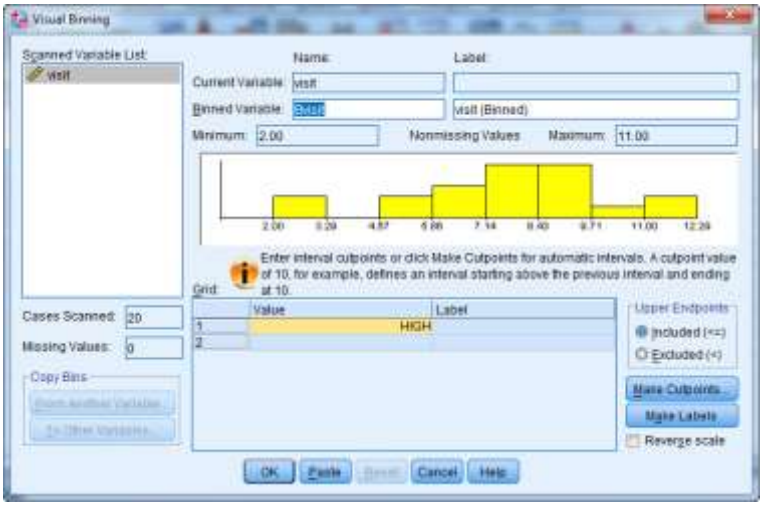
- فایل recoding.sav را باز کنید.
- از منوی اصلی و مسیر Transform ➔Visual Binning متغیر visit را به قسمت Variables to Bin منتقل نموده و با انتخاب Continue به پنجره اصلی وارد شوید.
- نام متغیر جدید (Binned Variable) را تعریف و گروه‌بندی را انجام دهید.



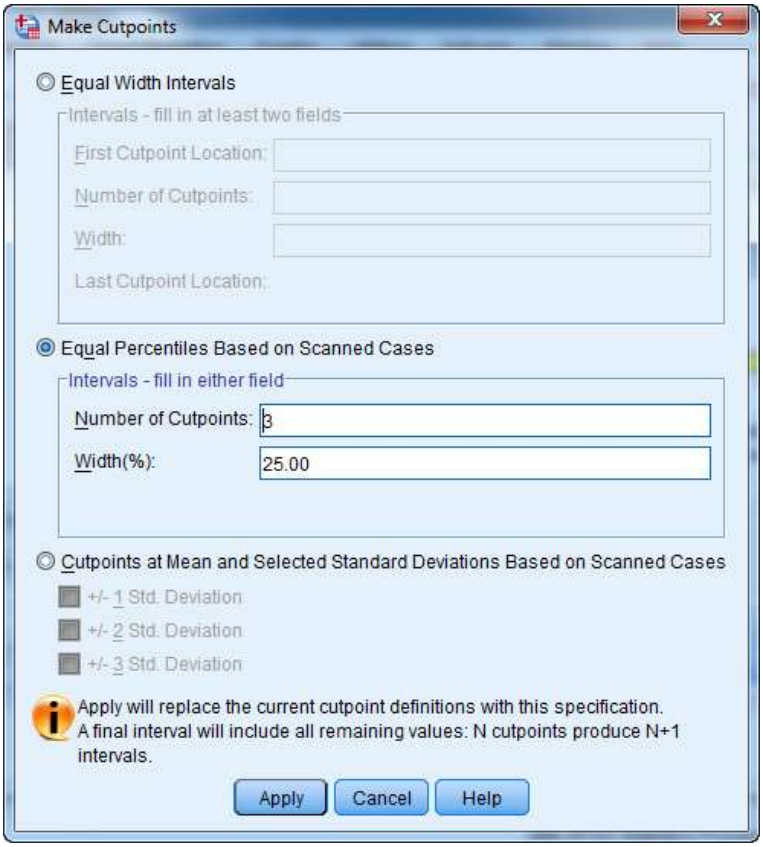
شکل ۳۰: مسیر دستور رمزگذاری دیداری



شکل ۳۱: انتخاب متغیر برای رمزگذاری دیداری



شکل ۳۲: پنجره اصلی رمزگذاری دیداری



شکل ۳۳: انتخاب روش گروه‌بندی برای رمزگذاری دیداری

ویدیوی ۷: رمزگذاری دیداری بر اساس صدک

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

ویدیوی ۸: رمزگذاری دیداری بر اساس انحراف معیار

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۸۵

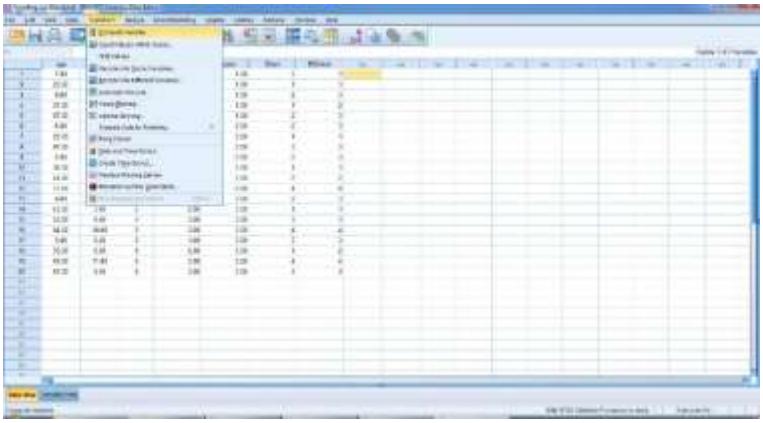
تبدیل داده‌ها: محاسبه متغیرهای جدید

- برای تک متغیرها (Single Variable) یا مجموعه‌ای از متغیرها (Set of Variables) کاربرد دارد.
- می‌توان با استفاده از این قابلیت، محاسبات جبری انجام داده یا ریشه دوم و میانگین متغیرها را محاسبه نمود.
- پس از انجام محاسبات، متغیر جدید به انتهای فهرست متغیرهای موجود اضافه می‌شود.
- چنانچه محاسبه به دلایلی مانند وجود متغیر مفقود انجام نگیرد، system missing marker (a blank cell) نمایش داده می‌شود.
- برای محاسبه متغیر جدید از مسیر Transform ➔ Compute Variable اقدام می‌شود.

□ نکته ۸۶

محاسبه متغیرهای جدید: عبارت جبری

- فایل recoding.sav را باز نموده و مقادیر استاندارد یا z متغیر age را از رابطه $z = \frac{age-mean}{SE}$ محاسبه نمایید.
- میانگین و انحراف معیار به ترتیب ۳۶/۹۵ و ۲۶/۵۳ می‌باشند.
- نتایج را با محاسبه مستقیم مقادیر z مقایسه نمایید.



شکل ۳۴: مسیر محاسبه متغیر جدید

ویدیوی ۹: محاسبه متغیر جدید (عبارت جبری)

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

ویدیوی ۱۰: محاسبه متغیر جدید (میانگین دو متغیر)

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۸۷

ضریب همبستگی پیرسون

- ضرایب همبستگی (Correlation Coefficients) شاخصی برای بررسی نحوه ارتباط متغیرهای پژوهش با یکدیگر می‌باشد.
- ضرایب همبستگی مختلفی قابل استفاده در علوم رفتاری، اجتماعی و پزشکی می‌باشند.
- ضریب همبستگی پیرسون (Pearson Product Moment Correlation) پرکاربردترین ضریب هست.

□ نکته ۸۸

ضریب همبستگی پیرسون: فروض

- متغیرها تقریباً فاصله‌ای هستند (approximately interval measurement).
- داده‌ها تقریباً از توزیع نرمال پیروی می‌کنند (approximately normally distributed).
- ضریب همبستگی پیرسون به داده‌های پرت حساس می‌باشد.

□ نکته ۸۹

ضریب همبستگی پیرسون: تاریخچه

- ضریب همبستگی پیرسون در سال ۱۸۹۶ توسط Karl Pearson توسعه داده شد.
- این ضریب، ادامه کارهای Sir Francis Galton در سال‌های ۱۸۸۶ و ۱۸۸۸ بود.

□ نکته ۹۰

ضریب همبستگی پیرسون: محدودیت‌ها

- ضریب پیرسون، درجه ارتباط خطی دو متغیر را نشان می‌دهد.
- چنانچه ارتباط دو متغیر غیرخطی باشد مثلاً U-shaped function باشد، ضریب پیرسون شدت ارتباط را کمتر از میزان واقعی برآورد خواهد کرد.
- چنانچه رابطه U-shaped function متقارن باشد، ضریب صفر محاسبه می‌شود.

□ نکته ۹۱

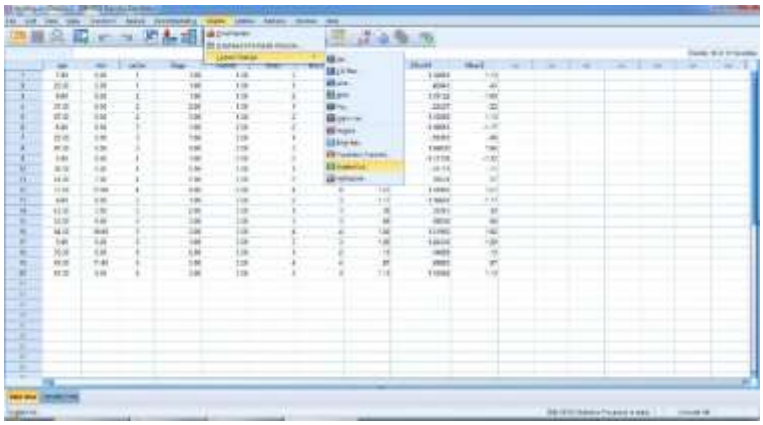
ضریب همبستگی پیرسون: تحلیل

- ارتباط دو متغیر به معنای تغییر همزمان آن‌ها می‌باشد.
- تغییر همزمان می‌تواند به معنای تغییر یک متغیر در نتیجه تغییر ایجاد شده در متغیر دیگر تفسیر شود (پیش‌بینی).
- توان دو ضریب همبستگی یا r^2 شدت ارتباط (the strength of the relationship,) را مشخص می‌سازد.

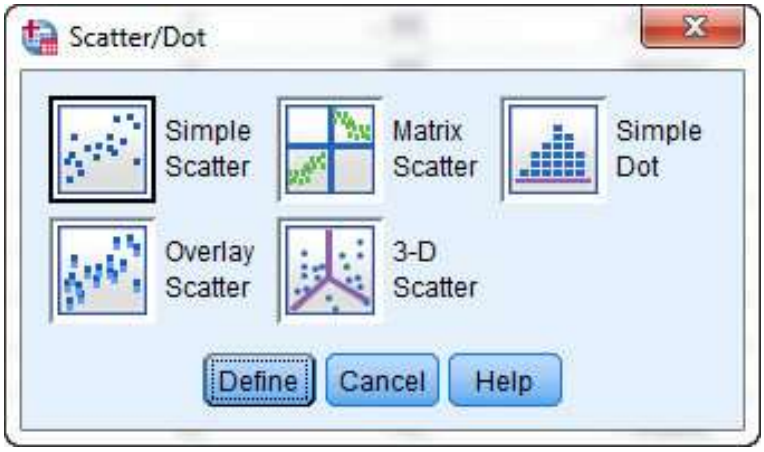
□ نکته ۹۲

مثال ضریب همبستگی پیرسون: ارتباط خطی

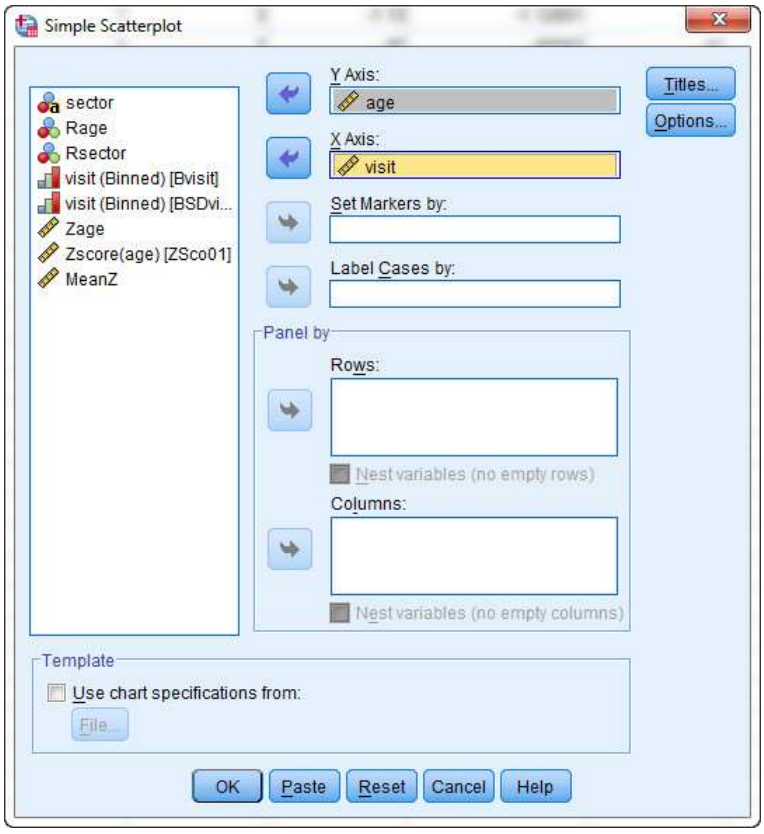
- ضریب همبستگی متغیرهای age و visit در فایل داده recoding.sav بررسی می‌شود.
- مسیر Graphs➔Legacy Dialogs➔Scatter/Dot انتخاب می‌شود.
- نمودار Simple Scatter انتخاب و متغیرها را به‌دلخواه در محورهای عمودی و افقی قرار می‌دهیم.



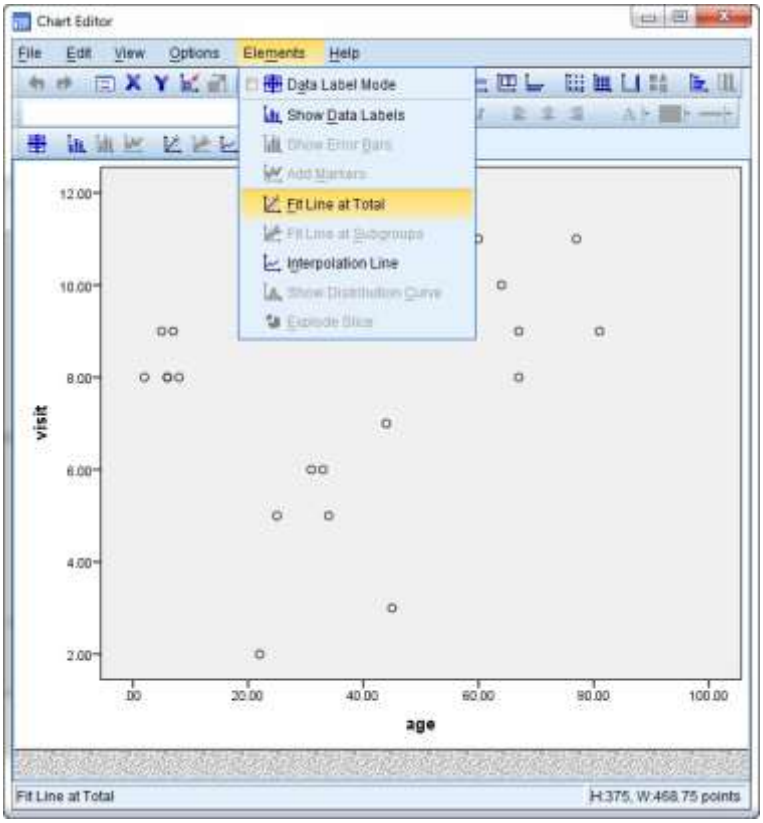
شکل ۳۵: مسیر بررسی ارتباط خطی



شکل ۳۶: انتخاب نوع ارتباط



شکل ۳۷: تعریف متغیرهای محورها



شکل ۳۸: برازش خط رگرسیون

ویدیوی ۱۱: بررسی خطی بودن ارتباط متغیرها

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۹۳

مثال ضریب همبستگی پیرسون: محاسبه

- از مسیر Analyze→Correlate→Bivariate برای محاسبه ضریب همبستگی اقدام نمایید.
- متغیر age و visit را به Variables Panel انتقال دهید.
- با انتخاب Flag significant correlations معنی‌داری با ستاره مشخص می‌شود.

جدول ۸: نتایج ضریب پیرسون

Correlations

		age	visit
age	Pearson Correlation	1	.312
	Sig. (2-tailed)		.180
	Sum of Squares and Cross-products	13372.950	380.550
	Covariance	703.839	20.029
	N	20	20
visit	Pearson Correlation	.312	1
	Sig. (2-tailed)	.180	
	Sum of Squares and Cross-products	380.550	110.950
	Covariance	20.029	5.839
	N	20	20

ویدیوی ۱۲: ضریب همبستگی پیرسون

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۹٤

ضریب همبستگی: آزمون‌های نا پارامتری

- چنانچه فروض ضریب همبستگی پیرسون تأمین نشود، استفاده از ضرایب همبستگی نا پارامتری (non-parametric correlation procedures) ضروری می‌گردد.
- آزمون‌های نا پارامتری، پارامتر نمونه را با پارامتر جامعه مقایسه نمی‌کنند و اجرای آن‌ها به فروض کمتری نیاز دارد.
- در این آزمون‌ها نیازی به نرمال بودن توزیع نیست.

□ نکته ۹۵

ضریب همبستگی: اسپیرمن و کندال تاو-بی

- این آزمون‌ها برای داده‌های طبقه‌ای و رتبه‌ای کاربرد دارد.
- ضرایب همبستگی اسپیرمن (Spearman rho.) و کندال تاو-بی (Kendall tau-b)، جزو آزمون‌های معروف نا پارامتری هستند.

- در آزمون‌های نا پارامتری ارتباط میان متغیرها ضروری است یکنواخت (monotonic relationship) باشد. ارتباط خطی نوعی رابطه یکنواخت می‌باشد.

□ نکته ۹۶

ضریب همبستگی اسپیرمن

- ضریب همبستگی اسپیرمن ابتدا توسط Sir Charles Spearman (1904) به‌منظور کاهش تأثیر داده‌های پرت بر دقت ضریب همبستگی پیرسون معرفی شد.
- چنانچه رتبه متغیرها محاسبه و سپس با استفاده از آن‌ها ضریب پیرسون برآورد گردد، عدد حاصل همان ضریب همبستگی اسپیرمن خواهد بود.

□ نکته ۹۷

ضریب همبستگی اسپیرمن: محاسبه

- در محاسبه این ضریب ابتدا اختلاف رتبه هر جفت داده محاسبه می‌شود. سپس اعداد به توان دو رسیده و جمع می‌شوند.
- اختلاف‌های بزرگ با توان دو بزرگ‌تر شده و ضریب همبستگی اسپیرمن آن‌ها کاهش می‌یابد.
- هر اندازه ضریب محاسباتی به ۱± نزدیک باشد، بیانگر تشابه بیشتر رتبه متغیرها هست (همبستگی و ارتباط بیشتر).

□ نکته ۹۸

ضریب همبستگی کندال تاو-بی

- Maurice G. Kendall (1938, 1948) ضریب همبستگی کندال تاو-بی را به‌عنوان جایگزینی برای رتبه‌بندی اسپیرمن معرفی کرد.
- این ضریب سه نوع tau-a، tau-b و tau-c دارد که IBM SPSS در قسمت Bivariate Correlation تنها tau-b را محاسبه می کند.

□ نکته ۹۹

ضریب همبستگی کندال تاو-بی: تفسیر

- چنانچه مقدار ضریب همبستگی کندال تاو-بی به ۱+ نزدیک باشد، می‌توان گفت مقادیر دو متغیر بسیار شبیه یکدیگر رتبه‌بندی شده‌اند.
- چنانچه ضریب به ۱- نزدیک باشد، بیانگر رتبه‌بندی معکوس متغیرها خواهد بود.
- ضریب صفر نشانگر استقلال متغیرها هست.

□ نکته ۱۰۰

ضریب همبستگی اسپیرمن و کندال تاو-بی: مثال

- فایل داده spearmankendall.sav را باز نمایید.
- از مسیر Analyze➔Correlate➔Bivariate برای محاسبه ضرایب همبستگی اقدام نمایید.
- متغیر improving (شاخص بهبود توان درمانی) و gender (جنسیت) را به قسمت Variables انتقال دهید.

جدول ۹: نتایج ضریب کندال و اسپیرمن

Correlations

			improving	gender
Kendall's tau_b	improving	Correlation Coefficient	1.000	.479**
		Sig. (2-tailed)	.	.001
		N	40	40
	gender	Correlation Coefficient	.479**	1.000
		Sig. (2-tailed)	.001	.
		N	40	40
Spearman's rho	improving	Correlation Coefficient	1.000	.531**
		Sig. (2-tailed)	.	.000
		N	40	40
	gender	Correlation Coefficient	.531**	1.000
		Sig. (2-tailed)	.000	.
		N	40	40

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

□ نکته ۱۰۱

ضریب همبستگی اسپیرمن و کندال تاو-بی: تحلیل

ضریب همبستگی اسپیرمن و کندال تاو-بی به ترتیب در یک آزمون دوطرفه ۰/۴۸ و ۰/۵۳ محاسبه شده که هر دو در سطح آماری یک درصد، معنی‌دار می‌باشند. به‌عبارت‌دیگر احتمال اینکه ارتباط میان متغیر بهبود

درمانی و جنسیت تصادفی یا شانسی باشد کمتر از یک درصد است و بهبود درمانی در مردان بهتر از زنان جواب می‌دهد.

ویدیوی ۱۳: ضریب همبستگی اسپیرمن و کندال

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۱۰۲

رگرسیون ساده خطی

- رگرسیون ساده خطی (Simple Linear Regression) و ضریب همبستگی پیرسون ارتباط نزدیکی با یکدیگر دارند.
- ضریب همبستگی میزان ارتباط متغیرها را مشخص نموده ولی رگرسیون، توانایی پیش‌بینی (Prediction) یک متغیر بر پایه متغیر دیگر را میسر می‌سازد.

□ نکته ۱۰۳

رگرسیون ساده خطی: تاریخچه

- فکر اولیه رگرسیون در سال ۱۸۸۶ ابتدا توسط Galton مطرح شد.
- دو سال بعد موضوع همگرایی (Covariation) ارائه گردید که اکنون با نام همبستگی (Correlation) شناخته می‌شود.

□ نکته ۱۰٤

رگرسیون ساده خطی: متغیرهای مستقل و وابسته

- در رگرسیون، متغیری که پیش‌بینی می‌شود (Being Predicted) را متغیر وابسته (Dependent) یا خروجی (Outcome) می‌نامند و معمولاً با Y نمایش می‌دهند.
- متغیر دیگر که برای پیش‌بینی (Predict) استفاده می‌شود را مستقل (Independent) یا پیش‌بینی کننده (Predictor) نامیده و با X مشخص می‌کنند.

□ نکته ۱۰۵

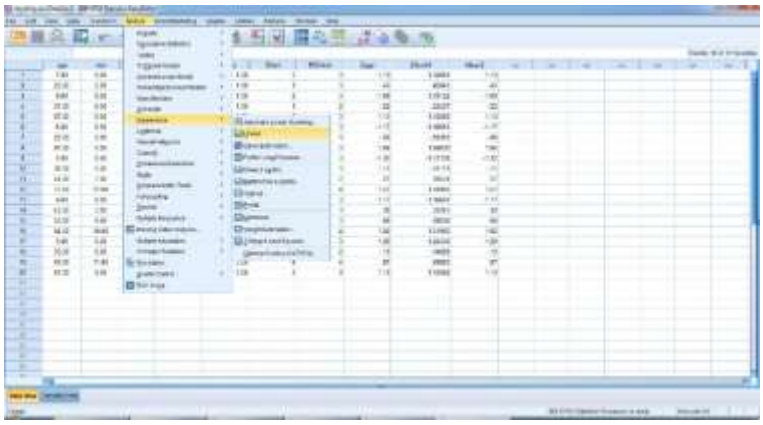
رگرسیون ساده خطی: نام‌گذاری

- رگرسیون می‌نامند زیرا مقادیر متغیر وابسته با استفاده از یک الگو (Model) پیش‌بینی می‌شود.
- ساده می‌باشد زیرا تنها از یک متغیر برای پیش‌بینی استفاده می‌شود.
- خطی است زیرا ضرایب الگو خطی هستند.

□ نکته ۱۰۶

رگرسیون ساده خطی: مثال

- فایل recoding.sav را باز کنید.
- از مسیر Analyze➔Regression➔Linear وارد شده و متغیر مراجعه به درمانگاه (visit) را به قسمت Dependent و متغیر سن (age) را به قسمت Independent(s) منتقل نمایید.
- در قسمت Statistics گزینه‌های Model fit، Descriptives و Part and partial correlations را انتخاب نمایید.



شکل ۳۹: مسیر دستور رگرسیون

جدول ۱۰: نتایج رگرسیون

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	6.499	.920		7.065	.000			
	age	.028	.020	.312	1.395	.180	.312	.312	.312

a. Dependent Variable: visit

□ نکته ۱۰۷

رگرسیون ساده خطی: تحلیل

ضریب همبستگی پیرسون متغیرهای سن و تعداد مراجعه در یک آزمون یک‌طرفه معنی‌دار نشده است. این موضوع در الگوی رگرسیون خطی نیز مشاهده می‌شود، به‌نحوی‌که ضریب متغیر سن، ۰/۰۳ محاسبه شده که با خطای معیار ۰/۰۲، آماره t مربوط ۰/۱۸می‌شود که در سطح ۵ درصد معنی‌دار نیست. به‌عبارت‌دیگر سن تأثیر معنی‌داری بر تعداد مراجعه به مراکز درمانی ندارد.

ویدیوی ١٤: رگرسیون ساده خطی

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۱۰۸

Centering the Predictor Variable

- چنانچه صفر عدد غیرمعمول (Unusual) یا غیر معتبر (Not valid) برای متغیر مستقل یا پیش‌بینی کننده (Predictor) باشد، از تکنیک Centering برای مفهوم بخشی به عرض از مبدأ مدل رگرسیون استفاده می‌شود.
- Centering تأثیری بر نتایج رگرسیون نمی‌گذارد.

□ نکته ۱۰۹

Centering: محاسبه

- Centering تبدیل متغیر اصلی به‌گونه‌ای که دارای میانگین صفر باشد، هست.
- این عمل با محاسبه انحراف داده‌ها از یک عدد مرجع (Reference Score) مانند میانگین انجام می‌گیرد.
- رابطه مورد استفاده به‌صورت (Deviation Score = Obtained Score−Mean) است.

□ نکته ۱۱۰

مثال: Centering

- ارتباط میان شاخص توده بدنی (Body Mass Index) یا BMI و ضربان قلب (Pulse Rate) بررسی می‌شود.
- شاخص توده بدنی در تشخیص اینکه آیا فرد اضافه وزن (Overweight) دارد یا خیر کاربرد دارد.
- Adolphe Quetelet این شاخص را در سال ۱۸۳۲ معرفی کرد و در سال ۱۹۷۲ توسط Ancel Keys، BMI نامیده شد.
- BMI نسبت وزن به کیلوگرم به توان دو قد به متر هست.
- طبقه‌بندی شاخص توده بدنی به‌صورت زیر است:
- کمتر از ۱۸/۵ لاغر (Underweight)، بین ۱۸/۵ تا ۲۴/۹ عادی (Normal)، بین ۲۵ تا ۲۹/۹ اضافه وزن (Overweight)، بین ۳۰ تا ۳۹/۹ چاق (Obese) و بیشتر از ۴۰ چاقی مفرط (Morbidly Obese) می‌باشد.
- با افزایش شاخص توده بدنی ریسک سلامت (Health Risk) افزایش می‌یابد.
- یکی از شاخص‌های ریسک سلامت، ضربان قلب (Heart Rate) است.
- احتمالاً شاخص توده بدنی، توانایی پیش‌بینی ضربان قلب را دارد.

□ نکته ۱۱۱

مثال سنترینگ: برآورد

- فایل BMI.sav را باز کنید.
- میانگین متغیر BMI را محاسبه نمایید.
- متغیر شاخص توده بدنی را Center کنید.
- رگرسیون ساده خطی بین متغیر ضربان قلب (pulse_rate) و شاخص توده بدنی (BMI) و شاخص توده بدنی Center شده (BMI_centered) را به‌طور جداگانه انجام داده و نتایج را مقایسه نمایید.

جدول ۱۱: نتایج رگرسیون سنتر شده

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	74.933	1.049		71.434	.000
	BMI_centered	.902	.158	.846	5.722	.000

a. Dependent Variable: ضربان قلب در دقیقه

□ نکته ۱۱۲

سنترینگ: تحلیل نتایج

ضریب زاویه هر دو رگرسیون ۰/۹ برآورد شد که نشان می‌دهد در جامعه مورد مطالعه با افزایش یک واحد در میانگین شاخص توده بدنی بر میزان ضربان قلب در دقیقه، یک واحد افزوده می‌شود. با جایگزینی عدد صفر در مقدار متغیر شاخص توده بدنی سنتر شده، ضربان قلب ۷۴/۹ به دست می‌آید که بیانگر ضربان قلب افراد با میانگین شاخص توده بدنی، ۲۹ است.

ویدیوی ۱۵: رگرسیون سنترینگ

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۱۱۳

رگرسیون خطی چند متغیره

رگرسیون خطی چند متغیره یا Multiple linear regression توسعه رگرسیون خطی ساده می‌باشد، به‌نحوی‌که مدل شامل دو یا چند متغیر پیش‌بینی کننده یا predictors هست. ضرایب برآوردی در رگرسیون خطی چند متغیره ضرایب جزئی رگرسیون (partial regression coefficients) نامیده می‌شوند، زیرا در محاسبه آن‌ها تأثیر متغیرهای دیگر دیده شده است.

در مواقعی با حذف متغیرهایی که معنی‌دار نشده‌اند، الگوی خلاصه ارائه می‌شود. این اقدام با ایراداتی همراه می‌باشد و لازم است حفظ یا حذف متغیرها بر اساس مبانی نظری مطالعه باشد. رگرسیون خطی را می‌توان

به روش‌های forward، backward و stepwise برآورد نمود. این روش‌ها تنها متغیرهای معنی‌دار را حفظ می‌کنند.

□ نکته ۱۱٤

روش‌های برآورد رگرسیون خطی چند متغیره

- در روش forward متغیرها یکی پس از دیگری وارد مدل شده و هر متغیری که معنی‌دار شود، در الگو حفظ می‌شود.
- در روش backward همه متغیرها ابتدا وارد شده و به‌تدریج متغیرهایی که توانایی توضیح دهندگی ندارند حذف می‌شوند.
- در روش stepwise برخلاف روش forward هنگامی‌که توانایی توضیح دهنگی متغیر کاهش یابد از الگو حذف شده و بعد تأثیر آن مجدد بررسی می‌شود.

□ نکته ۱۱۵

Automatic Linear Modeling

- روش جایگزین برای روش‌های اشاره شده، استفاده از روش Automatic Linear Modeling می‌باشد.
- این روش از تحلیل all-possible-subsets استفاده کرده و تمامی حالات مختلف را لحاظ می‌کند.
- چنانچه p تعداد متغیر باشد در آن صورت به تعداد (2^P)-1 زیرمجموعه خواهیم داشت. به‌عنوان‌مثال برای ۵ متغیر مستقل، ۳۱ زیرمجموعه وجود خواهد داشت.

□ نکته ۱۱٦

رگرسیون خطی چند متغیره: مثال

- فایل داده MLR.sav را باز کنید. این فایل شامل متغیر visit (تعداد مراجعه به درمانگاه) و age (سن بیمار) می‌باشد.
- متغیر جدید ageage که توان دو age می‌باشد را ایجاد نمایید.
- رگرسیون چند متغیره forward، backward، stepwise و Automatic Linear Modeling را انجام دهید.
- نتایج را تفسیر نمایید.

ویدیوی ۱۶: رگرسیون forward

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

ویدیوی ۱۷: رگرسیون backward

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

ویدیوی ۱۸: رگرسیون stepwise

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

ویدیوی ۱۹: رگرسیون اتوماتیک

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۱۱۷

رگرسیون خطی چند متغیره: تحلیل

بررسی نتایج روش‌های مختلف مشخص می‌سازد که در هر چهار روش، رابطه درجه دو مناسب‌ترین الگو برای پیش‌بینی تعداد مراجعه به درمانگاه با استفاده از سن بیمار می‌باشد، به‌گونه‌ای که با افزایش سن تا ۳۰ سالگی تعداد مراجعه کاهش و سپس افزایش می‌یابد. این در حالی است که در الگوی رگرسیون خطی ساده ارتباطی میان سن و تعداد مراجعه مشاهده نشد.

□ نکته ۱۱۸

رگرسیون خطی سلسله مراتبی

- رگرسیون خطی سلسله مراتبی (Hierarchical linear regression) توسعه رگرسیون خطی چند متغیره می‌باشد با این تفاوت که مدیریت وارد کردن متغیرها به‌جای اینکه به نرم‌افزار سپرده شود، در اختیار کاربر

می‌باشد. متغیرها در قالب زیرمجموعه‌ها (subsets) یا بلوک‌ها (blocks) به ترتیب اولویت وارد الگو می‌شوند به همین خاطر روش، سلسله مراتبی نامیده می‌شود.

- در رگرسیون خطی سلسله مراتبی متغیرهایی که پیش‌تر وارد می‌شوند به‌عنوان متغیرهای کمکی یا همپراش (covariates) برای متغیرهای بعدی عمل می‌کنند.
- در هر بلوک می‌توان روش برآورد الگو را بر اساس گزینه‌های موجود enter، backward، forward یا stepwise جداگانه مشخص کرد.

□ نکته ۱۱۹

رگرسیون خطی سلسله مراتبی: مثال

- به‌منظور مطالعه تأثیر روش پرتودرمانی بر وزن بیماران اقدام به جمع‌آوری اطلاعات مربوط به وزن ثانویه، وزن اولیه، جنسیت، سن، قد و روش معالجه برای ۱۰ بیمار گردید.
- با تقسیم متغیرهای مستقل به سه بلوک (بلوک ۱: وزن اولیه، بلوک ۲: جنسیت، سن و قد، بلوک۳: روش پرتودرمانی) الگوی رگرسیون خطی به روش سلسله مراتبی برآورد گردید.

جدول ۱۲: نتایج رگرسیون سلسله مراتبی

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-10.196	17.844		-.571	.583
	اولیه وزن	1.004	.238	.830	4.215	.003
2	(Constant)	-47.992	67.076		-.715	.506
	اولیه وزن	1.104	.368	.913	2.996	.030
	جنسیت	2.557	4.924	.185	.519	.626
	سن	.000	.209	-.001	-.002	.999
	قد	.163	.290	.145	.560	.600
3	(Constant)	6.155	46.277		.133	.901
	اولیه وزن	.841	.250	.696	3.368	.028
	جنسیت	.897	3.164	.065	.283	.791
	سن	.070	.134	.105	.524	.628
	قد	-.116	.207	-.104	-.562	.604
	پرتودرمانی روش	7.573	2.595	.536	2.918	.043

a. Dependent Variable: وزن ثانویه

ویدیوی ۲۰: رگرسیون سلسله مراتبی

قابل‌دسترس در سی دی و کانال تلگرام @drshahnavazi

□ نکته ۱۲۰

رگرسیون لوجیت دوحالتی

هنگامی که متغیر وابسته یا خروجی (outcome) به‌صورت طبقه‌ای (categorical) از قبیل قبولی یا ردی است از رگرسیون لوجیت یا لوجستیک استفاده می‌شود. هنگامی که متغیر وابسته شامل دو گروه می‌باشد نوع رگرسیون لوجستیک مورد استفاده، دوحالتی، دوتایی یا binary خواهد بود.