**کاربرد نرم افزارهای کامپیوتری SPSS**

**دکتر محمد علی هرمزی زاده**

سطرها: حجم نمونه

ستون ها: تعداد متغیرها

هر سطر اطلاعات مربوط به واحد تحلیل است.

واحدهای تحلیل، واحدهایی هستند که در جمع آوری اطلاعات درباره آنها اطلاعات جمع آوری می کنیم و اطلاعات آنها را بررسی و تحلیل می کنیم که می توانند عکس های یک روزنامه یا افراد یا SMS و غیره باشد. در کل چیزی که درباره آن اطلاعات جمع آوری و تحلیل می کنیم.

در SPSS اطلاعات سطر به سطر وارد می شود یعنی اول مشخصات واحد تحیل اول را وارد می کنیم و بعد به سراغ واحد تحلیل دوم در سطر دوم می رویم.

در VARIABLE VIEW هر سطر یک متغیر را نشان می دهد. برای متغیرهای اسمی و ترتیبی لازم است که کدها یا VALUESها مشخص شود. به عنوان مثال اگر متغیر فاصله ای باشد خودش عدد است و نیا به عدد برای شناسایی ندارد، نیازی به VALUES ندارد اما اگر ترتیبی بود مانند کد 1 یعنی جدول آنگاه لازم است که در VALUES وارد شود و دیگر مثال جنس که نیاز به کد دارد.

در MISSING معمولاض 9 یا 99 یا 999 را دز نظر می گیریم. باید توجه کرد که مثلاً اگر 99 را برای سن MISSING می گذاریم نباید کسی سنش 99 سال باشد.

DISCERT MISSING VALUSE برای این است که ممکن است چند نوع MISSING داشته باشیم و می توان تا 3 نوع MISSING وارد کرد. مثلاً کسی سنش را نمی گوید به شخص دیگر یا یادش نمی آید و غیره و برای ما فرق می کند و مهم است بین کسی که نمی داند و یا نمی گوید می توان 2 نوع کد متفاوت MISSING داد.

از طریق CTRL+HOME و CTRL+END حجم نمونه را بدست می آوریم.

ستون ها شامل متغیرهای مختلف است.

برای آشنایی با با فایل های SPSS به VARIABLE VIEW می رویم تا با متغیرها آشنایی پیدا کنیم.

NOMINAL متغیر اسمی

ORDINAL متغیر ترتیبی

LABLE قسمتی است که متغیر تعریف می شود

MISSING

RS OCCUPATIONAL PRESTIGE SCORE نمره منزلت شغلی فرد پاسخ دهنده

معمولاً شغل هایی که نساز به تخصص دارند از منزلت بالایی برخوردار هستند. درآمد، تحصیلات و احترام بیشتر و غیره از ویژگی های شغل های دارای منزلت بیشتر است.

می توانیم سه پاسخ را در قالب یک کد درج کنیم.

VALUE رده ها یا کدهای یک متغیر است.

کد به صورت دستی در لیست مخصوص به خودش تایپ می شود.

 در متغیرهای ترتیبی به آن کسانی که مقدار متغیرشان کم است کد کمتری بدهیم و به آن کسانی که مقدار متغیرشان بیشتر است کد بیشتری بدهیم. مانند اینکه شما در طول روز چقدر مطالعه می کنید؟

1. اصلاً مطالعه نمی کنم
2. خیلی کم
3. کم
4. متوسط تا حدی
5. زیاد
6. خیلی زیاد

هر چقدر میزان مطالعه زیاد می شود کدها هم زیاد می شود چرا که ممکن است نتایج برعکس شود. به عنوان مثال 1517 سطر عدد وارد کردیم و می خواهیم به کار پردازش و تحلیل بپردازیم اولین کار این است که از تک تک متغیرها جدول فراوانی FREQUENCY بگیریم.

دستورات آماری عمدتاً در ANALYSE قرار دارند.

ANALYSE → DESCRIPTIVE STATISTICS → FREQUENCIES.

یک DIALOGUE BOX باز می شود و می توانیم دستورهایمان را وارد کنیم. همه متغیرها در قسمت چپ وارد می شود. هر کدام از متغیرها را وارد کردیم گزینه فلش را می زنیم و به قسمت VARIABLESS منتقل می شود تا بتوانیم FRQUENCY بگیریم. بعد از اینکه OK کردیم به ما OUTPUT می دهد. در این پنجره خروجی ستون سمت چپ وجود دارد که در آن فهرست خروجی ها قرار دارد و در سمت راست خود خروجی قرار دارد.

TITLE عنوان

STATISTICS آمار

CUMULATIVE PERCENT درصد تجمعی

دستورالعمل کد گذاری متغیر برای پایان نامه

تعداد برادرها و خواهرها را به سه طبقه تقسیم کنیم.

1. کم (2-0)
2. متوسط (6-3)
3. زیاد (7 به بالا)
4. بی پاسخ

برای ضریب همبستگی یا تحلیل رگرسیون باید همان متغیر با اطلاعات اولیه باشد. 4 طبقه مانند بالا می سازیم.

برای این کار به منوی TRANSFORMمی رویم تا RECODE کنیم.

TRANSFOR → RECODE → INFO SAME VARIABLES OR INFO DIFFERENT VARIABLES.

فرق SAME با DIFFERENT: SAME در همان متغیر RECODE می شود و داده های اولیه برای همیشه از بین می رود. DIFFERENT به صورت یک متغیر دیگری درمی آورد و داده های اولیه دست نمی خورد.

وقتی DIFFRENT را انتخاب کردیم یک دیالوگ باکس باز می شود و در آن می توانیم متغیر را انتخاب کنیم. بعد از انتخاب متغیر، اسم جدید به آن می دهیم و پس از آن کدهای جدید را از OLD AND NEW انتخاب می کنیم. اگر یک کد را به یک کد انتخاب کنیم گزینه VALUE را انتخاب می کنیم.

اگر می خواهیم از یک عدد تا یک عدد، یکی را انتخاب کنیم گزینه RANGE را انتخاب می کنیم، در قسمت NEW VALUE کد را وارد و سپس گزینه ADD را می زنیم.

 قسمت RANG THROUGH HIGHES برای زمانی است که می خواهیم 7 به بالا را انتخاب کنیم. بعد از اینکه OK کردیم آخرین متغیر به عنوان مثال به نام SIBS2 شناخته شده و به آخر متغیرها اضافه می شود.

بعد از اینکه FREQUENCY گرفتیم جدول به صورت خلاصه ارائه می شود ولی متغیرها تعریف نمی شود که بر همین اساس در ستون DECIMALS دو رقم اعشاری که برای متغیر به وجود می آید را صفر می کنیم و سپس در قسمت VALUES متغیرها را تعریف می کنیم تا در جدول خلاصه شده مشکلی برای تعریف متغیرها و اعشار آنها نداشته باشیم.

به عنوان مثال اول یک FREQUENCY از سن می گیریم. بالاترین سن 89 و کمترین 18 سال است و تفاضل آنها را بدست می آوریم.

89 – 18 = 71 + 1 = 72 ÷ 4 = 18

بعد به ترتیب وارد نرم افزار SPSS می کنیم.

18 – 35

36 – 53

54 – 71

72 – 89

بی پاسخ

فرق بین دکمه OK و PAST در E DIALOGUE BOX: وقتی که از طریق منوی ANALYSE گزینه DESCRIPTIVE STATISTICS را انتخاب و گزینه FREQUENCY را انتخاب می کنیم دیالوگ باکس FREQUENCY باز می شود که بعد از انتخاب متغیرها و و ورود آنها به قسمت VARIABLE (S) اگر از گزینه OK استفاده کنیم فوراً دستور را اجرا می کند ولی اگر از گزینه PASTE استفاده کنیم دستوری برای ما نوشته می شود که عنوان آن SYNTAX است که شامل دستورات مختلفی است و می توان آن را جداگانه ذخیره کرد. این کار باعث کاهش زمان در اجرای متغیرها می شود.

PERCENTILE VALUES درصد متغیرها

دکمه STATISTICS مربوط به آماره هاست.

QUARTILES چارک ها

PERCENTILE صدک ها

CUT POINTS OF 10 EQUAL GROUP

قسمت CONTRAL TENDENCY شامل MEAN میانگین، SUM مجموع، MODE نما و MEDIAN میانه است.

قسمت DISGERSION شامل مواردی همچون:

STD. DEVIATION, MINIMUM, RANGE, VARIANCE, MAXIMUM, S.E MEAN است.

S.E MEAN اشتباه معیار

**اشتباه معیار چیست؟**

فرمول انحراف معیار

STD. DEV. =

STD. ERROR =

اگر انحراف معیار را تقسیم بر جذر تعداد نمونه کنیم اشتباه معیار به دست می آید. اشتباه معیار در حقیقت انحراف معیار میانگین است. یعنی اینکه اگر ما از یک جامعه آماری بیابیم نمونه های متعددی بگیریم.

X = مدت زمان تماشای تلوزیون

X1 = 5 ساعت

X2 = 10 ساعت

X3 = ساعت1

X4 = 0 ساعت

X5 = 8 ساعت

 میانگین انحراف معیار S = 2 n = 400

DISTRIBUTION شکل توزیع

SKEWNESS ضریب چولگی

KURTOSIS ضریب کشیدگی

ضریب چولگی نشان می دهد که آیا توزیع متقارن هست یا نیست.

KU < 0

کشیدگی منفی

KU > 0

کشیدگی مثبت است

KU = 0

کشیدگی متعادل است

SK < 0

چولگی منفی

SK > 0

چولگی مثبت

SK = 0

فاقد چولگی

کشیدگی وقتی منفی باشد به منزله این است که پراکندگی بیشتر است.

قسمت VALUES ARE GROUP MIDPOINTS مقادیر نقاط میانی گروه ها هستند.

زمانی که سن دقیق افراد را نداریم و فقط گروه های سنی را داریم و ناچار شدیم نقاط میانی را به عنوان سن افراد بنویسیم این قسمت را فعال می کنیم.

|  |  |  |
| --- | --- | --- |
| سن | فراوانی | نقطه میانگین MID POINT |
| 24-15 | 18 | 20 |
| 34-25 | 18 | 30 |
| 44-35  | 42 | 40 |

نقطه میانی تقریبی است بنابراین بهتر است این کار را نکنیم مگر اینکه ناچار باشیم.

دکمه CHART یا نمودار به صورت پیش فرض بر روی NONE قرار دارد.

BAR CHART نمودار میله ای که برای متغیرهای ترتیبی است.

PIE CHART نمودار دایره ای

HISTOGRAM هیستوگرام

قسمت FORMAT:

ORDER BY نحوه تنظیم طبقات

ASCENDING VALUES مقادیر صعودی. اگر به صورت صعودی باشد بر اساس فراوانی مرتب می شود نه بر حسب مقدار متغیر.

DESCENDING VALUES مقادیر نزولی. به صورت نزولی و بر اساس فراوانی مرتب می شود. زمانی که متغیر ما اسمی است و می خواهیم بر حسب فراوانی مرتب کنیم از این گزینه استفاده می کنیم.

قسمت VARIABLE MULTIPLE متغیرهای چندگانه

COMPARE VALUES اول آماره ها، بعد جدول و بعد نمودارها را می دهد.

ORGANIZE OUTPUT BY VARIABLES این گزینه همه چیز شامل آماره ها، جداول و نمودارها را می دهد.

با استفاده از این گزینه، جداولی که تعداد آنها بیش از یک رقم خاصی که می توان آنها را تنظیم کرد، حذف می شود. با این کار جداول کوچک که به کار محقق می آید نشان داده می شود.

تمرین:

PROB 2

1. خانوادگی
2. مالی
3. روانی
4. شغلی
5. تحصیلی
6. بیماری
7. مسکن

EDIT → OPTION → OUTPUT LABLE → NAMES AND LABLE

روی کاغذ دستورالعمل کدگذاری را می نویسیم.

1 → 2 مالی

2 →1 خانوادگی

3 → 4 شغلی

4 → 7 مسکن

5 → 5 تحصیلی

6 → 3 روانی

7 → 6 بیماری

8 → ELS بی پاسخ

ANALYZE → DESCRIPTIVE STATISTICS → DESCRIPTIVES → SAVE STANDARDIZED VALUES AS VARIABLE مقادیر استاندارد شده را به عنوان متغیر ذخیره کن

اگر میانگین جامعه و انحراف معیار کل جامعه را داشته باشیم از فرمول Z استفاده می کنیم در غیر اینصورت از فرمول T استفاده می کنیم.

z = پارامتر

ϻ = میانگین جامعه

σ = انحراف معیار کل جامعه

t = آماره

 = میانگین نمونه

s = انحراف معیار نمونه

آماره: ویژگی های عددی جامعه نمونه است.

پارامتر: ویژگی های عددی جامعه آماری است. مانند محاسبه میانگین کل سن جمعیت مردم ایران که به آن 31 = ϻ سال می گویند.

S = 10

ϻ = 30

xi = 40

ϻ = 20

18

.

.

.

89

سن فرد یک انحراف معیار از میانگین بیشتر است یا کمتر است.

در متغیر x واحد سنجش ما سال است. در متغیر z واحد سنجش انحراف معیار است.

سن این فرد مساوی با میانگین است. کسی که بالای میانگین باشد عدد استانداردش مثبت می شود و کسی که پائین میانگین باشد عدد استاندارد آن منفی است.

متغیر تعداد فرزندان x1 1000 – 10 →

سن x2 2000 – 90 →

درآمد x3 5800000000 – 12000000 →

اگر تیک SAVE را بزنیم در ستون داده ها نمره Z هر کدام از متغیرها را محاسبه کرده و یک متغیر جدید به دست می آورد.

DISPLAY ORDER ترتیب نمایش

ANALYZE → DESCRIPTIVE STATISTICS → DESCRIPTIVES → OPTIONS → DISPLAY ORDER → VARIABLE LIST. طبق لیست متغیرها

ALPHABET الفبایی

ASCENDING MEANS ترتیب صعودی

DESCENDING MEANS ترتیب نزولی

فرض کنیم یک پرسشنامه داریم و اسم 10 رسانه را در آن آورده ایم و از پاسخگو می پرسیم که از کدام رسانه بیشترین استفاده را دارد و درصد و جدول و رسانه ای که کمترین استفاده را دارد در ذیل جدول قرار گیرد به همین خاطر از گزینه های DESCENDING MEANS و ASCENDING MEANS استفاده می کنیم.

**دستور EXPLORE یا اکتشاف**

ANALYSE → DESCRIPTIVE → EXPLORE

LOWER BOUND حد پائین

UPPER BOUND حد بالا

فرض کنید جامعه آماری میلیونی داریم . یک نمونه 100 نفری از آن را انتخاب کردیم. میانگین تحصیلات کل جامعه چقدر است؟

N = 100000

N = 1000

تخمین دو نوع است:

1. نقطه ای: یعنی اینکه همین که میانگین نمونه است، میانگین جامعه هم هست.
2. فاصله ای

تخمین فاصله ای نمی گوید میانگین 88/12 است بلکه می گوید میانگین با فاصله مشخص بین این عدد تا آن عدد است با میزان اطمینان 95 درصد.

میزان اعتماد 95 درصدی میانگین کل جامعه ϻ عددی بین 73/12 تا 03/13 است.

%5 TRIMMED MEAN میانگین اصلاح شده 5%

زمانی که متغیر ما دارای چولگی باشد، یعنی مقادیر خیلی کم یا خیلی زیاد داشته باشد در آنجا آماره، میانگین خوبی نیست و محاسبه آن به سمت چولگی کشیده می شود. میانگین اصلاح شده 5/2 درصد از اعداد خیلی کوچک و 5/2 درصد از اعداد بزرگ را کنار گذاشته و 95 درصد که حالت کم یا زیاد هستند را میانگین آنها را در نظر می گیرد. 5 درصد افراد این میانگین حذف می شوند.

MEDIAN میانه

STEAM AND LEAF PLOT نمودار شاخ و برگ شبیه هیستوگرام است ولی 90 درصد چرخیده است.

DISPLAY → BOTH → STATISTICS → PLOTS

M – STIMATORS

نمودار ها و آزمون های نرمال بودن

PLOTS → NORMALITY PLOTS WITH TESTS

برای اینکه بفهمیم متغیر ما نرمال هست یا خیر از این گزینه استفاده می کنیم. پیش فرض این است که آنها 5 درصد را در نظر می گیرند.

فرضیه ما این است که توزیع سن نرمال است. وقتی که SIGNIFICANT که خطای ما در رد فرضیه صفر است اگر از 5 درصد کمتر باشد به راحتی رد می شود اما اگر بالای 5 درصد باشد فرضیه صفر رد نمی شود.

دستور CROSS TABS جداول دو بعدی به بالا می دهد.

ANALYZE → DISCRIPTIVE STATISTICS → CROSS TABS CELLS

روی CELLS می رویم یک دیالوگ باکس باز می شود.

OBSERVED COUNT فراوانی مشاهده شده

EXPECTED COUNT فراوانی مورد انتظار

اگر متغیر مستقل متغیر ستونی است باید از درصد ستونی اضافه کنیم.

اگر متغیر مستقل متغیر سطری است باید از درصد سطری یا ردیفی استفاده کنیم.

اگر متغیر مستقل مشخص نیست باید از TOTAL استفاده کنیم.

وقتی متغیر ستونی باشد باید درصد هم ستونی باشد.

در قسمت STATISTICS از آزمون های آماری استفاده می کنیم.

RESIDUAL باقیمانده ها

UNSTANDARDIZED استاندارد نشده

STANDARDIZE استاندارد شده

ADJ. STANDARDIZE استاندارد شده تعدیل شده. زیاد بکار می رود و در آزمون کای اسگوئر CHI – SQUARE مورد استفاده قرار می گیرد.

|  |  |  |  |
| --- | --- | --- | --- |
| جمع کل | زن | مرد |  |
| 80 | 30 | 50 | بلی |
| 60 | 40 | 20 | خیر |
| 140 | 70 | 70 | جمع کل |

50 – 40 = 10

30 – 40 = - 10

20 – 30 = - 10

40 – 30 = 10

جمع جبری آنها صفر می شود.

اگر متغیر ما اسمی باشد از 4 فرمولی که به ما در قسمت ANALYSE ارائه می دهد استفاده می کنیم.

NOMINAL اسمی

CONTINGENCY COEFFICIENT

ضرایب متقارن یا سیمتریک ضرایبی هستند که در آنها متغیر مستقل یا وابسته فرقی ندارند.

LAMBDA GOOGD MAN AND UNCERTAINTY COEFIFICIENT KRUSKALTAV ضریب اطمینان

بعضی از ضرایب جهت دار یا DIRECTIONAL هستند و بین متغیر وابسته و مستقل تفاوت قائل است.

ارتباط خطی دو نوع است:

1. مستقیم
2. معکوس

مستقیم

معکوس

Y

Y

X

X

ارتباط غیر خطی

احساس خوشبختی

احساس خوشبختی

تحصیلات

تحصیلات

یک متغیر اسمی و متغیر بعدی فاصله ای است. NUMINAL BY INTERVAL

ضریب ETA: ضریبی است که زمانی مورد استفاده قرار می گیرد که ما یک متغیر اسمی و یا ترتیبی داریم و متغیر بعدی ما فاصله ای و یا نسبی است. متغیر فاصله ای نسبی درآمد و متغیر اسمی شغل داریم که در آن برای اندازه گیری شدت، از ضریب ETA استفاده می کنیم.

رده شغلی، متغیر اسمی است.

ضریب KAPPA برای زمانی است که می خواهیم بدانیم آیا دو متغیر ترتیبی با هم یکسان و تعداد رده ها و کدهای آنها مساوی است آیا با هم موافق هستند یا خیر؟ در حقیقت توافق بین دو متغیر یکسان را به ما نشان می دهد.

ضریب RISK برای هر دو متغیر ارزشی است.

ضریب MC NEMAR هر دو متغیر ارزشی باشند ولی به صورت سنجش مکرر باشد. به عنوان مثال در دو زمان مختلف از پاسخ دهنده سئوال می کنیم که آیا روزنامه می خواند یا نه؟

ضریب کوکران زمانی به کار می رود که که ضریبی بین متغیرهای ما وجود دارد و ترتیبی بین رده های متغیر وجود داشته باشد.

SIG = 11/5

9/4

منطقه رد فرضیه H0

α = %5

منطقه قبول فرضیه H0

5/11 در منطقه رد H0 است و رابطه معنادار و قابل تعمیم است.

α = %5

SIG = 0/005

9/4 : کای اسکوئر جدول

5/11 : کای اسکوئر فرمول

زمانی می توانیم از آزمون کی اسکوئر استفاده کنیم که فراوانی مورد انتظار در 20 درصد خانه ها زیر 5 درصد باشد.

SIG پائین تر و یا مساوی آنها باشد فرضیه H0 رد می شود.

آزمون یک یا دو دامنه به فرضیه H1 بر می گردد. اگر فرضیه این باشد که زن ها بیشتر از مردان مطالعه می کنند این فرضیه، یک دامنه است. اکر بگوید میزان مطالعه زنان و مردان یکسان نیست، دو دامنه است به دلیل اینکه خودش دو معنی دارد یکی اینکه زن ها بیشتر از مردان و دیگری زنان بیشتر از مردان هستند.

آزمون دو دامنه

آزمون یک دامنه

1α

α - 1

α

α - 1

2α

آزمون یک دامنه H0 = ϻ1= ϻ2 2H1: ϻ1 > ϻ

آزمون دو دامنه H0 = ϻ1= ϻ2 2H1: ϻ1 ≠ ϻ

ANALYZE → COMPARE MEANS مقایسه میانگین ها

در جدول دو بعدی محاسبه نسبت ها و درصدها صورت می گیرد. در جداول فراوانی و دو بعدی مقایسه فراوانی ها درصدها و نسبت ها. چوت که در جدول، متغیرهای ما یا اسمی هستند و یا ترتیبی.

زمانی که متغیر اسمی یا ترتیبی مانند زن و مرد داریم یا یک متغیر فاصله ای نسبی مانند درآمد داریم، میانگین درآمد مردها و زن ها را بدست می آوریم و آنها ا مقایسه می کنیم که به آنها مقایسه میانگین ها می گویند.

اسمی – اسمی: آزمون کای دو – ضریب فی – V کرامر – لاندا

اسمی – ترتیبی: آزمون کای دو – ضریب فی – V کرامر – لاندا

ترتیبی – ترتیبی: آزمون کای دو – ضریب گاما – D سامرز – TB و TC کندال

کار همه این آزمون ها، مقایسه درصدهایا نسبت ها یا فراوانی هاست.

مقایسه میانگین:

اسمی یا ترتیبی – فاصله ای یا نسبی: آزمون T یا F – ضریب 1 تا که شدت را نشان می دهد.

وقتی متغیری فاصله ای یا نسبی شد می توان میانگین آن را محاسبه کرد.

در آزمون T نوع اول فقط یک گروه را مقایسه می کنیم.

در آزمون T نوع دوم دو گروه را مقایسه می کنیم.

در آزمون T نوع سوم، سه گروه را بررسی می کنیم.

DEPENDENT LIST لیست متغیرهای وابسته

INDEPENDENT LIST لیستم متغیرهای مستقل

به عنوان مثال در تعداد فرزندان به عنوان یک متغیر وابسته می توان میانگین آنها را اندازه گیری کرد.

در OPTION انواع آماره های مختلف قید شده است.

قسمت آماره های مربوط به لایه اول در OPTION: STATISTICS FOR FIRST LAYER

ANOVA TABLE AND EFA آزمون آنالیز واریانس آنووا

TEST FOR LINERITY آزمون خطی بودن

زمانی از آزمون خطی بودن استفاده می کنیم که متغیر مستقل باشد.

ANALYSIS OF VARIANCE (ANOVA) آزمون تحزیه واریانس یا تحلیل واریانس یا آنالیز واریانس. آماره آن آزمون F است. در این آزمون، واریانس به دو قسمت بین گروهی و درون گروهی تجزیه می شود.

آزمون واریانس جزو آزمون هایی است که مقایسه میانگین ها را انجام می دهد. آزمون F و T هم کار مقایسه میانگین ها را انجام می دهند.

کار آزمون T مقایسه دو گروه یا دو میانگین است در حالی که آزمون F مقایسه میان چند گروه و یا میانگین را انجام می دهد. پس اگر دو گروه داریم که می خواهیم میانگین آنها را با یکدیگر مقایسه کنیم از آزمون T استفاده می کنیم.

مستقل ← وابسته

جنس ← درآمد – آزمون T

شغل ← درآمد – آزمون F

متغیر مستقل نشان می دهد که ما از کدام آزمون استفاده کنیم.

آزمون خطی بودن یعنی اینکه با افزایش یک متغیر، آن یکی یا افزایش پیدا کرده یا کاهش پیدا می کند.

درآمد

تحصیلات دکتری

بی سوادی

بین تحصیلات و درآمد رابطه خطی مثبت وجود دارد. مستقیم

درآمد

تحصیلات دکتری

بی سوادی

رابطه خطی معکوس یا منفی. وقتی که متغیر اول زیاد می شود متغیر دوم کم می شود.

هر رابطه ای که به صورت خط راست نباشد غیر خطی است.

تماشای تلویزیون

بی سوادی

دیپلم

تحصیلات دکتری

متغیر اسمی، کیفی است و افزایش یا کاهش بردار نیست. مانند کم یا زیاد شدن شغل افراد.

**مقایسه میانگین هاCOMPARE MEANS :**

ANALYZE → COMPARE MEANS → MEANS

DEPENDENT LISTلیست متغیرهای وابسته که قابل میانگین گیری است.

INDEPENDENT LIST لیست متغیرهای مستقل که گروه ها را برای ما مشخص می کند.

در قسمت OPTION می توان واریانس و غیره را محاسبه کرد.

STATISTICS FOR FIRST LAYER

تحلیل واریانس یا آزمونF ANOVA TABLE AND ETA

آزمون خطی بودن TEST FOR LINEARITY

MEASURES OF ASSOCIATION ضرایب سنجش ارتباط

در آزمون F واریانس به دو قسمت تقسیم می شود، واریانس داخل گروه ها و واریانس بین گروه ها.

واریانس یعنی تغییر پذیری و اختلاف در متغیر.

هر چقدر واریانس درون گروه کمتر باشد و واریانس بین گروه ها بیشتر باشد متغیر مستقل تاثیر بیشتری بر متغیر وابسته دارد.

SUM F SQURES بخش مجذورات. صورت واریانس است.

مجذور: هر چیزی که به توان 2 برسد، مجذور می شود و به دلیل اینکه در فرمول نحوه محاسبه واریانس، تفاضل ها به توان 2 می رسد به آنها مجذورات می گویند.

واریانس: نوعی میانگین مجموع مجذورات است.

MEAN SQUAR یا MS یا میانگین مجذورات، اسم دیگر واریانس است.

هر چقدر F بزرگتر باشد واریانس بین گروه ها نسبت به درون گروه ها بزرگتر خواهد بود.

SIG یا سطح معناداری اگر کمتر از 5 صدم بود رابطه بین دو متغیر معنادار است و در کل جامعه آماری، با 95 درصد اطمینان بین دو متغیر رابطه وجود دارد.

ضریبETA : یک ضریبی است که زمانی بکار می رود که ما یک متغیر اسمی یا ترتیبی و یک متغیر فاصله ای نسبتی داشته باشیم. در این حالت برای محاسبه شدت ارتباط، دیگر نمی توان از هیچ آزمون دیگری استفاده کرد.

قدر مطلق ضریب اگر بین صفر تا 10 درصد باشد، علامت این است که ارتباط بسیار ضعیف است. اگر بین 10 تا 30 درصد باشد، ارتباط ضعیف است. اگر بین 30 تا 50 درصد باشد، شدت ارتباط متوسط است. اگر بین 50 تا 70 درصد باشد، ارتباط قوی است و اگر 70 درصد تا 1 باشد ارتباط بسیار قوی است. اگر 1 باشد ارتباط بین متغیرها به صورت کامل برقرار است.

مثبت

منفی

Y

Y

X

X

مانند عرضه و تقاضا که هر چقدر یک جنسی در بازار زیاد می شود قیمت آن کاهش پیدا می کند.

نحوه مقایسه گروه ها با یکدیگر:

اگر 10 گروه داشته باشیم:

**آزمون خطی بودن TEST FOR LINEARITY:**

آیا ارتباط خطی بودن بین دو متغیر وجود دارد یا خیر؟ یعنی اینکه هر دو متغیر باید کمی باشد و اگر یکی از آنها کیفی باشد ارتباط معنی نمی دهد. با بالا رفتن یک متغیر کمی، متغیر کمی بعدی هم افزایش پیدا می کند.

ارتباط:

1. مثبت
2. منفی

ارتباط:

1. خطی
2. غیر خطی

تماشای تلویزیون Y

تماشای تلویزیون Y

تحصیلات X

تحصیلات X

X

X

Y

Y

خط منفی

خط مثبت

امکان دارد که هم خطی بودن و هم انحراف از خطی بودن، معنادار شوند.

**آزمون های T:**

1. **نوع اول:** مقایسه میانگین یک نمونه با یک عدد ثابت است.
2. **نوع دوم:** مقایسه میانگین دو نمونه مستقل است مانند مقایسه میانگین حقوق کارمندان وزارت نفت و صدا و سیما.
3. **نوع سوم:** مقایسه میانگین دو نمونه وابسته است مانند مقایسه میانگین زن ها و شوهرها یا خواهرها و برادرها.

**آزمون T نوع اول:** زمانی بکار می رود که ما یک میانگین بیشتر نداریم و می خواهیم ببینیم آیا میانگین ما با آن عدد ثابت تفاوتی دارد یا خیر؟ آزمون T نوع اول مقایسه میانگین نمونه با یک عدد ثابت است.

ANALYZE → ONE SIMPLE T TEST

سئوال: آیا میانگین متغیر تحصیلات با عدد 12 تفاوت معنا دار دارد یا خیر؟ یا آیا تحصیلات در جامعه آمریکا بیش از 12 سال است؟

H0: ϻ = 12 ϻ: میزان تحصیلات در جامعه آمریکا

H1: ϻ ≠ 12 فرضیه صفر رد می شود و فرضیه یک را می پذیریم چرا که بیش از 12 شده است.

فرضیه H0 همیشه تفاوت را رد می کند و می گوید تفاوتی وجود ندارد.

تفاوتی که در کل جامعه وجود دارد واقعی است و اگر قابل تعمیم به کل جامعه باشد به آن تفاوت معنادار می گویند. اگر تفاوت معنادار باشد در مورد کل جامعه هم قابل صدق است.

فرضیه صفر، فرضیه بی تفاوتی است و همیشه تفاوتی وجود ندارد.

اگر تفاوت معنادار نباشد، هم ناشی از خطای نمونه گیری است و هم اینکه قابل تعمیم به جامعه نیست.

اگر SIG بیشتر از 5 درصد باشد آن موقع فرضیه H0 تائید می شود و تفاوت تعمیم داده نمی شود. اگر SIG کمتر از 5 درصد باشد آن موقع فرضیه H1 تائید می شود و تفاوت قابل تعمیم است.

**آزمون T نوع دوم:** مقایسه میانگین دو نمونه مستقل با هم

ANALYZE → INDEPENDENT SAMPLE T TEST

TEST VARIABLE متغیری که می خواهیم آزمایش کنیم

GROUPING VARIABLE متغیرها را گروه بندی می کنند

در آزمون های T و F که برای مقایسه میانیگن است قاعدتاً باید واریانس ها برابر باشد. اگر واریانس ها برابر باشد از یک فرمول می رویم و اگر نابرابر باشد از یک فرمول دیگر می رویم.

LEVENES TEST FOR EQUALITY OF VARIANCE آزمون لوین برای برابری واریانس ها

اگر SIG آزمون LEVENE از 5 درصد بیشتر بود از سطر بالا استفاده می کنیم اگر زیر 5 درصد بود از سطر پائین استفاده می کنیم.

همیشه باید به این نکته توجه کنیم که اگر SIG در همه آزمون ها زیر 5 درصد بود فرضیه صفر رد می شود.

در آزمون لوین، فرضیه صفر این است که واریانس ها نابرابر هستند که رد می شود.

**آزمون T نوع سوم:** مانند رابطه تحصیلات زن ها و شوهرها

ANALYZE → PAIRED SIMPLES T TEST

دستور EXPLORE

ANALYZE → DESCRIPTIVE STATISTICS → EXPLORE

DEPENDENT متغیر وابسته. متغیری است که در EXPLORE میانگین و میانگین 5 درصد و کلاً آماره های آن محاسبه می شود. باید سطح سنجش آن فاصله ای یا نسبی باشد که بتوان میانگین آن را حساب کرد.

FACTOR LIST به تفکیک گروه های فاکتور یا طبقه ها، میانگین ها و واریانس ها را به ما می دهد.

اگر در نظر بگیریم که درآمد، DEPENDENT یا متغیر وابسته باشد و FACTOR جنسیت باشد که در این حالت FACTOR میانگین و واریانس و انحراف معیار را به تفکیک زن و مرد به ما می دهد.

جدول CROSS TAB یا دو بعدی:

|  |  |  |
| --- | --- | --- |
|  | مرد | زن |
| درآمد کم | 15% | 25% |
| درآمد متوسط | 55% | 40% |
| درآمد زیاد | 30% | 35% |
| جمع کل | 100% | 100% |

در جدول دو بعدی فقط فراوانی و درصد داریم و نمی توانیم برای آنها انحراف معیار و واریانس و میانگین را محاسبه کنیم چرا که در جدول دو بعدی متغیرهای ما اسمی و ترتیبی هستند که نمی توان میانگین آنها را گرفت ولی می توان درصد آنها را محاسبه کرد.

جدول EXPLORE در SPSS انواع آماره ها را به ما می دهد.

دو نمونه جفت شده و دو نمونه وابسته: دو نمونه جفت شده یک گروه نمونه ثابت داریم که دو بار و به صورت پیش آزمون و پس آزمون، از آنها آزمون می گیریم. دو نمونه وابسته دو گروه هستند که جفت جفت به یکدیگر وابسته هستند.

آزمون آنوا ANOVA:

آزمون آنوا همان آزمون F است. تجزیه و تحلیل واریانس ANALYSIS OF VARIANCE (ANOVA)

ANALYZE → COMPARE MEANS → ONE – WAY ANOVA

در آزمون T می خواهیم دو میانگین را با هم مقایسه کنیم اما در آزمون F چند میانگین را با هم مقایسه می کنیم. مانند میزان اعتماد مردم کشور به صداوسیما که با متغیر فاصله ای، نمره بین 20 تا 80 بدست آمده است که هر چه به 20 نزدیک تر باشد میزان اعتماد کمتر است. به دلیل اینکه 31 استان داریم پس در نتیجه 31 میانگین داریم. مانند میانگین اعتماد مردم در استان آذربایجان شرقی 53، تهران 41 و غیره است و این میانگین ها با یکدیگر تفاوت دارند و برای اینکه ببینیم بین این میانگین ها تفاوت وجود دارد از آزمون F یا آنوا استفاده می کنیم. از این جهت که کار این آزمون مقایسه میانگین هاست شبیه آزمون T است و به همین دلیل در قسمت COMPARE MEANS وارد شده است.

آزمون لوین کار مقایسه واریانس ها را بر عهده دارد که بر همین اساس با آزمون آنوا که به آن آزمون تجزیه و تحلیل واریانس که کارش مقایسه بیش از 2 میانگین است می گویند.

علت نامگذاری آزمون تجزیه و تحلیل واریانس این است که در این آزمون کل واریانس به دو بخش درون گروهی و برون گروهی تقسیم می شود و هر چقدر واریانس بین گروه ها بیشتر باشد و واریانس درون گروه ها کمتر باشد نشان دهنده این است که تفاوت میانگین ها با یکدیگر بیشتر و معنادارتر است. مانند میزان اعتماد مردم کشور به صداوسیما اگر میانگین اعتماد در یک استان 22 و در استان دیگری 68 باشد علامت این است که بین استان ها تفاوت معنی دار وجود دارد اما در یک استان میانگین اعتماد مردم خیلی پائین و در حد 22 است و میانگین آن هم 22 می شود. همچنین در یک استان دیگری میانگین 67 است که در این حالت واریانس درون گروهی کم است ولی واریانس بین گروهی زیاد است. اگر این طور باشد آماره F عدد بزرگی می شود که این آماره از تقسیم واریانس بین گروه ها بر واریانس درون گروه ها به دست می آید.

ما در تحلیل واریانس از طریق تجزیه واریانس به دو بخش درون گروهی و برون گروهی اختلاف میانگین ها را محاسبه می کنیم.

در آزمون ANOVA متغیر وابسته باید متغیر فاصله ای نسبتی باشد که بتوانیم میانگین و واریانس آن را محاسبه کنیم. FACTOR یا متغیر مستقل هم می تواند اسمی یا ترتیبی باشد که چند گروه را شامل می شود چرا که می خواهیم میانگین متغیر وابسته را بر اساس تفکیک متغیر مستقل محاسبه کنیم. مانند نژاد که شامل سه طبقه است و در متغیر مستقل یا فاکتور قرار می گیرد را با تعداد فرزندان که در لیست متغیر وابسته یا DEPENDENT قرار دارد مقایسه کنیم. نتیجه حاصل شده هم درون گروهی و هم بین گروهی را بدست می آورد.

SUM OF SQUARES مجموع مجذورات یا صورت واریانس است.

صورت واریانس یا مجموع مجذورات یا SS

چرا به صورت واریانس مجموع مجذورات می گویند؟

مجدور یعنی اینکه به توان 2 رسیده است. وقتی که تفاضل ها به توان 2 برسد مانند و جمع بشود به آن مجموع مجذورات می گوئیم.

|  |  |  |  |
| --- | --- | --- | --- |
|  | سرخ پوست | سیاه پوست | سفید پوست |
| واریانس بین گروهی |  |  |  |
|  | 05294512 | 7145120169 | 11002247 |
| واریانس درون گروهی |  |  |  |

نحوه محاسبه df بین گروه ها:

 ? = 1 - تعداد گروه ها

 2 = 1 - 3 K – 1 =?

درجه آزادی اعضای گروه ها:

? = تعداد گروه ها - تعداد کل

N – K = ?

1509 – 3 = 1506

جمع دو درجه آزادی، درجه آزادی کل می شود.

F از تقسیم واریانس بین گروه ها بر درون گرو ها حاصل می شود.

SIG =0/002 به دلیل اینکه کمتر از 5 صدم شده، پس تفاوت معنادار است.

اگر در آزمون ANOVA و از طریق گزینه OPTION قسمت DESCRIPTIVE را فعال کنیم، انحراف استاندارد، اشتباه استاندارد و غیره را به ما می دهد.

آزمون همگونی واریانس TEST OF HEMOGENITY OF VARIANCES

آزمون لوین برای مقایسه واریانس هاست.

POST HOC یا آزمون های تعقیبی: زمانی بکار می رود که آزمون ANOVA معنی دار شده باشد و به ما می گوید که کدام گروه ها با کدام یک از گروه ها تفاوت معنادار دارد.

با فرض برابری واریانس ها EQUAL VARIANCE ASSUMED

با فرض نابرابری واریانس ها EQUAL VARIANCE NOT ASSUMED

در گزینه POST HOC تمامی آزمون های تعقیبی قرار دارد و زمانی بکار می رود که بین واریانس ها تفاوت معنادار نباشد. اگر تفاوت معنادار باشد از گزینه EQUAL VARIANCE NOT ASSUMED استفاده می کنیم که 4 آزمون تعقیبی دارد.

بررسی متغیر ترکیبی مستقل با یک متغیر وابسته سه چیز را نشان می دهد. به عنوان مثال تاثیر متغیر نژاد و جنس بر روی منزلت شغلی.

1. تاثیر متغیر اصلی نژاد
2. تاثیر متغیر اصلی جنس
3. تاثیر ترکیب نژاد و جنس

ممکن است وقتی از آزمون ONE – WAY ANOVA استفاده می کنیم نتایج با FACTORIAL ANOVA متفاوت باشد چون متغیر را با توجه به یک متغیر دیگر می سنجد.

برای FACTORIAL ANOVA از ANALYZE → GENERAL LINEAR MODEL →UNVARIABLE اقدام می کنیم.

RANOM FACTOR یعنی خود عامل یا فاکتور به صورت رندوم انتخاب شده است مانند اینکه از تمام استان ها انتخاب نکردیم بلکه بین آنها انتخاب کردیم.

COVIARIATE برای متغیرهای فاصله ای نسبتی است چون ما معمولاً ترتیبی داریم.

1. در هر متغیر اسمی یا ترتیبی: آزمون
2. یک متغیر اسمی یا ترتیبی و دیگری فاصله ای نسبتی: آزمون T و T و FACTORIAL
3. هر دو متغیر فاصله ای یا نسبتی: ضریب همبستگی پیرسون و تحلیل رگرسیون

ANALYZE → CORRELABLE ضریب همبستگی:

ضریب پیرسون یک ضریب متقارن SYMMETRIC است. رگرسیون نامتقارن است. در FACTORIAL هم نامتقارن است.

علت نسبت به معلول تقدم زمانی دارد و بدین شکل متغیر وابسته و مستقل را تمییز می دهیم.

**ضرایب همبستگی دو متغیره** :ANALYZE → CORRELATE → BIVARIATE برای پیدا کردن رابطه بین متغیرها مورد استفاده قرار می گیرد. مانند متغیرهای فاصله ای. جهت ارتباط، شدت ارتباط، معنی داری ارتباط.

بعضی متغیرها مانند جنس و سن را به هیچ وجه نمی توان به عنوان متغیر مستقل قرار داد.

BIVARIATE → CORRELATION COEFFICIENTS → PEARSON – SPEARMAN – KENDULS TU B

در این قسمت آزمونهای پیرسون، تابی کندال و اسپیرمن قرار دارد و بر اساس سنجش متغیرها، مشخص می کنیم که کدامیک از این آزمونها مناسب است.

آزمون معنی داریANALYZE →CORRELATE → BIVARIATE → TEST OF SIGNIFICANCE

ONE – TAILED یک دامنه

TWO – TAILED دو دامنه

FLAG SIGNIFICANCE CORRELATION ضرایب معنی داری را مشخص می کند.

BIVARIATE → OPERATION اطلاعات اضافه تر در این قسمت وجود دارد.

ANALYZE →CORRELATE → PARTIAL ضریب همبستگی جزئی یا ضریب همبستگی تفکیکی: ضریب همبستگی جزئی، ضریب همبستگی بین چند متغیر را به ما نشان می دهد. مانند اینکه آیا تحصیلات و منزلت اجتماعی تحت تاثیر برادرها و خواهرها هستند یا خیر؟

دو متغیری که می خواهیم بررسی کنیم در قسمت VARIABLES قرار می گیرد و متغیری که می خواهیم تاثیر آن را بر روی دو متغیر محاسبه کنیم در قسمت CONTROLLING FOR قرار می گیرد. مانند رابطه بین تحصیلات و منزلت اجتماعی از تعداد برادرها و خواهرها مستقل است. با این کار به جای رابطه کاذب، رابطه خالص بین دو متغیر مستقل به دست می آید.

52%+

منزلت شغلی

تحصیلات فرد

45%+

رابطه قوی و مثبت

تحصیلات همسر

با حذف تاثیر تحصیلات همسر، میزان تاثیر گذاری تحصیلات فرد بر منزلت شغلی به 45% درصد کاهش پیدا می کند.

گاهی اوقات رابطه دو متغیر، رابطه ای کاذب است و تاثیر بین دو متغیر صحیح نیست و از یک متغیر دیگری ناشی شده است.

رابطه کاذب

الف

ب

ج

ما فکر می کنیم بین دو متغیر رابطه وجود دارد ولی در اصل اینطور نیست و یک متغیر دیگری که آن را کنترل می کنیم باعث ایجاد رابطه بین دو متغیر ما شده است. مانند حوادث آتش سوزی که در دنیا اتفاق می افتد و محققین متوجه شدند بین تعداد آتش نشان ها و میزان خسارت، وابستگی عمیقی وجود دارد. هر چقدر آتش نشانان بیشتری اعزام می کنند، حجم آتش نیز بیشتر می شود.

+

تعداد آتش نشان ها

میزان خسارت

این رابطه کاذب است

حجم آتش سوزی

بین تحصیلات همسر فرد و منزلت شغلی او چه رابطه ای وجود دارد. پس از آن این رابطه را با یک سری متغیرهای دیگر کنترل کنیم تا نشان دهد کدامیک کاذب است. ضریب همبستگی گرفتن برای متغیرهای اسمی از ریشه اشتباه است و باید از آزمون آنوا استفاده کرد. مانند تحصیلات فرد، جنس و نژاد و غیره.

**تحلیل رگرسیون** ANALYZE → REGRESSION

 ضریب همبستگی پیرسون، شدت ارتباط را مشخص می کند.

اگر صفر شد، بین دو متغیر رابطه خطی وجود ندارد. r = 0

میزان درآمد 1500000 - 900000

میزان اعتماد 80 – 0

رابطه خطی یا مثبت است یا منفی. ارتباط بین دو متغیر می تواند خطی و یا غیر خطی باشد. خطی می تواند به صورت مثبت یا مستقیم و یا منفی و معکوس باشد. غیر خطی می تواند به صورت u شکل یا به صورت n شکل باشد.

مثبت (مستقیم ): r>0 عدد مثبت می شود. جهت خط در نمودار به سمت بالا است.

خطی

منفی (معکوس): r<0 عدد منفی می شود. جهت خط در نمودار به سمت پائین است.

ارتباط

U شکل r=0 مانند رابطه بین تحصیلات و تماشای تلویزیون

n شکل r=0

غیر خطی

در حالتی که در نمودار، نقطه های پراکنده وجود داشته باشد باز هم r=0 خواهد بود و رابطه خطی وجود ندارد.

**ضریب eta برای ارتباط غیرخطی:** اگرضریب eta و r صفر شد، اصلاً هیچ رابطه ای وجود ندارد اما اگر خلاف این بود، ارتباط غیر خطی وجود دارد.

ارتباط خطی یعنی اینکه ارتباط را به صورت یک خط مستقیم صعودی و یا نزولی وجود داشته باشد تا رابطه نشان داده شود.

**فرق بین ضریب همبستگی پیرسون r و رگرسیون b:**

اگر رابطه بین یک متغیر با یک متغیر دیگر خطی باشد، تحلیل رگرسیون خطی می شود و اگر رابطه بین یک متغیر با یک متغیر دیگر خطی باشد، تحلیل رگرسیون، غیر خطی می شود. در تحلیل رگرسیون خطی ساده، یک متغیر ساده با یک متغیر مستقل سنجیده می شود. مانند تاثیر تحصیلات بر درآمد.

سادهx→y

خطی

چند عاملی

تحلیل رگرسیون

غیر خطی

در تحلیل رگرسیونی برای اینکه رابطه بین متغیرها را متوجه شویم از داده های ریاضی استفاده می کنیم.

متغیر مستقل x ضریب رگرسیون y=a+b متغیر وابسته

تحلیل رگرسیون، یک تحلیل نامتقارن است و باید متغیر مستقل و وابسته مشخص شود وگرنه اگر جای آنها را عوض کنیم، یک تحلیل رگرسیونی جدید بوجود می آید. تحلیل رگرسیون به ما قدرت پیش بینی متغیر وابسته از روی متغیر مستقل را می دهد. تحلیل رگرسیونی در ادامه تحلیل ضریب همبستگی پیرسون بدست می آید.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | y درآمد |  x تحصیلات |
| 722500 | 25 | 4250 | -850 | -5 | 900 | 5 |
| 1562500 | 36 | 7500 | 1250 | 6 | 3000 | 16 |
| 62500 | 0 | 0 | 250 | 0 | 2000 | 10 |
| 422500 | 1 | 650 | -650 | -1 | 1100 | 9 |
| 2770000 | 62 | 12400 | 0 | 0 | 7000 | 40 |

y = a + bx

 = درآمدa + b (تحصیلات)

y = -250 + 200 (x)

درآمد = 250-+ 200 (تحصیلات)

a = عرض از مبدا یا عدد ثابت

a مقدار y را در زمانی که x مساوی با صفر است نشان می دهد.

a = اگر کسی تحصیلاتش صفر باشد ماهیانه 250 هزار تومان بدهکار می شود.

b یا رگرسیون = در ازای هر یک واحد افزایش x، y چقدر تغییر می کند.

اگر y منفی بود یعنی اینکه با افزایش تحصیلات، درآمد هم کم می شود. در ازای هر یک سال افزایش تحصیلات، درآمد 200 هزار تومان افزایش پیدا می کند.

ANALYZE → REGRESSION → LINEAR

INDEPENDENT متغیر مستقل

DEPENDENT متغیر وابسته

تحلیل رگرسیونی، نامتقارن است و باید متغیرهای مستقل و وابسته مشخص شوند. مانند سنجش منزلت شغلی به عنوان متغیر وابسته بر تحصیلات به عنوان متغیر مستقل.

R بزرگ یا MULTIPLE R یا همبستگی چندگانه زمانی رخ می دهد که می خواهیم تاثیر چند متغیر مستقل مانند تحصیلات، درآمد و غیره را محاسبه کنیم. چون در اینجا می توانستیم بیش از چند متغیر مستقل داشته باشیم در نتیجه از R بزرگ استفاده می کنیم در غیر اینصورت از r کوچک استفاده می کنیم.

R SQUARE یا مربع R: اگر ضریب پیرسون را به توان 2 برسانیم R SQUARE بدست می آید. وقتی که ما ضریبی را به توان 2 می رسانیم، مربع شده و تفسیر آن مناسب تر می شود.

r = 0/946 رابطه مثبت وجود دارد.y وx بین

نسبتی از یک واریانس متغیر است که واریانس دیگر می تواند آن را تبیین کند.

90 درصد از درآمد یک فرد به دلیل تحصیلات است و 1 درصد دیگر ناشی از کارهای دیگر است. وقتی r را به توان 2 برسانیم نسبتی از تغییرات وابسته که متغیر مستقل آن را به وجود آورده به دست می آید.

هر متغیری که تقدم زمانی داشته باشد، علت و هر متغیری که بعد از آن بیاید و تقدم زمانی نداشته باشد، معلول است.

**ADJUSTED R SQURE یا مجذور r تعدیل شده:** اگر تعداد متغیرها زیاد باشد، تغییر می کند.

در تحلیل رگرسیونی از آزمون آنوا استفاده می شود برای اینکه بین متغیرهای مستقل با وابسته، رابطه معناداری وجود دارد یا خیر.

**جدول COEFFICIENTS ضرایب رگرسیونی:**

ضرایب اگر به شکل معمولی خودشان باشند با B نشان داده می شوند، اگر ضرایب به صورت استاندارد شده دربیایند با BETA نوشته می شوند.

در این قسمت از آزمون T برای تک تک متغیرها استفاده کرده تا سطح معناداری بدست بیاید در حالی که آزمون آنوا برای کل متغیرها محاسبه می شود.

**معادله رگرسیونی:**

نمره منزلت شغلی = 13/079 + 2/295 + (تحصیلات)

نمره منزلت شغلی = 13/079 + 2/295 + (12)

نمره منزلت شغلی = 13/079 + 2/295 + (10)

نمره منزلت شغلی = 13/079 + 2/295 + (16)

**METHOD** روش ورود متغیرهای مستقل به معادله چگونه باشد؟ کدامیک اول یا کدامیک دوم باشد.

ENTER یا ورود: همه متغیرهای مستقل وارد معادله شوند. می خواهد معنی دار باشد یا نباشد.

STEPWISE یا گام به گام: ابتدا آن متغیری که رابطه اش با متغیر وابسته بیشتر است وارد شود پس از آن متغیرهای دیگر وارد شوند. هر کدام معنی دار هستند وارد شوند و هر کدام که نیستند گام به گام خارج شوند.

REMOVE همه متغیرها را خارج می کند مگر متغیری که خودمان وارد می کنیم.

BACKWARD رو به عقب: کل متغیرها را وارد معادله می کند و هر کدام که معنی دار نیست را یک به یک خارج می کند.

FORWARD رو به جلو: به این معناست که هیچکدام را وارد نمی کند. اولی را بررسی می کند ببیند رابطه معنی دار دارد یا خیر و بعد دومی و سومی و غیره.

CONSTANT عدد ثابت

نمره منزلت شغلی = 477/2 (تحصیلات) + 0/155 (سن)

y = 2/477 (x1) + 0/155 (x2)

 تعداد فرزندان= 1/791 + 0/036 - (سن) 0/094 (تحصیلات)

زمانی که از پرسشنامه برای گردآوری اطلاعات در یک پژوهش پیمایشی استفاده می کنیم بید به اعتبار VALIDITY و پایایی RELIABILITY آن دقت کنیم.

اعتبار: سنجش ما درست و مطابق با واقع باشد.

پایایی: تکرار سنجش نتایج یکسان به ما بدهد.

هر دو برای یک تحقیق لازم است و اگر تحقیقی پایایی و اعتبار نداشته باشد از درجه اعتبار ساقط است. مانند اینکه آیا آمار 78 درصدی اعتماد مردم به رادیو و تلویزیون که آن را بدست آورده ایم مطابق با واقعیت است یا خیر؟ یا چند درصد مردم از ماهواره استفاده می کنند که عدد 53 درصد بدست می آید و باید مشخص شود که آیا اعتبار دارد یا خیر؟

در تحقیق پیمایشی، زمانی که می خواهیم پایایی تحقیق را بسنجیم راه های مختلفی وجود دارد:

1. آزمون – آزمون مجدد
2. دو نیمه سازی
3. فرمهای مشابه یا موازی
4. همبستگی درونی: معمولاً بهترین شیوه است و در آن از ضریب آلفای کرونباخ استفاده می کنیم. فرض کنید که 5 سئوال داریم که این 5 سئوال برای سنجش میزان اعتماد مردم به اخبار صدا و سیما طراحی شده است.

ضریب آلفا، کل ضریب های همبستگی های سئوالات را با هم جمع کرده و به ما می گوید که به طور کل این 5 سئوال با یکدیگر همبستگی درونی دارند یا خیر عدد بین صفر تا یک بدست می آید که هر چه قدر به صفر نزدیک تر باشد پایایی ندارد ولی اگر به یک نزدیک تر باشد پایایی دارد.

ANALYZE → SCALE → RELIABILITY ANALYSIS

کادر ITEMS مربوط به سئوالات و گویه هایی است که می خواهیم پایایی آنها را بسنجیم. در بسیاری از موارد کل سئوالات پایان نامه را انتخاب کرده و به ITEMS می برند و پایایی آن را می سنجند که این کار اشتباه است بلکه باید ببینیم کدام سئوالات پرسشنامه بر اساس سنجش یک متغیر واحد طراحی شده اند باید ضریب پایایی آنها محاسبه شود.

SCALE مقیاس

چند سئوال که یک متغیر واحد را می سنجد یک مقیاس است.

زمانی که می خواهیم ضریب آلفا را محاسبه کنیم باید سئوالات را به قسمت ITEM بیاوریم که یک مقیاس باشند. ضریب آلفایی خوب است که بالای 70 درصد باشد اگر زیر 70 درصد باشد مشکل دارد.

در این حالت باید به ترتیب نزولی و صعودی سئوال ها دقت کنیم زمانی که نتیجه به سمت منفی رفته است در این حالت لازم است RECODE کنیم یا کد گذاری برخی از سئوال ها برعکس است.

گویه ها:

1. گویه مثبت: اخبار صدا و سیما واقعیت ها را بیان می کند. کاملاً موافقم 1 موافقم 2 نظری ندارم 3 مخالفم 4 کاملاً مخالفم 5
2. گویه منفی: اخبار صدا و سیما سانسور می شود. کاملاً موافقم 1 موافقم 2 نظری ندارم 3 مخالفم 4 کاملاً مخالفم 5

در گویه دوم باید کدگذاری برعکس شود تا بتوانیم پایایی را بدست بیاوریم. گویه های مثبت و منفی را عوض کنیم.

SCALE IF ITEM DELETED مقیاس در صورت حذف گویه یا سئوال: در صورت حذف هر گویه یا سئوالات کل مقیاس چه تغییری خواهد کرد. هر سئوالی یا گویه ای که با حذف ضریب آلفا افزایش پیدا می کند یا باید RECODE شده یا حذف شود. در این حالت یک جدول سوم هم به ما ارائه می شود.

یک مقیاس حداقل باید دو گویه داشته باشد.

باید خود سئوالات هم به ما اجازه RECODE شدن هم بدهد.

منوی TRANSFORM یا تغییر شکل داده ها: این منو برای تحقیقات کمّی لازم و واجب است.

دستور COMPUTE یا محاسبه کردن:

این امکان را به ما می دهد که با استفاده از متغیرهای فعلی، متغیرهای جدید بسازیم. متغیرهای موجود را با هم ترکیب کرده و به متغیر جدیدی می رسیم. مانند اینکه ممکن است بخواهیم دو یا چند متغیر را با هم جمع بزنیم و یک متغیر جدیدی بسازیم که نمره مجموع آنها باشد یا ممکن است که بخواهیم میانگین 5 متغیر را بگیریم به صورت یک متغیر جدید دربیاوریم ممکن است که بخواهیم فرمول نویسی یا استاندارد سازی کنیم که همه اینها در این قسمت اتفاق می افتد.

ساده ترین کار در COMPUTE این است که یک متغیر را کپی کنیم.

TARGET VARIABLE متغیر هدف: متغیر جدیدی است که می خواهیم بسازیم.

NUMERIC EXPRESSION عبارت عددی: به عنوان مثال متغیر جنس را در قسمت TARGET می نویسیم و یک متغیر به نام SEX را کپی می کنیم و اسم آن عوض می شود. برای بدست آوردن میانگین اعضای پاسخ دهنده در COMPUTE اسم متغیر جدید را وارد می کنیم.

FUNCTION GROUP توابع ریاضی و آماری است که می توانیم از آنها استفاده کنیم.

STATISTICAL توابع آماری را به ما نشان می دهد.

در قسمت FUNCTIONS AND SPECIAL VARIABLES جزئیات توابع آماری و یا هر چیزی که انتخاب کرده ایم قرار می گیرد.

وقتی MEAN را انتخاب می کنیم میانگین هر کدام از متغیرهایی که انتخاب کرده ایم به ما می دهد.

\*\* علامت توان است.

در قسمت IF به صورت شرطی می توانیم COMPUTE را انجام دهیم.

 *اگر سن بزرگتر از 18 بود محاسبه را انجام می دهد و غیره.*

*جنس – نژاد (جن)*

*مردان سفید پوست 1*

*مردان سیاه پوست 2*

*مردان سرخ پوست 3*

*زنان سفید پوست 4*

*زنان سیاه پوست 5*

*زنان سرخ پوست 6*

*در ادامه متغیر را استاندارد سازی می کنیم. 89 – 18 استاندارد سازی 100 – 0*

*این کار در دو مرحله تغییر مبدا و مقیاس صورت می گیرد.*

*تغییر مبدا: از 18 کم می کنیم.*

*تغییر مقیاس: همه مقیاس ها را در 71 ضرب و بر 100 تقسیم می کنیم. 71 – 0*

***مباحث آموزش داده شده در کلاس درس* SPSS *دکتر هرمزی زاده:***

1. *ورود داده ها*
2. *تعریف متغیرها*
3. *دستور* FREQUENCIES
4. *دستور* RECODE
5. *دستور* DESCRIPTIVES
6. *دستور* EXPLORE *یا آزمون نرمال بودن*
7. *دستور* CROSSTABS
8. *دستور* MEANS *آزمون خطی بودن*
9. *دستور* ONE – SAMPLE T – TEST
10. *دستور* INDEPENDENT - SAMPLE T – TEST
11. *دستور* PAIRED SAMPLES T – TEST
12. *دستور* ONE WAY ANOVA
13. *دستور* GENERAL LINEAR MODEL
14. *دستور* BIVARIATE CORRELATE
15. *دستور* PARTIAL CORRELATE
16. *دستور* LINEAR REGRESSION
17. *دستور* COMPUTE
18. *دستور* SCALE

***تمرین های امتحانی:***

*1. با استفاده از جدول فراوانی مشخص کنید که منزلت شغلی چند درصد از افراد بالاتر از 50 است.*

*این سئوال از طریق دستور* FREQUENCY *حل می شود. در جدول به دست آمده رقم 50 را بدست می آوریم و درصد فراوانی را از 100 کم می کنیم.*

*2. شماره ردیف پاسخگویانی را که سیاه پوست هستند و نمره منزلت شغلی آنها بالاتر از 50 است را پیدا کنید و بنویسید.*

*در مواردی پیش می آید که می خواهیم افراد خاصی را در داده هایمان پیدا کنیم مانند مثال فوق، بر همین اساس در منوی* DATA *دستور* SELECT CASES *را فعال می کنیم. وقتی که دیالوگ باکس باز شد 5 انتخاب وجود دارد.*

ALL CASES *همه را انتخاب کن*

IF CONDITION IS SATISFIED *اگر یک شرطی را داشته باشد*

RANDOM SAMPLE OF CASES *یک نمونه تصادفی از همه افراد انتخاب کن*

BASED ON TIME OR CASE RANGE *بر اساس زمان یا شماره افراد انتخاب کن*

USE FILTER VARIABLE *از یک متغیر فیلتر استفاده کن*

*در گزینه دوم وقتی روی* IF CONDITION IS SATISFIED *کلیک می کنیم دیالوگ باکس جدیدی باز می شود که دستور را در آن می نویسیم. مانند* RACE = 2 & PRESTG80

راه دوم از طریق منوی ANALYZE صورت می گیرد.

ANALYZE → REPORT → CASE SUMMARIES

راه دوم بسیار راحت تر است و شماره ردیف افراد را می دهد بدون اینکه دنبال آنها بگردیم.

شماره ردیف افرادی را پیدا کنید که تعداد برادرها و خواهرهای آنها بیش از 10 نفر است ولی تعداد فرزندان آنها صفر است.

ابتدا از طریق منوی DATA → SELECT CASES و دستور IF CONDITION IS SATISFIED عمل کرده و دستور را وارد می کنیم و سپس از طریق منوی ANALYZE → REPORT → CASE SUMMARIES تعداد را بدست می آوریم.

3. این سئوال مربوط به دستور RECODE می شود و متغیرهای سن و غیره را به یک تعداد مشخصی گروه تقسیم کرده و یک طبقه بندی جدید انجام می دهیم. نمره منزلت شغلی از 17 تا 86 است و آن را به صورت یک متغیر ترتیبی با 4 طبقه به صورت عرض مساوی دربیاورید.

اولین کاری که می کنیم این است که از متغیر منزلت FREQUENCY می گیریم. نمره بدست آمده عددی بین 17 تا 86 است. سپس فاصله طبقاتی را محاسبه می کنیم.

17 – 34

35 – 52

53 – 70

71 – 86

سپس از دستور TRANSFORM → RECODE INTO DIFFERENT VARIABLES وارد شده و متغیر منزلت شغلی را وارد کرده و گزینه OLD AND NEW را فعال می کنیم سپس گزینه RANGE را فعال کرده و ردیف طبقاتی بدست آمده را وارد می کنیم و از گزینه NEW VALUE گزینه COPY OLD VALUE را فعال می کنیم تا ردیف های جدید وارد شده در قسمت OLD AND NEW وارد شود و سپس گزینه CONTINUE را فعال کرده و اعداد بدست آمده نمایش داده می شود.

4. صدک های بیست و سوم، چهل و هفتم و هفتاد و سوم سن را بنویسید و آنها را توصیف کنید.

از دستور FREQUENCY و گزینه STATISTIC قسمت PERCENTILES و انتخاب و به ترتیب صدک ها را وارد می کنیم.

سن 23 درصد از افراد زیر 31 سال است و 71 درصد بالای 31 سال است.

سن 47 درصد از افراد زیر 40 سال و بقیه بالای 40 سال است.

سن 73 درصد از افراد زیر 58 سال و بقیه بالای 58 سال است.

5. رابطه بین نژاد RACE و احساس خوشبختی GENERAL HAPPINES را بررسی کنید.

احساس خوشبختی: متغیر ترتیبی

نژاد: متغیر اسمی

این سئوال از طریق جدول دو بعدی و CROSSTAB حل می شود و روش حل آن از طریق متغیرها تشخیص داده می شود.

ابتدا باید تشخیص دهیم کدام متغیر مستقل و کدامیک وابسته است و هر کدام که تقدم داشته باشد متغیر مستقل است. در قسمت STATISTIC گزینه آزمون کای اسکوئر را فعال و نتایج بدست آمده را توصیف می کنیم.

مقدار کای اسکوئر بدست آمده 797/24 و درجه آزادی آن 4 و سطح معنی داری آن 000/0 است و با 95 درصد اطمینان نتایج قابل تعمیم است.

6. همه متغیرها را بر حسب میانگین آنها به صورت نزولی مرتب کنید.

با دستورSTATISTIC DESCRIPTIVE این کار انجام می شود و از گزینه OPTION گزینه DECENDING MEAN را انتخاب می کنیم.

7. با استفاده از آزمون و نمودار بررسی کنید که آیا متغیر تحصیلات پاسخگو، توزیع نرمال دارد یا خیر؟

این مسئله از طریق دستور EXPLORE حل می شود.

ANALYZE → DESCRIPTIV STATISTICS → EXPLORE

ابتدا متغیر تحصیلات را در DEPENDENT LIST قرار می دهیم و در قسمت PLOTS گزینه NORMALITY PLOTS WITH TEST را فعال می کنیم. فرضیه صفر ما این است که توزیع این متغیر، نرمال است که در حال حاضر رد می شود چرا که توزیع نرمال نیست و این پاسخ توسط آزمون های کولموگروف اسمیرنوف و شاپیرو ویک رد شده است.

8. شدت ارتباط بین دو متغیر میزان احساس خوشبختی و نگرش نسبت به زندگی را بررسی کنید.

ابتدا باید سطح سنجش این دو متغیر را از طریق دستور FREQUENCY بدست بیاوریم. هم با دستور CROSSTAB و هم CORELLATE این پاسخ داده می شود. البته در دستور CROSSTAB همه آزمون های تای بی کندال و تای سی کندال و غیره وجود ندارد.

ANALYZE → CROSSTAB → ROW: GENERAL HAPPINES ,COLUMN: IS LIFE EXCITING OR DULL

از گزینه های STATISTIC و ORDINAL آزمونها را انتخاب می کنیم. بر همین اساس در این سئوال شدت ارتباط 327/0 هزارم است.

9. منزلت شغلی افراد فعال و منفعل را با هم مقایسه کنید.

دو گروه مستقل و دو میانگین هستند که می خواهیم آنها را با هم مقایسه کنیم.

اول از دستور FREQUENCY متغیرها را می شناسیم سپس از گزینه ANALYZE به گزینه COMPARE MEANS رفته و آزمون INDEPENDENT SAMPLE T –TEST را انتخاب می کنیم. منزلت شغلی در قسمت TEST VARIABLE قرار می گیرد و GROUPING VARIABLE هم فعال بودن و منفعل بودن می شود.

در قسمت DEFINE GROUPS گروه های اول و دوم را وارد می کنیم تا آزمون انجام شود.