The Asia Foundation

احصائیه تطبیقی برای تحلیل پالیسی

چاپ دوم

محمد شعیب حیدری و فهیم احمد یوسفزی

USAID FROM THE AMERICAN PEOPLE

Australian Government
Department of Foreign Affairs and Trade

همکاری آلمان

The Asia Foundation

احصائیه تطبیقی برای تحلیل پالیسی

محمد شعیب حیدری و فهیم احمد یوسفزی


بسم الله الرحمن الرحیم

احصائیه تطبیقی برای تحلیل پالیسی

نویسنده گان

محمد شعیب حیدری و فهیم احمد یوسفزی

دیزاین

سید راشد سادات و روح الله محمدی

چاپ دوم: کابل، 1397

صاحب امتیاز : بنیاد آسیا، دفتر افغانستان

(The Asia Foundation)

فهرست عناوین

بخش چهارم- فرضیه چیست؟

بخش پنجم- آزمون همبستگی و رگرسیون

سپاسگذاری

در پایـــان میخواهیـــم از همکاران ما در بنیاد آســـیا هر یک داکتر تبســـم اکســـیر، داکـــتر زاک وارن، عبدالتواب جلیلی، ســـید مسعود ســـادات، محمد شریح شـــیوان، شـــمیم سرابی و مهدی فروغ، اشـــتراک کننده برنامه آموزشـــی تحلیل دیتا داکـــتر احمد نوید شـــمس و محصلین رشـــته احصائیه پوهنځـــی اقتصاد پوهنتون کابـــل حمیدالله قرلـــق و بی بی رقیه شـــایق که وقـــت گرانبهای خویش را برای مـــرور و ارایه نظریات و پیشـــنهادات در غنامندی و بهتر شـــدن این اثر صرف نموده انـــد ابراز ســـپاس و امتنان نماییم.

مقدمه

انگیزه اساسی برای تهیه این کتاب در نخست ادامه مواد درسی برای کورس تحلیل دیتا توسط برنامه ستاتا و ضمیمه برای درک بهتر مفاهیمی که در رهنمای برنامه ستاتا تحت عنوان "تحلیل دیتا توسط برنامه استاتا" است، می باشد.

یکی از اهداف عمده این اثر بلند بردن ظرفیت در نهاد های دولتی و دیگر نهاد های که عملا با موضوعات احصائیه و اقتصاد سنجی (Econometrics) سر و کار دارند میباشد بنابراین، در این اثر کوشش صورت گرفته تا مفاهیم احصائیوی و اقتصاد سنجی (Econometrics) و استفاده از آنها به شکل عملی آن بیان گردد. همچنان مثالهای این اثر به شکلی ترتیب شده است تا برای خواننده قابل درک باشند و با استفاده از مفاهیم احصائیوی به تحلیل و بررسی مشکلات خویش بپردازند. تفاوت دیگر این اثر استفاده از کارتون ها است که در هنگام مطالعه این اثر خواننده علاقمندی بیشتر برای خواندن پیدا میکند.

بخشی از اهداف این اثر آشنایی خواننده با نقش احصائیه در طرح و تدوین پالیسی ها می باشد. باید یاد آور شد که نتایج بدست آمده از تحقیقات و سروی ها در شکل دهی و ساختن پالیسی ها با اهمیت می باشد و مورد استفاده قرار می گیرد. هر چند در این اثر در رابطه بین احصائیه و پالیسی سازی صحبت نشده است اما در تالیفات بعدی به این موارد پرداخته خواهد شد.

این کتاب به پنج بخش تقسیم گردیده است. در بخش اول مفاهیم اساسی که در علم احصائیه روزانه مورد استفاده قرار می گیرد به معرفی گرفته است، بخش دوم معیارهای گرایش مرکزی, بخش سوم معیارهای پراگنده گی، بخش چهارم معرفی فرضیه و ساختن فرضیه و در بخش پنجم کتاب آزمون های احصائیوی مانند: همبستگی، رگرسیون خطی و رگرسیون پروبیت و آزمون میانگین دو جامعه (t -test) به تشریح گرفته شده است.

این اثر برای اولین بار در سال 1395 به چاپ رسیده است و چاپ دوم آن در سال 1397 صورت گرفته است. در هردو دوره، کتاب خالی از اشتباهات در نوشتار و تشریح مفاهیم نمیباشد و ما را از آن بابت معذور بدارید. برای غامندی هر چه بیشتر کتاب در آینده ما خواهان نظریات و پیشنهادات خواننده گان و مستفدین محترم میباشیم.

بخش اول

معرفی مفاهیم اساسی

معرفی

در دنیـــای جدیـــد که عصر تکنالـــوژی و معلومات بـــوده، هر روز بـــه اهمیت علـــم احصائیه که بخش جدا ناشـــدنی و اســـاس مهـــم برای پیشرفـــت تکنالوژی و ارایـــه معلومات بوده، افـــزوده میشـــود. احصائیه بنیاد و اســـاس مهم برای پیشرفـــت در عرصه هـــای مختلف چون (زراعـــت، اقتصاد، بیولوژی، طب، صنعـــت، پلانگذاری، تعلیم و تربیه) و ســـایر علوم میباشـــد. احصائیه تقریبـــاً در تمام عرصه هـــای زندگی بشری کاربرد وســـیع دارد.

احصائیه چیست (Statistics)؟

در علـــم احصائیـــه موضوعـــات مورد بحث شـــامل جمع آوری، ترتیـــب و تنظیم، خلاصه ســـازی، تحلیـــل و ارایه دیتا و همچنـــان نتیجـــه گیری و تصمیم گیری بر اســـاس یافته ها از دیتاهای بدســـت آمده میباشـــد. احصائیه بیشـــتر به جمع آوری معلومـــات کمـــی به گونـــه سیســـتماتیک و تحلیل و تفســـیر معلومات گردآوری شـــده میپردازد. کلمـــه احصائیه شـــامل مفاهیم مانند:

1. حقایق کمی ، مثلا تعداد افراد که در یک منطقه زندگی میکنند.
2. مطالعه شیوه های جمع آوری دیتا و تحلیل و تفسیر حقایق بدست آمده میباشد.

به صورت عموم احصائیه به دو بخش تقسیم میگردد:

1. احصائیه تشریحی
2. احصائیه استنباطی

احصائیه تشریحی یا عبارت از آن بخشی از احصائیه می باشد که تنظیم، تشریح و نمایش دیتا را در بر می گیرد. مسائل مانند توزیع فریکونسی، میانگین، مود، میدیان (میانه)، معیار های پراگندگی مانند دامنه ی تغیرات، انحراف معیاری و وریانس شامل احصائیه تشریحی اند.

احصائیه استنباطی یا تحلیلی عبارت از آن بخشی از احصائیه است که معلومات بدست آمده از نمونه یک جمعیت را تحلیل نموده و پیش بینی و نتیجه گیری لازم را در مورد ارائه می کند. موضوعات مانند تیوری نمونه گیری، آزمون فرضیه ها، رگرسیون (وابستگی) کورلیشن (همبستگی) مربوط احصائیه استنباطی یا تحلیلی می شوند.

جمعیت (Population)

در مباحث احصائیوی تمام واحدهای که در یک مجموعه شامل و حداقل دارای یک صفت مشترک میباشند بنام جمعیت و یا جامعه یاد میشوند. یا به عبارت دیگر تمام مشاهدات شامل در یک مطالعه عبارت از جمعیت میباشند. به گونه مثال تمام کارمندان یک وزارتخانه، تمام شاگردان یک مکتب و یا تمام کتاب های موجود در یک کتابخانه عبارت از مثال های از جمعیت میباشد.

نمونه (Sample)

در مباحث احصائیوی نمونه به مفهوم یک نسبت انتخاب شده کوچکتر از جمعیت میباشد. یا یک مجموعه کوچک از مشاهدات انتخاب شده از جمعیت میباشد. تعداد واحدها یا مشاهدات در یک نمونه عبارت از حجم نمونه میباشد. در تصویر پائین نمونه و جمعیت نمایش داده شده است. جمعیت درتصویر تمام شاگردان صنف می باشد و سه نفر شاگرد که در نزدیک تخته ایستاده اند عبارت از نمونه است که از جمعیت یعنی صنف نماینده گی میکند.

شکل: ۱-۱ نمونه و جمعیت

بـــه ادامه مثال گذشـــته اگر در یـــک وزارتخانه بـــه تعداد 2,000 کارمنـــد موجود باشـــد و از میان آنهـــا 200 نفر برای یـــک تحقیق انتخـــاب گردند، 2,000 نفـــر عبارت از جمعیـــت و 200 نفر انتخاب شـــده عبارت از نمونه تمـــام کارمندان وزارتخانه (جمعیت) میباشـــد.

هـــرگاه یک جمعیت بزرگ باشـــد و وقت کافی، منابع مالی و امکان بررســـی تمام افراد شـــامل در یـــک جمعیت ممکن نباشـــد در آن صورت از نمونه گیری اســـتفاده می شود.

چرا نمونه از یک جامعه گرفته میشود؟

- بخاطر کمبود وقت
- مصارف زیاد
- غیر ممکن و غیر معقول بودن مطالعه یک جامعه در کل

تعیین حجم یا اندازه نمونه (Sample Size Selection)

بعـــد از این کـــه جمعیت مورد نظرتـــان را تعیین نمودیـــد در قدم بعدی تصمیم این اســـت که حجم نمونـــه تان به چه انـــدازه باشـــد تا نماینده گـــی از جمعیت تان کـــرده بتواند. فرض کنید که شـــما یـــک تحقیق در مورد اطفال زیر ســـن 18 ســـال که مشـــغول به کار هســـتند انجام دهیـــد. با فرض این کـــه مجموع اطفال زیر ســـن کارگر در افغانســـتان به 500,000 تـــن میرســـد و بنابر محدودیت وقت و مصـــارف زیاد نمیتوانید تمام 500,000 کودک کارگر زیر ســـن 18 ســـال را سروی کنیـــد کـــه آیا از کار خود راضی هســـتند و یا خیـــر، بنابراین مجبور هســـتید با تعدادی کمـــتر اطفال سروی را انجـــام دهیـــد تا از تمـــام جمعیت نماینده گـــی کرده بتواند. اما ســـوالی کـــه در این جـــا پیش می آید، این اســـت که حجم نمونـــه باید چه مقدار انتخاب شـــود؟

اساســـاً حجم نمونه مربوط به این میشـــود که شـــما به چـــه اندازه میخواهید نتایج بدســـت آمده از نمونـــه مورد نظرتان بـــا جامعه مطابقت داشـــته باشـــد و یا به چـــه اندازه دیتـــای تان دقیق باشـــد تا نماینـــده گی از تمـــام جمعیت مورد نظر نمـــوده بتواند.

در اینجا اساساً باید به دو نکته ذیل توجه کرد. 1 حاشیه خطا (Margin of error) 2 سطح اطمینان Confidence level

1. **حاشیه خطا (Margin of Error)**

حاشیه خطا عبارت از خطای نمونه گیری میباشد به طور مثال شما یک تعداد معین از اطفال کارگر را سروی نموده اید و نتیجه بدست آمده شما این است که 40 فیصد اطفال از کار خود راضی هستند در صورت که حاشیه خطای شما 5% باشد پس شما گفته میتوانید که 35 الی 45 فیصد مجموع اطفال نسبت به کار خویش رضایت دارند.

به یاد داشته باشید به هر اندازه که بخواهید حاشیه خطا در سروی تان کم باشد پس به نمونه بزرگتر ضرورت دارید و هراندازه که نمونه کوچکتر انتخاب نموده باشد حاشیه خطا بزرگتر میباشد.

2. سطح اطمینان (Confidence Level)

سطح اطمینان معمولاً در حدود 90% , 95 % و 99% تعیین میشود. سطح اطمینان برای ما این را بازگو میکند که به کدام اندازه بالای نتیجه بدست آمده خود اطمینان داریم.

به تعقیب مثال قبلی اگر سطح اطمینان 95% باشد پس گفته میتوانیم که با 95 فیصد اطمینان در حدود 35% الی 45% اطفال کارگر زیر سن 18 سال نسبت به کار خود رضایت دارند.

بعد از تصمیم در مورد حاشیه خطا و سطح اطمینان شما میتوانید که اندازه نمونه خود را تعیین کنید, جدول ذیل بیانگر تعیین حجم نمونه به اساس جمعیت مورد نظر با درنظرداشت درجه های مختلف سطح اطمینان و حاشیه خطا میباشد.

	سطح اطمینان Confidence Interval = 95%			سطح اطمینان Confidence Interval = 99%		
	حاشیه خطا Margin of error (MoE)			حاشیه خطا Margin of error (MoE)		
(Population size) تعداد جمعیت	5%	2.5%	1%	5%	2.5%	1%
100	80	94	99	87	96	99
500	217	377	475	285	421	485
1,000	278	606	906	399	727	943
10,000	370	1,332	4,899	622	2,098	6,239
100,000	383	1,513	8,762	659	2,585	14,227
500,000	384	1,532	9,423	663	2,640	16,055
1,000,000	384	1,534	9,512	663	2,647	16,317

شکل: ۱-۲ تعیین حجم نمونه به اساس جمعیت، حاشیه خطا و سطح اطمینان

از جدول فوق چنین نتیجه میگیریم که برای یک جمعیت که در حدود 500,000 نفر را در بر داشته باشد باید که نمونه به اندازه 384 نفر انتخاب کنیم (با درنظرداشت سطح اطمینان 95% و حاشیه خطا 5%).

در صورت که جمعیت شما بیشتر از 1,000,000 باشد برای تعیین حجم نمونه به لینک ذیل مراجعه نمایید.

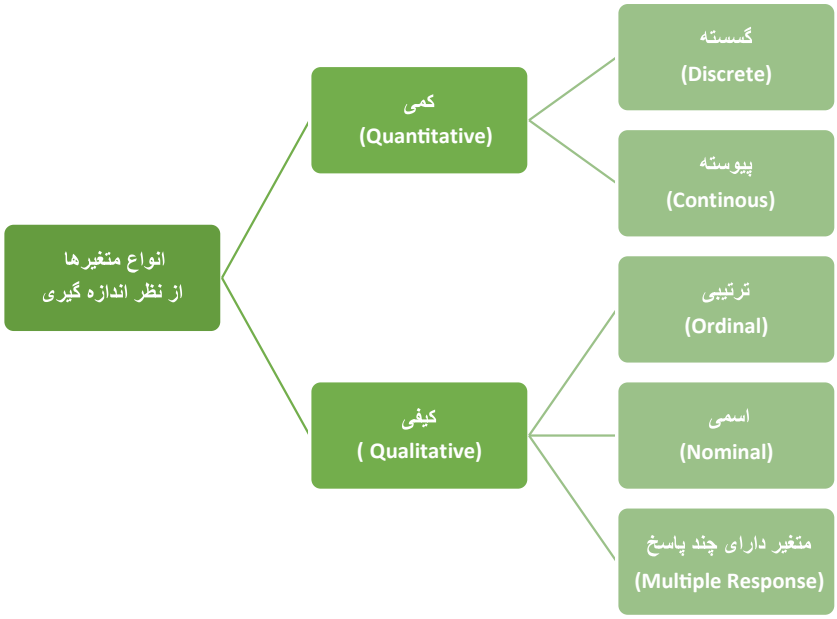
https://www.checkmarket.com/sample-size-calculator/

متغیر و انواع آن (Variable and Types of Variable)

متغیر چیست؟

یک مشخصه یا معلومات که در حال تغییر بوده و قیمت های متفاوت را به خود گرفته میتواند. اندازه قد افراد یک جامعه میتواند یک متغیر باشد اندازه قد هر فرد میتواند نظر به فرد دیگر متفاوت باشد. در شکل پائین اندازه قد افراد یک مثال خوب برای نشان دادن متغیر است به دلیل اینکه قد تمام افراد معمولاً یکسان نیست و نظر به هر فرد میتواند تغیر کند.

شکل: ۱-۳ افراد با اندازه قدهای متفاوت



1. **متغیـــر کیفـــی (Qualitative):** عبارت از متغیر اســـت که تغیـــر در آن در اندازه ی آن نبوده بلکـــه تغییر در نوع آن مـــی باشـــد و انجـــام دادن عملیه های ریاضی بـــا این متغیر هـــا معنی ندارد.

مانـــد: جنســـیت کـــه به مرد و زن تقســـیم شـــده و تغیر تنها در نوع جنســـیت اســـت نه ایـــن که مرد یـــا زن بودن جنســـیت از نگاه عـــددی دارای کدام مفهوم باشـــد.

2. **متغیـــر کمـــی (Quantitative):** عبـــارت از متغیر می باشـــد که تغیرات در انـــدازه آن صورت می گیـــرد و انجام دادن عملیـــه های حســـابی با ایـــن متغیرها ممکـــن و با مفهوم می باشـــد و یـــا متغیرهای که از اعداد تشـــکیل شـــده اســـت در مقابل متغیر کیفی مانند متغیر جنســـیت که مرد و زن می باشـــد از حروف تشـــکیل شـــده اســـت.

مانند: ســـن، عایـــد، تعداد کتـــاب که میتواند گفـــت 23 عدد کتاب کـــه عدد 23 دارای یک مفهوم اســـت بـــرای انجام دادن عملی های حســـابی

انواع متغیرهای کمی

- **متغیـر پیوسـته (Continuous):** نوعـی از متغیـر اسـت کـه ارزش هـای نـا محـدود را در بـر میگیـرد و بیـن دو واحـد آن هـر نقطـه یـا ارزشـی را میتـوان انتخـاب کـرد.
- به عنوان مثال، وزن یک متغیر پیوسته است
- **متغیر گسسته (Discrete):** متغیری است که ارزش های محدود را در بر میگیرد.

مانند: ماه های سال از 1 الی 12 است. و رنگهای رنگین کمان که از 1 الی 7 رنگ است.

انواع متغیرهای کیفی

- **متغیر ترتیبی (Ordinal):** در این نوع متغیر ترتیب پاسخ ها بسیار با اهمیت و دارای مفهوم می باشد. مانند: سطح تحصیل که از کم شروع به طرف سطح بلندتر ادامه میابد.

1 = دوره ابتداییه　　1 = بسیار ناراحت

2 = دوره متوسطه　　2 = ناراحت

3 = دوره لیسه　　3 = خوشحال

- **متغیر اسمی (Nominal):** ترتیب پاسخ ها در این نوع متغیر، دارای اهمیت و مفهوم نمی باشد، مانند: جنسیت، حالت مدنی

1= مرد　　1 = متاهل

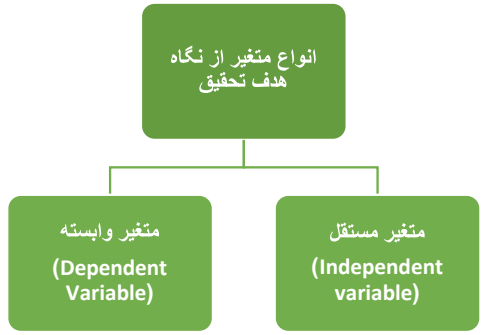
2= زن　　2 = مجرد

- **متغیر دارای چند پاسخ (Multiple Response Variable):** نوعی از متغیر است که دارای چند گزینه می باشد.به طور مثال: شما کدام وسایل ذیل را برای بدست آوردن اطلاعات و معلومات استفاده میکنید؟

1 = رادیو　　3 = شبکه های اجتماعی　　5 = روزنامه

2 = تلویزیون　　4= مسجد　　6 = شورای قریه

انواع متغیر از نگاه هدف تحقیق



- **متغیر وابسته (Dependent variable):** طوری که از نام آن پیداست، عبارت از متغیری است که قیمت ها یا نتیجه آن وابسته به قیمت های متغیر مستقل یا عوامل دیگر است.
- **متغیر مستقل (Independent variable):** این متغیر طوری که از نام آن پیداست به گونه مستقل قیمت گرفته میتواند و قیمت های آن وابسته به متغیر دیگر یا عوامل دیگر نیست.

بـــه طور مثال اگر مـــا بخواهیم یک نـــوع آزمایش را در مورد این که بالای ســـطح خوشـــحالی کدام عوامـــل تاثیر گذار میباشـــد انجام دهیم. در اینجا ســـطح خوشـــحالی عبـــارت از متغیر وابســـته بـــوده و مجموع عوامل دیگـــر که بالای ســـطح خوشـــحالی تاثیر گذار اند عبارت از متغیر های مســـتقل میباشـــند.

فریکونسی (Frequency)

در علـــم احصائیه فریکونســـی یک دیتا یا معلومات مشـــخص به معنی این اســـت کـــه این ارزش چند بار تکرار شـــده است.

بـــه گونه مثال، اگـــر در یکی از آمریـــت های وزارت صحت عامـــه 4 نفر از طبقه ذکـــور و 6 نفر از طبقه اناث باشـــد، در نتیجه فریکونســـی برای طبقـــه ذکور 4 و برای اناث 6 می باشـــد.

فریکونسی همیشه بوسیله حرف انگلیسی (f) نشان داده میشود.

جدول فریکونسی

یک جـــدول فریکونســـی عبارت از بیـــان معلومات احصائیوی به شـــکل ابتدایی اما ارزشـــمند میباشـــد. ســـاختار یک جدول فریکونســـی به گونه ی میباشـــد که گزینه های یک ســـوال به ترتیب نوشـــته شـــده و فریکونســـی هر کدام از گزینه ها شـــده ذکر میشـــود.

مثال:

شغل پاسخ دهنده گان سروی، 2015

وظیفه	فریکونسی	فیصدی
شاغل	3,946	41
متقاعد	79	0.8
خانم خانه	4,073	42.6
متعلم	627	6.6
بیکار	837	8.7
مجموعه	9,562	100

شکل: ۱-۴ جدول فریکونسی شغل پاسخ دهنده گان

تمرین

1. راجع به علم احصائیه معلومات دهید؟
2. جمعیت و نمونه چه است و برای هر کدام دو مثال بنوسید؟
3. چرا از نمونه استفاده میکنیم؟
4. با درنظر داشت 99% سطح اطمینان و 1% حاشیه خطا برای انجام تحقیق از یک جمعیت که تعداد شان به 100,000 خانم میرسد حد اقل به چه اندازه نمونه ضرورت است؟
5. انواع متغیر از نگاه اندازه گیری را نام بگیرید؟
6. برای متغیر اسمی و پیوسته مثال دهید؟
7. متغیر وابسته و مستقل را تعریف و برای هر کدام دو مثال دهید؟

بخش دوم

معیارهای گرایش مرکزی

(Measures of Central Tendency)

معرفی

برای مطالعه یک جمعیت و به خصوص یک مشخصه خاص آن جمعیت مانند سن، بررسی تمام افراد در یک جمعیت به گونه انفرادی مشکل به نظر میرسد که به یک نتیجه در مورد سن افراد برسیم، یعنی با مشاهده هر کدام از سن افراد نمیتوان گفت که آیا جمعیت مورد مطالعه نظر به سال گذشته چه تغییرات را در مشخصه سن افراد شاهد بوده است. اما بهتر خواهد بود تا یک سن را دریافته و به اساس آن در مورد جمعیت بتوان اظهار نظر کرد. یکی از مثال ها میتواند این باشد که اوسط سن فارغین پوهنتون طبی در سال 1385، 29 سال بوده اما در سال 1394 به 26 سال کاهش پیدا کرده است. در این مثال اوسط سن به نماینده گی از تمام سن محصلین به بحث گرفته میشود. زیرا یاد آوری سن هر یک از محصلین بی مفهوم بوده ما را به نتیجه نمی رساند.

یکی از معیارهایی که با دریافت آن میتوان به نماینده گی از جمعیت صحبت کرد، عبارت از معیارهای گرایش مرکزی میباشد.

تعریف

معیـــار گرایـــش مرکزی عبارت از معیار اســـت که مرکز دیتا را نشـــان داده و دیگـــر ارزش ها در یـــک مجموعه ازدیتاها در اطـــراف آن قـــرار دارند. میتـــوان برای فهم بهتر موضوع به دو فرد که در شـــکل پائین در مرکز ســـایر افراد ایســـتاده انـــد و موســـیقی می نوازند و ســـایر افـــراد در اطراف این دو نفـــر که اتن ملی را انجـــام میدهند از مثالـــی برای معیار های گرایش مرکزی یاد آور شـــد که شـــاخصهای هســـتند برای نشـــان دادن مرکز دیتا .

شکل: ۲-۱ معیارهای گرایش مرکزی

مثال: اوســـط یکـــی از معیارهـــای گرایش مرکزی بوده، فرضاً اوســـط ســـن کارمنـــدان وزارت مالیه در ســـال 1392، 22 ســـال بـــوده اما در ســـال 1393 این اوســـط به 25 ســـال افزایش پیدا کـــرده، این افزایش به این معنی اســـت که ســـن کارمنـــدان وزارت مالیـــه به صورت اوســـط نظر به ســـال 1392 در ســـال 1393 افزایش پیدا کرده اســـت.

در این اثر مود، میدیان و اوسط به معرفی گرفته میشوند.

مود (Mode): مود ساده ترین شاخص گرایش مرکزی است، مود عددی است که دارای بیشترین فریکونسی است، یا عددی که بیشتر از اعداد دیگر تکرار شده است. مود از طریق مشاهده ی عددی که بیشتر تکرار گردیده است تعیین می گردد. در شکل پائین گل ارغوانی مود است زیرا نسبت به دیگر گل ها بیشتر در تصویر تکرار شده است.

شکل: ۲-۲ نمایش مود در تصویر

مـــود در مطالعـــات بازاریابـــی حایز اهمیت میباشـــد، به عنوان مثـــال یک مدیر بازاریابـــی میتواند با اســـتفاده از مود دریابـــد که کـــدام رنگ لباس بیشـــترین تقاضا را در بازار دارا میباشـــد.

بـــه گونه مثال معلومات به دســـت آمـــده از یک دیتا ســـت وزارت معارف به کمـــک مود دریافته که متعلمین بیشـــتر علاقمند به خواندن کدام کتاب درســـی هســـتند.

کتاب های مورد علاقه شاگردان لیسه های شهر کابل، 1394

کتاب	فریکونسی (به نفر)
هندسه	8,555
دری	12,333
تاریخ	30,000
ریاضی	9,990

شکل: ۲-۳ جدول کتاب های مورد علاقه شاگردان برای مطالعه

از جدول داده شـــده در بالا این موضوع واضح میگردد که بیشـــترین شـــاگردان (30,000 شـــاگرد) در لیســـه های شـــهر کابـــل علاقه مند بـــه مطالعه کتاب تاریخ هســـتند.

مثـــال: در اعـــداد ذیل اندازه هـــای 8 جوره بوت میباشـــد که در یک روز از یک دوکان خریداری شـــده، مـــود عبارت از اندازه 44 میباشـــد زیرا نســـبت به دیگر اندازه ها زیادتر تکرار شـــده اســـت.

44, 45, 44, 41, 44, 43, 42, 40

مثـــال: در اعـــداد ذیل مود عبارت از اعداد 3 و 6 میباشـــد زیرا نســـبت به دیگر اعـــداد زیاد تر تکـــرار گردیده و دارای فریکونســـی مساوی اند.

2 6 7 6 1 3 6 3 7 5 3 1

همچنـــان در صـــورت که دیتا ما طبقه بندی شـــده باشـــد. هرطبقه و گـــروپ که بزرگ ترین فریکونســـی یـــا تکرار را دارا باشـــد، هـــمان طبقه یا گروپ، مود اســـت.

میانـــه (Median): میدیـــان نقطـــه ی وســـط یک سلســـله از ارزش ها یا اعداد میباشـــد، که سلســـله اعـــداد را به دو قســـمت تقســـیم کرده، 50% اعداد بزرگـــتر از میدیان و 50% دیگر کوچکـــتر از میدیان و خود میدیان در وســـط اعداد قـــرار میگیرد. بنابراین میدیان نقطه وســـطی اســـت یعنی عددی اســـت که سلســـله اعداد را به دو قســـمت مســـاوی تقســـیم می کند.

بـــه عبارت دگر اگر یک سلســـله ی اعـــداد را به ترتیب لســـت کنیم، میانه (میدیان) همان عددی اســـت که در وســـط این سلســـله اعداد قـــرار می گیرد.

میدیـــان زمانـــی منحیـــث معیار گرایـــش مرکزی مورد اســـتفاده قرار میگیـــرد که سلســـله اعداد دارای یـــک عدد فوق العـــاده کوچـــک یا بزرگ باشـــد. در مثـــال پائین یک درخت که در وســـط ســـایر درختها قـــرار دارد دیده میشـــود که ارتفاع آن نســـبت به ســـایر درختان بســـیار بلنـــد می باشـــددر این حالت بایـــد برای دریافـــت مرکز دیتـــا از میدیان استفاده شود.

شکل: ۲-۴ مورد استفاده میدیان

در اعداد ذیل میدیان عبارت از عدد (4.5) میباشد

8 6 3 1 2 4 5 7

مرحله اول : ترتیب اعداد به شکل نزولی یا صعودی.

1 2 3 4 5 6 7 8

مرحلـــه دوم : چون مجموع مشـــاهدات ما جفت اســـت بنـــاً حاصل جمع دو عدد که در وســـط قرار دارد را تقســـیم بر 2 میکنیـــم که عدد بدســـت آمده عبارت از میدیان میباشـــد. 4.5= 2/(4+5)

نـــوت : اگر مجموع مشـــاهدات ما طاق باشـــد عدد که در وســـط مشـــاهدات ما (بعـــد از ترتیب نمودن اعـــداد) قرار دارد عبارت از میدیان میباشـــد.

مثال: در سلسله اعداد ذیل میدیان را دریابید؟

8 6 11 18 10 22 12

مرحلـــه اول اعـــداد را ترتیـــب میکنیم از کم بـــه زیاد (و یا از زیـــاد به کم) و عددی که در وســـط سلســـله اعداد قرار

گیـــرد عبارت از میدیان می باشـــد.

به طور مثال در مثال قبلی میدیان عبارت از عدد 11 میباشد.

~~22 18 12~~ 11 ~~10 8 6~~

نکتـــه: میدیان زمانی مورد اســـتفاده قرار میگیرد که سلســـله اعـــداد دارای یک عدد فـــوق العاده کوچک یا بزرگ باشد.

مثـــال: در پائیـــن مقدار تیل کـــه ۱۰ موتـــر در وزارت مخابرات در یک ماه مصرف می رســـانند داده شـــده و به صورت مشـــخص موتـــر با شـــماره سریـــال ۲۳۳ نظر به ســـایر موترها به مقـــدار قابل ملاحظه ای تیل بیشـــتر مـــصرف میکند که در ایـــن حالت بهتر اســـت از میدیان بـــرای دریافت مرکز دیتا اســـتفاده گردد.

مصرف تیل وسایط نقلیه در وزارت ترانسپورت – 31 حمل 1393

شماره موتر	مقدار تیل مصرف شده (لیتر)
225	330
226	455
227	322
228	544
229	444
230	456
231	654
232	233
233	2,300
234	267

شکل: ۲-۵ جدول مصرف تیل موترها

اوســـط (Mean): مشـــهورترین شـــاخص گرایش مرکزی اوسط اســـت. اوســـط از طریق جمع کردن تمام اعداد و تقسیم حاصل جمع آن بر تعداد تمام مشـــاهدات به دســـت می آید. در شـــکل پائین ســـه مـــرد با اندازه ریـــش های متفاوت رســـامی شـــده و شـــخصی که در میان دو مرد دیگر ایســـتاده بلندی ریش وی نسبـــت به دو فرد دیگر در وســـط قرار دارد. و مثالی از اوســـط میباشد.

شکل: ۲-۶ نمایش اوسط توسط ریش های افراد با اندازه های متفاوت

در اعداد ذیل اوسط عبارت از عدد (3.3) میباشد.

2 7 5 4 1 2 3 5 3 1

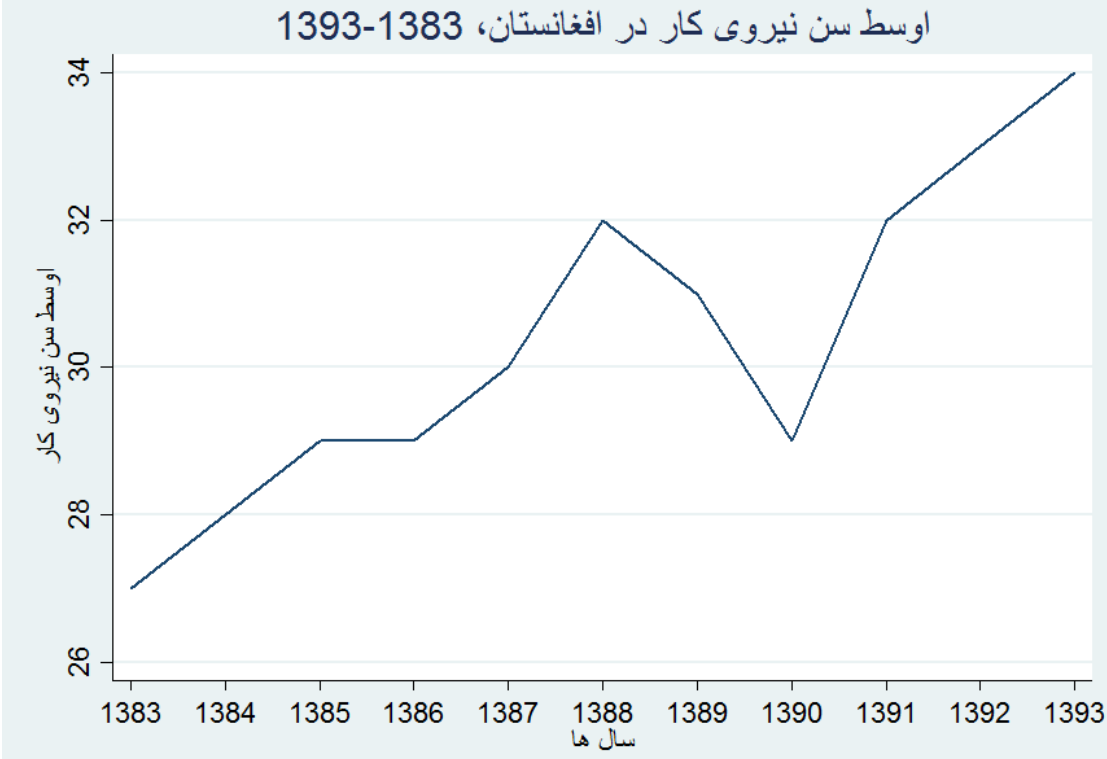
مرحله اول : تمام نمرات را جمع میکنیم .

33 = 2 + 7 + 5 + 4 + 1 + 2 + 3 5+ + 3 + 1

مرحله دوم : حاصل جمع تمام نمرات را تقسیم تعداد مشاهدات میکنیم

3.3 = 10 / 33

مثال: از اوسط میتوان استفاده های مختلف کرد، مثال ذیل اوسط سن نیروی کار در افغانستان را در جریان 10 سال نشان میدهد که اوسط سن نیروی کاری از 27 سال در سال 1383 به 34 سال در سال 1393 رسیده است.



شکل: ۲-۷ گراف اوسط سن نیروی کار در ۱۰ سال

مثال: مصارف دو ریاست وزارت تجارت (پلان و پالیسی و انکشاف سکتور خصوصی) برای 12 ماه سال در دسترس است، توسط اوسط میتوان دریافت که به صورت اوسط مصارف سالانه کدام ریاست بیشتر است:

ماه ها	پلان و پالیسی (افغانی)	انکشاف سکتور خصوصی (افغانی)
1	15,000	12,000
2	12,300	13,200
3	13,300	14,500
4	11,000	11,000
5	10,000	12,020
6	12,300	14,000
7	11,000	15,000
8	14,000	10,000
9	11,500	8,500
10	10,000	12,300
11	15,000	16,000

12	13,000	15,500
اوسط	12,835	12,367

شکل: ۲-۸ جدول مصارف دو ریاست وزارت تجارت

حالا نظر به جدول در بالا بعد از 12 ماه معلوم گردیده که مصارف ماهانه ریاست پلان و پالیسی به صورت اوسط نسبت به انکشاف سکتور خصوصی بیشتر است.

به عنوان مثال در سلسله اعداد پائین عواید ماهانه 5 آمریت ریاست زراعت ولایت لوگر داده شده در این میان عواید یکی از آمریت ها با 4 آمریت دیگر تفاوت زیادی دارد و در این خواهد بود تا از میدیان استفادگردد.

نام آمریت	عواید ماهانه
اداری	1,000
تحقیق	1,200
تهیه و تدارکات	1,400
کنترول و ارزیابی	800
مالی	170,000

شکل: ۲-۹ جدول عواید ماهانه ۵ آمریت ریاست زراعت ولایت لوگر

مثلا اگر بخواهیم دریابیم که به طور اوسط ریاست زراعت ولایت لوگر به چه اندازه عواید دارد به این منظور از اوسط (mean) استفاده می کنیم. در این صورت عواید تمام آمریت ها را با هم جمع میکنیم و تقسیم بر تعداد آمریت ها میکنیم که 34,880 میشود.

با وجود این که نتیجه گیری ما اشتباه نمیباشد، اما میتواند گمراه کننده باشد و ما را به راه اشتباه راهنمایی کند و موجب اتخاذ تصامیم اشتباه گردد زیرا 34,880 نسبت به عواید 4 آمریت (اداری، تحقیق , تهیه و تدارکات و کنترول و ارزیابی) بلند میباشد و نسبت به عواید آمریت مالی به مراتب کمتر میباشد.

پس در این جا بهتر این خواهد بود که از میدیان استفاده کنیم. میدیان بعد از محاسبه 1,400 بدست می آید.

تمرین

1. معیارهای گرایش مرکزی چیست؟
2. مود چیست و یک مثال دهید؟
3. در کدام موقع برای پیدا نمودن مرکز دیتا باید از میدیان استفاده کنیم؟
4. میدیان را توضیح دهید و در سلسله اعداد ذیل میدیان را دریابید؟

12, 23, 13, 18, 25, 15

5. عواید پنج شرکت هوایی به اساس معلومات وزارت مالیه در سال 1393 قرار جدول ذیل می باشد، برای دریافت مرکز دیتا باید از کدام یک از معیارهای گرایش مرکزی استفاده کرد؟

نام شرکت هوایی	عواید به میلیون افغانی
امارات	330
کام ایر	0.5
صافی	245
آریانا	340
افق شرق	430

شکل: ۲-۱۰ عواید ۵ شرکت هوایی در افغانستان

بخش سوم

معیارهای پراگندگی

(Measures of Dispersion)

معرفی

با مطالعه معیارهای گرایش مرکزی میتوان مرکز یک سلسله اعداد یا یک توزیع را دریافت، اما معیارهای گرایش مرکزی توانایی این را ندارند تا در مورد اینکه به کدام اندازه اعداد در دو طرف مرکز اعداد پراگنده هستند معلومات دهد. در یک سلسله اعداد یا یک توزیع تمام اعداد با هم مساوی نیستند. به معنی اینکه شاید بعضی اعداد از مرکز بسیار تفاوت داشته و یا بعضی کمتر تفاوت از مرکز را نشان دهند. این تفاوت بنام انحراف یاد شده که توسط معیارهای پراگندگی نشان داده میشود. انحراف کمتر به معنی این است که در یک سلسله اعداد تفاوت کمتر هر یک از اعداد از اوسط وجود دارد. از سوی دیگر انحراف بیشتر به معنی این است در یک سلسله اعداد تفاوت بین هر یک اعداد (مشاهدات) از اوسط زیاد میباشد و کمتر مشابه هم هستند.

مثال ذیل را در نظر داشته باشید:

مقایسه نمرات دو شاگرد، لیسه حبیبیه 1394

مضامین	نمرات احمد	نمرات فهیم
انگلیسی	85	68
تاریخ	90	75
دری	80	65
جغرافیه	25	67
ریاضی	65	70
مجموعه	345	345
اوسط	69	69

شکل: ۳-۱ جدول مقایسه نمرات دو شاگرد لیسه حبیبیه

در جدول بالا نمرات دو تن از شاگردان لیسه حبیبیه در 5 مضمون مختلف داده شده است، که مجموعه هر دوی آنها 345 و در نتیجه اوسط نمرات آنها مساوی به 69 می باشد. حقیقت این است که احمد در یکی از مضامین نمره 25 را گرفته ناکام مانده حال آنکه هر دوی آنها دارای مجموع نمرات و اوسط مساوی میباشند. از جدول بالا

میتـــوان چنیـــن نتیجه گرفت که نمرات فهیم یکـــی از دیگری کمتر انحـــراف دارد و برعکس نمرات احمـــد دارای انحراف بیشـــتر اســـت و انحراف کمتر مطلوب است.

معیارهـــای که در اینجا ما به منظور دانســـتن و آشـــنایی با انحـــراف مورد مطالعه قـــرار میدهیم عبارت انـــد از: دامنه تغییرات (**Range**)، انحراف (**Variance**) و انحـــراف معیاری (**Standard Deviation**).

دامنـــه تغییـــرات (Range): یکی از ابتدایی تریـــن معیارهای پراگندگی عبـــارت از دامنه تغییرات می باشـــد که تفاوت میان کوچکترین و بزرگترین عدد را در یک سلســـله اعداد نشـــان میدهد. دامنه تغییرات در شـــکل که در پائین رســـامی شـــده عبـــارت از تفاوت بیـــن ارتفاع قد بلندتریـــن پسر که کلاه زرد به سر داشـــته و قـــد کوتاه ترین پـــسر که پیراهن سرخ به تن دارد دریافت شـــده میتواند.

شکل: ۳-۲ دامنه تغییرات از کوتاه ترین تا بلندترین قد

فورمول دامنه تغیرات عبارت از:

دامنه تغیرات = بزرگترین عدد - کوچکترین عدد

Range=L-S

L=largest value

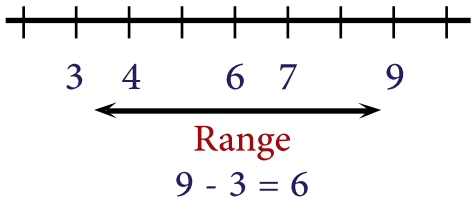
S=smallest value

مثال: در سلسله اعداد ذیل دامنه تغییرات را دریابید؟

3 ,4 ,6 ,7 ,9

Range=L-S

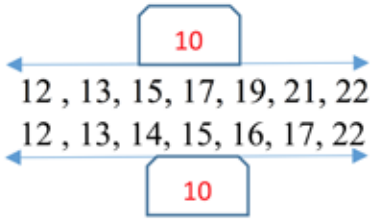
Range=9 - 3=6



بـــا وجود کـــه دامنه تغیـــرات یکی از معیارات پراگندگی اســـت امـــا منحیث یکـــی از بهترین معیار هـــای پراگندگی به شـــمار نمـــی رود. زیرا دامنـــه تغیرات تنها با کوچکتریـــن عدد و بزرگترین عدد یک سلســـله اعـــداد سرو کار دارد. به دو مثـــال ذیل توجـــه کنید با وجود که پراگندگی در ســـلسله اعداد اولی نســـبت به سلســـله اعداد دومی کمتر اســـت اما باز هم هـــر دو دیتا ســـت دارای دامنه تغیرات یکســـان اند.

A) 12 , 13, 15, 17, 19, 21, 22

B) 12 , 13, 14, 15, 16, 17, 22



انحراف معیاری (Standard Deviation)

هر یک انحراف معیار جمعیت و انحراف معیار نمونه دارای فرمول های جداگانه می باشد.

انحراف معیار جمعیت:

$$\sigma = \sqrt{\frac{1}{N}\sum(xi - \mu)^2}$$

انحراف معیار نمونه :

$$s = \sqrt{\frac{1}{N-1}\sum(xi - \bar{x})^2}$$

انحراف معیاری عبارت از یک معیاری برای اندازه گیری پراگندگی اعداد از اوسط آنها است.

سمبول که انحراف معیاری به وسیله آن نشان داده میشود:

الف: σ (سیگما حرف لاتین) میباشد برای نشان دادن انحراف معیاری یک جمعیت استفاده میشود.

ب: s، برای نشان دادن انحراف معیاری یک نمونه مورد استفاده قرار می گیرد.

فورمول برای دریافت انحراف معیاری بسیار ساده است؛ و عبارت از جذر دوم انحراف می باشد. شاید حالا برای شما این سوال مطرح گردد که انحراف چیست؟ به همین خاطر در مرحله اول انحراف را به مطالعه میگیریم.

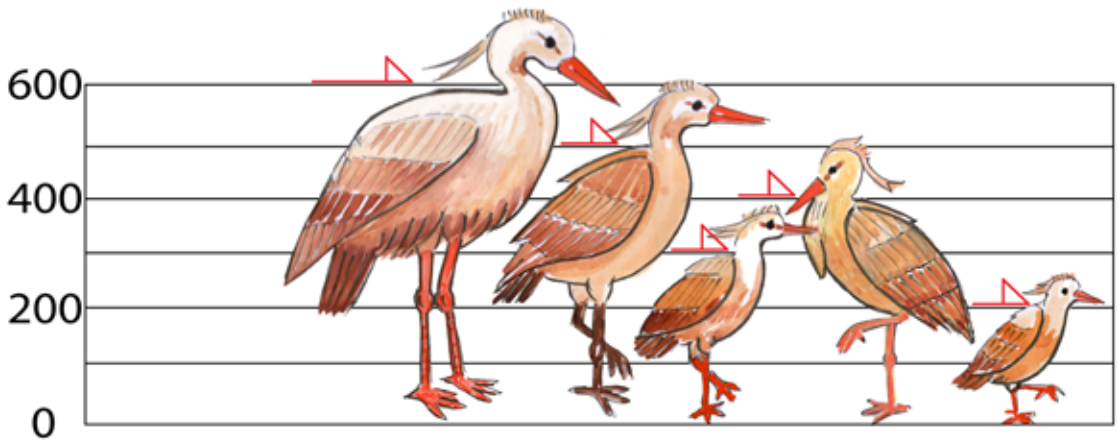
انحراف (Variance)

عبارت از اوسط مربعات تفاوت ها از اوسط یک سلسله اعداد یا یک توزیع می باشد. یا به عبارت دگر انحراف یا وریانس عبارت از مجموع مربعات تفاوت های یک سلسله ی اعداد از اوسط آن آعداد، تقسیم بر تعداد آن سلسله عددی است.

نگران نباشد شاید تعاریف بالا کمی پیچیده به نظر برسد اما برای فهم ساده تر انحراف به مراحل ذیل توجه کنید .

- در مرحله اول اوسط یک سلسله اعداد را دریابید.
- در مرحله بعد از هر عدد، اوسط بدست آمده را تفریق نموده و نتیجه آن را مربع نمایید (مربعات تفاوت ها از اوسط).
- مجموع آنها (مربعات تفاوت از اوسط) را تقسیم بر تعداد آنها میکنیم.
- اما به یاد داشته باشید در صورت پیدا نمودن انحراف نمونه باید که مجموع آنها را تقسیم (n-1) کنید.

مثال: برای فهم بیشتر به تصویر ذیل که قد (به میلی متر) چند حیوان (مرغابی ها) داده شده توجه کنید.



قد هریک (تا شانه): 600 میلی متر، 500 میلی متر، 300 میلی متر، 400 میلی متر و 200 میلی متر میباشد.

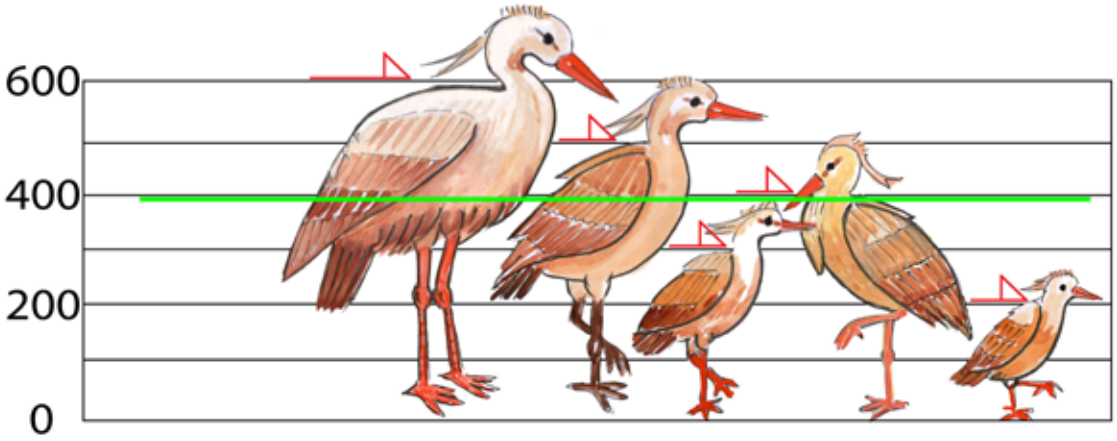
حالا اوسط، انحراف و انحراف معیاری قدهای داده شده در بالا را دریابید.

در مرحله اول اوسط را پیدا میکنیم.

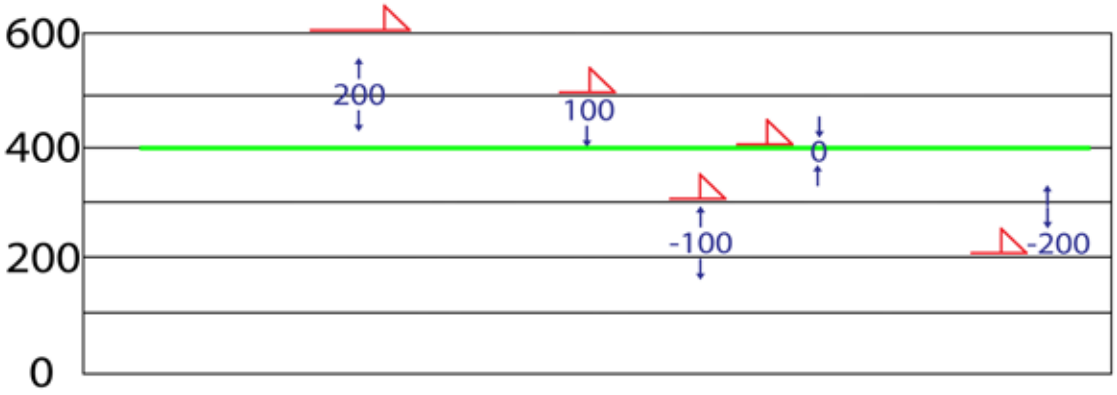
پاسخ:

$$\text{اوسط} = Mean = \frac{600+500+300+400+200}{5} = \frac{2000}{5} = 400$$

دیدیم که اوسط قدها عبارت از 400 میلی متر می باشد، حالا اوسط را توسط خط سبز نشانی میکنیم.



در ایـــن مرحله تفاوت هریـــک از قدها را از اوســـط پیدا میکنیم.



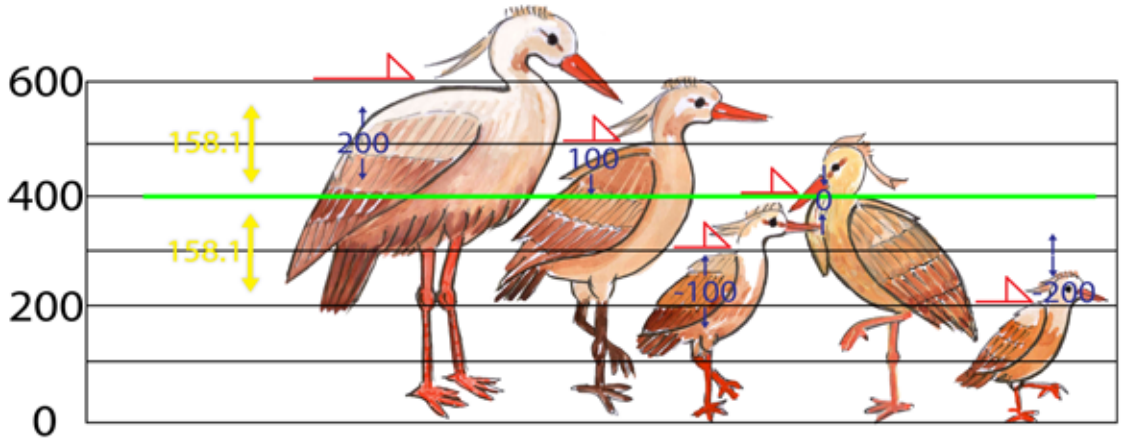
نکته: انحراف به سمبول S^2 نشان داده می باشد.

$$\text{انحراف} = Variance = \frac{200^2 + 100^2 + (-100)^2 + 0^2 + (-200)^2}{5-1} = \frac{100,000}{4} = 25,000$$

در نتیجه انحراف مساوی به 25,000 می باشد.

و انحراف معیاری عبارت از جذر مربع انحراف می باشد.

انحراف معیاری: $sd = \sqrt{25,000} = 158,1$



در نتیجـــه با اســـتفاده از انحراف معیاری، گفتـــه میتوانیم که کدام یک از مشـــاهدات (قدها) نورمال، کدام یک بســـیار بلند و کدام آنها بســـیار کوتاه می باشـــد.

مثال: یک نمونه دارای 5 مشاهده، که معلومات در مورد معاشات کارمندان دو وزارت خانه دولتی می باشد.

معاشات پنج نفر کارمند در دو وزرات خانه، حمل 1395

نام کارمندان	تجارت و صنایع (افغانی)	مخابرات و تکنالوژی(افغانی)
احمد	100,000	2,200
رشید	500	300
فرید	2,000	450
رحیم	40,000	1,000
حامد	3,200	2,500
انحراف معیاری (Std. Dev.)	42,921	1,008

شکل: ۳-۳ جدول معاشات ۵ کارمند وزارت تجارت و صنایع و مخابرات وتکناوژی

حالا گفته میتوانیم که معاشات کارمندان وزارت تجارت و صنایع نسبت به معاشات کارمندان وزارت مخابرات و تکنالوژی پراگنده تر می باشند. یا به عبارت دیگر معاشات کارمندان وزارت تجارت نسبت به وزارت مخابرات و تکنالوژی بیشتر پراگنده است.

تمرین

1. ضرورت دانستن معیار های پراگندگی را توضیح داده و انواع آن را نام بگیرید؟
2. چرا دامنه تغیرات نمیتواند یک معیار خوب برای نشان دادن پراگنده گی یک سلسله ی اعداد باشد؟
3. جدول ذیل نشان دهنده سن 11 نفر از مرد ها و زنان ولایت کابل میباشد, با استفاده از انحراف معیاری دریابید که پراگندگی در میان کدام گروه بیشتر میباشد؟

سن مرد ها	سن خانم ها
18	18
24	15
26	28
22	79
19	33
45	17
35	40
32	30
33	37
40	22
60	60

شکل: ۳-۴ جدول سن ۱۱ نفر مرد و زن ولایت کابل

4. معاش مامورین دیپارتمنت های تدارکات و دیپارتمنت تحقیق وزارت اقتصاد که هر یک دارای 11 نفر کارمند اند به ترتیب ذیل میباشد، با استفاده از انحراف معیاری دریابید که پراگندگی در معاشات کدام دیپارتمنت زیاد است؟

معاشات کارمندان دیپارتمنت های تدارکات و تحقیق وزارت اقتصاد، در ماه حمل 1395

شماره	نام کارمندان	تدارکات (افغانی)	تحقیق(افغانی)
1	احمد	12,000	20,000
2	فرهاد	15,000	20,000
3	نصرت	10,000	21,000
4	صادق	10,000	19,000
5	احمد شفیق	9,000	19,500
6	محد زمان	13,000	21,000
7	محمد امان	17,000	20,500
8	احمد فواد	12,000	19,300
9	احمد جمشید	12,000	21,000
10	احمد رشاد	15,000	20,000
11	محمد مسعود	9,000	21,500

شکل: ۳-۵ جدول معاشات ۱۱ نفر کارمند دو ریاست وزارت اقتصاد

بخش چهارم

فرضیه (Hypothesis) چیست؟

تصـــور کنید که شـــما فردا یک امتحان دارید. شـــما شـــب قبل تا ناوقـــت در بیـــرون از منزل هســـتید و تلویزیون می بینید. شـــما میدانید کـــه اگر روز قبل از امتحـــان درس بخوانید در نتیجه نمـــره بهتر را در امتحان فـــردا خواهید گرفت. بـــه نظر شـــما نتیجه امتحان فردای شـــما چگونـــه خواهد بود با فرض اینکه شـــب قبـــل از امتحان تا ناوقت شـــب در بیـــرون از منزل بودید؟

زمانیکه شـــما به این ســـوال پاســـخ میدهیـــد در واقع شـــما یک فرضیه را شـــکل میدهید. یک فرضیه عبارت از پیش بینی قابل اندازه گیری و مشـــخص می باشـــد. و فرضیـــه عبـــارت از ارایه علمی پیش بینی شـــما می باشـــد، توســـط یک فرضیه شـــما پیش بینی میکنید که چه اتفـــاق خواهد افتاد.

در حالـــت مشـــخص فرضیه، شـــما میتوانید به شـــکل ذیل باشـــید: " اگر من امشـــب درس نخوانم، پس نمره پائین تـــر را در امتحان فـــردا خواهم گرفت".

در شـــکل ذیـــل پائیـــن مثال خوبـــی بـــرای فهم بهتر فرضیه داده شـــده است دو فرد هســـتند که هر کدام آنهـــا با دیـــدن دود از فاصله دورتـــر فرضیات متفاوت دارنـــد. یکـــی از آنها فکـــر میکند کـــه در قریه آتش ســـوزی شـــده و فرد دیگر به این باور اســـت که گویا دیگدان برای آماده ســـاختن غذا روشـــن شـــده است.

شکل: ۴-۱ نمایش فرضیه

هدف یک فرضیه

معمـــولا دانشـــمندان در یک بخش مشـــخص علوم، بعد از مشـــاهدات و اندیشـــیدن به نتایج میرســـند که تـــا قبل از انجـــام آزمایـــش ها و ثابت شـــدن یا نشـــدن آن یـــک فرضیه اســـت. فرضیه، حدس و گمان اســـت که هنـــوز به اثبات نرســـیده و با اســـتفاده از احصائیه و تســـت هـــای احصائیوی میتواند یـــک فرضیه به اثبات برســـد یا اینکـــه به اثبات نرسد.

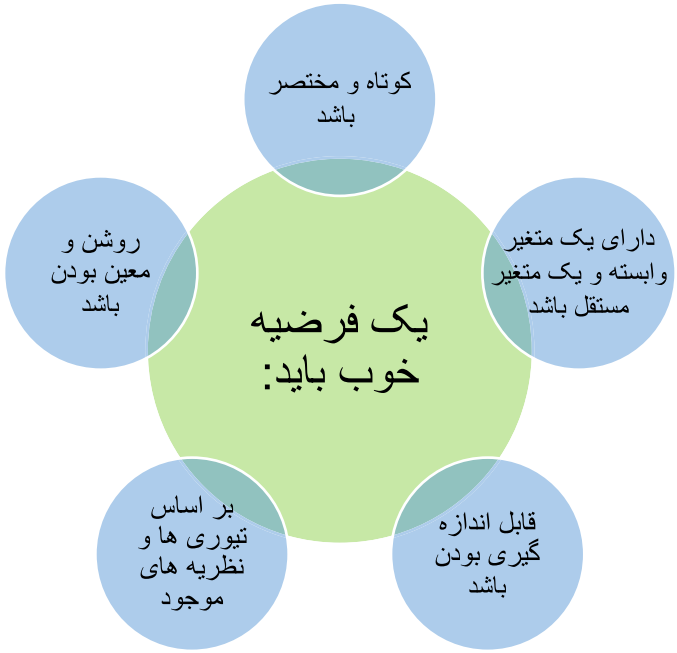
هـــدف از اســـتفاده فرضیـــه در یک تحقیق این اســـت که رابطـــه بین دو پدیـــده یا به زبـــان تحقیق رابطـــه بین دو متغیـــر، یـــک متغیر وابســـته و دیگری متغیر مســـتقل را بررســـی و مطالعه کند. یـــک فرضیه به ما کمـــک می کند تا اینکـــه چـــه نتیجـــه را در یک تحقیق بایـــد به بررســـی بگیریم و تمام تمرکـــز تحقیق بالای آن می باشـــد.

تغیرات در متغیر وابســـته، بســـتگی به تغیرات در متغیر مســـتقل دارد. یعنی متغیر وابســـته در پی تغییرات در متغییر آزاد از خـــود تغییـــر نشـــان می دهـــد. لازم به یاد آوری اســـت که متغیر وابســـته را بنـــام متغیر نتیجـــه (Outcome Variable) نیـــز یـــاد میکننـــد. و در مثال بـــالا نتیجه امتحان عبارت از متغیر وابســـته اســـت زیرا نتیجه آن وابســـته به متغیر مســـتقل اســـت. در مثال بالا متغیر مســـتقل عبـــارت از درس خواندن در شـــب قبل از امتحان می باشـــد.

بـــاز هم میتوان از مثال بالا اســـتفاده کـــرد و به این نکته اشـــاره کرد که یک فرضیـــه باید قابل آزمایش باشـــد، چونکه نمره امتحان شـــما قابل ســـنجش اســـت. شـــما میتوانید نمرات امتحان تـــان را که بـــرای آن درس نخوانـــده اید و نمرات امتحان را که شـــب قبل بـــرای آن درس خوانده اید مقایســـه کنید.

ویژگی های یک فرضیه خوب (Characteristics of a Good Hypothesis)

- دارای یک متغیر وابسته و یک متغیر مستقل باشد.
- قابل اندازه گیری بودن باشد.
- روشن و معین بودن باشد.
- بیان بر اساس تیوری ها و نظریه های موجود.
- کوتاه و مختصر باشد.



انواع فرضیه: (Types of Hypothesis)

1. فرضیه تحقیق یا متناوب(H_A) یا (H_1) - Alternative Hypothesis
2. فرضیه صفر (H_0) - Null Hypothesis

منظـــور از فرضیـــه تحقیق، فرضیه اي اســـت کـــه در آن به بیان تفـــاوت بین متغیرهـــا (رابطه) پرداخته مي شـــود، ولي فرضیـــه صفر بیان گـــر عدم وجـــود رابطه بین متغیرها اســـت.

مثال:

فرضیه تحقیق: افزایش فساد اداری باعث این میشود که رشد اقتصادی در یک اقصاد کاهش پیدا کند.

فرضیه صفر: میان سطح فساد اداری و میزان رشد اقتصادی هیچ رابطه ی وجود ندارد.

انکشاف فرضیه (Developing Hypothesis)

در هنگام انکشاف یا ساختن فرضیه نکات مهم که در پائین به آن اشاره میشود را در نظر گرفت:

- فرضیـه بایـد بـه صـورت یـک جملـه خـبری، روشـن و بـدون ابهـام بیـان شـود. چـون بیـان فرضیـه بـه شـکل خـبری قابـل فهـم تـر اسـت.

  مثال: شاگردانیکه باهوش تر اند فرصت بیشتر برای ادامه تحصیل دارند.
- فرضیـه بایـد رابطـه بیـن یـک یـا چنـد متغیـر را بیـان کنـد. رابطـه معمـولاً بـه شـکل مثبـت و مسـتقیم بیـان

میشود هر چند در بعضی مواقع رابطه به شکل معکوس نیز بیان میشود. مثال که در بالا به آن اشاره شد نشان دهنده رابطه مستقیم و مثبت است.

- فرضیه باید قابل آزمون باشد. فرضیه باید طوری انتخاب شود که بتوان با استفاده از دیتا و مشاهدات آنها را رد یا تایید کرد. در صورت تایید فرضیه بتوان نتایجی را از آن پیش بینی کرد.
- در نوشته های علمی کوتاه بودن و واضح بودن همیشه دارای اهمیت می باشد. در نتیجه یک فرضیه باید کوتاه و واضح باشد.

فرضیه نامناسب: اثر تدریس به شیوه سخنرانی و پرسش و پاسخ در یادگیری شاگردان کمتر از روش شاگرد محوری است.

در فرضیه بالا اختصار نوشته کردن مراعت نشده است.

فرضیه مناسب: تاثیر تدریس به روش سخنرانی در یادگیری شاگردان کمتر از روش آزمایشی است.

مثال اول:

در یک آزمایش که در لابراتوار ادویه جات وزارت صحت عامه بالای یک داروی جدید صورت گرفته:

فرضیه صفر (H_0): ادویه که جدید انکشاف و تولید شده است، از نگاه تاثیرگذاری با ادویه که در گذشته ساخته شده تفاوت ندارد.

فرضیه تحقیق (H_A): ادویه که جدید تولید شده، تاثیر گذاری آن متفاوت تر از ادویه همانند است که در گذشته موجود بوده است.

مثال دوم:

فرضیه صفر (H_0): روشن بودن چراغ های اتاق در شب و بیدار ماندن یک شاگرد تا دیروقت در شب باهمدیگر ارتباط ندارند.

فرضیه تحقیق (H_A): روشن ماندن چراغ های اتاق در شب، سبب بیدار ماندن شاگردان تا دیروقت در شب می شود.

تمرین

1. فرضیه چیست در باره آن معلومات دهید؟
2. در مورد کارهای روزمره وظیفه تان فکر نموده و نظر به تصور خودتان یک فرضیه بسازید؟
3. یک فرضیه بسازید و هم در فرضیه خویش متغیر مستقل و وابسته را نشان داده و قابل اندازه گیری بودن متغیرها را مشخص سازید؟
4. یک فرضیه متناوب و یک فرضیه صفری را انکشاف دهید؟
5. هدف یک فرضیه چیست با ذکر یک مثال توضیح دهید؟

بخش پنجم

آزمون های احصائیوی

آزمون همبستگی (Correlation)

اصطلاح همبستگی رابطه بین دو متغیر را نشان میدهد. به عنوان مثال زمانیکه والدین فرزندان شان را تشویق به درس خواندن بیشتر می نمایند،

در حقیقت آنها بر این باور هستند که تلاش بیشتر فرزندانشان باعث افزایش نمرات آنها در امتحانات میشود. به عنوان مثال این باور نزد مردم وجود دارد که میان افزایش سن افراد و فشار خون شان رابطه وجود دارد.

تعریف: همبستگی یک تحلیل احصائیوی است که هدف آن مطالعه رابطه بین دو متغیر می باشد، به عنوان مثال رابطه بین قد پدر و پسر، بارش باران و حاصلات زراعتی، معاش یک کارمند و کیفیت انجام کار.

تصویر که رسامی شده است میتواند نشان دهنده رابطه بین ثروت و خوشحالی باشد. یکی از افراد که ثروتمند تر است خوشحال تر بوده و شخص دیگر ثروتمند نیست ناراضی به نظر می رسد. و هدف اساسی از آزمون همبستگی نشان دادن رابطه بین متغیر ها است که در این شکل ثروت یک متغیر و سطح رضایت از زندگی یا خوشحالی متغیر دیگر می باشد.

شکل: ۵-۱ ارتباط همبستگی میان ثروت و خوشحالی

تحلیل احصائیوی همبستگی به اندازه گیری شدت و چگونگی موجودیت رابطه بین دو متغیر می پردازد. لازم به یادآوری می باشد که همبستگی رابطه علت و معلول را نشان نمیدهد. به طور مثال در انگلستان معلومات در مورد اشخاصی که مبتلا به سرطان بودند جمع آوری شد و در نتیجه میان استفاده از سگرت و سرطان شش رابطه همبستگی مثبت بدست آمد. اما این ارتباط همبستگی مثبت به معنای رابطه علت معلولی میان این دو متغیر نیست شاید یک عامل سومی دیگر مانند آلودگی هوا باعث افزایش مرض سرطان شده باشد.

همبستگی مثبت و منفی (Negative & Positive Correlation): همبستگی مثبت و منفی دو متغیر وابسته به جهت حرکت دو متغیر است.

اگر هر دو متغیر هم جهت حرکت کنند به این معنی که با افزایش یک متغیر، متغیر دیگر نیز افزایش پیدا کند و یا اینکه با کاهش یک متغیر، متغیر دیگر نیز کاهش پیدا کند در این صورت رابطه بین دو متغیر

مثبت و یا مستقیم گفته میشود. از مثال همبستگی مثبت میتوان به رابطه مستقیم بین افزایش قد و وزن، حاصلات زراعتی و بارش باران، نمرات مضمون ریاضی و احصائیه اشاره کرد.

اگر دو متغیر خلاف جهت همدیگر حرکت کنند، به این معنی که با افزایش یک متغیر، متغیر دیگر کاهش

پیـــدا کند و یـــا اینکه با کاهـــش یک متغیر متغیـــر دیگر افزایش پیـــدا کند، در ایـــن صورت **ارتباط همبســـتگی منفی یـــا غیر مســـتقیم** بین دو متغیر وجـــود دارد. مانند مقـــدار محصولات زراعتی در بـــازار و قیمت آنها، در بـــازار از مثال همبســـتگی منفی می باشد.

ضریب همبســـتگی(.Correlation Coefficient) شـــدت رابطه بین دو متغیر توســـط ضریب همبســـتگی نشـــان داده میشـــود و ســـمبول ضریب هبســـتگی عبارت از (r) بوده که قیمـــت 1- الـــی 1+ را گرفته می تواند. در هنگام تفســـیر ضریب همبســـتگی نـــکات ذیل دارای اهمیت می باشـــد:

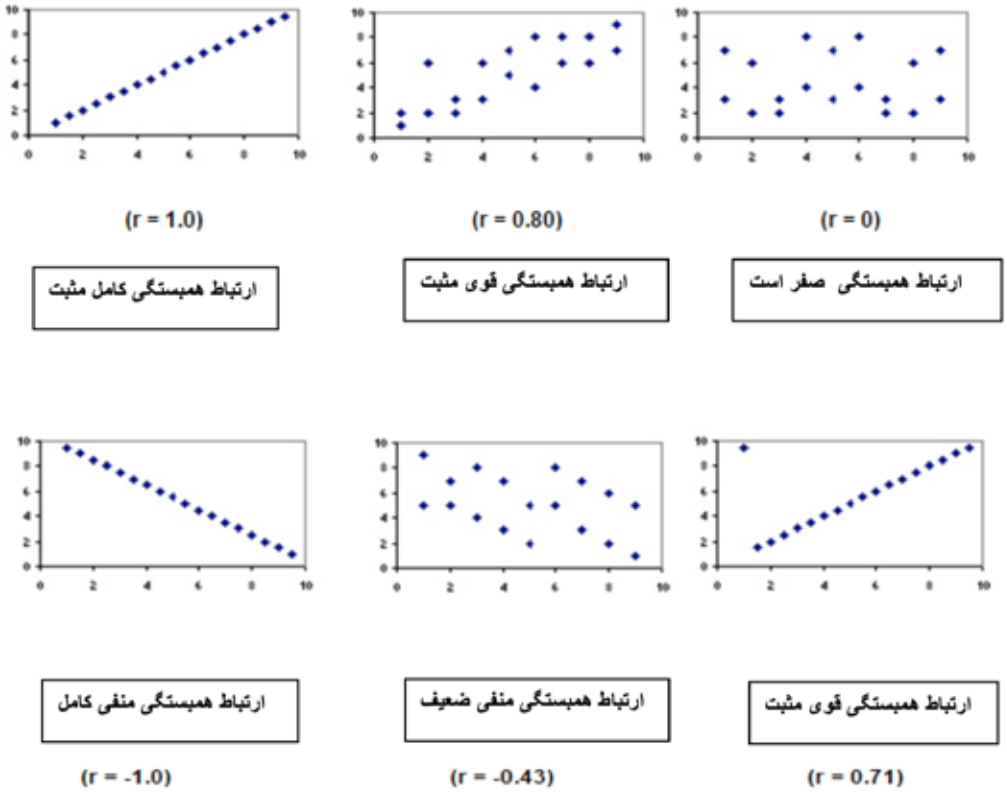
الف: هرگاه ضریب همبستگی r = +1 در اینصورت ارتباط کامل و مثبت بین دو متغیر وجود دارد.

ب: هرگاه ضریب همبستگی r = -1 باشد در اینصورت ارتباط همبستگی کامل و منفی بین دو متغیر وجود دارد.

ج: هرگاه ضریب همبستگی r =0 در این صورت هیچ ارتباط همبستگی بین دو متغیر وجود ندارد.

د: هـــرگاه قیمـــت مطلقه ضریب همبســـتگی بزرگتر از 0.5 (r>0.5) باشـــد در ایـــن صورت ارتباط همبســـتگی بین دو متغیر قوی می باشـــد.

در شکل پائین انواع مختلف ارتباط همبستگی بین متغیر ها را دیده میتوانید.



فرمول ارتباط همبستگی:

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} \quad , \quad X = x - \bar{x} \ , \ Y = y - \bar{y}$$

x عبارت از هر مشاهده جداگانه متغیر مستقل یا x

y عبارت از هر مشاهده جداگانه متغیر وابسته یا y می باشد.

مثال: ریاست زراعت ولایت پروان برای بررسی اینکه تاثیر افزایش کود اصلاح شده زراعتی و حاصلات زمین چگونه رابطه دارد، از آزمون همبستگی استفاده کرده و نتایج آن در پائین به تفسیر گرفته شده:

حاصلات زمین به تن = y

مقدار کود زراعتی به کیلوگرام = x

معلومات در مورد کود اصلاح شده زراعتی و حاصلات یک قطعه زمین در ولایت پروان

XY	y^2	Y= y - $\bar{y}$ X = x- 68	x^2	X= x-$\bar{x}$ X= x- 67	y	x
6	4	-2	9	-3	66	64
2	1	-1	4	-2	67	65
3	9	-3	1	-1	65	66
0	0	0	0	0	68	67
2	4	2	1	1	70	68
0	0	0	4	2	68	69
12	16	4	9	3	72	70
25	34	0	28	0	476	469

شکل: ۵-۲ معلومات در مورد کود اصلاح شده زراعتی و حاصلات زمین

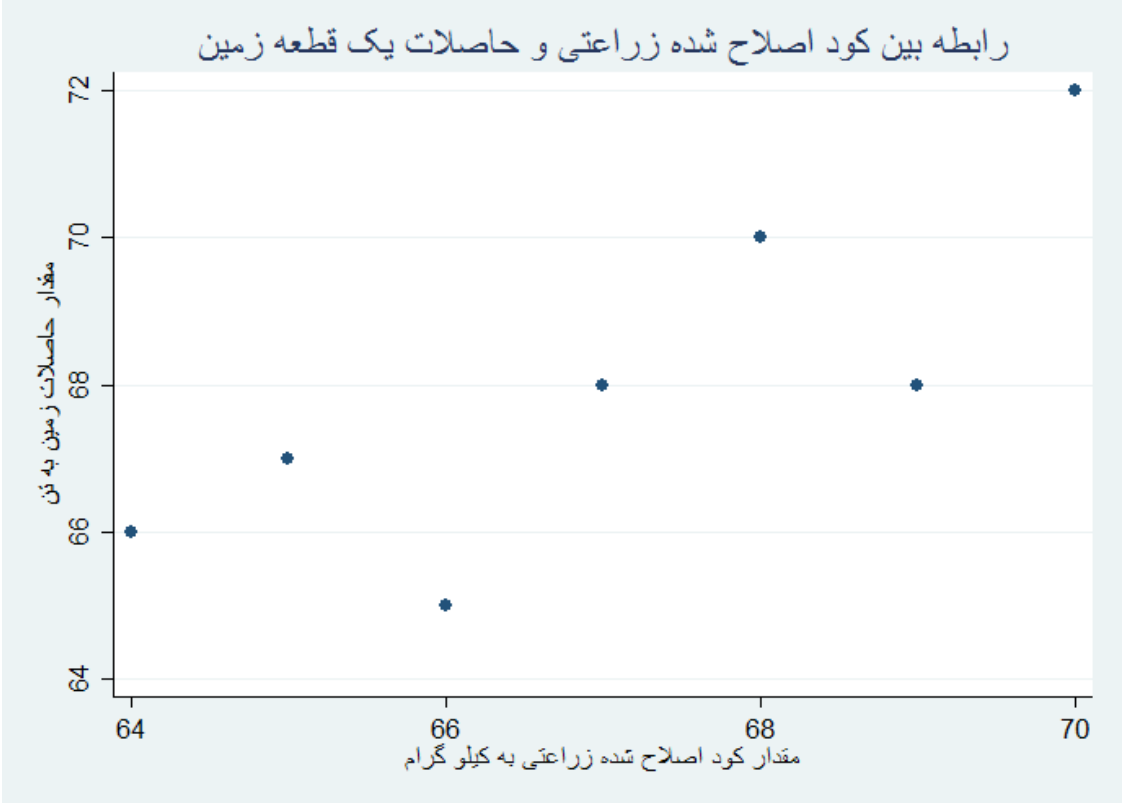
فرمول:

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} \quad , \quad X = x - \bar{x} \ , \ Y = y - \bar{y}$$

$$\bar{x} = \frac{469}{7} = 67 \ , \ \bar{y} = \frac{476}{7} = 68$$

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \sum Y^2}} = \frac{25}{\sqrt{28 * 34}} = \frac{25}{\sqrt{952}} = \frac{25}{30.85} = 0.81$$

تفسیر ضریب همبستگی (Correlation Coefficient): در این مثال بعد از سنجش ضریب همبستگی رابطه بین کود اصلاح شده زراعتی و حاصلات زمین مثبت بدست آمده، یعنی با افزایش کود اصلاح شده زراعتی حاصلات زمین نیز افزایش پیدا میکند. چون ارزش (r) علامه آن مثبت و ارزش آن بزرگتر از 0.5، یعنی +0.8 می باشد، علامه مثبت آن نشانه دهنده رابطه مستقیم و ارزش مطلقه آن نشان دهنده قوت و شدت رابطه می باشد.



شکل: ۵-۳ رابطه همبستگی میان کود اصلاح شده زراعتی و حاصلات زمین زراعتی

مثال: معلومات در مورد درجه حرارت و فروشات شرکت آیسکریم داده شده است، با استفاده از ارتباط همبستگی بگویید که چگونه رابطه بین فروشات آیسکریم و درجه حرارت وجود دارد؟

معلومات در مورد درجه حرارت و فروشات شرکت آیسکریم

فروشات آیسکریم y = به افغانی	درجه حرارت x = به سانتیگراد
215	14.2
325	16.4
185	11.9
332	15.2
406	18.5

22.1	522
19.4	412
25.1	614
23.4	544
18.1	421
22.6	445
17.2	408

شکل: ۵-۴ معلومات در مورد درجه حرارت و فروشات شرکت آیسکریم

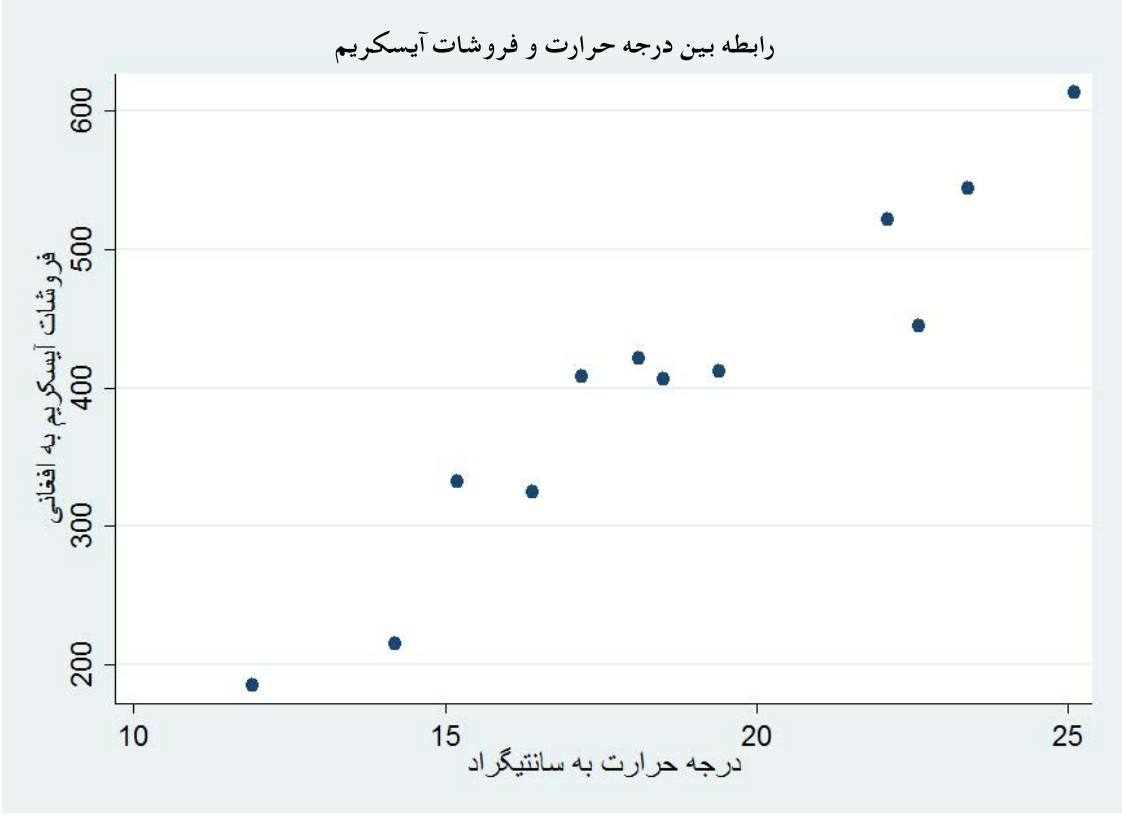
حل:

$$r = \frac{\sum XY}{\sqrt{\sum X^2 \ \sum Y^2}} = \frac{5,325}{\sqrt{177 * 174.757}} = 0.9575$$

دو شـــکل که در پائین کنار هم رســـامی شده خواســـته اســـت که رابطه بین دو متغیر را نشـــان دهد. یکی آن فروشات آیســـکریم و دیگـــری درجه حرارت بـــه این معنی که با گرم شـــدن هوا فروشـــات آیســـکریم افزایش پیـــدا می کند و برعکـــس با سردی هوا فروشـــات آن کاهش پیـــدا می کند.

شکل: ۵-۵ رابطه همبستگی میان درجه حرارت و فروشات آیسکریم

تفسیر ضریب همبستگی: در اینجا ضریب همبستگی عبارت از (r=+0.95) می باشد که علامه مثبت آن نشان دهنده رابطه مستقیم بین درجه حرارت و فروشات آیسکریم دارد، یعنی به هر اندازیکه درجه هوا گرم تر میشود در نتیجه فروشات آیسکریم نیز افرایش پیدا می کند، و چون ارزش ضریب همبستگی بزرگتر از 0.5 می باشد نشان دهنده رابطه قوی بین دو متغیر است.



شکل: ۵-۶ رابطه همبستگی میان درجه حرارت و فروشات آیسکریم

سطح معنی داری (P-Value): در هنگام مطالعه ارتباط همبستگی سطح معنی داری دارای اهمیت زیادی می باشد. سطح معنی داری به سمبول (P-Value) نشان داده شده و سطح معنی داری نشان دهنده این موضوع است که ضریب همبستگی به دست آمده دارای اعتبار می باشد و تصادفی و از روی چانس بدست نیامده است.

قیمت که (P-Value) میگیرد اگر کوچکتر از 0.05 (P-Value<0.05) باشد، در این صورت رابطه همبستگی بین دو متغیر از نگاه احصائیوی معنی دار می باشد.

رگرسیون (Regression)

در بحث گذشته به مطالعه ارتباط همبستگی پرادختیم که برای بررسی چگونگی ارتباط بین دو متغیر بود، حالا به مطالعه رگرسیون می پردازیم که برای مطالعه موجودیت رابطه بین دو متغیر یا بیشتر از آن می پردازد و تفاوت آن با ارتباط همبستگی عبارت از این است که در ارتباط همبستگی متغیرها به متغیرهای وابسته و مستقل تقسیم بندی نمی شود، حال آنکه در رگرسیون متغیر ها به دو گروپ تقسیم میشوند، یکی متغیر مستقل و دیگری وابسته و رگرسیون به مطالعه رابطه بین متغیر های مستقل و وابسته می پردازد.

در رگرسیون متغیر ها به دو دسته تقسیم میشوند، یکی متغیرهای مستقل (**Independent Variables**) که مستقلانه قیمت میگیرند و دیگر متغیرهای وابسته (**Dependent Variables**) که قیمت آنها وابسته به قیمت متغیرهای مستقل است.

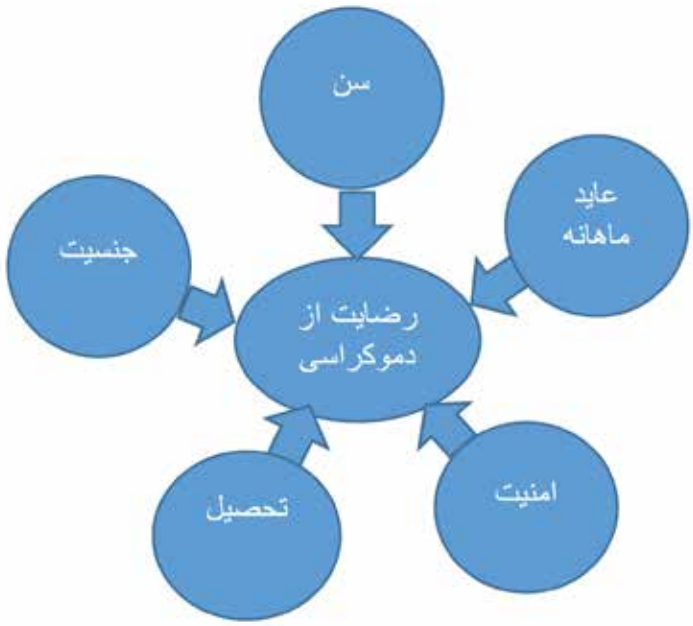
رگرسیون که یک تحلیل احصائیوی می باشد برای ساختن مدل های تحلیلی مورد استفاده قرار می گیرد. در هنگام ساختن مدل ها، متغیرها به یک متغیر وابسته و حد اقل یک متغیر مستقل تقسیم میشوند. بعد از ساختن مدل و انجام تحلیل رگرسیون معلوم می گردد که تغییرات در متغیر مستقل چگونه باعث تغییر در متغیر وابسته میشود.

در اینجا از انواع مختلف رگرسیون تنها دو نوع آن را به مطالعه می گیریم.

1. رگرسیون حداقل مربعات (OLS regression or Ordinary Least Square)
2. رگرسیون پروبیت (Probit Regression)

نکته: رگرسیون یکی از مباحث احصائیوی است که در مورد آن کتاب و معلومات بسیار موجود است و در اینجا تنها مباحث که برای ما در هنگام استفاده برنامه ستاتا (Stata) دارای اهمیت می باشد مورد مطالعه قرار می گیرد.

رگرسیون به طریقه حداقل مربعات (Ordinary Least Square or OLS Regression) برای آشنایی بیشتر با رگرسیون، در ابتدا یک مدل تحلیلی ساخته و متغیرها را به دو دسته مستقل و وابسته تقسیم می کنیم. به مثال ذیل توجه کنید!



معادله رگرسیون به شکل ذیل می باشد:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \ldots + u$$

در معادله رگرسیون، (x1, x2…xn) عبارت از متغیرهای مستقل می باشد که قیمت های مستقل را به خود میگیرند و (y) عبارت از متغیر وابسته می باشد که قیمت های آن وابسته به قیمت متغیرهای مستقل می باشد.

تحلیل رگرسیون در واقع تاثیرات متغیرهای مستقل را بالای متغیر وابسته طوری مطالعه میکند که تاثیرات متغیرهای مستقل دیگر ثابت بماند. به عبارت دیگر، ضریب های β برای هر متغیر مستقل، تاثیر خالص همان متغیر مستقل بالای متغیر وابسته می باشد. برای محاسبه β یکی از طریقه ها طریقه حداقل مربعات می باشد. در برنامه (Stata) توسط فرمان (reg) می باشد که در بخش های قبلی در مورد آن به تفصیل معلومات داده شده است که مطالعه آن قبل از مطالعه این بخش ضروری می باشد.

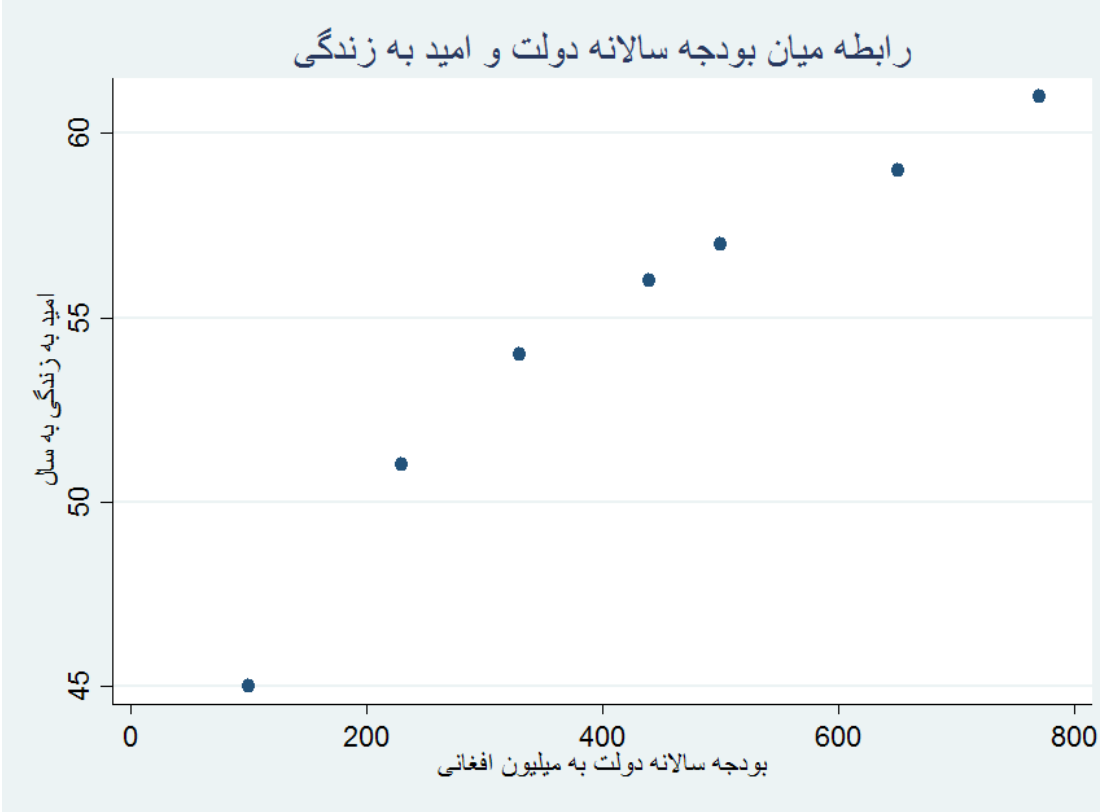
فرضیه های رگرسیون خطی (OLS Regression Assumptions)

در این بخش به تشریح فرضیه های رگرسیون خطی می پردازیم. بخاطر داشته باشید که ابتدا باید فرضیه رگرسیون خطی در نظر گرفته شده و بعد از آن فرمان رگرسیون انجام شود و از نتایج به دست آمده تفسیر رابطه بین متغیرها صورت گیرد.

1. **فرضیه رابطه خطی بین متغیر وابسته و متغیرهای غیر وابسته (Linearity):**

در رگرسیون خطی چند متغیره یکی از فرضیه ها این است که رابطه بین متغیرهای مستقل و وابسته باید خطی

باشـــد. و برای این کار میتوان از نمودار پراگندگی (Scatter-plot) اســـتفاده کرد.



شکل: ۵-۷ رابطه میان بودجه خدمات صحی و امید به زندگی

3. تغیر پذیری در تمام متغیرها (variability in all variables)

ایـــن موضوع بســـیار حایز اهمیت اســـت که هر یـــک از متغیرها باید تغیرپذیر باشـــند. زیـــر اگر متغیـــری از ابتدا تا انتهـــا قیمت یکســـان و تغییر نا پذیر داشـــته باشـــد، در این صورت متغیر گفته نمی شـــود و جـــا دادن چنین یک متغیر در رگرســـیون هیچ معنایی نخواهد داشـــت. پس تغیـــر پذیر بودن قیمت ها یا مشـــاهدات موجـــود در یک متغیر یکی از اساســـی ترین شرط ها برای رگرســـیون است.

مثـــال: اگـــر عاید ماهانه مد نظر باشـــد، باید چند عاید ماهانه با ارزشـــهای متفاوت داشـــته باشـــیم تـــا اینکه در مدل شامل گردد.

(100 افغانـــی، 230 افغانـــی، 450 افغانـــی و ....) این عاید چون دارای چنـــد عاید با ارزش های متفاوت اســـت، میتواند در مدل شـــامل گردد.

(100 افغانـــی، 100 افغانـــی، 100 افغانـــی) این عایـــد ماهانه به عنوان یک متغیر در مدل شـــامل شـــده نمیتواند، چون

کـــه تمـــام ارزش های آن 100 افغانی می باشـــد و تغییر را نشـــان نمیدهد و یـــا واریانس ندارد.

4. نمونه تصادفی (Random Sample)

نمونـــه که از جمعیت گرفته میشـــود و از آن متغیرها برای ســـاختن مدل رگرســـیون اســـتفاده میشـــود، باید به شـــکل تصادفـــی انتخاب شـــده باشـــد، و معلومات جمع آوری گردیده باشـــد.

5. مولتی کولینریتی "هم خطی چندگانه" (Multicollinearity)

مولتـــی کولینریتـــی به ســـخن ســـاده زمانی وجـــود دارد کـــه دو و یا بیشـــتر از دو متغیر مســـتقل با هـــم رابطه قوی همبســـتگی داشـــته باشـــند. برای آزمون این فرضیه در پروگـــرام ســـتاتا (Stata) میتوانیم از فرمان (pwcorr) اســـتفاده کنیـــم، در صورتی که ضریب همبســـتگی بزرگتر از 0.8 باشـــد، باید یکی از دو متغیر مســـتقل از مدل رگرســـیون بیرون کشیده شود.

مثال: در این مثال دو متحول مستقل (قد و سن) را مدنظر میگیریم

در ایـــن مدل میخواهیم بررســـی کنیم کـــه آیا رابطه بین این دو متحول مســـتقل وجـــود دارد یا خیر. بـــرای انجام این کار از آزمون همبســـتگی اســـتفاده میکنیم و رابطه همبســـتگی بین دو متحول مســـتقل قد و ســـن را بررســـی میکنیم. بعـــد از آزمون همبســـتگی بیـــن دو متغیر به این نتیجه می رســـیم کـــه ارتباط همبســـتگی قوی میان ایـــن دو متغیر وجود داشـــته اســـت و ضریـــب همبســـتگی (r =0.894) بزرگتر از 0.8 اســـت که در نتیجـــه یکی از ایـــن متغیر های مســـتقل را از مدل رگرســـیون نا گزیر بیرون می ســـازیم.

```
pwcorr height age, sig
```

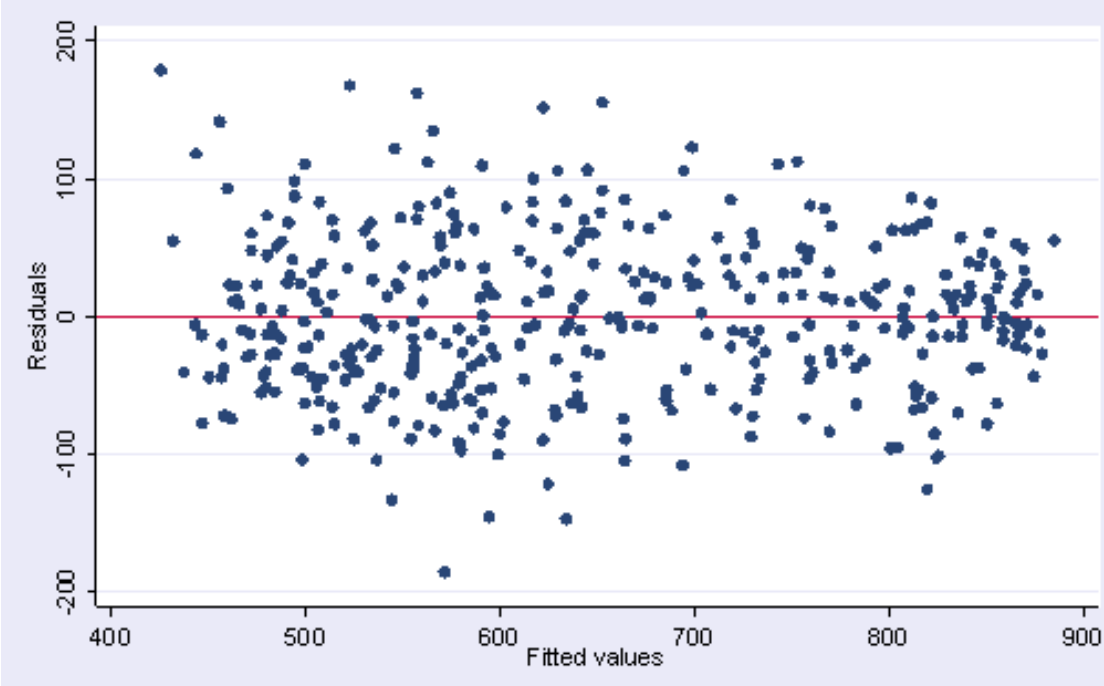
	height	age
height	1.0000	
age	0.8944 0.0066	1.0000

6. فرضیه برابری انحراف ها(Homoscedasticity) :

یکـــی دیگـــر از فرضیه های رگرســـیون طریقه حداقـــل مربعات عبـــارت از فرضیه برابـــری وریانس خطاها می باشـــد. خطـــا ها عبـــارت از تفـــاوت بین قیمت هـــای واقعی (y) و قیمت های تخمین شـــده آن به صورت اوســـط به اســـاس قیمت های (x) می باشـــد.

برای انجام این آزمون در برنامه ستاتا از فرمان ذیل استفاده میکنیم. البته به خاطر داشته باشید که در مرحله اول باید رگرسیون اجرا گردد و بعد از آن فرمان برای انجام آزمون برابری واریانس خطا ها انجام شود.

rvfplot, yline(0)



شکل: ۵-۹ فرضیه برابری انحراف خطا ها

در شکل بالا دیده میشود که نقطه ها در اطراف (بالا و پائین) خط سرخ با واریانس تقریبن مساوی پراگنده می باشند و فرضیه درست است. هرگاه نقطه ها در اطراف خط با واریانس تقریبن مشابه به حرکت نباشد، در این صورت مشکل نابرابری انحراف خطاها (Heteroskedasticity) پدید می آید.

داده های خارج از محدوده (Outlier) یکی از دلایل پدید آمدن مشکل انحراف خطاها می باشد که نتیجه رگرسیون چند متغیره خطی را تحت تاثیر قرار میدهد. یک مثال میتواند که برای فهم موضوع کمک کند.

به گونه مثال وزارت زراعت معلومات را در ارتباط به محصولات زراعتی پنج ولایت کشور در سال 1393 به نشر رسانده اما محصولات زراعتی یک ولایت نسبت به ولایت های دیگر تفاوت بسیار دارد و از حالت معمول متفاوت است. در جدول ذیل به محصولات زراعتی ولایت لوگر نگاه کنید.

محصولات گندم پنج ولایت افغانستان در سال 1393

نام ولایت	محصولات به تن
کابل	200
غزنی	230
میدان وردک	130
لوگر	38,000
پنجشیر	234

شکل: ۵-۸ داده های خارج از محدوده

7. متغیر وابسته پیوسته (Continuous Dependent Variable)

در مدل رگرسیون به طریقه ای حد اقل مربعات که ترتیب می گردد، باید متغیر وابسته یک متغیر گسسته باشد. از مثال متغیر پیوسته میتوان به قد درآمد نمرات امتحان فروشات اشاره کرد.

اما متغیرهای مستقل میتوانند که پیوسته یا دارای کتگوری (Categorical Variable) باشد، مثال های متغیر پیوسته در بالا ذکر گردید و مثال های متغیر دارای کتگوری عبارت اند از: جنسیت (مرد یا زن)، ملیت (تاجیک، هزاره، پشتون و ...)، محل زندگی (شهر یا دهات).

نکته: در اینجا به چند مثال رگرسیون پرداخته و موضوع مهم تفسیر نتیجه رگرسیون می باشد.

مثال: در این مثال اول یک رگرسیون ساده یعنی متشکل از دو متغیر، یکی وابسته و دیگری مستقل می باشد، به عنوان مثال میخواهم بدانیم که آیا با افزایش سطح تحصیل رضایت مردم از دموکراسی افزایش پیدا می کند یاخیر؟

برای رسیدن به این نتیجه مراحل ذیل ضروری می باشد:

1. **دانستن کودها و لیبل های متغیرها**
2. **انجام فرمان رگرسیون**
3. **تفسیر نتیجه رگرسیون**

1. تحلیل کودهای متغیرها:

متغیر (q32) نشان دهنده سطح رضایت مردم از دموکراسی می باشد که کودهای آن از 1 الی 4 می باشد. که از کود 1 شروع شده به طرف کود 2 رفته و به همین ترتیب تا کود 4 نشاندهنده رضایت کمتر از دموکراسی می

باشد.

متغیـــر (d10) دارای کودهـــای 1 الی 13 می باشـــد و بـــا هر واحد از کود 1 الی 13، نشـــان دهنده تحصیل بیشـــتر می باشد.

```
. labellist q32 d10 m6b d1
Q32:
           1 Very satisfied
           2 Somewhat satisfied
           3 Somewhat dissatisfied
           4 Very dissatisfied
          98 Refused
          99 Don't Know
D10:
           1 Never went to a school
           2 Informal schooling at home or at a literacy class
           4 Primary School, incomplete (classes 1 to 5)
           5 Primary School, complete (finished class 6)
           6 Secondary education, incomplete (classes 7 to 8)
           7 Secondary education, complete (finished class 9)
           8 High School incomplete (classes 10-11)
           9 High School complete (finished class 12)
          10 14th grade incomplete (class 13)
          11 14th grade complete (finished class 14)
          12 University education incomplete (have no degree diploma)
          13 University education complete (have degree diploma)
          98 Refused (vol.)
          99 Don't know (vol.)
M6B:
           1 Rural
           2 Urban
D1:
           1 Male
           2 Female
```

2. **اجرای فرمان رگرسیون در برنامه :**(Stata)اجرای فرمان رگرسیون به شکل ذیل می باشد،

reg [dependent variable] [independent variables]

reg q32 d10 if q32<98

3. **تفسیر نتیجه فرمان رگرسیون بعد از اجرای آن در برنامه (Stata)**

```
. reg q32 d10 if q32<98

      Source |       SS       df       MS              Number of obs =     9,468
-------------+------------------------------           F(1, 9466)    =      0.09
       Model |  .073044591     1  .073044591            Prob > F      =    0.7663
    Residual |  7823.77349 9,466  .826513151            R-squared     =    0.0000
-------------+------------------------------           Adj R-squared =   -0.0001
       Total |  7823.84654 9,467  .826433562            Root MSE      =    .90913

------------------------------------------------------------------------------
         q32 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d10 |   .0007893   .0026551     0.30   0.766    -.0044152    .0059938
       _cons |   2.399013    .013428   178.66   0.000     2.372691    2.425335
------------------------------------------------------------------------------
```

بعـــد از اجرای فرمان رگرســـیون برنامه ســـتاتا نتیجه که در بـــالا ملاحظه میکند را نشـــان میدهد. حالا نـــکات که برای تحلیـــل ما دارای اهمیت می باشـــد در اینجـــا به ترتیب به تفســـیر میگیریم.

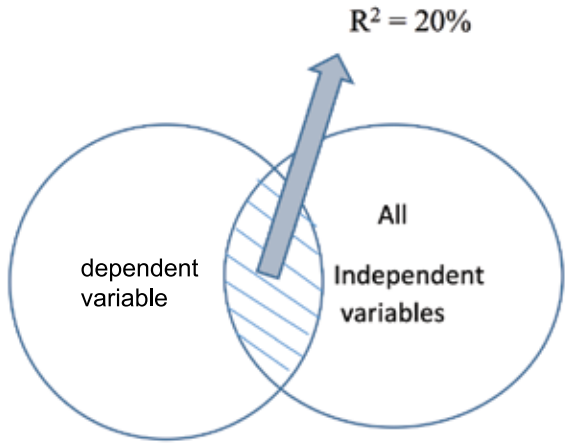
1. **ضریـب رگرسـیون (Regression Coefficient):** در مـودل هـای رگرسـیون سـاده تنهـا یـک ضریـب رگرسـیون وجـود دارد.

   ایـن ضریـب نشـان دهنـده آن اسـت کـه بـا یـک واحـد یـا یـک سـال افزایـش در سـطح تحصـل بـه انـدازه (β =0.00078) واحـد رضایـت از دموکراسـی کاهـش پیـدا مـی کنـد. یـا بـه تفسـیر سـاده تـر شـهروندان تحصیـل کـرده نسـبت بـه افـراد کـه تحصیـل کمـتر دارنـد رضایـت کمـتر از دموکراسـی نشـان داده انـد.

2. **P-value یـا P>|t|** همانگونـه کـه در مبحـث ارتبـاط همبسـتگی مطالعـه کردیـم نشـان دهنـده سـطح معنـی داری مـی باشـد و اگـر قیمـت آن از 0.05 کوچکـتر باشـد رابطـه بدسـت آمـده معنـی دار مـی باشـد. امـا در ایـن مثـال قیمـت آن بزرگـتر از 0.05 اسـت، پـس در نتیجـه رابطـه بدسـت آمـده معنـی دار نیسـت و نیـاز بـه تفسـیر آن نیسـت امـا در بـالا تنهـا بخاطـر کمـک بـرای یـاد گیـری نتیجـه رگرسـیون قیمـت β تفسـیر شـده اسـت.

3. **R-squared:** نـام دیگـر آن (Coefficient Determination) ضریـب تشـخیص مـی باشـد، کـه نشـان دهنـده توانایـی مـودل در توضیـح دادن تغییـرات در متغیـر وابسـته مـی باشـد. و قیمـت آن در اینجـا عبـارت از (0.000) مـی باشـد بـه معنـی اینکـه تقریبـاً ایـن مـودل بـه انـدازه 0% توانایـی تغییـرات در متغیـر وابسـته (رضایـت از دموکراسـی) را توسـط متغیـر مسـتقل (سـطح تحصیـل) دارا مـی باشـد.

نکتـــه: با اینکه ضریب تشـــخیص در مثال گذشـــته تقرین 0% بود، اما اگر ضریب تشـــخیص 20% باشـــد. میتوان آنرا به شکل ذیل توسط یک شکل گرافیکی به منظور فهم بهتر نشان داد.



مثـــال 2. در ایـــن مودل کـــه در اینجا به آن مـــی پردازیم برعـــلاوه تحصیل متغیرهای مســـتقل دیگر را نیـــز در مودل شـــامل می ســـازیم تا رابطه آنها را نیز با متغیر وابســـته (رضایت از دموکراســـی) به مطالعه بگیریم. مانند: ســـن (d2)، جنســـیت (d1)، عاید ماهانه (d18a) و شـــهری یـــا دهاتی بودن (m6b).

برای دانستن روابط بین متغیرهای مستقل و وابسته در این مدل باید مانند قبل سه مرحله را اجرا کنیم.

1. **دانستن کودها و لیبل های متغیرها**
2. **انجام فرمان رگرسیون**
3. **تفسیر نتیجه رگرسیون**

1. دانستن کود و لیبل متغیرها

- در متغیـــر (m6b) بـــا حرکت از کود 1 به 2 یعنی نظر شـــهری های نســـبت به یک موضوع، در ایـــن مثال رضایت شهری ها از دموکراسی.
- (d1) متغیـــر اســـت کـــه دارای دو کود 1 و 2 می باشـــد و حرکـــت از کود 1 بـــه 2 ،یعنی نظر زنـــان در مورد یک موضـــوع، در اینجا رضایت از دموکراســـی.

```
Q32:
           1 Very satisfied
           2 Somewhat satisfied
           3 Somewhat dissatisfied
           4 Very dissatisfied
          98 Refused
          99 Don't Know
D10:
           1 Never went to a school
           2 Informal schooling at home or at a literacy class
           4 Primary School, incomplete (classes 1 to 5)
           5 Primary School, complete (finished class 6)
           6 Secondary education, incomplete (classes 7 to 8)
           7 Secondary education, complete (finished class 9)
           8 High School incomplete (classes 10-11)
           9 High School complete (finished class 12)
          10 14th grade incomplete (class 13)
          11 14th grade complete (finished class 14)
          12 University education incomplete (have no degree diplo
          13 University education complete (have degree diploma)
          98 Refused (vol.)
          99 Don't know (vol.)
M6B:
           1 Rural
           2 Urban
D1:
           1 Male
           2 Female
```

2. انجام فرمان رگرسیون در برنامه (Stata)

reg [dependent variable] [independent variables]

reg q32 d10 d2 m6b d1 d18a if q32<98 & d18a>99

3. تفسیر نتیجه فرمان رگرسیون بعد از اجرای آن در برنامه (Stata)

```
. reg q32 d10 d2 m6b d1 d18a if q32<98 & d18a>99
```

Source	SS	df	MS		
				Number of obs	= 5,746
				F(5, 5740)	= 4.52
Model	19.2228333	5	3.84456665	Prob > F	= 0.0004
Residual	4880.17588	5,740	.850204857	R-squared	= 0.0039
				Adj R-squared	= 0.0031
Total	4899.39871	5,745	.852810916	Root MSE	= .92207

q32	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
d10	-.0037196	.0036565	-1.02	0.309	-.0108878	.0034486
d2	.0015388	.0009915	1.55	0.121	-.000405	.0034826
m6b	.0410946	.0318054	1.29	0.196	-.021256	.1034452
d1	-.1044162	.0260869	-4.00	0.000	-.1555564	-.053276
d18a	-7.95e-08	3.57e-07	-0.22	0.824	-7.80e-07	6.21e-07
_cons	2.481562	.0697972	35.55	0.000	2.344733	2.61839

1. تفسیر رابطه بین سن و رضایت از دموکراسی، یعنی با یک واحد (یک سال) افزایش در سن به اندازه (0.00153) واحد رضایت از دموکراسی کمتر می شود. تفسیر رابطه بین شهری بودن و دهاتی بودن، در اینجا رابطه بین این دو متغیر مستقیم می باشد، یعنی با افزایش هر واحد در متغیر شهری یا دهاتی بودن (m6b) باعث افزایش در متغیر (q32) رضایت از دموکراسی به اندازه (β=0.04109) واحد میشود، یعنی شهری ها بیشتر از کسانیکه در روستاها زندگی میکنند، از دموکراسی رضایت دارند.

2. تفسیر رابطه بین جنسیت و رضایت از دموکراسی، در اینجا رابطه بین این دو متغیر غیر مستقیم می باشد، یعنی یک واحد افزایش در متغیر جنسیت (d1) باعث کاهش در متغیر (q32) رضایت از دموکراسی به اندازه (β =0.1044) واحد میشود، یعنی زنان بیشتر از مردان از دموکراسی در کشور رضایت دارند.

3. R-squared: نام دیگر آن (Coefficient Determination) ضریب تشخیص می باشد، که نشان دهنده توانای مودل در توضیح دادن تغییرات در متغیر وابسته می باشد. و قیمت آن در اینجا عبارت از (0.0039) می باشد به معنی اینکه تقریبن این مودل به اندازه 0.39% توانایی تغییرات در متغیر وابسته (رضایت از دموکراسی) را توسط متغیرهای مستقل (سطح تحصیل، جنسیت، سن، عاید ماهانه و شهری ودهاتی) دارا می باشد.

4. در این مودل رابطه بین هیچ یک از متغیر های مستقل با متغیر وابسته معنی دار نمی باشد به استثنای رابطه بین جنسیت و رضایت از دموکراسی که قیمت P-value = 0.000 از 0.05 کوچکتر می باشد، و متباقی آنها از 0.05 بزرگتر می باشند.

پروبیت رگرسیون (Probit Regression)

از دو نـــوع رگرســـیون که مـــورد مطالعه ما قرار دارد، ابتدا رگرســـیون به طریقـــه حداقل مربعات را به بررســـی گرفتیم. در ایـــن جا رگرســـیون پروبیت را به تشریـــح می گیریم. این نوع رگرســـیون زمانیکـــه که متغیر وابســـته در یک تحلیل رگرســـیون دارای دو قیمـــت (Dummy variable) باشـــد مورد اســـتفاده قرار می گیرد و رگرســـیون بـــه طریقه حداقل مربعات در این حالت اســـتفاده نمی شـــود.

تفاوت بین رگرسیون به طریقه حداقل مربعات و پروبیت

رگرسیون پروبیت	رگرسیون حداقل مربعات
هنگامی استفاده میشود که متغیر وابسته دارای دو قیمت قیمت های 0 و 1 باشد.	هنگامی استفاده میشود که متغیر وابسته یک متغیر پیوسته (Continuous)

شکل: ۵-۱۰ تفاوت میان رگرسیون به طریقه خطی و رگرسیون پروبیت

مثال: با اســـتفاده از دیتای سروی مردم افغانســـتان در ســـال 2015، بنیاد آســـیا می خواهیم یک مدل تحلیلی ســـاخته و رابطه بین جنســـیت (d1)، ســـن (d2)، دهاتی و شـــهری بودن (m6b) و ســـطح تحصیل (d10)را با نظر مردم افغانستان در مـــورد اینکـــه کشـــور به کدام در حرکت اســـت را به بررســـی می گیریـــم. برای دریافـــت این رابطه ســـه مرحله را مانند رگرســـیون حداقل مربعـــات انجام میدهیم:

1. **دانستن کودها و لیبل های متغیرها**
2. **انجام فرمان رگرسیون**
3. **تفسیر نتیجه رگرسیون**

1. دانستن کود و لیبل متغیرها

کـــود ســـایر متغیرهـــا را از مثال های گذشـــته میدانید، در اینجـــا تنها کود متغیر وابســـته (q1) کـــه دارای دو کود 0 و 1 می باشـــد را مـــورد مطالعه قـــرار میدهیم، البته بخاطر داشـــته باشـــید که قبل از انجـــام فرمان رگرســـیون پروبیت بایـــد کودهـــای متغیر وابســـته را بـــه 0 و 1 تبدیل کنید که این موضوع به تفســـیر گرفته شـــده در فصـــل های قبلی و مطالعه آن ضروری می باشـــد.

```
. labellist q1
direction:
           0 Right direction
           1 Wrong Direction
```

2. انجام فرمان پروبیت رگرسیون

probit [dependent variable] [independent variables]

probit q1 d1 m6b d10 d2

3. تفسیر نتیجه فرمان رگرسیون بعد از اجرای آن در برنامه ستاتا

در تفسیر نتیجه رگرسیون پروبیت یک تفاوت وجود دارد، در بحث مربوط به رگرسیون به طریقه حداقل مربعات شما آموختید که در هنگام تفسیر نتیجه رگرسیون، یک واحد تغییر در متغیر مستقل با ثابت نگه داشته تاثیرات دیگر متغیرهای مستقل به چه تعداد واحد تغییرات (کاهش یا افزایش) در متغیر وابسته رونما می گردید. اما در تفسیر نتیجه رگرسیون پروبیت، این گونه بیان میشود که با یک واحد تغییر در متغیر مستقل با ثابت نگه داشتن تاثیرات دیگر متغیرهای مستقل، متغیر وابسته چند فیصد احتمال وقوع آن کاهش یا افزیش پیدا می کند. برای فهم بیشتر به تفسیر ذیل توجه کنید:

1. در اینجا به تفسیر رابطه بین جنسیت و نظر مردم در مورد اینکه افغاستان به کدام سمت در حرکت است می پردازیم، چونکه علامه ضریب رگرسیون (β=-0.06463) منفی می باشد، رابطه بین متغیر مستقل جنسیت (d1) و متغیر وابسته (q1) منفی بوده، یعنی زن ها نسبت به مردان 6.4 در صد بیشتر احتمال دارد تا بگویند که کشور به سمت درست در حرکت است.

```
. probit q1 d1 m6b d10 d2

Iteration 0:   log likelihood = -6025.1992
Iteration 1:   log likelihood = -5995.5281
Iteration 2:   log likelihood = -5995.5221
Iteration 3:   log likelihood = -5995.5221

Probit regression                               Number of obs     =      8,978
                                                LR chi2(4)        =      59.35
                                                Prob > chi2       =     0.0000
Log likelihood = -5995.5221                     Pseudo R2         =     0.0049
```

direction	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
d1	-.0646357	.0286046	-2.26	0.024	-.1206998	-.0085717
m6b	.2383293	.0359654	6.63	0.000	.1678383	.3088202
d10	-.0053066	.0042001	-1.26	0.206	-.0135387	.0029254
d2	.0030824	.0010977	2.81	0.005	.0009309	.005234
_cons	-.0068433	.077997	-0.09	0.930	-.1597145	.146028

آزمون میانگین دو جمعیت T-test

محققین در رشته های مختلف از t-test برای مقایسه اوسط دو گروپ استفاده میکنند.

مثلآ شما میخواهید تفاوت اوسط سن را در دو ولایت افغانستان دریابید و شما از هر ولایت درمورد سن 3 نفر پرسان میکنید و هر یک به ترتیب ذیل جواب میدهند.

سن سه نفر از باشنده گان ولایت کابل و مزار شریف

(n) تعداد مشاهدات	کابل	مزار شریف
1	20	29
2	28	18
3	36	22
اوسط	28	23

شکل: ۵-۱۰ سن سه نفر از باشنده گان دو ولایت کابل و مزار شریف

به نظر ساده دیده میشود که اوسط سن در شهر کابل(28 سال) نسبت به شهر مزار شریف (23 سال) بزرگتر است اما این تفاوت در اوسط سن دو ولایت میتواند که از نظر احصایوی منطقی نباشد، دلیل آن این است که تنها 3 نفر نمیتواند نماینده گی از تمام شهروندان این ولایت کند یعنی این تفاوت سنی میتواند به اساس چانس باشد.

اما اگر شما از هر شهر (کابل و مزار شریف) در مورد سن 300 باشنده آن بپرسید و هنوز هم یک تفاوت آشکارا در اوسط سنی این دو گروه مشاهده شود پس کمتر احتمال دارد که این تفاوت بر اساس چانس اتفاق بیافتد و به همین ترتیب به هر مقدار که تعداد نمونه ما زیاد باشد چانس اینکه نتیجه بدست آمده ما به اساس چانس نباشد زیادتر است.

به صورت عموم دو نوع T-test وجود دارد.

1. T-test یکطرفه
2. T-test دو طرفه

آزمون میانگین یک طرفه One sided T-test

در صورت که ما دیتای یک گروه را داشته باشیم و از گروه دومی را تنها اوسط آنرا داشته باشیم، بناً برای مقایسه اوسط دو گروه از T-test یکطرفه استفاده میکنیم.

ساختار انجام فرمان T-test در برنامه ستاتا به شکل ذیل میباشد.

ttest var1== [mean_var2]

در اینجا برای اجرای مثال عملی از دیتای سروی مردم افغانستان SAP استفاده میکنیم که شما میتوانید آنرا از لینک ذیل دریافت کنید.

http://asiafoundation.org/where-we-work/afghanistan/survey

یکی از سوالات که در این سروی از سروی شدگان پرسان گردیده است عبارت سن آنها میباشد حالا اگر بخواهیم اوسط سن این گروه را با اوسط سن یک گروه دیگر که به صورت فرضی آنرا 35 سال در نظر گرفتیم مقایسه

کنیم در صفحه فرمان Command Window فرمان ttest نوشته به تعقیب آن متغیر را که نشان دهنده سن است (d2) را نوشته کرده و بعد اوسط سن جمعیت دوم را که به صورت تخمینی 35 فرض نمودیم نوشته میکنیم.

نتیجه فرمان اجرا شده به شکل ذیل میباشد.

```
. ttest d2==35

One-sample t test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
d2	9,586	34.6039	.1299777	12.72587	34.34912	34.85869

```
    mean = mean(d2)                                   t =  -3.0474
Ho: mean = 35                         degrees of freedom =     9585

    Ha: mean < 35               Ha: mean != 35                 Ha: mean > 35
 Pr(T < t) = 0.0012         Pr(|T| > |t|) = 0.0023          Pr(T > t) = 0.9988
```

1. فرضیه H_0 بیانگر این موضوع است که اوسط سن گروه اولی مساوی به 35 (اوسط سن گروه دومی که به صورت تخمینی فرض نمودیم) میباشد.
2. فرضیه H_A بیانگر این موضوع است که اوسط سن گروه اولی مساوی به 35 (اوسط سن گروه دومی که به صورت تخمینی فرض نمودیم) نمیباشد.

اما از روی P-value قضاوت نموده میتوانیم که کدام فرضیه را تایید کنیم و کدام فرضیه را رد کنیم.

در صورت که P-value بزرگتر از 0.05 باشد پس گفته میتوانیم که فرضیه H_0 درست بوده و فرضیه H_A را رد میکنیم اما در صورت که P-vlaue کوچکتر از 0.05 باشد پس فرضیه H_A را تایید کرده و فرضیه H_0 را رد میکنیم.

آزمون میانگین دو طرفه Two sided T-test

در صورت که ما معلومات در مورد هر دو گروه داشته باشیم پس از T-test دو طرفه استفاده میکنیم.

به طور مثال در سال 2015 یکی از سوالات در سروی مردم افغانستان این بود که در مورد سن هر یک از پاسخ دهنده گان (مردان و زنان) سوال شده, حالا اگر بخواهیم بدانیم که اوسط سن مردان و زنان باهم برابر بوده پس از T-test دو طرفه استفاده میکنیم.

ساختار فرمان T-test دو طرفه در برنامه ستاتا به شکل ذیل میباشد.

ttest continuous variable = by (category variable)

ttest d2, by (d1)

و نتیجه فرمان به شکل ذیل در صفحه ستاتا نمایش داده میشود.

```
. ttest d2 , by(d1)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
Male	4,847	36.07427	.1946841	13.55399	35.6926	36.45594
Female	4,739	33.10002	.1689367	11.62967	32.76883	33.43122
combined	9,586	34.6039	.1299777	12.72587	34.34912	34.85869
diff		2.974252	.2582042		2.468117	3.480386

```
    diff = mean(Male) - mean(Female)                          t =  11.5190
Ho: diff = 0                                     degrees of freedom =  9584

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

در اینجا به چند نکته ذیل باید توجه شود.

1. فرضیه H_0 بیانگر این موضوع است که اوسط سن بین مردان و زنان یکسان میباشد.
2. برعکس فرضیه H_A بیانگر این موضوع است که اوسط سن مردان و زنان یکسان نمیباشد.

در مورد این که بدانیم کدام فرضیه ما درست میباشد و کدام فرضیه ما اشتباه است از روی P-value میتوانیم قضاوت کنیم.

در صورت که p-value بزرگتر از 0.05 باشد پس فرضیه H_0 را تایید میکنیم ,یعنی اوسط سن مردان و خانم ها یکبرابر است, و فرضیه H_A را رد میکنیم.

اما در صورت که P-value کوچکتر از 0.05 باشد پس فرضیه H_A را تایید میکنیم، یعنی اوسط سن مردان و زنان یکسان نیست, و فرضیه H_0 را رد میکنیم.

3. نوع دیگر ttest دو طرفه شباهت زیاد با هم دارد, در این دو نوع، شما دو گروه یا متغیر را مستقیماً با همدیگر مقایسه میکنید. مثلاً در تحقیقات طبی وقتی یک دانشمند یک دوا را کشف میکند، آنرا در قدم اول روی بعضی حیوانات آزمایشگاهی آزمایش میکند. درین صورت، بخاطر دریافت تاثیر این دوا، دو گروه میباشند که یکی دوا را گرفته و دیگری نگرفته (RCT). یک ttest به این داشمند کمک میکند تا بداند که آیا این دوا تاثیر داشته یا نداشته است، اگر داشته است این تاثیر قابل اعتبار است یا خیر.

اما به هر صورت نتیجه فرمان ttest به شکل بالا میباشد که باز هم نظر به قیمت p-value تصمیم میگیریم که فرضیه H_0 و یا فرضیه H_A را قبول کنیم.

- به یاد داشته باشید که در هنگام راپور باید قیمت های t و df (degree of freedom) را همراه با p-value گزارش دهید.

تمرین

1. در مورد آزمون همبستگی معلومات دهید؟
2. همبستگی مثبت و منفی را تعریف و برای هر یک مثال دهید؟
3. از فرمول همبستگی استفاده کنید و رابطه بین دو متغیر کمی که روزانه در محیط کار یا درسی با آنها روبرو هستید، را دریابید و چگونگی رابطه بین دو متغیر را به به صورت مفصل تشریح کنید؟
4. درمورد رگرسیون خطی تحقیق نمایید و معلومات خود را به صورت مفصل بنویسید؟
5. حالا بعد از مطالعه رگرسیون خطی بگویید که کدام زمان و چگونه از آن استفاده میکنید؟
6. از دیتاست سروی مردم افغانستان در سال 2015 استفاده کنید و یک مدل تحلیلی بسازید و معلومات ذیل را تکمیل کنید؟

الف: فرضیه صفر و فرضیه جاگزین بسازید.

ب: راجع به اندازه نمونه معلومات دهید.

ج: رابطه بین متغیر وابسته و مستقل هر یک را به صورت جداگانه تفسیر دهید با ذکر (Coefficients and P-values)

د: در مورد (Coefficient determinant) معلومات دهید و در این مدل آن را به تفسیر بگیرید.

7. چه زمانی از رگرسیون پروبیت استفاده میکنید؟
8. مدل تحلیلی برای رگرسیون پروبیت انکشاف دهید و معلومات ذیل را بعد از انجام فرمان رگرسیون تکمیل کنید؟

الف: فرضیه صفر و متناوب بسازید.

ب: راجع به اندازه نمونه معلومات دهید.

ج: رابطه بین متغیر وابسته و مستقل هر یک را به صورت جداگانه تفسیر دهید با ذکر (Coefficients and P-values)

د: در مورد (Coefficient determinant) معلومات دهید و در این مدل آن را به تفسیر بگیرید.

فهرست منابع

References

Acock, Alan C. (2014). A Gentle Guide to Stata. Texas: Stata Press Publication

Doughetry C. (2016). Elements of Econometrics. London: University of London

Gujarati, Damodar N. (2004). Basic Econometrics, Fourth Edition. New York: The McGraw-Hill Companies

Hansen, Bruce E. (2015). Econometrics. Wisconsin: John Wiley & Sons

Kerns, Jay G. (2011). Introduction to Probability and Statistics. Vienna: The McGraw-Hill Companies

Reagle, Derrick & Salvatore, Dominick. (2001). Statistics And Econometrics, Second Edition. New York: The McGraw-Hill Companies

Rumsey, Deborah. (2010). Statistics for Dummies. Indiana: Wiley Publishing. Inc.

Spiegel, Murray R. (1999). Theory and Problems of Statistics. New York: The McGraw-Hill Companies

Suseela, Tmt. N. & Sundaram Gnana G. (2005). Statistics for Higher Secondary. Tamilnadu: TEXTBOOK CORPORATION

Wooldridge, Jeffery M. (2012). Introductory Econometrics A Modern Approach, Fifth Edition. Michigan: South-West, Cengage Learning

انگیزه اساسی برای تهیه این کتاب در نخست ادامه مواد درسی برای کورس تحلیل دیتا توسط برنامه ستاتا و ضمیمه برای درک بهتر مفاهیمی که در رهنمای برنامه ستاتا تحت عنوان "تحلیل دیتا توسط برنامه ستاتا" است، می باشد.

بنیاد آسیا از اداره انکشاف بین المللی ایالات متحده امریکا (USAID)، وزارت امور خارجه و تجارت آسترالیا (DFAT)، و انجمن همکاری های بین المللی های آلمان (GIZ) بخاطر حمایت شان از برنامه های ارتقای ظرفیت تحقیقاتی ابراز سپاس و امتنان مینماید.