آمار و کاربرد آن در مدیریت۲

جهت آزمون کارشناسی ارشد در گرایش های رشته مدیریت و حسابداری

شامل آموزش دروس

تستهای آمار رشته های مدیریت از سال۱۳۹۸ الی ۱۴۰۲

مولف : علی زارعی

مقدمه

با توجه به توسعه روز افزون علم و آشنایی با روش های آماری و محاسبات آماری از جمله نیازهای علوم تجربی است. این کتاب بر اساس تجربه ی سال ها تدریس در دروس مختلف جهت سهولت در درس آموزش آمار و کاربرد آن در مدیریت نوشته شده است. این مجموعه شامل تدریس کلیه مطالبی است که در سالهای متفاوت در آزمون های مختلف مطرح شده می باشد و با توضیح ساده هر مطلب درسی به همراه مثال های کاربردی و نکات مهم می باشد. در این اثر سعی شده است تا پاسخ هر مثال به صورت کاملا ساده و روان بیان شود تا کلیه دانشجویانی که نیاز به منبعی در مورد آمارو کاربرد آن در مدیریت دارند بتوانند از مطالب آن استفاده کنند. . خواهشمند است هرگونه پیشنهاد در مورد بهبود این کتاب را به اینجانب با شماره ۰۹۱۲۳۲۱۱۰۷۸ اطلاع دهید.

با تشکر

علی زارعی

فصل چهارم

نمونه گیری

نمونه گیری

روشهای نمونه گیری برآوردگرها : به طور کلی انواع روش های نمونه گیری را می توان در دو بخش نمونه گیری احتمالی و غیر احتمالی جای داد.

۱- نمونه گیری غیر احتمالی

a) در دسترس b) داوطلب c) هدفمند d) سهمی e)بدون نظم(دیمی)

f) شبکه ای (گلوله برفی)

۲- نمونه گیری احتمالی

a)تصادفی ساده b)سیستماتیک (منظم) c) طبقه بندی d)خوشه ای

در نمونه گیری احتمالی هر یک از واحدهای جامعه می توانند با احتمالی مشخص در نمونه قرار گیرند. در حالیکه در نمونه گیری غیر احتمالی انتخاب نمونه براساس قوانین احتمالات صورت نمی گیرد و نمونه به کمک قضاوت انسانی تهیه می شود. بنابراین اشتباهات برآوردهای غیر احتمالی، اغلب غیر تصادفی و غیرقابل اندازه گیری است. در این روش ها هر قدر که حجم نمونه را بزرگ اختیار کنیم، نمونه ها اغلب نمی توانند معرف واقعی جامعه باشند. اما با این حال گاهی اوقات نمونه گیری غیراحتمالی بهترین روش نمونه گیری می باشد. مانند زمانی که امکان تهیه چارچوب نمونه گیری وجود نداشته باشد. یکی از مهم ترین کاربردهای نمونه گیری غیراحتمالی، استفاده در مطالعات مقدماتی پیمایش های بزرگ است. در چنین مواردی که هدف تعیین واریانس صفت مورد اندازه گیری (برای تعین حجم نمونه و روش نمونه گیری) یا تست پایایی ابزار اندازه گیری و... می باشد، احتیاجی به روش های نمونه گیری احتمالی نیست.

۱- نمونه گیری غیر احتمالی

a) در دسترس : این روش ضعیف ترین نوع نمونه گیری غیر احتمالی است. که در آن نمونه ها داوطلبانه و هدفمند هستند مانند گروهی از دانشجویان که در یک درس انتخاب واحد کرده اند این نمونه ها در جمع آوری اطلاعات اکتشافی مفید است ولی امکان برآورد خطا($e = \overline{x} - \mu$) در آن وجود ندارد. و چون معرف کل جامعه نیست روایی بیرونی ندارد

b) داوطلب : این روش نیز غیر احتمالی است چون نمونه ها بطور تصادفی انتخاب نشده اند و امکان دارد جهت دار باشد و برآورد با آنها برآورد صحیحی از جامعه ندهد.

مصاحبه بدون ساختار : این گونه مصاحبه به طور معمول از پژوهش هایی که در رابطه با سرگذشت زندگی افراد باشد بهره گیری می‌شود و کاربرد فراوانی در روش تحقیق جهت تحلیل آماری دارد. در روش مصاحبه بدون ساختار پژوهشگر سعی می کند که نقاط عقیده ها یا شرایط مصاحبه شوندگان را درک کند

مصاحبه ساختار یافته: در مصاحبه ساختار یافته سؤالات از پیش تعیین شده و برنامه‌ریزی شده است. سؤالات به ترتیب مشخصی پرسیده می‌شود .مصاحبه شونده در پاسخگویی به سؤالات کم‌ترین آزادی عمل ممکن را در مقایسه با انواع دیگر مصاحبه دارد. از انعطاف پذیری پایینی برخوردار است .

مصاحبه تاییدی : در مواقعی استفاده می شود که حجم نمونه بسیار زیاد بوده و شواهدی برای بررسی داده هایی که از قبل جمع آوری شده فراهم می کند.

c) قضاوتی(هدفمند) : معمولا در نمونه گیری قضاوتی انتخاب سنجیده واحدها به طریقی صورت می گیرد که هریک معرف بخشی از جامعه مورد نظر باشند. به عنوان مثال فرض کنید قصد داریم به بررسی وضعیت نوجوانانی که از خانه فرار کرده اند بپردازیم و مصاحبه با چنین افرادی ضروری است. از آنجایی که تهیه لیستی از این افراد ممکن نیست، محقق سعی می کند به مکان هایی مانند پارک ها، ترمینال ها و... مراجعه کرده و به طور عمدی افراد مورد نظر را در مکان های خاص شناسایی کند. این روش نمونه گیری با توجه به موضوع مورد بررسی براساس فهم نظری و تجربه پیشین محقق از جمعیت مورد مطالعه است. در نمونه گیری هدفمند ملاک تعیین حجم نمونه اشباع شدن است .

d) سهمیه ای: در این روش جامعه آماری به چند طبقه تقسیم شده سپس به اختیار سهمی به هر طبقه اختصاص داده می شود و در ادامه نمونه هایی را که دسترسی به آن ها ساده تر است به دلخواه انتخاب می کنند. البته ممکن است جامعه مورد مطالعه طبقه بندی نشود، در این حالت نیز آنقدر نمونه اختیار می شود تا حجم نمونه مورد نظر حاصل شود.

برای مثال فرض کنید می خواهیم به بررسی رضایت شغلی کارکنان صنایع بزرگ استان تهران بپردازیم و قبل از مرحله اصلی گردآوری داده ها احتیاج به یک مطالعه مقدماتی برای سنجش واریانس صفت مورد نظر، روایی و پایایی پرسشنامه داریم و از طرفی براساس تحقیقات قبلی می دانیم رضایت شغلی کارکنان صنایع مختلف با یکدیگر تفاوت دارد. با فرض این که حجم نمونه مقدماتی ۵۰ نفر در نظر گرفته شده باشد و با در نظر گرفتن نسبتی از کل کارکنان که در هر صنعت مشغول به کار هستند، سهمیه هر گروه را تعیین می کنیم.

در نمونه گیری سهمیه ای غالبا آن دسته از ویژگی هایی که که در تحقیقات پیشین هم تشخیص داده شده اند مورد مطالعه قرار می گیرند به همین دلیل نمونه گیری سهمیه ای تا حد زیادی به تحقیقات پیشین و یافته های بدست آمده در میدان تحقیق وابسته است. مشکل دیگر این روش عدم توجه به برخی ویژگی های جمعیت مورد نظر است. اما در مجموع با توجه به تشابه ساختار نمونه و جمعیت در نمونه گیری سهمیه ای، احتمال معرف بودن برآوردهای حاصل از آن بیش از برآوردهایی است که از نمونه گیری اتفاقی یا قضاوتی به دست می آید. و خطای آن کمتر است.

e) اتفاقی(بدون نظم یادیمی) : در این روش محقق افرادی را مورد مطالعه قرار می دهد که در دسترس هستند و مصاحبه گر در چارچوب تعداد و حجم نمونه افرادی را به طور اتفاقی انتخاب کرده و با آن ها مصاحبه می کند. به عنوان مثال محققی که می خواهد نگرش مردم را نسبت به یک نمایشگاه بسنجد، از میان افرادی که از نمایشگاه دیدن کرده اند به طور اتفاقی چند نفر را انتخاب می کند. با توجه به موضوع تحقیق مکان هایی که این نمونه گیری می تواند در ان صورت گیرد مانند ایستگاه اتوبوس، بازار، فروشگاه،

بیمارستان و... می باشد.نکته قابل توجه این است که در نمونه گیری قضاوتی محقق افراد خاص را انتخاب می کند اما در نمونه گیری اتفاقی، هرچند محقق در مکان های خاصی قرار می گیرد، اما به طور اتفاقی افرادی را در آن مکان ها انتخاب می کند. نمونه گیری اتفاقی عمدتا با مشاهده مشارکتی و آزمایشی توأم است و از طریق آن محقق همواره موضوعات و محیط هایی را که به آسانی قابل دسترسی باشند، مورد مطالعه قرار می دهد. نمونه گیری به روش اتفاقی راحتر است .

f) شبکه ای (گلوله برفی یا زنجیره ای): این روش در مقایسه با دیگر روش های ذکر شده چندان منظم نیست، کاربرد کمتری دارد و بیشتر در تحقیقات قوم نگاری مورد استفاده قرار می گیرد تا در مطالعات پیمایشی.

این فن شامل شناسایی برخی افراد مهم یک جمعیت و مصاحبه با آن ها است، سپس محقق به پیشنهاد این افراد برای مصاحبه به سراغ افراد دیگر می رود. در این روش هسته کوچک اصلی، با افزایش مرحله ای، رشد می کند و مانند گلوله برفی که با غلتیدن بر روی زمین بزرگ می شود، نمونه تحقیق نیز افزایش می یابد. این روش معمولا زمانی مورد استفاده قرار می گیرد که چارچوب نمونه گیری وجود ندارد و از طرفی افراد نمونه نسبت به یکدیگر شناخت دارند. و میتوان اعضای پنهان را جامعه را پیدا کرد. به عنوان مثال فرض کنید می خواهیم به علل اعتیاد از دیدگاه معتادان بپردازیم. در این صورت با شناسایی تعدادی از معتادان که حاضر به مصاحبه هستند، از آن ها نام نشانی افراد دیگر را که می شناسند جویا شده و به این ترتیب به تعداد مورد نیاز نمونه تهیه کنیم. از این روش هنگامی استفاده می شود که اعضای گروه هدف به نوعی با یکدیگر در ارتباط بوده و دارای ویژگی های مشترک باشند. مزیت این روش در این است که محقق می تواند به بررسی کل شبکه بپردازد اما از طرف دیگر او تنها محدود به کسانی خواهد بود که در این شبکه مرتبط به هم قرار دارند.

g) گروهی : در این روش تنها با گروه های معینی تماس حاصل شده و اطلاعات لازم گرفته می شود. در این روش محقق براساس تجربه شخصی و یا تجارب مشابه دیگران، یک گروه اجتماعی را معرف جامعه ای که به آن تعلق دارند، می یابد و با آن ها مصاحبه می کند. گروه مورد مطالعه می تواند اعضای شورای اسلامی روستا ، ریش سفیدان و... باشند. نتایج این نمونه ها هرچند ممکن است معرف جامعه باشند، اما نمی توان با اطمینان آماری به جامعه مورد مطالعه تعمیم داد.

۲- نمونه گیری احتمالی

a) تصادفی ساده : در این روش نمونه گیری واحدهای مورد انتخاب دارای شانس مساوی $\frac{n}{N}$ برای انتخاب شدن هستند. که در آن **n**تعداد نمونه و **N** تعداد جامعه است در اینجا قوانین احتمال است که معین می کند کدام واحدها یا افراد از جامعه انتخاب شوند. انتخاب یا از طریق قرعه کشی است ویا از طریق استفاده از جدول اعداد تصادفی.در نمونه گیری تصادفی افراد از فهرست جامعه انتخاب می شوند در روش قرعه کشی ابتدا کلیه واحدها یا افراد شماره بندی شده و یا اسامی آنها تهیه می شود و سپس به قید قرعه از بین آنها تعداد لازم برای نمونه انتخاب می شود. برای کنترل واریانس در گمارش تصادفی اثر یک متغیر در بین گروه های مورد بررسی توزیع می شود. این نمونه گیری معمولا به یکی از روشهای زیر انجام می شود

الف) نمونه گیری تصادفی ساده بدون جایگذاری : در این روش نمونه انتخابی کنار گذاشته شده و حذف میشود یک ویژگی مهم نمونه گیری تصادفی ساده بدون جایگذاری این است که احتمال استخراج هر واحد مشخص از جامعه در هر استخراجی مساوی با احتمال استخراج آن واحد مشخص در استخراج اول است .تعداد نمونه انتخابی از رابطه $\binom{N}{n}=\frac{N!}{n!(N-n)!}$ بدست می آید. و احتمال انتخاب از رابطه $p(A)=\frac{1}{\binom{N}{n}}$ بدست می آید

ب) نمونه گیری تصادفی ساده با جایگذاری:اگر در انتخاب n واحد نمونه پس از انتخاب هر واحد آن را به جامعه بر گردانیم و انتخاب بعدی انجام دهیم نمونه گیری تصادفی ساده با جایگذاری می نامند در این روش انتخاب هر واحد مستقل از انتخاب واحدهای دیگر است. همچنین در این طرح نمونه گیری احتمال انتخاب مجدد و دوباره نمونه ها وجود دارد.

b) سیستماتیک (منظم) : نمونه گیری سیستماتیک مشتمل بر گزینش واحدها به روشی سیستماتیک و در نتیجه به صورتی غیر تصادفی است. منظور از این فن نمونه گیری معمولا پخش کردن واحدها به طور یکنواخت بر روی چارچوب است عنصر تصادفی بودن اغلب به این ترتیب دخالت داده می شود که اولین واحد را به طور تصادفی انتخاب می کنند . در این صورت گزینش اولین واحد، بقیه واحدهای نمونه را معین می کنند. یعنی ابتدا $I=\frac{N}{n}$ را بدست می آوریم سپس یک عدد بطور تصادفی کمتر از I را انتخاب میکنیم. سپس به اندازه I به نمونه انتخابی اضافه میکنیم تا تعداد nنمونه انتخاب شود . مثلا از ۱۰۰ نفر بخواهیم نمونه ۵ نفری انتخاب کنیم اول از ۱ تا ۱۰۰ افراد را شماره گذاری میکنیم سپس $I=\frac{N}{n}=\frac{100}{5}=20$ را محاسبه میکنیم آنکاه عددی بین ۰ و ۵ انتخاب میکنیم مثلا ۲ سپس ۲۰ تا ۲۰ تا جلو میرویم یعنی انتخابهای ما شماره های ۲-۲۲-۴۲-۶۲-۸۲ میشود. هر گاه عنصرهای جامعه بطور تصادفی شماره گذاری شوند دقت نمونه گیری سیستماتیک (منظم) با نمونه گیری تصادفی ساده برابر است.

c) طبقه بندی : از این روش نمونه گیری در مواقعی استفاده میشود که ترکیب جامعه همگن نباشد و افراد دارای صفات غیر مشترکی باشند.در این روش نمونه گیری برای اجتناب از اشکالاتی که ممکن است در روش قبلی با آن مواجه شویم، افراد جامعه آماری را بسته به خصوصیاتی که آنها را از یکدیگر متمایز می سازد به طبقات مختلف تقسیم می کنیم. در نمونه گیری تصادفی طبقه ای، واحدهای جامعه مورد مطالعه در طبقه هایی که از نظر صفت متغیر همگن تر هستند، گروه بندی می شوند. به این ترتیب تغییرات در درون گروه ها حداقل می شود. در نمونه گیری طبقه بندی تغییرات داخل طبقات کم و تغییرات بین طبقات زیاد است. معمولا برای طبقه بندی واحدهای جامعه، متغیری به عنوان ملاک در نظر گرفته می شود که با صفت متغیر مورد مطالعه بستگی داشته باشد.سپس به تعداد مورد نیاز و متناسب با جمعیت هر یک از طبقات افراد نمونه را انتخاب می کنیم .انتخاب افراد می تواند هم به روش تصادفی باشد و هم به روش تصادفی سیستماتیک. بنابر این این روش در جمعیتهای نا همگن که توزیع جمعیت در گروهها و طبقات مختلف متفاوت است، استفاده میشود. نمونه گیری تصادفی طبقه ای هنگامی که بخواهیم نمونه

ای نسبتا کوچک از جامعه ای بزرگ که شامل چند زیر گروه عمده باشد انتخاب کنیم که نشانده دهنده هدف نمونه باشد و انتخاب بر اساس اصول احتمال باشد استفاده می شود در این روش اطلاعات از یک طبقه با طبقه دیگر متفاوت است مانند انتخاب دانشجویان یک دانشگاه در این روش واریانس (تغییر پذیری) هر طبقه باید از واریانس کل کمتر باشد در غیر اینصورت دقت کاهش پیدا می کند.

برای مثال به منظور بررسی نسبت قبول شدگان در مدارس تیز هوشان پایه ششم ابتدایی در شهر تهران و رابطه آن با محل جغرافیایی دبستان، می توان دبستان های شهر تهران را برحسب محل آن ها به پنج طبقه تقسیم کرد. طبقه اول دبستان های شمال غربی، طبقه دوم دبستان های شمال شرقی، طبقه سوم دبستان های مرکزی، طبقه چهارم دبستان های جنوب غربی و طبقه پنجم دبستان های جنوب شرقی. پس از آن از هر طبقه تعدادی دبستان به روش تصادفی ساده انتخاب می شود. در نمونه گیری طبقه ای حجم نمونه n را به شیوه های مختلف می توان میان طبقات تقسیم کرد. ساده ترین شیوه تقسیم مساوی تعداد نمونه میان طبقات است. سایر شیوه ها شامل انتساب بهینه و انتساب متناسب می باشند.

d) خوشه ای:نمونه گیری خوشه ای یک نوع نمونه گیری تصادفی است در بسیاری از مواقع می توان با اجرای روش انتخاب تصادفی گروهها یا خوشه هایی از واحدهای نمونه گیری به جای گرفتن نمونه تصادفی ساده ازکل جامعه است. یکی از مزیت های این طرح نمونه گیری این می باشد که در میزان هزینه به طور اساسی صرفه جویی صورت می پذیرد. نمونه گیری خوشه ای ما را از ساختن چارچوب برای تمامی جامعه بی نیاز می کند که این چارچوب خود اغلب یک کار پر خرج و خسته کننده ای است. به علاوه چون واحدهای یک خوشه ,مجاور هم هستند و بنابراین دسترسی به آنها آسان است, فرایند نمونه گیری خوشه ای بطور قابل توجهی به مقرون به صرفه است. و کاربرد آن در مواقعی است که اگر جامعه دارای تجانس (همگن) کامل باشد و توزیع جغرافیایی افراد یا جامعه مورد مطالعه پراکنده باشد ویا فهرست افراد در دسترس نباشد هزینه ها افزایش می یابد که با این روش در هزینه صرفه جویی میشود. در نمونه گیری خوشه ای تغییرات داخل خوشه ها زیاد و تغییرات بین خوشه ها کم است. به عبارتی در نمونه گیری خوشه ای واریانس (تغییر پذیری) درون طبقات زیاد و بین طبقات کم است . برای مثال فرض کنید می خواهیم درآمد هر خانوار را در یک شهر بزرگ برآورد کنیم، اگر از نمونه گیری تصادفی ساده استفاده کنیم نیاز به فهرست تمام خانوارهای شهر داریم. یافتن این چارچوب می تواند بسیار پرهزینه و یا غیرممکن باشد، استفاده از نمونه گیری طبقه ای نیز در چنین جامعه ای مستلزم داشتن فهرستی از خانوارها در هر طبقه می باشد. در این حالت می توانیم شهر را به نواحی معینی مانند بلوک ها (خوشه هایی از اعضا) تقسیم نماییم و یک نمونه تصادفی ساده را از بلوک های این جامعه انتخاب کرده و در آمد هر خانوار را در بلوک هایی که در نمونه واقع شده است اندازه بگیریم. این کار با استفاده از یک چارچوب که کلیه بلوک های شهر را فهرست نموده، به اجرا در می آید. دلیل دیگری که استفاده از نمونه گیری خوشه ای را پیشنهاد می دهد جمع آوری داده ها با هزینه ی کمتر است. برای مثال فرض کنید فهرستی از خانوارهای شهر در دسترس است و می خواهیم نمونه ای به روش تصادفی ساده از خانوارهای مختلف که در سطح شهر پراکنده اند تهیه کنیم. هزینه ی مصاحبه با خانوارها با توجه به هزینه رفت و آمد پرسشگر و دیگر هزینه ها بالا است. یک روش کاهش هزینه های رفت و آمد استفاده از روش نمونه گیری خوشه ای می باشد زیرا افراد درون یک خوشه باید از نظر جغرافیایی به هم نزدیک باشند. در نمونه گیری خوشه ای واحد تحلیل فرد نیست.

فصل پنجم

توزیع های نمونه گیری

برآورد فاصله ای

آزمون فرض

توزیع‌های نمونه‌گیری

برآوردگر: هر تابعی از نمونه که به پارامتر یا پارامترهای جامعه بستگی نداشته باشد برآوردگر یا آماره گویند مقدار عددی برآوردگر را برآورد گویند.

ویژگیهای برآوردگر:

۱- نااریب باشد یعنی $E\left(\hat{\theta}\right)=\theta$ **که** $b=\theta-E\left(\hat{\theta}\right)$ **را مقدار اریبی گویند. که در این صورت**

$$E\left(\hat{\theta}-\theta\right)^{۲}=E\left[\hat{\theta}-E\left(\hat{\theta}\right)\right]^{۲}+\left[\theta-E\left(\hat{\theta}\right)\right]^{۲}\Rightarrow MSE(\theta)=var(\theta)+b^{۲}(\theta)$$

۲- کارایی داشته باشد یعنی دارای کمترین واریانس باشد به عنوان مثال اگر $\theta_۱$ **و** $\theta_۲$ **برآوردگرهای یک آماره باشد و** $var(\theta_۱)\le var(\theta_۲)$ **باشد آنگاه** $\theta_۱$ **از** $\theta_۲$ **کاراتر است به طور کلی** $e=\frac{MSE(\theta_۲)}{MSE(\theta_۱)}$ **و اگر نا اریب باشد** $e=\frac{var(\theta_۲)}{var(\theta_۱)}$

اگر
- $e>۱$ **یعنی** $\theta_۱$ **کاراتر از** $\theta_۲$
- $e=۱$ **یعنی ارجحیت نسبت به یکدیگر ندارند**
- $e<۱$ **یعنی** $\theta_۲$ **کاراتر از** $\theta_۱$

۳- سازگاری: برآورد کنندهٔ $\hat{\theta}$ **را سازگار گویند. اگر با بزرگ شدن اندازهٔ نمونه از لحاظ احتمال به سمت پارامتر** θ **میل کنند یعنی**

$$\lim_{n\to\infty}\left|P_\theta-\hat{P_\theta}\right|=۰$$

مثال: اگر $X_۱,X_۲,X_۳,X_۴$ **یک نمونه ۴ تایی از جامعه‌ای با** $\mu=۲۰$ **و** $\sigma^{۲}=۱۰$ **باشد کدام یک از برآوردگرها مناسب است؟**

$$\bar{x}_۱=\frac{۱}{۴}\sum_{i=۱}^{۴}x_i \qquad \bar{x}_۲=\frac{۱}{۴}\left(x_۱+۲x_۲+x_۳\right)$$

$$E\left(\bar{x}_۱\right)=E\left(\frac{۱}{۴}\sum_{i=۱}^{۴}x_i\right)=\frac{۱}{۴}\sum_{i=۱}^{۴}E\left(x_i\right)=\frac{۴\times۲۰}{۴}=۲۰$$

$$E(\bar{x}_۲)=E\left[\frac{۱}{۴}(x_۱+۲x_۲+x_۳)\right]=\frac{۱}{۴}\left[E(x_۱)+۲E(x_۲)+E(x_۳)\right]=\frac{۴\times۲۰}{۴}=۲۰$$

بنابراین نا اریب هستند

$$var(\bar{x}_۱)=var\left[\frac{۱}{۴}\sum_{i=۱}^{۴}x_i\right]=\frac{۱}{۱۶}\left[\sum_{i=۱}^{۴}var(x_i)\right]=\frac{۴\times۱۰}{۱۶}=۲/۵$$

$$var(\bar{x}_۲)=var\left[\frac{۱}{۴}(x_۱+۲x_۲+x_۳)\right]=\frac{۱}{۱۶}\left[var(x_۱)+۴\,var(x_۲)+var(x_۳)\right]$$

$$=\frac{۱}{۱۶}=\left[۱۰+۴۰+۱۰\right]=۳/۷۵$$

بنابراین $var(\bar{x}_۱)<var(\bar{x}_۲)$ یعنی $\bar{x}_۱$ کاراتر از $\bar{x}_۲$ است.

ارشد مدیریت ۱۴۰۰

۶۰ به‌منظور برآورد پارامتری از جامعه از دو برآوردکننده استفاده شده است. برآوردکننده اول دارای اریبی صفر و انحراف معیار ۲ و برآوردکننده دوم دارای اریبی ۱ و انحراف معیار ۲ می‌باشد. کارایی برآوردکننده دوم به اول چقدر است؟

۱) ۰/۸
۲) ۰/۹
۳) ۱
۴) ۱/۲۵

$$\sigma_{\theta_۱}=۲ \quad b_{\theta_۱}=o \quad \sigma_{\theta_۲}=۲ \quad b_{\theta_۲}=۱$$

$$MSE(\theta_۱)=var(\theta_۱)+b^۲(\theta_۱)=۲^۲+o=۴$$

$$MSE(\theta_۲)=var(\theta_۲)+b^۲(\theta_۲)=۲^۲+۱=۵$$

$$e=\frac{MSE(\theta_۱)}{MSE(\theta_۲)}=\frac{۴}{۵}=۰/۸$$

گزینه ۱ صحیح است

توزیع میانگین نمونه‌ای:

۱- جامعه متناهی باشد یعنی اگر $\frac{n}{N}>۰/۰۵$ باشد جامعه را متناهی گویند که در این صورت نیاز به تصحیح واریانس است و $\frac{N-n}{N-۱}$ که ضریب تصحیح واریانس گویند.

$$E(\bar{x})=\mu \quad , \quad var(\bar{x})=\frac{N-n}{N-۱}.\frac{\sigma^۲}{n}$$

۲- جامعه نامتناهی باشد: اگر $\frac{n}{N}\le۰/۰۵$ باشد جامعه را نامتناهی گویند و ضریب تصحیح واریانس تقریباً

یک است.

$$E(\bar{x})=\mu \quad , \quad var(\bar{x})=\frac{\sigma_x^2}{n}$$

شرایط تقریب خوب به نرمال:

هرگاه توزیع فراوانی متقارن باشد اگر $n\geq ۳۰$ باشد توزیع $\bar{X}$ را تقریباً نرمال در نظر می‌گیریم و در صورتی‌که توزیع فراوانی چوله باشد اگر $n\geq ۱۰۰$ باشد توزیع $\bar{X}$ را تقریباً نرمال در نظر می‌گیریم.

نکته : چون میانگین توزیع نمونه گیری $\overline{X}$ برابر μ است $\overline{X}$ یک برآورد گر بدون سوگیری از μ است

مثال: یک نمونه به اندازه $n=۱۰۰$ به طور تصادفی از جامعه‌ای شامل یک میلیون عنصر با میانگین ۱۰ و واریانس ۳۶ انتخاب می‌کنیم.

الف) احتمال اینکه میانگین نمونه‌ای بزرگتر یا مساوی ۱۱/۲ باشد را بدست آورید.

ب) اگر حجم جامعه ۱۰۰۰ عنصر باشد احتمال آن که میانگین نمونه‌ای بیشتر از ۱۱/۲ باشد را بدست آورید.

الف) $\frac{n}{N}=\frac{۱۰۰}{۱۰۰۰۰۰۰}=۰/۰۰۰۱<۰/۰۵\Rightarrow$ نیاز به تصحیح ندارد $\Rightarrow z=\frac{\bar{x}-\mu}{\frac{\sigma_x}{\sqrt{n}}}$

$$P(\bar{x}\geq ۱۱/۲)=P\left(z\geq \frac{۱۱/۲-۱۰}{\frac{۶}{\sqrt{۱۰۰}}}\right)=P(z>۲)=۰/۰۲۲۸$$

ب) $\frac{n}{N}=\frac{۱۰۰}{۱۰۰۰}=۰/۱>۰/۰۵\Rightarrow$ نیاز به تصحیح دارد $\Rightarrow z=\frac{\bar{x}-\mu}{\sqrt{\frac{N-n}{N-۱}}\cdot\frac{\sigma_x}{\sqrt{n}}}$

$$P(\bar{x}\geq ۱۱/۲)=P\left(z\geq \frac{۱۱/۲-۱۰}{\sqrt{\frac{۱۰۰۰-۱۰۰}{۱۰۰۰-۱}}\times\frac{۶}{\sqrt{۱۰۰}}}\right)=P(Z\geq ۲/۱)=۱-\varphi(۲/۱)=$$

$۱-۰/۹۸۲۱=۰/۰۱۷۹$

مثال : دریک نمونه تصادفی ۴۰۰نفری توزیع نمرات خلاقیت کلامی دانش آموزان دارای توزیع نرمال با میانگین ۱۳۵ و انحراف استاندارد ۱۸ است تقریبا نمره خلاقیت چند دانش آموز کمتر یا مساوی ۱۷۱ است؟

۱) ۱۸۹ ۲) ۲۲۹ ۳) ۳۳۶ ۴) ۳۹۲

$$n=۴۰۰\ ,\ \sigma=۱۸\ ,\ \mu=۱۳۵$$

$$p(x\le ۱۷۱)=p(z\le \frac{۱۷۱-۱۳۵}{۱۸})=p(z\le ۲)=۰/۹۷۷۲$$

$$p=\frac{x}{n}\Rightarrow x=p.n=۰/۹۷۷۲\times ۴۰۰=۳۹۰/۸۸\approx ۳۹۱$$

گزینه۴ صحیح است

ارشد مدیریت ۱۴۰۱

۶۲- میانگین توزیع نمونه‌گیری واریانس نمونه‌های ۳تایی از جامعه‌ای نرمال با ۴ عضو برابر ۱۶ است. انحراف معیار توزیع نمونه‌گیری میانگین نمونه‌های ۳تایی از این جامعه چقدر است؟

۱) $\sqrt{۱۲}$

۲) ۱۲

۳) ۴

۴) ۲

$$var(\bar{x})=\sigma^{۲}_{\bar{x}}=\frac{\sigma^{۲}_{x}}{n}=\frac{۱۶}{۴}=۴\Rightarrow \sigma_{\bar{x}}=\sqrt{۴}=۲$$

گزینه ۴ صحیح است

ارشد مدیریت ۱۴۰۲

۵۶- میانگین توزیع نمونه‌گیری واریانس نمونه‌های تصادفی ۳ تایی از جامعه‌ای نرمال برابر ۲۰ و انحراف‌معیار توزیع نمونه‌گیری میانگین نمونه‌های ۴ تایی از این جامعه ۲ است. تعداد اعضای این جامعه، کدام است؟

۱) ۴

۲) ۵

۳) ۱۶

۴) ۲۵

$$\sigma^{۲}_{\bar{x}}=\frac{\sigma^{۲}_{x}}{n}\Rightarrow n=\frac{\sigma^{۲}_{x}}{\sigma^{۲}_{\bar{x}}}=\frac{۲۰}{۲^{۲}}=۵$$

گزینه ۲ صحیح است

توزیع دو جمله‌ای: اگر x دارای توزیع دو جمله‌ای با پارامترهای P و n باشد در صورتی‌که n بزرگ باشد x دارای توزیع تقریبی نرمال با میانگین $\mu=np$ و واریانس $var(x)=npq$ است.

تقریب نرمال بر توزیع دو جمله‌ای:

شرایط خوب برای استفاده از این تقریب آن است که ۱) $n>۲۰$ ۲) $nP>۵$ و $nq>۵$ که در این صورت باید تصحیح پیوستگی انجام گیرد. چون توزیع دو جمله‌ای گسسته و توزیع نرمال پیوسته است.

$$P(a \le x \le b) = P(a - ۰/۵ < x < b + ۰/۵) P = \left(\frac{a - ۰/۵ - nP}{\sqrt{nPq}} \right) < z < \left(\frac{b + ۰/۵ - nP}{\sqrt{nPq}} \right)$$

مثال: احتمال آن که در ۲۵ بار پرتاب یک سکه بین هفت تا ۱۵ بار شیر ظاهر شود را بدست آورید.

الف) با استفاده از توزیع دو جمله‌ای

ب) با استفاده از تقریب نرمال

$$P(۷ < x < ۱۵) = P = \left(\frac{۷ - ۰/۵ - ۱۲/۵}{\sqrt{۲۵ \times ۰/۵ \times ۰/۵}} \right) < z < \left(\frac{۱۵ + ۰/۵ - ۱۲/۵}{\sqrt{۲۵ \times ۰/۵ \times ۰/۵}} \right)$$

$$= P(-۲/۴ < z < ۱/۲) = \varphi(۱/۲) - \varphi(-۲/۴) = ۰/۸۷۶۷$$

توزیع نسبت: اگر $\bar{P} = \frac{x}{n}$ را به عنوان تبدیل خطی در نظر بگیریم هر n به اندازه کافی بزرگ باشد در این صورت $E(x) = \mu = P$ و $var(x) = \frac{\bar{P}\bar{q}}{n}$

مثال: اگر نسبت موفقیت در جامعه‌ای ۰/۸ باشد و از این جامعه یک نمونه ۱۰۰ تایی انتخاب کنیم احتمال اینکه نسبت موفقیتها بیش از ۰/۷۴ باشد را بدست آورید.

$$P(\bar{P} \ge ۰/۷۴) = P\left(z > \frac{\bar{P} - P}{\sqrt{\frac{Pq}{n}}} \right) = P\left(z > \frac{۰/۷۴ - ۰/۸}{\sqrt{\frac{۰/۸ \times ۰/۲}{۱۰۰}}} \right) = P(z > ۱/۵) = ۰/۰۶۶۸$$

توزیع اختلاف میانگینها:

$$E(\overline{x_۱} - \overline{x_۲}) = \mu_۱ - \mu_۲ \quad , \quad \sigma^۲_{\overline{x_۱} - \overline{x_۲}} = \sigma^۲_{\overline{x_۱}} + \sigma^۲_{\overline{x_۱}} - ۲\,cov(\overline{x_۱} - \overline{x_۲})$$

$$cov(\overline{x_۱} - \overline{x_۲}) = ۰ \Rightarrow \sigma^۲_{\overline{x_۱} - \overline{x_۲}} = \sigma^۲_{\overline{x_۱}} + \sigma^۲_{\overline{x_۱}} \Rightarrow \sigma_{\overline{x_۱} - \overline{x_۲}} = \sqrt{\sigma^۲_{\overline{x_۱}} + \sigma^۲_{\overline{x_۱}}} \Rightarrow \sigma_{\overline{x_۱} - \overline{x_۲}} = \sqrt{\frac{\sigma^۲_۱}{n_۱} + \frac{\sigma^۲_۲}{n_۲}}$$

توزیع نمونه‌ای اختلاف بین میانگینهای دو جامعه نرمال مستقل وقتی واریانس جامعه معلوم باشد

در این صورت کوواریانس بین دو جامعه صفر است بنابراین:

$$E = (\bar{x}_۱ - \bar{x}_۲) = \mu_۱ - \mu_۲$$

$$\text{var}\left(\bar{x}_1-\bar{x}_2\right)=\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\Rightarrow z=\frac{\left(\bar{x}_1-\bar{x}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

مثال: از دو جامعه آماری A و B با مشخصات $\mu_1=100$ و $\mu_2=105$ و $\sigma_1^2=\sigma_2^2=200$ و نمونه‌های $n_1=40$ و $n_2=50$ گرفته‌ایم احتمال آنکه اختلاف میانگین کمتر از یک باشد را بدست آورید.

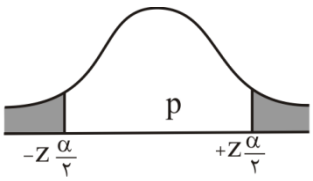
$$P\left(|\bar{x}_2-\bar{x}_1|<1\right)=P\left(-1<\bar{x}_2-\bar{x}_1<+1\right)=p\left(\frac{(-1)-(105-100)}{\sqrt{\frac{200}{40}+\frac{200}{50}}}\langle z\langle\frac{(+1)-(105-100)}{\sqrt{\frac{200}{40}+\frac{200}{50}}}\right)$$

$$=p(-2\langle z\langle-1/33)=\varphi(-2)-(\varphi-1/33)=1-\varphi(2)-(1-(\varphi(1/33))=\varphi(2)-(\varphi 1/33)=$$

$$=0/9772-0/9082=0/069$$

برآورد فاصله‌ای میانگین واقعی جامعه وقتی واریانس جامعه معلوم باشد.

در این صورت از توزیع نرمال (z) استفاده می‌کنیم که در سطح اطمینان P می‌توان به صورت زیر نوشت (α : سطح خطا و $\alpha+P=1$)



$$\pm Z_{\frac{\alpha}{2}}=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\Rightarrow \mu:\bar{x}\pm Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

نکته۱ : $L=2Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ **طول فاصله اطمینان و** $d=|\bar{x}-\mu|=Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ **خطای برآورد**

نکته۲: رابطه انحراف معیار و دامنه تغییرات $R=X_{max}-X_{min}$ $\quad\sigma_x\approx\frac{R}{4}\quad\&\quad\sigma_x\approx\frac{R}{6}$

نکته۳: روابط حدود بالا و پایین

$$LCL=\bar{x}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\quad,UCL=\bar{x}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\Leftarrow,\mu=\frac{UCL+LCL}{2}\quad,d=\frac{UCL-LCL}{2}$$

مثال: از جامعه با انحراف استاندارد ۱۰ یک نمونه ۲۵ گرفته‌ایم که میانگین نمونه ۷۸ بدست آمده است یک برآورد فاصله‌ای ۹۵٪ برای میانگین واقعی جامعه بنویسید.

$$\alpha = ۱ - P = ۱ - ۰/۹۵ = ۰/۰۵ \quad , \quad \frac{\alpha}{۲} = \frac{۰/۰۵}{۲} = ۰/۰۲۵ \quad , \quad Z_{۰/۰۲۵} = ۱/۹۶$$

$$\mu : \bar{x} \pm Z_{\frac{\alpha}{۲}} \frac{\sigma}{\sqrt{n}} = ۷۸ \pm ۱/۹۶ < \frac{۱۰}{\sqrt{۲۵}} \Rightarrow ۷۴/۰۸ < \mu < ۸۱/۹۲$$

حفظ شود $Z_{۰/۰۵} = ۱/۶۴۵ \quad Z_{۰/۰۲۵} = ۱/۹۶ \approx ۲$

ارشد مدیریت۱۴۰۰

۵۹- فرض کنید توزیع ضخامت پاک‌کن‌های تولیدی یک دستگاه به‌صورت $X \sim N(۱۱,۲)$ است. ضخامت ۹۰ درصد پاک‌کن‌های تولیدی در چه دامنه‌ای در دو طرف میانگین قرار می‌گیرد؟

(در توزیع نرمال استاندارد: $P(Z \geq ۱/۲۸) = ۰/۱$, $P(Z \leq -۱/۶۴) = ۰/۰۵$)

۱) $[۷/۷۲ , ۱۴/۲۸]$

۲) $[۸/۴۴ , ۱۳/۵۶]$

۳) $[۹/۳۶ , ۱۲/۶۴]$

۴) $[۹/۷۲ , ۱۲/۲۸]$

$$\mu = ۱۱ \quad \sigma = ۲ \quad p = ۹۰\%$$

$$\alpha = ۱ - P = ۱ - ۰/۹ = ۰/۱ \quad , \quad \frac{\alpha}{۲} = \frac{۰/۱}{۲} = ۰/۰۵ \quad , \quad Z_{۰/۰۵} = ۱/۶۴۵$$

$$\mu : \bar{x} \pm Z_{\frac{\alpha}{۲}} \sigma = ۱۱ \pm ۱/۶۴۵ \times ۲ \Rightarrow \mu : ۱۱ \pm ۳/۲۸ \Rightarrow ۷/۲۸ < \mu < ۱۴/۲۸$$

گزینه ۱ صحیح است

برآورد فاصله‌ای میانگین واقعی واریانس جامعه نامعلوم باشد در این صورت با استفاده از واریانس نمونه و با توجه به حجم نمونه از توزیع z یا t استفاده می‌شود.

الف) برآورد فاصله‌ای میانگین واقعی جامعه وقتی واریانس جامعه نامعلوم باشد و حجم نمونه بیش‌تراز ۳۰ باشد در این صورت طبق قضیه حد مرکزی از توزیع z استفاده می‌شود.

$$\mu : \bar{x} \pm Z_{\frac{\alpha}{۲}} \frac{s}{\sqrt{n}}$$

ب) حجم نمونه کمتر از ۳۰ باشد در این صورت به جای توزیع z از توزیع t با درجه آزادی $\nu = df = n - ۱$ **) استفاده می‌شود.**

$$\mu : \bar{x} \pm t_{\frac{\alpha}{۲}, df} . \frac{s}{\sqrt{n}}$$

مثال: در یک نمونه ۱۶ تایی میانگین نمونه ۶۵ و واریانس نمونه ۸۱ بدست آمده است یک برآورد فاصله‌ای ۹۵٪ برابری میانگین واقعی جامعه بنویسید.

$$n=۱۶<۳۰\rightarrow t$$

$$s^۲=۸۱\Rightarrow s=۹ \qquad \bar{x}=۶۵ \qquad n=۱۶$$

$$\mu:\bar{x}\pm t_{\frac{\alpha}{۲},df}.\frac{s}{\sqrt{n}}=۶۵\pm t_{.۰۵,۱۵۰}.\frac{۹}{\sqrt{۱۶}}=۶۵\pm ۱/۷۵۳\times\frac{۹}{۴}\Rightarrow ۶۱/۰۵۵<\mu<۶۸/۹۴۴$$

ارشد مدیریت۹۸

۴۷- نمونه تصادفی از قیمت کالایی در ۴۹ فروشگاه دارای میانگین ۱۵۰ و انحراف معیار ۱۴ واحد پول است. میانگین قیمت این کالا با ضریب اطمینان ۰/۹۵ در کدام بازه است؟ $(S_{-\infty}^{-۱/۹۶}=۰/۰۲۵)$

۱) (۱۴۶/۰۸,۱۵۳/۹۲) ۲) (۱۴۸/۰۸,۱۵۱/۹۲)

۳) (۱۴۹/۰۲,۱۵۰/۹۸) ۴) (۱۴۲/۲۶,۱۵۷/۸۴)

چون حجم نمونه بیشتر از ۳۰ است از توزیع نرمال استفاده می کنیم

$$\int_{-\infty}^{-۱/۹۶} f(z)=۰/۰۲۵\Rightarrow \quad Z_{./۰۲۵}=۱/۹۶$$

$$\mu:\bar{x}\pm Z_{\frac{\alpha}{۲}}\frac{S}{\sqrt{n}}=۱۵۰\pm ۱/۹۶\times\frac{۱۴}{\sqrt{۴۹}}\Rightarrow \mu:۱۵۰\pm ۳/۹۲\Rightarrow ۱۴۶/۰۸<\mu<۱۵۳/۹۲$$

گزینه ۱ صحیح است

برآورد فاصله‌ای یک‌طرفه

کافی است به جای $\frac{\alpha}{۲}$ از α استفاده کنیم

یک‌طرفه بالا $\mu:\bar{x}+z_{\alpha}\frac{\sigma}{\sqrt{n}}$

یک‌طرفه پائین $\mu:\bar{x}-z_{\alpha}\frac{\sigma}{\sqrt{n}}$

ارشد مدیریت ۱۴۰۰

۶۲- از یک نمونهٔ تصادفی ۱۶ تایی از دستگاه پرکننده شیشه‌های نوشابه، میانگین و واریانس به‌ترتیب ۲۵۰ و ۱۰۰ میلی لیتری حاصل شده است. با توجه به کارکرد غیر نرمال دستگاه، در سطح اطمینان ۹۶٪ میانگین مقدار نوشابه در شیشه‌های پر شده توسط دستگاه چقدر برآورد می‌شود؟

۱) $[۲۳۵/۸\ ,\ ۲۶۴/۲]$

۲) $[۲۴۵,\ ۲۶۵]$

۳) $[۱۵۰\ ,\ ۳۵۰]$

۴) $[۲۳۷/۵\ ,\ ۲۶۲/۵]$

چون گفته شده غیر نرمال از قضیه چپیشف حل می کنیم

$$P\left(|x-\mu|<k\sigma\right)\geq 1-\frac{1}{k^2}\Rightarrow P\left(|\bar{x}-\mu|<k\frac{s}{\sqrt{n}}\right)\geq 1-\frac{1}{k^2}$$

$$1-\frac{1}{k^2}=0/96\Rightarrow\frac{1}{k^2}=1-0/96=0/04\Rightarrow k^2=\frac{1}{0/04}=\frac{100}{4}=25\Rightarrow k=5$$

$$\mu:\bar{x}\pm k\frac{s}{\sqrt{n}}\Rightarrow\mu:250\pm5\times\frac{10}{\sqrt{16}}=250\pm12/5\Rightarrow 237/5\leq\mu\leq 262/5$$

برآورد فاصله‌ای نسبتها:

برای نسبتها اغلب از توزیع z استفاده می‌شود.

$$\bar{P}=\frac{x}{n}\quad,\quad \bar{q}=1-\bar{P}\quad,\sigma_{\bar{p}}=\sqrt{\frac{\bar{P}\bar{q}}{n}}$$

$$P:\bar{P}\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{P}\bar{q}}{n}}$$

مثال: از بین ۱۰۰ کالا ۲۰ کالا معیوب یک برآورد فاصله‌ای در سطح خطای ۵٪ برای نسبت کالاهای معیوب بنویسید

$$\bar{P}=\frac{x}{n}=\frac{20}{100}=0/2\qquad q=1-\bar{P}=1-0/2=0/8$$

$$P:\bar{P}\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{P}\bar{q}}{n}}=0/2\pm1/96\sqrt{\frac{0/2\times0/8}{100}}\Rightarrow 0/1216<P<0/2784$$

برآورد فاصله‌ای واریانس و انحراف معیار:

همانگونه که می‌دانیم $\frac{(n-1)s^2}{\sigma^2}$ **دارای توزیع خی دو با درجهٔ آزادی** $\nu=df=n-1$ **است**

$$\frac{(n-1)s^2}{\sigma^2}\sim\chi^2$$

$$\sigma_u^2=\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},df}}\quad,\quad \sigma_L^2=\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},df}}\quad,\quad \sigma_L^2<\sigma^2<\sigma_u^2$$

$$\sigma_u^2=s\sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2},df}}}\quad,\quad \sigma_L=s\sqrt{\frac{n-1}{\chi^2_{\frac{\alpha}{2},df}}}\quad,\quad \sigma_L<\sigma<\sigma_u$$

مثال: در یک نمونه ۱۰ تایی $s^2 = ۸۰$ بدست آمده است یک برآورد فاصله‌ای ۹۰٪ برای واریانس واقعی جامعه بنویسید.

$$df = n - ۱ = ۱۰ - ۱ = ۹$$

$$\alpha = ۱ - ۰/۹ = ۰/۱$$

$$\chi^2_{۰/۰۵,۹} = ۱۶/۹۱۹ \qquad \chi^2_{۰/۹۵,۹} = ۳/۳۲۵$$

$$\frac{\alpha}{۲} = \frac{۰/۱}{۲} = ۰/۰۵ \Rightarrow \chi^2_{۰/۰۵,۹} = ۱۶/۹۱۹ \Rightarrow \sigma^2_L = \frac{(n-۱)s^2}{\chi^2_{\frac{\alpha}{۲},df}} = \frac{۹ \times ۸۰}{۱۶/۹۱۹} = ۴۲/۵۵۵$$

$$۱ - \frac{\alpha}{۲} = ۱ - ۰/۰۵ = ۰/۹۵ \Rightarrow \chi^2_{۰/۹۵,۹} = ۳/۳۲۵ \Rightarrow \sigma^2_u = \frac{(n-۱)s^2}{\chi^2_{۱-\frac{\alpha}{۲},df}} = \frac{۹ \times ۸۰}{۳/۳۲۵} = ۲۱۶/۵۴$$

برآورد فاصله‌ای اختلاف میانگین دو جامعه نرمال مستقل وقتی واریانس جوامع معلوم باشد در این صورت بدون توجه به حجم نمونه از توزیع z استفاده می‌شود.

$$\mu_۲ - \mu_۱ : \bar{x}_۲ - \bar{x}_۱ \pm z_{\frac{\alpha}{۲}} \sqrt{\frac{\sigma^2_۱}{n_۱} + \frac{\sigma^2_۲}{n_۲}}$$

و اگر عدد صفر در بازهٔ بدست آمده قرار گیرد یعنی اختلاف بر اثر عامل تصادفی است

مثال: از دو جامعه نرمال مستقل با واریانس $\sigma^2_۱ = ۶۰$ و $\sigma^2_۲ = ۴۵$ نمونه‌های $n_۱ = ۱۰$ و $n_۲ = ۱۰$ گرفته‌ایم که $\bar{x}_۱ = ۴۰$ و $\bar{x}_۲ = ۴۵$ بدست آمده یک برآورد فاصله‌ای ۹۵٪ برای اختلاف میانگین دو جامعه بنویسید و توضیح دهدی اختلاف بر اثر عامل تصادفی است.

$$\mu_۲ - \mu_۱ : \bar{x}_۲ - \bar{x}_۱ \pm z_{\frac{\alpha}{۲}} \sqrt{\frac{\sigma^2_۱}{n_۱} + \frac{\sigma^2_۲}{n}}$$

$$\mu_۲ - \mu_۱ : ۴۵ - ۴۰ \pm ۱/۹۶ \sqrt{\frac{۶۰}{۱۰} + \frac{۴۵}{۱۵}} = ۵ \pm ۵/۸۸ \Rightarrow -۰/۸۸ < \mu_۲ - \mu_۱ < ۱۰/۸۸$$

چون عدد صفر در بازه قرار می‌گیرد بنابراین اختلاف بر اثر عامل تصادفی است.

برآورد فاصله اختلاف میانگین دو جامعه نرمال مستقل وقتی واریانس جوامع نامعلوم باشد:

با فرض برابری واریانسها دو حالت امکان دارد رخ دهد

الف)حجم نمونه‌ها بیشتر از ۳۰ باشد: در این صورت طبق قضیه حد مرکزی می‌توان از توزیع z با استفاده از واریانس نمونه‌ها برآورد فاصله‌ای را نوشت

$$\mu_۲ - \mu_۱ : \bar{x}_۲ - \bar{x}_۱ \pm z_{\frac{\alpha}{۲}} \sqrt{\frac{s^2_۱}{n_۱} + \frac{s^2_۲}{n_۲}}$$

ب)حجم نمونه‌ها کمتر از ۳۰ باشد: در این از توزیع t و درجهٔ آزادی $v = df = n_1 + n_2 - 2$ استفاده می‌شود

$$df = n_1 + n_2 - 2$$

$$s_p^2 = \frac{(n_1 -)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad , \quad sp = \sqrt{s_p^2}$$

$$\mu_2 - \mu_1 : \bar{x}_2 - \bar{x}_1 \pm t_{\frac{\alpha}{2},df}.sp.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

مثال: یک برآورد فاصله‌ای ۹۰٪ برای اختلاف میانگین دو جامعه A و B که نمونه‌های زیر از آنها بدست آمده است بنویسید.

A	B
$\bar{x}_1 = 100$	$\bar{x}_2 = 95$
$s_1^2 = 90$	$s_2^2 = 100$
$n_1 = 8$	$n_2 = 14$

$$df = n_1 + n_2 - 2 = 8 + 14 - 2 = 20$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{7 \times 90 + 13 \times 100}{20} = 96/5$$

$$s_p = \sqrt{s_p^2} = \sqrt{96/5} = 9/82$$

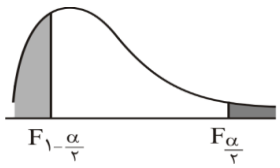
$$\mu_2 - \mu_1 : \bar{x}_2 - \bar{x}_1 \pm t_{\frac{\alpha}{2},df}.sp.\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \Rightarrow$$

$$\mu_2 - \mu_1 : 100 - 95 \pm 1/725 \times 9/82\sqrt{\frac{1}{8} + \frac{1}{14}} = 5 \pm 3/279 \Rightarrow 1/72 < \mu_2 - \mu_1 < 8/279$$

بنابراین اختلاف بین میانگین دو جامعه واقعی است.

برآورد فاصله‌ای نسبت دو واریانس:

همان‌گونه که می‌دانیم نسبت دو واریانس دارای توزیع F یا فیشر است که چوله به راست است با درجات آزادی $v_1 = df_1 = n_1 - 1 \quad , v_2 = df_2 = n_2 - 1$



$$v_1 = n_1 - 1 \quad , v_2 = n_2 - 1 \quad , F = \frac{\sigma_2^2}{\sigma_2^2}.\frac{s_1^2}{s_2^2} \Rightarrow P\left[F_{1-\frac{\alpha}{2},v_1,v_2} < F < F_{\frac{\alpha}{2},v_1,v_2}\right] = 1 - \alpha$$

$$\Rightarrow \frac{1}{F_{\frac{\alpha}{2},v_1,v_2}}\frac{s_1^2}{s_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}.\frac{1}{F_{1-\frac{\alpha}{2},v_1,v_2}}$$

می‌دانیم $F_{1-\frac{\alpha}{2},\nu_1,\nu_2}=\frac{1}{F_{\frac{\alpha}{2},\nu_2,\nu_1}}$ بنابراین: $\frac{1}{F_{\frac{\alpha}{2},\nu_1,\nu_2}}.\frac{s_1^2}{s_2^2}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{s_1^2}{s_2^2}.F_{\frac{\alpha}{2},\nu_2,\nu_1}$

مثال :یک فاصله اطمینان ۹۰٪ برای نسبت واریانس جامعه A و B بدست آورید و توضیح دهید آیا واریانس دو جامعه با احتمال ۹۰٪ برابر است

A	B
$n_1=5$	$n_2=7$
$s_1^2=25$	$s_2^2=13/5$
$\bar{x}_1=13$	$\bar{x}_2=15/5$

$$\frac{1}{F_{\frac{\alpha}{2},\nu_1,\nu_2}}.\frac{s_1^2}{s_2^2}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{s_1^2}{s_2^2}.F_{\frac{\alpha}{2},\nu_1,\nu_2}$$

$$\frac{1}{F_{.05,4,6}}.\frac{25}{13/3}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{25}{13/3}.F_{\frac{\alpha}{2},.05,6,4}$$

$$\frac{1}{4/5337}\times\frac{25}{13/5}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{25}{13/5}.6/1631$$

$$0/4084<\frac{\sigma_1^2}{\sigma_2^2}<11/4098$$

چون عدد یک در بازهٔ بدست آمده قرار می‌گیرد بنابراین با احتمال ۹۰٪ و واریانس می‌تواند برابر باشد.

برآورد فاصله‌ای اختلاف دو نسبت:

در صورتی‌که از n_1 عضو جامعه اول x_1 عضو دارای خاصیتی و از n_2 عضو جامعه دوم x_2 عضو دارای آن خاصیت باشد در این به دو صورت می‌توان برآورد فاصله‌ای برای اختلاف نسبتی نوشت

روش اول

$$\bar{P}_1=\frac{x_1}{n_1}\quad,\quad \bar{q}_1=1-\bar{P}_1\quad,\quad \bar{P}_2=\frac{x_2}{n_2}\quad,\quad \bar{q}_2=1-\bar{P}_2$$

$$P_1-P_2:\bar{P}-\bar{P}_2\pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\bar{P}_1(1-\bar{P}_1)}{n_1}+\frac{\bar{P}_2(1-\bar{P}_2)}{n_2}}$$

روش دوم

$$P=\frac{x_1+x_2}{n_1+n_2}\qquad q=1-P$$

$$P_1-P_2:\bar{P}-\bar{P}_2\pm Z_{\frac{\alpha}{2}}\sqrt{Pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

مثال: کالایی توسط دو دستگاه A و B تولید می‌شود که از ۱۰۰ کالای دستگاه A ۱۵ کالا معیوب و ۶۰ کالای دستگاه B، ۶ کالا معیوب است یک برآورد فاصله‌ای ۸۵٪ برای اختلاف نسبت کالاهای معیوب توسط دو دستگاه A و B بنویسید

$$\bar{P}_A = \frac{x_A}{n_A} = \frac{۱۵}{۱۰۰} = ۰/۱۵ \quad , \quad \bar{P}_B = \frac{x_B}{n_B} = \frac{۶}{۶۰} = ۰/۱$$

$$P_A - P_B : \bar{P}_A - \bar{P}_B \pm Z_{\frac{\alpha}{۲}} \sqrt{\frac{\bar{P}_A(۱-\bar{P}_A)}{n_A} + \frac{\bar{P}_B(۱-\bar{P}_B)}{n_B}}$$

$$P_A - P_B : ۰/۱۵ - ۰/۱ \pm ۱/۹۶ \sqrt{\frac{۰/۱۵ \times ۰/۸۵}{۱۰۰} + \frac{۰/۱ \times ۰/۹}{۶۰}}$$

$$۰/۰۵ - ۰/۱ < P_A - P_B < ۰/۰۵ + ۰/۱ \Rightarrow -۰/۰۵ < P_A - P_B < ۰/۱۵$$

بنابراین اختلاف نسبت کالاهای معیوب دو دستگاه با احتمال ۹۵٪ بر اثر عامل تصادفی است

آزمون فرض‌های آماری:

هرگاه ادعایی مطرح شود از آزمون فرض‌های آماری استفاده می‌شود آنچه ادعا شده اغلب H_1 در نظر گرفته می‌شود و نقیض آن در H_0 قرار می‌گیرد که در نتیجه تصمیم گرفته شده و واقعیت امکان دارد دو نوع خطا رخ دهد

۱- خطای نوع اول (α): احتمال رد H_0 در صورتی‌که H_0 درست باشد و با حرف α نشان می‌دهیم

$\alpha = P(H_0 \mid H_0 \text{ درست باشد})$

۲- خطای نوع دوم (β): احتمال پذیرش H_0 در صورتی‌که H_0 نادرست باشد و با حرف β نشان می‌دهیم

$\beta = P(H_0 \text{ پذیرش} \mid H_0 \text{ نادرست باشد})$

۳- توان آزمون (π): احتمال رد H_0 در صورتی‌که H_0 نادرست باشد و با حروف π نشان می‌دهیم. توان آزمون متمم خطای نوع دوم است $\pi = 1-\beta$ و

$\pi = P(H_0 \mid H_0 \text{ نادرست باشد})$

نکته۱: با افزایش حجم نمونه خطای نوع اول و دوم کاهش می یابد

نکته۲: مجموع خطای نوع اول و دوم ثابت است یعنی با افزایش خطای نوع اول خطای نوع دوم کاهش می یابد و بالعکس

نکته۳: در ساخت فرضیه علامت مساوی همیشه در H_0 قرار دارد یعنی کلمات حداقل- حداکثر و برابر

نتایج ممکن آزمون فرض

تصمیم / وضعیت	H_0 پذیرفته می شود	H_0 رد می شود
اگر H_0 درست باشد	تصمیم درست سطح اطمینان $=1-\alpha$ =احتمال	خطای نوع اول سطح آزمون $=\alpha$ =احتمال
اگر H_0 نادرست باشد	خطای نوع دوم β =احتمال	تصمیم درست توان آزمون $\pi = 1-\beta$ =احتمال

توجه ۱: هدف از محاسبه آماره ها بر اساس داده های حاصل ازنمونه توصیف نمونه وآزمون فرض پیرامون ویژگی های جامعه است.

توجه ۲: اگر توان آزمون بالا برود خطای نوع اول افزایش می یابد چون توان آزمون از رابطه $1-\beta$ = بدست می آید بنابر این (توان آزمون) $\beta = 1-$ در نتیجه اگر توان آزمون بالا برود خطای نوع دوم کاهش می یابد و چون در یک آزمون مجموع خطای نوع اول و دوم ثابت است با کاهش خطای نوع دوم خطای نوع اول افزایش می یابد

مراحل انجام آزمون فرض:

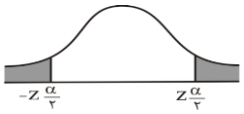
۱-نوشتن فرض آماری **۲-تعیین آماره آزمون و محاسبه مقدار آن**

۳-تعیین مقدار بحرانی از روی جدول **۴-تصمیم گیری**

آزمون فرض میانگین با مقدار ثابت از جامعه نرمال و واریانس جامعه معلوم باشد.

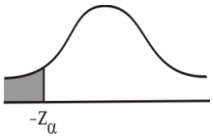
در این صورت از توزیع z استفاده می‌شود و آمارهٔ آزمون از رابطه $Z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ **بدست می‌آید.**

آزمون دو دامنه $\begin{cases}H_0:\mu=\mu_0\\H_1:\mu\neq\mu_0\end{cases}$ $z>z_{\frac{\alpha}{2}}$ **یا** $z<-z_{\frac{\alpha}{2}}\Rightarrow$ **رد می‌شود** H_0

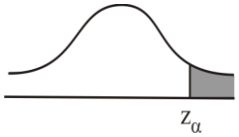


یا $|z|>z_{\frac{\alpha}{2}}\Rightarrow$ **رد می‌شود** H_0

آزمون یک طرفه چپ $\begin{cases}H_0:\mu\geq\mu_0\\H_1:\mu<\mu_0\end{cases}$ $z<-z_\alpha\Rightarrow$ **رد می‌شود** H_0



آزمون یک طرفه راست $\begin{cases}H_0:\mu\leq\mu_0\\H_1:\mu>\mu_0\end{cases}$ $z>z_\alpha\Rightarrow$ **رد می‌شود** H_0



نکته: هنگام نوشتن فرضیه در صورتی که پیشنه کافی نداشته آزمون دودامنه می نویسیم

مثال: از یک جامعه نرمال با واریانس ۶۴ و میانگین μ **یک نمونه ۲۵ تایی گرفته‌ایم که میانیگن نمونه** $\bar{x}=۸۰$ **بدست آمده است در سطح خطای ۵٪ آزمون کنید میانگین جامعه بیشتر از ۸۲ است.**

$$\begin{cases}H_0:\mu\leq ۸۲\\H_1:\mu>۸۲\end{cases}\qquad z=\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}=\frac{۸۰-۸۲}{\frac{۸}{\sqrt{۲۵}}}=-۱/۲۵$$

چون $z=-۱/۲۵\prec z_{./۰۵}=۱/۶۴۵$ **بنابراین** H_0 **پذیرفته می‌شود و با احتمال ۹۵٪ حداکثر میانگین جامعه ۸۲ است**

نکته :در آزمونهای یک طرفه به جای آلفا از نصف آلفا استفاده می شود

P _ مقدار (P-VALUE) : نشانده احتمالی است که ما مشاهده خواهیم کرد اگر فرض صفر صحیح باشد.در هنگام استفاده از P-مقدار اگر P-مقدار کمتر از آلفا تعیین شده باشد فرض صفر رد میشود

نکته : در استفاده از فاصله اطمینان برای بررسی معنی داری آزمون ها ی برابری میانگین دو جامعه اگر عدد صفر در فاصله اطمینان ۹۵٪ قرار گیرد فرض تایید می شود و $p\succ ۰/۰۵$ **است**

محاسبه حجم نمونه : $n=\frac{(Z_{\frac{\alpha}{2}})^2\times\sigma^2}{d^2}$ $\Leftrightarrow$ $n=\left(\frac{(Z_{\frac{\alpha}{2}})\times\sigma}{d}\right)^2$ $d=Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\Rightarrow$ **درحالتی که**

$$n = \frac{(Z_{\frac{\alpha}{2}} + Z_{1-\beta})^2 * \sigma^2}{d^2}$$

خطای نوع دوم در نظر گرفته شود

مثال : در صورتیکه خطای نوع اول و دوم هر دو برابر ۰/۵ باشد توان آزمون را بدست آورید

۱) ۰/۵ ۲) ۰/۹۵ ۳) ۰/۲۵ ۴) ۰/۰۵

گزینه۱ صحیح است چون $\text{توان آزمون} = 1-\beta = 1-0/5 = 0/5$

ارشد مدیریت۹۹

۵۵- در یک جامعه آماری با واریانس ۱ ، اگر هدف برآورد میانگین جامعه آماری باشد، با دقت برآورد ۰٫۱۰، حجم نمونه تحقیق در سطح خطای ۵٪ چقدر است؟

$$\left(\int_{-\infty}^{-1/96} f(z)\,dz = 0/025\right)$$

۱) ۳۸۴
۲) ۱۹۶
۳) ۹۵
۴) ۴۹

$$n = \left(\frac{(Z_{\frac{\alpha}{2}})\times\sigma}{d}\right)^2 \Rightarrow n = \left(\frac{1/96\times 1}{0/1}\right)^2 = 384$$

گزینه ۱ صحیح است

ارشد مدیریت ۱۴۰۰

۶۱- فاصله اطمینان برای میانگین وزن محصولات یک شرکت با احتمال ۹۵٪ براساس نمونه‌های تصادفی، (۴۹٫۸ , ۳۰٫۲) محاسبه شده است. اگر توزیع وزن محصولات این شرکت نرمال باشد و انحراف معیار وزن محصولات این شرکت ۳۰ باشد، حجم نمونه تصادفی این تخمین چقدر بوده است؟

(راهنمایی: $[(z \mid \alpha = 0/025) = 1/96 \; ; \; (z \mid \alpha = 0/05) = 1/64]$)

۱) ۲۵
۲) ۳۶
۳) ۶۴
۴) ۱۴۴

$$UCL = 49/8 \quad LCL = 30/2 \quad \sigma = 30$$

$$d = \frac{UCL - LCL}{2} = \frac{49/8 - 30/2}{2} = 9/8$$

$$n = \left(\frac{(Z_{\frac{\alpha}{2}})\times\sigma}{d}\right)^2 \Rightarrow n = \left(\frac{1/96\times 30}{9/8}\right)^2 = 36$$

گزینه ۲ صحیح است

آزمون فرض میانگین با مقدار ثابت از جامعه نرمال وقتی واریانس جامعه نامعلوم باشد : در این صورت از واریانس نمونه به جای واریانس جامعه استفاده می‌کنیم و بسته شرایط از توزیع z یا t استفاده می‌شود.

الف) حجم نمونه بیشتر از ۳۰ باشد:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \quad |z| > z_{\frac{\alpha}{2}} \Rightarrow \text{رد می‌شود } H_0$$

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad z < -z_{\alpha} \Rightarrow \text{رد می‌شود } H_0 \qquad \text{آماره آزمون } Z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad z > z_{\alpha} \Rightarrow \text{رد می‌شود } H_0$$

ب) حجم نمونه کمتر از ۳۰ باشد: در این صورت از توزیع t استیودنت با درجهٔ آزادی $df = n - 1$ استفاده می‌کنیم

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \quad |t| > t_{\frac{\alpha}{2},df} \Rightarrow \text{رد می‌شود } H_0$$

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \quad t < -t_{\alpha,df} \Rightarrow \text{رد می‌شود } H_0 \qquad \text{آماره آزمون } t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \quad t > t_{\alpha,df} \Rightarrow \text{رد می‌شود } H_0$$

مثال: در یک نمونه ۱۶ تایی میانگین نمونه ۴۳ و واریانس نمونه ۴۹ بدست آمده است در سطح خطای ۵٪ آزمون کنید آیا میانگین جامعه ۴۰ است. $df = n - 1 = 16 - 1 = 15$ $n = 16 < 30 \rightarrow t$

$$\begin{cases} H_0 : \mu = 40 \\ H_1 : \mu \neq 40 \end{cases} \qquad t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{43 - 40}{\frac{9}{\sqrt{16}}} = 1/333$$

چون $t = 1/333 \leq t_{0/025,15} = 2/131$ **، بنابراین** H_0 **پذیرفته می‌شود و با توجه به نمونه با احتمال ۹۵٪ میانگین جامعه ۴۰ است.**

ارشد مدیریت ۱۴۰۱

۶۴- برای آزمون $H_0 : \mu \leq 15$ در یک جامعهٔ نرمال از یک نمونهٔ تصادفی $n = 25$، $\bar{x} = 16$ و $S^2 = 0/16$ به‌دست‌آمده است. مقدار آمارهٔ آزمون کدام است؟

۱) ۰/۴

۲) ۵

۳) ۷/۵

۴) ۱۲/۵

$n=۲۵ \quad \bar{x}=۱۶ \quad s^۲=۰/۱۶ \Rightarrow s=۰/۴$

$$\begin{cases} H_۰ : \mu \le ۱۵ \\ H_۱ : \mu \succ ۱۵ \end{cases} \qquad t=\frac{\bar{x}-\mu_۰}{\frac{s}{\sqrt{n}}}=\frac{۱۶-۱۵}{\frac{۰/۴}{\sqrt{۲۵}}}=۱۲/۵$$

گزینه ۴ صحیح است

ارشد مدیریت ۱۴۰۲

۵۸- ادعا شده است: «میانگین جامعه‌ای برابر θ است». اگر از این جامعه نرمال نمونه تصادفی به حجم n انتخاب کنیم، آماره آزمون مناسب دارای چه توزیعی است؟

۱) $\chi^۲_n$

۲) $\chi^۲_{n-۱}$

۳) $t_{n-۱}$

۴) t_n

گزینه ۳ صحیح است

آزمون فرض نسبت: اغلب از توزیع z استفاده می‌شود

$$\begin{cases} H_۰ : P = P_۰ \\ H_۱ : P \ne P_۰ \end{cases} |z| > z_{\frac{\alpha}{۲}} \Rightarrow H_۰ \text{ رد می‌شود}$$

$$\begin{cases} H_۰ : P \ge P \\ H_۱ : P < P_۰ \end{cases} z < -z_\alpha \Rightarrow H_۰ \text{ رد می‌شود} \quad \text{آماره آزمون } Z=\frac{P-P_۰}{\sqrt{\frac{P_۰(۱-P_۰)}{n}}} \qquad P=\frac{x}{n}$$

$$\begin{cases} H_۰ : P \le P_۰ \\ H_۱ : P > P_۰ \end{cases} z > z_\alpha \Rightarrow H_۰ \text{ رد می‌شود}$$

مثال: در یک نمونه ۱۰۰ تایی تعداد ۱۸ کالا معیوب است در سطح خطای ۵٪ آزمون کنید نسبت کالای معیوب کمتر از ۲۰٪ است؟

$$P=\frac{x}{n}=\frac{۱۸}{۱۰۰}=۰/۱۸$$

$$\begin{cases} H_۰ : P \ge \%۲۰ \\ H_۱ : P < \%۲۰ \end{cases} \qquad Z=\frac{P-P_۰}{\sqrt{\frac{P_۰(۱-P_۰)}{n}}}=\frac{۰/۱۸-۰/۲}{\sqrt{\frac{۰/۲\times۰/۸}{۱۰۰}}}=-۰/۵$$

چون $z=-۰/۵ < -z_{/۰۵}=-۱/۶۴۵$، $H_۰$ پذیرفته می‌شود و با توجه به نمونه با احتمال ۹۵٪ نسبت کالای معیوب حداقل ۲۰٪ است.

ارشد مدیریت ۱۴۰۰

۶۳ ادعا شده است که نرخ بیکاری جوانان کشور کمتر از ۱۴ درصد است، فرضیه آماری H_0 کدام است؟

۱) $P > ۰/۱۴$

۲) $P \le ۰/۱۴$

۳) $P < ۰/۱۴$

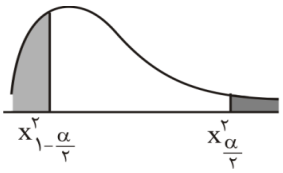
۴) $P \ge ۰/۱۴$

گزینه ۴ صحیح است چون علامت مساوی همیشه در H_o قرار می گیرد. $\begin{cases} H_۰ : P \ge \%۱۴ \\ H_۱ : P < \%۱۴ \end{cases}$

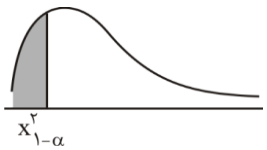
آزمون فرض واریانس با مقدار ثابت:

همانگونه که می‌دانیم $\frac{(n-۱)s^۲}{\sigma^۲}$ دارای توزیع خی دو $(\chi^۲)$ با درجه آزادی $\nu = df = n-۱$ است که این توزیع یک منحنی چوله به راست دارد آمار، آزمون در این آزمون از رابطهٔ $x^۲ = \frac{(n-۱)s^۲}{\sigma_۰^۲}$ بدست می‌آید.

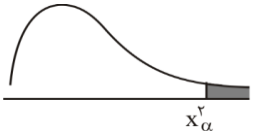
$\begin{cases} H_۰ : \sigma^۲ = \sigma_۰^۲ \\ H_۱ : \sigma^۲ \ne \sigma_۰^۲ \end{cases}$　$x^۲ > X^۲_{\frac{\alpha}{۲},df} \Rightarrow$　$x^۲ < X^۲_{۱-\frac{\alpha}{۲},df} \Rightarrow H_۰$ رد می‌شود



$\begin{cases} H_۰ : \sigma^۲ \ge \sigma_۰^۲ \\ H_۱ : \sigma^۲ < \sigma_۰^۲ \end{cases}$　$x^۲ < X^۲_{۱-\alpha,df} \Rightarrow H_۰$ رد می‌شود



$\begin{cases} H_۰ : \sigma^۲ \le \sigma_۰^۲ \\ H_۱ : \sigma^۲ > \sigma_۰^۲ \end{cases}$　$x^۲ < X^۲_{\alpha,df} \Rightarrow H_۰$ رد می‌شود



مثال: در یک نمونه ۱۶ تایی میانگین نمونه ۷۵ و واریانس نمونه ۴۴ بدست آمده است در سطح خطای ۵٪ آزمون کنید آیا واریانس جامعه بیش‌تر از ۴۰ است؟

$\begin{cases} H_۰ : \sigma^۲ \le ۴۰ \\ H_۱ : \sigma^۲ > ۴۰ \end{cases}$ آزمون آمار $x^۲ = \frac{(n-۱)s^۲}{\sigma_۰^۲} = \frac{۱۵ \times ۲۴}{۴۰} = ۱۶/۵$

چون $x^۲ = ۱۶/۵ > \chi^۲_{۰/۰۵,۱۵} = ۲۴/۹۹۶$، بنابرای با احتما ۹۵٪ با توجه به نمونه $H_۰$ پذیرفته می‌شود و واریانس جامعه حداقل ۴۰ است

ارشد مدیریت ۱۴۰۲

۵۹- به منظور آزمون فرضیه «پراکندگی وزن محصولات کارخانه، کمتر از ۶ است»، نمونه تصادفی ۲۰ تایی از این جامعه نرمال انتخاب شده است. با توجه به سطح زیرِ منحنی دنباله سمت راست توزیع χ^2 ارائه‌شده در جدول زیر، به ازای چه مقدار برای آماره آزمون فرضیه پژوهشی در سطح معناداری ۱۰ درصد رد می‌شود؟

df \ α	۰/۹۵	۰/۹	۰/۱	۰/۵
۱۹	۱۰/۱	۱۱/۶	۲۷/۲	۳۰/۱
۲۰	۱۰/۸	۱۲/۴	۲۸/۴	۳۱/۴

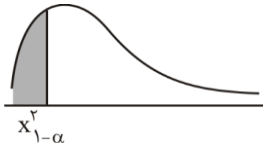
۱) $\chi^2 \le ۲۸/۴$

۲) $\chi^2 \le ۲۷/۲$

۳) $\chi^2 \ge ۱۲/۴$

۴) $\chi^2 \ge ۱۱/۶$

جهت ناریب بودن برای واریانس آزمون انجام میشود.



$$n=۲۰ \Rightarrow df = n-۱ = ۲۰-۱ = ۱۹ \quad , \quad ۱-\alpha = ۱-۰/۱ = ۰/۹$$

$$\Rightarrow \chi^2_{۰/۹,۱۹} = ۱۱/۶$$

$$\begin{cases} H_۰ : \sigma \ge ۶ \\ H_۱ : \sigma < ۶ \end{cases} \Leftrightarrow \begin{cases} H_۰ : \sigma^2 \ge ۳۶ \\ H_۱ : \sigma^2 < ۳۶ \end{cases}$$

گزینه ۴ صحیح است

آزمون فرض اختلاف میانگین دو جامعه نرمال مستقل وقتی واریانس جوامع معلوم باشد: در این صورت بدون توجه به حجم نمونه از توزیع z استفاده می‌شود و آمار، آزمون از رابطه

$$Z = \frac{\bar{x}_۱ - \bar{x}_۲ - (\mu_۱ - \mu_۲)}{\sqrt{\frac{\sigma_۱^2}{n_۱} + \frac{\sigma_۲^2}{n_۲}}}$$

بدست می‌آید که اگر آزمون برابری میانگینها انجام شود به جای $\mu_۱ - \mu_۲$ عدد صفر را قرار می‌دهیم

$$\begin{cases} H_۰ : \mu_۱ - \mu_۲ = a \\ H_۱ : \mu_۱ - \mu_۲ \ne a \end{cases} \quad |z| > z_{\frac{\alpha}{۲}} \Rightarrow H_۰ \text{ رد می‌شود}$$

$$\begin{cases} H_۰ : \mu_۱ - \mu_۲ \ge a \\ H_۱ : \mu_۱ - \mu_۲ < a \end{cases} \quad z < -z_\alpha \Rightarrow H_۰ \text{ رد می‌شود}$$

آماره آزمون $$Z = \frac{\bar{x}_۱ - \bar{x}_۲ - a}{\sqrt{\frac{\sigma_۱^2}{n_۱} + \frac{\sigma_۱^2}{n_۲}}}$$

$$\begin{cases} H_۰ : \mu_۱ - \mu_۲ \le a \\ H_۱ : \mu_۱ - \mu_۲ > a \end{cases} \quad z > z_\alpha \Rightarrow H_۰ \text{ رد می‌شود}$$

آزمون فرض اختلاف میانگین دو جامعه نرمال مستقل وقتی واریانس جوامع نامعلوم باشد. در این صورت با فرض برابر واریانسها دو حالت را در نظر می‌گیریم.

الف)حجم نمونه بیش از ۳۰ باشد. در این صورت از توزیع z استفاده می‌شود.

$$\begin{cases} H_۰ : \mu_۱ - \mu_۲ = a \\ H_۱ : \mu_۱ - \mu_۲ \ne a \end{cases} \quad |z| > z_{\frac{\alpha}{۲}} \Rightarrow H_۰ \text{ رد می‌شود}$$

$$\begin{cases} H_0 : \mu_1 - \mu_2 \geq a \\ H_1 : \mu_1 - \mu_2 < a \end{cases} \quad z < -z_\alpha \Rightarrow H_0 \text{ رد می‌شود}$$

آماره آزمون $$Z = \frac{\bar{x}_1 - \bar{x}_2 - a}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$\begin{cases} H_0 : \mu_1 - \mu_2 \leq a \\ H_1 : \mu_1 - \mu_2 > a \end{cases} \quad z > z_\alpha \Rightarrow H_0 \text{ رد می‌شود}$$

ب) حجم نمونه‌ها کمتر از ۳۰ باشد(**t** مستقل) :در این صورت از توزیع **t** با درجه آزادی $df = n_1 + n_2 - 2$ استفاده می‌شود.

$$\begin{cases} H_0 : \mu_1 - \mu_2 = a \\ H_1 : \mu_1 - \mu_2 \neq a \end{cases} \quad |t| > t_{\frac{\alpha}{2},df} \Rightarrow H_0 \text{ رد می‌شود}$$

$$df = n_1 + n_2 - 2$$

$$s_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$\begin{cases} H_0 : \mu_1 - \mu_2 \geq a \\ H_1 : \mu_1 - \mu_2 < a \end{cases} \quad t < -t_{\alpha,df} \Rightarrow H_0 \text{ رد می‌شود}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2 - a}{S_{\bar{x}_1 - \bar{x}_2}} \Rightarrow t = \frac{\bar{x}_1 - \bar{x}_2 - a}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\begin{cases} H_0 : \mu_1 - \mu_2 \leq a \\ H_1 : \mu_1 - \mu_2 > a \end{cases} \quad t > t_{\alpha,df} \Rightarrow H_0 \text{ رد می‌شود}$$

یاد آوری : در استفاده از فاصله اطمینان برای بررسی معنی داری آزمون ها ی برابری میانگین دو جامعه اگر عدد صفر در فاصله اطمینان ۹۵٪ قرار گیرد فرض تایید می شود و $p \succ ۰/۰۵$ است

مثال: از دو جامعه نرمال مستقل نمونه‌های زیر بدست آمده است:

A	$\bar{x}_1 = ۸۰$	$n_1 = ۱۱$	$s_1^2 = ۵۰$
B	$\bar{x}_2 = ۸۵$	$n_2 = ۱۶$	$s_2^2 = ۶۰$

در سطح معنی داری ۵٪ آزمون کنید میانگین دو جامعه برابر است.

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

$$df = n_1 + n_2 - 2 = ۱۱ + ۱۶ - ۲ = ۲۵$$

$$s_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{۱۰ \times ۵ + ۱۵ \times ۶۰}{۲۵} = ۵۶$$

$$t = \frac{\bar{x}_1 - \bar{x}_2 - a}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{۸۵ - ۸۰}{\sqrt{۵۶}\sqrt{\frac{1}{۱۱} + \frac{1}{۱۶}}} = +۱/۷$$

چون $|t| = |-۱/۷| > t_{./۰۲۵,۲۵} = ۲/۰۶۰$ بنابراین H_0 پذیرفته می‌شود و با احتمال ۹۵٪ با توجه به نمونه میانگین دو جامعه برابر است.

مثال : در گزارش یک پژوهش نتایج آزمون t با دو گروه مستقل نتایج به صورت $t_{۱۲۸} = ۳/۱۶$ گزارش شده است حجم نمونه چند نفر بوده است

۱) ۱۲۷ ۲) ۱۲۸ ۳) ۱۲۹ ۴) ۱۳۰

گزینه۴ صحیح است $df = n_۱ + n_۲ - ۲ \Rightarrow ۱۲۸ = n_۱ + n_۲ - ۲ \Rightarrow n_۱ + n_۲ = ۱۲۸ + ۲ = ۱۳۰$

اندازه اثر : در صورتیکه حد بالای میانگین کوچکتر مثلاگروه ۲ از حد پایین میانگین بزرگترمثلا گروه۱ مقدار کوچکتری داشته باشد همپوشانی ندارند اندازه اثر بزرگتر نشان دهنده تفاوت بیشتر دو گروه است هر چه عدد اندازه اثر کوچکتر باشد همپوشانی بیشتر است یکی از مزیتهای اندازه اثر مستقل بودن از حجم نمونه است.

آزمون فرض برابری میانگین‌های دو جامعه نرمال مستقل با فرض مجهول بودن واریانسها و برابر نبودن واریانسها در این صورت از توزیع t استفاده می‌شود با درجه آزادی

$\begin{cases} H_۰ : \mu = \mu_۰ \\ H_۱ : \mu \neq \mu_۰ \end{cases}$ $|t'| > t_{\frac{\alpha}{۲},df}$ $\Rightarrow$ رد می‌شود $H_۰$

$$df = \frac{\left(\frac{s_۱^۲}{n_۱} + \frac{s_۲^۲}{n_۲}\right)^۲}{\frac{\left(\frac{s_۱^۲}{n_۱}\right)^۲}{n_۱ - ۱} + \frac{\left(\frac{s_۲^۲}{n_۲}\right)^۲}{n_۲ - ۱}}$$

$\begin{cases} H_۰ : \mu \geq \mu_۰ \\ H_۱ : \mu < \mu_۰ \end{cases}$ $t' < -t_{\alpha,df}$ $\Rightarrow$ رد می‌شود $H_۰$

$$t' = \frac{\bar{x}_۱ - \bar{x}_۲}{\sqrt{\frac{s_۱^۲}{n_۱} + \frac{s_۲^۲}{n_۲}}}$$

$\begin{cases} H_۰ : \mu \leq \mu_۰ \\ H_۱ : \mu > \mu_۰ \end{cases}$ $t' > t_{\alpha,df}$ $\Rightarrow$ رد می‌شود $H_۰$

در کتاب آمار مهنسی لیبرمن درجه آزادی به صورت $df = \frac{\left(\frac{s_۱^۲}{n_۱} + \frac{s_۲^۲}{n_۲}\right)^۲}{\frac{\left(\frac{s_۱^۲}{n_۱}\right)^۲}{n_۱ + ۱} + \frac{\left(\frac{s_۲^۲}{n_۲}\right)^۲}{n_۲ + ۱}} - ۲$ بیان شده است

مقایسه‌های زوج شده (t وابسته):

هرگاه افراد مورد آزمایش ثابت باشد و دو آزمایش بر روی این افراد انجام شود از مقایسه‌های زوج شده استفاده می‌شود مانند نمرات افراد قبل از رفتن به کلاس درس در یک درس خاص که از توزیع t با درجهٔ آزادی $df = n - ۱$ استفاده می‌شود برای انجام این آزمون اختلاف دو حالت را بدست آورده و بررسی

میانگینهای آزمون را انجام می‌دهیم

$\begin{cases} H_0 : \mu_d = a \\ H_1 : \mu_d \neq a \end{cases}$ $|t| > t_{\frac{\alpha}{2},df} \Rightarrow$ H_0 رد می‌شود

$d_i = x_i - y_i \qquad df = n-1$

$\begin{cases} H_0 : \mu_d \geq a \\ H_1 : \mu_d < a \end{cases}$ $t < -t_{\alpha,df} \Rightarrow$ H_0 رد می‌شود $\qquad s_d^2 = \dfrac{\sum d_i^2 - n\bar{d}^2}{n-1}$

$t = \dfrac{\bar{d}-a}{\frac{s_d}{\sqrt{n}}}$

$\begin{cases} H_0 : \mu_d \leq a \\ H_1 : \mu_d > a \end{cases}$ $t > t_{\alpha,df} \Rightarrow$ H_0 رد می‌شود

مثال: سرعت خواندن ۶ دانشجو قبل از رفتن به کلاس تندخوانی و بعد از رفتن به کلاس در جدول زیر ثبت شده است آیا سرعت خواندن به طور متوسط افزایش یافته است. (در سطح خطای ۵٪)

قبل(x)	۲	۲/۵	۴	۵	۷	۶
بعد (y)	۳	۳	۴/۵	۱۰	۱۲	۹

$d_x = y_i - x_i$	۱	۰/۵	۰/۵	۵	۵	۳	۱۵
d_i^2	۱	۰/۲۵	۰/۲۵	۲۵	۲۵	۹	۶۰/۵

$df = n-1 = ۶-۱ = ۵ \qquad \bar{d} = \dfrac{\sum x_i}{n} = \dfrac{۱۵}{۶} = ۲/۵$

$s_d^2 = \dfrac{\sum x_i^2 - n\bar{d}^2}{n-1} = \dfrac{۶۰/۵ - ۶\times(۲/۵)^2}{۶-۱} = ۴/۶$

$t = \dfrac{\bar{d}-a}{\frac{sd}{\sqrt{n}}} = \dfrac{۲/۵}{\frac{\sqrt{۴/۶}}{\sqrt{۶}}} = ۲/۸۷$

چون $t = ۲/۸۷ > t_{./۰۵,۵} = ۲/۰۱۵$ بنابراین H_0 رد می‌شود و با احتمال ۹۵٪ و با توجه به $\begin{cases} H_0 : \mu_d \leq ۰ \\ H_1 : \mu_d > ۰ \end{cases}$

نمونه رفتن به کلاس مؤثر بوده است.

مثال : در یک آزمون **T**وابسته به حجم ۹ اگرمیانگین اختلاف نمرات قبل و بعد از آزمون ۰/۵ با انحراف استاندارد اختلاف نمرات ۰/۵باشد مقدار آماره آزمون را بدست آورید. $t = \dfrac{\bar{d}-a}{\frac{s_d}{\sqrt{n}}} = \dfrac{۰/۵}{\frac{۰/۵}{\sqrt{۹}}} = \sqrt{۹} = ۳$

مثال : در یک نمونه ۱۰۰ تایی اگرمیانگین اختلاف نمرات قبل و بعد از آزمون ۸ با انحراف استاندارد اختلاف نمرات ۱۰ باشد یک برآورد فاصله ای ۹۵٪ برای اختلاف نمرات بنویسید . $t_{./۰۲۵,۹۹} \approx ۲$

$t = \dfrac{\bar{d}-a}{\frac{s_d}{\sqrt{n}}} \Rightarrow d : \bar{d} \pm t_{\frac{\alpha}{2},df} . \dfrac{s_d}{\sqrt{n}} \Rightarrow d : ۱۰ \pm ۲ \times \dfrac{۸}{\sqrt{۱۰۰}} = ۱۰ \pm ۱/۶ \Rightarrow ۸/۴ \leq d \leq ۱۱/۶$

آزمون فرض اختلاف نسبتها:

از توزیع z استفاده می‌کنیم

$$\begin{cases} H_0 : P_1 - P_2 = a \\ H_1 : P_1 - P_2 \neq a \end{cases} |z| > z_{\frac{\alpha}{2}} \Rightarrow \text{رد می‌شود } H_0$$

$$P_1 = \frac{x_1}{n_1} \qquad P_2 = \frac{x_2}{n_2}$$

$$z = \frac{P_1 - P_2 - a}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}}$$

$$\begin{cases} H_0 : P_1 - P_2 \geq a \\ H_1 : P_1 - P_2 < a \end{cases} z < -z_\alpha \Rightarrow \text{ردمی‌شود } H_0$$

$$\hat{P} = \frac{x_1 + x_2}{n_1 + n_2} \qquad z = \frac{P_1 - P_2 - a}{\sqrt{\hat{P}(1-\hat{P}) * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\begin{cases} H_0 : P_1 - P_2 \leq a \\ H_1 : P_1 - P_2 > a \end{cases} z > z_\alpha \Rightarrow \text{رد می‌شود } H_0$$

ارشد مدیریت ۱۴۰۱

۶۳- ادعا شده است، نسبت ضایعات در کارخانه (۱) کمتر از نسبت ضایعات در کارخانه (۲) است. فرضیه H_0 کدام است؟

۱) $H_0 : p_1 \geq p_2$

۲) $H_0 : \mu_1 \geq \mu_2$

۳) $H_0 : p_1 \leq p_2$

۴) $H_0 : \mu_1 \leq \mu_2$

گزینه ۱ صحیح است علامت مساوی همیشه در H_0 قرار می گیرد. $\begin{cases} H_0 : P_1 - P_2 \geq 0 \\ H_1 : P_1 - P_2 < 0 \end{cases} \Rightarrow \begin{cases} H_0 : P_1 \geq P_2 \\ H_1 : P_1 < P_2 \end{cases}$

آزمون فرض نسبت دو واریانس:

همانگونه که می‌دانیم نسبت دو واریانس دارای توزیع **F** یا فیشر با درجات آزادی $\nu_1 = n_1 - 1$ و $\nu_2 = n_2 - 1$ است و آمارهٔ آزمون از رابطه $F = \left(\frac{\sigma_2^2}{\sigma_1^2}\right)\frac{s_1^2}{s_2^2}$ بدست می‌آید.

$$\begin{cases} H_0 : \frac{\sigma_1^2}{\sigma_2^2} = a \\ H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq a \end{cases} F_{1-\frac{\alpha}{2},\nu_1,\nu_2} \langle FF > F_{\frac{\alpha}{2},\nu_1,\nu_2} \Rightarrow \text{رد می‌شود } H_0$$

$$\begin{cases} H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} \geq a \\ H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} < a \end{cases} \quad F < F_{1-\alpha,\nu_1,\nu_2} \Rightarrow H_0 \text{ رد می‌شود}$$

$$\begin{cases} H_0 : \dfrac{\sigma_1^2}{\sigma_2^2} \leq a \\ H_1 : \dfrac{\sigma_1^2}{\sigma_2^2} > a \end{cases} \quad F > F_{\alpha,\nu_1,\nu_2} \Rightarrow H_0 \text{ رد می‌شود}$$

نکته: آمارهٔ آزمون $F = a_0 \dfrac{s_1^2}{s_2^2}$ در آزمون برابر واریانس‌ها $a = 1$ است.

مثال: در سطح خطای ۱۰٪ آزمون کنید آیا واریانس دو جامعه A و B برابر است

A	$n_1 = 10$	$\bar{x}_1 = 102$	$s_1^2 = 80$
B	$n_2 = 6$	$\bar{x}_2 = 105$	$s_2^2 = 50$

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases} \qquad F = \frac{s_1^2}{s_2^2} = \frac{80}{50} = 1/6$$

$$F_{\frac{\alpha}{2},\nu_1,\nu_2} = F_{0/05,9,5} = 4/7725$$

$$F_{1-\frac{\alpha}{2},\nu_1,\nu_2} = \frac{1}{F_{\frac{\alpha}{2},\nu_1,\nu_2}} = F_{0/05,5,9} = \frac{1}{3/4817} = 0/2872 = F_{0/95,9,5}$$

چون $F = 1/6 > F_{0/05,9,5} = 4/7725$ یا $F = 1/6 < F_{0/95,9,5} = 0/2872$ بنابراین H_0 با توجه به نمونه با احتمال ۹۵٪ پذیرفته می‌شود و واریانس دو جامعه برابر است.

ارشد مدیریت ۹۹

۵۶- به منظور آزمون فرضیه «ریسک سرمایه‌گذاری در بورس کمتر از مسکن است»، نمونه‌هایی تصادفی از هر گروه انتخاب شده است که اطلاعات آن به‌صورت جدول زیر است:

نوع سرمایه‌گذاری	تعداد نمونه	میانگین بازدهی	انحراف معیار بازدهی
بورس	۲۵	۱۰۰	۴
مسکن	۱۶	۶۴	۸

با فرض نرمال بودن توزیع سود سرمایه‌گذاری در بورس و مسکن، آماره آزمون این فرضیه کدام است؟

۱) ۰/۵۰

۲) ۰/۲۵

۳) ۱/۵۶

۴) ۲

$$F=\frac{s_1^2}{s_2^2}=\frac{4^2}{8^2}=0/25$$

گزینه ۲ صحیح است

آزمون کوکران برای همگنی واریانسها:

همانگونه که می‌دانیم برای آزمون ۲ واریانس از آزمون F استفاده می‌شود ولی اگر تعداد واریناس‌های مورد مقایسه بیش از ۲ باشد از آزمون کوکران استفاده می‌شود.

$$\begin{cases} H_0 : \sigma_1^2=\sigma_2^2=\ldots=\sigma_K^2 \\ H_1 : \sigma_i^2 \neq \sigma_j^2 \end{cases}$$

آماره آزمون $G=\frac{s_{max}^2}{s_1^2+s_2^2+\ldots+s_k^2}$

در صورتی‌که $G>g_\alpha$ **باشد فرض** H_0 **رد می‌شود و حداقل یکی از** σ^2 **ها متفاوت است.**

فصل ششم

ضریب همبستگی

رگرسیون

ضریب همبستگی: معیاری است برای تعیین ارتباط بین یک یا چند متغیر مستقل (x) با یک متغیر وابسته (y) در صورتی که ارتباط بین دو متغیر همسو باشد مانند قد و وزن ضریب همبستگی مثبت و در صورتی که ارتباط بین دو متغیر غیر همسو باشد مانند رابطه فشار هوا و فاصله از سطح زمین ضریب همبستگی منفی می‌شود در آمار پارامتریک از ضریب همبستگی پیرسن استفاده می‌کنیم که عددی بین ۱- و ۱+ است و بدون واحد است ضریب همبستگی نمونه‌ای را با حرف r و ضریب همبستگی جامعه را با حرف ρ نشان می‌دهیم.

$$r=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2.\sum(y_i-\bar{y})^2}}\Rightarrow r=\frac{n\sum xy-(\sum x.\sum y)}{\sqrt{\sum(x_i-\bar{x})^2.\sum(y_i-\bar{y})^2}}$$

$$S_{pxy}=S_{xy}=\sum(x_i-\bar{x})(y_i-\bar{y})=\sum x_iy_i-n\bar{x}\bar{y}=\sum x_iy_i-\frac{\sum x_i\sum y_i}{n}$$

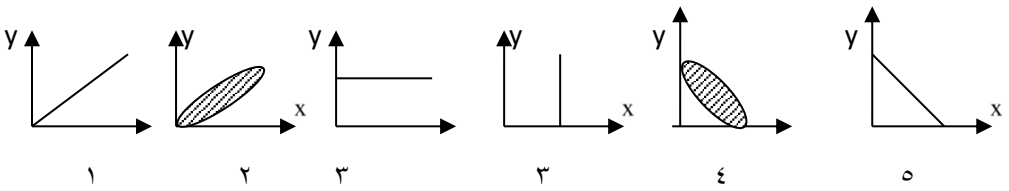
$$S_{xx}=\sum(x_i-\bar{x})^2=\sum x_i^2-n\bar{x}^2=\sum x_i^2-\frac{(\sum x_i)^2}{n}$$

$$S_{yy}=\sum(y_i-x\bar{y})^2=\sum y_i^2-n\bar{y}^2=\sum y_i^2-\frac{(\sum y_i)^2}{n}$$

با جایگذاری در رابطه ضریب همبستگی می‌توان نوشت

$$r=\frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}\Rightarrow r=\frac{s_{xy}}{\sqrt{N\sigma_x^2N\sigma_y^2}}\Rightarrow r=\frac{s_{xy}}{N\sigma_x\sigma_y}\quad ,cov_{xy}=\frac{s_{xy}}{n}\Rightarrow r=\frac{cov_{xy}}{\sigma_x\sigma_y}$$

$$|r|\le 1\Rightarrow -1\le r\le +1\Rightarrow\begin{cases}r=+1 & \text{همبستگی خطی مستقیم کامل}\\ 0<r<1 & \text{همبستگی مستقیم}\\ r=0 & \text{عدم همبستگی خطی}\\ -1<r<0 & \text{همبستگی معکوس}\\ r=-1 & \text{همبستگی خطی معکوس کامل}\end{cases}$$



توجه : متغیری که تغییرات آن اندازه گیری می شود را متغیر وابسته گویند.

ضریب تعیین: مجذور ضریب همبستگی را ضریب تعیین گویند که نشان دهندهٔ آن است که چنددرصد تغییرات متغیروابسته توسط متغیرمستقل بیان می‌شود. $V=r_{xy}^2$ضریب تعیین که اغلب بر حسب درصد بیان میشود یعنی $\%V=(r_{xy}^2).(100)$

مثال :ضریب همبستگی بین دو متغیر x,y را تعیین کنید و نوع آن را مشخص کنید

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})\times(y-\bar{y})$
۲	۱۵	۲−۶=−۴	۱۵−۵=۱۰	$(-۴)^۲=۱۶$	$(۱۰)^۲=۱۰۰$	−۴×۱۰=−۴۰
۴	۱۰	۴−۶=−۲	۱۰−۵=۵	$(-۲)^۲=۴$	$(۵)^۲=۲۵$	−۲×۵=−۱۰
۶	۵	۶−۶=۰	۵−۵=۰	$(۰)^۲=۰$	$(۰)^۲=۰$	۰×۰=۰
۸	۰	۸−۶=۲	۰−۵=−۵	$(۲)^۲=۴$	$(-۵)^۲=۲۵$	۲×−۵=−۱۰
۱۰	−۵	۱۰−۶=۴	−۵−۵=−۱۰	$(۴)^۲=۱۶$	$(-۱۰)^۲=۱۰۰$	۴×−۱۰=−۴۰
$\sum x=۳۰$	$\sum y=۲۵$	۰	۰	$SS_x=۴۰$	$SS_y=۲۵۰$	$S_{xy}=-100$

$$\bar{x}=\frac{\sum x}{n}=\frac{۳۰}{۵}=۶ \quad , \quad \bar{y}=\frac{\sum y}{n}=\frac{۲۵}{۵}=۵$$

همبستگی معکوس کامل

$$r=\frac{S_{xy}}{\sqrt{SS_x\times SS_y}}=\frac{-۱۰۰}{\sqrt{۴۰\times۲۵۰}}=\frac{-۱۰۰}{\sqrt{۱۰۰۰۰}}=\frac{-۱۰۰}{۱۰۰}=-۱$$

نکته۱ : به طور کلی اگر فواصل x,y مساوی باشد ضریب همبستگی یک است و علامت آن بستگی به حاصل ضرب تغییرات دارد.

x	y	x	y	x	y	x	y
۲	۱۵	۳	۴	۲	۱۰	۳	۵
۴	۱۰	۶	۸	۲	۱۵	۶	۵
۶	۵	۹	۱۲	۲	۲۰	۴۵	۵
۸	۰	۱۲	۱۶	۲	۳۰	۱۰	۵
۱۰	۵−	۱۵	۲۰	۲	۵۰	۷۰	۵
r =-۱		r =۱		r =۰		r =۰	

آزمون فرض ضریب همبستگی با مقدار ثابت: این آزمون جهت نشان دادن میزان تاثیر متغیر مستقل بر متغیر وابسته را نشان می دهد که به دو روش صورت می‌گیرد

۱- با استفاده از آزمون t : آماره آزمون از رابطه $t=\frac{r\sqrt{n-۲}}{\sqrt{۱-r^۲}}$ **که نقطه بحرانی با درجه آزادی**

$df = n-2$ بدست می آید یعنی اگر $t \succ t_{\frac{\alpha}{2},n-2}$ باشد فرض صفر رد می شود و ضریب همبستگی معنی دار است با اطمینان $p = 1-\alpha$ و متغیر مستقل و وابسته رابطه دارند. در صورتیکه فرض صفر پذیرفته شود $(H_0 : \rho = 0)$ رابطه میان ضریب همبستگی و توان آماری ناودیسی هست.

$\begin{cases} H_0 : \rho = \rho_0 \\ H_1 : \rho \neq \rho_0 \end{cases}$ $|t| > t_{\frac{\alpha}{2},df} \Rightarrow H_0$ رد می‌شود

$\begin{cases} H_0 : \rho \geq \rho_0 \\ H_1 : \rho < \rho_0 \end{cases}$ $t < -t_{\alpha,df} \Rightarrow H_0$ رد می‌شود

$df = n-2$

آماره آزمون $t = \dfrac{(r-\rho_0)\sqrt{n-2}}{\sqrt{1-r^2}}$

$\begin{cases} H_0 : \rho \leq \rho_0 \\ H_1 : \rho > \rho_0 \end{cases}$ $t > t_{\alpha,df} \Rightarrow H_0$ رد می‌شود

۲- با استفاده از توزیع z:

در این صورت با استفاده از تبدیل فیشر برای $n > 20$، $z = \frac{1}{2} Ln \frac{1+r}{1-r}$ دارای توزیع نرمال با میانگین $E(z) = \frac{1}{2} Ln \frac{1+\rho}{1-\rho}$ و واریانس $var(z) = \frac{1}{n-3}$ است

$\begin{cases} H_0 : \rho = \rho_0 \\ H_1 : \rho \neq \rho_0 \end{cases}$ $|z^*| > z_{\frac{\alpha}{2}} \Rightarrow H_0$ رد می‌شود

$\begin{cases} H_0 : \rho \geq \rho_0 \\ H_1 : \rho < \rho_0 \end{cases}$ $z^* < -tz_\alpha \Rightarrow H_0$ رد می‌شود $\quad z^* = \dfrac{\frac{1}{2} Ln \frac{1+r}{1-r} - \frac{1}{2} Ln \frac{1+\rho_0}{1-\rho_0}}{\frac{1}{\sqrt{n-3}}}$

$\begin{cases} H_0 : \rho \leq \rho_0 \\ H_1 : \rho > \rho_0 \end{cases}$ $z^* > z_\alpha \Rightarrow H_0$ رد می‌شود

مثال: در صورتی که همبستگی هوش و عملکرد تحصیلی ۰/۹ باشد فرض صفر بودن همبستگی در جامعه به کدام صورت نوشته می شود؟

۱) $H_0 : \rho_{xy} \neq 0$ ۲) $H_0 : \rho_{xy} = 0$ ۳) $H_0 : \rho_{xy} \geq 0$ ۴) $H_0 : \rho_{xy} \leq 0$

گزینه۲ صحیح است چون علامت مساوی در H_0 قرار میگیرد و گزینه ۳ جهت حداقل ۰/۹ است

مثال: در صورتی که در یک نمونه ۱۸ تایی ضریب همبستگی دو متغیر xوy برابر ۰/۸ باشد آماره آزمون را بدست آورید و در سطح معنی داری ۵٪ آزمون کنید معنی دار بودن ضریب همبستگی را.

$$r = ۰/۸\ ,\ N = ۱۸ \Rightarrow df = n - ۲ = ۱۸ - ۲ = ۱۶$$

$$t = \frac{r_{xy}\sqrt{N-۲}}{\sqrt{۱-r^{۲}}} = \frac{۰/۸\times\sqrt{۱۸-۲}}{\sqrt{۱-۰/۸^{۲}}} = \frac{۰/۸\times\sqrt{۱۶}}{\sqrt{۰/۳۶}} = \frac{۳/۲}{۰/۶} = ۵/۳۳$$

چون $t = ۵/۳۳ \succ t_{۰/۰۲۵,۱۸} = ۲/۱۰۱$ **فرض صفر رد می شود ضریب همبستگی معنی دار است با احتمال ۹۵٪**

مثال : اگر به ازای سه نمره $x_۱, x_۲, x_۳$ **در متغیر X و** $y_۱, y_۲, y_۳$ **در متغیر Y روابط زر برقرار باشد همبستگی چند خواهد بود؟**

$$(x_۱ - \bar{x}).(y_۱ - \bar{y}) = +۱$$
$$(x_۲ - \bar{x}).(y_۲ - \bar{y}) = ۰$$
$$(x_۳ - \bar{x}).(y_۳ - \bar{y}) = +۱$$

۱) -۱　　۲) +۰/۵　　۳) ۰　　۴) +۱

گزینه ۳ صحیح است.در واقع جمع اعداد $s_{xy} = 0$ **برابر صفر است بنابراین ضریب همبستگی صفر است**

نکته: آزمون ضریب همبستگی به شدت تابع حجم نمونه است. تا جائی که امکان دارد حجم نمونه را افزایش می دهیم تا آزمون با صحت بالاتری باشد

نکته۲ : اگر مقادیر xوy استاندار شده باشد $r = \frac{\sum Z_x . Z_y}{N}$

نکته۳:چون $t_{df}^{۲} = F_{۱,df} \Leftarrow$

$$t = \frac{r\sqrt{n-۲}}{\sqrt{۱-r^{۲}}} \xrightarrow{t^{۲}=F} F = \left(\frac{r\sqrt{n-۲}}{\sqrt{۱-r^{۲}}}\right)^{۲} \Rightarrow F = \frac{r^{۲}.(n-۲)}{۱-r^{۲}}$$

نکته۴: $COV(ax, by) = a.b.COV(x, y)$

نکته۵: پیش بینی نمره استاندارد $Z_{y'} = Z_x . r_{xy}$

نکته۶ : همیشه علامت s_{xy}, r, cov, b **یکی است**

مثال: متغیر X دارای توزیع نرمال $\bar{x} = ۳۶, S_x = ۴$ **و متغیر Y دارای توزیع نرمال با** $\bar{y} = ۷۰, S_y = ۷$ **است . اگر همبستگی دو متغیر** $r_{xy} = -۱$ **باشد . کواریانس بین دو متغیر کدام است؟**

۱) -۹　　۲) -۱۰　　۳) -۲۸　　۴) -۳۶

گزینه۳ درست است

$s_x = ۴, s_y = ۷, r_{xy} = -۱$

$$r_{xy} = \frac{cov(x,y)}{s_x.s_y} \Rightarrow -۱ = \frac{cov(x,y)}{۴\times۷} \Rightarrow cov(x,y) = -۱\times۴\times۷ = -۲۸$$

ارشد مدیریت۱۳۹۹

۵۷- اگر $\bar{x}=۴$ ، $\bar{y}=۵$ ، $s_x^۲=۱۶$ ، $s_y^۲=۲۵$ و $cov(x,y)=-۱۵$ باشد، چند درصد تغییرات y به وسیلهٔ x بیان می‌شود؟

۱) ۰/۷۵

۲) ۰/۴۷

۳) ۰/۵۶

۴) ۰/۲۵

$$s_x^۲ = ۱۶ \Rightarrow s_x = ۴, s_y^۲ = ۲۵ \Rightarrow s_y = ۵$$

$$r_{xy} = \frac{cov(x,y)}{s_x.s_y} \Rightarrow r_{xy} = \frac{-۱۵}{۴\times۵} \Rightarrow r_{xy} = -۰/۷۵$$

$$\%r_{xy}^۲ = (-۰/۷۵)^۲ \times ۱۰۰ = \%۵۶/۲۵$$

گزینه ۳ صحیح است .

ارشد مدیریت ۱۴۰۰

۶۴- اگر $\bar{x}=۱۶$ ، $\bar{y}=۱۶$ ، $\sigma_x^۲=۹$ ، $\sigma_y^۲=۱۶$ و $Cov(x,y)=-۹$ باشد. چند درصد تغییرات y به‌وسیله x بیان می‌شود؟

۱) ۷۵-

۲) ۶/۳-

۳) ۵۰

۴) ۵۶/۳

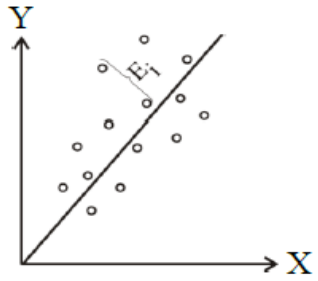
$$s_x^۲ = ۹ \Rightarrow s_x = ۳ \ , s_y^۲ = ۱۶ \Rightarrow s_y = ۴ \ , cov(x,y) = -۹$$

$$r_{xy} = \frac{cov(x,y)}{s_x.s_y} \Rightarrow r_{xy} = \frac{-۹}{۳\times۴} = -۰/۷۵$$

$$\%r_{xy}^۲ = (-۰/۷۵)^۲ \times ۱۰۰ = \%۵۶/۲۵$$

رگرسیون خطی :

رگرسیون به معنی بازگشت برای تعیین رابطه بین یک چند متغیر مستقل (پیش بین) با یک متغیر وابسته (ملاک)است به عبارتی رگرسیون مطالعه چگونگی بازگشت نمره های y به x می باشد . در رگرسیون خطی ساده رابطه بین مستقل و یک متغیر وابسته را بررسی می‌کنیم در رگرسیون دو متغیر باید کمی (مقیاس فاصله ای یا نسبی) باشند و کاربرد رگرسیون خطی پیش بینی متغیر ملاک بر اساس متغیر پیش بین است .پارامترهای مدل به روش کمترین مربعات (L.S) بدست می‌آید رگرسیون خطی به صورت $y_i = a + bx_i + e_i$ است که $E_i \sim N\left(۰, \sigma_e^۲\right)$ هستند.



$$e_i = y_i - \hat{y}_i$$

$$Q = \sum E_i^2 = \sum (y_i - a - b_i x_i)^2$$

$$\begin{cases} \frac{\partial Q}{\partial a} = -2\sum (y_i - a - b_i x_i) = 0 \\ \frac{\partial Q}{\partial b} = -2\sum (y_i - a - b_i x_i) = 0 \end{cases} \Rightarrow \begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \Rightarrow \begin{cases} b = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \Rightarrow b = \frac{n\sum xy - (\sum x.\sum y)}{n\sum x^2 - (\sum x)^2} \end{cases}$$

توجه ۱: در صورتیکه $\sum (y_i - \hat{y}_i)^2 = 0$ باشد بیشترین رابطه بین متغیر مستقل و وابسته وجود دارد

توجه۲: تکرار آزمایش بهترین روش تشخیص برای ثبات تاثیر متغیر مستقل بر متغیر وابسته است.

مراحل نوشتن معادله خط رگرسیون :

۱) $b_{xy} = \frac{S_{xy}}{Sxx} = \frac{n\sum xy - (\sum x.\sum y)}{n\sum x^2 - (\sum x)^2}$ ۲) $a = \bar{y} - b\bar{x}$ ۳) $y = a + b\,x$

یادآوری

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$S_{yy} = \sum (y_i - x\bar{y})^2 = \sum y_i^2 - n\bar{y}^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

محاسبه مجموع مربعات خطا و واریانس خطا به صورت زیر محاسبه می‌شود.

$$\sum e_i^2 = ss_E = s_{yy} - b^2 s_{xx} = s_{yy} - \frac{s_{xy}^2}{s_{xx}}$$

واریانس خطا $\sigma_e^2 = \frac{ss_E}{n-2}$

نکته ۲: برای پیش‌بینی کردن میانگین پاسخ به جای متغیر مستقل مقدار داده شده را قرار دهیم

$$E\left(y \middle| x = x^*\right) = y^* = \hat{a} + \hat{b}\,x^*$$

ارشد مدیریت۱۳۹۸

۵۶- قیمت (x) و تقاضای کالایی (y) در ۵ بازار مختلف جمع‌آوری شده است. شیب خط رگرسیون کدام است؟

x	۱۲	۱۰	۱۳	۱۶	۹
y	۲۰	۱۵	۱۱	۷	۲۲

۱) $-۱/۷$　　۲) $-۱/۶$　　۳) $-۱/۹$　　۴) $-۱/۴$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})\times(y-\bar{y})$
۱۲	۲۰	$۱۲-۱۲=۰$	$۲۰-۱۵=۵$	$(۰)^۲=۰$	$۰\times۵=۰$
۱۰	۱۵	$۱۰-۱۲=-۲$	$۱۵-۱۵=۰$	$(-۲)^۲=۴$	$-۲\times۰=۰$
۱۳	۱۱	$۱۳-۱۲=۱$	$۱۱-۱۵=-۴$	$(۱)^۲=۱$	$۱\times(-۴)=-۴$
۱۶	۷	$۱۶-۱۲=۴$	$۷-۱۵=-۸$	$(۴)^۲=۱۶$	$۴\times-۸=-۳۲$
۹	۲۲	$۹-۱۲=-۳$	$۲۲-۱۵=۷$	$(-۳)^۲=۹$	$-۳\times۷=-۲۱$
$\sum x=۶۰$	$\sum y=۷۵$	۰	۰	$S_{xx}=۴۰$	$S_{xy}=-۵۷$

$$\bar{x}=\frac{\sum x}{n}=\frac{۶۰}{۵}=۱۲ \quad ,\bar{y}=\frac{\sum y}{n}=\frac{۷۵}{۵}=۱۵$$

$$b=\frac{S_{xy}}{S_{xx}}=\frac{-۵۷}{۳۰}=-۱/۹$$

گزینه ۳ صحیح است

ارشد مدیریت ۱۴۰۰

۶۵- اگر معادله رگرسیونی دو متغیر x و y به‌صورت $y=۳x+۵۶$ حاصل شده باشد و به ازای مقدار $x=۱۰$ مقدار واقعی $y=۹۰$ باشد، کدام مورد دربارهٔ همبستگی دو متغیر درست است؟

۱) $۰<\rho<۱$

۲) $-۱<\rho<۰$

۳) $\rho=۱$

۴) $\rho=-۱$

$$y=۳x+۵۶$$

$$E\left(y|x=x^*\right)=y^*=\hat{a}+\hat{b}x \Rightarrow y^*=۳\times۱۰+۵۶=۸۶$$

گزینه ۱ صحیح است چون مقدار برآورد شده با مقدار واقعی برابر نیست و همچنین شیب خط ۳+ است بنابراین همبستگی عددی بین صفر و یک است.

ارشد مدیریت ۱۴۰۱

۶۵- براساس ۵ نقطهٔ (x, y) داریم: $\sum y^2 = ۴۵۱ ، \sum x^2 = ۱۴۵ ، \sum xy = ۱۹۵ ، \sum y = ۴۵ و \sum x = ۲۵$ شیب معادلهٔ خط رگرسیون چقدر است؟

۱) $-۱/۲۵$

۲) $-۱/۵$

۳) $۲/۷۵$

۴) $۱/۲۵$

$$\bar{x} = \frac{\sum x}{n} = \frac{۲۵}{۵} = ۵ \quad , \bar{y} = \frac{\sum y}{n} = \frac{۴۵}{۵} = ۹$$

$$s_{xy} = \sum xy - n\bar{x}\bar{y} = ۱۹۵ - (۵ \times ۵ \times ۹) = -۳۰$$

$$s_{xx} = \sum x^2 - n\bar{x}^2 = ۱۴۵ - ۵ \times ۵^2 = ۲۰$$

$$b = \frac{s_{xy}}{s_{xx}} = \frac{-۳۰}{۲۰} = -۱/۵$$

گزینه ۲ صحیح است

ارشد مدیریت ۱۴۰۲

۶۰- به‌منظور بررسی وجود رابطه خطی بین دو متغیر، ۱۶ نمونه تصادفی انتخاب شده است که براساس آنها، میانگین مجذورات تغییرات ناشی از خطا و آماره آزمون به‌ترتیب ۶ و ۲۱ به‌دست آمده است. چند درصد از تغییرات متغیر وابسته، به‌وسیله متغیر مستقل تبیین می‌شود؟

۱) ۳۶

۲) ۴۵

۳) ۶۰

۴) ۶۷

$$F = ۲۱ \quad SS_E = ۶ \quad n = ۱۶$$

$$F = \frac{R^2.(n-۲)}{۱-R^2} \Rightarrow ۲۱ = \frac{R^2.(۱۶-۲)}{۱-R^2} \Rightarrow \frac{۲۱}{۱۴} = \frac{R^2}{۱-R^2}$$

$$۱/۵ = \frac{R^2}{۱-R^2} \Rightarrow ۱/۵ - ۱/۵R^2 = R^2 \Rightarrow ۱/۵ = ۲/۵R^2 \Rightarrow R^2 = \frac{۱/۵}{۲/۵} = ۰/۶ = \%۶۰$$

گزینه ۳ صحیح است

پیش فرض های رگرسیون:

۱- بین متغیرهای پیش بین و ملاک رابطه خطی وجود داشته باشد. که با نمودار پراکنش یا ضریب همبستگی پیرسون

۲- متغیرهای مستقل و وابسته باید دارای مقیاس فاصله ای باشند

۳-واریانس خطا نرمال یا ثابت باشد با بررسی نمودار صورت می گیرد

۴-توزیع فراوانی باقیمانده ها نرمال باشد با آزمون کولموگروف اسمیرنوف

۵-باقیمانده ها از هم مستقل باشند تست دوربین واتسون (عدد نزدیک به ۲)

۶-بین متغیرهای مستقل رابطه همخطی وجود نداشته باشد با مقدار تلرانس یا VIF بررسی می شود

مثال: داده‌های زیر مربوط به نتایج یک بررسی در زمینه انجماد شتاب داده شده شمشها است که y عمق انجماد و x معرف زمان سرد شدن تعیین کنید الف) ضریب همبستگی ب) معادله خط رگرسیون ج) امید ریاضی عمق انجماد پس از ۵۰ دقیقه د) واریانس خطا

عمق انجماد(y)	۵	۶	۷/۵	۸/۵	۱۰	۱۲
زمان سرد شدن Y	۸	۱۴	۱۸/۵	۲۳/۵	۳۰	۳۸

جواب

x	Y	$X^{۲}$	$y^{۲}$	Xy
۸	۵	۶۴	۲۵	۴۰
۱۴	۶	۱۹۶	۳۶	۸۴
۱۸/۵	۷/۵	۳۴۲/۲۵	۵۶/۲۵	۱۳۸/۷۵
۲۳/۵	۸/۵	۵۵۲/۲۵	۷۲/۲۵	۱۹۹/۷۵
۳۰	۱۰	۹۰۰	۱۰۰	۳۰۰
۳۸	۱۲	۱۴۴۴	۱۴۴	۴۵۶
۱۳۲	۴۹	۳۴۹۸/۵	۴۳۳/۵	۱۲۱۸/۵

$$\bar{x}=\frac{\sum_{i=۱}^{n} x}{n}=\frac{۱۳۲}{۶}=۲۲$$

$$\bar{y}=\frac{\sum_{i=۱}^{n} y_i}{n}=\frac{۴۹}{۶}=۸/۱۷$$

$$s_{xy}=\sum_{i=۱}^{n} x_i y_i - n\bar{x}\bar{y}=۱۲۱۸/۵-۶\times۲۲\times۸/۱۷=۱۴۰/۰۶$$

$$s_{xx}=\sum_{i=۱}^{n} x_i^{۲} - n\bar{x}^{۲}=۳۴۹۸/۵-۶\times۲۲^{۲}=۵۹۴/۵$$

$$s_{yy}=\sum_{i=۱}^{n} y_i^{۲} - n\bar{y}^{۲}=۴۳۳/۵-۶\times۸/۱۷^{۲}=۳۳$$

الف) $r=\frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}=\frac{۱۴۰/۰۶}{\sqrt{۵۹۴/۵\times۳۳}}=\frac{۱۴۰/۰۶}{۱۴۰/۰۶۶}=۰/۹۹۹$

ب)
$$\hat{b}=\frac{s_{xy}}{s_{xx}}=\frac{۱۴۰/۰۶}{۵۹۴/۵}=۰/۲۳۵۶$$
$$\hat{a}=\bar{y}-b\bar{x}=۸/۱۷-۰/۲۳۵۶\times۲۲=۲/۹۸$$
$$\hat{y}=\hat{a}+\hat{b}x=۲/۹۸+۰/۲۳۵۶x$$

ج) $E\left(y|x=x^*\right)=\hat{a}+\hat{b}x^*=۲/۹۸+۰/۲۳۵۶\times۵۰=۱۴/۷۶$

د)
$$ss_E=s_{yy}-b^۲s_{xx}=۳۳-۰/۲۳۵۶^۲\times۵۹۴/۵=۰/۰۰۰۹$$
$$s^۲_{y|x}=\sigma_e^۲=\frac{ss_E}{n-۲}=\frac{۰/۰۰۰۹}{۶-۲}=۰/۰۰۰۲۲۵$$

برآورد فاصله‌ای برای شیب خط رگرسیون

$$\sigma_b^۲=\frac{\sigma_e^۲}{s_{xx}}=\frac{\sigma_e^۲}{\sum_{i=۱}^{n}(x_i-\bar{x})^۲}$$
$$B:\hat{b}\pm t_{\frac{\alpha}{۲},df}\sigma_b\Rightarrow\beta:\hat{b}\pm t_{\frac{\alpha}{۲},df}\frac{\sigma_e}{\sqrt{s_{xx}}}$$

برآورد فاصله‌ای برای عرض از مبداء

$$\sigma_a^۲=\sigma_e^۲\left(\frac{۱}{n}+\frac{\bar{x}^۲}{s_{xx}}\right)\Rightarrow\sigma_a^۲=\sigma_e^۲\left(\frac{۱}{n}+\frac{\bar{x}^۲}{\sum(x-\bar{x})^۲}\right)$$
$$a:\hat{a}\pm t_{\frac{\alpha}{۲},df}\cdot\sqrt{\sigma_e^۲\left(\frac{۱}{n}+\frac{\bar{x}^۲}{s_{xx}}\right)}\Rightarrow a:\hat{a}\pm t_{\frac{\alpha}{۲},df}\sigma_e\cdot\sqrt{\frac{۱}{n}+\frac{\bar{x}^۲}{s_{xx}}}$$

برآورد نقطه‌ای و فاصله‌ای برای میانگین به ازای مقدار مشخص x

$$E\left(y|x=x^*\right)=y^*=\hat{a}+\hat{b}x^*$$
$$y^*:\hat{a}+\hat{b}x^*\pm t_{\frac{\alpha}{۲},df}\sigma_e\sqrt{\frac{۱}{n}+\frac{\left(x^*-\bar{x}\right)^۲}{s_{xx}}}$$

فاصله اطمینان پیش‌بین برای مشاهدهٔ آتی از متغیر وابسته

$$y^{\cdot}:\hat{a}+\hat{b}x^{*}\pm t_{\frac{\alpha}{2},df}\sigma_e\sqrt{1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{s_{xx}}}$$

برآورد نقطه‌ای و فاصله‌ای برای متغیر مستقل x به ازای یک مشاهده از متغیر وابسته y

$$x'=\frac{y'-a}{b}$$

$$x'=\frac{y'-a}{b}\pm t_{\frac{\alpha}{2},df}.\frac{\sigma_e}{b}\sqrt{1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{s_{xx}}}$$

آزمون فرض برای شیب خط رگرسیون:

$$\begin{cases}H_{\cdot}:\beta=\beta_{\cdot}\\ H_1:\beta\neq\beta_{\cdot}\end{cases}\quad |t|>t_{\frac{\alpha}{2},df}\Rightarrow \text{رد می‌شود } H_{\cdot}$$

$$\begin{cases}H_{\cdot}:\beta\geq\beta_{\cdot}\\ H_1:\beta<\beta_{\cdot}\end{cases}\quad t<-t_{\alpha,df}\Rightarrow \text{رد می‌شود } H_{\cdot}$$

$$df=n-2$$
$$t=\frac{b-B_{\cdot}}{\frac{\sigma_e}{\sqrt{s_{xx}}}}$$

$$\begin{cases}H_{\cdot}:\beta\leq\beta_{\cdot}\\ H_1:\beta>\beta_{\cdot}\end{cases}\quad t>t_{\alpha,df}\Rightarrow \text{رد می‌شود } H_{\cdot}$$

آزمون فرض برای عرض از مبداء

$$\begin{cases}H_{\cdot}:A=A_{\cdot}\\ H_1:A\neq A_{\cdot}\end{cases}\quad |t|>t_{\frac{\alpha}{2},df}\Rightarrow \text{رد می‌شود } H_{\cdot}$$

$$\begin{cases}H_{\cdot}:A\geq A_{\cdot}\\ H_1:A<A_{\cdot}\end{cases}\quad t<-t_{\alpha,df}\Rightarrow \text{رد می‌شود } H_{\cdot}$$

$$df=n-2$$
$$t=\frac{a-A_{\cdot}}{\sigma_e\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{s_{xx}}}}$$

$$\begin{cases}H_{\cdot}:A\leq A_{\cdot}\\ H_1:A>A_{\cdot}\end{cases}\quad t>t_{\alpha,df}\Rightarrow \text{رد می‌شود } H_{\cdot}$$

مثال: به منظور به دست آوردن رسوب از یک ماده خاص y در یک محلول مفروض از معرف شیمیایی، x، استفاده می‌شود خلاصه اطلاعات به شرح زیر است.

$$\sum_{i=۱}^{۱۵} x_i = ۹ \quad , \quad \sum_{i=۱}^{۱۵} (x_i - \bar{x})^۲ = ۲۱$$
$$\sum_{i=۱}^{۱۵} y_i = ۱۳۵ \quad , \quad \sum_{i=۱}^{۱۵} (y_i - \bar{y})^۲ = ۱۱۹$$
$$\sum_{i=۱}^{۱۵} (x_i - \bar{x})(y_i - \bar{y}) = ۴۲$$

تعیین کنید.

الف) ضریب همبستگی

ب) معادله خط رگرسیون

پ) امید ریاضی مقدار رسوب به ازای $x = ۷$

ت) فاصله اطمینان ۹۵٪ برای امید ریاضی مقدار رسوب نظیر $x = ۷$

ج) فاصله پیش‌بینی ۹۵٪ مقدار رسوب نظیر $x = ۷$

د) آزمون فرض معنادار بودن ضریب همبستگی در سطح خطای ۵٪

ذ) آزمون فرض معنادار شیب خط در سطح خطای ۵٪

هـ) آزمون فرض معنادار بودن عرض از مبداء در سطح خطای (۵٪ $t_{۰/۰۲۵,۱۳} = ۲/۱۶$)

الف) $r = \dfrac{s_{xy}}{\sqrt{s_{xx} s_{yy}}} = \dfrac{۴۲}{\sqrt{۲۱ \times ۱۱۹}} = ۰/۸۴$

ب)
$$\hat{b} = \frac{s_{xy}}{s_{xx}} = \frac{۴۲}{۲۱} = ۲ \qquad \bar{x} = \frac{\sum_{i=۱}^{n} x_i}{n} = \frac{۹۰}{۱۵} = ۶ \qquad \bar{y} = \frac{\sum_{i=۱}^{n} y_i}{n} = \frac{۱۳۵}{۱۵} = ۹$$
$$\hat{a} = \bar{y} - \hat{b}\,\bar{x} = ۹ - ۲ \times ۶ = -۳$$
$$\hat{y} = \hat{a} + \hat{b}\,x = -۳ + ۲x$$

پ) $y^* = \hat{a} + \hat{b}\,x^* = -۳ + ۲ \times ۷ = ۱۱$

$$SS_E = s_{yy} - b^2 s_{xx} = ۱۱۹ - ۲^2 \times ۲۱ = ۳۵$$

$$s^2_{y|x} = \sigma_e^2 = \frac{SS_E}{n-۲} = \frac{۳۵}{۱۵-۲} = ۲/۶۹ \Rightarrow \sigma_e = ۱/۶۴$$

ت) $$y^* = \hat{a} + \hat{b}x^* \pm t_{\frac{\alpha}{۲},df}\,\sigma_e \sqrt{\frac{۱}{n} + \frac{(x^* - \bar{x})^2}{s_{xx}}}$$

$$y^* = (-۳ + ۲\times ۷) \pm ۲/۱۶ \times ۱/۶۴ \sqrt{\frac{۱}{۱۵} + \frac{(۷-۶)^2}{۲۱}} = ۱۱ \pm ۱/۱۹۷$$

$$۹/۸۰۳ < y^* < ۱۲/۱۹۷$$

$$y^{\cdot} = a + bx^{\cdot} \pm t_{\frac{\alpha}{۲},df}\,\sigma_e \sqrt{۱ + \frac{۱}{n} + \frac{(x_{\cdot} - \bar{x})^2}{s_{xx}}}$$

ج) $$y^{\cdot} = (-۳ + ۲\times ۷) \pm ۲/۱۶ \times ۱/۶۴ \sqrt{۱ + \frac{۱}{۱۵} + \frac{(۷-۶)^2}{۲۱}} = ۱۱ \pm ۳/۷۴$$

$$۷/۲۶ < y^{\cdot} < ۱۴/۷۴$$

د) $$\begin{cases} H_{\cdot} : \rho = ۰ \\ H_۱ : \rho \neq ۰ \end{cases} \qquad t = \frac{(r - \rho_{\cdot})\sqrt{n-۲}}{\sqrt{۱-r^2}} = \frac{۰/۸۴\sqrt{۱۵-۲}}{\sqrt{۱-۰/۸۴^2}} = ۵/۵۸۱۹$$

چون $t = ۵/۵۸۱۹ > t_{./۰۲۵,۱۳} = ۲/۱۶$ است بنابراین $H_{\cdot}$ رد می‌شود و ضریب همبستگی در سطح اطمینان ۹۵٪ معنی‌دار است.

ذ) $$\begin{cases} H_{\cdot} : \beta = ۰ \\ H_۱ : \beta \neq ۰ \end{cases} \qquad t = \frac{b - \beta_{\cdot}}{\frac{\sigma_e}{\sqrt{s_{xx}}}} = \frac{۲-۰}{\frac{۱/۶۴}{\sqrt{۲۱}}} = ۵/۵۸۸۵$$

چون $t = ۵/۵۸۸۵ > t_{./۰۲۵,۱۳} = ۲/۱۶$ است بنابراین $H_{\cdot}$ رد می‌شود و شیب خط در سطح اطمینان ۹۵٪ معنادار است.

هـ) $$\begin{cases} H_{\cdot} : A = ۰ \\ H_۱ : A \neq ۰ \end{cases} \qquad t = \frac{a - A_{\cdot}}{\sigma_e \sqrt{\frac{۱}{n} + \frac{x^2}{s_{xx}}}} = \frac{-۳-۰}{۱/۶۴\sqrt{\frac{۱}{۱۵} + \frac{۶۲}{۲۱}}} = -۱/۳۷$$

چون $|t| = ۱/۳۷ > t_{./۰۲۵,۱۳} = ۲/۱۶$ بنابراین $H_{\cdot}$ تایید می‌شود و عرض از مبداء در سطح اطمینان ۹۵٪ معنی‌دار نیست و برابر صفر است.

نکته: برای تعیین معادله خط رگرسیون ابتدا دیاگرام پراکنش را رسم می‌کنیم و مدل را تشخیص و با تغییر متغیر مناسب آن به مدل خطی تبدیل می‌کنیم به عنوان مثال اگر تشخیص دهیم رابطهٔ بین x و y نمایی است با لگاریتم گرفتن از طرفین به مدل خطی تبدیل می‌شود و مدل بدست می‌آوریم.

$$y = ab^{*} \Rightarrow Lny = Lna = xLnb \qquad y^{*} = a^{*} + b^{*}x$$

$$\Rightarrow a^{*} = e^{a^{*}} \quad , \quad b^{*} = e^{b^{*}} \Rightarrow y\left(e^{a^{*}}\right).\left(e^{b^{*}}\right)^{x}$$

نکته: همیشه یک تابع خطی و درجه یک برحسب x، برای y وجود دارد یعنی:

$$(x,y) \sim \left(\mu_X, \mu_y, \sigma^{2}x, \sigma^{2}y, P\right)$$

$$\Rightarrow y - \mu_y = \rho\frac{\sigma_y}{\sigma_x}(x - \mu_X)$$

نکته: در خروجی SP_{SS} در ستون Sigt در آزمون فرض‌ها می‌توان بدون داشتن جدول t با مقایسهٔ عدد Sigt و α در مورد B_j تصمیم‌گیری کرد یعنی اگر این مقدار احتمال کوچکتر یا مساوی α باشد فرض صفر را رد می‌کنیم و $B_j \neq ۰$ است.

مثال : داده های مقابل مفروض است اگر $VaR(e_i) = \sigma^{2}, E(e_i) = ۰$ باشد رگرسیون خطی Y بر حسب X کدام است؟

X	Y
۲	۱
۰	۳
۴	۵

۱) $y = x + ۱$ ۲) $y = \frac{۱}{۲}(x + ۱)$ ۳) $y = \frac{۱}{۲}(x - ۱)$ ۴) $y = -\frac{۱}{۲}(x - ۱)$

جواب: گزینه ۲ صحیح است

x	y	xy	x^{2}	y^{2}
۱	۲	۲	۱	۴
۳	۰	۰	۹	۰
۵	۴	۲۰	۲۵	۱۶
$\sum x = ۹$	$\sum y = ۶$	$\sum xy = ۲۴$	$\sum x^{۲} = ۳۵$	$\sum y^{۲} = ۲۰$

۱) $b_{xy} = \frac{n\sum xy - (\sum x.\sum y)}{n\sum x^{۲} - (\sum x)^{۲}} = \frac{(۳\times ۲۲) - (۹\times ۶)}{۳\times ۳۵ - ۹^{۲}} = \frac{۶۶ - ۵۴}{۲۴} = \frac{۱۲}{۲۴} = \frac{۱}{۲}$

$$\bar{x}=\frac{\sum x}{n}=\frac{۹}{۳}=۳\ ,\ \bar{y}=\frac{\sum y}{n}=\frac{۶}{۳}=۲$$

$$۲)a=\bar{y}-b\bar{x}=۲-(\frac{۱}{۲}\times ۳)=\frac{۴}{۲}-\frac{۳}{۲}=\frac{۱}{۲}$$

$$۳)y=a+bx=\frac{۱}{۲}+\frac{۱}{۲}x=\frac{۱}{۲}(۱+x)$$

مثال: اصل حداقل مجذورات (LS) تفاوت کدام دو نمره را کمینه می سازد؟

۱) ε, y ۲) X, ε

۳) $\hat{y}, \varepsilon$ ۴) $\hat{y}, \varepsilon$

گزینه ۴ صحیح است چون $e_i = y_i - \hat{y}_i$ **خطا باید حداقل شود**

نکته ۷: خط رگرسیون از نقطه $\bar{x}$ **و** $\bar{y}$ **می‌گذرد.**

نکته ۸: رابطه شیب خط رگرسیون و ضریب همبستگی $b = r.\frac{S_y}{S_x}$

با توجه به نکته ۱و۲ نتیجه می گیریم $(y-\bar{y}) = r.\frac{S_y}{S_x}(x-\bar{x})$

نکته۹: رابطه واریانس خطای پیش بینی (استاندارد) و ضریب همبستگی $S_{xy} = S_y\sqrt{1-r_{xy}^2}$

نکته ۱۰: ضریب رگرسیون تفکیکی استاندارد شده که معیاری برای تعیین مقدار تاثیر متغیر پیش بین روی متغیرملاک است و عددی بین ۱- و۱+ است $\beta = \frac{S_x}{S_y}.b$

توجه۱ : در رگرسیون یک متغیره $df = n-2$ **و در حالت کلی** $df = n-k-1$ **است که k تعداد متغیر مستقل می باشد**

توجه۲: در صورتی‌که بخواهیم معادله خط x برحسب y را بنویسیم به ترتیب زیر عمل می‌کنیم.

۱) $\hat{b}^* = \frac{s_{xy}}{s_{yy}}$ ۲) $\hat{a}^* = x - \hat{b}^* y$ ۳) $\hat{x} = \hat{a}^* + \hat{b}^* y$

نکته۱۱: ضریب همبستگی واسطهٔ هندسی b و b^* **است یعنی** $r = \sqrt{b.b^*}$

مثال: اگر $r_{xy} = ۰/۵$ **باشد و تمام نمرات دو متغیر XوY به صورت نمره های استاندارد باشند عرض از مبدا و شیب خط رگرسیون برابر چند است؟**

۱) $۱-۰$ ۲) $۰/۵-۱$ ۳) $۰/۵-۰$ ۴) $۱-۰/۵$

گزینه۳ صحیح است چون نرمال استاندارد است بنابر این میانگین صفر و واریانس یک است

$$s_x = s_y = ۱ \Rightarrow b = r\frac{s_y}{s_x} \Rightarrow b = ۰/۵ \times \frac{۱}{۱} = ۰/۵ \Rightarrow a = \bar{y} - b\ \bar{x} = ۰ - (۰/۵ \times ۰) = ۰$$

نکته : $E(ax+by) = aE(x) + bE(y)$ و $\text{var}(ax+by) = \text{var}(x) + \text{var}(y) + ۲a.b.\text{cov}(x,y)$

مثال: اگر $\sigma_x^۲ = ۲ , \sigma_y^۲ = ۳$ **,** $\sigma_{xy} = ۱$ **باشد . حاصل** $\sigma_{x+y}^۲$ **چند است ؟**

۱) ۳　　۲) ۵　　۳) ۶　　۴) ۷

گزینه ۴ صحیح است. $\text{var}(ax+by) = \text{var}(x) + \text{var}(y) + ۲a.b.\text{cov}(x,y) = ۲ + ۳ + ۲ \times ۱ = ۷$

متغیر تعدیل‌کننده : متغیری است که به صورت مستقیم بر جهت یا میزان رابطه متغیرهای مستقل و وابسته می‌تواند موثر باشد. اثرات این متغیر قابل مشاهده و اندازه‌گیری است. مانند رابطه هوش با افزایش میزان یاد گیری دانش آموزان یا متغیر جنسیت در بررسی رابطه ی روش تدریس و یادگیری دانشجویان یک متغیر تعدیل کننده است

خطای پیش بینی : اختلاف مقدار واقعی (y) و مقدار پیش بینی را خطای پیش بینی گویند و با $\hat{y}$ **نشان میدهیم**

ضریب رگرسیون تفکیکی استاندار د شده : اگر به جای نمره های خام از داد های استاندارد شده z استفاده کنیم از ضریب رگرسیون تفکیکی استاندار د شده استفاه میکنیم و با حرف β **نشان میدهیم.**

و از رابطه $\beta = b\frac{s_y}{s_x}$ **بدست می آید. که در آن b شیب خط رگرسیون و** s_x **انحراف معیار متغیر پیش بین (x) و** s_y **انحراف معیار متغیر ملاک (y)است**

$$s_y = \sqrt{\frac{\sum y^۲ - \frac{(\sum y)^۲}{n}}{n}} \quad , \quad s_x = \sqrt{\frac{\sum x^۲ - \frac{(\sum x)^۲}{n}}{n}}$$

مثال : در صورتی که شیب خط رگرسیون ۴ و انحراف معیار متغیر پیش بین ۱۰ و انحراف معیار متغیر ملاک ۵ است. ضریب رگرسیون تفکیکی استاندار د شده کدام است ؟

الف)۱　　ب)۲　　ج)۳　　د)۴

جواب گزینه ۲ صحیح است.

$$\beta = b\frac{s_y}{s_x} = ۴ \times \frac{۵}{۱۰} = ۲$$

آزمون خطی بودن (Lack of fit): در صورتی‌که به ازای یک x چند y وجود داشته باشد از آزمون خطی یا Lack of fit استفاده می‌شود.

$$\begin{cases} H_۰ : y = a + bx \\ H_۱ : y \neq a + bx \end{cases}$$

که در این حالت متوسط y مقادیر x_i **را با** $\bar{y}_i$ **نشان می‌دهیم.**

$$\begin{array}{lll} x_1 \to y_{11} & x_2 \to y_{21} & x_n \to y_{n1} \\ \quad y_{12} & \quad y_{22} & \quad y_{n2} \\ \quad \vdots & \quad \vdots & \quad \vdots \\ \quad y_1 k_1 & \quad y_2 k_2 & \quad y_n x_n \\ \quad \bar{\bar{y}}_1 & \quad \bar{\bar{y}}_2 & \quad \bar{\bar{y}}_n \end{array}$$

$$SS_E = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\left(y_{ij} - \hat{y}_i\right)^2$$

$$SS_{PE} = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\left(y_{ij} - \hat{y}_i\right)^2$$

$$SS_{Lof} = \sum_{i=1}^{n} k_i\left(\bar{y}_i - \hat{y}_i\right)^2 \quad , \quad SS_{Lof} = SS_E - SS_{PE}$$

که جدول آنالیز واریانس آن به صورت زیر است.

منبع تغییرات	SS	df	$Ms_R = \frac{ss}{df}$	F
رگرسیون	SS_R	۱	$MS_R = SS_R$	$F = \frac{MS_{Lof}}{MS_{PE}}$
Lof	SS_{Lof}	$n-2$	$MS_{Lof} = \frac{SS_{Lof}}{n-2}$	
PE	SS_{PE}	$N-n$	$MS_{PE} = \frac{SS_{PE}}{N-n}$	
کل	S_{yy}	$N-1$		

اگر $F > F_{\alpha,n-2,N-n}$ باشد H_0 رد می‌شود و آزمون خطی بودن مردود است و در حالت خاص اگر $k_1 = k_2 = ...k_n$ باشد در این صورت رابطه **F** را می‌توان به صورت زیر نوشت

$$F = \frac{k\sum_{i=1}^{n}\frac{\left(\bar{y}_i - \hat{y}_i\right)^2}{n-2}}{\sum_{i=1}^{n}\sum_{j=1}^{k}\frac{\left(\bar{y}_j - \hat{y}_i\right)^2}{n(k-1)}}$$

که در این صورت اگر $F > F_{\alpha,n-2,n(k-1)}$ باشد فرض H_0 رد می‌شود.

شرط کافی برای آن که دو متغیر تصادفی **x** و **y** دارای توزیع نرمال باشند آن است که دو متغیر رابطه خطی داشته باشند.

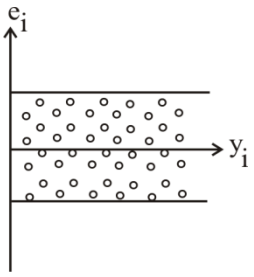
$$f_{xy}(u,v)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-P^2}}\,ExP\left[-\left[\frac{\frac{(u-\mu_x)^2}{\sigma^2}-2\rho\frac{(u-\mu_x)(v-\mu_x)}{\sigma_x\sigma_y}+\frac{(v-\mu_y)^2}{\sigma_y^2}}{2-\left(1-\rho^2\right)}\right]\right]$$

که در این صورت $E(y|x)=\mu_x+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\quad,\quad \sigma^2=\sigma_y^2\left(1-\rho^2\right)$

تشخیص در تحلیل رگرسیون از روی نمودارها:

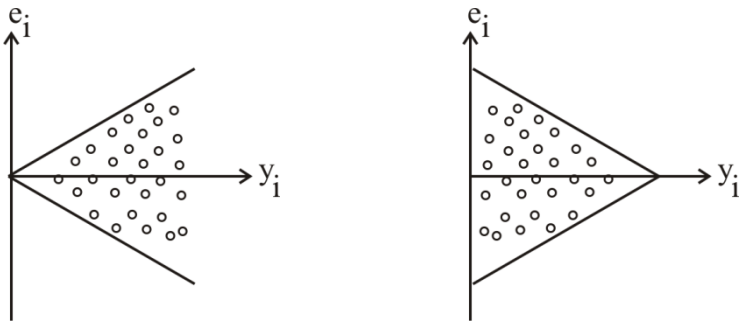
در صورتی‌که $\left(\hat{y}_i, e_i\right)$ را رسم کنیم حالات زیر امکان دارد رخ دهد.

۱) داده ها بین دو خط موازی قرار گیرد در این صورت

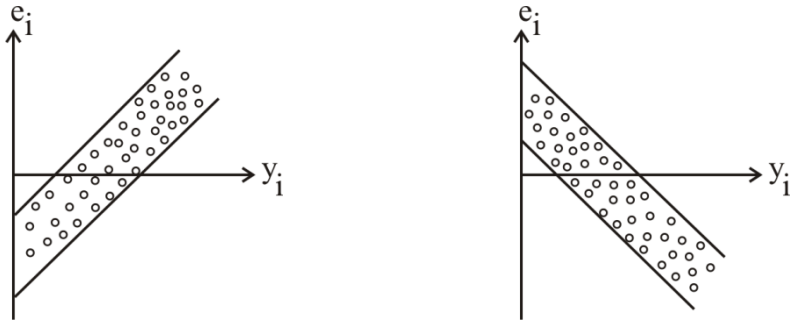


فرضیه‌های مدل برقرار است $\Rightarrow e_i=0 \Rightarrow \hat{y}_i$

۲) داده ها به صورت دوکی باشد در این صورت واریانس خطا ثابت نیست



۳) داده بین دو خط موازی شیب دار باشد در این صورت مدل باید بازسازی شود



تحلیل پس مانده‌ها:

همانگونه که می‌دانیم $e_i \sim N\left(0, \sigma_e^2\right)$ **می‌باشد بنابراین** $\frac{e_i}{\sigma}$ **دارای توزیع نرمال با میانگین صفر و واریانس یک است در صورتی‌که** $d_i = \frac{e_i}{\sqrt{MSE}}$ **باشد اگر** $P\left(-z_{\frac{\alpha}{2}} < d_i < z_{\frac{\alpha}{2}}\right)$ **باشد نرمال بودن را می‌پذیریم یعنی اگر** $P\left(-1/96 < d_i < 1/96\right) = 0/95$ **باشد با احتمال ۹۵٪ نرمال بودن را می‌پذیریم**

رگرسیون چند متغیره:

مدل رگرسیونی که بیش از یک متغیر مستقل داشته باشد رگرسیون چند متغیره گویند که مدل آن به صورت $y = B_0 + B_1x_1 + B_2x_2 + B_3x_3 + ...B_kx_k + \varepsilon$ **می‌باشد که برای برآورد پارامترها آن به فرم ماتریسی نمایش می‌دهیم.**

$$e = y - y' \Rightarrow SS_{res} = \sum (y - y')^2 = \sum e^2 \Rightarrow SS_{res} = \sum y'^2 - \frac{(\sum y)^2}{n}$$

$$B = \begin{bmatrix} B_0 \\ B_1 \\ \vdots \\ B_k \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{bmatrix} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad x = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{1k} \\ 1 & x_{21} & x_{22} & x_{2k} \\ . & & & \\ . & & & \\ 1 & x_{n1} & x_{n2} & x_{nk} \end{bmatrix}_{n*k}$$

k: تعداد متغیرهای مستقل و n: تعداد داده‌ها

لازم به توضیح است که در ماتریس x اگر مقدار ثابت B_0 **در ماتریس B در نظر گرفته نشود عدد یک را در ستون اول نمی‌نویسیم که پارامترها به روش کمترین مربعات به صورت زیر محاسبه می‌شود.**

$$\hat{b} = (x'x)^{-1}x'y$$

$$cov(b) = E\left(\left[b - E(b)\right] * \left[b - E(b)\right]'\right) \Rightarrow cov(b) = \sigma^2 (x'x)^{-1}$$

$$SS_E = S_{yy} - SS_R = y'y - b'x'y$$

$$\sigma^2 = MSE = \frac{SS_E}{n-p}, p = k+1$$

ماتریس $(X'X)^{-1}$ **یک ماتریس متقارن است اگر اعضای این ماتریس را به صورت** c_{ij} **نمایش دهیم آنگاه واریانس** $\hat{B_i}$ **از رابطه**

$$(x'x)^{-1}=\begin{bmatrix} c_{00} & c_{01} & c_{02} & c_{0k} \\ c_{10} & c_{11} & c_{12} & c_{1k} \\ \vdots & & c_{32} & c_{2k} \\ \vdots & & & \\ c_{k0} & c_{k1} & c_{k2} & c_{kk} \end{bmatrix}$$

$$\hat{\tilde{y}} = xb = x(x'x)^{-1}xy = Hy$$

ماتریس $H = x(x'x)^{-1}x$ یک ماتریس خود توان است.

آزمون فرض ضرایب انفرادی رگرسیون

$$\begin{cases} H_0 : \beta_j = \beta_j \\ H_1 : \beta_j \neq \beta_j \end{cases} \quad |t| > t_{\frac{\alpha}{2},df} \Rightarrow \text{رد می‌شود } H_0$$

$$\begin{cases} H_0 : \beta_j \geq \beta_j \\ H_1 : \beta_j < \beta_j \end{cases} \quad t < -t_{\alpha,df} \Rightarrow \text{رد می‌شود } H_0$$

$$\begin{cases} H_0 : \beta_j \leq \beta_j \\ H_1 : \beta_j > \beta_j \end{cases} \quad t > t_{\alpha,df} \Rightarrow \text{رد می‌شود } H_0$$

$$df = n - 2$$

$$t = \frac{b_j - \beta_{j0}}{\sqrt{c_{jj} \times \hat{\sigma}^2}}$$

برآورد فاصله ضرایب رگرسیون:

$$B_j : b_j \pm t_{\frac{\alpha}{2},df}\sqrt{\hat{\sigma}^2 c_{jj}}$$

برآورد میانگین پاسخ و برآورد فاصله‌ای میانگین پاسخ:

$$\tilde{y}^* = x^{*\prime} b$$

$$\Rightarrow E(y^*) : \tilde{y}^* \pm t_{\frac{\alpha}{2},df}\sqrt{\hat{\sigma}^2 x^{*\prime}(x'x)^{-1}x^*}$$

پیش‌بینی مشاهدات جدید و فاصلهٔ اطمینان پیش‌بینی جدید:

$$\tilde{y}^{\cdot} = x^{\cdot\prime} b$$

$$\Rightarrow y^{\cdot} : \tilde{y}^{\cdot} \pm t_{\frac{\alpha}{2},df}\sqrt{\hat{\sigma}^2\left(1 + x^{\cdot\prime}(x'x)^{-1}x^{\cdot}\right)}$$

آزمون معنادار بودن رگرسیون: با استفاده از تحلیل واریانس این آزمون را انجام می‌دهیم

منبع تغییرپذیری SOV	مجموع مربعات SS	درجات آزادی df	میانگین مربعات $MS=\frac{SS}{df}$	آمارهٔ آزمون F
رگرسیون	SS_R	K	$MS_R=\frac{SS_R}{k}$	$F=\frac{MS_R}{MS_E}$
خطا	SS_E	$n-k-1=n-P$	$MS_E=\frac{SS_E}{n-P}=\hat{\sigma}^2$	
کل	$SS_T=S_{yy}$	$n-1$		

در صورتی‌که $F>F_{\alpha,k-2,n-P}$ باشد H_0 رد می‌شود و حداقل یکی از متغیرها نقش قابل توجهی دارد که در آن:

$$S_{yy}=y'y-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}=\sum_{i=1}^{n}y_i^2-n\bar{y}^2$$

$$SS_R=b'x'y-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}=b'x'y-n\bar{y}^2$$

$$SS_E=S_{yy}-SS_R=y'y-b'x'y$$

$$R^2=\frac{SS_R}{S_{yy}}=\frac{S_{yy}-SS_E}{S_{yy}}=1-\frac{SS_E}{S_{yy}}$$

رابطهٔ F با ضریب تعیین

$$F=\frac{MS_R}{MS_E}=\frac{\frac{SS_R}{k}}{\frac{SS_E}{n-k-1}}=\frac{(n-k-1)\frac{SS_R}{S_{yy}}}{k.\frac{SS_E}{S_{yy}}}=\frac{n-k-1}{k}=\frac{R^2}{1-R^2}$$

$$\Rightarrow F=\frac{n-k-1}{k}.\frac{R^2}{1-R^2}=\frac{n-p}{k}.\frac{R^2}{1-R^2}$$

نکته: MS_R و MS_E یک برآوردگر ناریب برای σ^2 هستند

ضریب تعیین اصلاح شده $\left(\bar{R}^{۲}\right)$:

$$\bar{R}^{۲}=۱-\frac{n-۱}{n-P}\left(۱-R^{۲}\right)$$

$$\bar{R}^{۲}=۱-\frac{n-۱}{S_{yy}}MS_E \quad , \quad MS_E=\frac{SS_E}{n-P} \quad ,\bar{R}^{۲}=۱-\frac{\sum\frac{\left(y-\hat{y}\right)^{۲}}{n-P}}{\sum\frac{\left(y-\hat{y}\right)^{۲}}{n-۱}}$$

مثال: در مدل خطی $y=B_۱X_۱+B_۲X_۲+B_۳X_۳+\varepsilon$ تعیین کنید:

۱) ضرایب خط رگرسیون　　۲) واریانس خطا　　۳) آزمون فرض $B_۱=۰$ در سطح خطای ۵٪

۴) جدول آنالیز واریانس

y_i	$x_{i۱}$	$x_{i۲}$	$x_{i۳}$
۱	۰	۰	۱
۲	۱	۱	۰
-۱	۱	۰	۰
۰	۱	-۱	۰
۰	۰	۰	۰

۱) جواب: چون $B_۰$ در مدل وجود ندارد در ستون اول **X** عدد یک وجود ندارد.

$$X=\begin{bmatrix}۰ & ۰ & ۱\\ ۱ & ۱ & ۰\\ ۱ & ۰ & ۰\\ ۱ & -۱ & ۰\\ ۰ & ۰ & ۰\end{bmatrix}\Rightarrow X'X=\begin{bmatrix}۳ & ۰ & ۰\\ ۰ & ۲ & ۰\\ ۰ & ۰ & ۱\end{bmatrix}\Rightarrow (X'X)^{-۱}=\begin{bmatrix}۱/۳ & ۰ & ۰\\ ۰ & ۱/۲ & ۰\\ ۰ & ۰ & ۱\end{bmatrix}$$

$$x'y=\begin{bmatrix}۰ & ۱ & ۱ & ۱ & ۰\\ ۱ & ۱ & ۰ & -۱ & ۰\\ ۱ & ۰ & ۰ & ۰ & ۰\end{bmatrix}\begin{bmatrix}۱\\ ۲\\ -۱\\ ۰\\ ۰\end{bmatrix}=\begin{bmatrix}۱\\ ۲\\ ۱\end{bmatrix}$$

$$\hat{B}=(x'x)^{-۱}x'y=\begin{bmatrix}۱/۳ & ۰ & ۰\\ ۰ & ۱/۲ & ۰\\ ۰ & ۰ & ۱\end{bmatrix}\begin{bmatrix}۱\\ ۲\\ ۱\end{bmatrix}=\begin{bmatrix}۱/۳\\ ۱\\ ۱\end{bmatrix}=\begin{bmatrix}B_۱\\ B_۲\\ B_۳\end{bmatrix}$$

(۲

$$y'y=\sum y_i^۲=۶$$

$$b'x'y=\begin{bmatrix}۱/۳ & ۱ & ۱\end{bmatrix}\begin{bmatrix}۱\\ ۲\\ ۱\end{bmatrix}=\frac{۱۰}{۳}$$

$$SSE=y'y-b'x'y=۶-\frac{۱۰}{۳}=۸/۳$$

$$\sigma^۲=\frac{SSE}{n-p}=\frac{۸/۳}{۵-(۳+۱)}=۸/۳$$

(۳

$$t=\frac{b_۱}{\sqrt{\sigma^۲Cjj}}=\frac{۱/۳}{\sqrt{۸/۳\times ۱/۳}}=۰/۳۵$$

چون $t=۰/۳۵ \triangleright t_{./۰۲۵,۱}=۱۲/۷۰۶$، بنابراین $H_۰$ پذیرفته می‌شود و $B_۱$ با احتمال ۹۵٪ برابر صفر است.

(۴

منبع تغییرات	ss	df	$Ms=\frac{ss}{df}$	F
رگرسیون	$SS_R=\frac{۷}{۳}$	۳	$\frac{۷}{۹}$	$F=\frac{۷/۹}{۸/۳}=\frac{۷}{۲۴}$
خطا	$SS_R=\frac{۸}{۳}$	۱	$\frac{۸}{۲}$	
کل	$SS=۵$	۴		

آزمون معنادار بودن رگرسیون: با استفاده از جدول **ANOVA** رگرسیون این آزمون را انجام می‌دهیم

منبع تغییرپذیری SOV	مجموع مربعات SS	درجات آزادی df	میانگین مربعات $MS=\frac{SS}{df}$	آمارهٔ آزمون F
رگرسیون	SS_{reg}	k	$MS_{reg}=\frac{SS_{reg}}{k}$	$F_{م}=\frac{MSreg}{MS_{res}}$
خطا	SS_{res}	$n-k-1$	$MS_{res}=\frac{SS_{res}}{n-k-1}=\hat{\sigma}^2$	
کل	$S_{yy}=SS_T$	$n-1$		

اگر $F_{م} \succ F_{\alpha,k,N-K-1}$ فرض صفر رد میشود و حداقل یکی از شیب خط ها معنی دار است

نکته۱۲ :مجذور اتا(η^2) شدت دو متغیر وقتی متغیر وابسته کمی و متغیر مستقل کیفی باشد را نشان میدهد

نکته۱۳: میزان تغیرات **y**نسبت به **x**را ضریب تعیین ($V=R^2$)مشخص میکند.

ضریب بیگانگی $=1-R^2$ و $R^2=\frac{SS_{reg}}{SS_T}=1-\frac{SS_{res}}{SS_T}$

نکته۱۴ : به منظور بررسی مسئله چند گانگی خطی از دو آماره رواداری یا تحمل و عامل تورم واریانس (**VIF**) استفاده میشود و رابطه $VIF=\frac{1}{1-R^2}$ بدست می آید

مثال :نتایج حاصل از سه متغیر پیش بین با تعداد ۱۰عدد در جدول آنوا زیر ثبت شده . جدول را کامل کنید و توضیح دهید فرض صفر تایید میشود؟ ضریب تعیین ضریب بیگانگی را بدست آورید.

منبع تغییرپذیری SOV	مجموع مجذورات SS	درجات آزادی df	میانگین مربعات $MS=\frac{SS}{df}$	آمارهٔ آزمون F
رگرسیون	۲۰			
خطا				
کل	۵۰			

جواب:

منبع تغییرپذیری SOV	مجموع مربعات SS	درجات آزادی df	میانگین مربعات $MS=\frac{SS}{df}$	آمارهٔ آزمون F
رگرسیون	۲۰	$K=3$	$MS_{reg}=\frac{20}{3}=6/67$	$F_{م}=\frac{MSreg}{MS_{res}}$
خطا	۳۰=۲۰-۵۰	۶	$MS_{res}=\frac{30}{6}=5$	$F_{م}=\frac{6/7}{5}=1/34$
کل	۵۰	$n-1=10-1=9$		

چون $F_{م} = ۱/۳۴ \prec F_{./۰۵,۳,۶} = ۴/۷$ فرض صفر تایید میشود و از شیب خط ها معنی دار نیست با احتمال ۹۵٪

ضریب بیگانگی $= ۱ - R^{۲} = ۱ - ۰/۴ = ۰/۶$ $R^{۲} = \frac{SS_{reg}}{SS_T} = \frac{۲۰}{۵۰} = ۰/۴$

فرض های رگرسیون چند متغیری :

۱- نمره ها مستقل باشند ۲- نمره های موجود در هر خانه از جامعه ای انتخاب شده باشند

۳- خطی باشند برای این کار تمام متغیرهای مستقل دیگر را ثابت نگه میدارند

محا سبات رگرسیون چند متغیری:

$$۱) b_۱ = \frac{r_{y۱} - r_{y۲} r_{۱۲}}{۱ - r_{۱۲}^{۲}} \left(\frac{S_y}{S_۱} \right)$$

$$b_۲ = \frac{r_{y۲} - r_{y۱} r_{۱۲}}{۱ - r_{۱۲}^{۲}} \left(\frac{S_y}{S_۲} \right)$$

$$۲) a = \bar{y} - b_۱ \bar{x}_۱ - b_۲ \bar{x}_۲$$

$$۳) y = a + b_۱ x_۱ + b_۲ x_۲ + ...$$

ضرایب استاندارد شده:

$$\beta_۱ = b_۱ \frac{S_y}{S_۱} \quad , \quad \beta_۲ = b_۲ \frac{S_y}{S_۲}$$

مجذور همبستگی (ضریب تعیین): $R^{۲} = \frac{r_{y۱}^{۲} + r_{y۲}^{۲} - ۲ r_{y۱} r_{y۲} r_{۱۲}}{۱ - r_{۱۲}^{۲}}$

ضریب تعیین تعدیل شده : در نتایج پژوهش ها درج می شود چون تاثیر حجم نمونه در آن نشان داده شده است

$$R_{تعدیل}^{۲} = ۱ - \left[(۱ - R^{۲}) . (\frac{N - ۱}{N - K - ۱}) \right]$$

مثال : اگردر نتایج حاصل از سه متغیر پیش بین با تعداد ۱۰عدد ضریب تعیین ۰/۸ باشد ضریب تعیین تعدیل شده را بدست آورید.

$$N = ۱۰, K = ۳, R^{۲} = ۰/۸$$

$$R_{تعدیل}^{۲} = ۱ - \left[(۱ - R^{۲}) . (\frac{N - ۱}{N - K - ۱}) \right] = ۱ - \left[(۱ - ۰/۸) . (\frac{۱۰ - ۱}{۱۰ - ۳ - ۱}) \right] = ۱ - ۰/۰۳ = ۰/۹۷$$

آماره آزمون F : $F=\frac{n-k-1}{k}.\frac{R^2}{1-R^2}=\frac{n-p}{k}.\frac{R^2}{1-R^2}$ یا $F=\frac{\frac{R^2}{K}}{\frac{1-R^2}{N-K-1}}$

مثال : در سوال قبل آماره آزمون را بدست آورید

$$F=\frac{\frac{R^2}{K}}{\frac{1-R^2}{N-K-1}}=\frac{\frac{0/8}{3}}{\frac{1-0/8}{10-3-1}}=\frac{\frac{0/8}{3}}{\frac{0/2}{6}}=\frac{4/8}{0/6}=8$$

چون $F_{0/05,3,6}=4/7 \prec F_{m}=8$ فرض صفر رد میشود و حداقل یکی از شیب خط ها معنی دار است با احتمال ۹۵٪

نکته : به منظور بررسی مسئله چند گانگی خطی از دو آماره رواداری یا تحمل و عامل تورم واریانس (VIF) استفاده میشود

انواع روش های ورود داده ها در رگرسیون :

۱-همزمان(Enter)　　۲- گام به گام (Stepwise)　　۳- روش حذف (Remove)

۴-روش پس رونده (Backward)　　۵-پیش رونده (Forward)

رگرسیون سلسله مراتبی : رگرسیون سلسله مراتبی راهی است برای نشان دادن اینکه آیا کمیت‌های مورد علاقه شما پس از قرار گرفتن سایر کمیت‌ها در مدل، مقدار قابل توجهی از واریانس کمیت وابسته وابسته) منظور همان ضریب تعیین R^2 است (را توضیح می‌دهند یا خیر. در این روش، چندین مدل رگرسیون را با اضافه کردن کمیت‌هایی به مدل قبلی در هر مرحله می‌سازید. رگرسیون سلسله مراتبی (Hierarchical regression) نشان می‌دهد متغیرهای پیش‌بین پس از ورود متغیرهای دیگر به صورت معنادار تغییرات متغیر وابسته را توجیه می‌کنند. این چارچوب بیشتر برای مدل‌های مقایسه‌ای مناسب است تا روش‌های آماری. یکی از کاربردهای اساسی این روش در حل مسائل متغیر تعدیل کننده است .به منظور بررسی نقش تعدیل کنندگی یک متغیر مثلا z در رابطه بین دو متغیر x و y از رگرسیون سلسله مراتبی استفاده می شود. رگرسیون سلسله‌مراتبی یا ترتیبی این امکان را فراهم می‌آورد که تاثیر چند متغیر مستقل بر یک متغیر وابسته طی چند مرحله مشخص شود.

تحلیل مسیر (Path analysis): تحلیل مسیر تکنیکی آماری است که با استفاده از معادلات رگرسیون خطی استاندارد (معادلات رگرسیون خطی بر حسب ضرایب رگرسیون استاندارد) به میزان انطباق یک مدل علی نظری با واقعیت (داده‌ها) می‌پردازد. در این روش از ضریب بتای استاندارد رگرسیون جهت تعیین جهت و شدت روابط میان متغیرها استفاده می‌شود. مقدار آماره تی نیز معناداری روابط را نشان می‌دهد. به عبارت دیگر تحلیل مسیر تکنیکی برای آزمون تجربی مدل علی نظری است. علاوه بر این اگر مدل علی نظری با داده‌های جمعیتی معین انطباق داشته باشد می‌توان انواع اثرات (اثر مستقیم و غیرمستقیم و کاذب و خالص) تک تک متغیرهای مستقل بر متغیر وابسته مدل علی نظری را نیز حساب

کرد. هدف تحلیل مسیر به دست آوردن برآوردهای کمی روابط علی (همکنشی یکجانبه یا کواریته) بین مجموعه ای از متغیرهاست. ساختن یک مدل علی لزوماً به معنای وجود روابط علی در بین متغیرهای مدل نیست بلکه این علیت بر اساس مفروضات همبستگی و نظر و پیشینه تحقیق استوار است. اغلب برای انجام محاسبات مربوط به تحلیل مسیر می‌توان از نرم‌افزارهای Amos یا Lisrel یا SPSS استفاده کرد. تحلیل مسیر جهت و شدت روابط متغیرهای تحقیق را نشان می‌دهد. مقادیری که جهت و میزان تاثیر میان متغیرها را نشان می‌دهند ضریب مسیر نامیده می‌شوند ضرایب مسیر همان ضریب استاندارد شده (β) رگرسیون هستند. بنابراین برای تحلیل مسیر باید از رگرسیون خطی ساده استفاده شود. تحلیل مسیر تنها بر روی متغیرهای قابل مشاهده انجام پذیر است و اگر بخواهید بین ابعاد تحلیل مسیر را اجرا کنید باید میانگین سوالات هر بعد را حساب کنید تا متغیر پنهان به یک متغیر قابل مشاهده تبدیل شود.

فصل هفتم

تحلیل واریانس

تحلیل واریانس: هرگاه میانگین بیش از دو جامعه را بخواهیم مقایسه کنیم به جای آزمون t از تحلیل واریانس استفاده می‌کنیم زیرا برای مقایسه K جامعه با یکدیگر باید $\binom{k}{۲}=\frac{k(k-۱)}{۲}$ آزمون t انجام دهیم .

تحلیل واریانس یک عامله: در این تحلیل مدل در نظر گرفته شده به صورت $x_{ij}=\mu+\alpha_i+e_{ij}$ یامی‌باشد که در آن μمیانگین کل α_iاثر تیمار و e_{ij}خطا می‌باشد. که جدول آنالیز واریانس تحلیل یک عاملی به صورت زیر است

حداقل یکی از $\begin{cases} H_۰ : \alpha_۱=\alpha_۲=...\alpha_k=۰ \\ H_۱ : \alpha_i \text{ها متفاوت است} \end{cases}$ **یاحداقل یکی از** $\begin{cases} H_۰ : \mu_۱=\mu_۲=...\mu_k \\ H_۱ : \mu_i \text{ها متفاوت است} \end{cases}$

جدول آنوا ANOVA

متغیرات	SS	df	$Ms=\frac{SS}{df}$	F
بین گروهها (تیمار)	SSb	$k-۱$	$MS_b=\frac{SS_b}{k-۱}$	$F=\frac{MS_b}{MSw}$
درون گروهها (خطا)	$SSw=ss_{res}$	$k(n-۱)=N-k$	$MSw=\frac{SS_w}{k(n-۱)}$	
کل	$SS=SS_T$	$N-۱$		

اگر $F_{م}>F_{\alpha,k-۱,k(n-۱)}$ **باشد** $H_۰$ **رد می‌شود و حداقل یکی از** μ_i **ها متفاوت است**

نکته۱: $t_{df}^{۲}=F_{۱,df}$

K: تعداد تیمار n_i**: تعداد هر تیمار** **N: تعداد کل** T_i**: جمع هر تیمار** **T: جمع کل**

$\bar{x}_{..}$: میانگین کل $\bar{x}_{..}=\frac{\sum_{i=۱}^{k}\sum_{j=۱}^{n}x_{ij}}{N}, \bar{x}_{i.}=\sum_{j=۱}^{n}\frac{x_{ij}}{n_i} \Rightarrow \bar{X}_{..}=\frac{\sum_{i=۱}^{k}\bar{x}_{i.}}{k}$

$$S_{Tr}=SS_B=n\sum_{i=۱}^{k}(\bar{x}_{i.}-\bar{x}_{..})^{۲}=\sum_{i=۱}^{k}\frac{T_i^{۲}}{n_i}-\frac{T^{۲}}{N}$$

$$SS_W=\sum_{i=۱}^{k}\sum_{j=۱}^{n}(x_{ij}-\bar{x}_{i.})^{۲}=\sum_{i=۱}^{k}(n_i-۱)s_i^{۲} \quad ,SS_t=\sum_{i=۱}^{k}\sum_{j=۱}^{n}(x_{ij}-\bar{x}_{..})^{۲}=\sum_{i=۱}^{k}\sum_{j=۱}^{n}x_{ij}^{۲}-\frac{T^{۲}}{N}$$

$$\eta^{۲}=\frac{SS_B}{SS_T}$$

نکته ۱: در صورتیکه $\eta^2 \leq ۰/۵$ باشد مجموع مجذورات درون گروهی بیشتر از مجموع مجذورات بین گروهی است و بالعکس

نکته۲: تفاوت های فردی در طرح اندازه گیری مکرر یک عاملی کنترل میشود(تکرار سنجش)

توجه : محدود η^2 به صورت زیر است

۱- η^2 حدود ۰/۰۱ یعنی اثر کم ۲- η^2 حدود۰/۰۶ یعنی اثر متوسط ۳- η^2 حدود ۰/۱۴ یعنی اثر زیاد

مثال: در سطح معنی‌دار ۵٪ آزمون کنید که آیا اختلاف معنی‌داری بین میانگینها سه جامعه زیر وجود دارد.

A	۳	۴	۲	۵	۱۴
B	۲	۱	۶	۳	۱۲
C	۴	۹	۷	۶	۲۶

$$T = \sum_{i=۱}^{k} T_i = ۱۴+۱۲+۲۶ = ۵۲$$

$$SS_{Tr} = \sum_{i=۱}^{k} \frac{T_i^۲}{n_i} - \frac{T_۲}{N} = \frac{۱۴^۲}{۴} + \frac{۱۲^۲}{۴} + \frac{۲۶^۲}{۴} - \frac{۵۲^۲}{۱۲} = ۲۸/۶۷$$

$$SS_T = \sum_{i=۱}^{k}\sum_{j=۱}^{n} x_{ij}^۲ - \frac{T_۲}{N} = ۳^۲ + ۴^۲ + ۲^۲ + ۵^۲ + ... + ۷^۲ + ۶^۲ - \frac{۵۲^۲}{۱۲} = ۶۰/۶۷$$

$$SS_E = SS_T - SS_{Tr} = ۶۰/۶۷ - ۲۸/۶۷ = ۳۲$$

منبع	SS	df	$Ms = \frac{SS}{df}$	F
تیمار	۶/۲۸۷	۳−۱=۲	$\frac{۲۸/۶۷}{۲} = ۱۴/۳۳۵$	$F = \frac{MS_{Tr}}{MS_E} = \frac{۱۴/۳۳۵}{۳/۲} = ۴/۴۷۹$ چون $F = ۴/۴۷۹ > F_{/۰۵,۲,۱۰} = ۴/۱۰۲۸$
خطا	۳۲	۱۲−۲=۱۰	$\frac{۳۲}{۱۰} = ۳/۲$	
کل	۶۷/۶۰	۱۲−۱=۱۱		

چون $F = ۴/۴۷۹ > F_{/۰۵,۲,۱۰} = ۴/۱۰۲۸$ است بنابراین $H_۰$ رد می‌شود و یکی از μ_iها با احتمال ۹۴٪ متفاوت است

مثال : درآزمون تحلیل واریانس برای مقایسه سه گروه مستقل در صورتیکه $SS_t = ۶۷۰$ و $SS_B = ۱۰۰$ و تعداد هر گروه مساوی ۲۰ نفر باشد اندازه F چند است ؟

۱) ۵۰　　۲) ۱۰　　۳) ۵　　۴) ۲

متغیرات	SS	df	$Ms = \frac{SS}{df}$	F
بین گروهها (تیمار)	۱۰۰	۳−۱=۲	$MS_b = \frac{۱۰۰}{۲} = ۵۰$	$F = \frac{MS_b}{MSw} = \frac{۵۰}{۱۰} = ۵$
درون گروهها (خطا)	۶۷۰−۱۰۰=۵۷۰	۶۰−۳=۵۷	$MSw = \frac{۵۷۰}{۵۷} = ۱۰$	گزینه ۳ صحیح است
کل	۶۷۰	۶۰−۱=۵۹		

تحلیل واریانس دو عامله یک مشاهده به ازای هر ترکیب (بدون تاثیر متقابل) : مدل در نظر گرفته شده به صورت $x_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$ یا می‌باشد که در آن رابطه خطی زیر برقرار است

K: تعداد تیمار (ستون)　　n : تعداد بلوکها(سطر)

حداقل یکی از $\begin{cases} H_۰ : \alpha_۱ = \alpha_۲ = ... \alpha_k = ۰ \\ H_۱ : \text{حداقل یکی از } \alpha_i \text{ها مخالف صفر است} \end{cases}$　　**حداقل یکی از** $\begin{cases} H_۰ : \beta_۱ = \beta_۲ = ... \beta_n = ۰ \\ H_۱ : \text{حداقل یکی از } \beta_j \text{ها مخالف صفر است} \end{cases}$

نوع درس	نوع درس	وسیله آموزش		
		رادیو	کامپیوتر	تلویزیون
	ریاضی	۴	۳	۵
	فیزیک	۳	۷	۴
	ادبیات	۵	۶	۵
	عربی	۵	۴	۶

جدول تحلیل واریانس رده بندی دو طرفه بدون تأثیر متقابل

متغیرات	SS	df	$Ms = \frac{SS}{df}$	F
تیمار (ستون)	SS_A	$k-۱$	$MS_A = \frac{SS_A}{k-۱}$	$F_A = \frac{MS_A}{MS_w}$
بلوک (سطر)	SS_B	$n-۱$	$MS_B = \frac{SS_B}{n-۱}$	$F_B = \frac{MS_B}{MS_w}$
خطا	$SS_{res} = SS_w$	$(n-۱)(k-۱)$	$MS_w = \frac{SS_w}{(n-۱)(k-۱)}$	
کل	$SS = SS_T$	$N-۱$		

اگر $F_A > F_{\alpha,k-۱,(n-۱)(k-۱)}$ **باشد** $H_۰$ **رد می‌شود و حداقل یکی از** α_i **ها مخالف صفر است**

اگر $F_B > F_{\alpha,n-۱,(n-۱)(k-۱)}$ **باشد** $H_۰$ **رد می‌شود و حداقل یکی از** B_j **ها مخالف صفر است**

جدول تحلیل واریانس رده بندی دو طرفه بدون تأثیر متقابل

متغیرات	SS	Df	$SS^* = Ms = \frac{SS}{df}$	F
تیمار	$SS_۴ = SS_{Tr}$	$k-۱$	$MS_{Tr} = \frac{SS_{Tr}}{k-۱}$	$F_{Tr} = \frac{MS_{Tr}}{MS_E}$
بلوک	$SS_۳ = SS_B$	$n-۱$	$MS_B = \frac{SS_B}{n-۱}$	$F_B = \frac{MS_B}{MS_E}$
خطا	$SS_۲ = SS_E$	$(n-۱)(k-۱)$	$MS_E = \frac{SS_E}{(n-۱)(k-۱)}$	
کل	$SS = SS_T$	$N-۱$		

اگر $F_B > F_{\alpha,k-۱,(n-۱)(k-۱)}$ **باشد** $H_۰$ **رد می‌شود و حداقل یکی از** α_i **ها مخالف صفر است**

اگر $F_B > F_{\alpha,n-۱,(n-۱)(k-۱)}$ **باشد** $H_۰$ **رد می‌شود و حداقل یکی از** B_j **ها مخالف صفر است**

$$SS_{Tr} = n\sum_{i=۱}^{k}(\bar{x}_{i.} - \bar{x}_{..})^۲ = \sum_{i=۱}^{k}\frac{T_{i.}^۲}{n_i} - \frac{T_{..}^۲}{N}$$

$$SS_B = k\sum_{j=۱}^{n}(\bar{x}_{.j} - \bar{x}_{..})^۲ = \sum_{j=۱}^{n}\frac{T_j^۲}{k} - \frac{T_{..}^۲}{N}$$

$$SS_T = \sum_{i=۱}^{k}\sum_{j=۱}^{n}(x_{ij} - \bar{x}_{..})^۲ = \sum_{i=۱}^{k}\sum_{j=۱}^{n}x_{ij}^۲ - \frac{T_{..}^۲}{N}$$

مثال: در سطح خطای ۵٪ آزمون کنید در تعداد فروش کالای خاص، با توجه فصول مختلف و نوع وسیله تبلیغاتی مختلف تفاوت معنی داری وجود دارد؟

فصل	وسیله تبلیغ			
	رادیو	روزنامه	تلویزیون	T_j
بهار	۴	۳	۵	۱۲
تابستان	۳	۷	۴	۱۴
پاییز	۵	۶	۵	۱۶
زمستان	۵	۴	۶	۱۵
$T_{i.}$	۱۷	۲۰	۲۰	

$$T = \sum_{i=۱}^{k}T_{i.} = ۲۰+۲۰+۱۷ = ۵۷$$

$$SS_{Tr}=\sum_{i=1}^{k}\frac{T_{i.}^{2}}{n_i}-\frac{T^{2}}{N}=\frac{17^{2}}{4}+\frac{20^{2}}{4}+\frac{20^{2}}{4}-\frac{57^{2}}{4}=1/5$$

$$SS_{B}=\sum\frac{T_{j}^{2}}{k}-\frac{T^{2}}{N}=\frac{12^{2}}{3}+\frac{14^{2}}{3}+\frac{16^{2}}{3}+\frac{15^{2}}{3}-\frac{57^{2}}{2}=2/91$$

$$SS_{T}=\sum\sum x_{ij}^{2}-\frac{T^{2}}{N}=4^{2}+3^{2}+++-\frac{57^{2}}{12}=16/25$$

$$SS_{E}=SS_{T}-SS_{A}-SS_{Tr}=16/25-1/5-2/91=11/84$$

منبع تغییرات(تیمار)	SS	Df	$MS=\frac{SS}{df}$	MS
فصل (تیمار)	۱/۵	۲	۰/۷۵	$F_{Tr}=\frac{0/75}{1/97}=0/25$
بلوک	۲/۹۱	۳	۰/۹۷	$F_{B}=\frac{0/97}{1/97}=0/49$
خطا	۱۱/۸۴	۶	۱/۹۷	
کل	۱۶/۲۵	۱۱		

چون $F_{Tr}=0/25\not> F_{.05,2,6}=5/1433$ ، $H_{.}$ پذیرفته می‌شود و وسیلهٔ تبلیغ با احتمال ۹۵٪ مؤثر نیست

چون $F_{B}=0/49\not> F_{.05,3,6}=4/7571$ ، $H_{.}$ پذیرفته می‌شود و فصل با احتمال ۹۵٪ مؤثر نیست

تحلیل واریانس دو عامله یا m مشاهده به ازای هر ترکیب (طرح فاکتوریل) : مدل در نظر گرفته شده به صورت $x_{ijr}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\varepsilon_{ijr}$ یا می‌باشد که رابطه‌های زیر برقرار است. طرح فاکتوریل کامل، که به عنوان طرح کاملا متقاطع نیز شناخته می‌شود، به یک طرح آزمایشی اطلاق می‌شود که از دو یا چند عامل تشکیل شده است که هر عامل دارای مقادیر یا «سطوح» ممکن مجزای متعددی است. با استفاده از این طرح می توان تمامی ترکیبات ممکن سطوح عامل را در هر تکرار بررسی کرد. اثرمتقابل موقعی اتفاق می افتد که اثر یک متغیر مستقل بر متغیر وابسته بسته به سطح یک متغیر مستقل دیگر تغییر کند. این اصطلاح بیشتر برای روش های طراحی آزمایش به کار می رود .اثر متقابل بر حسب تعداد متغیرهای اصلی و برهم کنش ها به دو سطحی، سه سطحی و.....

	نوع درس	وسیله آموزش		
		رادیو	کامپیوتر	تلویزیون
نوع درس	ریاضی	۴	۳	۵
		۳	۵	۳
		۵	۹	۲
	فیزیک	۳	۷	۴
		۶	۴	۲
		۶	۵	۳
	ادبیات	۵	۶	۵
		۳	۸	۶
		۷	۹	۹
	عربی	۵	۴	۶
		۶	۷	۷
		۸	۶	۹

جدول تحلیل واریانس رده بندی دو طرفه با تاثیر متقابل جدول آنوا **ANOVA**

منبع متغیرات	SS	df	$SS^* = Ms = \frac{SS}{df}$	F
تیمار	SS_A	$k-1$	$MS_A = \frac{SS_A}{k-1}$	$F_A = \frac{MS_A}{MS_E}$
بلوک	SS_B	$n-1$	$MS_B = \frac{SS_B}{n-1}$	$F_B = \frac{MS_B}{MS_E}$
تاثیر متقابل	$SS_{A.B}$	$(n-1)(k-1)$	$MS_{A.B} = \frac{SS_{A.B}}{(k-1)(n-1)}$	$F_{A.B} = \frac{MS_{A.B}}{MS_W}$
خطا	SS_w	$kn(m-1)$	$MS_w = \frac{SS_w}{kn(m-1)}$	
کل	$SS_T = SS_{yy}$			

اگر $F_A > F_{\alpha,k-1,k_n(m-1)}$ باشد فرض H_0 (یعنی $\alpha_i = 0$) رد می‌شود و حداقل یکی از تیمارها متفاوت است اگر $F_B > F_{\alpha,n-1,k_n(m-1)}$ باشد فرض H_0 (یعنی $\beta_j = 0$) رد می‌شود و حداقل یکی از بلوک‌ها متفاوت است

اگر $F_{A.B} > F_{\alpha,(k-1(n-1)),k_n(m-1)}$ باشد فرض H_0 (یعنی $(\alpha\beta)_{ij} = 0$) رد می‌شود و تاثیر متقابل وجود دارد

$$SS_{Tr} = nm\sum_{i=1}^{k}(\bar{x}_{i..} - \bar{x}_{...})^2 = \sum_{i=1}^{k}\frac{T_{i..}^2}{n_i} - \frac{T^2}{N}$$

$$SS_B = km\sum_{j=1}^{m}\left(\bar{x}_{.j.} - \bar{x}_{...}\right)^2 = \sum_{j=1}^{n}\frac{T_{.j.}^2}{k} - \frac{T^2}{N}$$

$$SS_{Tr.B} = m\sum_{i=1}^{k}\sum_{j=1}^{n}\left(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...}\right)^2$$

$$SS_E = \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m}\left(\bar{x}_{ijk} - \bar{x}_{ij.}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m}x_{ijr}^2 - \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}x_{ij.}^2}{m}$$

$$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m}\left(x_{ijr} - \bar{x}_{...}\right)^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m}x_{ijr}^2 - \frac{T^2}{N}$$

$$SS_{Tr.B} = SS_T - SS_{Tr} - SS_B - SS_E$$

مثال: در سطح خطای ۵٪ آزمون کنید

الف) تبلیغات در میانگین فروش مؤثر است

ب) تخفیف در میانگین فروش مؤثر است

ج) تاثیر متقابلی بین قیمت گذاری و سیاست فروش وجود دارد

قیمت گذاری	سیاست فروش	
	با تبلیغ	بدون تبلیغ
با تخفیف	۱۰ ۱۱	۶ ۵
بدون تخفیف	۶ ۷	۵ ۴

ابتدا جمع در هر سلول را محاسبه می‌کنیم و مانند بدون تکرار مشاهدات SS_B, SS_{Tr} را محاسبه کنید.

	با تبلیغ	بدون تبلیغ	$T_{.j}$
با تخفیف	۲۱	۱۱	۳۲
بدون تخفیف	۱۳	۹	۲۲
$T_{i.}$	۳۴	۲۰	

$$SS_{Tr}=\sum_{i=1}^{k}\frac{T_i^2}{n_i}-\frac{T^2}{N}=\frac{34^2}{4}+\frac{20^2}{4}-\frac{54^2}{8}=24/5$$

$$SS_B=\sum_{j=1}^{n}\frac{T_j^2}{k}-\frac{T^2}{N}=\frac{32^2}{4}+\frac{22^2}{4}-\frac{54^2}{8}=12/5$$

$$SS_T=\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m}x_{ijr}^2-\frac{T^2}{N}=10^2+11^2+...5^2+4^2-\frac{54^2}{2}=43/5$$

$$SS_E=\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{r=1}^{m}x_{ijr}^2-\frac{\sum_{i=1}^{k}\sum_{j=1}^{n}x_{ij\cdot}^2}{m}=10^2+11^2+...5^2+4^2-\frac{11^2+21^2+9^2+13^2}{2}=2$$

$$SS_{Tr.B}=SS_T-SS_{Tr}-SS_B-SS_E=43/5-24/5-12/5-2=4/5$$

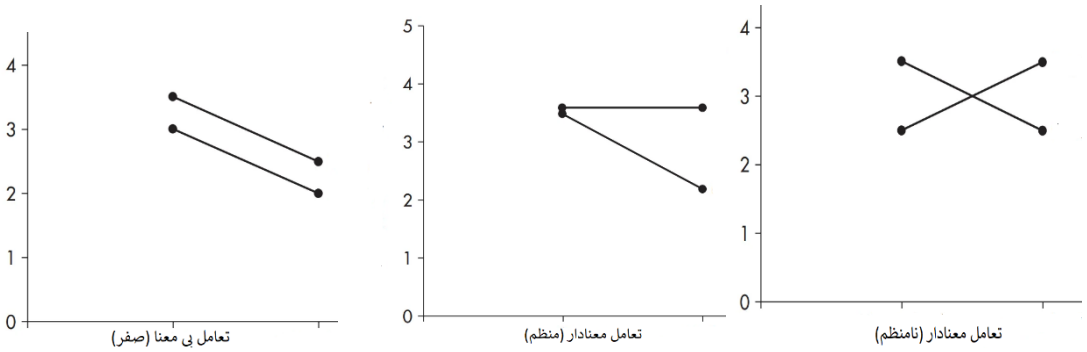
منبع تغییرات(تیمار)	SS	df	$MS=\frac{SS}{df}$	MS
تیمار	۲۴/۵	۱	۲۴/۵	$F_{Tr}=\frac{24/5}{0/5}=49$
بلوک	۱۲/۵	۱	۱۲/۵	$F_B=\frac{12/5}{0/5}=25$
تأثیر متقابل	۴/۵	۱	۴/۵	$F_{Tr.B}=\frac{4/5}{0/5}=9$
خطا	۲	۴	۰/۵	
کل	۴۳/۵	۷		

چون $F_{Tr}=49>F_{.05,1,4}=7/7086$ می‌باشد بنابراین H_0 رد می‌شود و با اطمینان ۹۵٪ تبلیغ مؤثر است.

چون $F_B=25>F_{.05,1,4}=7/7086$ می‌باشد بنابراین H_0 رد می‌شود و با اطمینان ۹۵٪ شیوهٔ قیمت گذاری موثر است.

چون $F_{Tr.B}=9>F_{.05,1,4}=70/7086$ می‌باشد بنابراین H_0 رد می‌شود و با اطمینان ۹۵٪ تأثیر متقابل وجود دارد.

نکته : اگر دو خط یکدیگر را قطع کنند تعامل (تاثیر متقابل) معنا دار است و اگر دو خط موازی یکدیگر باشند تعامل بی معنا است . و در صورتی که اثر تعاملی نامنظم وجود داشته باشد باید تفسیر متغیرهای مستقل با احتیاط صورت گیرد.



متغیر کنترل: به متغیر مزاحمی گفته می شود که محقق تصمیم می گیرد اثر آن را خنثی یا حذف کند.

محقق به دو روش می تواند اثر متغیر مزاحم را کنترل کند:

۱-کنترل پژوهشی: با ثابت نگه داشتن متغیر در یک سطح، اثر آن خنثی می شود. مثلا فقط از بین دانش آموزان دختر نمونه گیری انجام می شود.

۲-کنترل آماری: به کمک روش آماری تحلیل کواریانس، اثر متغیر مزاحم مطالعه و از روی نتایج حذف می شود.

فصل هشتم

ناپارامتری

آزمون‌های ناپارامتری:

هر گاه توزیع جامعه مشخص نباشد از آزمون‌های ناپارامتری (آمار توزیع آزاد) استفاده می‌شود که بر اساس تعداد می‌باشد. که دقت آن نسبت به آمار پارامتری کمتر است.

آزمون علامت یک نمونه‌ای:

از این آزمون برای آزمون میلنه میلنه یا چارک‌ها و اگر توزیع جامعه متقارن باشد آزمون میانگین هم می‌توان استفاده کرد و سپس $(X_i - \theta_0)$ را محاسبه می‌کنیم اگر مثبت باشد علامت + و اگر منفی باشد علامت – قرار می‌دهیم که پارامترهای توزیع دوجمله‌ای می‌باشد که برای میانه $P = \frac{1}{2}$ و برای چارکها $\frac{1}{4}$ یا $\frac{3}{4}$ قرار می‌دهیم که احتمال موردنظر است و اگر $nq > 5, nP > 5$ باشد می‌توان از تقریب توزیع نرمال استفاده کرد.

$$z = \frac{x - np_0}{\sqrt{np_0 q_0}}$$

مثال: در یک جامعه تاکنون میانه برابر ۴۰ بوده است در سال جاری به علت تغییرات احساس می‌شود میانه تغییر کرده است به این منظور نمونه‌ای ۱۴تایی انتخاب کرده و اعداد

۴۰ – ۴۵ – ۴۷ – ۴۰ – ۳۳ – ۴۲ – ۵۰ – ۳۹ – ۴۰ – ۳۷ – ۴۹ – ۴۳ – ۴۰ – ۴۶

بدست آمده است در سطح معنی‌داری $\alpha = 0/1$ این اعداد را آزمون کنیم.

جواب: ابتدا $(X_i - \theta_0)$ را بدست می‌آوریم و علامت را تعیین می‌کنیم و اگر حاصل صفر شود علامت نمی‌گذاریم.

$$++-++-+++$$

بنابراین $n = 10, p = \frac{1}{2}$ یعنی $B(10, \frac{1}{2})$ بنابراین چون تعداد مثبت‌ها بیش از $np = 10 \times \frac{1}{2} = 5$ است بنابراین؛

$$P(X \geq 8) = 0/1 = 1 - P(X \leq 7) = 1 - 0/945 = 0/055 < 0/1$$

چون تعداد علامت مثبت در نمونه یعنی آمارهٔ آزمون برابر ۷ و در ناحیه بحرانی قرار نمی‌گیرد و H_0 پذیرفته می‌شود یا می‌توان بر اساس منفی بحث کرد یعنی اولین جایی که احتمال در $n=10$ و $p=0/5$ کمتر از ۰/۱ باشد.

$$P(X \leq C) = 0/1 = C_{0/1, 10, 0/5} = 2$$

چون تعداد منفی‌ها بیش از ۲ است بنابراین در ناحیهٔ رد قرار نمی‌گیرد و H_0 پذیرفته می‌شود.

کاربردهای آزمون مربع کای دو (خی‌دو): همانگونه که می‌دانیم توزیع خی دو مجموع مربعات نرمال استاندارد است $(X^2 = \sum_{i=1}^{df} Z_i^2)$ که میانگین آن برابر درجه آزادی و واریانس دو برابر درجه آزادی است توزیع خی‌دو دارای منحنی چوله به راست که در آزمون فرض‌های زیر استفاده می‌شود.

۱ـ جدولهای توافقی :

که با نام‌های آزمونهای استقلال و آزمونهای همگونی (نسبتها) مطرح می‌شود برای مواقعی که r سطر و c

ستون داشته باشیم که فراوانی مشاهده شده O_{ij}و فراوانی مورد انتظار e_{ij}می‌باشد که از رابطهٔ $e_{ij}=\frac{O_i \times O_{ij}}{n}$ بدست می‌آید و آمارهٔ آزمون از رابطهٔ $X^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-e_{ij})^2}{e_{ij}}$ بدست می‌آید و با عدد بحرانی χ^2_{α} با درجهٔ آزادی $df=(r-1)(c-1)$ مقایسه می‌شود.در صورتی که $X^2>\chi^2_{\alpha,df}$ باشد H_0 رد می‌شود و در نتیجه داده مستقل نیستند یا می‌توان گفت نسبتها برابر نیست لازم به توضیح است این آزمون در مواقعی بکار می‌رود که هیچ یک از e_{ij}ها کمتر از ۵ نباشد در این مواقع برخی خانه‌ها را با یکدیگر ادغام می‌کنیم.

نکته: در صورتی که درجه آزادی برابر یک شود از تصحیح یتس استفاده می‌شود و آمارهٔ آزمون از رابطهٔ $X^2=\sum\frac{(|F_{oi}-F_{ei}|-0/5)^2}{F_{ei}}$ بدست می‌آید.

مثال: در سطح خطای ۵٪ آزمون کنید میزان تحصیلات و جنسیت مستقل از یکدیگر هستند.

تحصیلات / جنسیت	A	B	C	D
خانم	۵۵	۲۰	۱۵	۲۰
آقا	۳۵	۲۰	۲۵	۱۰

تحصیلات / جنسیت	A	B	C	D	جمع
خانم	۵۵	۲۰	۱۵	۲۰	۱۱۰
آقا	۳۵	۲۰	۲۵	۱۰	۹۰
جمع	۹۰	۴۰	۴۰	۳۰	

تحصیلات / جنسیت	A	B	C	D
خانم	$\frac{۱۱۰\times۹۰}{۲۰۰}=۴۹/۵$	$\frac{۱۱۰\times۴۰}{۲۰۰}=۲۲$	$\frac{۱۱۰\times۴۰}{۲۰۰}=۲۲$	$\frac{۱۱۰\times۳۰}{۲۰۰}=۱۶/۵$
آقا	$\frac{۹۰\times۹۰}{۲۰۰}=۴۰/۵$	$\frac{۹۰\times۴۰}{۲۰۰}=۱۸$	$\frac{۹۰\times۴۰}{۲۰۰}=۱۸$	$\frac{۹۰\times۳۰}{۲۰۰}=۱۳/۵$

$$X^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-e_{ij})^2}{e_{ij}}=\frac{(55-49/5)^2}{49/5}+\frac{(20-22)^2}{22}+\frac{(15-22)^2}{22}+\frac{(20-16/5)^2}{16/5}$$

$$+\frac{(35-40/5)^2}{40/5}+\frac{(20-18)^2}{20}+\frac{(25-18)^2}{18}+\frac{(10-13/5)^2}{13/5}=8/092$$

$$df=(r-1)(C-1)=(2-1)(4-1)=3$$

چون $\chi^2_{.05,3}=7/815<X^2=8/092$ است بنابراین $H_.$ رد می‌شود و با احتمال ۹۵٪ با توجه به نمونه جنسیت و میزان تحصیلات مستقل از یکدیگر نیستند.

مثال: در سطح خطای ۵٪ آزمون کنید بازمانیهای لامپهای خلاء و بلوک بالا و بلوک پائین مستقل از یکدیگر هستند.

نوع بازمانی O_{ij}

	A	B	C	D
بلوک بالا	۵۵	۲۰	۱۵	۲۰
بلوک پائین	۳۵	۲۰	۲۵	۱۰

	A	B	C	D	جمع
بلوک بالا	۵۵	۲۰	۱۵	۲۰	۱۱۰
بلوک پائین	۳۵	۲۰	۲۵	۱۰	۹۰
جمع	۹۰	۴۰	۴۰	۳۰	

	A	B	C	D
بلوک بالا	$\frac{۱۱۰\times۹۰}{۲۰۰}=۴۹/۵$	$\frac{۱۱۰\times۴۰}{۲۰۰}=۲۲$	$\frac{۱۱۰\times۴۰}{۲۰۰}=۲۲$	$\frac{۱۱۰\times۳۰}{۲۰۰}=۱۶/۵$
بلوک پائین	$\frac{۹۰\times۹۰}{۲۰۰}=۴۰/۵$	$\frac{۹۰\times۴۰}{۲۰۰}=۱۸$	$\frac{۹۰\times۴۰}{۲۰۰}=۱۸$	$\frac{۹۰\times۳۰}{۲۰۰}=۱۳/۵$

$$X^2=\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-e_{ij})^2}{e_{ij}}=\frac{(55-49/5)^2}{49/5}+\frac{(20-22)^2}{22}+\frac{(15-22)^2}{22}+\frac{(20-16/5)^2}{16/5}$$

$$+\frac{(35-40/5)^2}{40/5}+\frac{(20-18)^2}{20}+\frac{(25-18)^2}{18}+\frac{(10-13/5)^2}{13/5}=8/092$$

$$df=(r1)(C-1)=(2-1)(4-1)=3$$

چون $X^2=8/092>\chi^2_{.05,3}=7/815$ است بنابراین $H_.$ رد می‌شود و با احتمال ۹۵٪ بازمانیهای لامپهای خلاء و بلوک مستقل از یکدیگر نیستند.

ضریب توافقی **C**: اگر در آزمون استقلال دو صفت وابسته باشند به کمک ضریب توافقی **C** می‌توان شدت این بستگی را مشخص کرد. $C=\sqrt{\frac{X^2}{X^2+n}}$

ضریب فی : $\Phi=\sqrt{\frac{\chi^2}{n}}$

ضریب کریمر :که تعمیم ضریب فی است $V=\sqrt{\frac{X^2}{N.\min[(K-1),(L-1)]}}$

آزمون نیکویی برازش:

از این آزمون در مواقعی استفاده می‌شود که بخواهیم بدانیم داده‌ها از توزیع خاصی تبعیت می‌کنند یا خیر. که **Foi** فراوانی مشاهده شده و $F_{ei}=Np$ فراوانی مورد انتظار است و آمارهٔ آزمون از رابطهٔ $X^2=\sum_{i=1}^{k}\frac{(F_{oi}-F_{ei})^2}{F_{ei}}$ و درجه آزادی از رابطهٔ **df=k-p-۱** بدست می‌آید که در آن **k** تعداد طبقات و **p** تعداد پارامتری که باید برآورد شود. در صورتی که $X^2>X^2_{\alpha,df}$ باشد $H_.$ رد می‌شود و فراوانی نظری و تجربی برابر نیست و داده‌ها از آن توزیع تبعیت نمی‌کنند.

مثال: تعداد کالای تولیدی مشاهده توسط چهار دستگاه به صورت زیر نسبت شده است

A	B	C	D
۲۵۰	۱۱۰	۹۰	۵۰

مدیر کارخانه مدعی است که تولید کالا به نسبت اعداد زیر می‌باشد

A	B	C	D
۸	۷	۶	۴

در سطح اطمینان ۵٪ آزمون کنید تردیدی برای ادعای فوق وجود دارد.

دستگاه	F_{oi}	$F_{ei}=n.p$
A	۲۵۰	$۵۰۰\times\frac{۱۳}{۲۵}=۲۶۰$
B	۱۱۰	$۵۰۰\times\frac{۷}{۲۵}=۱۴۰$
C	۹۰	$۵۰۰\times\frac{۶}{۲۵}=۱۴۰$
D	۵۰	$۵۰۰\times\frac{۴}{۲۵}=۸۰$
	۵۰۰	۵۰۰

$$X^{۲}=\sum_{i=۱}^{k}\frac{(F_{oi}-F_{ei})^{۲}}{F_{ei}}=\frac{(۲۵۰-۲۶۰)^{۲}}{۲۶۰}+\frac{(۱۱۰-۱۴۰)^{۲}}{۱۴۰}+\frac{(۹۰-۱۲۰)^{۲}}{۱۲۰}$$

$$+\frac{(۵۰-۸۰)^{۲}}{۸۰}=۲۵/۵۶$$

$$df=k-p-۱=۴-۰-۱=۳$$

چون $X^{۲}_{./۰۵,۳}=۷/۸۱۵<X^{۲}=۲۵/۵۶$ است بنابراین $H_{.}$ رد می‌شود و با احتمال ۹۵٪ نظر مدیر صحیح نیست.

نکته: در صورتی که درجه آزادی برابر یک شود از تصحیح یتس استفاده می‌شود و آمارهٔ آزمون از رابطهٔ $X^{۲}=\sum\frac{(|F_{oi}-F_{ei}|-۰/۵)^{۲}}{F_{ei}}$ بدست می‌آید.

مثال: از ۶۰ کالای تولید شده یک کارخانه در یک ماه، ۱۵ کالا معیوب است در سطح خطای ۵٪ آزمون کنید کالاهای معیوب از توزیع برنولی با پارامتر ۰/۲ تبعیت می‌کند.

کیفیت	f_{oi}	P	$F_{ei}=NP$
معیوب	۱۵	۰/۲	۱۲
سالم	۴۵	۰/۸	۴۸
	۶۰		

$\Rightarrow df=k-p-۱=۲-۰-۱=۱$ نیاز به تصحیح یتس دارد.

$$X^2=\sum_{i=1}^{k}\frac{(|F_{oi}-F_{ei}|-0/5)^2}{F_{ei}}=\frac{(|15-12|-0/5)^2}{12}+\frac{(|45-48|-0/5)^2}{48}=0/65$$

چون $X^2=0/65\not> X^2_{0/05,1}=3/841$ بنابراین H_0 پذیرفته می‌شود و با احتمال ۹۵٪ تعداد کالاهای معیوب از توزیع برنولی با پارامتر p=۰/۲ تبعیت می‌کند.

مثال: نمونه‌ای شامل ۲۰۰ صفحه تایپ شده را بررسی کرده‌ایم و تعداد اشتباهات در صفحات ثبت شده است در سطح خطای ۰/۰۱ آزمون کنید آیا تعداد اشتباهات از توزیع پواسون تبعیت می‌کند

تعداد اشتباهاتX	۰	۱	۲	۳	۴	۵
F_{oi} تعداد صفحات	۵۰	۶۰	۵۰	۲۴	۱۲	۴

همانگونه که می‌دانیم در توزیع پواسون پارامتر توزیع برابر میانگین توزیع است.

$$\mu=\frac{\sum F_i X_i}{N}=\frac{0\times50+1\times60+2\times50+3\times24+4\times12+5\times4}{200}=1/5$$

با استفاده از رابطهٔ پواسون $P(X=x)=\frac{e^{-\mu}\mu^x}{x!}$ احتمالها را بدست می‌آوریم

X	F_{oi}	$P=\frac{e^{-\mu}\mu^x}{x!}$	$F_{ei}=NP$
۰	۵۰	۰/۲۲۳۱	۴۴/۶۲
۱	۶۰	۰/۳۳۴۷	۶۶/۹۴
۲	۵۰	۰/۲۵۱۰	۵۰/۲
۳	۲۴	۰/۱۲۵۵	۲۵/۱
۴	۱۲	۰/۰۴۷۱	۹/۴۲
۵	۴	۰/۰۱۷۶	۳/۵۲

چون فراوانی نظری ردهٔ آخر کمتر از ۵ است بهتر است دو ردهٔ آخر را با هم ادغام کنیم

X	F_{oi}	F_{ei}
۰	۵۰	۴۴/۶۲
۱	۶۰	۶۶/۹۴
۲	۵۰	۵۰/۲
۳	۲۴	۲۵/۱
۴≥	۱۶	۹/۴۲

$$X^{۲}=\sum_{i=۱}^{k}\frac{(F_{oi}-F_{ei})^{۲}}{F_{ei}}=\frac{(۵۰-۴۴/۶۲)^{۲}}{۴۴/۶۲}+\frac{(۶۰-۶۶/۹۴)^{۲}}{۶۶/۹۴}++\frac{(۵۰-۵۰/۲)^{۲}}{۵۰/۲}$$

$$+\frac{(۲۴-۲۵/۱)^{۲}}{۲۵/۱}+\frac{(۱۶-۱۲/۹۴)^{۲}}{۱۲/۹۴}=۲/۱۴$$

$$df=k-p-۱=۵-۱-۱=۳$$

چون $\chi^{۲}_{./.۱,۳}=۱۱/۳۴۵\not< X^{۲}=۲/۱۴$ بنابراین فرض $H_{.}$ پذیرفته می‌شود و با احتمال ۹۹٪ تعداد اشتباهات از توزیع پواسون تبعیت می‌کند.

آزمون استقلال در جدول ۲×۲:

در جدول ۲×۲ می‌توان به صورت زیر آماره آزمون را بدست آورد:

حالت اول

حالت دوم		
a	b	a+b
c	d	c+d

$$X^{۲}=\frac{(ad-cb)^{۲}(a+b+c+d)}{(a+b)(a+c)(b+d)(c+d)}$$

اگر $X^{۲}>X^{۲}_{\alpha,۱}$ باشد فرض $H_{.}$ رد می‌شود و مستقل نیستند.

همچنین برای این آزمون می‌توان از آمارهٔ Z استفاده کرد.

$$Z=\frac{(ad-cb)\sqrt{a+b+c+d}}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$$

اگر $|Z|>Z_{\alpha/2}$ باشد فرض H_0 رد می‌شود و مستقل نیستند.

آزمون رتبه علامت‌دار: در صورتی که داده به صورت زوجی باشد استفاده می‌شود که ابتدا اختلاف دو حالت را به دست می‌آوریم سپس قدرمطلق اختلافها را حساب می‌کنیم و آنها را از کوچک به بزرگ رتبه می‌دهیم و در صورتی که اعداد یکسان داشته باشیم رتبه متوسط می‌دهیم و سپس مجموع رتبه‌های مثبت $\left(\overset{+}{T}\right)$ یا مجموع رتبه‌های منفی $\left(\overset{-}{T}\right)$ را بدست می‌آوریم اگر $n\geq 15$ باشد توزیع $\overset{+}{T}$ یا $\overset{-}{T}$ تقریباً نرمال است با میانگین و واریانس $Var(T)=\frac{n(n+1)(2n+1)}{24}$, $E(T)=\frac{n(n+1)}{4}$ که در این صورت آماره آزمون به صورت $Z=\frac{T-E(T)}{\sqrt{Var(T)}}$ بدست می‌آید و بسته به شرایط آزمون، آزمون صورت می‌گیرد.

مثال: ورقه‌های ۱۵ دانشجو توسط دو استاد به طور مستقل شده و نمرات آنها در جدول زیر ثبت شده است در سطح ۵٪ آزمون کنید آیا نمرات دو استاد تفاوت معنی‌دار دارد.

استاد اول (x)	۶	۱۲	۱۹	۱۴	۷	۱۵	۱۰	۱۳	۱۱	۵	۱۸	۲۰	۱۲	۱۷	۱۴
استاد دوم (y)	۸	۱۳	۱۷	۱۳	۶	۱۸	۱۲	۱۱	۱۵	۸	۲۰	۱۹	۱۴	۱۵	۱۶
d=x-y	-۲	-۱	۲	۱	۱	-۳	-۲	۲	-۴	-۳	-۲	-۱	-۲	۲	-۲
قدر مطلق مرتب‌شده	۱	۱	۱	۱	۲	۲	۲	۲	۲	۲	۲	۲	۳	۳	۴
رتبه فرض	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
رتبه اصلی	۲/۵	۲/۵	۲/۵	۲/۵	۲/۵	۸/۵	۸/۵	۸/۵	۸/۵	۸/۵	۸/۵	۸/۵	۱۳/۵	۱۳/۵	۱۵

$$T^{+}=۲/۵+۲/۵+۸/۵+۸/۵+۸/۵=۳۰/۵$$

$$E(T)=\frac{n(n+1)}{4}=\frac{۱۵\times ۱۶}{4}=۶۰$$

$$Var(T)=\frac{n(n+1)(2n+1)}{24}=\frac{15\times16\times31}{24}=310$$

$$Z=\frac{T-E(T)}{\sqrt{Var(T)}}=\frac{30/5}{\sqrt{310}}=\frac{-29/5}{17/6}=-1/67$$

$$\begin{cases}H_0:\mu_1=\mu_2\\ H_1:\mu_1\neq\mu_2\end{cases}\qquad \begin{cases}H_0:\bar{d}=0\\ H_1:\bar{d}\neq0\end{cases}$$ یا

چون $|z|=1/67\not>z_{./025}=1/96$ بنابراین H_0 پذیرفته می‌شود و نمرات در استاد در سطح اطمینان ۹۵٪ با یکدیگر تفاوت معنی‌دار ندارد.

آزمون مجموع رتبه‌ها:

آزمون **U** من ـــ وتینی یا آزمون ویلکاکسون جایگزین آزمون **t** در آزمون پارامتری برای مقایسه یکسان بودن میانگین دو جامعه استفاده می‌شود. برای انجام این آزمون ابتدا دو جامعه را با یکدیگر ادغام کرد. و از کوچک به بزرگ رتبه می‌دهیم و در صورت یکسان بودن اعداد رتبه متوسط می‌دهیم سپس مجموع رتبه‌ها هریک از جوامع را بدست می‌آوریم (R_1,R_2) سپس آماره‌های U_1,U_2 را بدست آورده و **min** آنها را انتخاب می‌کنیم. اگر $n_1>8,n_2>8$ باشد در این صورت **U** تقریباً دارای توزیع نرمال می‌باشد.

$$E(U)=\frac{n_1n_2}{2}$$

$$Var(U)=\frac{n_1n_2(n_1+n_2+1)}{12}$$

و آمارهٔ آزمون از رابطهٔ $Z=\frac{U-E(U)}{\sqrt{Var(U)}}$ بدست می‌آید.

آزمون مجموع رتبه‌ها آزمون H :

آزمون **H** یا آزمون کروسکال ـــ والیس تعمیم‌یافته آزمون **U** من ویتمنی است و برای مقایسه میانگین **k**جامعه‌به کار می‌رود و جایگزین تحلیل واریانس در آمار پارامترنی است برای انجام این جوامع رابا یکدیگر ادغام می‌کنیم و از کوچک‌به بزرگ مرتب می‌کنیم و رتبه می‌دهیم مانند حالتهای قبل سپس مجموع رتبه‌های هر جامعه را تعیین می‌کنیم (R_i) که در این صورت آمارهٔ آزمون از رابطه $H=\frac{12}{n(n+1)}\sum_{i=1}^{k}\frac{R^2_i}{n_i}-3(n+1)$ بدست می‌آید اگر حجم نمونه‌ها حداقل برابر ۵ باشد آمارهٔ **H** دارای توزیع خی دو **x^2** با درجه آزادی **df=k-1** است اگر $H>x^2_{d,df}$ باشد. H_0 رد می‌شود و میانگین حداقل یک جامعه متفاوت است.

تجزیه و تحلیل فریدمن: در این آزمون داده‌ها حداقل‌بلید دارای مقیاس رتبه‌ای‌باشـــند و تعمیم‌یافته آزمون رتبه علامت‌دار اســـت یعنی روی افراد‌ثلبت k آزمایش انجام شـــود برای انجام این آزمون‌به هر جامعه رتبه می‌دهیم و انتظار داریم که فراوانی رتبه‌ها در kجامعه برابرباشـــد. بنابراین مجموع رتبه‌های هر جامعه را بدست می‌آوریم (R_i) سپس آمارهٔ آزمون را از رابطهٔ $X^2 = \frac{12}{nk(k+1)}\sum_{i=1}^{k} R_i^2 - 3n(k+1)$ بدست می‌آوریم که n هر جامعه و kتعداد جوامع است اگر $X^2 > x^2_{\alpha,K-1}$ باشد H_0 رد می‌شود و حداقل یکی از جوامع متفاوت است.

مثال: تحقیق کنید آیا نتیجه ارزشیابی ۳ استاد یکسان است در سطح خطای ۵٪

نمرات

دانشجو	I	II	III
۱	۴۸	۷۵	۱۰۰
۲	۳۰	۱۰۰	۱۰۰
۳	۱۰۰	۸۶	۹۶
۴	۳۳	۵۸	۶۰
۵	۶۰	۶۵	۹۰
۶	۲۰	۶۰	۴۰

رتبه ها

دانشجو	I	II	III
۱	۱	۲	۳
۲	۱	۲/۵	۲/۵
۳	۱	۲	۳
۴	۱	۲	۳
۵	۲	۱	۳
۶	۱	۳	۲
Ri	۷	۱۲/۵	۱۶/۵

$$X^{۲}=\frac{۱۲}{nk(k+۱)}\sum_{i=۱}^{n}R_i^{۲}-۳n(k+۱)$$

$$n=۶ \quad k=۳$$

$$X^{۲}=\left[\frac{۱۲}{۶\times۳(۳+۱)}\left[۷^{۲}+۱۲/۵^{۲}+۱۶/۵^{۲}\right]\right]-۳\times۶\times(۳+۱)=۷/۵۸۳$$

$$df=k-۱=۳-۱=۹$$

چون $X^{۲}_{./۰۵,۲}=۵/۹۹۱ > X^{۲}=۷/۵۸۳$ **بنابراین** $H_.$ **رد می‌شود و ارزشیابی اساتید در سطح خطای ۵٪ یکسان نیست.**

آزمون ردیف (گردش):

از این آزمون می‌توان برای آزمون استقلال یا تصادفی بودن ردیف استفاده می‌شود. در صورتی که $n_۱\geq۱۰, n_۲\geq۱۰$ **باشد می‌توان از تقریب نرمال استفاده کرد:**

$$E(R)=\frac{۲n_۱n_۲}{n_۱+n_۲}+۱$$

$$Var(R)=\frac{۲n_۱n_۲(۲n_۱n_۲-n_۱-n_۲)}{(n_۱+n_۲)^{۲}(n_۱+n_۲-۱)}$$

$$Z=\frac{R-E(R)}{\sqrt{Var(R)}}$$

که در آن R تعداد ردیفها یا تغییرات است $n_۱, n_۲$ **تعداد هریک اعضا است.**

نکته: اگر $n_۱, n_۲$ **کمتر از ۱۰ باشند از جدول مخصوص آزمون ردیف استفاده می‌شود.**

مثال: مراجعات بر جنسیت به یک بانک به صورت زیر است در سطح خطای ۵٪ آزمون کنید آیا ورود افراد از لحاظ جنسیت تصادفی است.

$$R=۸$$

$$E(R)=\frac{۲n_۱n_۲}{n_۱+n_۲}+۱=\frac{۲\times۱۱\times۱۲}{۱۱+۱۲}+۱=۱۲/۴۷$$

$$Var(R)=\frac{۲n_۱n_۲(۲n_۱n_۲-n_۱-n_۲)}{(n_۱+n_۲)^{۲}(n_۱+n_۲-۱)}=\frac{۲\times۱۱\times۱۲(۲\times۱۱\times۱۲-۱۱-۱۲)}{(۱۱+۱۲)^{۲}(۱۱+۱۲-۱)}=۵/۴۶۷$$

$$Z=\frac{R-E(R)}{\sqrt{Var(R)}}=\frac{۸-۱۲/۴۷}{\sqrt{۵/۴۶۷}}=-۱/۹۱۱$$

چون $|Z|=۱/۹۱۱ \not> Z_{./۰۲۵}=۱/۹۶$ **بنابراین** $H_.$ **پذیرفته می‌شود و ورود افراد تصادفی است با احتمال ۹۵٪.**

ضریب همبستگی رتبه‌ای اسپیرمن:

در محاسبهٔ ضریب همبستگی پواسون فرض اصلی این است که نمونه‌ها از جامعه نرمال دوبعدی باشد در صورتی که متغیرها کیفی باشند یا کمی و به متعلق به جامعه نرمال نباشند از ضریب همبستگی رتبه‌ای اسپیرمن استفاده می‌شود که برای محاسبهٔ این ضریب به مقادیر y,x از کوچک به بزرگ رتبه می‌دهیم و در صورت یکسان بودن رتبه متوسط می‌دهیم سپس اختلاف رتبه‌های زوج y,xمحاسبه می‌کنیم.

سپس از رابطهٔ $(d_i = r_{xi} - r_{yi})$ **ضریب همبستگی محاسبه می‌شود و اگر** $n > ۱۰$

$$r_s = ۱ - \frac{۶\sum_{i=۱}^{k} d_i^۲}{n(n^۲ - ۱)}$$

باشد می‌توان از توزیع نرمال تقریب زد و آزمون فرض ضریب همبستگی را انجام می‌دهیم.

$$E(r_s) = ۰$$

$$Var(r_s) = \frac{۱}{n-۱} \Rightarrow z = \frac{r_s - ۰}{\sqrt{\frac{۱}{n-۱}}} \Rightarrow z = r_s\sqrt{n-۱}$$

لازم به توضیح است اگر nکوچک باشد می‌توان از توزیع t برای آزمون فرض استفاده کرد که کمتر استفاده می‌شود.

$$t = \frac{r_s\sqrt{n-۲}}{\sqrt{۱-r_s^۲}}$$

مثال: ضریب هبمستگی اسپیرمن برای دو متغیر y,x را محاسبه کنید.

x	۶	۹	۷	۵	۴	
y	-۲	-۵	۴	۶	۷	
r_x	۳	۵	۴	۲	۱	
r_y	۲	۱	۳	۴	۵	
$d = r_x - r_y$	۱	۴	۱	-۲	-۴	
$d^۲$	۱	۱۶	۱	۴	۱۶	$\sum d_i^۲ = ۳۸$

$$r_s = ۱ - \frac{۶\sum_{i=۱}^{k} di^۲}{n(n^۲ - ۱)} = ۱ - \frac{۶ \times ۳۸}{۵(۲۵ - ۱)} = -۰/۹$$

آزمون کولموگروف اسمیرنوف (ks):

این آزمون برای همگونی یا نیکویی برازش بین دو فراوانی تجمعی نسبی در آمار ناپارامتری به کار می‌رود. برای انجام این آزمون فراوانی تجمعی نسبی مشاهده شده (F_o) **و فراوانی تجمعی نسبی نظر**

(F_e) را بدست می‌آوریم سپس اختلاف آنها را محاسبه می‌کنیم که آمارهٔ آزمون $D_n = \max|F_o - F_e|$ می‌باشد و اگر $D_n > D_{\alpha,n}$ باشد. H_0 رد می‌شود و از آن توزیع تبعیت نمی‌کند.

حدود تلرانس طبیعی آزاد ـ توزیع:

در صورتی که توزیع جامعه مشخص نباشد برای حدود تلرانس دوطرفه، تعداد مشاهدات لازم به طوری که با احتمال، $(1-\alpha)100$ درصد از توزیع بین کوچکتر بین مشاهده و بزرگترین مشاهده قرار گیرد و تقریباً از رابطهٔ $n \simeq (\frac{2-\alpha}{\alpha})(\frac{x^2_{1-\gamma,4}}{4}) + \frac{1}{2}$ بدست می‌آید.

و برای حدود تلرانس یک طرفه تعداد مشاهدات لازم به طوری که با احتمال γ، $(1-\alpha)100$ درصد از توزیع بین کوچکترین مشاهده و بزرگترین مشاهده قرار گیرد. تقریباً از رابطهٔ $n = \frac{\log(1-\gamma)}{\log(1-\alpha)}$ بدست می‌آید.

مثال: تعداد مشاهدات لازم به منظور حدود تلرانس آزاد ـــ توزیع دو طرفه که با احتمال ۹۰٪ دست کم ۹۵٪ درصد توزیع را در برگیرد را بدست آورید.

$$1-\alpha = 0/95 \Rightarrow \alpha = 0/05 \quad , \gamma = \%90 \qquad X^2_{0/1,4} = 7/77944$$

$$n \approx (\frac{2-\alpha}{\alpha})(\frac{x^2_{1-\gamma,4}}{4}) + \frac{1}{2} = (\frac{2-0/05}{0/05})(\frac{X^2_{0/1,4}}{4}) + \frac{1}{2} = 76$$

مثال: تعداد مشاهدات لازم به منظور حدود تلرانس آزاد ـــ توزیع یک طرفه که با احتمال ۰/۹۸ دست کم ۹۰٪ درصد توزیع را در برگیرد را بدست آورید.

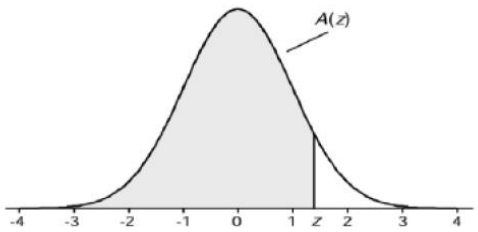
$$1-\alpha = 0/9 \Rightarrow \alpha = 0/1 \ , \gamma = 0/98$$

$$n = \frac{\log(1-\gamma)}{\log(1-\alpha)} = \frac{\log(0/02)}{\log(0/9)} = \frac{-1/69}{-0/04} \simeq 42$$

جدول توزیع نرمال استاندارد

TABLE A.1

Cumulative Standardized Normal Distribution



$A(z)$ is the integral of the standardized normal distribution from $-\infty$ to z (in other words, the area under the curve to the left of z). It gives the probability of a normal random variable not being more than z standard deviations above its mean. Values of z of particular importance:

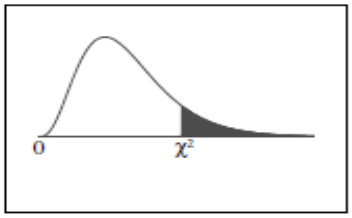
z	$A(z)$	
1.645	0.9500	Lower limit of right 5% tail
1.960	0.9750	Lower limit of right 2.5% tail
2.326	0.9900	Lower limit of right 1% tail
2.576	0.9950	Lower limit of right 0.5% tail
3.090	0.9990	Lower limit of right 0.1% tail
3.291	0.9995	Lower limit of right 0.05% tail

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999							

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	**0.50**	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
two-tails	**1.00**	**0.50**	**0.40**	**0.30**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.002**	**0.001**
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

TABLE A.3

F Distribution: Critical Values of *F* (5% significance level)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.42	19.43	19.44	19.45
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.71	8.69	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.87	5.84	5.82	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.64	4.60	4.58	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.96	3.92	3.90	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.53	3.49	3.47	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.24	3.20	3.17	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.03	2.99	2.96	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.86	2.83	2.80	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.74	2.70	2.67	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.64	2.60	2.57	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.55	2.51	2.48	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.48	2.44	2.41	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.42	2.38	2.35	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.37	2.33	2.30	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.33	2.29	2.26	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.29	2.25	2.22	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.26	2.21	2.18	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.22	2.18	2.15	2.12
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.20	2.16	2.12	2.10
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.17	2.13	2.10	2.07
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.15	2.11	2.08	2.05
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.13	2.09	2.05	2.03
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.11	2.07	2.04	2.01
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.09	2.05	2.02	1.99
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.08	2.04	2.00	1.97
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.06	2.02	1.99	1.96
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.05	2.01	1.97	1.94
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.04	1.99	1.96	1.93
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.04	1.99	1.94	1.91	1.88
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.95	1.90	1.87	1.84
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.89	1.85	1.81	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.86	1.82	1.78	1.75
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.89	1.84	1.79	1.75	1.72
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.88	1.82	1.77	1.73	1.70
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.86	1.80	1.76	1.72	1.69
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.85	1.79	1.75	1.71	1.68
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.78	1.73	1.69	1.66
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.82	1.76	1.71	1.67	1.64
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.80	1.74	1.69	1.66	1.62
250	3.88	3.03	2.64	2.41	2.25	2.13	2.05	1.98	1.92	1.87	1.79	1.73	1.68	1.65	1.61
300	3.87	3.03	2.63	2.40	2.24	2.13	2.04	1.97	1.91	1.86	1.78	1.72	1.68	1.64	1.61
400	3.86	3.02	2.63	2.39	2.24	2.12	2.03	1.96	1.90	1.85	1.78	1.72	1.67	1.63	1.60
500	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.77	1.71	1.66	1.62	1.59
600	3.86	3.01	2.62	2.39	2.23	2.11	2.02	1.95	1.90	1.85	1.77	1.71	1.66	1.62	1.59
750	3.85	3.01	2.62	2.38	2.23	2.11	2.02	1.95	1.89	1.84	1.77	1.70	1.66	1.62	1.58
1000	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.89	1.84	1.76	1.70	1.65	1.61	1.58

منابع

آمار و کاربرد آن در مدیریت جلد اول	دکتر عادل آذر	انتشارات سمت
آمار و کاربرد آن در مدیریت جلد دوم	دکتر عادل آذر	انتشارات سمت
آمار و کاربرد آن در مدیریت جلد اول	دکتر خدیجه جمشیدی	دانشگاه پیام نور
آمار و کاربرد آن در مدیریت جلد دوم	دکتر خدیجه جمشیدی	دانشگاه پیام نور
آمار و احتمال	شلدون راس	دانشگاه فردوسی مشهد
احتمال و کاربرد آن	دکتر علی عمیدی	دانشگاه پیام نور
آمار مهندسی لیبرمن	دکتر هاشم محلوجی	مرکز نشر دانشگاهی