
بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

رشته‌های علوم پزشکی و شاخه‌های مرتبط

مطابق با سرفصل‌های مصوب

مطالعات پزشکی

مطالعات پزشکی

● طرح‌های مطالعاتی در پزشکی را می‌توان به دو دسته تقسیم کرد:

● ۱- مطالعات مشاهده‌ای: آن دسته از مطالعاتی هستند که بدون مداخله پژوهشگر انجام می‌گیرند و معمولاً مطالعه در یک دوره زمانی معین به انجام می‌رسد.

● ۲- مطالعات تجربی: به مطالعاتی گفته می‌شود که پژوهشگر به طور مستقیم در طرح و اجرای آزمایش دخالت دارد. مطالعات تجربی که بر روی انسان صورت می‌گیرد را کارآزمایی‌های بالینی (Clinical Trials) می‌گویند.

مطالعات پزشکی

مطالعات تجربی

(Experimental Studies)

کارآزمایی های بالینی

(Clinical trials)

- کار آزمایی شاهد دار

(Controlled trials)

- کار آزمایی بدون شاهد

(Uncontrolled trials)

مطالعات توصیفی

(Observational Studies)

- مجموعه موردی

(Case-Series)

- شاهد موردی

(Case-Control)

- مقطعی

(Cross-Sectional)

- همگروهی

(Cohort)

مطالعات پزشکی

① مطالعه مجموعه مورد: هنگامی است که خصوصیات معین و نکات جالب و غیر معمول یک گروه از بیماران در یک مقاله به صورت توصیفی گزارش شود.

② مطالعه شاهد موردی: هنگامی است که یک گروه از بیماران در یک دوره زمانی تحت مطالعه قرار می گیرند و جهت مطالعه رو به گذشته است. موردها بیماران و شاهد‌ها افراد سالم در مطالعه هستند. صحت این مطالعه به صحت مدارک پزشکی جمع آوری شده وابسته است.

مطالعات پزشکی

● مطالعات مقطعی: در این نوع مطالعه برای تعیین وضعیت یک بیماری یا هر فرایند وابسته به زمان (مانند تصادفات) اطلاعات گروهی از افراد را در یک زمان مشخص و کوتاه جمع‌آوری کرده و تحلیل می‌کنند.

● مطالعات همگروهی: این نوع مطالعه که برای بررسی علل یک وضعیت، روند یک بیماری یا عوامل خطر آن بیماری می‌باشد، یک مطالعه طولی است. در این مطالعه پژوهشگر در طول زمان مطالعه می‌تواند مدارکی قوی برای رابطه علت و معلولی به دست آورد.

مطالعات پزشکی

- کارآزمایی‌های شاهد دار: مطالعاتی را می‌گویند که یک دارو یا روش درمان، با دارو یا روش درمان دیگری مقایسه می‌گردد. در این نوع مطالعات همواره دو گروه از افراد در مطالعه قرار دارند. نحوه اختصاص گروه شاهد و گروه درمان، این مطالعات را به انواع دیگری تقسیم بندی می‌کند.
- کارآزمایی‌های بدون شاهد: به مطالعاتی گفته می‌شود که در آن پژوهشگران بر دارو یا روش آزمایش، بدون مقایسه با درمان دیگری، توجه می‌کنند. هدف هر آزمایش تعیین تفاوت مداخله پزشک (یا روش درمان) با روش قبل است.

مفاهيم

آمار علم جمع آوری، خلاصه کردن و طبقه بندی داده‌ها است و روش‌های تحلیلی را برای پیش بینی، برآورد و تصمیم گیری در شرایط مختلف ارائه می‌دهد.



از واژه آمار ممکن است معنی لیست و فهرستی از اشیاء یا افراد به ذهن متبادر شود که با علم آمار تفاوت دارد.

آمار علم تصمیم گیری در شرایط عدم اطمینان است.

کاربرد

● ما هنگامی به استفاده از علم آمار روی می آوریم که:

- (۱) بخواهیم داده‌ها را به صورت یک مجموعه خلاصه کرده و برای بهره‌برداری، توصیف نماییم.
- (۲) یا بخواهیم ادعایی را که در مورد پارامترهای یک جامعه مطرح است، آزمون کنیم و اطمینان دهیم که در شرایطی که یک پروژه، دقیقاً به همان شکل تکرار شود، همان یافته‌ها بدست خواهد آمد.

آمار استنباطی

- احتمال
- توابع توزیع
- نمونه گیری
- برآوردها
- آزمون ها

آمار توصیفی

- ساماندهی داده ها
- شاخص های
- مرکزی و پراکندگی
- نمودارها

● **آمار توصیفی:** به مجموعه روش‌هایی که برای سازماندهی و خلاصه کردن و توصیف اطلاعات به کار می‌رود، گفته می‌شود. (رسم نمودار - تهیه جدول‌های توزیع فراوانی - شاخص‌های مرکزی و پراکندگی)

● **آمار استنباطی:** به مجموعه روش‌هایی که به وسیله آنها نتایج به دست آمده از نمونه را به جامعه تعمیم می‌دهند، گفته می‌شود. (احتمال - توابع توزیع - نمونه‌گیری - برآوردها - آزمون فرض)

مفاهیم مقدماتی

● **جامعه آماری:** متشکل از اعضاء و اشیائی است که حداقل در یک ویژگی مشترک باشند و به خوبی تعریف و مشخص شده باشند. یک جامعه آماری ممکن است مجموعه‌ای از انسانها، حیوانات، اشیاء و... باشد. تعداد اعضاء جامعه را N نشان می‌دهند.

● **نمونه:** به بخشی از جامعه آماری می‌گویند که همه ویژگی‌های جامعه را داشته باشد. در صورتی که جمع آوری اطلاعات از همه جمعیت امکان پذیر نباشد، به روش‌های علمی یک نمونه برای مطالعه انتخاب می‌کنند. تعداد اعضاء نمونه را n نشان می‌دهند.

مفاهیم مقدماتی

● **صفت آماری:** در یک جامعه آماری هر عضو تحت مطالعه، دارای ویژگی‌هایی است که از آنها برای توصیف یک عضو استفاده می‌کنیم. در آمار به هر ویژگی یک صفت می‌گویند. مثلاً سن، وزن، فشارخون و... صفتهایی برای یک بیمار هستند.

● **صفت مشخصه:** صفتهایی هستند که در همه‌ی اعضای جامعه مشترک‌اند و حیطه‌ی مطالعه و تحقیق را معلوم می‌کنند. مانند بیماران زن مبتلا به سرع با سن بیشتر از سی سال بستری شده در بیمارستان A

مفاهیم مقدماتی

● متغیرها و داده‌ها: هر ویژگی مربوط به یک عضو جامعه را که عملاً مورد پرسش یا اندازه‌گیری قرار می‌گیرد، یک صفت متغیر یا به اختصار یک متغیر (Variable) می‌گویند. بنابراین داده‌ها (Data) مقادیر اندازه‌گیری شده متغیرها هستند.

متغیرها

کمی

کیفی

● متغیرهای کمی: متغیرهایی هستند که قابل شمارش و اندازه‌گیری‌اند و حاصل سنجش آنها یک مقدار کمی است مانند وزن، قد، فشار خون و مانند اینها.

● متغیرهای کیفی: متغیرهایی هستند که غیر قابل شمارش و اندازه‌گیری‌اند و حاصل سنجش آنها یک حالت و وضعیت است. مثل رنگ چشم، گروه خون، مهارت و از این دست.

مفاهیم مقدماتی

متغیرها

کمی

کیفی

پیوسته

گسسته

رتبه‌ای

اسمی

مفاهیم مقدماتی

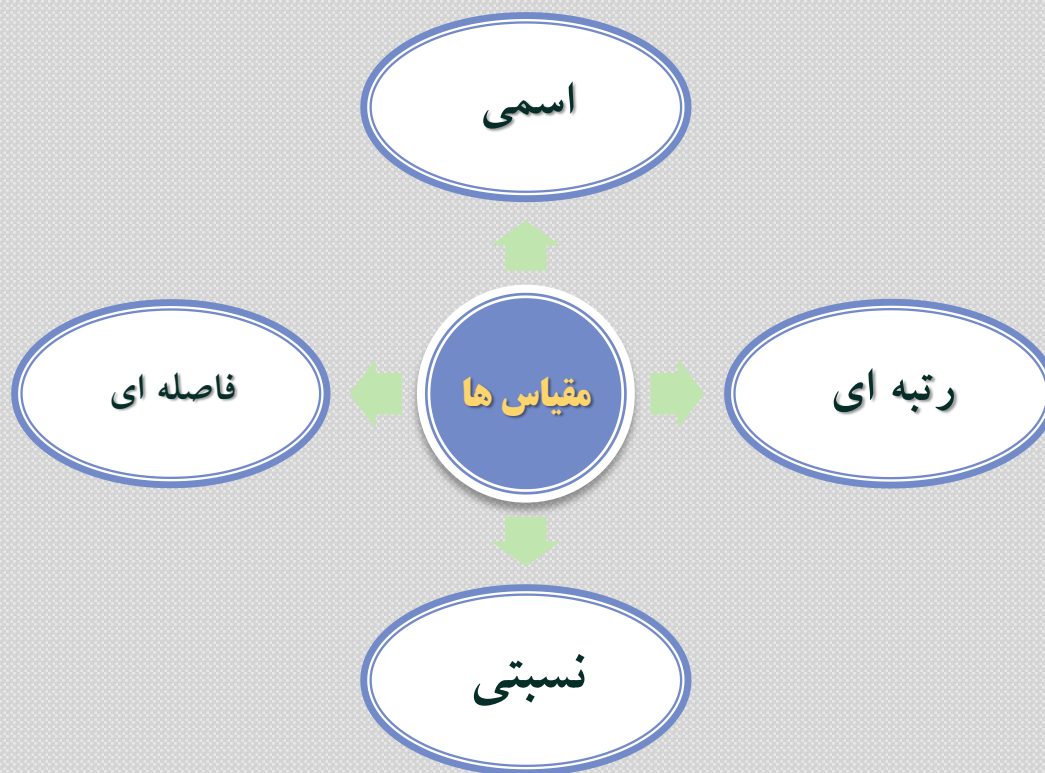
متغیرهای کمی:

دامنه وسیعی از تحلیل‌های آماری به متغیرهای کمی اختصاص دارند. متغیرهای کمی متغیرهایی هستند که همیشه نتیجه اندازه‌گیری آنها، یک عدد است. مثل تعداد فرزندان یک خانواده، دمای بدن و مانند آن.

در این میان متغیرهایی که فقط قابل شمارش هستند متغیر کمی **گسسته** و متغیرهایی را که اندازه پذیرند و مقادیر آنها در فاصله‌ای از اعداد حقیقی قرار دارد، متغیر کمی **پیوسته** می‌گوییم.

مقیاس‌های اندازه‌گیری

● برای سنجش متغیرها از مقیاس‌های مختلفی استفاده می‌شود.



مقیاس های اندازه گیری

- مقیاس اسمی: در این مقیاس، هر عدد نشان دهنده وضعیت یا حالت متغیر است و هیچ کمّیتی را بیان نمی کند
- مقیاس رتبه‌ای: در این مقیاس، هر عدد نشان دهنده ترتیب و توالی سطوح متغیر است.
- مقیاس فاصله‌ای: در این مقیاس، اعداد بیانگر یک کمّیت‌اند ولی مبداء معینی ندارند.
- مقیاس نسبتی: در این مقیاس، اعداد دقیقاً بیانگر کمّیتی هستند که شامل یک مبداء معین اس.

دسته بندی اطلاعات

آمار توصیفی

● در آمار توصیفی برای توصیف بهتر اطلاعات سه کار عمده انجام می دهیم:

دسته بندی اطلاعات و تهیه جدول توزیع فراوانی

به دست آوردن شاخص مناسبی که بتواند ویژگی داده ها را به خوبی آشکار کند.

رسم نمودار مناسب که توصیف بصری اطلاعات طبقه بندی شده را نشان دهد.

دسته بندی اطلاعات

● اگر داده‌های مطالعه از یک متغیر کیفی هستند، هنگام گردآوری اطلاعات با دنباله‌ای از وضعیت‌ها یا حالت‌ها رو برو هستید. در این حال کافی است سطوح متغیر و فراوانی هر سطح را در یک جدول قرار دهید.

مثال ۱

گروه خون، متغیری کیفی است. اطلاعات مربوط به گروه خون ۱۰۰ نفر را در اختیار داریم. اطلاعات را در جدولی به صورت زیر دسته بندی کرده‌ایم.

گروه خون	AB	O	B	A
فراوانی fi	۳۵	۲۰	۳۲	۱۳

به چنین جدولی جدول توزیع فراوانی می‌گویند.

دسته بندی اطلاعات

● اگر داده‌ها مقادیر کمی ولی گسسته هستند، به طور معمول دارای تکرار خواهند بود. برای دسته بندی کردن این گونه داده‌ها مقادیر مشاهده شده و فراوانی آنها را در کنار یکدیگر ثبت می‌کنیم تا جدول توزیع فراوانی به دست آید.

مثال ۲

فراوانی f_i	تعداد دندان‌ها
۲	۰
۹	۱
۱۸	۲
۱۰	۳
۷	۴
۴	۵ و بیشتر

تعداد دندان‌های ترمیم شده ۵۰ بیمار مراجعه کننده به یک مرکز دندانپزشکی را به صورت جدول توزیع فراوانی زیر مرتب کرده‌ایم.

دسته بندی اطلاعات

- در صورتی که داده‌ها کمی پیوسته باشند، باید مراحل زیر را برای دسته بندی آنها دنبال کنید.
- ۱- تعدادی طبقه را به تناسب تعداد داده‌ها و یا با استفاده از دستور زیر تعیین کنید:

$$k = 1 + 3/322 \log(n)$$

- ۲- دامنه تغییرات داده‌ها را از رابطه:

$$R = \text{Max}(x_i) - \text{Min}(x_i)$$

محاسبه کنید.

دسته بندی اطلاعات

۳- فاصله طبقه‌ها را از رابطه زیر به دست آورید:

$$c = R/k \quad (\text{با تقریب اضافی})$$

۴- حدود هر طبقه را تعیین کنید و داده‌هایی که در هر طبقه قرار می‌گیرند، شمارش کرده و به عنوان فراوانی آن طبقه ثبت کنید.

مثال ۳

داده‌های زیر فشار خون ۲۵ بیمار بستری بر حسب میلیمتر جیوه است که می‌خواهیم آن‌ها را دسته بندی کنیم. (اعداد با تقریب کمتر از یک گرد شده اند)

۱۰۰	۱۰۱	۹۷	۱۰۴	۱۰۲	۱۱۰	۱۰۳	۱۰۶	۱۱۰	۱۰۴	۱۰۳	۹۸	۱۰۵
۱۰۰	۱۰۹	۱۰۳	۱۰۴	۹۹	۹۸	۱۰۹	۱۰۵	۱۰۳	۱۱۰	۱۰۴	۱۰۵	

دسته بندی اطلاعات

برای این داده ها ۵ طبقه در نظر می گیریم که پس از محاسبات لازم، دامنه تغییرات را ۱۳ و فاصله هر طبقه را پس از تقریب، ۳ به دست می آوریم.

$$c = \frac{r}{k} = \frac{13}{5} = 2.6 \approx 3$$

حدود طبقه	فراوانی f_i
۹۷ - ۱۰۰	۴
۱۰۰ - ۱۰۳	۴
۱۰۳ - ۱۰۶	۱۱
۱۰۶ - ۱۰۹	۱
۱۰۹ - ۱۱۲	۵

اگر حدود هر طبقه را تعیین کنیم و داده هایی را که در هر طبقه قرار می گیرند، شمارش نماییم، جدول توزیع فراوانی به صورت جدول مقابل به دست خواهد آمد. این جدول شامل $k = 5$ طبقه است.

جدول توزیع فراوانی

در یک جدول توزیع فراوانی می‌توانیم فراوانی‌های دیگری را نیز به دست آوریم.

- فراوانی مطلق: به تعداد داده‌ها در هر طبقه فراوانی مطلق آن طبقه می‌گویند و آن را با f_i نشان می‌دهند.
- فراوانی نسبی: در صورتی که فراوانی‌های مطلق را بر کل فراوانی‌ها تقسیم کنیم، فراوانی نسبی Γ_i به دست می‌آید.
- فراوانی تجمعی: به مجموع فراوانی‌های مطلق طبقه‌های قبل و همان طبقه، فراوانی تجمعی آن طبقه می‌گویند و آن را با F_i نمایش می‌دهند.
- فراوانی تجمعی نسبی: می‌توان از تقسیم فراوانی‌های تجمعی بر تعداد داده‌ها، این فراوانی را به دست آورد. (R_i)
- نماینده طبقه: نقطه وسط هر طبقه را نماینده آن طبقه می‌گویند و آن را با x'_i نشان می‌دهند.

جدول توزیع فراوانی

● جدول توزیع فراوانی مثال قبل (فشار خون ۲۵ بیمار بستری) به صورت زیر است:

حدود طبقات	f_i	ri	Fi	Ri	x'_i
۹۷ - ۱۰۰	۴	۰,۱۶	۴	۰,۱۶	۹۸/۵
۱۰۰ - ۱۰۳	۴	۰,۱۶	۸	۰,۳۲	۱۰۱/۵
۱۰۳ - ۱۰۶	۱۱	۰,۴۴	۱۹	۰,۷۶	۱۰۴/۵
۱۰۶ - ۱۰۹	۱	۰,۰۴	۲۰	۰,۸	۱۰۷/۵
۱۰۹ - ۱۱۲	۵	۰,۲۰	۲۵	۱	۱۱۰/۵



تمرین

دانشجو پس از پایان این فصل باید بتواند به
سوالات این قسمت پاسخ دهد

۱- نوع هریک از متغیرهای زیر را معلوم کنید. (متن درس)

میزان علاقمندی به درس	درآمد خانواده
مارک اتومبیل	موسیقی مورد علاقه
تعداد کتابهایی که در یکماه گذشته مطالعه کرده اید	مدت تماشای تلویزیون در روز
تعداد اتاقهای منزل	شغل
سطح سواد	تعداد بیماران مراجعه کننده به یک کلینیک در روز
رنگ چشم	فشار خون
سود سالانه یک شرکت	رضایت از ازدواج
رشته تحصیلی	وزن نوزاد هنگام تولد
قد نوزاد هنگام تولد	تعداد تصادفات در یک روز
تعداد خطاهای یک سیستم در طول یک روز کاری	رتبه شما در کلاس

تمرین

- ۲- در یک جدول توزیع فراوانی اگر فراوانی تجمعی طبقه آخر ۲۵ و فراوانی مطلق یک طبقه قبل آن ۳ باشد، فراوانی تجمعی و فراوانی نسبی طبقه ماقبل آخر را حساب کنید. (متن درس)
- ۳- تفاوت متغیرهای کیفی و کمی را توضیح دهید. (متن درس)
- ۴- آیا متغیرهای کیفی و کمی قابل تبدیل به یکدیگر هستند؟ (متن درس)
- ۵- چه تفاوتی بین صفت‌های مشخصه و متغیر وجود دارد؟ (متن درس)
- ۶- تعداد دندان‌های ترمیمی ۴۶ بیمار مراجعه کننده به یک مرکز دندانپزشکی در یک روز به شرح زیر است.

۴	۳	۰	۳	۴	۱	۱	۲	۳	۵	۳	۳	۴	۱	۲	۲	۰	۵	۱	۳	۴	۵	۲
۱	۱	۲	۲	۱	۲	۸	۰	۵	۴	۷	۱	۱	۴	۱	۲	۳	۴	۵	۲	۲	۰	۳

جدول توزیع فراوانی شامل فراوانی‌های مطلق، تجمعی و نسبی را تشکیل دهید.
(متن درس)

پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

**رشته‌های علوم پزشکی و شاخه‌های مرتبط
مطابق با سرفصل‌های مصوب**

شاخص‌های توصیفی

شاخص های توصیفی

● شاخص یا معیار:

مشخص کننده های عددی شاخص هایی هستند که برای تعیین میزان تمرکز یا تفرق داده ها از آنها استفاده می کنیم و به طور کلی دو دسته هستند.

۱- شاخص های گرایش به مرکز: شاخص هایی هستند که میزان تمرکز داده ها را معین می کنند.

۲- شاخص های پراکندگی: به شاخص هایی که میزان پراکندگی داده ها از مرکز را نمایش می دهند، گفته می شود.

شاخص‌های توصیفی

شاخص‌های پراکندگی

- واریانس
- انحراف معیار
- دامنه تغییرات
- ضریب تغییرات
- ضریب چولگی

شاخص‌های مرکزی

- میانگین
- میانه
- نما
- چندکها

شاخص‌های مرکزی

شاخص‌های مرکزی

شاخص‌هایی عددی هستند که میزان گرایش به مرکز داده‌ها را مشخص می‌کنند.

● میانگین حسابی (Mean): یکی از مهمترین شاخص‌های مرکزی است که از رابطه زیر به دست می‌آید:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + x_2 + \dots + x_n)$$

● در صورتی که داده‌ها را دسته بندی کرده باشیم؛ فرمول میانگین به صورت زیر تغییر خواهد کرد:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k f_i x_i = \frac{1}{n} (f_1 x_1 + f_2 x_2 + \dots + f_k x_k)$$

شاخص های مرکزی

ویژگی های میانگین:

۱- همیشه مجموع اختلاف داده ها از میانگین حسابی، صفر است.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

۲- هرگاه به هر یک از داده ها عدد ثابتی را مانند a اضافه یا کم کنیم، میانگین داده های حاصل، برابر با میانگین داده های قبلی به اضافه (یا منهای) عدد ثابت a خواهد بود.

$$Y_i = X_i \pm a \rightarrow \bar{Y} = \bar{X} \pm a$$

شاخص های مرکزی

۳- هرگاه عدد ثابت a را در هریک از داده ها ضرب کنیم، میانگین داده های حاصل، برابر میانگین داده های قبلی ضربدر عدد ثابت a خواهد بود.

$$Y_i = aX_i \Rightarrow \bar{Y} = a\bar{X}$$

۴- درمورد تقسیم هم به صورت ضرب عمل می کنیم.

$$Y_i = \frac{X_i}{a} \rightarrow \bar{Y} = \frac{\bar{X}}{a}$$

شاخص های مرکزی

۵- اگر یک مجموعه n_1 تائی از داده های یک صفت با میانگین $\bar{X}_1$ و یک مجموعه n_2 تائی از داده ها با میانگین $\bar{X}_2$ داشته باشیم، آنگاه میانگین کل این دو مجموعه برابر است با:

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

این موضوع برای k جمعیت و به صورت زیر قابل تعمیم است:

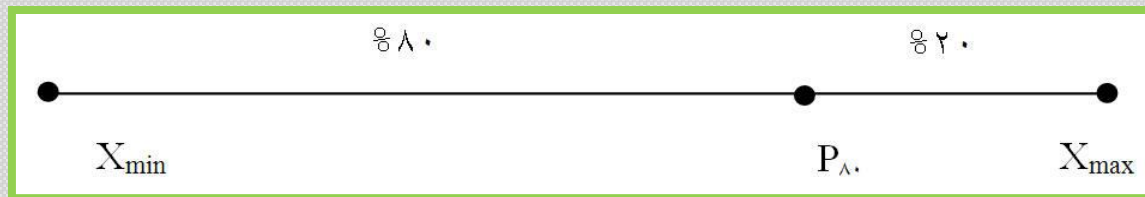
$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + \dots + n_k \bar{X}_k}{n_1 + n_2 + \dots + n_k}$$

شاخص های مرکزی

● میانه: (Median) عددی است در بین داده‌ها که اگر آنها را مرتب کرده باشیم، (بطور غیر نزولی) نیمی از داده‌ها کمتر از آن واقع شوند. به عبارت دیگر داده‌ای که در وسط قرار می‌گیرد و آن را با Md نشان می‌دهند.

البته باید به تعداد داده‌ها که فرد یا زوج باشند نیز توجه داشت.

● چندک: (Percentile) هر چندک با یک نسبت (p) معین می‌شود که با ضرب آن در تعداد داده‌ها مکان قرار گرفتن چندک مشخص می‌شود.



شاخص های مرکزی

چندکها در داده های پیوسته طبقه بندی شده از رابطه زیر به دست می آید:

$$Q = L_i + \frac{n.p - F_{i-1}}{f_i} \times C$$

نما: نما یا مد (Mode) داده ای است از بین داده ها که فروانی آن بیشتر از سایر داده ها باشد و آن را با Mo نمایش می دهیم .
اگر داده ها طبقه بندی شده باشند، نما از رابطه زیر محاسبه می شود:

$$Mo = l_i + \frac{D_1}{D_1 + D_2} \times C$$

$$D_1 = f_i - f_{i-1}$$

$$D_2 = f_i - f_{i+1}$$

که در این رابطه:

شاخص های مرکزی

مثال: شاخص های مرکزی را برای داده های جدول زیر به دست آورده ایم.

تعداد دندانها	فراوانی f_i	$x_i \cdot f_i$	F_i
۰	۲	$۰ \times ۲ = ۰$	۲
۱	۹	$۱ \times ۹ = ۹$	۱۱
۲	۱۸	$۱۸ \times ۲ = ۳۶$	۲۹
۳	۱۰	$۱۰ \times ۳ = ۳۰$	۳۹
۴	۷	$۴ \times ۷ = ۲۸$	۴۶
۵	۴	$۴ \times ۵ = ۲۰$	۵۰
	۵۰	۱۲۳	

- در این مثال، نما ۲ است. زیرا فراوانی ۲ بیشتر از سایر داده ها است.
- میانگین $۲/۴۶$ از تقسیم ۱۲۳ بر ۵۰ به دست می آید.
- میانۀ از معدل دو عدد بیست و پنجم و بیست و ششم به دست می آید؛ یعنی $md = ۲$
- بقیه چندک ها را می توان به همین صورت به دست آورد.

شاخص های مرکزی

مثال: برای داده های جدول فراوانی زیر شاخص های مرکزی را به دست آورده ایم.
 میانگین $6/48$ ؛ که از تقسیم 324 بر 50 به دست آمده است.

همچنین نما $7/5$ و در طبقه چهارم است که آن را از رابطه زیر به دست آورده ایم:

حدود طبقات	فراوانی	نماینده		Fi
۱ - ۳	۸	۲	$8 \times 2 = 16$	۸
۳ - ۵	۵	۴	$4 \times 5 = 20$	۱۳
۵ - ۷	۱۳	۶	$13 \times 6 = 78$	۲۶
۷ - ۹	۱۵	۸	$15 \times 8 = 120$	۴۱
۹ - ۱۱	۹	۱۰	$10 \times 9 = 90$	۵۰
	۵۰		۳۲۴	

$$Mo = l_i + \frac{D_1}{D_1 + D_2} \times C \Rightarrow 7 + \frac{2}{2 + 6} \times 2 = 7/5$$

Li = کران پائین طبقه ای است که فراوانی مطلق آن ماکزیمم است. (طبقه نمادار)

D_1 = تفاضل فراوانی مطلق طبقه نمادار با طبقه قبل از آن است، یعنی $D_1 = f_i - f_{i-1} = 15 - 13 = 2$

D_2 = تفاضل فراوانی مطلق طبقه نمادار با طبقه بعد از آن است. یعنی $D_2 = f_i - f_{i+1} = 15 - 9 = 6$

C = فاصله طبقات

شاخص های پراکندگی

شاخص های پراکندگی

شاخص هایی عددی هستند که میزان پراکندگی داده ها حول میانگین را نشان می دهند.

X_i	f_i	X_i	f_i
۲	۲	۴	۵
۳	۳	۵	۱۰
۴	۱۵	۶	۷۰
۵	۱۰	۷	۱۰
۶	۱۰	۸	۵
۷	۴۰		
۸	۱۵		
۹	۳		
۱۰	۲		

مثال: دو گروه از بیماران را از نظر مصرف تعداد قرص های مسکن در یک ماه مورد بررسی قرار داده ایم که نتایج در دو جدول زیر آمده است.

برای این دو گروه، میانگین، میانه و نما یکسان و برابر ۶ است.

در این دو مجموعه داده، تفاوت هایی را مشاهده می کنید که با شاخص های مرکزی قابل بیان نیستند.

شاخص های پراکندگی

● واریانس (Variance): واریانس یکی از معیارهای بسیار مناسب در بیان میزان پراکندگی داده‌ها است. زیرا برخلاف بعضی از شاخص‌ها، از همه داده‌ها برای محاسبه آن استفاده می‌شود. واریانس را می‌توان از یکی از دو رابطه زیر به دست آورد:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1)$$

$$S^2 = \overline{x^2} - \bar{x}^2 \quad (2)$$

مثال: برای داده‌های زیر واریانس را به دست می‌آوریم. ۳ ۲ ۴ ۳ ۶ ۳ ۵ ۶
ابتدا میانگین داده‌ها را حساب می‌کنیم و سپس از فرمول (۱) واریانس را محاسبه می‌کنیم، داریم:

$$\bar{x} = \frac{1}{8} (3 + 2 + 4 + 3 + 6 + 3 + 5 + 6) = \frac{32}{8} = 4$$

$$S^2 = \frac{1}{8} [(3-4)^2 + (2-4)^2 + (4-4)^2 + (3-4)^2 + (6-4)^2 + (3-4)^2 + (5-4)^2 + (6-4)^2] = \frac{1}{8} (16) = 2$$

توجه کنید: در این مثال داده‌ها دسته‌بندی نشده‌اند.

شاخص های پراکندگی

مثال:

برای داده‌های جدول توزیع فراوانی زیر واریانس را محاسبه کنید.

ابتدا با تشکیل دو ستون آخر جدول مقادیر زیر را به دست می‌آوریم:

$$\bar{X} = \frac{125}{50} = 2.5 \quad \overline{X^2} = \frac{407}{50} = 8.14$$

سپس واریانس از فرمول (۲) به صورت زیر بدست می‌آید.

$$S^2 = 8.14 - (2.5)^2 = 1.89$$

X_i	f_i	$f_i X_i$	$f_i X_i^2$
۰	۳	۰	$۰ \times ۳ = ۰$
۱	۷	۷	$۱ \times ۷ = ۷$
۲	۲۱	۴۲	$۴ \times ۲۱ = ۸۴$
۳	۶	۱۸	$۹ \times ۶ = ۵۴$
۴	۷	۲۸	$۱۶ \times ۷ = ۱۱۲$
۵	۶	۳۰	$۲۵ \times ۶ = ۱۵۰$
جمع	۵۰	۱۲۵	۴۰۷

توجه کنید: در این مثال داده‌ها گسسته و دسته بندی شده‌اند.

شاخص های پراکندگی

خواص واریانس

$$x_1 = x_2 = \dots = x_n \Rightarrow S_x^2 = 0$$

۱- اگر همه داده ها باهم برابر باشند واریانس داده ها صفر است و برعکس

$$S^2 \geq 0$$

۲- واریانس داده ها مقداری مثبت است.

$$Y_i = a \pm X_i \rightarrow S_Y^2 = S_X^2$$

۳- اگر به داده ها عدد ثابتی مانند a را اضافه یا کم کنیم، واریانس تغییری نمی کند.

$$Y_i = ax_i \Rightarrow S_Y^2 = a^2 S_X^2$$

۴- اگر داده ها را در عدد ثابتی مانند a ضرب کنیم، واریانس آنها در مجذور a ضرب می شود.

$$Y_i = \frac{X_i}{a} \rightarrow S_Y^2 = \frac{S_X^2}{a^2}$$

به همین صورت در مورد تقسیم بر عدد ثابتی مانند a داریم:

شاخص های پراکندگی

Std Deviation

انحراف معیار

انحراف معیار از جذر مثبت واریانس به دست می آید و آن را با S نشان می دهیم و می نویسیم: $S = \sqrt{S^2}$

انحراف معیار مثال قبل به این صورت به دست می آید.
 $S = \sqrt{19/57} = 4/42$

ضریب تغییرات

ضریب تغییرات داده ها که آن را با CV نشان می دهیم از نسبت انحراف معیار به میانگین به دست می آید و به صورت زیر می نویسیم.

$$C.V = \frac{S}{\bar{x}} \times 100$$

مثال: اگر در یک نمونه، میانگین داده ها ۵ و واریانس داده ها ۴ باشد؛ ضریب تغییرات داده ها به

صورت زیر به دست می آید:

$$C.V = \frac{2}{5} \times 100 = 40$$

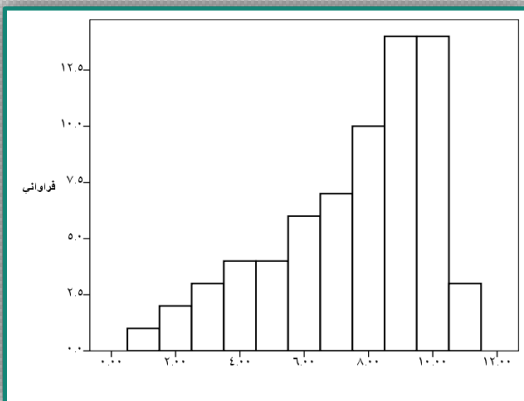
شاخص های پراکندگی

Skewness

تقارن و چولگی

ممکن است دو مجموعه داده، از لحاظ گرایش به مرکز و پراکندگی تفاوت چندانی با هم نداشته باشند ولی از نظر تقارن با یکدیگر متفاوت باشند. شاخص های دیگری برای اندازه گیری میزان تقارن یا عدم تقارن داده ها وجود دارند.

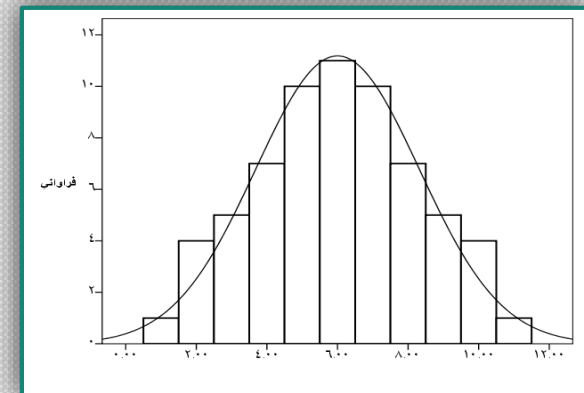
عدم تقارن (چولگی)



توجه داشته باشید:
در توزیع های متقارن
تقریباً میانگین، میانه و
نما بر هم منطبق اند.

$$Mo = \bar{x} = Md$$

تقارن در توزیع داده ها



شاخص های پراکندگی

ضریب چولگی پیرسن

یکی از شاخص های تعیین میزان تقارن و عدم تقارن داده ها، ضریب چولگی پیرسن است. اگر $\bar{X}$ میانگین، Md میانه و S انحراف معیار یک مجموعه داده باشند؛ ضریب چولگی که منسوب به پیرسن است، از رابطه زیر بدست می آید:

$$SK = \frac{3(\bar{X} - Md)}{S} \quad \text{ضریب چولگی}$$

$$S = 4/42$$

$$\bar{X} = 18/27$$

$$Md = 19/4$$

مثال: اگر میانگین، میانه و انحراف معیار داده ها را به صورت مقابل در اختیار داشته باشیم؛ ضریب چولگی به صورت زیر به دست می آید.

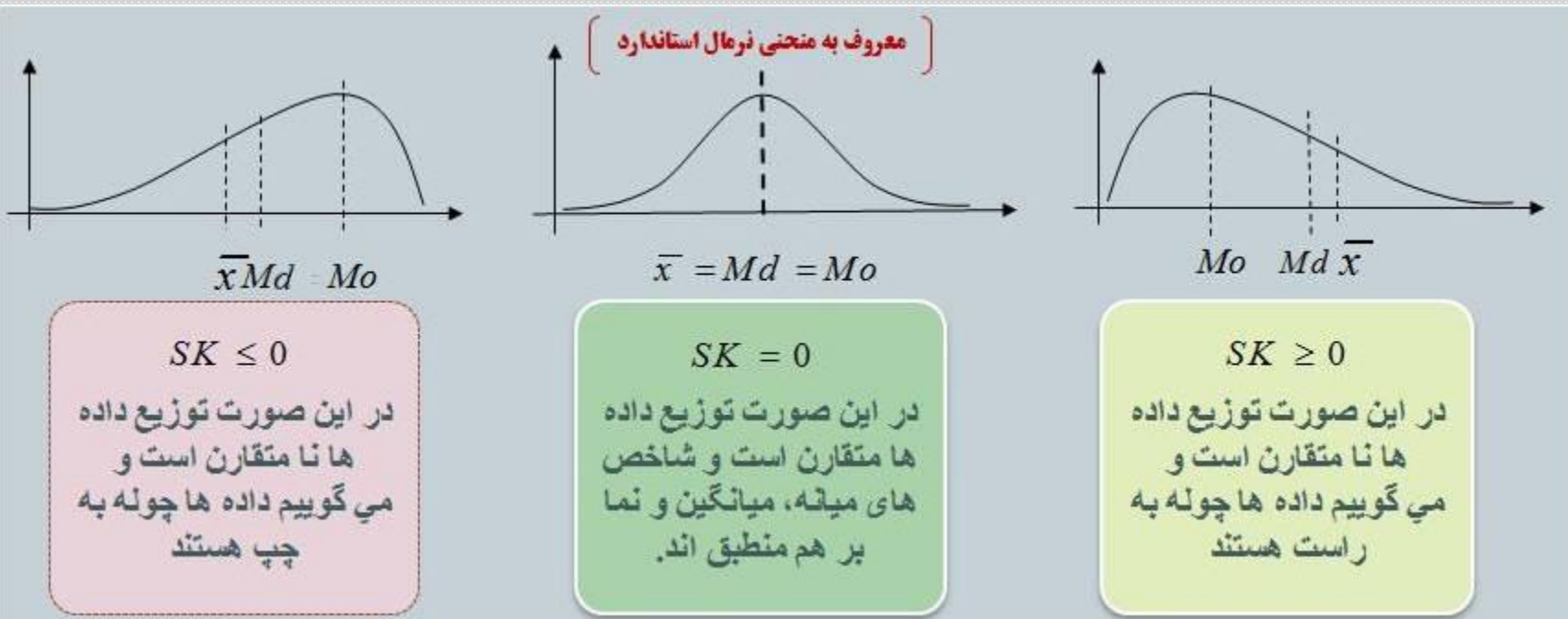
$$Sk = \frac{3(\bar{X} - Md)}{S} = \frac{3(18/27 - 19/4)}{4/42} = -0/77$$

ضریب چولگی پیرسن

شاخص های پراکندگی

تفسیر چولگی

ضریب چولگی ممکن است مقداری منفی، مثبت یا صفر باشد. بر این اساس وضعیت قرار گرفتن سه شاخص مهم مرکزی، میانگین، میانه و نما را در حالت های مختلف ببینید.



شاخص های پراکندگی

محدوده تجربی مشاهدات

اگر داده‌ها تقریباً متقارن باشند، در صورتی که میانگین و انحراف معیار را در اختیار داشته باشیم، می‌توان به طور تجربی، فاصله‌های زیر را برای داده‌ها در نظر گرفت:

در حدود ۶۸ درصد داده‌ها در فاصله $(\bar{x} - s, \bar{x} + s)$ قرار دارند.

در حدود ۹۵ درصد داده‌ها در فاصله $(\bar{x} - 2s, \bar{x} + 2s)$ قرار دارند.

در حدود ۹۹ درصد داده‌ها در فاصله $(\bar{x} - 3s, \bar{x} + 3s)$ قرار دارند.

مثال: توزیع فشار خون بیماران دیابتی تقریباً متقارن است. اگر میانگین فشار سیستولیک آنها ۱۵ و واریانس $2/25$ باشد، تقریباً ۶۸ درصد این بیماران فشار خونسشان در چه حدودی است؟
تقریباً ۶۸ درصد فشار خون در فاصله $(13/5, 16/5)$ قرار دارند، زیرا:

$$(\bar{x} - s, \bar{x} + s) = (15 - 1/5, 15 + 1/5) = (13/5, 16/5)$$

نمودارهای آماري

نمودارهای آماری

برای توصیف بهتر داده‌های آماری از نمودارها استفاده می‌کنیم. غالباً این نمودارها در ارتباط با داده‌های پیوسته به کار می‌گیرند.

منظور از نمایش نموداری داده‌ها، تجسم عینی اطلاعات نهفته در آنهاست.

تعدادی از نمودارهای مهم آماری

- ۱- نمودار دایره ای
- ۲- نمودار ستونی
- ۳- نمودار مستطیلی یا هیستوگرام
- ۴- نمودار چند ضلعی فراوانی مطلق
- ۵- نمودار چند ضلعی فراوانی تجمعی

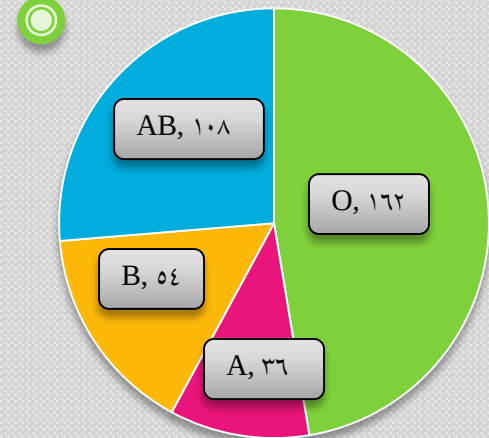
نمودارهای آماری

نحوه رسم نمودار دایره‌ای را در این مثال ببینید.

مثال: گروه خون ۱۰۰ بیمار بستری در یک بیمارستان را در اختیار داریم، نمودار دایره‌ای داده‌ها را رسم می‌کنیم.

گروه خون	فراوانی f_i	فراوانی نسبی f_i	قطاع دایره (درجه)
O	۴۵	۰/۴۵	۱۶۲
A	۱۰	۰/۱	۳۶
B	۱۵	۰/۱۵	۵۴
AB	۳۰	۰/۳	۱۰۸

$$۰/۱۵ \times ۳۶۰ = ۵۴$$

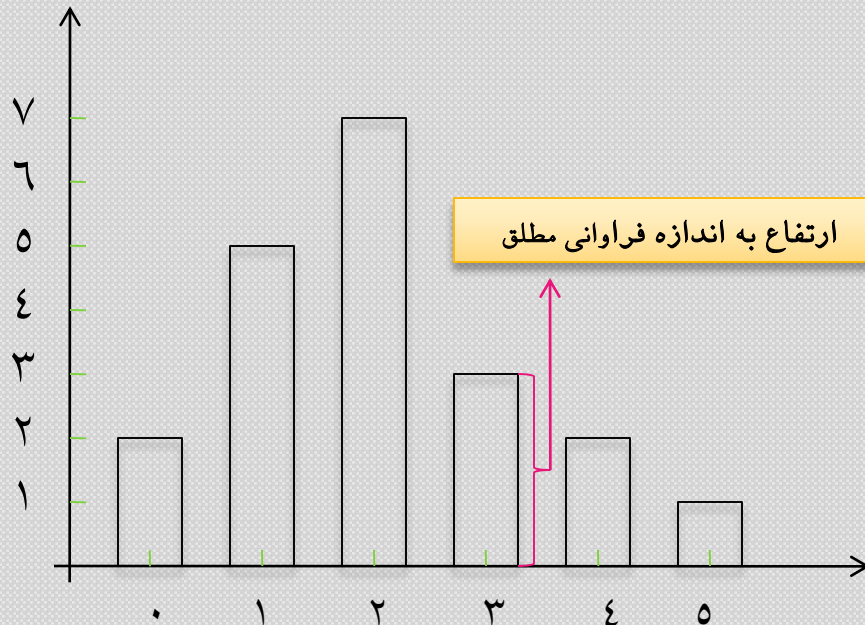


نمودارهای آماری

روش رسم نمودار ستونی

مثال برای داده های زیر نمودار ستونی رسم کنید.

۳ ۱ ۲ ۲ ۰ ۱ ۲ ۳ ۲ ۱ ۱ ۵ ۲ ۰ ۱ ۲ ۲ ۴ ۴ ۳



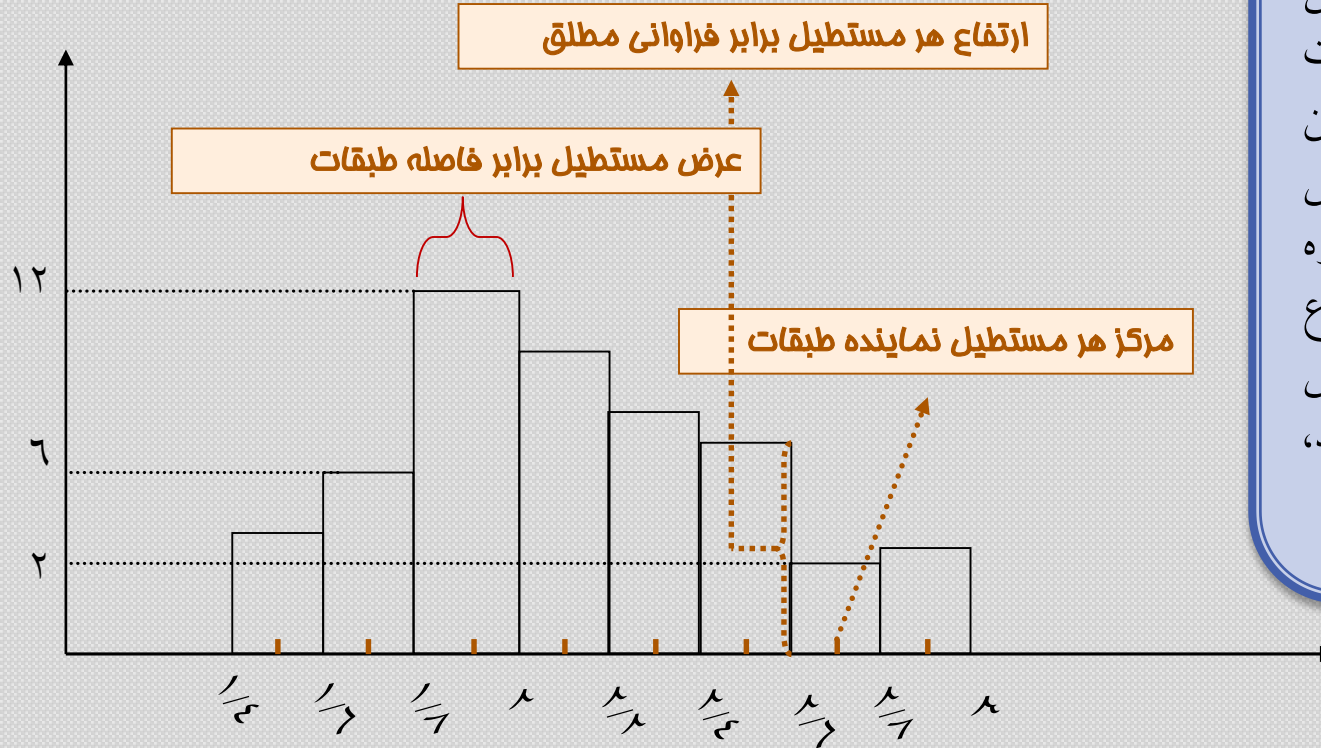
x	fi
۰	۲
۱	۵
۲	۷
۳	۳
۴	۲
۵	۱

برای رسم نمودار ستونی کافی است مقادیر متغیر را روی محور افقی تعیین کنید سپس مستطیل هایی با فاصله و به ارتفاع فراوانی مطلق، برای هر مقدار متغیر رسم کنید.

ابتدا جدول توزیع فراوانی را تشکیل می دهیم

نمودارهای آماری

روش رسم نمودار مستطیلی

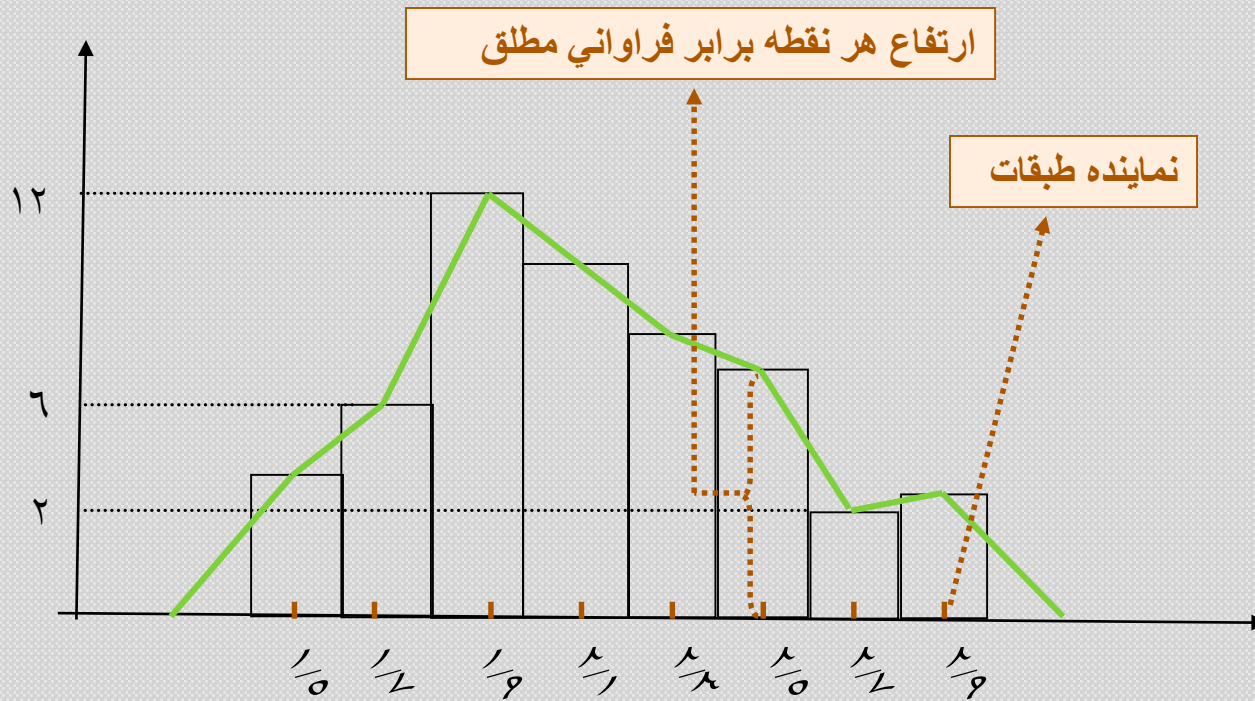


برای رسم هیستوگرام فراوانی یا نمودار مستطیلی داده ها ابتدا حدود طبقات را روی محور افقی تعیین می کنیم. بعد مستطیل هائی که عرض آنها به اندازه فاصله طبقات و ارتفاع آنها به اندازه فراوانی مطلق طبقه مربوطه باشد، رسم می کنیم.

نمودارهای آماری

روش رسم نمودار چند ضلعی

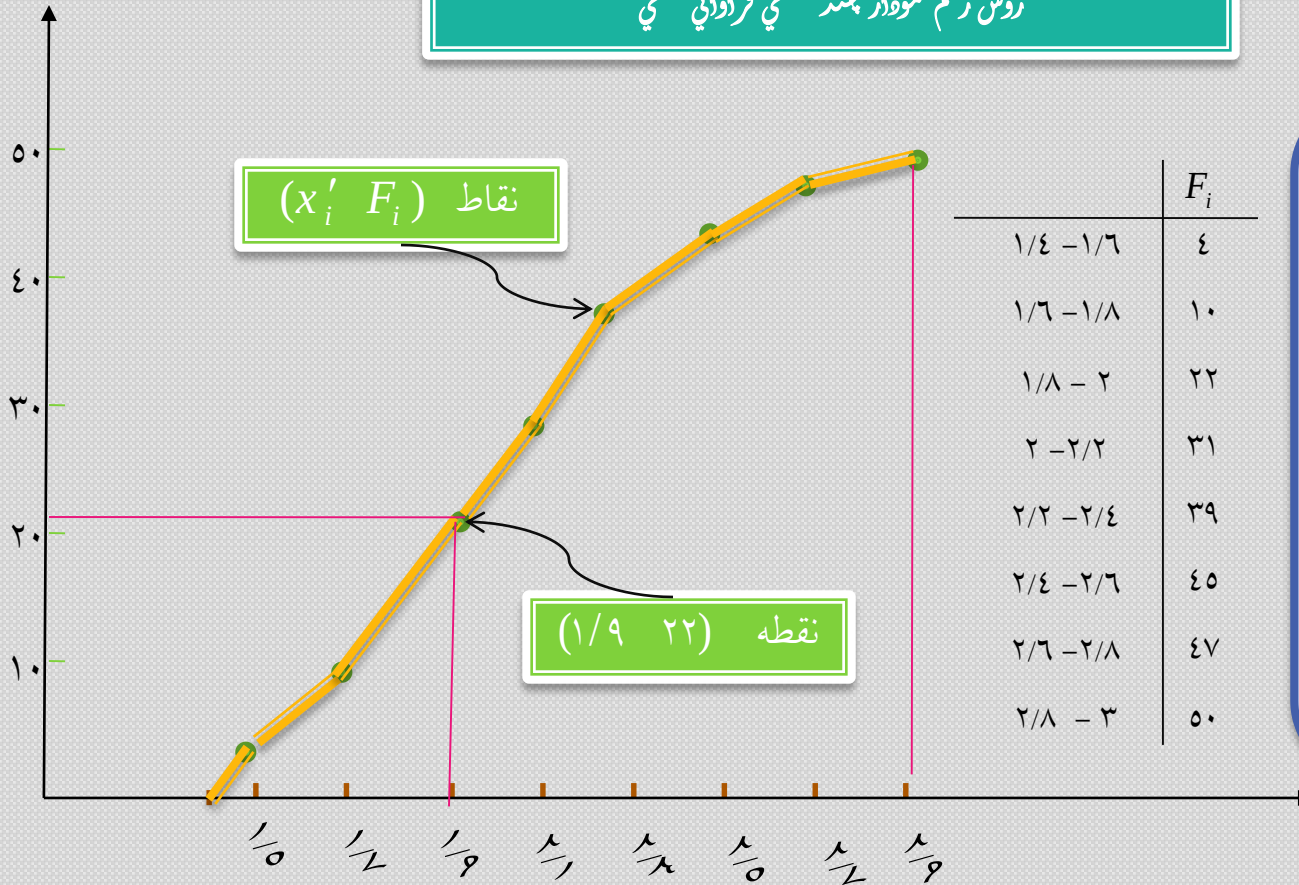
برای درک بهتر نمودار چند ضلعی، مستطیل‌ها را روی این نمودار قرار داده‌ایم.



برای رسم چند ضلعی فراوانی داده‌ها نماینده طبقات را روی محور افقی تعیین می‌کنیم. بعد نقاطی که ارتفاع آنها به اندازه فراوانی مطلق طبقه مربوطه باشد، مشخص کرده و آنها را به یکدیگر متصل می‌کنیم.

نمودارهای آماری

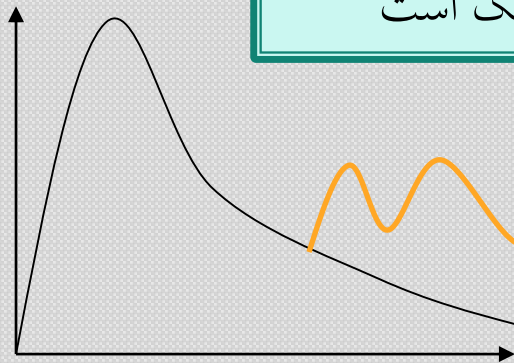
روش رسم نمودار چند ضلعی فراوانی تجمعی



برای رسم چند ضلعی فراوانی تجمعی داده‌ها نماینده طبقات را روی محور افقی تعیین می‌کنیم. بعد نقاطی که ارتفاع آنها به اندازه فراوانی تجمعی طبقه مربوطه باشد، مشخص کرده و آنها را به یکدیگر متصل می‌کنیم.

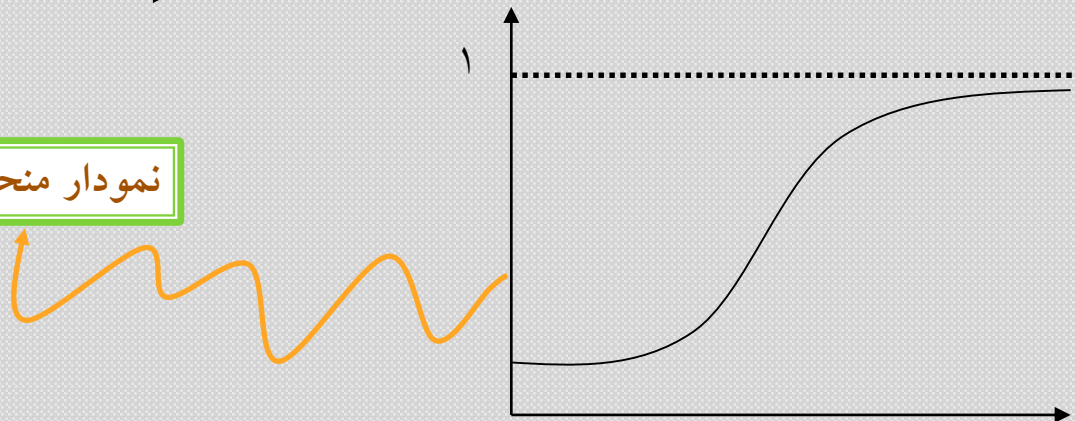
نمودارهای آماری

- همواره نمودار فراوانی داده ها، بالای محور یعنی در سمت مثبت است.
- میانه مساحت منحنی فراوانی را به دو قسمت مساوی تبدیل می کند.
- نمودار منحنی فراوانی تجمعی غیر نزولی و بین صفر تا یک است



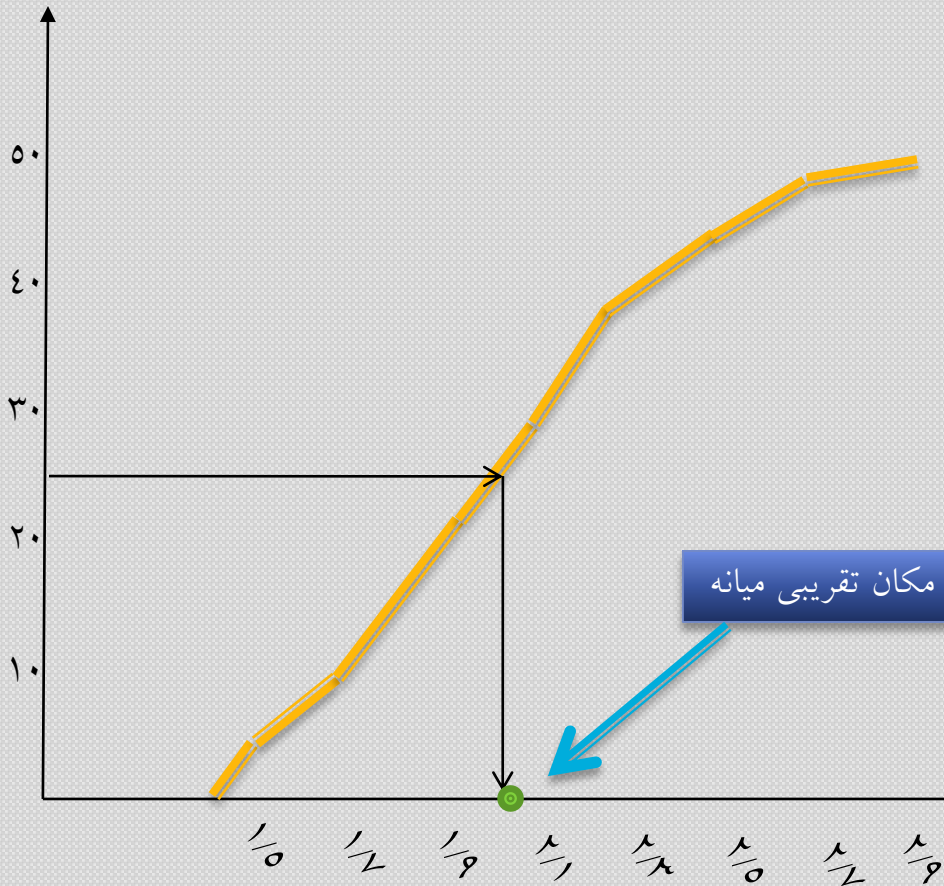
نمودار منحنی فراوانی

نمودار منحنی فراوانی تجمعی



نمودارهای آماری

تعیین چندک ها به کمک
نمودار فراوانی تجمعی



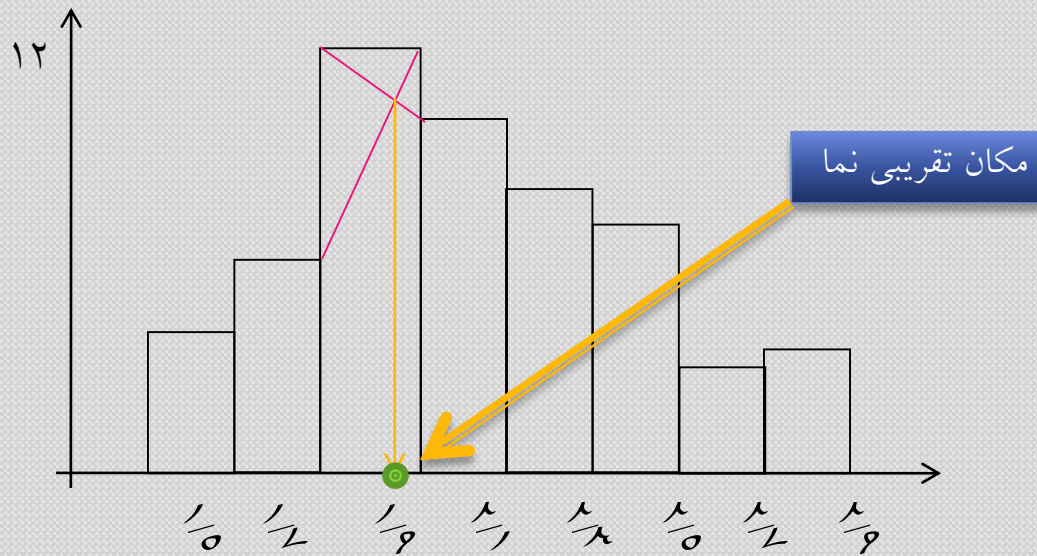
با ضرب نسبت مربوط به هر چندک در تعداد داده ها ($n \times p$) می توان مکان آن را رو نمودار فراوانی تجمعی مشخص کرد. به عنوان مثال میانه را روی نمودار از رابطه زیر به دست آورده ایم:

$$n \times p = 50 \times \frac{1}{2} = 25$$

نمودارهای آماری

تعیین نما روی نمودار
مستطیلی

برای تعیین نما روی نمودار مستطیلی،
ستون بلندتر را در نظر گرفته و مانند
شکل مکان تقریبی نما را معلوم می کنیم.



تمرین

تمرین

۱- تفاوت سه شاخص میانگین، میانه و نما را با مثال توضیح دهید

۲- تعداد توقفهای ۱۰ روز یک رایانه به صورت زیر ثبت شده است:

۴ ۳ ۰ ۲ ۳ ۳ ۱ ۳ ۰ ۱

الف - میانه و نمای این داده ها را بدست آورید
ب- ضریب تغییرات آنها را معین کنید

۳- تعداد دندانهای ترمیمی ۴۶ بیمار مراجعه کننده به یک مرکز دندانپزشکی در یک روز به شرح زیر است.

۲ ۵ ۴ ۳ ۱ ۵ ۰ ۲ ۲ ۱ ۴ ۳ ۳ ۵ ۳ ۲ ۱ ۱ ۴ ۳ ۰ ۳ ۴
۳ ۰ ۲ ۲ ۵ ۴ ۳ ۱ ۴ ۱ ۷ ۴ ۵ ۰ ۸ ۲ ۱ ۲ ۲ ۱ ۱

الف - میانه و نمای داده ها را محاسبه کنید
ب- ضریب تغییرات داده ها چند درصد است؟

تمرین

۴- جدول زیر سطح هموگلوبین در ۲۵ زن تحت مطالعه را نشان می دهد.

حدود طبقات	f_i
۷ - ۹	۴
۹ - ۱۱	۷
۱۱ - ۱۳	۱۱
۱۳ - ۱۵	۳

- الف- میانگین سطح هموگلوبین در بین زنان چقدر است؟ $\hat{\mu}$
- ب- ضریب چولگی داده ها را به دست آورده و تفسیر کنید. $\hat{\mu}_3$
- ج- چارک سوم داده ها را معلوم کنید. $\hat{\mu}_3$
- د- نمودار چند ضلعی فراوانی تجمعی را رسم کنید. $\hat{\mu}_3$

۵- توزیع سنی یک نمونه ۵۰ نفری از بیماران در جدول زیر آمده است.

حدود طبقات	فراوانی
۲۰ - ۴۰	۱۲
۴۰ - ۶۰	۳۰
۶۰ - ۸۰	۸

- الف- برای داده ها میانه و نما را محاسبه کنید. $\hat{\mu}$
- ب- ۹۵٪ سن بیماران در چه فاصله ای قرار دارد. $\hat{\mu}_3$
- ج- فراوانی های نسبی و تجمعی را معلوم کنید. $\hat{\mu}_3$

تمرین

- ۶- داده های مربوط به یک مطالعه را ۵ برابر کرده و با عدد ۳ جمع کرده ایم. میانگین جدید ۲۰/۵ و واریانس جدید ۱۰۰ است. میانگین و واریانس اولیه را به دست آورید. $\# \langle \hat{\mu} \rangle \tilde{\$}$
- ۷- اگر در یک مطالعه، میانگین و واریانس داده ها به ترتیب ۱۵ و ۱۶ باشند و سپس داده ها را با عدد ۳ جمع کرده و بر ۲ تقسیم کنیم؛ $\# \langle \hat{\mu} \rangle \tilde{\$}$
- الف- کدامیک از این شاخص ها تغییر نمی کنند؟ میانه - دامنه تغییرات - ضریب تغییرات - نما.
- ب- میانگین و واریانس جدید را به دست آورید. $\# \langle \hat{\mu} \rangle \tilde{\$}$
- ۸- کدامیک از شاخص های زیر به واحد اندازه گیری متغیرها وابسته نیست؟ $\# \langle \hat{\mu} \rangle \tilde{\$}$
- میانگین - واریانس - ضریب تغییرات - میانه - انحراف معیار
- ۹- یک بیمارستان شامل ۴ بخش است که میانگین و واریانس سن بیماران هر بخش در جدول زیر آمده است.

- الف- ضریب تغییرات کل را برای سن به دست آورید. $\# \langle \hat{\mu} \rangle \tilde{\$}$
- ب- سن ۹۵ درصد بیماران در چه حدودی قرار دارد؟ $\# \langle \hat{\mu} \rangle \tilde{\$}$

بخش	۱	۲	۳	۴
میانگین	۲۸	۳۳	۴۱	۳۸
واریانس	۱۲	۹	۲۰	۱۶

پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

**رشته های علوم پزشکی و شاخه های مرتبط
مطابق با سرفصل های مصوب**

احتمال

احتمال

- احتمال: اندازه ای است عددی که به یک پدیده تصادفی نسبت می دهیم.
- آزمایش تصادفی: هر آزمایشی که از قبل نتوان نتیجه آن را به طور قطع بیان کرد، یک آزمایش تصادفی است.

مثال:

- ❖ پرتاب یک تاس
- ❖ نتیجه یک عمل جراحی
- ❖ مدت زمانی که طول می کشد یک بیمار بهبود پیدا کند.
- ❖ تعداد تکرار آزمایش تا یافتن یک مورد بیماری
- ❖ تعیین گروه خون یک بیمار
- فضای نمونه: به تمامی حالات ممکن که در یک آزمایش تصادفی رخ می دهد، فضای نمونه می گویند. فضای نمونه را با S نشان می دهند.

احتمال

فضای نمونه مربوط به نتیجه یک عمل جراحی $S = \{ \text{بهبودی} , \text{عدم بهبودی} \}$

فضای نمونه پرتاب یک تاس (مکعب شش وجهی) $S = \{1, 2, 3, 4, 5, 6\}$

فضای نمونه آزمایش گروه خون یک فرد $S = \{A, O, B, AB\}$

فضای نمونه انتخاب عددی در فاصله صفر تا دو $S = \{ x \mid 0 \leq x \leq 2 \}$

توجه داشته باشید که هر فضای نمونه یک مجموعه است که ممکن است گسسته یا پیوسته باشند.

پیشامد

به هر زیر مجموعه فضای نمونه، یک **پیشامد** می گوئیم و پیشامدها را با حروف A و B و... نشان می دهیم

احتمال

احتمال

فرض کنید فضای نمونه S دارای n عضو است و عضوها از نظر رخ دادن همشانس هستند. اگر A پیشامدی از این فضای نمونه باشد، احتمال وقوع A را به صورت زیر تعریف می‌کنیم.

$$P(A) = \frac{\text{حالتهایی که امکان رخداد } A \text{ وجود دارد}}{\text{همه حالتهایی که ممکن است رخ دهد}} \quad P(A) = \frac{\text{تعداد عضوهای } A}{\text{تعداد عضوهای } S} \quad P(A) = \frac{n(A)}{n(S)}$$

مثال: پرتاب تاس، مثال واضحی از فضای نمونه‌ای است که عضوهای آن هم شانساند. یعنی احتمال وقوع هر طرف آن $\frac{1}{6}$ است و فضای نمونه آن به صورت زیر است:

$S = \{1, 2, 3, 4, 5, 6\}$

احتمال هر یک از پیشامدهای زیر را بدست می‌آوریم.

$$A: \text{پیشامد اینکه عددی فرد رو شود} \rightarrow A = \{1, 3, 5\} \rightarrow n(A) = 3 \rightarrow P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = \frac{1}{2}$$

$$B: \text{پیشامد رو شدن عددی کمتر از 5} \rightarrow B = \{1, 2, 3, 4\} \rightarrow n(B) = 4 \rightarrow P(B) = \frac{n(B)}{n(S)} = \frac{4}{6} = \frac{2}{3}$$

احتمال

قوانین احتمال

(۱) احتمال پیشامد حتمی یک (۱۰۰ درصد) است. $P(S) = ۱$

(۲) احتمال پیشامد غیر ممکن صفر است. $P(\phi) = ۰$

(۳) احتمال هر پیشامد مانند A ، غیر منفی است. $۰ \leq P(A) \leq ۱$

(۴) اگر A و B دو پیشامد هم ارز باشند، احتمال های آنها با هم مساوی است.

(۵) اگر A زیر مجموعه B باشد، احتمال A کوچکتر یا مساوی احتمال B است. $A \subset B \Rightarrow P(A) \leq P(B)$

(۶) اگر A' ، متمم پیشامد A باشد. $P(A') = ۱ - P(A)$, $P(A) = ۱ - P(A')$

(۷) اگر A و B دو پیشامد دلخواه باشند. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

قانون
جمع
احتمال

احتمال

مثال: ۶۰ درصد بیماران در یک بیمارستان بیمار قلبی، ۵۰ درصد بیمار کلیوی و ۳۰ درصد هر دو بیماری را دارند. اگر بیماری را به تصادف انتخاب کنیم، احتمال اینکه:
الف) بیمار قلبی یا کلیوی باشد چقدر است؟
ب) بیماری نوع دیگری داشته باشد؟

$$\begin{aligned} P(F \cup W) &= P(F) + P(W) - P(F \cap W) \\ &= ۰/۶ + ۰/۵ - ۰/۳ = ۰/۸ \end{aligned}$$

الف) اگر پیشامد F را انتخاب بیمار قلبی و W را پیشامد انتخاب بیمار کلیوی در نظر بگیریم؛ داریم:

$$\begin{aligned} P(F \cup W)' &= ۱ - P(F \cup W) \\ &= ۱ - ۰/۸ = ۰/۲ \end{aligned}$$

ب) پیشامد اینکه بیماری نوع دیگری داشته باشد، متمم این پیشامد است که یا بیمار کلیوی باشد یا بیمار قلبی، پس:

استقلال پیشامدها: دو پیشامد A و B را مستقل می‌گویند هرگاه احتمال رخ دادن پیشامد A به پیشامد B و بالعکس ربطی نداشته باشد. یعنی:

$$P(A \cap B) = P(A) \times P(B)$$

قانون
ضرب
احتمال

اگر $A_1, A_2, \dots, A_n$ پیشامدهای مستقلی از فضای نمونه S باشند به طوری که امکان رخ دادن همزمان آنها وجود داشته باشد، آنگاه احتمال اتفاق افتادن همزمان آنها از رابطه زیر که (قانون ضرب) به دست می‌آید.

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2 / A_1) \cdot P(A_3 / A_1, A_2) \dots P(A_n / A_1, A_2, \dots, A_{n-1})$$

مثال

جعبه‌ای شامل ۳۰ اسم که ۵ تای آن اسم دختر و بقیه پسراند. ۳ اسم پشت سرهم به تصادف از این جعبه برمی‌داریم و یکی بعد از دیگری قرائت می‌کنیم. احتمال این که دو اسم اول دختر و سومی اسم پسر باشد چقدر است؟

حل: اگر پیشامدهای S_1, S_2, S_3 را به ترتیب پیشامدهای انتخاب اسم های دختر در استخراج اول و دوم و اسم پسر در استخراج سوم در نظر بگیریم، داریم:

$$P(S_1 \cap S_2 \cap C_3) = P(S_1) \cdot P(S_2 / S_1) \cdot P(C_3 | S_1, S_2) = \frac{5}{30} \times \frac{4}{29} \times \frac{25}{28} = \frac{25}{1218}$$

توجه: همان گونه که می بینید، این گونه پیشامدها مستقل نیستند زیرا احتمال آنها به یکدیگر وابسته است. در این مثال در صورتی که اسامی برداشت شده در هر مرحله را به جای خود برگردانیم، آنوقت پیشامدها مستقل بودند.

احتمال

مثال: اطلاعات مربوط به ۱۰۰ بیمار بستری در یک بیمارستان قلب به صورت جدول زیر در اختیار است.
از بین اسامی این بیماران، یک نفر به تصادف انتخاب می‌کنیم؛
الف- احتمال اینکه سیگاری نباشد چقدر است؟ پاسخ: ۲۵٪
ب- اگر بدانیم کم سیگار مصرف می‌کند، احتمال اینکه عوارض قلبی شدید داشته باشد چقدر است؟ پاسخ: ۳۰٪
ج- چند درصد از بیماران غیر سیگاری دارای عوارض شدید اند؟ پاسخ: ۲۴٪

عوارض مشاهده شده قلبی					
مصرف سیگار	جمع	ضعیف	متوسط	شدید	
	۳۵	۱۸	۱۱	۶	زیاد
	۴۰	۱۰	۱۸	۱۲	کم
	۲۵	۱۴	۵	۶	عدم مصرف
	۱۰۰	۴۲	۳۴	۲۴	جمع

احتمال

احتمال تجربی

در مباحث آماری یکی از موضوعات مهمی که مطرح است دانستن توزیع احتمال متغیرهایی است که در مطالعه شرکت دارند. همواره این سوال وجود دارد که چگونه می توان یک الگوی احتمال برای یک آزمایش پیدا کرد؟ یا چگونه می توان قضاوت کرد که فلان مدل احتمال برای یک آزمایش الگویی برازنده است؟

فرض کنید A پیشامدی باشد که بخواهیم احتمال آن را در یک آزمایش تصادفی حساب کنیم. برای این کار آزمایش را n بار تکرار می کنیم. فرض کنید K تعداد دفعاتی باشد که پیشامد A رخ می دهد. در این صورت فراوانی نسبی پیشامد A در n بار آزمایش را **احتمال تجربی** پیشامد A می گوئیم. این احتمال با افزایش تکرار آزمایش پایدار می شود.

$$P(A) = \lim_{n \rightarrow \infty} \frac{k}{n}$$

فراوانی های نسبی در جدول توزیع فراوانی را می توان احتمال های تجربی انگاشت. البته با افزایش حجم نمونه این احتمالات دقیقتر می شوند.

احتمال

مثال: فرض کنید می خواهیم الگوی احتمال انتخاب ۳ کاندیدای ریاست جمهوری را در یک دوره انتخابات تعیین کنیم. بر این اساس در مورد سه کاندیدای A و B و C از ۱۰۰۰ نفر نظر خواهی کرده ایم. نتایج جمع آوری شده به صورت زیر است:

نام کاندیدا	C	B	A
تعداد رای	۲۱۰	۳۸۰	۴۱۰
فراوانی نسبی	۰/۲۱	۰/۳۸	۰/۴۱

احتمال تجربی انتخاب هر کاندیدا

بر اساس این الگوی احتمال اگر جمعیت واجد شرایط رای گیری در کشور ۱۸۷۰۰۰۰۰ نفر باشند، تقریباً چند نفر به کاندیدای B رای خواهند داد.

پاسخ ۷۱۰۶۰۰۰ نفر

مثال:

به منظور تعیین نسبت جنسیت در سال ۱۹۸۹ یک مطالعه در ایرلند نشان داد از ۵۱۶۵۹ تولد ۲۶۶۱۳ تولد پسر و بقیه دختر بودند. بنا براین احتمال تجربی تولد یک پسر در ایرلند ۰/۵۱۵ و احتمال تجربی تولد یک دختر ۰/۴۸۵ است.

جالب است بدانید بر اساس احتمال تجربی، نسبت جنسیت در تمام نژادها روی کره زمین، یکسان و در برابر هر ۱۰۰ دختر ۱۰۵ پسر متولد می شود!



احتمال

در محاسبه احتمالاتی که روی فضای نمونه گسسته تعریف می شوند، نیاز به دانستن بعضی قواعد شمارش داریم. قواعد متعددی برای شمارش وجود دارند که در زیر به بعضی از آنها که مورد استفاده بیشتری دارند، اشاره می کنیم.



تعداد حالت هایی که می توان از بین n شیء r شی را کنار یکدیگر قرار داد، ترتیب های r تایی از n شیء مختلف ($r \leq n$) می گوئیم و از رابطه زیر به دست می آید:

$$P_n^r = \frac{n!}{(n-r)!} \quad \begin{cases} n! = n \times (n-1) \times (n-2) \times \dots \times 2 \times 1 \\ 5! = 5 \times 4 \times 3 \times 2 \times 1 = 120 \end{cases}$$

مثال

از بین ۵ نفر به چند حالت می توان
الف- صف های ۳ تایی ساخت؟
ب- صف های ۵ تایی ساخت؟

انتخاب با ترتیب

$$P_5^3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = \frac{5 \times 4 \times 3 \times 2!}{2!} = 5 \times 4 \times 3 = 60 \quad \text{الف- صف های ۳ تایی}$$

$$5! = 120$$

ب- تعداد صف های ۵ تایی

احتمال

اصل ترکیب

در بسیاری از مباحث آماری موضوع انتخاب مطرح است. انتخاب r شیء از بین n شیء بدون در نظر گرفتن ترتیب آنها، **ترکیب** نام دارد. تعداد ترکیب های ممکن r شیء از بین n شیء را به صورت زیر می نویسیم:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

(در ترکیب، ترتیب قرار گرفتن اشیاء اهمیت ندارد)

انتخاب بدون ترتیب

مثال: به چند طریق می توان از بین ۷ دانشجو ۳ نفر از آنها را برای شرکت در یک کنفرانس انتخاب کرد.

$$\binom{7}{3} = \frac{7!}{3!(7-3)!} = \frac{7 \times 6 \times 5 \times 4!}{6} = 35$$

مثال:

سه زن و چهار مرد به تصادف در یک صف ایستاده اند.
احتمال هر یک از پیشامدهای زیر را حساب می کنیم.

$$P(B) = \frac{n(B)}{n(S)} = \frac{3! \times 4!}{7!} = \frac{1}{7}$$

B پیشامد اینکه سه زن کنار هم باشند؛

C پیشامد اینکه اول صف دو مرد، بعد سه زن و در انتها دو مرد باشند؛

$$P(C) = \frac{n(C)}{n(S)} = 2 \times \frac{2! \times 3! \times 2!}{7!} = \frac{1}{105}$$

مثال:

از بین ۱۰ نفر که ۴ تای آنها زن هستند، ۳ نفر انتخاب می‌کنیم. احتمال هر یک از پیشامدهای زیر را به دست آورید.

$$P(A) = \frac{n(A)}{n(S)} = \frac{\binom{6}{3}\binom{4}{0}}{\binom{10}{3}} = \frac{20}{120} = \frac{1}{6}$$

A پیشامد این که زن انتخاب نشود.

$$P(B) = \frac{n(B)}{n(S)} = \frac{\binom{4}{1}\binom{6}{2} + \binom{6}{3}}{\binom{10}{3}} = \frac{60 + 20}{120} = \frac{2}{3}$$

B پیشامد این که حداکثر یک زن انتخاب شود.



تمرین

تمرین

۱- چهار نفر داوطلب شغل معاونت اداری هستند. اگر احتمال انتخاب شدن فرد A دو برابر احتمال انتخاب شدن فرد B باشد و B و C تقریباً شانس یکسان داشته باشند و احتمال انتخاب شدن فرد C دو برابر فرد D باشد احتمالات زیر را حساب کنید.

الف- احتمال اینکه فرد A انتخاب شود چقدر است؟ $\# \hat{U} \#$

ب- احتمال اینکه D انتخاب نشود. $\# \hat{U} \#$

ج- احتمال اینکه فرد A یا B یا D انتخاب شوند چقدر است؟ $\# \hat{U} \#$

۲- بطور تصادفی از بین ۵ کودک واکسینه شده و ۷ کودک واکسینه نشده، ۳ کودک انتخاب می کنیم:

الف- در چند حالت بیشتر از یک کودک واکسینه انتخاب می شود؟ $\# \hat{U} \#$

ب- احتمال اینکه کودک واکسینه شده انتخاب نشود چقدر است؟ $\# \hat{U} \#$

۳- ۲۰ داده آماری با میانگین ۱۲ و انحراف معیار ۲ را با ۱۰ داده آماری دیگر با میانگین ۹ و انحراف معیار ۳ در نظر بگیرید.

الف) ضریب تغییرات کل داده ها را حساب کنید. $\# \hat{U} \#$

ب) یک فاصله تجربی ۹۵ درصدی برای هر دسته از داده ها و سپس برای کل داده ها، تعیین کنید. $\# \hat{U} \#$

تمرین

- ۴- بطور تصادفی از بین اسامی ۵ دانشجوی تهرانی و ۴ دانشجوی شهرستانی
- الف- ۳ اسم با جایگذاری بر می‌داریم. احتمال اینکه هر ۳ نفر تهرانی باشند چقدر است؟ $\$ \langle \tilde{U} \rangle \#$
- ب- ۳ اسم یکجا بر می‌داریم. احتمال اینکه دو تهرانی انتخاب شود چقدر است؟ $\$ \langle \tilde{U} \rangle \#$
- ج- ۳ اسم یکی پس از دیگری و بدون جایگذاری بر می‌داریم. احتمال اینکه دو نفر اول شهرستانی و نفر سوم تهرانی باشد، چقدر است؟ $\$ \langle \tilde{U} \rangle \#$

جنسیت	زن	مرد
سیگاری	۵	۲۵
غیر سیگاری	۱۵	۳۵
ترک کرده	۲	۱۸

- ۵- از یک جمعیت ۱۰۰ نفری اطلاعات زیر در دست است.
- یک نفر به تصادف از این جمعیت انتخاب می‌کنیم،
- الف - اگر بدانیم مرد است، احتمال اینکه ترک کرده باشد؟ $\$ \langle \tilde{U} \rangle \#$
- ب - احتمال اینکه یا زن باشد یا سیگاری نباشد؟ $\$ \langle \tilde{U} \rangle \#$
- ۶- دو تاس را با هم پرتاب می‌کنیم، احتمال اینکه حداقل یک ۶ بیاید چقدر است؟ $\$ \langle \tilde{U} \rangle \#$

- ۷- احتمال پسرزایی در یک خانواده ۶/۰ است. اگر این خانواده ۴ فرزند داشته باشند:
- الف) در صورتی که بدانیم دو فرزند اول دختر است، احتمال اینکه فرزند سوم پسر و چهارم دختر باشد؟ $\$ \langle \tilde{U} \rangle \#$
- ب) احتمال اینکه اولی پسر و سه فرزند بعدی دختر باشند چقدر است؟ $\$ \langle \tilde{U} \rangle \#$

تمرین

۸- در بررسی شیوع بیماری گال (بیماری پوستی با لکه های پررنگ) در بین ۱۰۰۰ سرباز در یک پادگان آموزش نظامی:

تعداد ۱۵ نفر از ناحیه دست؛ ۱۸ نفر از ناحیه پا و ۶ نفر هم از ناحیه دست و هم از ناحیه پا مبتلا هستند. چند درصد سربازان حداقل از یک ناحیه در گیر این بیماری اند؟ چه تعداد این بیماری را ندارند؟ احتمال اینکه یک سرباز فقط از ناحیه دست مبتلا باشد چقدر است؟ £ ¤ ¢

پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

**رشته‌های علوم پزشکی و شاخه‌های مرتبط
مطابق با سرفصل‌های مصوب**

توابع احتمال

توابع احتمال

می توان پیشامدهای یک آزمایش تصادفی را به جای اینکه با حالت و وضعیت شان نشان دهیم، با یک مقدار عددی نمایش دهیم. برای این کار روی فضای نمونه **متغیرهای تصادفی** را تعریف می کنیم و به کمک آنها بیان توصیفی پیشامدها را به مقادیر عددی تبدیل می نماییم.

مثال: جنسیت سه نوزاد هنگام تولد را یک آزمایش تصادفی در نظر بگیرید. فضای نمونه آن شامل ۸ عضو است و اگر متغیر تصادفی X را تعداد دخترها در سه تولد در نظر بگیریم، در این صورت مقایر X به ترتیب ۰ و ۱ و ۲ و ۳ خواهد بود. دلیل این امر را در جدول زیر مشاهده کنید.

X متغیر تصادفی	۰	۱	۲	۳
S فضای نمونه	AAA	AAB ABA BAA	ABB BAB BBA	BBB

B را دختر و A را پسر در نظر گرفته ایم.

توابع احتمال

تابع احتمال

به تابعی که بتواند به هر مقدار متغیر تصادفی، احتمال آن را نسبت دهد، یک تابع احتمال گسسته می‌گوییم. چنین تابعی را با $f(x)$ نشان می‌دهیم. ❖ از این جهت به آن تابع احتمال گسسته می‌گوییم که فضای نمونه آزمایش، یک فضای نمونه گسسته است.

یک تابع احتمال گسسته را می‌توان به صورت یک جدول احتمال یا به صورت یک ضابطه ریاضی بیان کرد.

X متغیر تصادفی	۰	۱	۲	۳
S فضای نمونه	AAA	AAB ABA BAA	ABB BAB BBA	BBB
مقادیر احتمال	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

جدول احتمال مقابل را می‌توان به صورت این فرمول نیز نوشت:

$$f(x) = \frac{\binom{3}{x}}{2^3} \quad x = 0, 1, 2, 3$$

توابع احتمال

ویژگی‌های تابع احتمال گسسته

هر تابع احتمال باید ویژگی‌های زیر را داشته باشد:

۱- برای تمامی مقادیر X ، مثبت باشد. $\forall x \quad f(x) = P(X = x) \geq 0$

۲- مجموع احتمالات برابر یک باشد. $\sum_x f(x) = 1$

مثال: بررسی کنید تابعی که با ضابطه زیر تعریف شده
یک تابع توزیع احتمال است؟

$$f(x) = P(X = x) = \frac{\binom{4}{x}}{16} \quad x = 0, 1, 2, 3, 4$$

$$\sum_{x=0}^4 f(x) = \sum_{x=0}^4 P(X = x) = f(0) + f(1) + f(2) + f(3) + f(4)$$

$$= \frac{\binom{4}{0}}{16} + \frac{\binom{4}{1}}{16} + \frac{\binom{4}{2}}{16} + \frac{\binom{4}{3}}{16} + \frac{\binom{4}{4}}{16}$$

$$= \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} + \frac{1}{16} = 1$$

باید مجموع احتمالات برای تمامی نقاط X برابر یک باشد.

توابع احتمال

میانگین

امید ریاضی یک توزیع را **میانگین توزیع** (میانگین جامعه) می‌گوییم و آن را با μ_x نشان می‌دهیم. امید ریاضی شاخصی برای میزان گرایش به مرکز یک توزیع است. همچنین امید ریاضی به مفهوم میزان مورد انتظار X نیز تعبیر شده و از رابطه زیر محاسبه می‌شود:

$$E(x) = \sum_x x \cdot P(X = x) = \sum_x x \cdot f(x)$$

مثال: برای متغیر تصادفی X با تابع احتمال زیر، امید ریاضی (میانگین) را حساب می‌کنیم.

X	۰	۱	۲	۴
$f(x)$	۰/۲	۰/۳	۰/۱	۰/۴
$x \cdot f(x)$	۰	۰/۳	۰/۲	۱/۶

$$E(x) = \sum_x x \cdot f(x) = ۰ + ۰/۳ + ۰/۲ + ۱/۶ = ۲/۳$$

توابع احتمال

واریانس

واریانس متغیر تصادفی X شاخصی برای تعیین میزان پراکندگی توزیع مقادیر X نسبت به میانگین است و آن را با σ_x^2 نشان داده و به این صورت تعریف می کنیم:

$$Var(x) = \sigma_x^2 = E(x - \mu_x)^2 = E(x^2) - E^2(x)$$

مثال:

در تابع احتمال مقابل، واریانس را محاسبه می کنیم.

X	۰	۱	۲	۴
$f(x)$	۰/۲	۰/۳	۰/۱	۰/۴
$x \cdot f(x)$	۰	۰/۳	۰/۲	۱/۶
$x^2 \cdot f(x)$	۰	۰/۳	۰/۴	۶/۴

$$Var(x) = E(x^2) - E^2(x) = 7/12 - (2/3)^2 = 2/9$$

$$E(x) = (0 \times 0/2) + (1 \times 0/3) + (2 \times 0/1) + (4 \times 0/4) = 2/3$$

$$E(x^2) = (0)^2 \times (0/2) + (1)^2 \times (0/3) + (2)^2 \times (0/1) + (4)^2 \times (0/4) = 7/12$$

توابع احتمال

- در میان توابع احتمال مختلف توابعی هستند که از اهمیت ویژه ای برخوردارند و کاربردهای بیشتری هم دارند مانند توابع توزیعی که در زیر به آنها اشاره می کنیم.

توزیع دو جمله ای

آزمایش هایی را در نظر بگیرید که فقط شامل دو حالت باشند. یک حالت آن را موفقیت و حالت دیگر را عدم موفقیت. مانند: سالم / معیوب - زن / مرد یا شیر / خط. فرض کنید احتمال موفقیت در هر آزمایش مقدار P باشد. اگر چنین آزمایش هایی را n بار به طور مستقل تکرار کنیم، و متغیر تصادفی X ، تعداد موفقیت باشد، X دارای توزیع احتمال دو جمله ای است و تابع احتمال آن به صورت زیر است:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, 2, \dots, n$$

$$E(X) = np$$

$$Var(X) = np(1-p)$$

میانگین و واریانس توزیع دو جمله ای عبارت اند از:

توابع احتمال

ویژگی‌های توزیع دو جمله ای

- ۱- دارای دو پارامتر n و p است.
- ۲- هر چه p و q به یکدیگر نزدیکتر باشند توزیع دو جمله ای متقارن‌تر است.
- ۳- با افزایش مقدار n ضریب چولگی به صفر میل می‌کند.
- ۴- توزیع دو جمله‌ای برای n های بزرگ مطلوب نیست.
اگر نسبت جامعه به نمونه $\frac{N}{n}$ حداقل ۱۰ باشد، می‌توان از نمونه‌گیری بدون جایگذاری استفاده کرد. (می‌شود فرض کرد پیشامدها مستقل هستند).
- ۵- اگر n بزرگ و p کوچک باشد (یا $10 < np < 1$) از تقریب پواسن استفاده می‌شود.
- ۶- بدون توجه به p ، اگر n بزرگ باشد به طوری که $np > 5$ باشد، از تقریب نرمال استاندارد استفاده می‌شود.

توابع احتمال

مثال

اگر متغیر تصادفی X را تعداد فرزندان پسر در یک خانواده ۴ فرزندی در نظر بگیریم، مطلوب است:

الف- تابع احتمال X را معلوم کنید.

ب- احتمال اینکه همه فرزندان دختر باشند چقدر است؟

ج- انتظار می‌رود این خانواده چند پسر داشته باشند؟

الف- متغیر تصادفی X ، دارای توزیع دوجمله ای به صورت مقابل است:

$$f(x) = \frac{\binom{4}{x}}{16} \quad x = 0, 1, 2, 3, 4$$

ب- احتمال اینکه همه فرزندان دختر باشند یعنی تعداد پسران صفر باشد.

$$P(x = 0) = f(0) = \frac{\binom{4}{0}}{16} = \frac{1}{16}$$

ج- برای محاسبه میانگین داریم:

$$E(x) = \sum_{x=0}^{x=4} x \cdot f(x) = (0 \times \frac{1}{16}) + (1 \times \frac{4}{16}) + (2 \times \frac{6}{16}) + (3 \times \frac{4}{16}) + (4 \times \frac{1}{16}) = 2 = np$$

توابع احتمال

مثال

تقریباً ۱۰٪ از درمانهای ریشه دندان با شکست مواجه می شود. اگر متغیر تصادفی X تعداد شکست در درمان ریشه ۴ دندان باشد،

الف- تابع احتمال X را معلوم کنید.

ب- احتمال اینکه حداقل یک درمان ریشه، موفقیت آمیز باشد چقدر است؟

ج- انتظار می رود در بین ۲۰ درمان ریشه دندان چند دندان با موفقیت درمان شود؟

الف- متغیر تصادفی X ، دارای توزیع دوجمله ای است:

$$f(x) = \binom{4}{x} \times (0.1)^x \times (0.9)^{4-x} \quad x = 0, 1, 2, 3, 4$$

$$P(x \geq 1) = 1 - P(x = 0) = 1 - f(0) = 1 - \binom{4}{0} (0.1)^0 (0.9)^4 = 1 - 0.67 = 0.33 \quad \text{ب-}$$

$$E(x) = np = 20 \times 0.1 = 2$$

ج- برای محاسبه میانگین داریم:

توابع احتمال

توزیع پواسن

آزمایش‌هایی را در نظر بگیرید که در یک فاصله زمانی یا ناحیه مکانی، متوسط تعداد موفقیت مشخص و برابر مقدار λ باشد. اگر متغیر تصادفی X ، تعداد موفقیت در یک فاصله زمانی که به تصادف انتخاب می‌کنیم، باشد، X دارای توزیع پواسن است و تابع احتمال آن به صورت زیر خواهد بود.

$$f(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

در یک چهارراه در هر ۲۴ ساعت به طور متوسط ۳ تصادف رخ می‌دهد. اگر تعداد تصادفات در یک روز که به طور تصادفی در نظر گرفته ایم، کمیت X باشد؛ توزیع احتمال X چیست؟ احتمال اینکه دقیقاً چهار تصادف رخ دهد را حساب کنید.

مثال

$$f(x) = \frac{e^{-3} 3^x}{x!} \quad x = 0, 1, 2, \dots$$

$$P(X = 4) = f(4) = \frac{e^{-3} 3^4}{4!}$$

متوسط تعداد تصادفات را $\lambda = 3$ داریم. پس توزیع احتمال متغیر تصادفی X ، توزیع پواسن است.

توابع احتمال

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

میانگین و واریانس توزیع پواسن با هم برابر و مساوی λ است.

تقریب احتمالات دو جمله ای

محاسبه احتمال‌های دو جمله‌ای وقتی تکرار آزمایش‌ها (n) زیاد می‌شود، تقریباً غیر ممکن است. این محاسبه وقتی مشکل‌تر می‌شود که احتمال موفقیت (p) به صفر یا یک نزدیک باشد. در چنین مواقعی می‌توانیم احتمالات توزیع دو جمله‌ای را با استفاده از توزیع پواسن و با تبدیل زیر محاسبه کنیم:

$$\lambda = np \rightarrow p = \frac{\lambda}{n}$$

توابع احتمال

مثال

می‌دانیم ۲٪ افراد یک جمعیت آلوده به ویروس HIV هستند. اگر ۲۰۰ نفر آنها را مورد آزمایش قرار دهیم، احتمال اینکه دو نفر مبتلا مشاهده شود چقدر است؟ میانگین و واریانس X را به دست آورید.

آزمایش فوق در شرایط توزیع دو جمله‌ای صدق می‌کند و به دلیل بالا بودن تکرار آزمایش می‌توان برای محاسبه احتمالات از توزیع پواسن استفاده کرد.

$$\lambda = n \times p = 200 \times 0.02 = 4$$

$$P(X = 2) = f(2) = \frac{e^{-\lambda} \lambda^2}{2!} = \frac{e^{-4} 4^2}{2!} = \frac{8}{e^4}$$

احتمال اینکه دو نفر مبتلا باشند،

$$E(X) = \text{Var}(X) = \lambda = 4$$

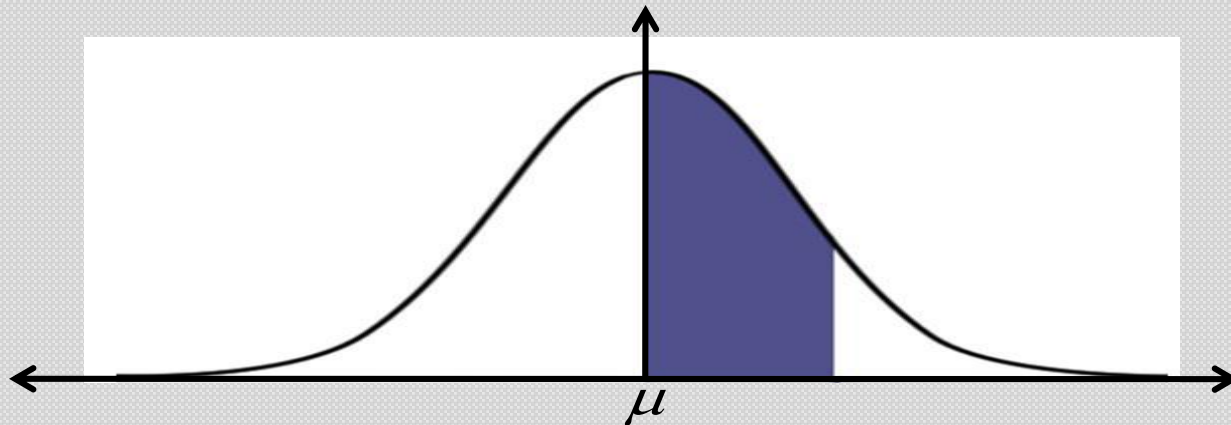
میانگین و واریانس:

توابع احتمال

توزیع نرمال

یکی از مهمترین توزیع ها در مباحث آماری است و کاربردهای بسیاری در تحقیقات پزشکی و مهندسی دارد. به تجربه ثابت شده است که در دنیای اطراف ما توزیع بسیاری از متغیرهای طبیعی از همین تابع پیروی می کنند. فرمول آن بر حسب، دو پارامتر میانگین و واریانس بیان می شود. ضابطه تابع توزیع نرمال به صورت زیر است:

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} \cdot e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} \quad -\infty < x < +\infty$$



توابع احتمال

ویژگی های توزیع نرمال

(۱) توزیعی است پیوسته که از منهای بینهایت تا مثبت بینهایت گسترده است.

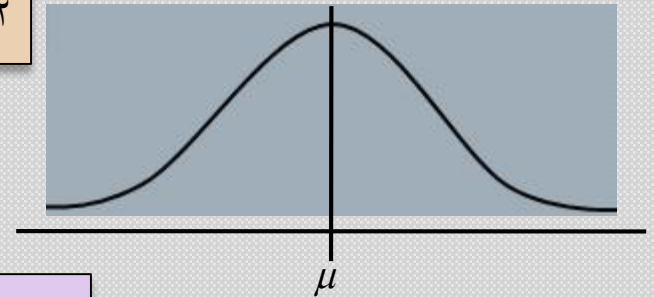
(۲) مساحت زیر منحنی نرمال برابر یک است. $P(-\infty < x < +\infty) = 1$

(۴) میانگین آن μ و واریانس آن σ^2 است.

(۳) توزیعی متقارن است در نتیجه ضریب چولگی آن صفر است. $Sk = 0$

(۵) سه شاخص مهم مرکزی یعنی میانگین، میانه و نما بر هم منطبقند. $\mu = Md = Mo$

(۶) محور تقارن آن μ میانگین است. به همین دلیل $P(x < \mu) = 0.5 = P(x > \mu)$



توابع احتمال

توزیع نرمال استاندارد

در یک توزیع نرمال اگر میانگین صفر و واریانس یک باشد توزیع را نرمال استاندارد شده می گویند. ضابطه تابع توزیع نرمال استاندارد به صورت زیر است:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad -\infty < x < +\infty$$

امتمالات توزیع نرمال استاندارد
به منظور استفاده کاربران در
جداولی تنظیم و ارائه شده است.

$$Z = \frac{x - \mu}{\sigma}$$

می توان احتمالات هر توزیع نرمال دیگری را
با استاندارد کردن و به کمک تغییر متغیر Z از
جداول نرمال استاندارد به دست آورد.

در توزیع نرمال استاندارد می توان از قواعد
زیر برای محاسبه احتمالات استفاده کرد:

۱) $P(z > a) = 1 - P(z < a)$

۲) $P(z > a) = P(z < -a)$

۳) $P(a < z < b) = P(z < b) - P(z < a)$

توابع احتمال

مثال

قد افراد در یک منطقه، دارای توزیع نرمال با میانگین ۱۷۰ و واریانس ۱۰۰ سانتیمتر است. اگر یک فرد از افراد این منطقه را به تصادف انتخاب کنیم، احتمال اینکه:

الف- قد او کمتر از ۱۶۴ سانتیمتر باشد؛

ب- در فاصله (۱۷۵ ۱۶۵) باشد، چقدر است؟

$$\begin{aligned}P(X < 164) &= P\left(\frac{X - \mu}{\sigma} < \frac{164 - 170}{10}\right) \\&= P(Z < -0.6) = 0.2743\end{aligned}$$

$$\begin{aligned}P(165 < X < 175) &= P\left(\frac{165 - 170}{10} < Z < \frac{175 - 170}{10}\right) \\&= P(-0.5 < Z < 0.5) \\&= P(Z < 0.5) - P(Z < -0.5) \\&= 0.6915 - 0.3085 = 0.3830\end{aligned}$$

برای محاسبه این احتمالات از جدول توزیع نرمال استاندارد استفاده کرده ایم.



توابع احتمال

تقریب احتمالات دوجمله ای با نرمال

دیدیم که در توزیع دوجمله ای هنگامی که تعداد تکرار آزمایش افزایش می یابد محاسبه احتمالات تقریباً غیر ممکن است. در چنین مواقعی استفاده از تقریب پواسن یک از روش مناسب برای محاسبه تقریبی احتمالات دوجمله ای است. یک روش دیگر، استفاده از توزیع نرمال استاندارد است مشروط به اینکه $np > 5$ باشد و نسبت موفقیت و شکست تقریباً برابر باشند.

$$Z = \frac{X - np}{\sqrt{npq}}$$

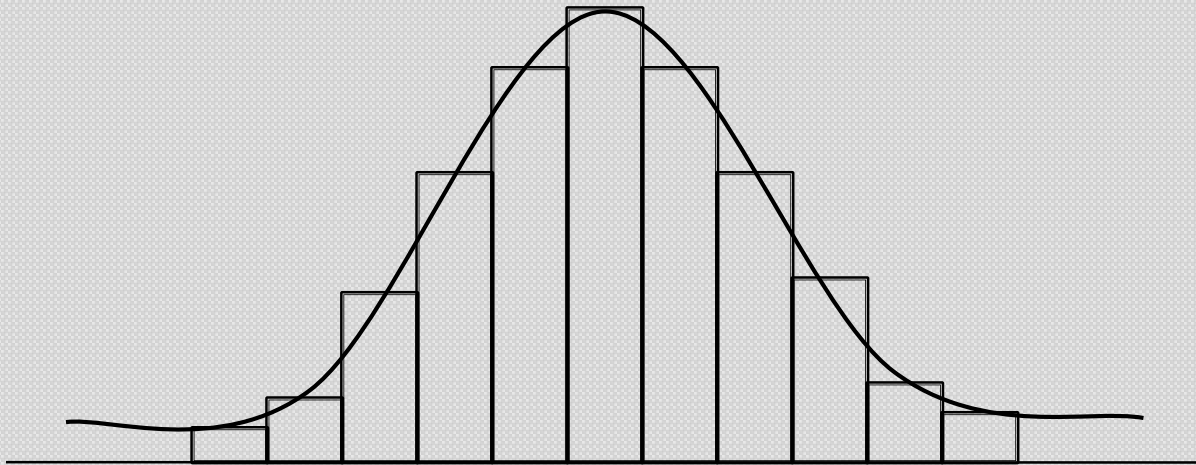
برای این کار از تبدیل زیر استفاده کنید.

که در آن $np = \mu_x$ میانگین و $\sqrt{npq} = \sigma_x$ انحراف معیار توزیع دوجمله ای است.

توابع احتمال

از آنجایی که توزیع نرمال یک توزیع پیوسته و توزیع دوجمله‌ای گسسته است، بهتر است بر اساس یک قانون تجربی به نام *تصحیح پیوستگی* احتمالات را دقیقتر به دست آورد.

به عنوان مثال اگر به جای فاصله $7 \leq X \leq 10$ ، فاصله $6.5 \leq X \leq 10.5$ را در نظر بگیریم، احتمال دوجمله‌ای با سطح زیر منحنی نرمال مربوطه بهتر تطبیق پیدا می‌کند.



توابع احتمال

مثال

۳۰ درصد پذیرفته شدگان دانشگاه، متاهل هستند. از یک نمونه شامل ۱۰۰۰ دانشجوی،
الف- احتمال اینکه متاهل‌ها کمتر از ۲۸۰ نفر باشند، چقدر است؟
ب- احتمال اینکه دقیقاً ۳۱۶ نفر باشند، چقدر است؟

با فرض این که X تعداد متاهل‌ها در نمونه ۱۰۰۰ تایی باشد، دراین صورت X دارای توزیع دو جمله‌ای با $n=1000$ و $p=0.3$ است.

$$\begin{cases} \sqrt{npq} = \sqrt{1000 \times 0.3 \times 0.7} = 14/5 \\ n.p = 300 \end{cases} \quad \text{میانگین و انحراف معیار توزیع به صورت زیر اند:}$$

$$P(X < 280/5) = P\left(\frac{X - \mu}{\sigma} < \frac{280/5 - 300}{14/5}\right) = P(Z < -1/35) = 0.0875 \quad \text{الف-}$$

$$P(X = 316) = P\left(\frac{315/5 - 300}{14/5} < Z < \frac{316/5 - 300}{14/5}\right) \quad \text{ب-}$$

$$= P(1/0.7 < Z < 1/14) = P(Z < 1/14) - P(Z < 1/0.7) = 0.8729 - 0.8577 = 0.0152$$



توابع احتمال

مثال

می‌دانیم ۱۰ درصد از بیماری‌های صرع به دارو مقاوم هستند.
چقدر احتمال دارد در بین ۲۰ بیمار که اخیراً دارو مصرف کرده‌اند؛
الف- کمتر از ۱۵ نفر درمان شوند؟
ب- دقیقاً ۱۷ نفر درمان شوند؟

آزمایش فوق در شرایط توزیع دو جمله‌ای صدق می‌کند و به دلیل بالا بودن تکرار آزمایش می‌توان برای محاسبه احتمالات از توزیع نرمال استفاده کرد.

$$E(x) = n \times p = 20 \times 0.5 = 10$$

الف-

$$\sigma_x^2 = n \times p \times q = 10 \times 0.5 \times 0.5 = 2.5 \longrightarrow \sigma_x = 1.58$$

$$P(x < 15) = P\left(\frac{x - \mu}{\sigma} < \frac{15 - 10}{1.58}\right) = P(z < 3.16) = 0.9992$$

ب- از تصحیح پیوستگی استفاده می‌کنیم؛ داریم:

$$P(x = 17) = P(16.5 < x < 17.5) = P\left(\frac{16.5 - 10}{1.58} < \frac{x - \mu}{\sigma} < \frac{17.5 - 10}{1.58}\right)$$

$$= P(-4.12 < z < 4.75) = P(z < 4.75) - P(z < -4.12) = 0.9999 - 0.0001 = 0.9998$$



تمرین

تمرین

$$f(x) = \frac{x+1}{k} \quad x = 0, 1, 2$$

۱- قانون احتمال یک متغیر تصادفی از تابعی به صورت:

پیروی می‌کند. مقدار k چقدر است؟ $\$ \rightarrow \#$

۲- یک عمل جراحی با احتمال $0/8$ موفقیت آمیز است. در روز آینده قرار است ۵ بیمار مورد عمل جراحی قرار گیرد.

اگر X تعداد کسانی باشد که بهبود پیدا می‌کنند؛ قانون احتمال X چیست؟ $\$ \rightarrow \#$
میانگین و واریانس آن را معلوم کنید.

۳- در یک جمعیت، نسبت پرفشاری خون در زنان بالای ۳۵ سال ۳۰ درصد است. چقدر احتمال دارد در بین ۶ زن انتخابی ۲ نفر مبتلا به پرفشاری خون باشد؟ $\$ \rightarrow \#$

۴- در هر ساعت به طور متوسط دو نفر به بخش اورژانس یک بیمارستان مراجعه می‌کنند. اگر X تعداد مراجعه کنندگان در دو ساعت آینده باشد؛

الف) تابع احتمال X چیست؟ $\$ \rightarrow \#$

ب) چقدر احتمال دارد در ۲ ساعت فقط ۲ نفر مراجعه کنند. $\$ \rightarrow \#$

ج) مقدار مورد انتظار و واریانس آن را به دست آورید. $\$ \rightarrow \#$

- ۵- نسبت پسرزایی در یک خانواده ۳۵ درصد است. اگر این خانواده ۴ فرزند داشته باشند، چقدر احتمال دارد همه فرزندان پسر باشند. #£ùò Ñ\$
- ۶- به طور متوسط در هر ۶ ساعت ۴ تصادف در یک بزرگراه رخ می‌دهد. چقدر احتمال دارد در سه ساعت حداقل یک تصادف رخ دهد. #£ùò Ñ\$
- ۷- احتمال ابتلا به HIV در یک جمعیت در سال ۱ در ۱۰۰۰ است. اگر در یک سال معین ۲۰۰۰ نفر را تست کنیم. احتمال اینکه یک نفر مبتلا باشد، چقدر است؟ #£ùò Ñ\$
- ۸- در یک جمعیت ۴ درصد افراد، سیگاری‌اند. اگر تصادفاً ۵۰ نفر را تست کنیم. احتمال اینکه در بین آنها بیشتر از یک نفر سیگاری باشد، چقدر است؟ #£ùò Ñ\$
- ۹- غلظت هموگلوبین خون مردان دارای توزیع نرمال با میانگین $14_{mg/dl}$ و واریانس $1_{mg/dl}$ است. در این صورت:

چند درصد از مردان در این جمعیت غلظت هموگلوبین خون آنها بین $(13 - 16)_{mg/dl}$ است؟ #£ùò Ñ\$

۱۰- ۱۰٪ از سربازانی که آموزش‌های سخت نظامی می‌بینند به افتادگی پا (Foot Drop) دچار می‌شوند. اگر از بین این افراد، ۳۰ نفر به تصادف انتخاب کنیم چقدر احتمال دارد همه سالم باشند؟ #£ùò Ñ\$

۱۱- در یک اپیدمی ۴۰ درصد افراد یک جمعیت به مالاریا مبتلا شدند. اگر ۳۰ نفر از این جامعه به

صورت تصادفی، انتخاب کنیم، احتمال‌های زیر را بدست آورید. $\$ \rightarrow \#$

الف) کمتر از ۹ نفر مبتلا باشد،

ب) بیشتر از ۱۲ نفر مبتلا باشد،

ج) تعداد مبتلایان بین ۱۰ تا ۱۵ نفر باشد،

د) دقیقاً ۱۰ نفر مبتلا باشند. $\$ \rightarrow \#$

۱۲- با فرض اینکه تعداد ضربان قلب در یک جمعیت دارای توزیع نرمال با میانگین $80 / \text{min}$ و واریانس

۱۰۰ است. تعداد ضربان چند درصد افراد بین ۸۰ تا ۹۰ بار در دقیقه است. $\$ \rightarrow \#$

۱۳- ۵ درصد از افراد در یک جمعیت تیزهوش هستند. اگر ۶۰ نفر از این جمعیت را به تصادف انتخاب

کنیم و X تعداد کسانی باشد که تیزهوش‌اند،

الف - قانون احتمال X چیست؟ $\$ \rightarrow \#$

ب - احتمال اینکه هیچ فردی در این جمعیت تیزهوش نباشد چقدر است؟ $\$ \rightarrow \#$

ج - میانگین و واریانس X را به دست آورید. $\$ \rightarrow \#$

جدولها



جدول نرمال استاندارد

نحوه استفاده از جدول

هر بخش از جدول مقابل شامل ۳ ستون است. در ستون اول مقدار Z و در ستون دوم احتمال های کمتر از مقدار Z را مشاهده می کنید. ستون سوم اختصاص دارد به مقادیر بیشتر از Z یا کمتر از $-Z$ به عنوان مثال:

$$P(Z < 0.27) = 0.6064$$

$$P(Z > 0.27) = 0.3936 \text{ و در ستون بعدی}$$

که معادل $P(Z < -0.27)$ است.

(ادامه جدول در صفحه بعد)

Z	Cum p	Tail p	Z	Cum p	Tail p
0.00	0.5000	0.5000	0.40	0.6554	0.3446
0.01	0.5040	0.4960	0.41	0.6591	0.3409
0.02	0.5080	0.4920	0.42	0.6628	0.3372
0.03	0.5120	0.4880	0.43	0.6664	0.3336
0.04	0.5160	0.4840	0.44	0.6700	0.3300
0.05	0.5199	0.4801	0.45	0.6736	0.3264
0.06	0.5239	0.4761	0.46	0.6772	0.3228
0.07	0.5279	0.4721	0.47	0.6808	0.3192
0.08	0.5319	0.4681	0.48	0.6844	0.3156
0.09	0.5359	0.4641	0.49	0.6879	0.3121
0.10	0.5398	0.4602	0.50	0.6915	0.3085
0.11	0.5438	0.4562	0.51	0.6950	0.3050
0.12	0.5478	0.4522	0.52	0.6985	0.3015
0.13	0.5517	0.4483	0.53	0.7019	0.2981
0.14	0.5557	0.4443	0.54	0.7054	0.2946
0.15	0.5596	0.4404	0.55	0.7088	0.2912
0.16	0.5636	0.4364	0.56	0.7123	0.2877
0.17	0.5675	0.4325	0.57	0.7157	0.2843
0.18	0.5714	0.4286	0.58	0.7190	0.2810
0.19	0.5753	0.4247	0.59	0.7224	0.2776
0.20	0.5793	0.4207	0.60	0.7257	0.2743
0.21	0.5832	0.4168	0.61	0.7291	0.2709
0.22	0.5871	0.4129	0.62	0.7324	0.2676
0.23	0.5910	0.4090	0.63	0.7357	0.2643
0.24	0.5948	0.4052	0.64	0.7389	0.2611
0.25	0.5987	0.4013	0.65	0.7422	0.2578
0.26	0.6026	0.3974	0.66	0.7454	0.2546
0.27	0.6064	0.3936	0.67	0.7486	0.2514
0.28	0.6103	0.3897	0.68	0.7517	0.2483
0.29	0.6141	0.3859	0.69	0.7549	0.2451
0.30	0.6179	0.3821	0.70	0.7580	0.2420
0.31	0.6217	0.3783	0.71	0.7611	0.2389
0.32	0.6255	0.3745	0.72	0.7642	0.2358
0.33	0.6293	0.3707	0.73	0.7673	0.2327
0.34	0.6331	0.3669	0.74	0.7704	0.2296
0.35	0.6368	0.3632	0.75	0.7734	0.2266
0.36	0.6406	0.3594	0.76	0.7764	0.2236
0.37	0.6443	0.3557	0.77	0.7794	0.2206
0.38	0.6480	0.3520	0.78	0.7823	0.2177
0.39	0.6517	0.3483	0.79	0.7852	0.2148



جدول نرمال استاندارد

Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p
0.80	0.7881	0.2119	1.20	0.8849	0.1151	1.60	0.9452	0.0548
0.81	0.7910	0.2090	1.21	0.8869	0.1131	1.61	0.9463	0.0537
0.82	0.7939	0.2061	1.22	0.8888	0.1112	1.62	0.9474	0.0526
0.83	0.7967	0.2033	1.23	0.8907	0.1093	1.63	0.9484	0.0516
0.84	0.7995	0.2005	1.24	0.8925	0.1075	1.64	0.9495	0.0505
0.85	0.8023	0.1977	1.25	0.8944	0.1056	1.65	0.9505	0.0495
0.86	0.8051	0.1949	1.26	0.8962	0.1038	1.66	0.9515	0.0485
0.87	0.8078	0.1922	1.27	0.8980	0.1020	1.67	0.9525	0.0475
0.88	0.8106	0.1894	1.28	0.8997	0.1003	1.68	0.9535	0.0465
0.89	0.8133	0.1867	1.29	0.9015	0.0985	1.69	0.9545	0.0455
0.90	0.8159	0.1841	1.30	0.9032	0.0968	1.70	0.9554	0.0446
0.91	0.8186	0.1814	1.31	0.9049	0.0951	1.71	0.9564	0.0436
0.92	0.8212	0.1788	1.32	0.9066	0.0934	1.72	0.9573	0.0427
0.93	0.8238	0.1762	1.33	0.9082	0.0918	1.73	0.9582	0.0418
0.94	0.8264	0.1736	1.34	0.9099	0.0901	1.74	0.9591	0.0409
0.95	0.8289	0.1711	1.35	0.9115	0.0885	1.75	0.9599	0.0401
0.96	0.8315	0.1685	1.36	0.9131	0.0869	1.76	0.9608	0.0392
0.97	0.8340	0.1660	1.37	0.9147	0.0853	1.77	0.9616	0.0384
0.98	0.8365	0.1635	1.38	0.9162	0.0838	1.78	0.9625	0.0375
0.99	0.8389	0.1611	1.39	0.9177	0.0823	1.79	0.9633	0.0367
1.00	0.8413	0.1587	1.40	0.9192	0.0808	1.80	0.9641	0.0359
1.01	0.8438	0.1562	1.41	0.9207	0.0793	1.81	0.9649	0.0351
1.02	0.8461	0.1539	1.42	0.9222	0.0778	1.82	0.9656	0.0344
1.03	0.8485	0.1515	1.43	0.9236	0.0764	1.83	0.9664	0.0336
1.04	0.8508	0.1492	1.44	0.9251	0.0749	1.84	0.9671	0.0329
1.05	0.8531	0.1469	1.45	0.9265	0.0735	1.85	0.9678	0.0322
1.06	0.8554	0.1446	1.46	0.9279	0.0721	1.86	0.9686	0.0314
1.07	0.8577	0.1423	1.47	0.9292	0.0708	1.87	0.9693	0.0307
1.08	0.8599	0.1401	1.48	0.9306	0.0694	1.88	0.9699	0.0301
1.09	0.8621	0.1379	1.49	0.9319	0.0681	1.89	0.9706	0.0294
1.10	0.8643	0.1357	1.50	0.9332	0.0668	1.90	0.9713	0.0287
1.11	0.8665	0.1335	1.51	0.9345	0.0655	1.91	0.9719	0.0281
1.12	0.8686	0.1314	1.52	0.9357	0.0643	1.92	0.9726	0.0274
1.13	0.8708	0.1292	1.53	0.9370	0.0630	1.93	0.9732	0.0268
1.14	0.8729	0.1271	1.54	0.9382	0.0618	1.94	0.9738	0.0262
1.15	0.8749	0.1251	1.55	0.9394	0.0606	1.95	0.9744	0.0256
1.16	0.8770	0.1230	1.56	0.9406	0.0594	1.96	0.9750	0.0250
1.17	0.8790	0.1210	1.57	0.9418	0.0582	1.97	0.9756	0.0244
1.18	0.8810	0.1190	1.58	0.9429	0.0571	1.98	0.9761	0.0239
1.19	0.8830	0.1170	1.59	0.9441	0.0559	1.99	0.9767	0.0233



پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

**رشته‌های علوم پزشکی و شاخه‌های مرتبط
مطابق با سرفصل‌های مصوب**

بیر آورد آمارى

آمار استنباطی

● روشهای آماری برای تعمیم نتایج نمونه به جامعه:

- ۱- روش برآورد کردن (Estimations):
- ۲- روش آزمون فرض (Hypothesis test)

روشهای تعمیم نتایج نمونه به جامعه

آزمون

برآورد

ناپارامتری

پارامتری

فاصله ای

نقطه ای

برآورد و حدود اطمینان

یکی از مباحث آمار استنباطی موضوع برآورد است. همیشه موضوع برآورد یا تخمین با عدم قطعیت همراه است. به منظور برآورد یک پارامتر، از برآوردگر استفاده می‌کنند. یک برآوردگر باید دارای دو ویژگی باشد:

نااریب باشد.

حداقل پراکندگی را داشته باشد. (واریانس آن حداقل باشد)

کاربرد: این روش بیشتر در مطالعات توصیفی (مقطعی) که قرار است نتایج توصیفی آن مطالعه، به کل جامعه تعمیم داده شود، مورد استفاده قرار می‌گیرد.

هدف: برآورد یا تخمین چیزی (پارامتر) است که در جامعه به طور ثابت وجود دارد و محقق از آن خبر دقیق ندارد.

برآورد کردن

برآورد کردن به دو روش صورت می گیرد:

برآورد فاصله ای

برآورد نقطه ای

- ۱- برآورد نقطه ای: وقتی است که با استفاده از یک نمونه و توسط یک برآوردگر به نام **آماره** مقداری برای یک پارامتر برآورد می کنیم.
آماره: کمیت هایی هستند که از یک نمونه و بر اساس یک ضابطه به دست می آیند و برآوردی برای پارامتر جامعه محسوب می شوند.
مثال: مقدار فشار خون در یک جمعیت پارامتر مجهولی است (μ) که با داشتن یک نمونه توسط آماره ($\bar{X}$) مقداری برای آن برآورد می کنیم.

برآورد نقطه ای

برآورد (نمونه)

$\bar{x}$	میانگین نمونه
s^2	- واریانس نمونه
$\bar{p}$	- نسبت نمونه
n	- برآورد حجم نمونه
or	- برآورد نسبت خطر
$\hat{S}_t$	- برآورد تابع بقا

پارامتر (جامعه)

μ	- میانگین
σ^2	- واریانس
p	- نسبت
n	- حجم نمونه
OR	- نسبت خطر
S_t	- تابع بقا

برآوردهای فوق برآوردهای نااریب و دارای کمترین واریانس هستند.

برآورد نقطه ای

مثال: جامعه‌ای فرضی دارای ۱۰ عضو است که متغیر اندازه گیری شده در این جمعیت دارای مقادیر: ۰ ۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۹ است. در این جمعیت پارامترهای میانگین $\mu = 4/5$ و واریانس $\sigma^2 = 9/17$ می باشد. همچنین نسبت اعداد زوج $p = 0/5$ است.

حال اگر یک نمونه ۳ تایی به تصادف انتخاب کنیم و اعداد ۲ ۷ ۹ انتخاب شده باشند، برآوردهای میانگین، واریانس و نسبت اعداد زوج به ترتیب برابر اند با:

$$\bar{p} = \frac{1}{3}, \quad S^2 = 8/75, \quad \bar{x} = 6$$

برآورد

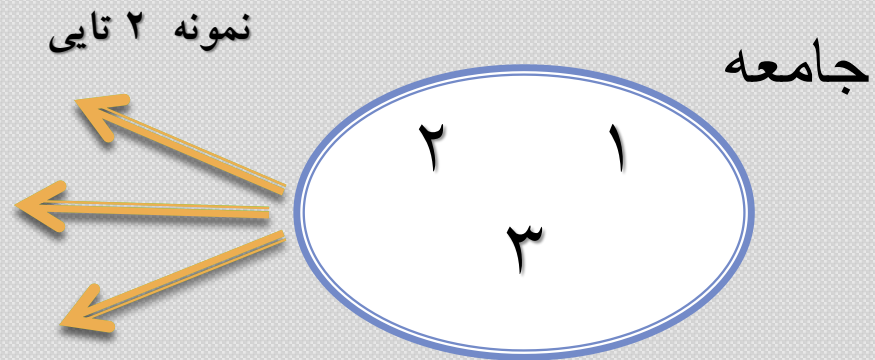
هر آماره به خاطر نمونه‌گیری مقداری "عدم اعتماد" در خود دارد. مفهوم عدم اعتماد (Uncertainty): اگر از جامعه‌ای به دفعات نمونه‌هایی به حجم مساوی انتخاب و برای هر نمونه آماره‌ها را به دست آوریم، لزوماً مساوی نخواهند بود. به این عدم اعتماد یا واریانس ناشی از نمونه‌گیری، خطای معیار آماره یا به اختصار خطای معیار (SE) می‌گویند.

SD (انحراف معیار) و SE (خطای معیار) با یکدیگر متفاوتند به طوری که انحراف معیار پراکندگی متغیر و خطای معیار پراکندگی برآوردگر را نشان می‌دهد.

برآورد

مثال: جامعه‌ای را در نظر بگیرید که فقط سه عضو دارد. اگر تمامی نمونه‌های دوتایی ممکن (با جایگذاری) از این جامعه را در نظر بگیریم برآوردهای میانگین به صورت زیراند:

نمونه‌ها	آماره $\bar{x}$
۱ و ۱	۱
۲ و ۱	۱/۵
۳ و ۱	۲
۲ و ۲	۲
۳ و ۲	۲/۵
۳ و ۳	۳



در این مثال احتمال اینکه مقدار برآورد شده با پارامتر مطابقت داشته باشد $3 \square 1$ است که احتمال کمی است. پس باید هنگام تعمیم آماره به جامعه عدم اعتماد موجود در نمونه را در محاسبات منظور کرده و احتمال مطابقت را بالا برد.

برآورد فاصله ای

دیدیم که هیچ تضمینی وجود ندارد که یک مقدار برآورد شده (برآورد نقطه‌ای) دقیقا با پارامتر جامعه تطابق داشته باشد. هرچند با افزایش حجم نمونه این مشکل تا حدودی مرتفع می‌شود ولی روش مناسب‌تر ایجاد یک فاصله به ترتیب زیر است:

۲- برآورد فاصله‌ای: یا فاصله اطمینان برای یک پارامتر، حدودی است که با استفاده از آماره مورد نظر و با سطح اطمینان بالایی پارامتر جامعه را در بر می‌گیرد.

● قاعده کلی پیدا کردن حدود اطمینان برای یک پارامتر به صورت زیر است:

$$\boxed{\text{آماره}} \pm \boxed{\text{ضریب اطمینان در تعمیم}} \times \boxed{\text{SE (آماره)}}$$

پارامتر

برآورد فاصله ای

ایجاد یک فاصله اطمینان یا برآورد فاصله ای برای یک پارامتر به عوامل زیر وابسته است:

۱- سطح اطمینانی که محقق را مجاب کند که فاصله ایجاد شده، پارامتر را در بر می گیرد.

۲- توزیع نمونه ای آماره که به کمک آن ضریب اعتماد تعیین می شود.

۳- واریانس آماره یا همان خطای معیار نمونه گیری (SE)

۴- حجم نمونه ای که برای برآورد انتخاب می شود. (n)

در مطالعات پزشکی توزیع بیشتر اندازه های بیولوژیک مثل هموگلوبین، کلسترول، قند خون، اندازه قلب، وزن، قد و مانند اینها دارای توزیع نرمال هستند و حد نرمال آنها را $\mu \pm 2\sigma$ در نظر می گیرند.

فاصله اطمینان برای میانگین (μ)

از آنجا که برآوردگر $\bar{x}$ یک برآوردگر مناسب برای میانگین جامعه است می توان برای پارامتر میانگین (μ) فاصله اطمینان را به صورت زیر تعیین کرد:
چون آماره $\bar{x}$ پارامتر μ را با خطای معیاری به یکی از دو حالت زیر برآورد می کند.

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

بنابراین فاصله اطمینان برای میانگین جامعه به صورت زیر خواهد بود:

$$\mu : \bar{x} \pm (\text{ضریب تعمیم}) \times \frac{\sigma}{\sqrt{n}} \quad \text{در صورتی که واریانس معلوم باشد.}$$

$$\mu : \bar{x} \pm (\text{ضریب تعمیم}) \times \frac{s}{\sqrt{n}} \quad \text{در صورتی که واریانس معلوم نباشد.}$$

فاصله اطمینان میانگین

● قضیه حد مرکزی:

در یک نمونه n تایی از جامعه ای با میانگین μ_x و واریانس σ_x^2 اگر تعداد نمونه به حد کافی بزرگ باشد، توزیع آماره $\bar{X}$ نرمال است و میانگین و واریانس آن به صورت زیر معرفی می شوند:

$$\begin{cases} \mu_{\bar{X}} = \mu_x \\ \sigma_{\bar{X}}^2 = \frac{\sigma_x^2}{n} \end{cases}$$

به عبارتی - توزیع جامعه هرچه باشد - وقتی یک نمونه n تایی از آن جامعه انتخاب می کنیم و میانگین آن ($\bar{X}$) را به دست می آوریم در صورتی که n به حد کافی بزرگ باشد، توزیع $\bar{X}$ نرمال خواهد بود. ♦ در آمار حد کافی وقتی است که تعداد نمونه ها بیشتر از ۲۵ باشد.

فاصله اطمینان میانگین

در فاصله اطمینان برای میانگین منظور از S یا σ انحراف معیار صفت اندازه گیری شده است و ضریب تعمیم نیز از جداول توزیع نرمال (Z) یا توزیع استودنت (t) به دست می آید.

به عنوان مثال:

● میزان اطمینان ۹۵٪	توزیع نرمال	ضریب تعمیم ۱/۹۶
● میزان اطمینان ۹۹٪	توزیع استودنت	ضریب تعمیم ۲/۵۸

مثال: مطالعه‌ای به منظور تعیین طول مدت بستری بیماران مبتلا به زخم اثنی عشر با همکاری چند بیمارستان صورت گرفت. در یک نمونه تصادفی ۶۴ تایی از این بیماران میانگین مدت بستری ۸/۲۵ روز با انحراف معیار ۵ به دست آمد. فاصله‌ای که با اطمینان ۹۵٪ میانگین واقعی را در بر داشته باشد، تعیین کنید.

فاصله اطمینان میانگین

حل:

$$\mu: \bar{X} \pm 1/96 \times \frac{s}{\sqrt{n}} = 8/25 \pm (1/96 \times \frac{5}{\sqrt{64}})$$

که حد پایین آن ۷/۰۲۵ و حد بالای آن ۹/۴۷۵ به دست می آید.

تفسیر: می توان گفت با اطمینان ۹۵٪ میانگین واقعی طول مدت بستری این بیماران در این فاصله قرار دارد.

$$P(7/025 < \mu < 9/475) = 0/95$$

فاصله اطمینان نسبت

● اگر p نسبت مجهولی در جامعه باشد $\bar{p} = \frac{x}{n}$ یک آماره برای برآورد کردن آن است که در آن x تعداد پیامد مشاهده شده و n تعداد نمونه آزمایش شده است. خطای معیار این آماره به صورت زیر است:

$$SE(\bar{p}) = \sqrt{\frac{\frac{x}{n}(1 - \frac{x}{n})}{n}}$$

● بنابراین فاصله اطمینان برای نسبت به صورت زیر است:

$$P: \bar{p} \pm z_{\frac{1-\alpha}{2}} \times \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}}$$

فاصله اطمینان نسبت

مثال:

به منظور برآورد نسبت دانش آموزان مبتلا به عیب دید چشم، یک نمونه ۱۴۰ تایی از دانش آموزان را انتخاب کرده و مشاهده می شود ۲۰ نفر از آنان دچار عیب دید بینایی هستند. اولاً حدودی تعیین کنید که با اطمینان ۹۵٪ نسبت واقعی مبتلایان را در بر گیرد. ثانياً با احتمال ۹۹٪ این نسبت را در بر گیرد.

$$P : \bar{p} \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \frac{20}{140} \pm (1.96 \times \sqrt{\frac{\frac{20}{140} \times \frac{120}{140}}{140}})$$

فاصله اطمینان نسبت

● فاصله ۹۹٪ نیز به صورت زیر خواهد بود.

$$P : \bar{p} \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \frac{20}{140} \pm (2 / 58 \times \sqrt{\frac{\frac{20}{140} \times \frac{120}{140}}{140}})$$

مثال: عنوان یک طرح تحقیقاتی به ای صورت است:
”بررسی میزان شیوع سیگار کشیدن و عوامل مرتبط با آن در...“
نتایج این بررسی در جدول زیر آمده است:

فاصله اطمینان نسبت

وضعیت سیگار کشیدن	تعداد	درصد
سیگاری	۱۳۶	۱۲,۲٪
غیر سیگاری	۹۷۶	۸۷,۸٪
جمع	۱۱۱۲	۱۰۰٪

● حدود اطمینان ۹۵٪ برای میزان شیوع کشیدن سیگار در این جمعیت برابر است با:

$$P : \bar{p} \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} = \frac{136}{1112} \pm (1.96 \times \sqrt{\frac{\frac{136}{1112} \times \frac{976}{1112}}{1112}})$$

$$(10/2 < P < 14/2)\%$$

فاصله اطمینان اختلاف بین دو نسبت

● فرض کنید می‌خواهیم نسبت مجهولی را در دو جمعیت با یکدیگر مقایسه کنیم و حدود اطمینان را برای اختلاف دو نسبت به دست آوریم.

جامعه ۲

نسبت P_2

$$\bar{p}_2 = \frac{x_2}{n_2} \text{ آماره}$$

جامعه ۱

نسبت P_1

$$\bar{p}_1 = \frac{x_1}{n_1} \text{ آماره}$$

فاصله اطمینان اختلاف بین دو نسبت

در این صورت فرمولی که طبق آن می توان فاصله اطمینان برای اختلاف بین این دو نسبت را به دست آورد، به صورت زیر است:

$$P_1 - P_2 : (\bar{p}_1 - \bar{p}_2) \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}}$$

مثال: ۲۰۰ نفر از بیمارانی را که از یک بیماری خاص رنج می برند به صورت تصادفی به دو گروه تقسیم کرده اند. از گروه اول که با روش معمول درمان شده اند ۷۸ نفر و از گروه دوم که با روش جدید درمان شده اند ۹۰ نفر در مدت ۳ روز بهبود یافته اند. پزشک معالج خواستار برآورد اختلاف بین دو نسبت فوق است.

فاصله اطمینان اختلاف بین دو نسبت

$$\bar{p}_1 = \frac{78}{100} \quad \text{نسبت بهبودی در روش معمول}$$

$$\bar{p}_2 = \frac{90}{100} \quad \text{نسبت بهبودی در روش جدید}$$

$$P_1 - P_2 : (\bar{p}_1 - \bar{p}_2) \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{\bar{p}_1 \bar{q}_1}{n_1} + \frac{\bar{p}_2 \bar{q}_2}{n_2}}$$

$$\left(\frac{78}{100} - \frac{90}{100}\right) \pm 1.96 \times \sqrt{\frac{(0.78)(0.22)}{100} + \frac{(0.9)(0.1)}{100}}$$
$$(-0.22 < P_1 - P_2 < -0.02)$$



تمرین

۱- به منظور برآورد یک صفت از یک برآوردگر استفاده می‌کنیم. یک برآوردگر باید دارای چه ویژگی‌هایی باشد؟ $\# \langle \hat{\theta} \rangle \tilde{\$}$

۲- تفاوت پارامتر و آماره چیست؟ $\# \langle \hat{\theta} \rangle \tilde{\$}$

۳- منظور از خطای معیار برآورد چیست؟ $\# \langle \hat{\theta} \rangle \tilde{\$}$

۴- در تعیین حدود اطمینان برای یک پارامتر چه عواملی موثراند؟ $\# \langle \hat{\theta} \rangle \tilde{\$}$

۵- مقادیر یک متغیر در یک جمعیت ۵ عضوی به ترتیب: ۳ ۵ ۷ ۹ ۱ است. یک نمونه ۳ تایی به تصادف از این جمعیت انتخاب کرده‌ایم که مقادیر ۳ ۷ بدست آمده است.

الف- برآوردهای میانگین و واریانس را به دست آورید. $\# \langle \hat{\theta} \rangle \tilde{\$}$

ب- فاصله‌ای بسازید که با اطمینان ۶۸ درصد میانگین واقعی را در بر داشته باشد. $\# \langle \hat{\theta} \rangle \tilde{\$}$

۶- در یک سال از ۶۰۰ بیمار سرطانی ۱۲۰ نفر آنها فوت شده‌اند (نسبت کشندگی ۲۰٪ است).

فاصله اطمینان ۹۵٪ برای نسبت کشندگی واقعی بیماری سرطان بسازید. $\# \langle \hat{\theta} \rangle \tilde{\$}$

۷- در یک دوره از ۶۰۰ بیمار دیابتی ۷۵ نفر آنها با یک روش جدید درمان، بهبود یافته‌اند. (نسبت بهبود ۱۲/۵٪ است).

فاصله اطمینان ۹۵٪ برای نسبت بهبودی واقعی این بیماری در روش جدید بسازید. $\# \langle \hat{\theta} \rangle \tilde{\$}$

- ۸- مراجعه کنندگان به بخش قلب یک بیمارستان دارای فشار خون، با انحراف معیار ۹ میلی متر جیوه هستند و توزیع فشار خون نرمال است. در یک نمونه ۳۶ تایی میانگین فشار خون $\bar{x} = 138$ را میلی متر جیوه بدست آورده ایم. **#£ùò Ñ\$**
- فاصله ای بدست آورید که با احتمال ۹۵٪ میانگین واقعی فشار خون را در بر داشته باشد.
- ۹- یک محقق در نمونه شامل ۱۰۰ نفر مقدار میانگین تعداد دفعات سرما خوردگی در سال را «۶» و انحراف معیار را «۳» بدست آورده است. فاصله ای بسازید که با اطمینان ۹۹ درصد میانگین واقعی تعداد دفعات بیماری در سال را در بر داشته باشد. **#£ùò Ñ\$**
- ۱۰- در یک تحقیق در مورد بیماری فشار خون، از بین ۵۵ نفر زن و ۷۳ نفر مرد به ترتیب ۸ و ۱۱ نفر داروی فشار خون مصرف می کردند.
- الف- برای نسبت کسانی که در هر یک از جمعیت های زن و مرد داروی فشار خون مصرف می کنند، یک فاصله اطمینان ۹۵ درصد بسازید. **#£ùò Ñ\$**
- ب- برای اختلاف بین نسبت مصرف کنندگان داروی فشار خون در دو جمعیت زن و مرد، یک فاصله اطمینان ۹۹ درصد بسازید. **#£ùò Ñ\$**
- ۱۱- در تنها زایشگاه یک شهر ۸۴ نوزاد در یک ماه گذشته متولد شده اند که ۳۷ تای آنها دختر بوده است. یک فاصله اطمینان ۹۹ درصدی برای نسبت جنسیت در این شهر تعیین کنید. **#£ùò Ñ\$**

پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

رشته‌های علوم پزشکی و شاخه‌های مرتبط

۲ واحد درسی

مطابق با سرفصل‌های مصوب

نمونه گیری و بر آورد حجم نمونه

روشهای جمع آوری اطلاعات

منظور از جمع آوری اطلاعات، اطلاعاتی است که بر اساس موضوع تحقیق و تعیین متغیرهای مطالعه از جامعه آماری برداشت می‌شود. در آمار دو روش کلی برای جمع آوری اطلاعات مطرح است :

نمونه گیری

مقصود از نمونه گیری یعنی انتخاب بخشی از جامعه که نماینده همه جامعه باشد. یعنی انتخاب به گونه ای صورت گیرد که تا حد امکان همه خصوصیات و جزئیات جامعه در نمونه ای که بر می داریم، مشاهده شود.

سرشماری

یعنی جمع آوری اطلاعات از تمامی اعضای یک جامعه آماری؛ که عملاً این کار بسیار پرهزینه ، وقت گیر و در بعضی مواقع غیر ممکن است. حتی گاهی کسب اطلاع منجر به انهدام واحدهای جامعه میگردد. به همین دلیل تقریباً در همه کشورها، هر ۱۰ سال یک بار اقدام به سرشماری عمومی می کنند.

روشهای جمع آوری اطلاعات

روشهای نمونه گیری بسیار متنوع است و بر اساس توزیع جمعیت و نوع مطالعه و بسیاری ملاحظات دیگر، روشهای متفاوتی در جمع آوری اطلاعات به طریق نمونه گیری وجود دارد. به این نکته نیز توجه داشته باشید که جمع آوری اطلاعات ممکن است از نظر نوع ابزار جمع آوری به یکی از صورت های پرسشنامه، مصاحبه، مشاهده، جامعه سنجی، کتابخانه و ... انجام شود.

روشهای نمونه گیری

طبقه ای

خوشه ای

سیستماتیک

تصادفی ساده

روشهای جمع آوری اطلاعات

نمونه گیری تصادفی ساده

قرعه کشی

وقتی تعداد عضوهای جامعه تحت مطالعه، در حدی باشد که بتوان به هر یک از اعضای جامعه شماره‌ای نسبت داد، به طوری که شانس انتخاب هر یک از اعضا یکسان باشد، از نمونه گیری تصادفی ساده استفاده کنید.

در این روش ابتدا باید قادر باشید فهرستی از تمامی اعضای جمعیت در اختیار بگیرید و از آن به عنوان چهارچوب نمونه گیری استفاده کنید. در این روش به هر عضو شماره‌ای نسبت می‌دهید. سپس با قرعه کشی و یا استفاده از جدول اعداد تصادفی نمونه‌های مورد نظرتان را به تعداد مورد لزوم انتخاب می‌کنید.

نمونه گیری تصادفی ساده را می‌توانید برای انتخاب تعدادی از بیماران در یک بیمارستان یا انتخاب تعدادی از دانش‌آموزان یک مدرسه و از این قبیل به کار گیرید.

این روش نمونه گیری برای جوامعی گسترده و با حجم زیاد، مناسب نیست.

روشهای جمع آوری اطلاعات

نمونه گیری سیستماتیک

- این روش نمونه گیری که به روش نمونه گیری منظم نیز معروف است دارای دو مرحله است .
- ❖ در مرحله اول شما باید یک نمونه گیری تصادفی ساده انجام دهید تا اولین عضو انتخابی جامعه معلوم شود.
 - ❖ در مرحله دوم نمونه ها را با فاصله های یکسان از یکدیگر انتخاب کنید.

مثال:

فرض کنید در یک روزنامه ۴۰۰۰ نفر مشترک هستند و هر مشترک دارای یک شماره اشتراک است. مدیر این روزنامه مایل است ۲۵۰ نفر از مشترکین را به عنوان نمونه انتخاب و از آنان در باره اضافه کردن یک صفحه بهداشت خانواده نظرخواهی کند. برای انتخاب ۲۵۰ نمونه، او ابتدا باید از بین اعداد ۱ تا ۱۶ (۱۶ از تقسیم ۴۰۰۰ بر ۲۵۰ به دست آمده است)، یک عدد را به تصادف انتخاب نماید. فرض کنید آن عدد ۵ باشد در این مرحله مشترک شماره ۵ از بین ۱۶ نفر اول انتخاب شده است. مشترک بعدی ، مشترک شماره بیست و یک خواهد بود (زیرا $21 = 5 + 16$). سپس مشترک شماره ۳۷ انتخاب بعدی و به همین صورت ادامه می دهد تا مشترک شماره ۳۹۸۹ که دویست و پنجاهمین عضو نمونه است، انتخاب شود.

روشهای جمع آوری اطلاعات

نمونه گیری طبقه ای

اگر ساختار جامعه‌ای که می‌خواهید از آن نمونه گیری کنید به گونه‌ای است که به چند طبقه طوری تقسیم شده است که تقریباً اعضای درون هر طبقه از نظر صفت مورد بررسی همگن هستند، می‌توانید از نمونه‌گیری طبقه‌ای استفاده کنید. در این روش می‌توان با یک نمونه کم در هر طبقه، برآورد دقیقی از صفت تحت بررسی تهیه کرد. برآورد هایی که در هر طبقه همگن تهیه می‌شوند سرانجام ترکیب شده و برآوردی دقیق برای صفت مورد نظر فراهم می‌کنند.

مثال

یک موسسه تحقیقاتی می‌خواهد متوسط مدت زمانی که بیماران دیابتی در ۱۲ بیمارستان یک شهرستان بستری می‌شوند را برآورد نماید. هر بیمارستان به عنوان یک طبقه در نظر گرفته شده از هر طبقه تعدادی بیمار دیابتی به عنوان نمونه انتخاب شده و مدت زمان بستری هر یک ثبت می‌شود. در این مثال فرض بر این است که صفت مورد بررسی در هر بیمارستان یکسان است.

روشهای جمع آوری اطلاعات

مثال

یک موسسه دارای ۵۰۰ کارمند در ۴ طبقه شغلی است. مدیر موسسه مایل است میزان بهره‌وری در هر طبقه شغلی را بررسی کند. برای این منظور قصد دارد از روش نمونه‌گیری طبقه‌ای، یک نمونه ۸۰ تایی از کارمندان را انتخاب و مورد بررسی قرار دهد. روش انتخاب تصادفی ساده در هر گروه به نسبت حجم آنها در جدول زیر آمده است.

طبقه‌های شغلی	۱	۲	۳	۴
حجم هر طبقه	۷۰	۱۰۰	۱۵۰	۱۸۰
تعداد نمونه	$\frac{۷۰}{۵۰۰} \times ۸۰ = ۱۱$	$\frac{۱۰۰}{۵۰۰} \times ۸۰ = ۱۶$	$\frac{۱۵۰}{۵۰۰} \times ۸۰ = ۲۴$	$\frac{۱۸۰}{۵۰۰} \times ۸۰ = ۲۹$

روشهای جمع آوری اطلاعات

نمونه گیری خوشه ای

وقتی حجم جامعه خیلی بزرگ نباشد بکارگیری شیوه هایی نظیر تصادفی ساده و سیستماتیک بدون اشکال است اما وقتی حجم جامعه بزرگ می شود این روش ها با مشکلاتی مواجه هستند.

در نمونه گیری خوشه ای برعکس نمونه گیری طبقه ای دلیلی وجود ندارد که صفت تحت بررسی در هر خوشه همگن باشد. بنابراین اگر صفت مورد بررسی در یک جمعیت به گونه ای است که نمی توان جمعیت را به طبقات همگن تقسیم کرد از نمونه گیری خوشه ای استفاده می کنیم.

مثال:

فرض کنید در یک یک موسسه می خواهیم تعداد فرزندان کارمندان را به عنوان یک صفت، مورد بررسی قرار دهیم. این موسسه شامل ۸ بخش است که هر بخش را به عنوان یک خوشه تلقی می کنیم. ابتدا به صورت تصادفی از بین ۸ بخش (خوشه) تعدادی را انتخاب می کنیم. این تعداد خوشه به تعداد نمونه بستگی دارد. سپس تعداد فرزندان همه کارمندان را ثبت می کنیم. چنین نمونه گیری را نمونه گیری خوشه ای یک مرحله ای می گویند.

روشهای جمع آوری اطلاعات

ممکن است نمونه گیری خوشه‌ای دو مرحله‌ای یا سه مرحله‌ای یا بیشتر باشد.

مثال

فرض کنید می‌خواهید میانگین تعداد فرزندان خانواده‌ها را در یک استان که دارای ۷ شهرستان و هر شهرستان دارای تعدادی بخش است، برآورد کنید. اگر در مرحله اول از بین ۷ شهرستان تعدادی را به تصادف انتخاب کنید و سپس از بین بخش‌های مختلف آن، تعدادی را به عنوان نمونه برگزینید و اطلاعات را از تمامی خانواده‌های ساکن در بخش‌های انتخابی جمع‌آوری نمایید، نمونه گیری خوشه‌ای دو مرحله‌ای انجام شده است. اینکه تمام خانواده‌ها را در هر بخش مورد بررسی قرار دهیم یا تعدادی را به تصادف انتخاب کنیم، بستگی به حجم نمونه‌گیری داشته و نمونه گیری از دو مرحله‌ای به نمونه گیری خوشه‌ای ۳ مرحله‌ای تبدیل می‌شود و بسته به ترکیب جمعیت به همین ترتیب می‌توان نمونه گیری را به n مرحله خوشه بندی کرد.

برآورد حجم نمونه

همواره یکی از مشکلات پژوهشگران در ابتدای تحقیق این است که از جامعه تحت مطالعه چند نمونه انتخاب کنند تا نتایج بدست آمده را بتوان با اطمینان بالایی به عموم جامعه نسبت داد. از طرفی حجم نمونه مناسب از نظر تامین دقت نتایج و صرفه جویی در وقت و هزینه نیز اهمیت دارد. بزرگ بودن نمونه باعث هزینه و وقت زیاد و کوچک بودن آن موجب عدم دقت کافی در برآوردها می شود. روشهای متفاوتی بر اساس ملاحظات آماری برای برآورد حجم نمونه وجود دارد.

در تعیین حجم نمونه عوامل زیر موثرند:

واریانس

اگر واریانس صفتی که می خواهید آن را به وسیله نمونه گیری برآورد کنید زیاد باشد، طبیعی است که به حجم نمونه ی بیشتری نیاز خواهید داشت.

سطح اطمینان

انتظاری که از نمونه می رود این است که بتواند پارامترها را با دقت کافی برآورد کند و این انتظار اطمینان بالایی را می طلبد و افزایش اطمینان، افزایش حجم نمونه را در پی خواهد داشت.

حجم جامعه

از آنجایی که باید صفت مورد بررسی به جامعه تعمیم داده شود، در یک جمعیت با حجم زیاد باید تعداد نمونه ها به قدری بزرگ باشند که ویژگی های جمعیت در آنها مشاهده شود.

توزیع جامعه

یکی دیگر از عوامل موثر بر تعداد حجم نمونه توزیع احتمالی صفتی است که در جامعه وجود دارد.

برآورد حجم نمونه

سطح اطمینان و دقت توسط محقق تعیین می شود اما به طور معمول واریانس داده ها از قبل مشخص نیست. می توان با استفاده از روش های زیر واریانس را معین کرد:

از اطلاعات به دست آمده در مطالعات قبلی استفاده کرد.

بر اساس یک نمونه گیری اولیه (پیش آزمون) واریانس را برآورد کرد.

گاهی در پرسشنامه ها که پاسخ ها کمی هستند و از کمترین و بیشترین مقادیر آنها اطلاع داریم، می توان واریانس را از رابطه زیر محاسبه کرد:

$$\sigma^2 = \frac{\max(x_i) - \min(x_i)}{6}$$

مثال

در مورد طیف پنج گزینه ای لیکرت، اگر ماهیت متغیر به گونه ای باشد که مجاز باشیم پاسخ ها را کمی در نظر بگیریم، واریانس را به صورت زیر به دست می آوریم:

$$s^2 = \frac{5-1}{6} = 0.667$$

برآورد حجم نمونه

حالت اول:

اگر متغیر مطالعه کمی است و حجم جامعه (N) زیاد است،

برای برآورد تعداد نمونه ابتدا باید واریانس (σ^2) را با یکی از راههایی که در بالا توضیح داده شد، به دست آورید. در گام بعدی باید دقت برآورد را تعیین کنید یعنی تعیین کنید تا چه میزان در اندازه گیری صفت مورد نظر خطا مجاز است (d^2). سپس میزان اطمینان را (که اغلب اوقات ۹۵٪ در نظر گرفته می شود) مشخص کنید و در نهایت از رابطه زیر تعداد نمونه لازم را معلوم کنید:

$$n = \frac{(z_{1-\frac{\alpha}{2}})^2 \times \sigma^2}{d^2}$$

در این رابطه، $z_{1-\frac{\alpha}{2}} = 1/96$ ، مقداری از جدول توزیع نرمال استاندارد و با توجه به اطمینان ۹۵٪ است.

برآورد حجم نمونه

مثال

فرض کنید مدیر یکی از شبکه‌های سیما مایل است بداند میانگین زمان تماشای تلویزیون در بین افراد ساکن یک منطقه چقدر است؟ برای این بررسی او به چه تعداد نمونه نیاز دارد تا با اطمینان ۹۵ درصد برآورد درستی از میانگین انجام داده باشد. او مقدار خطای اندازه‌گیری را ۵ دقیقه در نظر گرفته است و از مطالعات قبل می‌داند انحراف معیار زمان تماشای تلویزیون ۴۰ دقیقه است.

$$n = \frac{(1/96)^2 \times (40)^2}{5^2} = 245/9 \approx 246$$

برآورد حجم نمونه

حالت دوم:

متغیر مطالعه کمی و حجم جامعه (N) محدود است

پس از مشخص کردن واریانس، دقت برآورد و میزان اطمینان برای برآورد تعداد نمونه از رابطه زیر استفاده کنید:

$$n = \frac{N (z_{\frac{1-\alpha}{2}})^2 \times \sigma^2}{(N-1)d^2 + (z_{\frac{1-\alpha}{2}})^2 \times \sigma^2}$$

مثال

یک محقق برای بررسی وضعیت چرخه دانش در یک دانشگاه با ۲۴۰ عضو هیئت علمی می‌خواهد تعداد نمونه لازم را مشخص نماید. او در یک پیش آزمون انحراف معیار را ۱/۲ به دست آورده است و می‌خواهد این تحقیق را با اطمینان ۹۵٪ به سرانجام برساند. تعداد نمونه را چقدر در نظر گیرد وقتی مایل است خطای اندازه گیری در حد ۳٪ باشد.

$$n = \frac{240 \times (1/96)^2 \times (1/2)^2}{239 \times (0/03)^2 + (1/96)^2 \times (1/2)^2} = 90$$

برآورد حجم نمونه

حالت سوم:

متغیر مطالعه دارای مقیاس رتبه ای و ارزشی هستند و حجم جامعه (N) زیاد است

پس از مشخص کردن واریانس، دقت برآورد و میزان اطمینان برای برآورد تعداد نمونه از رابطه زیر استفاده کنید:

$$n = \frac{(z_{1-\frac{\alpha}{2}})^2 \times pq}{d^2}$$

وقتی می خواهیم نسبتی (مانند p) را در یک جمعیت برآورد کنیم. برای تعیین حجم نمونه از این فرمول استفاده می کنیم.

که در این رابطه p نسبت موفقیت و $q = 1 - p$ است. مقدار p را مانند واریانس از اطلاعات گذشته یا از پیش آزمون به دست می آوریم و اگر هیچ یک از این ها مقدور نباشد، مقدار آنرا ۰/۵ در نظر می گیریم. زیرا بیشترین مقدار واریانس برای نسبت وقتی است که $p = ۰/۵$ باشد.

برآورد حجم نمونه

مثال

محقق می خواهد با اطمینان ۹۵ درصد نسبت استفاده کنندگان از خدمات اینترنتی را در یک شهرستان برآورد کند. او می داند این نسبت در سال گذشته ۲۱ درصد بوده است. محقق به چه تعداد نمونه نیاز خواهد داشت تا در تحقیق او، خطای اندازه گیری بیشتر از ۵ درصد نباشد.

$$n = \frac{(1/96)^2 \times (0/21)(0/79)}{(0/05)^2} = 256$$

برآورد حجم نمونه

حالت چهارم:

متغیر مطالعه دارای مقیاس رتبه ای و ارزشی هستند و حجم جامعه (N) کم است

پس از مشخص کردن واریانس، دقت برآورد و میزان اطمینان برای برآورد تعداد نمونه از رابطه زیر استفاده کنید:

$$n = \frac{N (z_{\frac{1-\alpha}{2}})^2 \times pq}{(N-1)d^2 + (z_{\frac{1-\alpha}{2}})^2 \times pq}$$

مثال

محقق می خواهد حجم نمونه را برای یک تحقیق پرسشنامه ای با طیف پنج گزینه ای لیکرت از جامعه ای با تعداد ۱۵۰ عضو، برآورد نماید. اگر سطح اطمینان در این تحقیق ۹۵٪ باشد و از نسبت موفقیت (p) اطلاعی در دست نباشد، حجم نمونه لازم را چه مقدار اختیار کند تا خطای اندازه گیری بیشتر از ۶٪ نشود.

$$n = \frac{150 \times (1/96)^2 \times (0/5)^2}{149 \times (0/06)^2 + (1/96)^2 \times (0/5)^2} = 97$$



تمرین

- ۱- روش نمونه گیری سیستماتیک را توضیح دهید
 - ۲- تفاوت‌های عمده روش نمونه‌گیری طبقه‌ای با دو روش نمونه‌گیری تصادفی ساده و سیستماتیک را بنویسید
 - ۳- از یک جمعیت با سه طبقه به ترتیب به حجم‌های ۵۶۰ و ۴۸۵ و ۹۵۵ می‌خواهیم به روش نمونه‌گیری طبقه‌ای ۲۳۰ نمونه انتخاب کنیم. سهم هر طبقه را مشخص کنید.
 - ۴- برای برآورد نسبت افراد سیگاری با اطمینان ۹۵ درصد، در یک جمعیت با تعداد ۳۱۰ نفر چه تعداد نمونه لازم است وقتی می‌خواهیم خطای برآورد از ۳ درصد بیشتر نباشد. (درمورد واریانس اطلاعی نداریم).
 - ۵- ۳۶ درصد زنان تهرانی چاقی مفرط دارند. (قد و وزن متناسب نیست). محقق می‌خواهد این موضوع را روی ۱۴۵۵ زن شاغل در یک موسسه بررسی نماید.
- الف- با اطمینان ۹۵ درصد و مقدار خطای ۳ درصد، چه تعداد نمونه لازم دارد تا این نسبت را برآورد کند؟

- ۵- در یک سال از ۶۰۰ بیمار سرطانی ۱۲۰ نفر آنها فوت شده اند (نسبت کشندگی ۲۰٪ است). با توجه به این اطلاعات چه مقدار حجم نمونه لازم است تا بتوان مقدار واقعی نسبت کشندگی بیماری سرطان را در سطح ۹۵٪ اطمینان، برآورد کرد. (خطای برآورد را یک پنجم مقدار قبلی در نظر بگیرید).
- ۶- به منظور برآورد میانگین فشار خون ۷۵ زن شاغل در یک موسسه به چه تعداد حجم نمونه نیاز است تا میزان خطای برآورد از ۲ میلی متر بیشتر نباشد. می‌دانیم واریانس فشار خون در مطالعات قبلی ۸ بوده است
- ۷- یک محقق می‌خواهد میانگین هزینه درمان برای یک بیماری را با اطمینان ۹۵ درصد و خطای کمتر از ۱۰ برآورد کند. حجم نمونه را برای او معین کنید.
- ۸- حجم نمونه لازم برای برآورد نسبت مرگ و میر نسبی چقدر باشد تا دقت برآورد ۰/۰۴ و میزان اطمینان ۹۹ درصد باشد. از مطالعات قبلی می‌دانیم این نسبت ۰/۰۹ بوده است.
- ۹- به منظور برآورد میزان شیوع سل ریوی در یک استان، در سطح اطمینان ۹۵ درصد، به چه تعداد نمونه احتیاج است، اگر این میزان در مطالعه قبلی ۳ در ۱۰۰۰ برآورد شده باشد و بخواهیم خطای برآورد از یک چهارم میزان قبلی بیشتر نباشد؟

پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

رشته‌های علوم پزشکی و شاخه‌های مرتبط

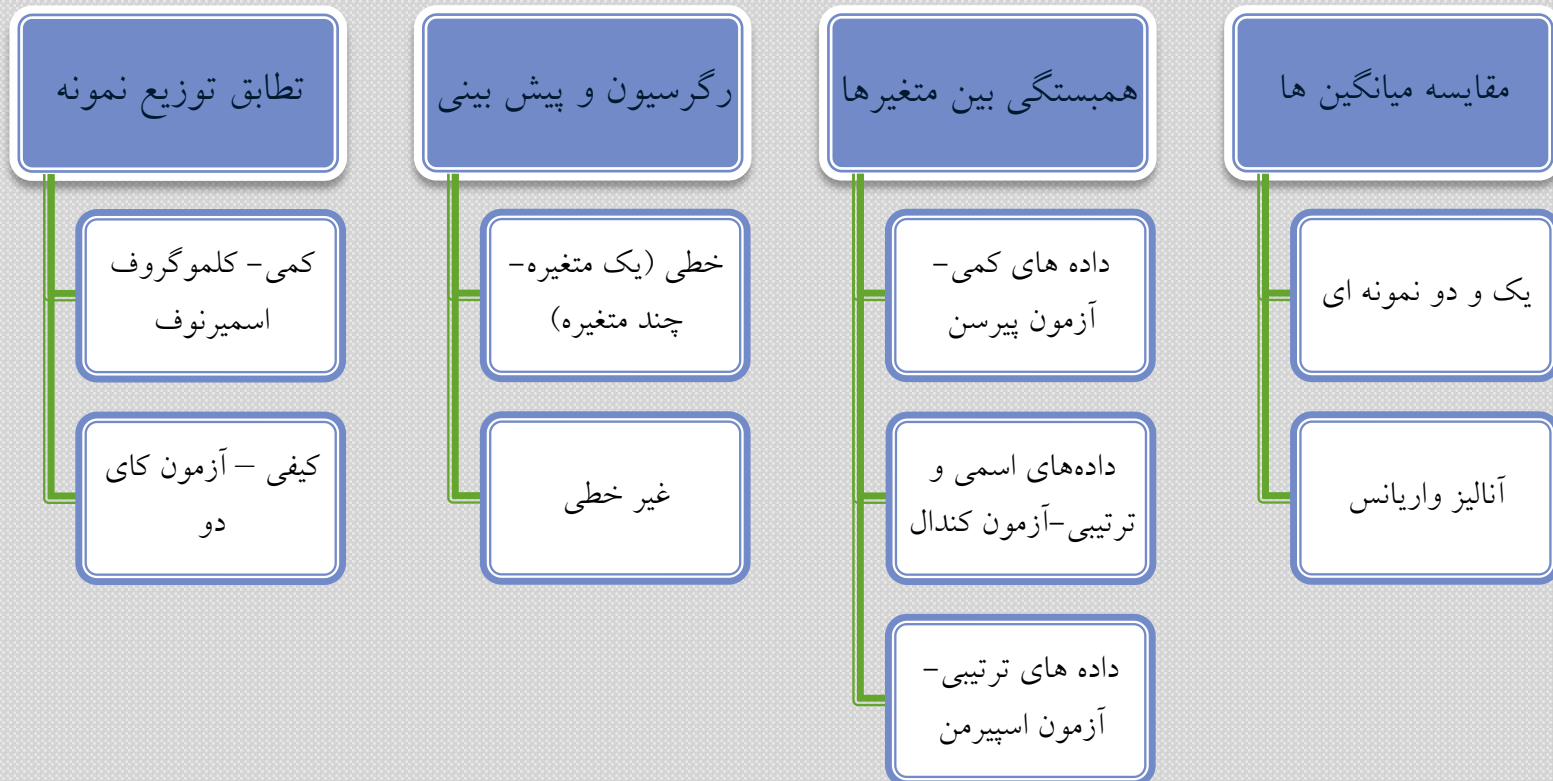
۲ واحد درسی

مطابق با سرفصل‌های مصوب

آزمون فرضی ۱

آزمون فرضیه

● هر تحقیق با طرح یک یا چند پرسش کلی شروع می‌شود. معمولاً سؤالات هر تحقیق را می‌توان در ۴ دسته کلی زیر قرار داد:



آزمون فرضیه

- آزمون فرض: (Hypothesis Testing) روندی است که در طی مراحل انجام آن و با کمک روش‌های آماری می‌خواهیم درستی یا نادرستی یک فرضیه (ادعا) را آزمون کنیم.
- فرضیه: ادعایی است در باره پارامترهای یک یا چند جمعیت که ممکن است درست و یا غلط باشد.

مثال:

۱. میانگین مدت اقامت بستری در یک بیمارستان ۵ روز است.
۲. عامل سیگار بر سرطان ریه مؤثر است.
۳. برنامه‌های آموزش بهداشت منجر به بهبود وضع سلامت جامعه می‌شود.

آزمون فرضیه

- هر پاسخی که به پرسش‌های یک تحقیق داده می‌شود، یک فرضیه است. این فرضیه ها منجر به یک یا گاهی چند آزمون آماری می‌شود.
انواع فرضیه:

فرضیه صفر: آنچه در جمعیت موجود است و درستی آن به اثبات رسیده است و به سادگی از آن نمی‌گذریم مگر با اطمینان بالایی درستی آن به مخاطره بیفتد و آن را با H_0 نشان می‌دهیم.
فرضیه یک: فرضیه مخالف فرض صفر است و آن را با H_1 نشان می‌دهیم.

فرض صفر: میانگین سن مردانی که اولین بار نشانه ریزش مو را مشاهده می‌کنند، ۳۶ سال است.
فرض یک: میانگین سن مردانی که اولین بار نشانه ریزش مو را مشاهده می‌کنند، بیشتر از ۳۶ سال است.

مثال

این دو فرضیه را به صورت خلاصه زیر می‌نویسند:

$$\begin{cases} H_0: \mu = 36 \\ H_1: \mu > 36 \end{cases}$$

آزمون فرضیه

شاخص آماری آزمون: شاخصی است که مقدار آن را به کمک یک نمونه تصادفی به دست می‌آوریم و به عنوان معیاری برای رد یا پذیرش فرض صفر به کار می‌رود. این شاخص به صورت کلی زیر است:

$$Q_{H0} = \frac{\text{مقدار پارامتر تحت فرض صفر} - \text{مقدار برآورد شده پارامتر}}{\text{خطای معیار برآورد (انحراف معیار برآوردگر)}}$$

سطح معنی داری: میزان خطایی است که ممکن است در رد کردن فرضیه صفر مرتکب شویم. به عبارتی احتمال رد کردن فرض صفر وقتی فرض صفر درست باشد و آن را با α نشان می‌دهیم. بدیهی است سطح اطمینان $1 - \alpha$ خواهد بود.

آزمون فرضیه

در اغلب پژوهش‌ها سطح معنی‌داری α را مقدار ۰/۱ یا ۰/۰۵ یا ۰/۰۱ در نظر می‌گیرند. این مقدار را محقق قبل از انجام آزمون تعیین می‌کند.

- در انجام هر آزمون آماری ممکن است دو نوع خطا صورت گیرد:
- ۱- خطای نوع اول (α) = احتمال رد فرض صفر است وقتی فرض صفر درست باشد که همان سطح معنی‌داری آزمون است.
 - ۲- خطای نوع دوم (β) = احتمال رد فرض یک است وقتی فرض یک درست باشد.

	فرض صفر درست	فرض صفر غلط
رد فرض صفر	خطای نوع اول α	توان آزمون $1 - \beta$
پذیرش فرض صفر	سطح اطمینان $1 - \alpha$	خطای نوع دوم β

آزمون فرضیه

توان آزمون

توان یا قدرت آزمون عبارت است از احتمال رد کردن فرض صفر وقتی که در حقیقت فرض صفر نادرست باشد. روشن است که در هر استنباط آماری خطر ارتکاب هر یک از دو خطای نوع اول و دوم وجود دارد. در عمل پژوهشگر قبل از انجام آزمون آماری خطای نوع اول α و تعداد نمونه n را مشخص می‌کند که خطای نوع دوم β نیز به طبع آن مشخص خواهد شد. از آنجا که بین α و β رابطه معکوس وجود دارد، بدون توجه به حجم نمونه (n)، افزایش در مقدار α باعث کاهش در مقدار β خواهد شد. لازم است که پژوهشگر به نوعی سازش که تعادل بین احتمال این دو خطا را به حد مطلوب برساند، دست یابد.

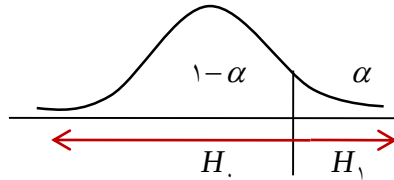
اگر محقق بخواهد که احتمال هر دو نوع خطا را کاهش دهد که به طور ضمنی افزایش توان آزمون را نیز در پی خواهد داشت، راهی جز افزایش حجم نمونه نخواهد داشت.

آزمون فرضیه

تعیین فرضیه ها

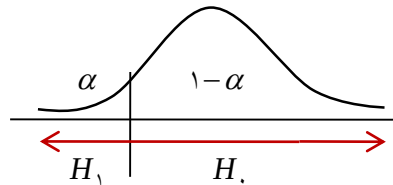
در تعیین فرضیه‌های صفر و یک، حالات مختلفی پیش می‌آید که منجر به آزمون‌های یکطرفه یا دو طرفه می‌شود. در آزمون‌های یکطرفه برعکس فرض H_0 که با مساوی همراه است، فرض H_1 با عبارت «کوچکتر» یا «بزرگتر» همراه است.

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$



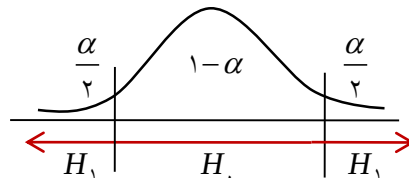
آزمون یکطرفه راست: در این نوع آزمون H_1 به اندازه α در دنباله راست توزیع نمونه گیری شاخص آزمون قرار دارد و سطح $1-\alpha$ در سمت چپ مربوط به H_0 است.

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$$



آزمون یکطرفه چپ: در این نوع آزمون H_1 به اندازه α در دنباله چپ توزیع نمونه گیری شاخص آزمون قرار دارد و سطح $1-\alpha$ در سمت راست، مربوط به H_0 است.

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$



آزمون دوطرفه: در این نوع آزمون H_1 به اندازه $\frac{\alpha}{2}$ در دنباله چپ و به اندازه $\frac{\alpha}{2}$ در دنباله راست توزیع نمونه گیری قرار دارد و سطح $1-\alpha$ در وسط مربوط به H_0 است.

آزمون فرضیه

مثال

آزمون
یکطرفه راست

میانگین طول عمر در یک جمعیت ۶۳ سال است.
فردی ادعا می کند این مقدار در سال گذشته افزایش داشته است.

$$\begin{cases} H_0: \mu = 63 \\ H_1: \mu > 63 \end{cases}$$

آزمون
یکطرفه چپ

نسبت افراد سیگاری در یک جمعیت ۲۶٪ است.
محقق می خواهد بررسی کند آیا این نسبت اخیراً کاهش یافته است؟

$$\begin{cases} H_0: P = 0.26 \\ H_1: P < 0.26 \end{cases}$$

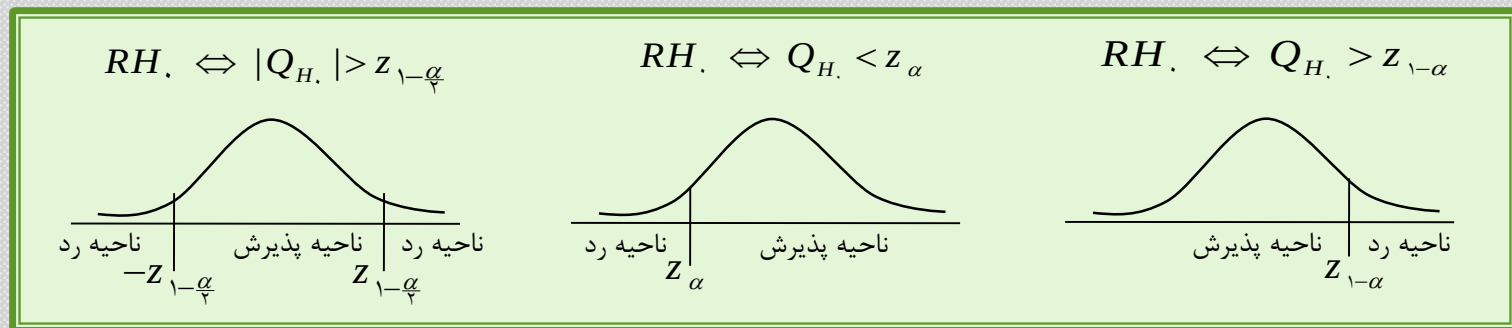
آزمون
دو طرفه

میانگین مصرف آب یک مجتمع مسکونی در هر ماه ۳۸۵۰ لیتر است.
مدیر مجتمع می خواهد بررسی کند آیا این میزان تغییر یافته است؟

$$\begin{cases} H_0: \mu = 3850 \\ H_1: \mu \neq 3850 \end{cases}$$

آزمون فرضیه

استراتژی رد فرض صفر: هنگامی می‌توان فرض صفر را رد کرد که داده‌های مورد بررسی با شرایطی که فرضیه صفر واقعا صحیح است، مطابقت نداشته باشد. این کار با قرار گرفتن مقدار شاخص آماری آزمون در ناحیه رد، مورد تایید قرار می‌گیرد و ناحیه رد نیز بر اساس توزیع احتمال آن و با توجه به سطح معنی داری آزمون تعیین می‌شود.



آزمون فرضیه

P-Value

P-Value که آن را احتمال معنی داری برای یک آزمون نیز می‌گویند، کوچکترین مقداری است که می‌توان برای آلفا در نظر گرفت تا فرض صفر رد شود. از طرفی آلفا سطح خطایی است که محقق برای رد فرض صفر در ابتدای آزمون در نظر می‌گیرد ولی P-Value مقدار خطای محاسبه شده در رد فرض صفر است. مزیت محاسبه P-Value این است که اخذ تصمیم در مورد رد یا پذیرش فرض صفر را به محقق واگذار می‌کند. هر چند از قبل مقدار آن را معلوم کرده است ولی می‌تواند در مورد مقدار آن تجدید نظر کند. مثلاً در یک آزمون اگر مقدار خطای محاسبه شده برابر $P\text{-Value}=0/04$ باشد، فرض صفر با آلفای کمتر از این مقدار پذیرفته است ولی اگر محقق آلفا را بیشتر از این مقدار در نظر بگیرد، فرض صفر رد می‌شود.

P-Value معمولاً در نرم افزارهای تحلیل آماری مورد استفاده قرار گرفته و محاسبه می‌شود تا محقق با مشاهده آن بتواند در مورد رد یا پذیرش فرض صفر به صورت زیر تصمیم‌گیری نماید:

$$R H, \Leftrightarrow P\text{-Value} < \alpha$$

آزمون فرضیه

مراحل انجام یک آزمون آماری

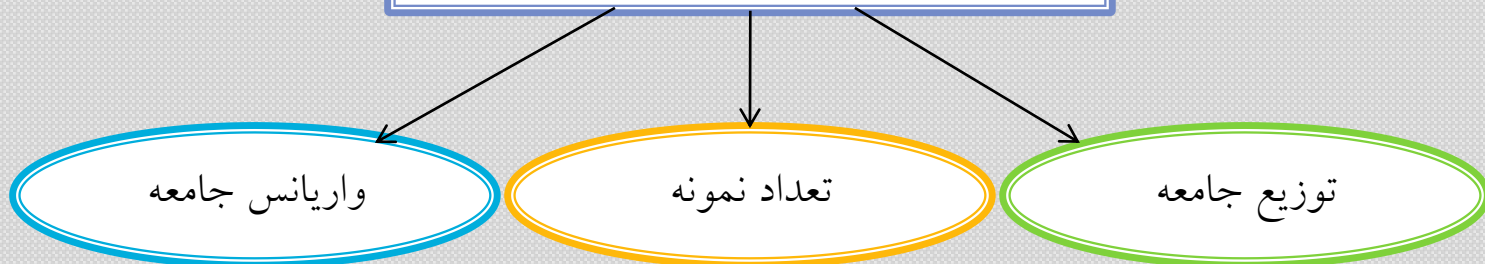
(۱) فرضیه‌های آزمون را تشکیل دهید.

(۲) سطح معنی داری آزمون (α) را مشخص کنید.

(۳) شاخص آماری آزمون (Q_H) را محاسبه و توزیع آن را معلوم کنید.

(۴) قاعده تصمیم‌گیری را برای رد یا پذیرش فرضیه H ، بکار گرفته و تصمیم‌گیری نمایید.

توزیع شاخص آزمون به سه عامل بستگی دارد:



آزمون های میانگین

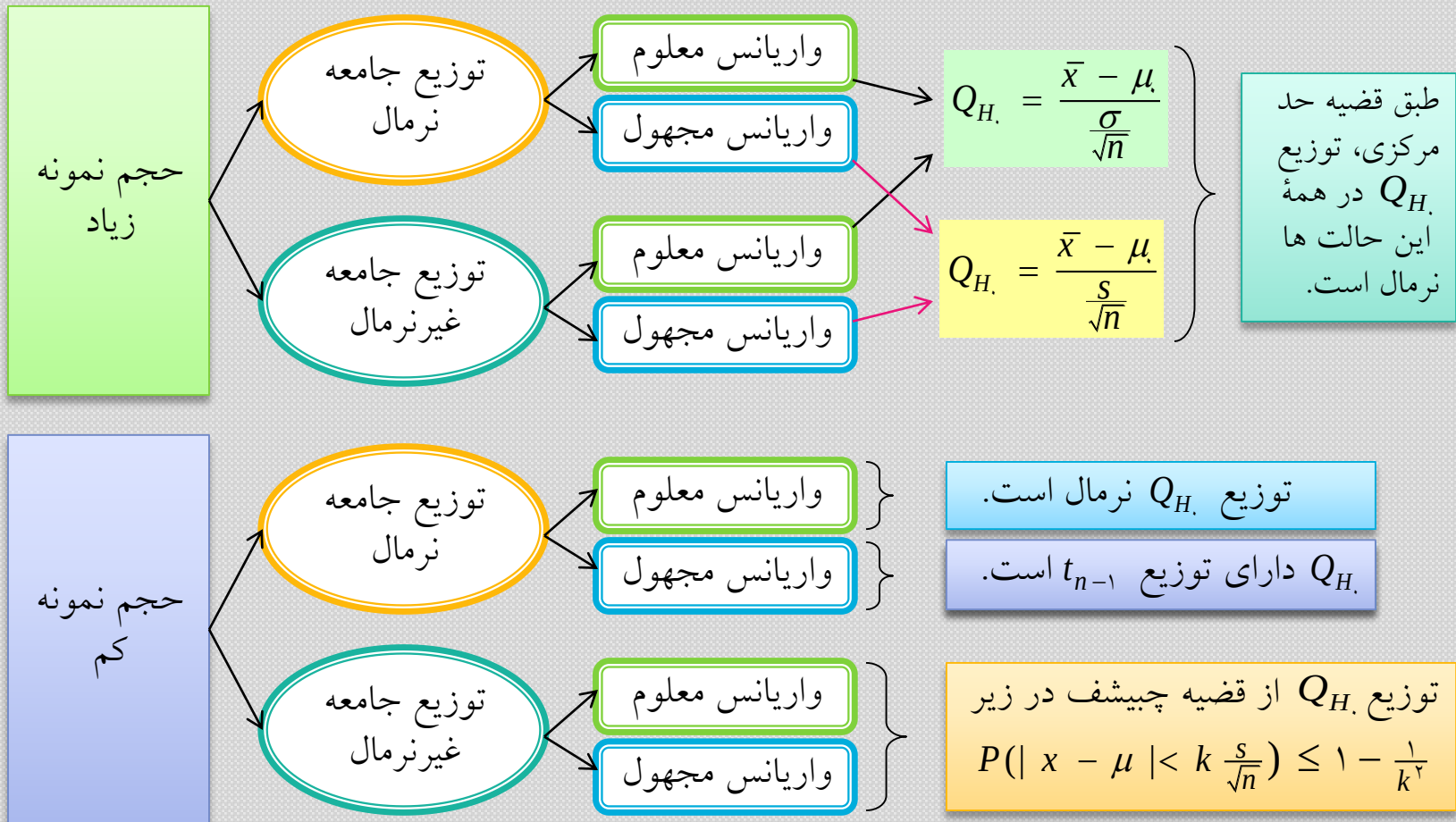
مقدمه

در بسیاری از مطالعات پزشکی بحث بر سر این موضوع است که آیا بین دو یا چند گروه اختلافی وجود دارد یا خیر؟ این موضوع در مطالعات تجربی که در آنها عملکرد افراد تحت مطالعه، در شرایط متفاوت با هم مقایسه می شود، بیشتر نمود پیدا می کند. به طور معمول بحث اختلاف بین دو یا چند گروه، بیشتر به اختلاف بین میانگین های آنها منجر می شود. زیرا معمولاً در مطالعات تجربی که در فصل اول با آنها اشاره کردیم، متغیرها کمی هستند. مثلاً فرض کنید می خواهید فشار خون دو گروه زن و مرد را با یکدیگر مقایسه کنید. طبیعی است که باید از هر یک از جمعیت های زن و مرد، نمونه هایی را انتخاب کنید و فشار خون آنها را اندازه بگیرید و سپس میانگین فشار خون هر گروه را محاسبه و با هم مقایسه کنید.

در اغلب مطالعات آماری موضوع مقایسه میانگین ها مطرح است.

آزمون‌های میانگین

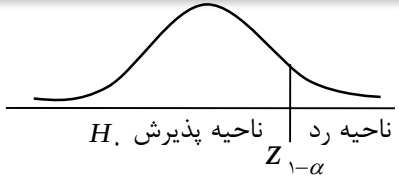
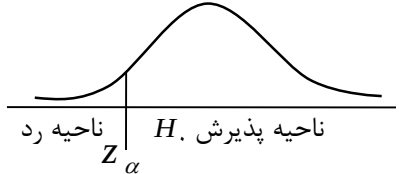
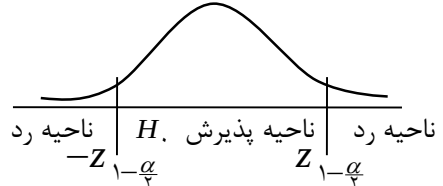
قبل از انجام آزمون‌های مربوط به میانگین در مورد توزیع شاخص آزمون به حالات زیر توجه کنید.



آزمون‌های میانگین

اگر فرضیه‌ای در خصوص میانگین یک جامعه مطرح شود. می‌توان آزمون را با انجام مراحل اشاره شده در قبل انجام داد. در یک آزمون فرض مربوط به میانگین، حالت‌های مختلف را در نظر می‌گیریم:

۱- توزیع جامعه نرمال و واریانس معلوم:

نوع آزمون	شاخص	توزیع	استراتژی	ناحیه رد و پذیرش
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} > z_{1-\alpha}$	
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} < z_{\alpha}$	
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} > z_{1-\frac{\alpha}{2}}$	

آزمون‌های میانگین

مثال

میانگین طول عمر در یک جمعیت ۶۳ سال با واریانس ۱۶ سال است و توزیع طول عمر نیز نرمال است. فردی ادعا می‌کند میانگین طول عمر در سال گذشته افزایش داشته است. محقق در یک نمونه ۳۶ تایی از افراد، میانگین سن را ۶۵ سال به دست می‌آورد. در سطح ۵٪ آیا ادعای فوق پذیرفته یا رد می‌شود؟

حل:

۱- فرضیه‌ها را تشکیل می‌دهیم.

$$\begin{cases} H_0: \mu = 63 \\ H_1: \mu > 63 \end{cases}$$

۲- سطح معنی داری ۵٪ تعیین شده است.

۳- توزیع شاخص آزمون نرمال (Z) و مقدار آن ۳ است:

$$Q_{H_1} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{65 - 63}{\frac{4}{\sqrt{36}}} = 3$$

۴- استراتژی: فرض صفر را وقتی رد می‌کنیم که مقدار شاخص آزمون از مقدار Z (در جدول) بیشتر باشد.

$$RH_0 \Leftrightarrow Q_{H_1} > Z_{1-\alpha}$$

تصمیم: چون مقدار شاخص آزمون ۳ از ۱/۹۵ بیشتر باشد.

بیشتر است؛ فرض صفر رد شده و ادعای فرد مورد قبول است.

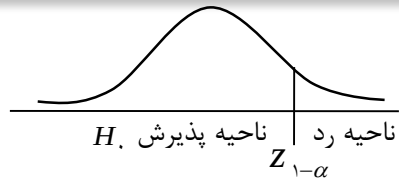
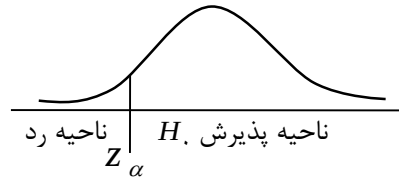
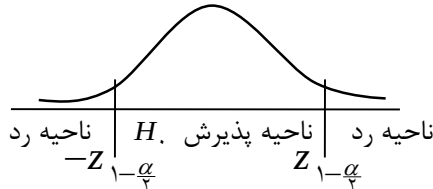
چون شاخص آزمون توزیع نرمال دارد، مقدار Z را با توجه به سطح معنی داری از جدول توزیع نرمال استاندارد به دست آورده‌ایم



آزمون‌های میانگین

۲- توزیع جامعه نرمال و واریانس نا معلوم:

در این وضعیت دو حالت پیش می‌آید. در صورتی که حجم نمونه به اندازه کافی بزرگ باشد ($n > 25$) شاخص آزمون دارای توزیع نرمال (Z) است و اگر تعداد نمونه کم باشد، از توزیع دیگری به نام t استفاده می‌کنیم.

نوع آزمون	شاخص	توزیع	استراتژی	ناحیه رد و پذیرش
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} > z_{1-\alpha}$	
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} < z_{\alpha}$	
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} > z_{1-\frac{\alpha}{2}}$	

آزمون‌های میانگین

مثال

وزن نوزادانی که مادران آنها با رژیم غذایی خاصی تغذیه شده‌اند، دارای توزیع نرمال با میانگین ۲۸۰۰ گرم است. در یک مطالعه جدید، یک گروه ۳۶ نفری از مادران باردار را تحت رژیم جدید غذایی قرار داده و میانگین وزن نوزادان آنها را ۲۸۵۰ با انحراف معیار ۱۸۰ گرم به دست آورده‌ایم. در سطح معنی‌داری ۰.۵٪ تغییر وزن نوزادان را بیازمایید.

فرض‌های آزمون به صورت زیر است:

$$\begin{cases} H_0: \mu = 2800 \\ H_1: \mu \neq 2800 \end{cases}$$

مقدار آماره آزمون را به دست می‌آوریم:

$$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{2850 - 2800}{\frac{180}{\sqrt{36}}} = 1.66$$

چون تعداد نمونه به اندازه کافی بزرگ است، توزیع شاخص آزمون، نرمال و استراتژی رد فرض صفر به صورت مقابل است:

$$RH_0 \Leftrightarrow |Q_{H_0}| > z_{1-\alpha/2}$$

از جدول توزیع نرمال استاندارد داریم: $Z_{0.975} = 1/96$

تصمیم: با توجه به اینکه مقدار آماره آزمون از مقدار جدول کوچکتر است، فرض صفر رد نمی‌شود.



آزمون‌های میانگین

۳- توزیع جامعه غیر نرمال و حجم نمونه به اندازه کافی بزرگ

در این وضعیت چون حجم نمونه به اندازه کافی بزرگ است، بنابر قضیه حد مرکزی توزیع شاخص آماری آزمون نرمال (Z) است.

مثال

میانگین زمان بستری شدن پس از یک عمل جراحی ۷ روز است. محقق به کمک پرونده های ۳۰ بیمار درمی‌یابد که میانگین زمان بستری ۵/۶ روز با واریانس ۹ است. در سطح معنی داری ۵٪ آیا مدت زمان بستری کاهش یافته است؟

حل

چون حجم نمونه به اندازه کافی بزرگ است، توزیع شاخص آماری آزمون نرمال (Z) خواهد بود. فرض‌های آزمون را می‌نویسیم و با توجه به مفروضات، مقدار شاخص آزمون را به دست می‌آوریم.

$$Q_{H_1} = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{5/6 - 7}{\frac{3}{\sqrt{30}}} = -2/55$$

$$RH_1 \Leftrightarrow Q_{H_1} < Z_{\alpha} \quad -2/55 < -1/64 \Rightarrow \text{فرض صفر رد می‌شود.}$$





تمرین

- ۱- در یک آزمون آماری تفاوت فرضیه صفر و یک چیست؟
- ۲- خطای معیار برآورد و انحراف معیار متغیر چه تفاوتی دارند؟
- ۳- منظور از سطح معنی داری آزمون چیست؟
- ۴- توان یک آزمون آماری به چه معنی است؟
- ۵- روش مناسب برای کاهش خطاها و افزایش توان آزمون چیست؟
- ۶- متوسط سن اولین ریزش مو در مردانی که دچار تاسی می شوند، ۲۸ سال با واریانس ۹ سال است. اخیرا پس از یک برنامه بهداشتی در یک نمونه ۳۶ تایی از مردان مبتلا به تاسی، میانگین سن اولین ریزش مو ۳۱ سال به دست آمده است. در سطح معنی داری ۵٪ آزمون کنید آیا این برنامه موثر بوده است؟
- ۷- میانگین و انحراف معیار نمره هوشی کودکان در یک منطقه به ترتیب ۸۰ و ۷ است. یک نمونه ۲۵ تایی از کودکان این منطقه را انتخاب و مورد آزمون هوش قرار می دهیم و میانگین نمره هوشی آنها را ۸۳ به دست می آوریم.
- در سطح معنی داری ۵٪ آیا می توان پذیرفت میانگین نمره هوشی کودکان افزایش داشته است؟

۸- وزن شیشه های شربت تولیدی یک شرکت داروسازی دارای توزیع نرمال با میانگین ۸۰ و واریانس ۹ گرم است. فردی مدعی است میانگین وزن شیشه ها کاهش یافته است. در یک نمونه ۲۰ تایی از این شیشه ها،

میانگین وزن آنها ۷۸ به دست آمد. در سطح ۵٪ آیا ادعای فرد را می پذیرید یا رد می کنید؟

۹- کلسیم سرم در بیماران مبتلا به آرترید روماتوئید دارای توزیع نرمال با میانگین ۹/۹ و انحراف معیار ۶/۲ است. پژوهشگری در یک نمونه ۱۶ تایی از این بیماران کلسیم سرم آنها را ۹/۲ مشاهده کرده است. آزمون کنید آیا میانگین نمونه مطالعه شده با میانگین جمعیت برابری دارد؟

۱۰- میانگین هموگلوبین پلاسما در زنان بالغ و سالم، ۱۴ گرم با انحراف معیار ۲ گرم است.

الف- یک فاصله اطمینان ۹۹ درصد برای میانگین واقعی بسازید.

ب- در یک نمونه ۴۹ تایی $\bar{x} = 13/5$ مشاهده شده است. آیا تفاوت معنی دار است؟

۱۱- ضریب خطر بین کمبود سلینیوم و ابتلا به آرتروز زانو در خانمهای بالای ۶۰ سال برابر ۱/۳۴ به دست آمده است. اگر احتمال معنی داری $P-Value < 0.001$ باشد، آیا می پذیرید که کمبود سلینیوم بر ابتلا به آرتروز زانو در خانمهای بالای ۶۰ سال موثر نیست (فرض صفر)؟ چرا؟

۱۲- در صورتی که ضریب خطر بین کمبود سلینیوم و پوکی استخوان در خانمهای بالای ۶۰ سال برابر ۱/۱۱ به دست آمده است. اگر احتمال معنی داری $P-Value = 0.06$ باشد، آیا می پذیرید که کمبود سلینیوم بر پوکی استخوان در خانمهای بالای ۶۰ سال موثر است (فرض یک)؟ چرا؟

جدول‌ها

جدول نرمال استاندارد

www.M-Mirzadeh.Blogfa.Com

تهیه کننده: محمدرضا میرزاده

نحوه استفاده از جدول

هر بخش از جدول مقابل شامل ۳ ستون است. در ستون اول مقدار Z و در ستون دوم احتمال های کمتر از مقدار Z را مشاهده می کنید. ستون سوم اختصاص دارد به مقادیر بیشتر از Z یا کمتر از $-Z$ به عنوان مثال:

$$P(Z < 0.27) = 0.6064$$

$$P(Z > 0.27) = 0.3936 \text{ و در ستون بعدی}$$

که معادل $P(Z < -0.27)$ است.

(ادامه جدول در صفحه بعد)

Z	Cum p	Tail p	Z	Cum p	Tail p
0.00	0.5000	0.5000	0.40	0.6554	0.3446
0.01	0.5040	0.4960	0.41	0.6591	0.3409
0.02	0.5080	0.4920	0.42	0.6628	0.3372
0.03	0.5120	0.4880	0.43	0.6664	0.3336
0.04	0.5160	0.4840	0.44	0.6700	0.3300
0.05	0.5199	0.4801	0.45	0.6736	0.3264
0.06	0.5239	0.4761	0.46	0.6772	0.3228
0.07	0.5279	0.4721	0.47	0.6808	0.3192
0.08	0.5319	0.4681	0.48	0.6844	0.3156
0.09	0.5359	0.4641	0.49	0.6879	0.3121
0.10	0.5398	0.4602	0.50	0.6915	0.3085
0.11	0.5438	0.4562	0.51	0.6950	0.3050
0.12	0.5478	0.4522	0.52	0.6985	0.3015
0.13	0.5517	0.4483	0.53	0.7019	0.2981
0.14	0.5557	0.4443	0.54	0.7054	0.2946
0.15	0.5596	0.4404	0.55	0.7088	0.2912
0.16	0.5636	0.4364	0.56	0.7123	0.2877
0.17	0.5675	0.4325	0.57	0.7157	0.2843
0.18	0.5714	0.4286	0.58	0.7190	0.2810
0.19	0.5753	0.4247	0.59	0.7224	0.2776
0.20	0.5793	0.4207	0.60	0.7257	0.2743
0.21	0.5832	0.4168	0.61	0.7291	0.2709
0.22	0.5871	0.4129	0.62	0.7324	0.2676
0.23	0.5910	0.4090	0.63	0.7357	0.2643
0.24	0.5948	0.4052	0.64	0.7389	0.2611
0.25	0.5987	0.4013	0.65	0.7422	0.2578
0.26	0.6026	0.3974	0.66	0.7454	0.2546
0.27	0.6064	0.3936	0.67	0.7486	0.2514
0.28	0.6103	0.3897	0.68	0.7517	0.2483
0.29	0.6141	0.3859	0.69	0.7549	0.2451
0.30	0.6179	0.3821	0.70	0.7580	0.2420
0.31	0.6217	0.3783	0.71	0.7611	0.2389
0.32	0.6255	0.3745	0.72	0.7642	0.2358
0.33	0.6293	0.3707	0.73	0.7673	0.2327
0.34	0.6331	0.3669	0.74	0.7704	0.2296
0.35	0.6368	0.3632	0.75	0.7734	0.2266
0.36	0.6406	0.3594	0.76	0.7764	0.2236
0.37	0.6443	0.3557	0.77	0.7794	0.2206
0.38	0.6480	0.3520	0.78	0.7823	0.2177
0.39	0.6517	0.3483	0.79	0.7852	0.2148



جدول نرمال استاندارد

Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p
0.80	0.7881	0.2119	1.20	0.8849	0.1151	1.60	0.9452	0.0548
0.81	0.7910	0.2090	1.21	0.8869	0.1131	1.61	0.9463	0.0537
0.82	0.7939	0.2061	1.22	0.8888	0.1112	1.62	0.9474	0.0526
0.83	0.7967	0.2033	1.23	0.8907	0.1093	1.63	0.9484	0.0516
0.84	0.7995	0.2005	1.24	0.8925	0.1075	1.64	0.9495	0.0505
0.85	0.8023	0.1977	1.25	0.8944	0.1056	1.65	0.9505	0.0495
0.86	0.8051	0.1949	1.26	0.8962	0.1038	1.66	0.9515	0.0485
0.87	0.8078	0.1922	1.27	0.8980	0.1020	1.67	0.9525	0.0475
0.88	0.8106	0.1894	1.28	0.8997	0.1003	1.68	0.9535	0.0465
0.89	0.8133	0.1867	1.29	0.9015	0.0985	1.69	0.9545	0.0455
0.90	0.8159	0.1841	1.30	0.9032	0.0968	1.70	0.9554	0.0446
0.91	0.8186	0.1814	1.31	0.9049	0.0951	1.71	0.9564	0.0436
0.92	0.8212	0.1788	1.32	0.9066	0.0934	1.72	0.9573	0.0427
0.93	0.8238	0.1762	1.33	0.9082	0.0918	1.73	0.9582	0.0418
0.94	0.8264	0.1736	1.34	0.9099	0.0901	1.74	0.9591	0.0409
0.95	0.8289	0.1711	1.35	0.9115	0.0885	1.75	0.9599	0.0401
0.96	0.8315	0.1685	1.36	0.9131	0.0869	1.76	0.9608	0.0392
0.97	0.8340	0.1660	1.37	0.9147	0.0853	1.77	0.9616	0.0384
0.98	0.8365	0.1635	1.38	0.9162	0.0838	1.78	0.9625	0.0375
0.99	0.8389	0.1611	1.39	0.9177	0.0823	1.79	0.9633	0.0367
1.00	0.8413	0.1587	1.40	0.9192	0.0808	1.80	0.9641	0.0359
1.01	0.8438	0.1562	1.41	0.9207	0.0793	1.81	0.9649	0.0351
1.02	0.8461	0.1539	1.42	0.9222	0.0778	1.82	0.9656	0.0344
1.03	0.8485	0.1515	1.43	0.9236	0.0764	1.83	0.9664	0.0336
1.04	0.8508	0.1492	1.44	0.9251	0.0749	1.84	0.9671	0.0329
1.05	0.8531	0.1469	1.45	0.9265	0.0735	1.85	0.9678	0.0322
1.06	0.8554	0.1446	1.46	0.9279	0.0721	1.86	0.9686	0.0314
1.07	0.8577	0.1423	1.47	0.9292	0.0708	1.87	0.9693	0.0307
1.08	0.8599	0.1401	1.48	0.9306	0.0694	1.88	0.9699	0.0301
1.09	0.8621	0.1379	1.49	0.9319	0.0681	1.89	0.9706	0.0294
1.10	0.8643	0.1357	1.50	0.9332	0.0668	1.90	0.9713	0.0287
1.11	0.8665	0.1335	1.51	0.9345	0.0655	1.91	0.9719	0.0281
1.12	0.8686	0.1314	1.52	0.9357	0.0643	1.92	0.9726	0.0274
1.13	0.8708	0.1292	1.53	0.9370	0.0630	1.93	0.9732	0.0268
1.14	0.8729	0.1271	1.54	0.9382	0.0618	1.94	0.9738	0.0262
1.15	0.8749	0.1251	1.55	0.9394	0.0606	1.95	0.9744	0.0256
1.16	0.8770	0.1230	1.56	0.9406	0.0594	1.96	0.9750	0.0250
1.17	0.8790	0.1210	1.57	0.9418	0.0582	1.97	0.9756	0.0244
1.18	0.8810	0.1190	1.58	0.9429	0.0571	1.98	0.9761	0.0239
1.19	0.8830	0.1170	1.59	0.9441	0.0559	1.99	0.9767	0.0233



جدول توزیع t

نحوه استفاده از جدول

در ستون اول درجه آزادی از ۱ تا ۱۲۰ و بی نهایت آمده است.
در سطر اول بالای جدول مقادیر معنی داری (آلفا) از ۰/۴ تا ۰/۰۰۵ آمده است.
احتمالات جدول، احتمالات تجمعی و دنباله سمت راست توزیع است.
به عنوان مثال:

مقدار t با ۱۰ درجه آزادی و سطح معنی داری ۰/۰۵ برابر است با

$$t_{(10, 0.05)} = 1.812$$

ν	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.689
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.660
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291



پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

رشته‌های علوم پزشکی و شاخه‌های مرتبط

۲ واحد درسی

مطابق با سرفصل‌های مصوب

آزمون فرضی ۲

آزمون‌های میانگین

آزمون اختلاف میانگین دو جامعه

فرض کنید اطلاعات مربوط به دو جمعیت به صورت زیر در اختیار است می‌خواهیم فرضیه یکسان بودن میانگین های دو جمعیت را آزمون کنیم.

جامعه ۲

میانگین μ_2

آماره $\bar{x}_2 = \frac{1}{n_2} \sum x_i$

تعداد نمونه n_2

جامعه ۱

میانگین μ_1

آماره $\bar{x}_1 = \frac{1}{n_1} \sum x_i$

تعداد نمونه n_1

اگر فرضیه های یک پژوهش به مقایسه میانگین های یک صفت در دو جامعه اشاره داشته باشد، می‌توانیم با حفظ تمامی مفروضاتی که در آزمون های مربوط یک جمعیت مطرح شد و تنها با توجه به مساوی یا نامساوی بودن واریانس‌ها مراحل انجام آزمون را مانند قبل پیگیری کرد.

آزمون‌های میانگین

آماره آزمون بسته به اینکه واریانس‌های دو جامعه با هم مساوی فرض شوند یا نه، به دو صورت کلی زیر است:

$$Q_{H.} = \frac{\bar{X}_1 - \bar{X}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

۱- شاخص آزمون وقتی واریانس‌های دو جمعیت با هم مساوی باشند.

$$Q_{H.} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

۱- شاخص آزمون وقتی واریانس‌های دو جمعیت با هم مساوی نباشند.

در آزمون‌های مربوط به اختلاف میانگین دو جمعیت، با توجه به معلوم یا نامعلوم بودن واریانس‌ها و حجم نمونه‌ها، توزیع شاخص آزمون در نهایت منجر به یکی از توزیع‌های Z یا t خواهد شد.

آزمون‌های میانگین

مثال

یک فرضیه پژوهشی چنین مطرح شده است «آیا میانگین هموگلوبین خون مردان و زنان یکسان است؟ از تحقیقات گذشته مشخص شده است که توزیع هموگلوبین نرمال و واریانس دو جامعه مساوی و مقدار آن $\sigma^2 = 4$ است. برای این منظور یک نمونه ۱۲ تایی از مردان و یک نمونه ۸ تایی از زنان را انتخاب و بر اساس این نمونه‌ها $\bar{x}_1 = 11/5$ ، $\bar{x}_2 = 13/5$ را به دست می‌آوریم. در سطح $\alpha = 0/05$ فرض برابری هموگلوبین خون زنان و مردان را آزمون کنید.

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases} \quad \text{۱- فرضیه های آزمون را تشکیل می‌دهیم.}$$

$$Q_{H_1} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{13/5 - 11/5}{2 \sqrt{\frac{1}{12} + \frac{1}{8}}} = 2/19 \quad \begin{array}{l} \text{۲- سطح معنی داری } 5\% \text{ تعیین شده است.} \\ \text{۳- شاخص آزمون دارای توزیع نرمال است.} \\ \text{است و مقدار آن } 2/19 \text{ است.} \end{array}$$

$$RH_1 \Leftrightarrow |Q_{H_1}| > z_{1-\frac{\alpha}{2}} \quad \begin{array}{l} \text{۴- استراتژی: فرض صفر را وقتی رد می‌کنیم که مقدار شاخص} \\ \text{آزمون از مقدار } Z \text{ جدول بیشتر باشد.} \end{array}$$

تصمیم: چون مقدار شاخص آزمون ۲/۱۹ از $z_{0/975} = 1/96$ بیشتر است؛ پس فرض صفر رد می‌شود. یعنی اختلاف قابل ملاحظه‌ای بین میانگین هموگلوبین خون مردان و زنان وجود دارد.



آزمون‌های میانگین

مثال: یکی از فرضیه‌های تحقیق مربوط به برابری فشار خون سیستولیک مردان و زنان است. با توجه به نرمال بودن توزیع فشار خون و اطلاعات جدول مقابل و با فرض نامساوی بودن واریانس‌ها، در سطح معنی داری ۵٪ فرضیه فوق را آزمون کنید.

	n	$\bar{x}$	s^2
مردان	۱۱	۱۲۵	۱۲۱
زنان	۱۰	۱۳۲	۲۲۵

حل: باتوجه به حجم کم نمونه‌ها و نامعلوم بودن واریانس‌ها، شاخص آزمون دارای توزیع t با $n_1 + n_2 - 2$ درجه آزادی است پس داریم:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

$$Q_{H_0} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{125 - 132}{\sqrt{\frac{121}{11} + \frac{225}{10}}} = -1/31$$

$$RH_0 \Leftrightarrow |Q_{H_0}| > t_{1-\alpha/2, 19}$$

$$|-1/31| > 2/09$$

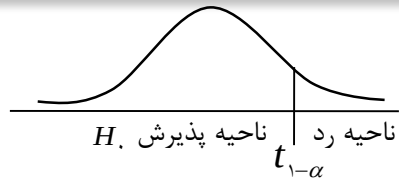
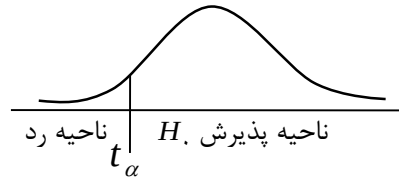
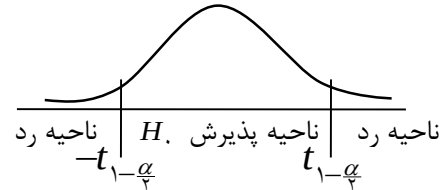
فرض صفر رد نمی‌شود یعنی فشار خون سیستولیک در زنان و مردان برابر است.



آزمون‌های میانگین

آزمون t یک نمونه ای

اگر توزیع جامعه آماری نرمال باشد هنگامی که تعداد نمونه ها کمتر از ۳۰ و واریانس جامعه نا معلوم است، شاخص آماری آزمون دارای توزیع t است. منحنی این توزیع مشابه توزیع نرمال و مشخصه این توزیع درجه آزادی آن است که با $n - 1$ معین می شود.

نوع آزمون	شاخص	توزیع	استراتژی	ناحیه رد و پذیرش
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} > t_{1-\alpha}$	
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} < t_{\alpha}$	
$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$	$Q_{H_0} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$	نرمال	$RH_0 \Leftrightarrow Q_{H_0} > t_{1-\frac{\alpha}{2}}$	

آزمون‌های میانگین

مثال

تعداد دفعات ابتلا به آنفولانزا در یک جمعیت دارای توزیع نرمال با میانگین ۴ بار در یک سال است. محقق می‌خواهد بداند پس از اجرای یک طرح واکسیناسیون، میانگین تعداد دفعات ابتلا در سال کاهش داشته است؟ او در یک نمونه ۱۶ تایی از افراد، میانگین تعداد دفعات ابتلا را ۳/۴ با انحراف معیار ۲ به دست می‌آورد. در سطح معنی داری ۰.۰۵ آیا واکسن فوق در پیشگیری موثر بوده است؟

حل:

$$\begin{cases} H_0: \mu = 4 \\ H_1: \mu < 4 \end{cases} \quad \text{۱- فرضیه ها را تشکیل می دهیم.}$$

$$Q_{H_1} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{3/4 - 4}{\frac{2}{\sqrt{16}}} = -1/2 \quad \begin{array}{l} \text{۲- سطح معنی داری } 0.05 \text{ تعیین شده است.} \\ \text{۳- شاخص آماری دارای توزیع } t \text{ با } 15 \text{ درجه آزادی} \\ \text{است و مقدار آن } 1/2 \text{ است.} \end{array}$$

$$RH_1 \Leftrightarrow Q_{H_1} < -t_{n-1, \alpha} \quad \begin{array}{l} \text{۴- استراتژی: فرض صفر را وقتی رد می کنیم که مقدار} \\ \text{شاخص آزمون از مقدار } t \text{ جدول کمتر باشد.} \end{array}$$

تصمیم: چون مقدار شاخص آزمون $-1/2$ از $-t_{15, 0.05} = -1.753$ کمتر نیست؛ فرض صفر رد نمی‌شود یعنی واکسن اثر قابل ملاحظه‌ای در پیشگیری از آنفلانزا نداشته است.

چون شاخص آزمون دارای توزیع t است، مقدار t را با توجه به تعداد نمونه و سطح معنی داری از جدول توزیع t به دست آورده‌ایم



آزمون‌های میانگین

آزمون t زوجی

وقتی می‌خواهید یک شرکت کننده در آزمون را در دو حالت مورد آزمایش قرار دهید، مانند هنگامی که می‌خواهید فشار خون افراد را در دو حالت ایستاده و نشسته مقایسه کنید، با اندازه‌های تکراری رو برو هستید. چنین آزمایشاتی نمونه‌های وابسته از داده‌ها را در اختیار می‌گذارند که به آنها داده‌های زوج شده می‌گوییم. این گونه آزمایش‌ها به طور معمول دارای دو متغیر هستند. یک متغیر مربوط به حالت اول از یک نمونه و متغیر دیگر مربوط به حالت دوم از همان نمونه است. برای آزمودن فرض برابری مقادیر متغیر در دو حالت، از آزمون t زوجی استفاده می‌کنیم.

شاخص آماری آزمون: آماره این آزمون مبتنی بر متغیر d_i است که از اختلافات زوج‌ها و به صورت زیر به دست می‌آید.

$$d_i = x_i - y_i$$

$$Q_{H_0} = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

اینک متغیر d_i متغیر تصادفی و شاخص آزمون t زوجی به صورت زیر خواهد بود:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i \quad \text{و} \quad s_d = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_i - \bar{d})^2} \quad \text{است.}$$

آزمون‌های میانگین

توزیع شاخص آزمون: در صورتی که حجم نمونه به اندازه کافی بزرگ باشد، شاخص آزمون دارای توزیع نرمال و در صورتی که تعداد نمونه کم و واریانس مجهول باشد، شاخص آماری آزمون دارای توزیع t با $n - 1$ درجه آزادی است. (n تعداد نمونه انتخابی است).

فرضیه های آزمون: از آنجا که از هر نمونه دو متغیر اندازه گیری می شود، اساس آزمون بر اختلاف بین مقادیر زوج شده است پس فرضیه ها را به صورت زیر می نویسیم:

$$\begin{cases} H_0: \bar{d} = 0 \\ H_1: \bar{d} \neq 0 \end{cases} \quad \text{یا به طور معادل} \quad \begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

استراتژی رد صفر: فرض صفر را وقتی رد می کنیم که مقدار شاخص آماری آزمون از مقدار t جدول، در سطح معنی داری α بیشتر باشد.

$$RH_0 \Leftrightarrow |Q_{H_0}| > t_{n-1, 1-\frac{\alpha}{2}}$$

آزمون‌های میانگین

مثال

محقق می‌خواهد تاثیر یک داروی فشار خون را روی بیماران مبتلا به فشار خون سیستولیک آزمون کند. او ۱۲ بیمار را به تصادف انتخاب کرده و فشار خون آنها را قبل و بعد از مصرف دارو اندازه گرفته است. در سطح ۵٪ تاثیر دارو را آزمون کنید.

ردیف	قبل	بعد	d_i	d_i^2
۱	۱۴۰	۱۱۰	۳۰	۹۰۰
۲	۱۵۰	۱۴۰	۱۰	۱۰۰
۳	۱۸۰	۱۹۰	-۱۰	۱۰۰
۴	۱۶۰	۱۷۰	-۱۰	۱۰۰
۵	۱۴۰	۱۱۰	۳۰	۹۰۰
۶	۱۴۰	۱۲۰	۲۰	۴۰۰
۷	۱۵۰	۱۱۰	۴۰	۱۶۰۰
۸	۱۷۰	۱۹۰	-۲۰	۴۰۰
۹	۱۵۰	۱۵۰	۰	۰
۱۰	۱۶۰	۱۴۰	۲۰	۴۰۰
۱۱	۱۷۰	۱۳۰	۴۰	۱۶۰۰
۱۲	۱۷۰	۱۴۰	۳۰	۹۰۰
جمع			۱۸۰	۷۴۰۰

$$\begin{cases} H_0: \bar{d} = 0 \\ H_1: \bar{d} \neq 0 \end{cases}$$

$$\bar{d} = \frac{180}{12} = 15 \quad s_d^2 = \left(\frac{1}{n} \sum_{i=1}^n d_i^2 \right) - \bar{d}^2 = 392$$

$$Q_{H_0} = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{15}{\frac{19.8}{\sqrt{12}}} = 2/6$$

تصمیم: با توجه به اینکه مقدار شاخص

آزمون از $t_{11, 0.975} = 2/2$ جدول بزرگتر است، فرض صفر رد می‌شود. یعنی داروی فوق در کاهش فشار خون موثر بوده است.



آزمون‌های نسبت

آزمون نسبت یک نمونه ای

اگر فرضیه‌ای در خصوص نسبت مجهولی در یک یا دو جامعه مطرح باشد، مانند:

- نسبت افراد سیگاری در یک جمعیت ۱۸ به ۱۰۰۰ نفر است.
 - میزان شیوع بیماری A در بین زنان یک منطقه افزایش داشته است.
 - ضریب هوشی در بین زنان و مردان به یک اندازه است.
- می‌توان آزمون‌های مربوط به نسبت را با انجام مراحل اشاره شده در قبل انجام داد.

چون هنگام برآورد کردن هر نسبتی، باید حجم نمونه به اندازه کافی بزرگ باشد؛ بنابراین در یک آزمون فرض مربوط به نسبت، شاخص آزمون طبق قضیه حد مرکزی همواره دارای توزیع نرمال (Z) خواهد بود.

$$Q_{H.} = \frac{\bar{p} - p.}{\sqrt{\frac{\bar{p} \cdot \bar{q}}{n}}}$$

شاخص آزمون نسبت برای یک جمعیت

اگر $\bar{p}$ نسبت برآورد شده در یک جمعیت باشد، داریم:
 $1 - \bar{p} = \bar{q}$

آزمون‌های نسبت

مثال

میزان شیوع آسم در بین زنان در یک منطقه ۱۵٪ بوده است. مرکز بهداشت در یک نمونه ۲۱۵ نفری از زنان دریافته است که ۳۹ نفر آنها سابقه آسم دارند. آیا بر اساس این نمونه می‌توان گفت میزان شیوع آسم در بین زنان این منطقه تغییر کرده است؟

حل:

$$\begin{cases} H_0: P = P_0 \\ H_1: P \neq P_0 \end{cases} \Rightarrow \begin{cases} H_0: P = 0.15 \\ H_1: P \neq 0.15 \end{cases}$$

$$\bar{p} = \frac{x}{n} = \frac{39}{215} = 0.18$$

$$Q_{H_1} = \frac{\bar{p} - p_0}{\sqrt{\frac{\bar{p} \cdot \bar{q}}{n}}} = \frac{0.18 - 0.15}{\sqrt{\frac{0.18 \times 0.82}{215}}} = 1.15 \quad RH_1 \Leftrightarrow |Q_{H_1}| > Z_{0.975}$$
$$|1.15| < 1.96$$

فرض صفر رد نمی‌شود یعنی میزان شیوع آسم در بین زنان تغییر نداشته است.



آزمون‌های نسبت

آزمون مقایسه دو نسبت

چنانچه P_1 و P_2 به ترتیب معرف نسبت صفت مورد مطالعه در جامعه اول و جامعه دوم باشند، آزمونی که به منظور مقایسه این دو نسبت صورت می گیرد یک آزمون دو دامنه بوده و فرضیه های آزمون به صورت زیر خواهد بود:

$$\begin{cases} H_0 : P_1 = P_2 \\ H_1 : P_1 \neq P_2 \end{cases}$$

$$Q_{H_0} = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{\bar{p}_1 \cdot \bar{q}_1}{n_1} + \frac{\bar{p}_2 \cdot \bar{q}_2}{n_2}}}$$

شاخص آزمون مقایسه نسبت برای دو جمعیت

استراتژی رد فرض صفر به صورت زیر است:

$$RH_0 \Leftrightarrow |Q_{H_0}| > z_{1-\frac{\alpha}{2}}$$

آزمون‌های نسبت

مثال

در یک کارآزمایی بالینی، ۲۰ نفر از ۲۴۰ نفری که واکسن آنفلانزا دریافت کرده اند و ۸۰ نفر از ۲۲۰ نفری که شبه واکسن دریافت کرده اند، مبتلا شده‌اند. آیا می‌توان گفت نسبت ابتلا در دو گروه با هم یکسان است؟ (تاثیر واکسن قابل ملاحظه نیست).

حل:

$$\begin{cases} H_0: P_1 = P_2 \\ H_1: P_1 \neq P_2 \end{cases}$$

$$\bar{p}_1 = \frac{x_1}{n_1} = \frac{20}{240} = 0.083 \Rightarrow \bar{q}_1 = 0.917$$

$$\bar{p}_2 = \frac{x_2}{n_2} = \frac{80}{220} = 0.36 \Rightarrow \bar{q}_2 = 0.64$$

$$Q_{H_0} = \frac{\bar{p}_1 - \bar{p}_2}{\sqrt{\frac{\bar{p}_1 \cdot \bar{q}_1}{n_1} + \frac{\bar{p}_2 \cdot \bar{q}_2}{n_2}}} = \frac{0.083 - 0.36}{\sqrt{\frac{0.083 \times 0.917}{240} + \frac{0.36 \times 0.64}{220}}} = -7/6$$

$RH_0 \Leftrightarrow |Q_{H_0}| > Z_{0.975}$
 $| -7/6 | > 1/96$

دلیل چنین تصمیمی به خاطر منفی شدن Q_{H_0} است. در صورتی که اگر همین مقدار مثبت باشد، نتیجه کاملاً بر عکس خواهد شد.

فرض صفر رد می‌شود. یعنی واکسن، میزان شیوع آنفلانزا را به طور قابل ملاحظه‌ای کاهش می‌دهد.



تمرین

تمرین

۱- میانگین و انحراف معیار نمره هوشی ۲۰ نفر از کودکان منطقه A به ترتیب ۸۰ و ۷ است. از بین کودکان منطقه B یک نمونه ۲۵ تایی انتخاب کرده و میانگین نمره هوشی آنان را ۷۸ به دست می آوریم. با فرض مساوی بودن واریانس نمره هوشی در دو منطقه و در سطح ۰.۰۵٪، آیا می توان فرضیه تاثیر منطقه را روی نتیجه آزمایش هوش مردود شناخت؟

۲- به منظور مقایسه میزان اثر آنتی بیوتیک یک مشتق جدید با آموکسی سیلین در ۸ محیط کشت از این مشتق و در ۱۰ محیط دیگر آموکسی سیلین قرار داده شد و هاله عدم رشد آنها تعیین گردید. در گروه مشتق جدید هاله ۱۸ میلی متر و در آموکسی سیلین ۱۱ میلی متر به دست آمد. با فرض نرمال بودن دو صفت و تساوی واریانسها (۴) در سطح ۰.۰۵٪ آیا تفاوت معنی داری در میزان اثر این دو آنتی بیوتیک وجود دارد؟

۳- در یک جمعیت میانگین کلرید سرم ۱۰۰ میلی اکی والان با انحراف معیار ۳ می باشد.

الف- چه نسبتی بیشتر از ۱۰۳ میلی و کمتر از ۹۷ میلی هستند؟

ب- در یک نمونه ۳۶ تایی میانگین کلرید سرم $\bar{x} = 95$ مشاهده شده، آیا فرض برابری میانگین این نمونه با میانگین جمعیت را می توان پذیرفت؟

- ۴- یک متخصص پوست ادعا می کند روشی ترکیبی برای درمان عفونت های پوستی ابداع کرده است که نسبت بهبودی را از ۵۰٪ قبل تا ۸۰٪ افزایش می دهد. او از ۲۵ بیمار پوستی ۲۰ نفر آنها را با روش ابداعی خود مداوا کرده است. آیا ادعای او را می توان در سطح ۵٪ پذیرفت.
- ۵- ۱۳ درصد از نوزادانی که مادران آنها در طول بارداری قرص آهن مصرف می کرده اند، دچار زردی نوزادی شدند. و در مقابل ۱۰ درصد از نوزادان مادرانی که قرص آهن مصرف نکرده اند نیز دچار این عارضه شده اند. آیا قرص آهن می تواند یک عامل خطر بروز زردی در نوزادان تلقی شود؟ (با ۵٪ خطای نوع اول)
- ۶- میزان شیوع آسم در بین زنان یک منطقه ۱۵ درصد بوده است. مرکز بهداشت در یک نمونه ۲۱۵ تایی از زنان دریافت کرده است که ۳۹ نفر آنها سابقه آسم دارند. آیا بر اساس این نمونه می توان گفت شیوع آسم در بین زنان این منطقه تغییر یافته است؟ (آلفا ۵ درصد)

تمرین

۷- نمره هوشی ۱۵ کودک را قبل و بعد از یک برنامه آموزشی در جدول زیر مشاهده می کنید. در سطح ۰.۵٪، آیا می توان پذیرفت برنامه آموزشی روی هوش کودکان تاثیر مثبت داشته است؟

قبل	۶۷	۷۴	۸۱	۷۴	۸۰	۷۸	۷۴	۶۹	۸۲	۸۱	۶۹	۸۸	۷۵	۶۸	۷۰
بعد	۸۰	۷۸	۹۰	۷۱	۸۵	۶۹	۷۷	۷۱	۸۰	۸۸	۸۰	۸۷	۸۰	۷۰	۸۰

۸- به منظور بررسی اثر یک داروی فشار خون ۱۰ نفر از بیماران مبتلا به فشار خون را به تصادف انتخاب و فشار خون آنها را قبل و بعد از مصرف دارو اندازه گرفته ایم. می دانیم توزیع صفت، توزیع نرمال است. آیا این دارو در کاهش فشار خون موثر است؟ (سطح معنی داری ۰.۵٪)

فشار خون قبل از مصرف دارو	۱۵	۱۴/۴	۱۴/۱	۱۵/۲	۱۷/۴	۱۵/۲	۱۵/۲	۱۶/۱	۱۴/۵	۱۵	۱۴/۸
فشار خون بعد از مصرف دارو	۱۴/۷	۱۲	۱۲/۳	۱۳/۵	۱۵	۱۴/۱	۱۲/۸	۱۱/۹	۱۳/۱	۱۲	۱۲

۹- در یک نمونه ۱۵۰۰ تایی از ساکنین یک منطقه، تعداد مسلولین ۳۱ نفر گزارش شده است. این نسبت قبلا ۱۲ در ۱۰۰۰ نفر بوده است. آیا این اطلاعات، افزایش میزان شیوع سل را تایید می کنند؟

جدول‌ها

جدول نرمال استاندارد

نحوه استفاده از جدول

هر بخش از جدول مقابل شامل ۳ ستون است. در ستون اول مقدار Z و در ستون دوم احتمال های کمتر از مقدار Z را مشاهده می کنید. ستون سوم اختصاص دارد به مقادیر بیشتر از Z یا کمتر از $-Z$ به عنوان مثال:

$$P(Z < 0.27) = 0.6064$$

$$P(Z > 0.27) = 0.3936 \text{ و در ستون بعدی}$$

که معادل $P(Z < -0.27)$ است.

(ادامه جدول در صفحه بعد)

Z	Cum p	Tail p	Z	Cum p	Tail p
0.00	0.5000	0.5000	0.40	0.6554	0.3446
0.01	0.5040	0.4960	0.41	0.6591	0.3409
0.02	0.5080	0.4920	0.42	0.6628	0.3372
0.03	0.5120	0.4880	0.43	0.6664	0.3336
0.04	0.5160	0.4840	0.44	0.6700	0.3300
0.05	0.5199	0.4801	0.45	0.6736	0.3264
0.06	0.5239	0.4761	0.46	0.6772	0.3228
0.07	0.5279	0.4721	0.47	0.6808	0.3192
0.08	0.5319	0.4681	0.48	0.6844	0.3156
0.09	0.5359	0.4641	0.49	0.6879	0.3121
0.10	0.5398	0.4602	0.50	0.6915	0.3085
0.11	0.5438	0.4562	0.51	0.6950	0.3050
0.12	0.5478	0.4522	0.52	0.6985	0.3015
0.13	0.5517	0.4483	0.53	0.7019	0.2981
0.14	0.5557	0.4443	0.54	0.7054	0.2946
0.15	0.5596	0.4404	0.55	0.7088	0.2912
0.16	0.5636	0.4364	0.56	0.7123	0.2877
0.17	0.5675	0.4325	0.57	0.7157	0.2843
0.18	0.5714	0.4286	0.58	0.7190	0.2810
0.19	0.5753	0.4247	0.59	0.7224	0.2776
0.20	0.5793	0.4207	0.60	0.7257	0.2743
0.21	0.5832	0.4168	0.61	0.7291	0.2709
0.22	0.5871	0.4129	0.62	0.7324	0.2676
0.23	0.5910	0.4090	0.63	0.7357	0.2643
0.24	0.5948	0.4052	0.64	0.7389	0.2611
0.25	0.5987	0.4013	0.65	0.7422	0.2578
0.26	0.6026	0.3974	0.66	0.7454	0.2546
0.27	0.6064	0.3936	0.67	0.7486	0.2514
0.28	0.6103	0.3897	0.68	0.7517	0.2483
0.29	0.6141	0.3859	0.69	0.7549	0.2451
0.30	0.6179	0.3821	0.70	0.7580	0.2420
0.31	0.6217	0.3783	0.71	0.7611	0.2389
0.32	0.6255	0.3745	0.72	0.7642	0.2358
0.33	0.6293	0.3707	0.73	0.7673	0.2327
0.34	0.6331	0.3669	0.74	0.7704	0.2296
0.35	0.6368	0.3632	0.75	0.7734	0.2266
0.36	0.6406	0.3594	0.76	0.7764	0.2236
0.37	0.6443	0.3557	0.77	0.7794	0.2206
0.38	0.6480	0.3520	0.78	0.7823	0.2177
0.39	0.6517	0.3483	0.79	0.7852	0.2148

جدول نرمال استاندارد

Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p
0.80	0.7881	0.2119	1.20	0.8849	0.1151	1.60	0.9452	0.0548
0.81	0.7910	0.2090	1.21	0.8869	0.1131	1.61	0.9463	0.0537
0.82	0.7939	0.2061	1.22	0.8888	0.1112	1.62	0.9474	0.0526
0.83	0.7967	0.2033	1.23	0.8907	0.1093	1.63	0.9484	0.0516
0.84	0.7995	0.2005	1.24	0.8925	0.1075	1.64	0.9495	0.0505
0.85	0.8023	0.1977	1.25	0.8944	0.1056	1.65	0.9505	0.0495
0.86	0.8051	0.1949	1.26	0.8962	0.1038	1.66	0.9515	0.0485
0.87	0.8078	0.1922	1.27	0.8980	0.1020	1.67	0.9525	0.0475
0.88	0.8106	0.1894	1.28	0.8997	0.1003	1.68	0.9535	0.0465
0.89	0.8133	0.1867	1.29	0.9015	0.0985	1.69	0.9545	0.0455
0.90	0.8159	0.1841	1.30	0.9032	0.0968	1.70	0.9554	0.0446
0.91	0.8186	0.1814	1.31	0.9049	0.0951	1.71	0.9564	0.0436
0.92	0.8212	0.1788	1.32	0.9066	0.0934	1.72	0.9573	0.0427
0.93	0.8238	0.1762	1.33	0.9082	0.0918	1.73	0.9582	0.0418
0.94	0.8264	0.1736	1.34	0.9099	0.0901	1.74	0.9591	0.0409
0.95	0.8289	0.1711	1.35	0.9115	0.0885	1.75	0.9599	0.0401
0.96	0.8315	0.1685	1.36	0.9131	0.0869	1.76	0.9608	0.0392
0.97	0.8340	0.1660	1.37	0.9147	0.0853	1.77	0.9616	0.0384
0.98	0.8365	0.1635	1.38	0.9162	0.0838	1.78	0.9625	0.0375
0.99	0.8389	0.1611	1.39	0.9177	0.0823	1.79	0.9633	0.0367
1.00	0.8413	0.1587	1.40	0.9192	0.0808	1.80	0.9641	0.0359
1.01	0.8438	0.1562	1.41	0.9207	0.0793	1.81	0.9649	0.0351
1.02	0.8461	0.1539	1.42	0.9222	0.0778	1.82	0.9656	0.0344
1.03	0.8485	0.1515	1.43	0.9236	0.0764	1.83	0.9664	0.0336
1.04	0.8508	0.1492	1.44	0.9251	0.0749	1.84	0.9671	0.0329
1.05	0.8531	0.1469	1.45	0.9265	0.0735	1.85	0.9678	0.0322
1.06	0.8554	0.1446	1.46	0.9279	0.0721	1.86	0.9686	0.0314
1.07	0.8577	0.1423	1.47	0.9292	0.0708	1.87	0.9693	0.0307
1.08	0.8599	0.1401	1.48	0.9306	0.0694	1.88	0.9699	0.0301
1.09	0.8621	0.1379	1.49	0.9319	0.0681	1.89	0.9706	0.0294
1.10	0.8643	0.1357	1.50	0.9332	0.0668	1.90	0.9713	0.0287
1.11	0.8665	0.1335	1.51	0.9345	0.0655	1.91	0.9719	0.0281
1.12	0.8686	0.1314	1.52	0.9357	0.0643	1.92	0.9726	0.0274
1.13	0.8708	0.1292	1.53	0.9370	0.0630	1.93	0.9732	0.0268
1.14	0.8729	0.1271	1.54	0.9382	0.0618	1.94	0.9738	0.0262
1.15	0.8749	0.1251	1.55	0.9394	0.0606	1.95	0.9744	0.0256
1.16	0.8770	0.1230	1.56	0.9406	0.0594	1.96	0.9750	0.0250
1.17	0.8790	0.1210	1.57	0.9418	0.0582	1.97	0.9756	0.0244
1.18	0.8810	0.1190	1.58	0.9429	0.0571	1.98	0.9761	0.0239
1.19	0.8830	0.1170	1.59	0.9441	0.0559	1.99	0.9767	0.0233

جدول توزیع t

نحوه استفاده از جدول

در ستون اول درجه آزادی از ۱ تا ۱۲۰ و بی نهایت آمده است.
در سطر اول بالای جدول مقادیر معنی داری (آلفا) از ۰/۴ تا ۰/۰۰۵ آمده است.
احتمالات جدول، احتمالات تجمعی و دنباله سمت راست توزیع است.
به عنوان مثال:

مقدار t با ۱۰ درجه آزادی و سطح معنی داری ۰/۰۵ برابر است با

$$t_{(10, 0.05)} = 1.812$$

ν	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.689
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.660
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

رشته‌های علوم پزشکی و شاخه‌های مرتبط

۲ واحد درسی

مطابق با سرفصل‌های مصوب

آزمون های نایارامتری

آزمونهای کی دو

χ^2 Chi-Square

آزمون همتوزیعی
کی دو

در بسیاری از تحقیقات، محقق علاقه‌مند به شمارش افراد، اشیا و یا پاسخ‌هایی است که در طبقه‌های خاصی قرار می‌گیرند. مثلاً تعدادی از بیماران را می‌توان بر حسب گروه خونی به ۴ دسته A، O، B و AB طبقه‌بندی کرد. یا مثلاً ممکن است افراد مختلف را بر حسب این که آیا "موافق"، "بی تفاوت" و یا "مخالف" کار بیرون از منزل زنان هستند، دسته‌بندی کرد. در این حال محقق می‌تواند این فرضیه را که پاسخ‌های ذکر شده از نظر فراوانی با یکدیگر تفاوت معنی‌داری خواهند داشت، آزمون کند. در این آزمون باید تعداد طبقه‌ها ۳ یا بیشتر باشد. این آزمون از جمله آزمون‌های تطابق توزیع است.

در حالتی که تنها دو طبقه وجود دارد، آزمون دو جمله‌ای (مقایسه نسبت‌ها) معادل این آزمون است.

این آزمون نسبت به حجم نمونه حساس است و هر چه حجم نمونه بیشتر باشد، توان آزمون بیشتر است. بنابراین حجم نمونه باید از ۵۰ بیشتر باشد.

آزمونهای کی دو

فرضیه ها:

در این آزمون فرضیه ها به صورت زیر مطرح هستند:

$$\begin{cases} H_0 : & \text{تفاوتی بین فراوانی طبقه ها وجود ندارد.} \\ H_1 : & \text{حداقل دو طبقه از نظر فراوانی متفاوت اند.} \end{cases}$$

آماره آزمون:

$$Q_{H_0} = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

آماره آزمون کی دو به این صورت است:

که در آن دو نوع فراوانی مد نظر قرار گرفته و با یکدیگر مقایسه می شود.

فراوانی مورد انتظار E_i

فراوانی مشاهده شده O_i

آزمونهای کی دو

توزیع آماره آزمون:

آماره این آزمون دارای توزیع کی دو با $n - 1$ درجه آزادی است که مقادیر آن در سطح معنی داری α در جدولی به نام جدول توزیع کی دو آمده است. (n تعداد طبقات است).

استراتژی رد فرض صفر:

هرچه مقادیر مشاهده شده دور از انتظار باشند، مقدار این آماره بزرگ و در نتیجه فرض صفر به مخاطره می افتد و بر عکس تفاوت کمتر بین مقادیر مشاهده شده و مقادیر مورد انتظار، آماره این آزمون را کوچک کرده و فرض صفر را نمی توان رد کرد.

استراتژی رد فرض صفر به این صورت است.

$$RH. \Leftrightarrow Q_{H.} > \chi^2_{(\alpha, n-1)}$$

مقدار آماره آزمون را محاسبه کرده و در صورتی که مقدار آن از مقدار جدول کی دو بیشتر باشد فرض صفر را رد می کنیم. در غیر این صورت فرض صفر را نمی توان رد کرد.

آزمونهای کی دو

مثال

یکسان بودن افراد در چهار گروه خونی مختلف موضوع مورد بررسی یک محقق است. او ۱۰۰ نفر را انتخاب و گروههای خونی آنان را به صورت جدول زیر ثبت کرده است.

AB	B	O	A	گروه خونی
۱۸	۲۱	۳۸	۲۳	مقادیر مشاهده شده

سوال این است که آیا تعداد گروه خونی در این جمعیت مساوی است؟

حل: فرضیه های آزمون به صورت زیر است:

$\begin{cases} H_0 : \text{تفاوتی بین فراوانی گروههای خونی وجود ندارد.} \\ H_1 : \text{حداقل دو گروه خونی از نظر فراوانی متفاوت اند.} \end{cases}$

آزمونهای کی دو

اگر قرار باشد گروههای خونی در این جمعیت به طور یکسان توزیع شده باشند باید در بین ۱۰۰ نفر در هر گروه خونی ۲۵ نفر قرار گیرند که مقدار مورد انتظار در هر گروه خونی است.

AB	B	O	A	گروه خونی
۱۸	۲۱	۳۸	۲۳	مقادیر مشاهده شده O_i
۲۵	۲۵	۲۵	۲۵	مقادیر مورد انتظار E_i
$\frac{(18 - 25)^2}{25} = 1/96$	$\frac{(21 - 25)^2}{25} = 0/64$	$\frac{(38 - 25)^2}{25} = 6/76$	$\frac{(23 - 25)^2}{25} = 0/16$	مقادیر آماره $\frac{(O_i - E_i)^2}{E_i}$

$$Q_{H.} = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = 1/96 + 0/64 + 6/76 + 0/16 = 9/52 \quad \text{مقدار آماره کی دو}$$

$$RH. \Leftrightarrow Q_{H.} > \chi^2_{(0/05, 3)} \quad \text{استراتژی رد فرض صفر به صورت}$$

مقدار کی دو در جدول با ۳ درجه آزادی و سطح معنی داری ۰/۰۵ برابر ۷/۸۲ است. بنابراین فرض صفر رد می شود. یعنی توزیع گروههای خونی در این جمعیت یکسان نمی باشد.



آزمونهای کی دو

آزمون استقلال کی دو

- جدولی را که در آن داده های آماری در k سطر و L ستون قرار دارند، یک جدول توافقی است.
- در این جدول می توان تعداد $k \times l$ ترکیب مختلف از مقادیر را مشاهده کرد.
- در چنین جدولی می توان تأثیر یک متغیر کیفی (اسمی یا ترتیبی) را بر روی یک متغیر کیفی دیگر بررسی کرد.

صفت
B

	B_1	B_2	...	B_l
A_1	d_{11}	d_{12}	...	d_{1l}
A_2	d_{21}	d_{22}	...	d_{2l}
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
A_k	d_{k1}	d_{k2}	...	d_{kl}

صفت A

آزمونهای کی دو

شرایط آزمون استقلال کی دو در جدول توافقی

- ۱- فراوانی مورد انتظار در هیچ طبقه ای نباید صفر باشد.
- ۲- جمع فراوانی طبقاتی که فراوانی آنها کمتر از ۵ است، نباید بیشتر از ۲۰٪ فراوانی کل باشد.
- ۳- عضویت افراد در هر دو گروه با هم همپوشانی نداشته باشد. یعنی هیچ فرد یا موردی نتواند در بیشتر از یک ترکیب از طبقات قرار گیرد.

در بحث ارتباط بین دو صفت کیفی می توان دو حالت برای انتخاب افراد در نظر گرفت

روش دوم: ابتدا جامعه را از نظر یکی از صفت های مورد بررسی به گروهای تقسیم کرده و سپس از هر گروه به تصادف تعدادی از افراد را انتخاب کرد.

روش اول: ابتدا افراد را به صورت تصادفی انتخاب کرد و سپس آنها را بر حسب دو صفت مورد بررسی در ترکیب های مختلف جدول قرار داد.

آزمونهای کی دو

مثال حالت اول :

ابتدا تعداد ۱۰۰ نفر از مردان و زنان را به تصادف انتخاب می‌کنیم. سپس آنها را بر حسب اینکه چه جنسیتی دارند و در چه گروه خونی قرار دارند، در هر یک از ترکیب‌های جدول به صورت زیر می‌نویسیم: 2×4

	O	AB	B	A
زن	۱۴	۱۵	۲۱	۲۰
مرد	۷	۱۰	۵	۸

مثال حالت دوم :

ابتدا تعداد ۷۰ زن به تصادف از بین زنان و ۳۰ مرد را به تصادف از بین مردان انتخاب می‌کنیم. سپس از آنها بر حسب اینکه در چه گروه خونی قرار دارند، سوال کرده و در هر یک از ترکیب‌های جدول می‌نویسیم:

آزمونهای مختلفی برای تعیین رابطه بین دو متغیر وجود دارد که یکی از این آزمون‌ها که برای تعیین رابطه بین متغیرهای کیفی مناسب است، آزمون کی دو است. البته می‌توان یکی از متغیرها را ترتیبی یا کمی طبقه‌بندی شده در نظر گرفت.

آزمونهای کی دو

شاخص آماری آزمون: در آزمون کی دو فراوانی‌های مشاهده شده (O_{ij}) با فراوانی‌های مورد انتظار (E_{ij}) با یکدیگر مقایسه می‌گردد و در نهایت شاخص آماری آزمون کی دو (پیرسن) به صورت زیر به دست می‌آید:

$$Q_H = \sum_{j=1}^l \sum_{i=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

که در این رابطه فراوانی مشاهده شده O_{ij} ، فراوانی هر ترکیب از خانه‌های جدول است و فراوانی‌های مورد انتظار E_{ij} را بر اساس قانون احتمالات به صورت زیر محاسبه می‌کنیم.

$$E_{ij} = \frac{O_{i.} \times O_{.j}}{n} \quad \text{که در آن } O_{i.} \text{ جمع فراوانی های سطر } i \text{ ام و } O_{.j} \text{ جمع فراوانی های ستون } j \text{ ام است.}$$

	O	AB	B	A	جمع
زن	۱۴	۱۵	۲۱	۲۰	۷۰
مرد	۷	۱۰	۵	۸	۳۰
جمع	۲۱	۲۵	۲۶	۲۸	۱۰۰

مثلاً: مقدار مورد انتظار برای زنان گروه خونی O به صورت زیر است:

$$E_{11} = \frac{O_{1.} \times O_{.1}}{n} = \frac{70 \times 21}{100} = 14.7$$

آزمونهای کی دو

فرضیه های آزمون: در آزمون استقلال کی دو می خواهیم این فرضیه را که بین دو متغیر ارتباطی وجود ندارد، آزمون کنیم. پس می توان فرضیه صفر و فرض مقابل آن را به صورت زیر نوشت:

$$\begin{cases} H_0 : & \text{رابطه ای بین دو متغیر وجود ندارد.} \\ H_1 : & \text{بین دو متغیر رابطه وجود دارد.} \end{cases}$$

توزیع شاخص آزمون: شاخص آماری آزمون دارای توزیع کی دو با $(k-1)(l-1)$ درجه آزادی است.

استراتژی آزمون: فرضیه صفر وقتی رد می شود که مقدار شاخص آزمون از مقدار کی دو جدول بیشتر باشد. در این صورت می گوئیم دو متغیر بر یکدیگر تاثیرگذارند.

$$RH_0 \Leftrightarrow Q_{H_0} > \chi^2_{(1-\alpha, (k-1)(l-1))}$$

آزمونهای کی دو

تعیین مقدار همبستگی در جدول توافقی

اگر در یک جدول توافقی، فرض صفر رد شده و وجود رابطه بین دو صفت محرز است، می توان شدت این رابطه را به کمک ضریبی به نام **ضریب توافق چوپروف** تعیین کرد. این ضریب به صورت زیر محاسبه می شود:

$$\rho_{\tau} = \sqrt{\frac{Q_{H.}}{n\sqrt{df}}} = \sqrt{\frac{Q_{H.}}{n\sqrt{(k-1)(l-1)}}$$

تعیین شدت همبستگی مشروط به این است که انتخاب افراد از نوع اول باشد.

که در این فرمول n تعداد نمونه، k تعداد سطر و l تعداد ستون در جدول توافقی است. همچنین $Q_{H.}$ مقدار آماره آزمون استقلال است که از رابطه زیر محاسبه می کنیم:

$$Q_{H.} = \sum_{j=1}^l \sum_{i=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

مقدار این ضریب برای دو صفت مستقل، صفر و برای دو صفت وابسته حداکثر یک خواهد بود.

آزمونهای کی دو

مثال

محقق می خواهد اثر نیروزایی نوعی ویتامین را بر افراد بررسی کند. او به ۱۰۰ نفر ویتامین مورد نظر و به ۱۰۰ نفر دیگر (گروه شاهد) شبه دارو تجویز می کند. جدول زیر حاوی نتایج این تجربه است. فرضیه مورد بحث را در سطح معنی داری ۵ درصد بیازمایید. (نحوه انتخاب افراد از نوع دوم است)

گروه	نتیجه آزمایش انرژی زایی		
	بدون تغییر	کاهش	افزایش
موردها	۵۶	۸	۳۶
شاهدها	۷۰	۱۰	۲۰

حل:

گروه	نتیجه آزمایش انرژی زایی			جمع
	بدون تغییر	کاهش	افزایش	
موردها	۵۶	۸	۳۶	۱۰۰
شاهدها	۷۰	۱۰	۲۰	۱۰۰
جمع	۱۲۶	۱۸	۵۶	۲۰۰

مقادیر مورد انتظار را برای هر ترکیب را حساب می کنیم.
مثلاً: مقدار مورد انتظار برای گروه موردها که بدون تغییر بوده اند به صورت زیر است:

$$E_{11} = \frac{O_{1.} \times O_{.1}}{200} = \frac{100 \times 126}{200} = 63$$

آزمونهای کی دو

فرضیه های آزمون به صورت زیر است.

$$\begin{cases} H_0 : & \text{رابطه ای بین دو متغیر وجود ندارد.} \\ H_1 : & \text{بین دو متغیر رابطه وجود دارد.} \end{cases}$$

برای محاسبه مقدار شاخص آزمون از رابطه زیر استفاده می کنیم:

$$Q_{H_1} = \sum_{j=1}^l \sum_{i=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\begin{aligned} Q_{H_1} &= \frac{(56 - 63)^2}{63} + \frac{(8 - 9)^2}{9} + \frac{(36 - 28)^2}{28} + \frac{(70 - 63)^2}{63} + \frac{(10 - 9)^2}{9} + \frac{(20 - 38)^2}{38} = \\ &= 0.87 + 0.11 + 2/9 + 0.78 + 0.11 + 2/9 = 7/58 \end{aligned}$$

$$RH_1 \Leftrightarrow Q_{H_1} > \chi^2_{(1-\alpha, (k-1)(l-1))}$$

فرض صفر را رد می کنیم چون مقدار شاخص آزمون ۷/۵۸ از مقدار ۵/۹۹ $\chi^2_{(0.95, 2)}$ بیشتر است. یعنی ویتامین مورد نظر بر انرژی زایی افراد تاثیر گذار است.



آزمونهای کی دو

مثال

به منظور مطالعه بستگی دو صفت رنگ مو و رنگ چشم افراد، و بیان شدت همبستگی در یک جمعیت، نمونه‌ای به حجم ۹۵ نفر به صورت تصادفی از این جمعیت انتخاب و بر حسب این دو صفت آنها را در جدول زیر ثبت کرده‌ایم. فرض عدم رابطه بین دو متغیر را با خطای نوع اول یک درصد بیازمایید.

رنگ مو	رنگ چشم		
	سیاه	سبز و آبی	سایر
روشن	۳۲	۱۴	۶
تیره	۱۲	۲۲	۹

حل:

مقادیر مورد انتظار را برای هر ترکیب را حساب می‌کنیم.
مثلاً: مقدار مورد انتظار برای افرادی با رنگ چشم روشن که رنگ موی سیاه دارند، به صورت زیر است:

$$E_{11} = \frac{O_{1.} \times O_{.1}}{n} = \frac{52 \times 44}{95} = 24/1$$

رنگ مو	رنگ چشم			جمع
	سیاه	سبز و آبی	سایر	
روشن	۳۲ ۲۴/۱	۱۴ ۱۹/۷	۶ ۸/۲	۵۲
تیره	۱۲ ۱۹/۹	۲۲ ۱۶/۳	۹ ۶/۸	۴۳
جمع	۴۴	۳۶	۱۵	۹۵

آزمونهای کی دو

فرضیه های آزمون به صورت زیر است.

$$\begin{cases} H_0 : & \text{رابطه ای بین رنگ چشم و رنگ موی افراد وجود ندارد.} \\ H_1 : & \text{بین دو متغیر رنگ چشم و رنگ مو رابطه وجود دارد.} \end{cases}$$

مقدار شاخص آزمون را حساب می کنیم:

$$Q_{H_0} = \sum_{j=1}^l \sum_{i=1}^k \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$Q_{H_0} = \frac{(32 - 24/1)^2}{24/1} + \frac{(14 - 19/7)^2}{19/7} + \frac{(6 - 8/2)^2}{8/2} + \frac{(12 - 19/9)^2}{19/9} + \frac{(22 - 16/3)^2}{16/3} + \frac{(9 - 6/8)^2}{6/8} = 10.67$$

$$RH_0 \Leftrightarrow Q_{H_0} > \chi^2_{(1-\alpha, (k-1)(l-1))}$$

فرض صفر را رد می کنیم چون مقدار شاخص آزمون ۱۰/۶۷ از ۹/۲۱ $\chi^2_{(0.99, 2)}$ بیشتر است. یعنی رنگ چشم و رنگ مو به هم ارتباط دارند.



آزمونهای کی دو

برای تعیین شدت همبستگی از ضریب چوپوروف استفاده می کنیم، داریم:

$$\rho_{\tau} = \sqrt{\frac{Q_{H.}}{n\sqrt{(k-1)(l-1)}}} = \sqrt{\frac{10/67}{95 \times \sqrt{2}}} = 0/28$$

شدت همبستگی برای این دو متغیر رنگ چشم و رنگ مو در
افراد این جمعیت ۲۸ درصد است.



تمرین

آزمونهای کی دو

۱- در یک بیمارستان طی ماههای گذشته ۱۷۳ عمل جراحی صورت گرفته است که توزیع تعداد عمل جراحی در هر گروه از بیماری ها در جدول زیر آمده است. با سطح معنی داری ۵ درصد، آزمون کنید آیا عمل جراحی برای همه بیماری ها به یک اندازه بوده است؟ χ^2 $\rightarrow$ χ^2

بیماری ها	قلبی عروقی	داخلی	چشم	زیبایی	ارولوژی
تعداد عمل انجام شده	۲۳	۳۱	۱۸	۴۴	۵۷

۲- در انستیتو پاستور ایران ۲۶ نفر را که گرگ هاری از ناحیه سر و گردن گاز گرفته بود، به دو گروه ۱۲ و ۱۴ تایی تقسیم کرده و گروه اول را با واکسن استاندارد پاستور و گروه دوم را علاوه بر واکسن با چند دوز سرم ضدهاری نیز معالجه کردند. χ^2 $\rightarrow$ χ^2

رنگ چشم	نتیجه درمان	
	زنده	مرده
واکسن	۹	۳
واکسن + سرم	۱۲	۲

الف- با توجه به اطلاعات جدول زیر، آیا اضافه کردن سرم در مرگ و میر بیماران موثر بوده است؟ (۵درصد)
 ب- آیا ضریب توافق چوپوروف برای این آزمون مفید است؟ χ^2 $\rightarrow$ χ^2

آزمونهای کی دو

جنسیت	اثر قرص		
	بی اثر	کم اثر	موثر
زن	۱۳	۱۵	۱۷
مرد	۱۱	۱۴	۲۰

۳- به منظور بررسی اثر تسکین دهنده‌گی یک نوع قرص مسکن بر جنسیت، ۴۵ زن و ۴۵ مرد را به تصادف انتخاب و نتیجه را در جدول زیر آورده ایم. در سطح معنی داری ۵ درصد، آزمون کنید آیا تاثیر این قرص در جنسیت افراد متفاوت است؟ χ^2 $\rightarrow$ χ^2

سواد سرپرست	تعداد فرزند		
	۱ و ۰	۲ و ۳	۴ و بیشتر
بی سواد	۳	۹	۸
تا دیپلم	۳	۱۱	۵
بالا تر	۷	۴	۳

۴- از بین ۵۳ خانوار انتخابی در یک جمعیت تعداد فرزند و سطح سواد سرپرست خانواده را پرسیده و نتیجه را در جدول زیر آورده ایم. χ^2 $\rightarrow$ χ^2

الف- با توجه به اطلاعات جدول زیر، آیا سطح سواد و تعداد فرزند به هم مرتبط اند؟ (با خطای یک درصد)

ب- اگر ضریب توافق چوپوروف برای این آزمون مفید است، شدت همبستگی را محاسبه کنید. χ^2 $\rightarrow$ χ^2

جدولها

جدول نرمال استاندارد

نحوه استفاده از جدول

هر بخش از جدول شامل ۳ ستون است. در ستون اول مقدار Z و در ستون دوم احتمال های کمتر از مقدار Z را مشاهده می کنید. ستون سوم اختصاص دارد به مقادیر بیشتر از Z یا کمتر از $-Z$ به عنوان مثال:

$$P(Z < 0.27) = 0.6064$$

$$P(Z > 0.27) = 0.3936 \text{ و در ستون بعدی}$$

$$P(Z < -0.27) \text{ که معادل است.}$$

پس برای احتمالات مقادیر منفی از ستون سوم استفاده کنید.

(ادامه جدول در صفحه بعد)

Z	Cum p	Tail p	Z	Cum p	Tail p
0.00	0.5000	0.5000	0.40	0.6554	0.3446
0.01	0.5040	0.4960	0.41	0.6591	0.3409
0.02	0.5080	0.4920	0.42	0.6628	0.3372
0.03	0.5120	0.4880	0.43	0.6664	0.3336
0.04	0.5160	0.4840	0.44	0.6700	0.3300
0.05	0.5199	0.4801	0.45	0.6736	0.3264
0.06	0.5239	0.4761	0.46	0.6772	0.3228
0.07	0.5279	0.4721	0.47	0.6808	0.3192
0.08	0.5319	0.4681	0.48	0.6844	0.3156
0.09	0.5359	0.4641	0.49	0.6879	0.3121
0.10	0.5398	0.4602	0.50	0.6915	0.3085
0.11	0.5438	0.4562	0.51	0.6950	0.3050
0.12	0.5478	0.4522	0.52	0.6985	0.3015
0.13	0.5517	0.4483	0.53	0.7019	0.2981
0.14	0.5557	0.4443	0.54	0.7054	0.2946
0.15	0.5596	0.4404	0.55	0.7088	0.2912
0.16	0.5636	0.4364	0.56	0.7123	0.2877
0.17	0.5675	0.4325	0.57	0.7157	0.2843
0.18	0.5714	0.4286	0.58	0.7190	0.2810
0.19	0.5753	0.4247	0.59	0.7224	0.2776
0.20	0.5793	0.4207	0.60	0.7257	0.2743
0.21	0.5832	0.4168	0.61	0.7291	0.2709
0.22	0.5871	0.4129	0.62	0.7324	0.2676
0.23	0.5910	0.4090	0.63	0.7357	0.2643
0.24	0.5948	0.4052	0.64	0.7389	0.2611
0.25	0.5987	0.4013	0.65	0.7422	0.2578
0.26	0.6026	0.3974	0.66	0.7454	0.2546
0.27	0.6064	0.3936	0.67	0.7486	0.2514
0.28	0.6103	0.3897	0.68	0.7517	0.2483
0.29	0.6141	0.3859	0.69	0.7549	0.2451
0.30	0.6179	0.3821	0.70	0.7580	0.2420
0.31	0.6217	0.3783	0.71	0.7611	0.2389
0.32	0.6255	0.3745	0.72	0.7642	0.2358
0.33	0.6293	0.3707	0.73	0.7673	0.2327
0.34	0.6331	0.3669	0.74	0.7704	0.2296
0.35	0.6368	0.3632	0.75	0.7734	0.2266
0.36	0.6406	0.3594	0.76	0.7764	0.2236
0.37	0.6443	0.3557	0.77	0.7794	0.2206
0.38	0.6480	0.3520	0.78	0.7823	0.2177
0.39	0.6517	0.3483	0.79	0.7852	0.2148

جدول نرمال استاندارد

Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p
0.80	0.7881	0.2119	1.20	0.8849	0.1151	1.60	0.9452	0.0548
0.81	0.7910	0.2090	1.21	0.8869	0.1131	1.61	0.9463	0.0537
0.82	0.7939	0.2061	1.22	0.8888	0.1112	1.62	0.9474	0.0526
0.83	0.7967	0.2033	1.23	0.8907	0.1093	1.63	0.9484	0.0516
0.84	0.7995	0.2005	1.24	0.8925	0.1075	1.64	0.9495	0.0505
0.85	0.8023	0.1977	1.25	0.8944	0.1056	1.65	0.9505	0.0495
0.86	0.8051	0.1949	1.26	0.8962	0.1038	1.66	0.9515	0.0485
0.87	0.8078	0.1922	1.27	0.8980	0.1020	1.67	0.9525	0.0475
0.88	0.8106	0.1894	1.28	0.8997	0.1003	1.68	0.9535	0.0465
0.89	0.8133	0.1867	1.29	0.9015	0.0985	1.69	0.9545	0.0455
0.90	0.8159	0.1841	1.30	0.9032	0.0968	1.70	0.9554	0.0446
0.91	0.8186	0.1814	1.31	0.9049	0.0951	1.71	0.9564	0.0436
0.92	0.8212	0.1788	1.32	0.9066	0.0934	1.72	0.9573	0.0427
0.93	0.8238	0.1762	1.33	0.9082	0.0918	1.73	0.9582	0.0418
0.94	0.8264	0.1736	1.34	0.9099	0.0901	1.74	0.9591	0.0409
0.95	0.8289	0.1711	1.35	0.9115	0.0885	1.75	0.9599	0.0401
0.96	0.8315	0.1685	1.36	0.9131	0.0869	1.76	0.9608	0.0392
0.97	0.8340	0.1660	1.37	0.9147	0.0853	1.77	0.9616	0.0384
0.98	0.8365	0.1635	1.38	0.9162	0.0838	1.78	0.9625	0.0375
0.99	0.8389	0.1611	1.39	0.9177	0.0823	1.79	0.9633	0.0367
1.00	0.8413	0.1587	1.40	0.9192	0.0808	1.80	0.9641	0.0359
1.01	0.8438	0.1562	1.41	0.9207	0.0793	1.81	0.9649	0.0351
1.02	0.8461	0.1539	1.42	0.9222	0.0778	1.82	0.9656	0.0344
1.03	0.8485	0.1515	1.43	0.9236	0.0764	1.83	0.9664	0.0336
1.04	0.8508	0.1492	1.44	0.9251	0.0749	1.84	0.9671	0.0329
1.05	0.8531	0.1469	1.45	0.9265	0.0735	1.85	0.9678	0.0322
1.06	0.8554	0.1446	1.46	0.9279	0.0721	1.86	0.9686	0.0314
1.07	0.8577	0.1423	1.47	0.9292	0.0708	1.87	0.9693	0.0307
1.08	0.8599	0.1401	1.48	0.9306	0.0694	1.88	0.9699	0.0301
1.09	0.8621	0.1379	1.49	0.9319	0.0681	1.89	0.9706	0.0294
1.10	0.8643	0.1357	1.50	0.9332	0.0668	1.90	0.9713	0.0287
1.11	0.8665	0.1335	1.51	0.9345	0.0655	1.91	0.9719	0.0281
1.12	0.8686	0.1314	1.52	0.9357	0.0643	1.92	0.9726	0.0274
1.13	0.8708	0.1292	1.53	0.9370	0.0630	1.93	0.9732	0.0268
1.14	0.8729	0.1271	1.54	0.9382	0.0618	1.94	0.9738	0.0262
1.15	0.8749	0.1251	1.55	0.9394	0.0606	1.95	0.9744	0.0256
1.16	0.8770	0.1230	1.56	0.9406	0.0594	1.96	0.9750	0.0250
1.17	0.8790	0.1210	1.57	0.9418	0.0582	1.97	0.9756	0.0244
1.18	0.8810	0.1190	1.58	0.9429	0.0571	1.98	0.9761	0.0239
1.19	0.8830	0.1170	1.59	0.9441	0.0559	1.99	0.9767	0.0233

جدول توزیع t

نحوه استفاده از جدول

در ستون اول درجه آزادی از ۱ تا ۱۲۰ و
بی نهایت آمده است.
در سطر اول بالای جدول مقادیر معنی
داری (آلفا) از ۰/۴ تا ۰/۰۰۵ آمده است.
به عنوان مثال:
مقدار t با ۱۰ درجه آزادی و سطح معنی
داری ۰/۰۵ برابر است با

$$t_{(10, 0.05)} = 1.812$$

v	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.656	127.321	318.289	636.578
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.328	31.600
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.894	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.689
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.660
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

جدول کی دو

نحوہ استفادہ از جدول

در ستون اول درجہ آزادی از ۱ تا ۱۰۰ آمده است.
در سطر اول بالای جدول مقادیر معنی داری (آلفا)
از ۰/۱ تا ۰/۰۰۱ آمده است.

به عنوان مثال :

مقدار χ^2 با ۱۰ درجہ آزادی و سطح معنی داری
۰/۰۵ برابر است با

$$\chi^2_{(10, 0.05)} = 18.31$$

جدول توزیع کی دو

درجہ آزادی	0.1	0.05	0.025	0.01	0.005	0.001
1	2.71	3.84	5.02	6.63	7.88	10.83
2	4.61	5.99	7.38	9.21	10.60	13.82
3	6.25	7.81	9.35	11.34	12.84	16.27
4	7.78	9.49	11.14	13.28	14.86	18.47
5	9.24	11.07	12.83	15.09	16.75	20.51
6	10.64	12.59	14.45	16.81	18.55	22.46
7	12.02	14.07	16.01	18.48	20.28	24.32
8	13.36	15.51	17.53	20.09	21.95	26.12
9	14.68	16.92	19.02	21.67	23.59	27.88
10	15.99	18.31	20.48	23.21	25.19	29.59
11	17.28	19.68	21.92	24.73	26.76	31.26
12	18.55	21.03	23.34	26.22	28.30	32.91
13	19.81	22.36	24.74	27.69	29.82	34.53
14	21.06	23.68	26.12	29.14	31.32	36.12
15	22.31	25.00	27.49	30.58	32.80	37.70
16	23.54	26.30	28.85	32.00	34.27	39.25
17	24.77	27.59	30.19	33.41	35.72	40.79
18	25.99	28.87	31.53	34.81	37.16	42.31
19	27.20	30.14	32.85	36.19	38.58	43.82
20	28.41	31.41	34.17	37.57	40.00	45.31
21	29.62	32.67	35.48	38.93	41.40	46.80
22	30.81	33.92	36.78	40.29	42.80	48.27
23	32.01	35.17	38.08	41.64	44.18	49.73
24	33.20	36.42	39.36	42.98	45.56	51.18
25	34.38	37.65	40.65	44.31	46.93	52.62
26	35.56	38.89	41.92	45.64	48.29	54.05
27	36.74	40.11	43.19	46.96	49.65	55.48
28	37.92	41.34	44.46	48.28	50.99	56.89
29	39.09	42.56	45.72	49.59	52.34	58.30
30	40.26	43.77	46.98	50.89	53.67	59.70
32	42.58	46.19	49.48	53.49	56.33	62.49
34	44.90	48.60	51.97	56.06	58.96	65.25
36	47.21	51.00	54.44	58.62	61.58	67.98
38	49.51	53.38	56.90	61.16	64.18	70.70
40	51.81	55.76	59.34	63.69	66.77	73.40
42	54.09	58.12	61.78	66.21	69.34	76.08
44	56.37	60.48	64.20	68.71	71.89	78.75
46	58.64	62.83	66.62	71.20	74.44	81.40
48	60.91	65.17	69.02	73.68	76.97	84.04
50	63.17	67.50	71.42	76.15	79.49	86.66
55	68.80	73.31	77.38	82.29	85.75	93.17
60	74.40	79.08	83.30	88.38	91.95	99.61
65	79.97	84.82	89.18	94.42	98.10	105.99
70	85.53	90.53	95.02	100.43	104.21	112.32
75	91.06	96.22	100.84	106.39	110.29	118.60
80	96.58	101.88	106.63	112.33	116.32	124.84
85	102.08	107.52	112.39	118.24	122.32	131.04
90	107.57	113.15	118.14	124.12	128.30	137.21
95	113.04	118.75	123.86	129.97	134.25	143.34
100	118.50	124.34	129.56	135.81	140.17	149.45

پایان

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

آمار و روش تحقیق

رشته‌های علوم پزشکی و شاخه‌های مرتبط

۲ واحد درسی

مطابق با سرفصل‌های مصوب

همبستگی و دگرسیون

مقدمه

وقتی شما صفت‌های متعددی از یک جمعیت را مطالعه می‌کنید ممکن است بخواهید بدانید بین این صفت‌ها رابطه‌ای وجود دارد یا خیر؟ مثلاً می‌دانیم بطور معمول بین قد و وزن افراد در یک جامعه رابطه مستقیم وجود دارد. یعنی اشخاص بلند قد وزن بیشتری نسبت به افراد کوتاه قد دارند. در چنین حالتی می‌گویند بین قد و وزن همبستگی وجود دارد. بر حسب اینکه متغیرها کمی یا کیفی باشند، انواع متفاوتی از همبستگی‌ها وجود دارند که در ادامه به بعضی از آنها اشاره خواهیم کرد. بررسی فرایند همبستگی را با مراحل زیر پی می‌گیریم:

- ۱- برآورد میزان همبستگی بین دو متغیر به کمک نمونه
- ۲- فاصله اطمینان برای پارامتر همبستگی در جامعه
- ۳- آزمون همبستگی متغیرهای کمی پیرسن
- ۴- تعیین معادله همبستگی (خط رگرسیون)
- ۵- آزمون همبستگی کی دو برای متغیرهای کیفی

همبستگی

ضریب همبستگی

ضریب همبستگی شاخصی است که میزان رابطه بین متغیرها را نشان می‌دهد. یک ملاک مناسب برای تعیین همبستگی دو متغیر کمی **ضریب همبستگی پیرسن** است که در جامعه آن را با ρ_{xy} و در نمونه آن را با r_{xy} نشان می‌دهیم.

$$\rho_{xy} = \frac{\overline{E(xy)} - E(x)E(y)}{\sqrt{[E(x^2) - E^2(x)][E(y^2) - E^2(y)]}}$$

ضریب همبستگی
پیرسن در جامعه

r_{xy} برآوردی نقطه‌ای از ρ_{xy} است.

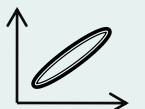
$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sqrt{s_x^2 \cdot s_y^2}}$$

ضریب همبستگی
پیرسن در نمونه

ضریب همبستگی همواره بین منفی یک و یک است. $-1 \leq r_{xy} \leq 1$

همبستگی

بر حسب اینکه مقدار ضریب همبستگی مثبت، منفی یا صفر باشد می توان نتیجه را به صورت زیر بیان کرد:

ضریب همبستگی	نمودار پراکنش	تفسیر
$\rho_{xy} = 1$		همبستگی دو متغیر کامل و مستقیم است
$0 < \rho_{xy} < 1$		همبستگی دو متغیر ناقص و مستقیم است
$\rho_{xy} = 0$		دو متغیر همبستگی ندارند. (ناهمبسته)
$-1 < \rho_{xy} < 0$		همبستگی دو متغیر ناقص و معکوس است
$\rho_{xy} = -1$		همبستگی دو متغیر کامل و معکوس است

مثلا

همبستگی بین دو متغیر قد و وزن مثبت و مستقیم ولی برعکس همبستگی بین دو متغیر سن و ایمنی بدن، همبستگی معکوس یا منفی است.

همبستگی

برای محاسبه میزان همبستگی بین دو متغیر، کافی است یک نمونه n تایی از هرکدام از متغیرها داشته باشیم. سپس با قرار دادن داده‌ها در فرمول ضریب همبستگی نمونه، میزان همبستگی دو متغیر را بدست آوریم.

$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sqrt{s_x^2 \cdot s_y^2}}$$

هر یک از اجزای فرمول ضریب همبستگی نمونه به صورت زیر است:

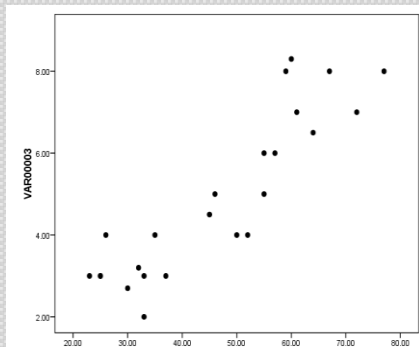
$$\bar{x} = \frac{1}{n} \sum x_i \quad \bar{y} = \frac{1}{n} \sum y_i \quad \overline{xy} = \frac{1}{n} \sum x_i y_i \quad s_y^2 = \overline{y^2} - \bar{y}^2 \quad s_x^2 = \overline{x^2} - \bar{x}^2$$

همبستگی

نمودار پراکنش

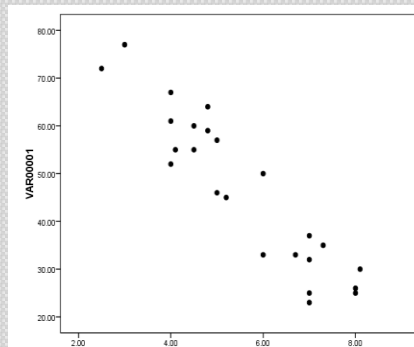
اگر هر یک از متغیرهای X و Y را روی یکی از محورهای مختصات قرار دهیم و نقاط توأم آنها را در صفحه مشخص نماییم می‌توان تصویر بهتری از رابطه آنها را مشاهده کرد. به این نمودار، **نمودار پراکنش** می‌گویند.

$$r_{xy} \approx +1$$



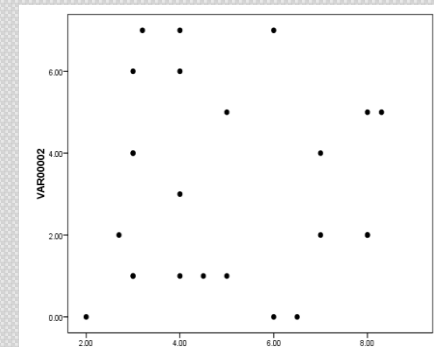
همبستگی مستقیم

$$r_{xy} \approx -1$$



همبستگی معکوس

$$r_{xy} \approx 0$$



عدم همبستگی

در مطالعه ارتباط بین دو متغیر باید بین مفهوم همبستگی و رابطه علت و معلولی بین آنها تفاوت گذاشت.

همبستگی

مثال

با توجه به داده های جدول زیر بررسی کنید آیا بین تعداد سیگار و ساعات خواب افراد سیگاری رابطه وجود دارد؟
(X = تعداد سیگار و Y = ساعات خواب است.)

x	y
۱	۸
۲	۶
۳	۷
۴	۶
۵	۴

حل:

کافی است ضریب همبستگی را برای داده های فوق به دست آوریم.

x	y	x^2	y^2	xy
۱	۸	۱	۶۴	۸
۲	۶	۴	۳۶	۱۲
۳	۷	۹	۴۹	۲۱
۴	۶	۱۶	۳۶	۲۴
۵	۴	۲۵	۱۶	۲۰
۱۵	۳۱	۵۵	۲۰۱	۸۵

جمع

$$\bar{x} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{31}{5} = 6.2$$

$$\bar{x}^2 = \frac{55}{5} = 11$$

$$\bar{y}^2 = \frac{201}{5} = 40.2$$

$$\overline{xy} = \frac{85}{5} = 17$$

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{17 - 18.6}{\sqrt{(11 - 9)(40.2 - 38.44)}} = \frac{-1.6}{1.88} = -0.85$$

با این نتیجه بین تعداد سیگار و ساعات خواب رابطه معکوس وجود دارد.

فاصله اطمینان ضریب همبستگی

با در دست داشتن یک نمونه و ایجاد یک حدود اطمینان می توان نسبت به ضریب همبستگی در جامعه قضاوت کرد.

جهت ایجاد حدود اطمینان برای پارامتر ضریب همبستگی نیاز به توزیع نمونه ای r_{xy} است. اگر توزیع توام دو صفت در جامعه نرمال باشد، آنگاه با تبدیل زیر می توان برای ضریب همبستگی جامعه یک فاصله اطمینان ساخت.

$$\omega = \frac{1}{2} \ln \frac{1+r}{1-r}$$

ثابت می شود ω دارای توزیع نرمال با میانگین و واریانسی به این صورت است:

$$E(\omega) = \mu_{\omega} = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$$

$$\text{var}(\omega) = \sigma_{\omega}^2 = \frac{1}{n-3}$$

بنابراین اگر یک نمونه n تایی از دو متغیر در اختیار داشته باشیم می توان حدود اطمینان برای ω را به صورت زیر نوشت:

$$\mu_{\omega} : \frac{1}{2} \ln \frac{1+r}{1-r} \pm z_{1-\frac{\alpha}{2}} \times \frac{1}{\sqrt{n-3}}$$

برای تبدیل ω به r نیازی به محاسبه لگاریتم نبوده و می توان از **جدول مربوط** به این تبدیل استفاده کرد.

مثال

در مثال مربوط به تعداد سیگار و خواب افراد، یک فاصله اطمینان برای ضریب همبستگی در جامعه به دست آورید.

حل: میانگین و انحراف معیار ω به صورت زیر است:

$$\begin{cases} \omega = \frac{1}{2} \ln \frac{1+r}{1-r} = \frac{1}{2} \ln \frac{1-0/85}{1+0/85} = -1/26 \\ s_{\omega} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{2}} = 0/707 \end{cases}$$

$$\mu_{\omega} : -1/26 \pm 1/96 \times 0/707 = (-2/65 \quad 0/13)$$

با استفاده از جدول تبدیل $\rho : (-0/99 \quad 0/13)$



همبستگی

آزمون ضریب همبستگی را به این منظور انجام می دهیم که مطمئن باشیم همبستگی بین دو متغیر ناشی از شانس و تصادف نیست و دلیل کافی برای تحلیل های بعدی همبستگی وجود دارد. بنابراین می توان مراحل انجام آزمون را مانند قبل، برای ضریب همبستگی نیز انجام داد.

فرضیه های آزمون: فرضیه های این آزمون به صورت زیر مطرح می شوند:

$$\begin{cases} H_0: \rho = 0 & \text{دو متغیر ناهمبسته اند} \\ H_1: \rho \neq 0 & \text{دو متغیر به هم بستگی دارند} \end{cases}$$

شاخص آماری آزمون: با توجه به ضابطه ای که برای هر شاخص آزمونی وجود دارد، شاخص آماری آزمون را به صورت زیر می نویسیم:

توزیع شاخص آزمون:

$$Q_H = \frac{\omega - 0}{1} \sqrt{n - 3}$$

همانگونه که قبلا نیز اشاره شد، توزیع این شاخص نرمال است.

همبستگی

استراتژی رد فرض صفر: با توجه به اینکه توزیع شاخص آزمون نرمال است، فرض صفر را وقتی رد می‌کنیم که مقدار آن با توجه به سطح معنی‌داری از مقدار جدول نرمال بیشتر باشد.

$$RH. \Leftrightarrow |Q_{H.}| > Z_{1-\frac{\alpha}{2}}$$

توان این آزمون به شدت به حجم نمونه وابسته است و به تجربه ثابت شده است که تعداد نمونه‌ها نباید کمتر از ۵ باشد.

در مثال قبل فرض عدم همبستگی دو متغیر تعداد سیگار و ساعات خواب را آزمون کنید.
آماره آزمون برابر است با:

$$Q_{H.} = \frac{\omega - 0}{1} = \frac{-1/26}{0.707} = -1/78$$
$$\sqrt{n-3}$$

چون $|-1/78| = 1/78$ از $Z_{0.975} = 1/96$ بزرگتر نیست، پس فرض صفر رد نمی‌شود. یعنی وابستگی دو متغیر نسبت به هم معنی‌دار نیست.



با توجه به اطلاعات جدول زیر؛ ابتدا بررسی کنید آیا بین دو متغیر رابطه وجود دارد؟ سپس ضریب همبستگی را در سطح ۰.۵٪ آزمون کنید.

ردیف	x	y	x ^۲	y ^۲	xy
۱	۰	۳	۰	۹	۰
۲	۲	۵	۴	۲۵	۱۰
۳	۳	۶	۹	۳۶	۱۸
۴	۴	۸	۱۶	۶۴	۳۲
۵	۵	۸	۲۵	۶۴	۴۰
۶	۶	۱۰	۳۶	۱۰۰	۶۰
جمع	۲۰	۴۰	۹۰	۳۰۸	۱۶۰

$$\bar{x} = \frac{20}{6} = 3.33$$

$$\bar{y} = \frac{40}{6} = 6.67$$

$$\bar{x}^2 = \frac{90}{6} = 15$$

$$\bar{y}^2 = \frac{308}{6} = 51.33$$

$$\overline{xy} = \frac{160}{6} = 26.67$$

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{s_x^2 \cdot s_y^2}} = \frac{26.67 - (3.33)(6.67)}{\sqrt{(15 - 11/1)(51.33 - 44/5)}} = \frac{4/46}{5/16} = 0.864$$

برای آزمون ضریب همبستگی داریم:

$$\omega = \frac{1}{2} \ln \frac{1+r}{1-r} = \frac{1}{2} \ln \frac{1+0.864}{1-0.864} = 1/31 \rightarrow Q_H = \frac{\omega - 0}{\frac{1}{\sqrt{n-3}}} = \frac{1/31}{0.577} = 2/27$$

چون ۲/۳۹ از ۱/۹۶ = Z_{۰.۹۷۵} بزرگتر است، پس فرض صفر رد می‌شود. یعنی دو متغیر نسبت به هم وابسته اند.



رگرسیون

رگرسیون

در صورتی که بین متغیرها رابطه وجود داشته باشد، می توان آن را با الگوهای ریاضی بیان کرد. معمولاً چنین الگویی ممکن است از نوع خطی یا غیر خطی باشد. اگر بتوان الگوی همبستگی را به صورت یک معادله خط نوشت، به آن معادله رگرسیون خطی می گویند.

هدف: در رگرسیون هدف آن است که با استفاده از معادله رگرسیون و به کمک یک نمونه تصادفی و بعضی روشهای آماری، رفتار متغیر وابسته را با آگاهی از مقادیر و مشخصات متغیرهای مستقل، پیش بینی کنیم.

در رگرسیون خطی ساده اگر Y را متغیر وابسته و X را متغیر مستقل در نظر بگیریم، می توان معادله خط رگرسیون را به صورت زیر نوشت:

$$Y = aX + b$$

در این معادله a و b را پارامترهای خط گفته و مانند هر پارامتر دیگری می توان آنها را برآورد کرد.

برآورد کردن پارامترهای خط رگرسیون: روشهای متفاوتی برای برآورد کردن دو پارامتر a و b وجود دارد. یکی از این روشها، روش «کمترین مربعات خطا» است که در آن این دو پارامتر به روش زیر برآورد می شود:

$$\hat{a} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} \quad \hat{b} = \bar{y} - \hat{a}\bar{x}$$

رگرسیون

مثال

x	y	
۱	۸	با توجه به داده های جدول زیر، بررسی انجام شده نشان داد بین تعداد سیگار
۲	۶	و ساعات خواب افراد سیگاری رابطه وجود دارد. معادله خط رگرسیون را به
۳	۷	دست آورید.
۴	۶	(X = تعداد سیگار متغیر مستقل و Y = ساعات خواب متغیر وابسته است)
۵	۴	

حل:

با توجه به داده های فوق پارامترهای خط رگرسیون را برآورد کرده و معادله خط را می نویسیم.

x	y	x^2	y^2	xy
۱	۸	۱	۶۴	۸
۲	۶	۴	۳۶	۱۲
۳	۷	۹	۴۹	۲۱
۴	۶	۱۶	۳۶	۲۴
۵	۴	۲۵	۱۶	۲۰
۱۵	۳۱	۵۵	۲۰۱	۸۵

جمع

$$\bar{x} = \frac{15}{5} = 3$$

$$\bar{y} = \frac{31}{5} = 6.2$$

$$\bar{x}^2 = \frac{55}{5} = 11$$

$$\bar{y}^2 = \frac{201}{5} = 40.2$$

$$\overline{xy} = \frac{85}{5} = 17$$

$$\hat{a} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x^2} = \frac{17 - 18.6}{11 - 9} = \frac{-1.6}{2} = -0.8$$

$$\hat{b} = \bar{y} - \hat{a}\bar{x} = 6.2 + 0.8 \times 3 = 8.6$$

معادله خط رگرسیون به این صورت است: $y = -0.8x + 8.6$

رگرسیون

در رگرسیون هدف آن است که با استفاده از معادله خط رگرسیون رفتار متغیر وابسته را با آگاهی از مقادیر و مشخصات متغیرهای مستقل، پیش بینی کنیم.

مثال

با توجه به نتایج مثال قبل معادله خط رگرسیون را به صورت زیر دست آورده ایم:
در صورتی که فرد سیگاری روزی $x = 7$ سیگار استعمال کند، چه مدت خواب برای او پیش بینی می کنید.

$$y = -0.1x + 3/8$$

(X = تعداد سیگار متغیر مستقل و Y = ساعات خواب متغیر وابسته است)

حل:

با قرار دادن مقدار $x = 7$ در معادله خط رگرسیون داریم:

$$y = -0.1(7) + 3/8 = 2/9$$



تمرین

تمرین

- ۱- اگر همبستگی بین دو متغیر کامل باشد می توان گفت یکی از آنها معلول دیگری است؟ چرا؟
- ۲- در آزمون ضریب همبستگی فرض صفر بر چه اساسی مطرح شده است؟
- ۳- به منظور بررسی رابطه بین سن (X) و فشار خون (Y)، اطلاعات مربوط به ۱۰ نفر از مراجعه کنندگان به یک درمانگاه را به صورت جدول زیر ثبت کرده ایم.

x	y
۱۰	۱۲/۱
۱۸	۱۲/۶
۲۲	۱۲
۲۹	۱۳
۳۵	۱۲٫۸
۴۰	۱۴/۵
۴۳	۱۳/۲
۴۵	۱۶/۱
۴۶	۱۳
۵۰	۱۴/۶

- الف- ضریب همبستگی بین این دو متغیر را تعیین و تفسیر کنید.
- ب- یک فاصله اطمینان ۹۵ درصد برای ضریب همبستگی به دست آورید
- ج- در صورت معنی داری ضریب همبستگی معادله ضریب همبستگی (معادله خط رگرسیون) را بنویسید
- د- برای فردی با سن ۴۴ سال چه میزان فشار خون پیش بینی می کنید؟

- ۴- اگر معادله خط رگرسیون بین دو متغیر به صورت $2y - 3x - 1 = 0$ باشد، آیا می توان گفت بین دو متغیر همبستگی معنی دار است؟ چرا؟

- ۵- اگر همبستگی بین دو متغیر تقریباً صفر باشد، آیا می توان گفت بین آنها هیچ رابطه ای وجود ندارد؟ چرا؟

جدول‌های آماری

جدول‌ها

جدول نرمال استاندارد

نحوه استفاده از جدول

هر بخش از جدول مقابل شامل ۳ ستون است. در ستون اول مقدار Z و در ستون دوم احتمال های کمتر از مقدار Z را مشاهده می کنید. ستون سوم اختصاص دارد به مقادیر بیشتر از Z یا کمتر از $-Z$ به عنوان مثال:

$$P(Z < 0.27) = 0.6064$$

$$P(Z > 0.27) = 0.3936 \text{ و در ستون بعدی}$$

که معادل $P(Z < -0.27)$ است.

(ادامه جدول در صفحه بعد)

Z	Cum p	Tail p	Z	Cum p	Tail p
0.00	0.5000	0.5000	0.40	0.6554	0.3446
0.01	0.5040	0.4960	0.41	0.6591	0.3409
0.02	0.5080	0.4920	0.42	0.6628	0.3372
0.03	0.5120	0.4880	0.43	0.6664	0.3336
0.04	0.5160	0.4840	0.44	0.6700	0.3300
0.05	0.5199	0.4801	0.45	0.6736	0.3264
0.06	0.5239	0.4761	0.46	0.6772	0.3228
0.07	0.5279	0.4721	0.47	0.6808	0.3192
0.08	0.5319	0.4681	0.48	0.6844	0.3156
0.09	0.5359	0.4641	0.49	0.6879	0.3121
0.10	0.5398	0.4602	0.50	0.6915	0.3085
0.11	0.5438	0.4562	0.51	0.6950	0.3050
0.12	0.5478	0.4522	0.52	0.6985	0.3015
0.13	0.5517	0.4483	0.53	0.7019	0.2981
0.14	0.5557	0.4443	0.54	0.7054	0.2946
0.15	0.5596	0.4404	0.55	0.7088	0.2912
0.16	0.5636	0.4364	0.56	0.7123	0.2877
0.17	0.5675	0.4325	0.57	0.7157	0.2843
0.18	0.5714	0.4286	0.58	0.7190	0.2810
0.19	0.5753	0.4247	0.59	0.7224	0.2776
0.20	0.5793	0.4207	0.60	0.7257	0.2743
0.21	0.5832	0.4168	0.61	0.7291	0.2709
0.22	0.5871	0.4129	0.62	0.7324	0.2676
0.23	0.5910	0.4090	0.63	0.7357	0.2643
0.24	0.5948	0.4052	0.64	0.7389	0.2611
0.25	0.5987	0.4013	0.65	0.7422	0.2578
0.26	0.6026	0.3974	0.66	0.7454	0.2546
0.27	0.6064	0.3936	0.67	0.7486	0.2514
0.28	0.6103	0.3897	0.68	0.7517	0.2483
0.29	0.6141	0.3859	0.69	0.7549	0.2451
0.30	0.6179	0.3821	0.70	0.7580	0.2420
0.31	0.6217	0.3783	0.71	0.7611	0.2389
0.32	0.6255	0.3745	0.72	0.7642	0.2358
0.33	0.6293	0.3707	0.73	0.7673	0.2327
0.34	0.6331	0.3669	0.74	0.7704	0.2296
0.35	0.6368	0.3632	0.75	0.7734	0.2266
0.36	0.6406	0.3594	0.76	0.7764	0.2236
0.37	0.6443	0.3557	0.77	0.7794	0.2206
0.38	0.6480	0.3520	0.78	0.7823	0.2177
0.39	0.6517	0.3483	0.79	0.7852	0.2148

جدول نرمال استاندارد

Z	Cum p	Tail p	Z	Cum p	Tail p	Z	Cum p	Tail p
0.80	0.7881	0.2119	1.20	0.8849	0.1151	1.60	0.9452	0.0548
0.81	0.7910	0.2090	1.21	0.8869	0.1131	1.61	0.9463	0.0537
0.82	0.7939	0.2061	1.22	0.8888	0.1112	1.62	0.9474	0.0526
0.83	0.7967	0.2033	1.23	0.8907	0.1093	1.63	0.9484	0.0516
0.84	0.7995	0.2005	1.24	0.8925	0.1075	1.64	0.9495	0.0505
0.85	0.8023	0.1977	1.25	0.8944	0.1056	1.65	0.9505	0.0495
0.86	0.8051	0.1949	1.26	0.8962	0.1038	1.66	0.9515	0.0485
0.87	0.8078	0.1922	1.27	0.8980	0.1020	1.67	0.9525	0.0475
0.88	0.8106	0.1894	1.28	0.8997	0.1003	1.68	0.9535	0.0465
0.89	0.8133	0.1867	1.29	0.9015	0.0985	1.69	0.9545	0.0455
0.90	0.8159	0.1841	1.30	0.9032	0.0968	1.70	0.9554	0.0446
0.91	0.8186	0.1814	1.31	0.9049	0.0951	1.71	0.9564	0.0436
0.92	0.8212	0.1788	1.32	0.9066	0.0934	1.72	0.9573	0.0427
0.93	0.8238	0.1762	1.33	0.9082	0.0918	1.73	0.9582	0.0418
0.94	0.8264	0.1736	1.34	0.9099	0.0901	1.74	0.9591	0.0409
0.95	0.8289	0.1711	1.35	0.9115	0.0885	1.75	0.9599	0.0401
0.96	0.8315	0.1685	1.36	0.9131	0.0869	1.76	0.9608	0.0392
0.97	0.8340	0.1660	1.37	0.9147	0.0853	1.77	0.9616	0.0384
0.98	0.8365	0.1635	1.38	0.9162	0.0838	1.78	0.9625	0.0375
0.99	0.8389	0.1611	1.39	0.9177	0.0823	1.79	0.9633	0.0367
1.00	0.8413	0.1587	1.40	0.9192	0.0808	1.80	0.9641	0.0359
1.01	0.8438	0.1562	1.41	0.9207	0.0793	1.81	0.9649	0.0351
1.02	0.8461	0.1539	1.42	0.9222	0.0778	1.82	0.9656	0.0344
1.03	0.8485	0.1515	1.43	0.9236	0.0764	1.83	0.9664	0.0336
1.04	0.8508	0.1492	1.44	0.9251	0.0749	1.84	0.9671	0.0329
1.05	0.8531	0.1469	1.45	0.9265	0.0735	1.85	0.9678	0.0322
1.06	0.8554	0.1446	1.46	0.9279	0.0721	1.86	0.9686	0.0314
1.07	0.8577	0.1423	1.47	0.9292	0.0708	1.87	0.9693	0.0307
1.08	0.8599	0.1401	1.48	0.9306	0.0694	1.88	0.9699	0.0301
1.09	0.8621	0.1379	1.49	0.9319	0.0681	1.89	0.9706	0.0294
1.10	0.8643	0.1357	1.50	0.9332	0.0668	1.90	0.9713	0.0287
1.11	0.8665	0.1335	1.51	0.9345	0.0655	1.91	0.9719	0.0281
1.12	0.8686	0.1314	1.52	0.9357	0.0643	1.92	0.9726	0.0274
1.13	0.8708	0.1292	1.53	0.9370	0.0630	1.93	0.9732	0.0268
1.14	0.8729	0.1271	1.54	0.9382	0.0618	1.94	0.9738	0.0262
1.15	0.8749	0.1251	1.55	0.9394	0.0606	1.95	0.9744	0.0256
1.16	0.8770	0.1230	1.56	0.9406	0.0594	1.96	0.9750	0.0250
1.17	0.8790	0.1210	1.57	0.9418	0.0582	1.97	0.9756	0.0244
1.18	0.8810	0.1190	1.58	0.9429	0.0571	1.98	0.9761	0.0239
1.19	0.8830	0.1170	1.59	0.9441	0.0559	1.99	0.9767	0.0233

جدول تبدیل r به ω

نحوه استفاده از جدول

هنگام ایجاد فاصله اطمینان برای ضریب همبستگی و آزمون های مربوط به ضریب همبستگی از تبدیل

$$\omega = \frac{1}{2} \ln \frac{1+r}{1-r}$$

استفاده می کنند. که نیاز است مجدداً از تبدیل عکس آن مقدار r را مشخص کنند. در این جدول مقادیر r و ω در مقابل هم آمده اند.

$$\omega = \frac{1}{2} \ln \frac{1+r}{1-r}$$

جدول تبدیل r به ω

r	ω	r	ω	r	ω	r	ω
۰.۰۰	۰.۰۰۰	۰.۲۵	۰.۲۵۵	۰.۵۰	۰.۵۴۹	۰.۷۵	۰.۹۷۳
۰.۰۱	۰.۰۱۰	۰.۲۶	۰.۲۶۶	۰.۵۱	۰.۵۶۳	۰.۷۶	۰.۹۹۶
۰.۰۲	۰.۰۲۰	۰.۲۷	۰.۲۷۷	۰.۵۲	۰.۵۷۶	۰.۷۷	۱.۰۲۰
۰.۰۳	۰.۰۳۰	۰.۲۸	۰.۲۸۸	۰.۵۳	۰.۵۹۰	۰.۷۸	۱.۰۴۵
۰.۰۴	۰.۰۴۰	۰.۲۹	۰.۲۹۹	۰.۵۴	۰.۶۰۴	۰.۷۹	۱.۰۷۱
۰.۰۵	۰.۰۵۰	۰.۳۰	۰.۳۱۰	۰.۵۵	۰.۶۱۸	۰.۸۰	۱.۰۹۹
۰.۰۶	۰.۰۶۰	۰.۳۱	۰.۳۲۱	۰.۵۶	۰.۶۳۳	۰.۸۱	۱.۱۲۷
۰.۰۷	۰.۰۷۰	۰.۳۲	۰.۳۳۲	۰.۵۷	۰.۶۴۸	۰.۸۲	۱.۱۵۷
۰.۰۸	۰.۰۸۰	۰.۳۳	۰.۳۴۳	۰.۵۸	۰.۶۶۲	۰.۸۳	۱.۱۸۸
۰.۰۹	۰.۰۹۰	۰.۳۴	۰.۳۵۴	۰.۵۹	۰.۶۷۸	۰.۸۴	۱.۲۲۱
۰.۱۰	۰.۱۰۰	۰.۳۵	۰.۳۶۵	۰.۶۰	۰.۶۹۳	۰.۸۵	۱.۲۵۶
۰.۱۱	۰.۱۱۰	۰.۳۶	۰.۳۷۷	۰.۶۱	۰.۷۰۹	۰.۸۶	۱.۲۹۳
۰.۱۲	۰.۱۲۱	۰.۳۷	۰.۳۸۸	۰.۶۲	۰.۷۲۵	۰.۸۷	۱.۳۳۳
۰.۱۳	۰.۱۳۱	۰.۳۸	۰.۴۰۰	۰.۶۳	۰.۷۴۱	۰.۸۸	۱.۳۷۶
۰.۱۴	۰.۱۴۱	۰.۳۹	۰.۴۱۲	۰.۶۴	۰.۷۵۸	۰.۸۹	۱.۴۲۲
۰.۱۵	۰.۱۵۱	۰.۴۰	۰.۴۲۴	۰.۶۵	۰.۷۷۵	۰.۹۰	۱.۴۷۲
۰.۱۶	۰.۱۶۱	۰.۴۱	۰.۴۳۶	۰.۶۶	۰.۷۹۳	۰.۹۱	۱.۵۲۸
۰.۱۷	۰.۱۷۲	۰.۴۲	۰.۴۴۸	۰.۶۷	۰.۸۱۱	۰.۹۲	۱.۵۸۹
۰.۱۸	۰.۱۸۲	۰.۴۳	۰.۴۶۰	۰.۶۸	۰.۸۲۹	۰.۹۳	۱.۶۵۸
۰.۱۹	۰.۱۹۲	۰.۴۴	۰.۴۷۲	۰.۶۹	۰.۸۴۸	۰.۹۴	۱.۷۳۸
۰.۲۰	۰.۲۰۳	۰.۴۵	۰.۴۸۵	۰.۷۰	۰.۸۶۷	۰.۹۵	۱.۸۳۲
۰.۲۱	۰.۲۱۳	۰.۴۶	۰.۴۹۷	۰.۷۱	۰.۸۸۷	۰.۹۶	۱.۹۴۶
۰.۲۲	۰.۲۲۴	۰.۴۷	۰.۵۱۰	۰.۷۲	۰.۹۰۸	۰.۹۷	۲.۰۹۲
۰.۲۳	۰.۲۳۴	۰.۴۸	۰.۵۲۳	۰.۷۳	۰.۹۲۹	۰.۹۸	۲.۲۹۸
۰.۲۴	۰.۲۴۵	۰.۴۹	۰.۵۳۶	۰.۷۴	۰.۹۵۰	۰.۹۹	۲.۶۴۷

پایان

انتخاب آزمون آماری مناسب

از کدام آزمون آماری استفاده کنیم؟

این سوال اغلب محققین جوان یا دانشجویانی است که در حال گذراندن پایان نامه هستند و داده‌ها را جمع آوری کرده‌اند ولی هنوز موفق به انتخاب آزمون مناسب برای طرح خود نشده‌اند.

حوزه آمار تا آنجا توسعه یافته است که ما را قادر می‌سازد تا تقریباً برای هر نوع طرح تحقیقی که مطرح می‌شود چندین آزمون آماری در اختیار داشته باشیم. با توجه به چنین امکاناتی، باید در عین حال منطقی را نیز برای انتخاب نوع آزمون رعایت کنیم. همواره محقق پس از در اختیار داشتن داده‌ها با این سوال رو برو است که از چه آزمونی برای فرضیات مطرح شده در طرح خود به جهت رسیدن به نتیجه منطقی و مطلوب استفاده کند؟ هر چند سوال تحقیق و فرضیه‌های مطرح شده تا حدودی به او کمک می‌کنند ولی دانستن اینکه چه عواملی در انتخاب یک آزمون دخالت دارند، از اهمیت ویژه‌ای برخوردار است. البته بهترین راه برای یادگیری اینکه از کدام آزمون آماری استفاده کنید این است که خودتان یک آماردان با تجربه باشید. در اینجا هم تجربه بزرگترین معلم است. در حقیقت هیچ روشی وجود ندارد که بتواند به شما یاد دهد که از چه چیزی و در چه زمانی استفاده کنید. مگر اینکه فرصت کافی و تجربه لازم برای انتخاب ابزار مناسب را داشته باشید. این پاورپوینت صرفاً راه و روش استفاده از ابزارها را به شما پیشنهاد می‌دهد ولی همه‌ی کار نیست و فقط برای شروع کار مناسب است. ممکن است برای یک طرح آزمونهای متفاوتی وجود داشته باشد.

انتخاب آزمون آماری مناسب

موضوعات اساسی که در انتخاب یک آزمون آماری دخالت دارند

پرسش تحقیق

سوالاتی که در ابتدا، یک محقق را برای رسیدن به جواب آن، وادار به تحقیق می‌کند.

فرضیه تحقیق

هر فرضیه آماری مبتنی بر یک آماره (شاخص آزمون) و هر آماره، توزیع خاصی دارد و توزیع‌های آماری در تعیین آزمون، تعیین کننده هستند.

ماهیت داده‌ها

مشخص کردن متغیرهای مطالعه و نوع داده‌هایی که محقق از پرسشنامه یا هر منبع دیگری به دست آورده است.

پرسش تحقیق ممکن است در یکی از دسته سوالات اساسی زیر قرار گیرد. این سوالات ما را به سمت انتخاب نوع آزمونی که مناسب تحقیق باشد، هدایت می‌کند.

انتخاب آزمون آماری مناسب

پرسش‌های کلی

□ سوال‌های مقایسه‌ای

- ۱- آیا می‌خواهید توزیع یک نمونه انتخابی را با یک توزیع نظری خاص مقایسه کنید؟
یا توزیع داده‌ها را با توزیع جمعیتی که داده‌ها از آن استخراج شده‌اند، مقایسه کنید؟
- ۲- آیا قصد دارید دو گروه را با یکدیگر مقایسه کنید؟
- ۳- آیا می‌خواهید بیشتر از دو گروه را با یکدیگر مقایسه کنید؟

□ سوال همبستگی

- ۳- آیا می‌خواهید رابطه همبستگی بین دو یا چند متغیر را بررسی کنید؟

□ سوال رگرسیون

- ۴- آیا می‌خواهید رفتار یک متغیر را از روی تغییرات یک یا چند متغیر دیگر الگوسازی و پیش‌بینی کنید؟

انتخاب آزمون آماری مناسب

پاسخ های کلی

□ آزمون های مقایسه‌ای

• اگر سوال تحقیق به دنبال آن است که نمونه ای که در اختیار محقق است، از یک توزیع آماری خاص پیروی می‌کند، آزمون‌های یک نمونه‌ای مانند آزمون یک نمونه‌ای t ، آزمون مربع کای، آزمون دو جمله‌ای و آزمون‌های کولموگروف - اسمیرنوف، آزمون‌های مناسبی هستند.

• اگر سوال یک تحقیق به دنبال اختلاف معنی داری بین دو گروه است، از آزمون‌های دو نمونه‌ای مانند آزمون t زوجی یا مستقل، آزمون کی دو، آزمون من ویتنی، آزمون مک نمار و آزمون‌های مشابه استفاده کنید.

• اگر در پی اختلاف معنی داری بین بیش از دو گروه هستید، از آزمون‌های آنالیز واریانس یا طرح آزمایشات استفاده کنید.

□ آزمون های همبستگی

اگر در پی یافتن رابطه همبستگی بین دو یا چند متغیر هستید از آزمون‌های همبستگی مانند آزمون کی دو، پیرسن، اسپیرمن و تاو کندانال و ... استفاده کنید.

□ رگرسیون

اگر به دنبال پیش بینی مقادیر یک متغیر، از روی اطلاعات مربوط به یک یا چند متغیر دیگر هستید، مباحث مربوط به رگرسیون (خطی، غیر خطی، لجستیک و ...) را دنبال کنید.

انتخاب آزمون آماری مناسب

آزمون های مقایسه ای در نرم افزار spss

□ اگر سوال تحقیق از دسته سوالات مقایسه ای است می توانید بر اساس تعداد گروه های مقایسه و اینکه ماهیت متغیر، مستقل است یا وابسته، از هر یک از آزمون های ارائه شده در جدولی که در پاورپوینت بعدی آمده است، آزمون مورد نظر را انتخاب کنید. ضمناً ما سعی کرده ایم برای هر آزمون روش استفاده آن را در نرم افزار spss حداقل با یک مثال و در قالب یک فایل آموزشی به اختصار بیان کنیم. برای استفاده از این فایل ها از لینک های ارائه شده در ستون انتهایی جدول استفاده کنید تا به این فایل ها هدایت شوید.

□ برای سوالات مربوط به همبستگی و رگرسیون جداول جداگانه ای در نظر گرفته شده است که در آینده ارائه خواهیم کرد.

یک نمونه‌ای	همه		آزمون دو جمله ای	فایل آموزش در spss	
	ترتیبی و اسمی		آزمون همتوزیعی کی دو	فایل آموزش در spss	
	همه		آزمون تصادفی بودن runs	فایل آموزش در spss	
	کمی		آزمون t یک نمونه ای	فایل آموزش در spss	
			کلموگرف – اسمیرنوف	فایل آموزش در spss	
مقایسه دو گروه	کمی	مستقل	آزمون t مستقل	فایل آموزش در spss	
		وابسته	آزمون t زوجی	فایل آموزش در spss	
	رتبه‌ای	مستقل	آزمون من ویتنی	فایل آموزش در spss	
		وابسته	آزمون ویلکاکسن یا علامت	فایل آموزش در spss	
	اسمی	مستقل	آزمون کی دو	فایل آموزش در spss	
		وابسته	آزمون مک نمار	فایل آموزش در spss	
	مقایسه بیشتر از دو گروه	کمی	مستقل	آنالیز واریانس یک طرفه	فایل آموزش در spss
			وابسته	آنالیز واریانس با مشاهدات تکراری	به زودی
رتبه‌ای		مستقل	آزمون کروسکال والیس	فایل آموزش در spss	
		وابسته	آزمون فریدمن	فایل آموزش در spss	
اسمی		مستقل	آزمون کی دو	فایل آموزش در spss	
		وابسته	آزمون کوکران	فایل آموزش در spss	