موسسه آموزش عالی آزاد ماهان
www.mahan.ac.ir

به نام خداوند بخشنده مهربان

آمادگی آزمون دکتری

آمار و روش تحقیق

مجموعه علوم ارتباطات و علوم اجتماعی

مؤلف:

دکتر طهمورث شیری

ماهان
مشاوران صعود

شیری، طهمورث (۱۳٤٦)
آمار و روش تحقیق/ دکتر طهمورث شیری
مشاوران صعود ماهان: ۱٤۰۱
۳۱۲ص: جدول، نمودار (آمادگی آزمون دکتری علوم ارتباطات و علوم اجتماعی)
ISBN: 978-600-458-898-0
فهرست‌نویسی بر اساس اطلاعات فیپا.
فارسی ـ چاپ اوّل
آمار و روش تحقیق
دکتر طهمورث شیری
ج ـ عنوان
کتابخانه ملی ایران ۳۲٤٥۱۲۳

انتشارات مشاوران صعود ماهان

- **نام کتاب:** .... آمار و روش تحقیق
- **مدیران مسئول:** .... مجید و هادی سیّاری
- **مولف:** .... دکتر طهمورث شیری
- **مسئول برنامه‌ریزی و تولید محتوا** .... سمیه بیگی
- **ناشر:** .... مشاوران صعود ماهان
- **نوبت و تاریخ چاپ:** .... اوّل /۱۴۰۱
- **تیراژ:** .... ۱۰۰۰ نسخه
- **قیمت:** ۳/۳۹۰/ ۰۰۰ریال
- **شابک:** .... ISBN۹۷۸-۶۰۰-۴۵۸-۸۹۸ -۰

انتشارات مشاوران صعود ماهان:تهران ـ خیابان ولیعصر، بالاتر از تقاطع ولیعصر مطهری، پلاک ۲۰۵۰
تلفن: ۸۸۱۰۰۱۱۳ و ۸۸۴۰۱۳۱۳

مقدمه ناشر

بنام خدا

ایمان داریم که هر تغییر و تحول بزرگی در مسیر زندگی بدون تحول معرفت و نگرش میسر نخواهد بود. پس بیایید با اندیشه توکل، تفکر، تلاش و تحمل در توسعه دنیای فکریمان برای نیل به آرامش و آسایش توأمان اوّلین گام را برداریم؛ چون همگی یقین داریم دانایی، توانایی می‌آورد.

شاد باشید و دلی را شاد کنید.

برادران سیاری

مقدمه مؤلف

ایده تشکیل جلساتی برای هدایت فارغ‌التحصیلان مقطع کارشناسی ارشد که قصد ورود به دوره دکتری در رشته ارتباطات اجتماعی را دارند در سال ۱۳۷۹ برای اوّلین بار در ایران، توسط برادران ارجمند سیاری مطرح و از همان سال کلاسهایی به صورت فشرده تشکیل شده که بسیار موفقیت‌آمیز بوده و نتیجه آن شد که در سالهای ۷۹- ۸۰همه شرکت‌کنندگان در کلاسهای ماهان مؤفق به ورود به دوره دکتری شدند. از آن پس کلاسهای کنکور دکتری تا به امروز در مؤسسه آموزش عالی آزاد ماهان با موفقیت روز افزون ادامه داشته است. مهمترین نکته در این دوره‌ها تفاوتی است که با کلاسهای رسمی دانشگاهی دارند. در این دوره‌ها سعی می‌شود به نکات کلیدی مباحث تأکید شود (چه در آزمونهای تشریحی سالهای قبل یا آزمونهای تستی سالهای اخیر). متن حاضر مجموعه‌ای گردآوری شده از مباحث آمار و روش تحقیق است که از چند کتاب مطرح که منابع اصلی آزمون دکتری هستند، تهیه شده است. اسامی این کتب (به دلیل رعایت امانت) در انتهای بحث آمده است. پس در واقع نمی‌توان متن حاضر را کتاب درسی صرف دانشگاهی دانست؛ بلکه مطالب آن جمع‌آوری و خلاصه شدهٔ مباحث و نکات کلیدی آمار و بحث گستردهٔ روش‌های تحقیق در علوم اجتماعی و ارتباطات از منابع مورد اشاره است. امید آن می‌رود عزیزانی که امکان مطالعه منابع نسبتاً زیاد دوره دکتری در رشته‌های ارتباطات اجتماعی را ندارند، با تلخیص صورت گرفته (و البته شرکت در کلاسهای درسی ماهان) موفق به پاسخگویی به پرسش‌های آزمون دوره دکتری گردند. بدیهی است مطالعه متن حاضر تنها با حضور مستمر در کلاسهای درسی، امکان ورود به دوره دکتری را افزایش می‌دهد.

دکتر طهمورث شیری

فهرست مطالب

کلیات

تعریف علم آمار: آمار، علمی است که خواص جامعه را بررسی می‌کند.

یا آمار علم جمع‌آوری داده‌ها، آنالیز داده‌ها و استنتاج آنان است.

مفاهیم اوّلیه[1]:

جامعه آماری (جمعیت آماری) با N نشان داده می‌شود؛ یعنی کل افرادی که قرار است مورد مطالعه قـرار گیرنـد. دو تعریف از جامعه آماری وجود دارد.

تعریف محض از جامعه آماری: کلیه افراد، اشیا و عناصری که حداقل یک صفت مشترک دارند.

درتعریف کاربردی؛ جامعه آماری همان کل افراد مورد مطالعه ما هستند.

انواع جامعه آماری: جامعه آماری یا محدود و قابل شمارش است یا نامحدود و غیرقابل شمارش

حجم جامعه(N) به تعداد کل مقادیر صفت در جامعه اطلاق می‌شود.

حجم نمونه(n)، جزئی از کل است.

هر دوی آنها (جامعه و نمونه) به لحاظ آماری قابل مطالعه هستند. تلاش این است کـه نمونـه بـه لحـاظ شـباهت و خواص به جامعه آماری نزدیک باشد.

آماره: شاخصه‌های آماری مورد مطالعه در نمونه (مانند واریانس نمونه)

پارامتر: شاخصه‌های آماری مورد مطالعه در جامعه آماری (مانند واریانس جامعه)

فلسفه نمونه‌گیری: مطالعه پارامترها از طریق آماره‌ها است.

فرد جامعه: به هر کدام از افراد یا اشیای مورد مطالعه می‌گویند. در واقع واحد تحلیل است.

مفهوم concept: سمبل یا نمادی که از تجرید معانی در ذهن ایجاد می‌شود.

متغیر variable: صفتی است که در جامعه ارزش‌های مختلفی به خودش بگیرد (حداقل ۲ ارزش داشته باشد.) یـا صفتی که در جامعه از فردی به فرد دیگر تغییر کند.

مانند سن، درآمد و... مثلاً درآمد، یک مفهوم است ولی میزان درآمد متغیر است.

در تحقیقات کمی بیشتر با متغیرها سروکار داریم و در تحقیقات کیفی به مفاهیم نیاز داریم.

انواع متغیر:

متغیرهای کمّی: صفاتی که بوسیله اعداد و در قالب کمیّت بیان می‌شود. انواع متغیرهای کمی شامل:

الف) پیوسته: به‌وسیله اعداد حقیقی بیان می‌شود؛ مانند سن (با اعشار نشان داده می‌شود)

ب)ناپیوسته: به‌وسیله اعداد طبیعی نشان داده می‌شود؛ . مانند تعداد فرزندان

متغیرهای کیفی: درباره حالات، خصوصیات و چگونگی‌ها بحث می‌کند.

---

[1] – مباحث برگرفته از کتاب روشهای آماری ، از مرحوم استاد دکتر کریم منصورفر و از انتشارات دانشگاه تهران است.

انواع سطوح اندازه‌گیری متغیرها

۱)اسمی nominal مثل مرد – زن ـ که در آن معرف‌ها یا مصادیق هر کدام از دیگری مستقل است و تعلق افراد بـه هر کدام از این مصادیق را نشان می‌دهد.
۲)رتبه‌ای ordinal ـ که در آن مصادیق هر متغیر به صورت مرتب بیان می‌شود، مانند میزان علاقه به مهاجرت
۳) فاصله‌ای (Interval) که در آن صفر بکار رفته (یا مبدأ شروع) قراردادی است.
۴) نسبی (Ratio) که در آن صفر بکار رفته حقیقی است.

فراوانی frequency

تعداد مشاهدات هر متغیر را فراوانی می‌گویند. پس متغیرها دارای فراوانی‌های متفاوت هستند.

$f_i \leftarrow$ نمونه
$F_i \leftarrow$ جامعه

دامنه تغییر (Renge)

به تفاوت بین ماکزیمم (بیشینه) تا مینیمم (کمینه)، دامنه تغییر می‌گویند. مثلا اگر بخواهیم سن ۲۰ نفر از افراد یک کلاس را مطالعه نماییم و اعداد به شرح زیر باشند:

۱۹ – ۲۵ – ۲۲ – ۲۰ – ۱۸ – ۱۷ – ۲۲ – ۲۶ – ۳۲ – ۲۸ – ۲۵ – ۲۶
۲۷ – ۲۸ – ۲۹ – ۲۹ – ۲۵ – ۲۴ – ۲۳ – ۲۲

دامنه تغییر متغیر سن برای داده‌های فوق برابر است با: $R = ۳۲ - ۱۷ = ۱۵$

حال اگر بخواهیم این داده‌ها را در قالب طبقات نمایش دهیم، همیشه تعداد طبقات برابر است با:

$$\text{تعداد طبقات} = \frac{\text{دامنه تغییر}}{\text{فاصله طبقات}}$$

$x_{min} = ۱۷$
$x_{max} = ۳۲$

پس اگر بخواهیم داده‌ها را با فاصله ۵ سال در جدول نمایش دهیم، تعداد طبقات برابر است با ۳

$$\text{تعداد طبقات} = \frac{۱۵}{۵}$$

مثال: توزیع سن بیست نفر از افراد یک جامعه

CL سطوح طبقات	فراوانی مطلق	فراوانی نسبی	مرکز گروه	فراوانی تجمعی	درصد فراوانی	فراوانی تجمعی درصدی	چگالی فراوانی نسبی
۵-۱۰	۵	۰/۲۵	۷/۵	۵	۲۵	۲۵	۰/۰۵
۱۰-۱۵	۸	$\frac{۸}{۲۰} = ۰/۴$	۱۲/۵	۱۳	۴۰	۶۵	۰/۰۸
۱۵-۲۰	۷	$\frac{۷}{۲۰} = ۰/۳۵$	۱۷/۵	۲۰	۳۵	۱۰۰	۰/۰۷
مجموع فراوانی مطلق	$\sum_n f_i = ۲۰$	۱	—	—	۱۰۰		

فراوانی نسبی: نسبت فراوانی مطلق هر طبقه به کل

فراوانی نسبی وزن هر طبقه را نشان می‌دهد.

جمع فراوانی نسبی همیشه یک است.

فراوانی تجمعی در مبحث چارک‌ها و میانه هم استفاده می‌شود.

فراوانی تجمعی: جمع فراوانی مطلق هر طبقه با طبقه یا طبقات قبل

فراوانی تجمعی آخرین طبقه همواره برابر با n (حجم نمونه) است.

درصد فراوانی: فراوانی مطلق تقسیم بر کل × ۱۰۰

جمع درصد فراوانی همیشه ۱۰۰ است.

نکته: (درصد فراوانی تجمعی کاربرد زیادی دارد مثلاً: در طیف تور ستون و در مبحث ضریب ابهام گویه‌ها استفاده می‌شود درباره این مبحث در آینده سخن گفته خواهد شد)

چگالی فراوانی نسبی: تقسیم فراوانی نسبی هر طبقه (گروه) بر فاصله طبقاتی همان گروه

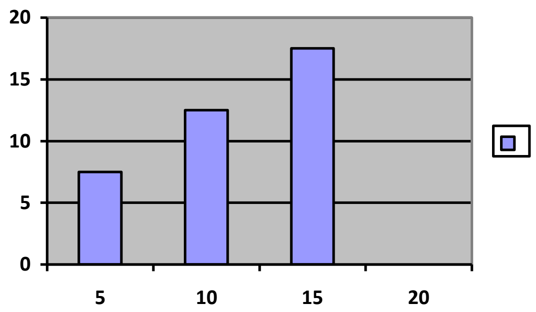
برخی نمودارها در توصیف آماری:

۱- هیستوگرام (مستطیلی) ← کمی پیوسته (این نمودار برای متغیرهای کمی پیوسته کاربرد دارد.)

۲- پولیگون (چند ضلعی) ← کمی ناپیوسته (این نمودار برای متغیرهای ناپیوسته کاربرد دارد.)

۳- اوجایو (ogive) ← این نمودار برای فراوانی‌های تجمعی کاربرد دارد.

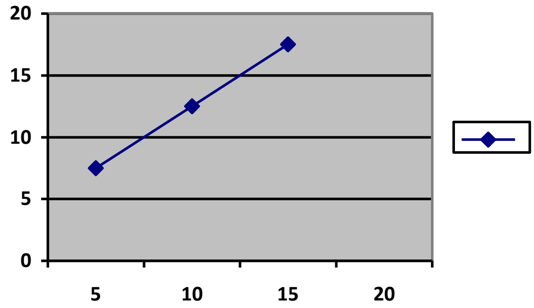
۴- دایره‌ای (Pie) ← برای متغیرهای در سطح اسمی و گاه رتبه‌ای کاربرد دارد.



نمودار هیستوگرام

نمودار میله‌ای فقط مرکز گروه را نشان می‌دهد.

نمودار اوجایو همواره صعودی است.



نمودار اوجایو (ogive)

برای اینکه فراوانی‌ها را در محیط دایره بسنجیم، لازم است به درجه تبدیل شوند و چون محیط دایره ۳۶۰ درجه است لذا:

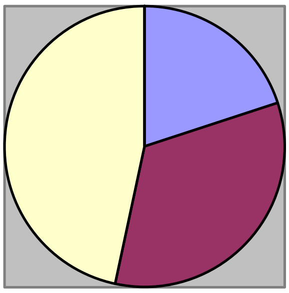
اگر n کوچک تر از ۳۶۰ بود، $\frac{۳۶۰}{n}$ می‌شود تا نشان دهد هر مشاهده در محیط دایره چند درجه است.

و اگر n بزرگتر از ۳۶۰ بود، $\frac{n}{۳۶۰}$ می‌شود.

مثلاً برای مسئله جدول قبل که تعداد کل افراد مورد مطالعه ۲۰ نفر بود، $۱۸ = \frac{۳۶۰}{۲۰}$ یعنی هر فرد روی محیط دایره ۱۸ درجه را شامل می‌شود.

مثلاً برای طبقه اوّل که سنین ۵ تا ۱۰ سال را شامل می‌شود و تعداد فراوانی‌های آن ۵ است، ۹۰ درجه را شامل می‌شود؛ چون $۵ \times ۱۸^\circ = ۹۰^\circ$

نمودار دایره‌ای



شاخص‌های مرکزی (مشخص‌کننده‌های مرکزی)

یعنی چند شاخص داریم که می‌خواهیم مرکز اینها را بدانیم؛ مثل میانگین، مد و میانه

۱- مُد (mode)

۲- میانه md یا me (median)

۳- میانگین $\bar{x}$ برای آماره و μ برای پارامتر (Average)

مد: مقداری از متغیر که دارای بیشترین فراوانی است (تعداد مشاهدات آن حداکثر است.).

به عنوان مثال، مد(نما) برای داده‌های زیر برابر است با:

$$۱۷-۱۸-۱۸-۱۸-۱۷-۱۵-۱۹-۱۶$$
$$MO=۱۸$$

چون عدد ۱۸ بیش از بقیه تکرار شده است.

برای محاسبه مقدار مد(نما) برای جدول زیر در گام اوّل لازم است ببینم کدام طبقه دارای بیشترین فراوانی است.

Cl	Fi	تجمعی
۵−۱۰	۵	۵
۱۰−۱۵	۸	۱۳
۱۵−۲۰	۷	۲۰

$d_۱$: تفاوت فراوانی طبقه دارای مد از طبقه قبل

$d_۲$: تفاوت فراوانی طبقه دارای مد از طبقه بعد

$$mo=L+h\left(\frac{d_۱}{d_۱+d_۲}\right)$$
$$mo=۱۰+۵\left(\frac{۳}{۳+۱}\right)=۱۳/۷۵$$

یعنی کسانی که ۱۳ سال و ۹ ماهه هستند یا ۱۳ سال و ۷۵ صدم سال سن دارند، بیشترین تکرار را دارند. در این مثال:

حتما مُد باید بین ۱۰ تا ۱۵ باشد یعنی در طبقه‌ای که مُد در آن قرار دارد.

میانه: مقداری از متغیر که تعداد مشاهدات یا فراوانی‌ها را به ۲ قسمت مساوی تقسیم می‌کند.

(برای داده‌های پراکنده) ۹-۶-۱۲-۵-۷-۵ برای محاسبه میانه، ابتدا لازم است داده‌ها به صورت صعودی یا نزولی مرتب شود.

۱۲-۹-۷-۶-۵-۵- در اینجا تعداد اعداد ۶ است پس میانه بین اعداد سوم و چهارم قرار دارد. پس میانگین اعداد سوم و چهارم را محاسبه می‌کنیم:

$$Me=\frac{۶+۷}{۲}=۶/۵$$

پس Me یا Md $\frac{n+۱}{۲}$

$$me=\frac{۶+۱}{۲}=۳/۵$$ یعنی بین عدد سوم و چهارم

(یعنی ۳ نفر بالای ۶/۵ سال سن دارند و ۳ نفر دیگر پایین ۶/۵ سال)

برای داده‌های طبقه‌بندی شده مقدار میانه برابر است با:

برای اینکه بدانیم میانه در کدام گروه یا طبقه قرار دارد، از فراوانی تجمعی استفاده می‌کنیم. چون تعداد کل افراد ۲۰ نفر است پس $\frac{n}{۲}=۱۰$ می‌شود لذا باید ببینیم نفر دهم در کدام طبقه قرار دارد. براساس جدول فوق، نفر دهم در طبقه دوم قرار دارد.

$$me = L + h\left(\frac{\frac{n}{۲}-f_a}{f_i}\right)$$

برای مثال جدول فوق میانه برابر است با:

$$me = ۱۰ + ۵\left(\frac{۱۰-۵}{۸}\right)$$

$$me = ۱۳/۱۲۵ \approx ۱۳$$

یعنی نصف افراد بین ۵ تا ۱۳ سال و نصف دیگر بین ۱۳ تا ۲۰ سال هستند.

انواع میانگین:

۱- میانگین حسابی (ساده) μ یا $\bar{x} = \frac{\sum f_i}{n}$

۲- میانگین وزنی؛ مثل محاسبه معدل با در نظر گرفتن واحد مقادیر صفت $\rightarrow \frac{\sum f_i xi}{n}$

۳- میانگین هندسی برابر است با: ریشه nام حاصل‌ضرب اعداد؛ مثلاً میانگین هندسی برای ۳ عدد ۷، ۵، ۴

$$\bar{x}_G = \sqrt[n]{x_۱ \times x_۲ \times \ldots \times x_n}$$

$$\sqrt[۳]{۴ \times ۵ \times ۷}$$

$$n = ۳$$

۴- میانگین هارمونیک $\leftarrow \bar{x}_H$

عکس میانگین هارمونیک n عدد برابر است با عکس تعداد، ضرب در مجموع عکس آن اعداد

$$\frac{۱}{\bar{x}_H} = \frac{۱}{n}\left(\frac{۱}{x_۱}+\frac{۱}{x_۲}+\ldots+\frac{۱}{x_n}\right)$$

مثال برای اعداد ۷ و ۵و ۴

$$\frac{۱}{\bar{x}_H} = \frac{۱}{۳}\left(\frac{۱}{۴}+\frac{۱}{۵}+\frac{۱}{۷}\right)$$

$$۴ \quad ۵ \quad ۷ \rightarrow n = ۳$$

$$\frac{۱}{۳}\times\left(\frac{۳۵+۲۸+۲۰}{۱۴۰}\right) = \frac{۱}{۳}\times\frac{۸۳}{۱۴۰} = \frac{۸۳}{۴۲۰} \rightarrow$$ عکس میانگین هارمونیک

$\bar{x}_h = \frac{۴۲۰}{۸۳} \simeq ۵/۰۶$ یعنی میانگین هارمونیک برابر است با

چارک‌ها:

۳ تا چارک داریم.

مثلاً ۲۵٪ افراد چه قدر از درآمد را دارند.

یا ۷۵٪ افراد در چه سنی قرار می‌گیرند.

مثال:

رابطه بین سن خریداران و تیراژ روزنامه

نکته: اگر سن خریداران روزنامه را از بزرگ به کوچک یا کوچک به بزرگ مرتب کنیم و اگر بخواهیم وضعیت سن خریداران روزنامه‌ها را مطالعه نماییم؛ در صورتی که سن مخاطبان را با هم جمع و بر تعداد تقسیم کنیم، میانگین سن مخاطبان به‌دست آمده است. در صورتی‌که سن مخاطبان را از کوچکترین تا بزرگترین مرتب نماییم و ببینیم ۵۰٪ اوّل آنها تا چه سنی را شامل می‌شوند، آن سن بیانگر میانه اعداد است و در صورتی‌که ببینیم در بین مخاطبان کدام مقطع سنی بیشترین تعداد را دارد، به mode یا نما رسیده‌ایم.

نکته: میانه نقطه ۵۰٪ است.

مثال: توزیع نمرات آزمون زبان ۸۰ نفر از دانشجویان

پس میانه همان چارک دوم است.

x_i	f_i	فراوانی تجمعی
۱۰−۲۰	۱۰	۱۰
۲۰−۳۰	۳۰	۴۰
۳۰−۴۰	۲۰	۶۰
۴۰−۵۰	۱۰	۸۰
	۸۰	

$$me = L + h\left(\frac{\frac{n}{۲} - f_a}{f_i}\right)$$ میانه

$$Q_۱ = L + H\left(\frac{\frac{n}{۴} - f_a}{f_i}\right)$$ چارک اوّل

$$Q_۱ = ۲۰ + ۱۰\left(\frac{۲۰-۱۰}{۳۰}\right) = ۲۳/۳$$ چارک اوّل ما

همانگونه که مشاهده می‌شود شیوه محاسبه برای چارک‌های اول، دوم(میانه) و سوم یکی است فقط در داخل پرانتز برای محاسبه چارک اول از $\frac{n}{۴}$ برای چارک دوم از $\frac{n}{۲}$ و چارک سوم از $\frac{۳n}{۴}$ استفاده می‌شود.

نمره یک چهارم افراد از ۱۰ تا ۲۳/۳ شده است. یعنی ۲۵ درصد اوّل (۲۰نفر) دارای نمره ۱۰ تا ۲۳/۳ شده‌اند.

چارک سوم $Q_۳ = L + h\left(\frac{\frac{۳n}{۴} - fa}{f_i}\right)$

$$Q_۳ = ۳۰ + ۱۰\left(\frac{۶۰ - ۴۰}{۲۰}\right) = ۴۰$$

یعنی نمره ۷۵٪ از افراد آزمون دهنده بین ۱۰ تا ۴۰ است و مابقی (۲۵ درصد باقیمانده) بین ۴۰ تا ۵۰ می‌باشد.

انحراف چارکی: هرچه فاصله چارک اوّل تا سوم کمتر باشد، داده‌ها همگن‌تر است. انحراف چارکی به تفاضل چارک سوم از چارک اوّل اشاره دارد. و در واقع یکی از شاخص‌های مطالعه پراکندگی است.

انحراف چارکی = $\frac{Q_۳ - Q_۱}{۲}$

خواص میانگین:

در بررسی خواص میانگین می‌خواهیم بدانیم که اگر مقادیر صفت x_i یا فراوانی‌ها f_iها را دستکاری کنیم، میانگین چه تغییری می‌کند؟

خاصیت اوّل: اگر ۴ عمل اصلی را روی مقادیر صفت اعمال کنیم (بر عددی تقسیم یا در آن ضرب یا با آن جمع یا از آن تفریق کنیم)، میانگین به همان نسبت تغییر می‌کند.

$$\begin{cases} ۴.۶.۸ \\ \bar{x} = ۶ \end{cases}$$

اگر همه را تقسیم بر ۲ کنیم، می‌شود اعداد: ۴، ۳، ۲ به دست می‌آید.

پس میانگین هم می‌شود ۳ یعنی تقسیم بر ۲ می‌شود. $\bar{X} = \frac{۲+۳+۴}{۳} = ۳$

خاصیت دوم: اگر فراوانی‌ها را بر عددی ثابت تقسیم، ضرب، جمع یا کسر کنیم، میانگین تغییری نمی‌کند.

مثلاً در اینجا فراوانی‌ها در ۲ ضرب شده است.

xi	fi	fi×xi	fi×۲	fi×xi
۱۵	۲	۳۰	۴	۶۰
۱۸	۳	۵۴	۶	۱۰۸
۲۰	۵	۱۰۰	۱۰	۲۰۰
	$\sum f_i = ۱۰$	۱۸۴	$\sum f_i = ۲۰$	۳۶۸

$\bar{x} = \frac{۱۸۴}{۱۰} = ۱۸/۴$ $\bar{x} = \frac{۳۶۸}{۲۰} = ۱۸/۴$

مبحث پراکندگی‌ها

الف ـ انحراف متوسط[1] (A.D) یکی از شاخص‌های پراکندگی است که در آن مجموع قدرمطلق تفاوت مقادیر صفت از میانگین محاسبه می‌شود.

$$\bar{x}=\frac{۴+۶+۱۰+۸}{۴}=۷$$

نکته: مجموع انحرافات مقادیر صفت از میانگین همیشه صفر است.

حال برای اینکه صفر نشود، قدرمطلق این انحراف را در نظر می‌گیریم؛ یعنی بدون علامت

فرمول انحراف متوسط :

$$A.D=\frac{|x_۱-\bar{x}|+|x_۲-\bar{x}|+\ldots+|x_n-\bar{x}|}{n}$$

$$A.D=\frac{|۴-۷|+|۶-۷|+|۱۰-۷|+|۸-۱۰|}{۴}$$

$$AD=\frac{۸}{۲}=۴$$

یعنی هرکدام از افراد به طور متوسط ۲ نمره با میانگین اختلاف دارند.

ب) واریانس: برای اینکه تفاضل مقادیر صفت از میانگین صفر نشود، می‌توانیم آنها را به توان ۲ برسانیم.

واریانس $S^۲=\frac{(۴-۷)^۲+(۶-۷)^۲+(۱۰-۷)^۲+(۸-۷)^۲}{۴}=\frac{۲۰}{۴}=۵$

نکته: به میانگین قدر مطلق انحرافات می‌گوییم انحراف متوسط

نکته: به میانگین مجذور انحرافات می‌گوییم واریانس = $S^۲$

انحراف معیار $S=\sqrt{s^۲}=\sqrt[۲]{۵}=۲/۲$

انحراف معیار / جذر واریانس است.

واریانس نمونه $S^۲$ واریانس جامعه V

$$S^۲=\frac{\sum(x_۱-\bar{x})^۲+(x_۲-\bar{x})^۲+\ldots(x_n-\bar{x})^۲}{n}$$

نکته: واریانس = مجموع مجذور انحراف مقادیر صفت از میانگین بر تعداد افراد

واریانس برای محاسبه پراکندگی بهترین است.

نکته: نسبت انحراف معیار به انحراف متوسط معمولاً $۱/۲۵\le\frac{S}{A.D}<۱$ است.

(همیشه انحراف معیار از انحراف متوسط بزرگتر است.)

[1] - Average deviation

واریانس در تمام محاسبات آماری کاربرد دارد؛ چون شاخص پراکندگی است و هرچه به صفر نزدیک‌تر باشد؛ یعنی پراکندگی کمتر است.

خواص واریانس:

۱- اگر از تمامی مقادیر صفت یک مقدار ثابت مانند A کسر یا به آن اضافه کنیم، مقدار واریانس تغییری نمی‌کند. (برعکس میانگین)

$$X = ۲,۴,۶$$
$$\bar{x} = ۴$$
$$S^۲ = \frac{(۲-۴)^۲ + (۴-۴)^۲ + (۶-۴)^۲}{۳}$$
$$S^۲ = \frac{۸}{۳} = ۲/۶۶$$

حال اگر به مقادیر صفت ۳ واحد اضافه نماییم، می‌شود:

۵ ۷ ۹

$\bar{x} = ۷$ میانگین

$$s^۲ = \frac{۸}{۳} = ۲/۶۶$$

و مجدداً واریانس همان ۲/۶۶ می‌شود.

۲- اگر تمامی مقادیر صفت را بر مقدار ثابتی مانند A تقسیم یا در آن ضرب نماییم، در صورت تقسیم واریانس متغیر اصلی تا $A^۲$ برابر کوچکتر (یعنی تقسیم بر $A^۲$ برابر می‌کنیم.) و در صورت ضرب $A^۲$ برابر بزرگتر می‌شود؛ در ابتدا مقادیر صفت را در عدد ۲ ضرب می‌کنیم می‌شود .

$$۴,۸,۱۲ \qquad \bar{X} = ۸$$
$$B^۲ = \frac{(۴-۸)^۲ + (۸-۸)^۲ + (۱۲-۸)^۲}{۳} = ۱۰/۶۶$$

پس مشاهده می‌شود که واریانس ۴ برابر بزرگتر شد.

حال مقادیر صفت را بر عدد ۲ تقسیم کنیم:

$$S^۲ = ۲/۶۶$$
$$a = ۲ \qquad a^۲ = ۲^۲ = ۴ \qquad x = ۱ \quad ۲ \quad ۳ \qquad s^۲ = \frac{(۱-۲)^۲ + (۲-۲)^۲ + (۳-۲)^۲}{۳} = ۰/۶۶$$
$$S^۲ = \frac{۲/۶۶}{۴} = ۰/۶۶$$

پس واریانس $a^۲$ برابر یعنی $(۲^۲)$ برابر کوچکتر می‌شود.

۳- اگر تمام فراوانی‌ها را در یک عدد ثابت تقسیم یا ضرب کنیم، واریانس تغییری نمی‌کند.

۴-اگر چند جامعه با واریانس‌ها و میانگین‌های مختلف داشته باشیم، برای محاسبه واریانس کل لازم است هم واریانس بین گروهها و هم واریانس درون گروهها را محاسبه نماییم.

برای ۴ کلاس روش تحقیق، میانگین و واریانس به شرح زیر است:

جامعه A $\bar{x}_A = ۱۵$ $S_A^۲ = ۲۰$

جامعه B $\bar{x}_B = ۱۴$ $S_B^۲ = ۳۰$

جامعه C $\bar{x}_C = ۱۲$ $S_C^۲ = ۴۰$

جامعه D $\bar{x}_D = ۱۱$ $S_D^۲ = ۴۰$

کدام کلاس کمترین پراکندگی را دارد؟

در کلاس A واریانس پایین‌تر است؛ لذا پراکندگی کمتری دارد.

اما بیشترین پراکندگی در کلاس C و D است.

سؤال بعدی: واریانس کل چقدر است؟

برای محاسبه واریانس کل باید آن را به عوامل اولیه تجزیه کرد.

$$S_t^۲ = S_B^۲ + S_w^۲$$

در تجزیه واریانس با ۲ واریانس سروکار داریم:

- واریانس بین گروه‌ها یا همان واریانس میانگین‌ها
- واریانس درون گروه‌ها یا همان میانگین واریانس‌ها

میانگین واریانس‌ها یا واریانس درون گروه‌ها $S_w^۲ = \frac{s_۱^۲ + s_۲^۲ + \ldots + s_n^۲}{n} \Rightarrow s_w^۲ = \frac{۲۰+۳۰+۴۰+۴۰}{۴} = ۳۲/۵$

اما برای محاسبه واریانس میانگین‌ها لازم است که اوّل میانگین کل محاسبه گردد:

$$\bar{x}_t = \frac{۱۵+۱۴+۱۲+۱۱}{۴} = \frac{۵۲}{۴} = ۱۳$$

میانگین نمره هر کلاس را از میانگین کل کم می‌کنیم و اعداد به‌دست آمده را جمع کرده، تقسیم بر تعداد می‌کنیم:

$$S_B^۲ = \frac{(۱۵-۱۳)^۲ + (۱۴-۱۳)^۲ + (۱۲-۱۳)^۲ + (۱۱-۱۳)^۲}{۴}$$

واریانس بین گروهها $s_B^۲ = ۲/۵$

$$S_B^۲ = \frac{(\bar{x}_۱ - \bar{x}_t) + (\bar{x}_۲ - \bar{x}_t) + \ldots + (\bar{x}_n - \bar{x}_t)^۲}{n}$$

واریانس کل $\Rightarrow s^۲t = ۲/۵ + ۳۲/۵ = ۳۵$

آنالیز واریانس، بحث بسیار مهمی است.

آزمون F مقایسه میانگین‌ها برای چند گروه است و متأثر از آنالیز واریانس است.

واریانس بین گروه‌ها تقسیم بر واریانس درون گروه‌ها مقدار F را می‌دهد. (درباره F بعداً بحث خواهد شد.)

در نمونه‌گیری طبقه‌ای و خوشه‌ای هم آنالیز واریانس کاربرد دارد.

اندازه نسبی پراکندگی: مقایسه پراکندگی ۲ متغیر با ۲ واحد مختلف است. در واقع، شاخص دیگری است که پراکندگی ۲ یا چند متغیر را می‌خواهیم مطالعه کنیم که واحد آنها با هم برابر نیست؛

مثلاً درآمد ← ریال وزن ← کیلو محاسبه می‌شود.

$R.D = \frac{S}{\bar{x}} \times ۱۰۰$ اندازه نسبی پراکندگی

مثال:

انحراف معیار وزن گروهی از افراد ۱۰و میانگین وزن آنها ۲۰ است، همچنین انحراف معیار درآمد افراد ۵ و میانگین درآمد آنها ۲۰ است کدام صفت همگن‌تر است؟

میانگین وزن آنها = ۲۰

وزن $= \frac{۱۰}{۲۰} \times ۱۰۰ = \%۵۰$

درآمد $= \frac{۵}{۲۰} \times ۱۰۰ = \%۲۵ \rightarrow$ به صفر نزدیک‌تر است.

انحراف معیار درآمد =۵

میانگین درآمد = ۲۰

چون شاخص اندازه نسبی پراکندگی درآمد کمتر از پراکندگی وزن است؛ لذا همگن‌تر هستند و از نظر درآمدی بیشتر شبیه هم هستند.

نمرات استاندارد[۱]

نمرات درصدی به صورت ترتیبی بیان می‌شود و در صورت توزیع درصدی نمی‌توان گفت که با تغییر نمره درصدی همان تغییر در نمره خام ایجاد می‌شود لذا نمرات اگر به نمرات استاندارد تبدیل شود این مقایسه امکانپذیر است. نمرات استاندارد وضعیت نمرات را نسبت به میانگین نشان میدهند و دارای دو ویژگی هستند.

۱- از نمرات خام بدست می‌آیند.

۲- با مقیاس فاصله‌ای هستند پس همه محاسبات ریاضی در آن امکان‌پذیر است.

$$Z = \frac{X - \bar{X}}{S}$$

مثال: نمرات خام ۲۵و ۲۰و ۱۵ و۱۰و ۵

برای اینکه این نمرات به نمرات استاندارد تبدیل شوند باید میانگین و واریانس و سپس انحراف معیار آنها محاسبه گردد.

$\bar{X} = ۱۵ \quad S^۲ = ۵۰ \quad S = ۷/۰۷$

نکته۱: اگر نمره خام بالای میانگین باشد نمره Z مثبت می‌شود.

نکته۲: اگر نمره خام پایین میانگین باشد نمره Z منفی می‌شود.

[۱] برگرفته از کتاب احتمالات و آمار کاربردی، تألیف استاد دکتر علی دلاور است.

مثلاً برای آنکه میزان انحراف از میانگین را در منحنی نرمال برای نمره ۱۲ تا میانگین بدانیم به شیوه مقابل عمل می‌کنیم.

$$Z_{۱۲}=\frac{۱۲-۱۵}{۷/۰۷}=-۱/۴۱$$

چون عدد ۱۲ از میانگین کوچکتر است مقدار Z منفی می‌شود.

نکته: یکی از موارد استفاده از نمره Z مقایسه بین دو نمره خام است که شاید ظاهراً با هم فرق نکنند ولی وقتی استاندارد شوند تفاوت داشته باشندو بالعکس.

منحنی نرمال

براساس نمرات Z منحنی طبیعی استاندارد وتشریح می‌گردد ویژگیهای منحنی نرمال:

۱- متقارن است و حداکثر ارتفاع آن در میانگین قرار دارد.

۲- میانگین، میانه و نما با هم برابرند.

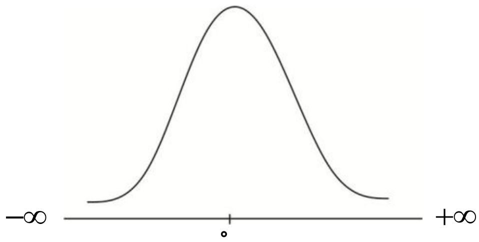
۳- منحنی دو نقطه عطف دارد که نسبت به خطی که نشانگر میانگین، میانه و نما است قرینه است.

۴- دنباله‌های منحنی با محور X موازی است پس از $-\infty$ (منهای بینهایت) تا $+\infty$ (مثبت بینهایت) ادامه دارد اما در عمل بین ۳- تا ۳+ انحراف استاندارد است.

ممکن است دو یا چند توزیع طبیعی باشند ولی میانگین و انحراف استاندارد آنها متفاوت باشد و همچنین منحنی‌هایشان با هم فرق کند.

خواص منحنی نرمال:

منحنی نرمال از توزیع داده‌ها بحث می‌کند. از خواص منحنی نرمال برای تحلیل داده‌های اجتماعی استفاده می‌شود.



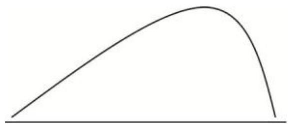
اگر توزیع نرمال باشد:

نکته: میانگین، میانه و نما با هم برابرند.

نکته: در توزیع نرمال میانگین صفر و واریانس یک است $S^{۲}=۱$

Skewness چولگی بیان کننده مقدار انحرافی است که منحنی به طرفین دارد.

Kurtosis کشیدگی (از ارتفاع بحث می‌کند.) $\bar{X}<Me<Mo$



انواع چولگی:

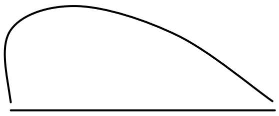
۱) چولگی منفی

(z منفی است.)

$\bar{x}<me<mo$

۲) $Mo<Me<\bar{x}$

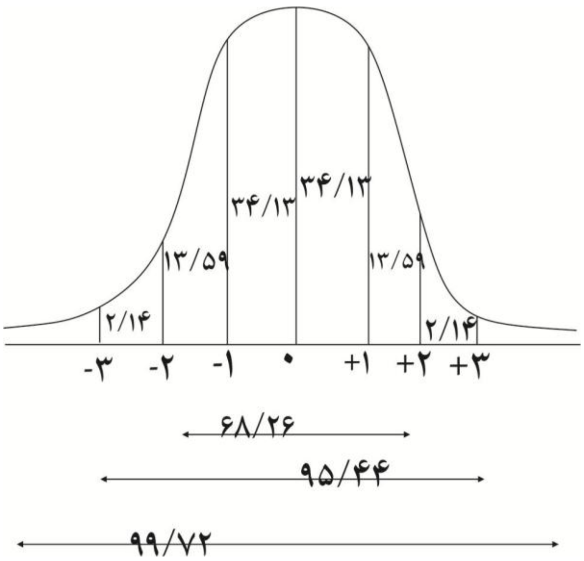
چولگی مثبت (z مثبت است.)



$Mo < Me < \bar{X}$

داده استاندارد $z = \frac{x-\bar{x}}{s}$ انحراف معیار

(از خواص منحنی نرمال برای آزمون فرضیه‌ها نیز استفاده می‌کنیم.) مقدار z با ۹۰درصد اطمینان برای بررسی همانندی یا ناهمانندی دو مؤلفه (مانند دو متغیر یا جامعه با نمونه)، ۱/۶۸ و با ۹۵ درصد ۱/۹۶ و با ۹۹ درصد اطمینان ۲/۵۸ است.



پس از استاندارد کردن نمرات خام، عدد بدست آمده مقدار انحراف استاندارد از میانگین را نشان می‌دهد برای پیداکردن مساحت زیر منحنی باید به جدول سطح زیر منحنی نرمال مراجعه شود.

مثال: توزیع نمرات درس آمار ۷ نفر از دانشجویان

۱۰–۱۲–۱۳–۱۵–۸–۱۲–۱۴

$\bar{X} = \frac{۸۴}{۷} = ۱۲ \qquad S^۲ = ۴/۸۵$

$S = ۲/۲$

برای پیداکردن این نکته که چند درصد افراد نمرهای بین ۸ تا میانگین دارند به روش زیر اقدام می‌شود.

$Z_۸ = \frac{۸-۱۲}{۲/۲} = \frac{-۴}{۲/۲} = ۱/۸۱$

سپس جدول سطح زیر منحنی نرمال مراجعه می‌کنیم.

در این جدول پنج ستون وجود دارد که ستون اول مقدار **Z** را نشان می‌دهد ستون دوم (**A**) سطح تا میانگین ستون سوم (**B**) سطح بزرگتر (یعنی سطح تا میانگین +۰/۵) ستون چهارم (**C**) سطح کوچکتر (یعنی فاصله مقدار **Z** تا میانگین منهای ۰/۵) و (**y**) ارتفاع را نشان می‌دهد.

برای مثال فوق مقادیر برابر است با

Z	A	B	C	y
۱/۸۱	۰/۴۶۴۹	۰/۹۶۴۹	۰/۰۳۵۱	۰/۰۷۷۵

طبق جدول فوق اگر اعداد را در عدد ۱۰۰ ضرب کنیم نشان می‌دهد که:

۴۶/۴۹ درصد افراد نمره‌ای بین ۸ تا ۱۲ دارند.

۹۶/۴۹ درصد افراد نمره‌ای بین ۸ تا بالاترین دارند.

۳ درصد افراد نمره‌ای کمتر از ۸ دارند.

مساحت ارتفاع حدود ۷ درصد است.

روش‌های تحلیل آماری:

الف) کمی

۱- همبستگی ← هم تغییری

۲- وابستگی

۳- تفاوت آماری

ب) کیفی

آمار استنباطی

شامل: همبستگی، وابستگی، تفاوت آماری و آزمون‌های آماری

زمانی از آمار استنباطی بحث می‌کنیم که بخواهیم:

- روابط بین متغیرها را بررسی کنیم.
- استنباط پارامترها از طریق آماره‌ها را مطالعه کنیم.

اگر فرضیه‌ها را به ۲ قسمت فرضیه برای روش‌های کمی و فرضیه برای روش‌های کیفی تقسیم کنیم، فرضیه‌های کمی به ۳ دسته تقسیم می‌شوند:

- فرض همبستگی
- فرض وابستگی
- فرض تفاوت آماری

مثال:

- بین سن و مهاجرت همبستگی وجود دارد.
- بین سن و مهاجرت وابستگی وجود دارد.
- بین سن و مهاجرت رابطه (تفاوت) معنادار وجود دارد.

❊ در بحث همبستگی بررسی می‌کنیم که آیا با تغییر y, x هم تغییر می‌کند یا بالعکس. مبنای ضریب همبستگی کوواریانس است؛ یعنی واریانس دوطرفه، در بحث از هم تغییری تفاوتی ندارد x اوّل تغییر کند یا y. رابطه اینها ۲ طرفه است.

برای آزمون همبستگی از ضرایب همبستگی استفاده می‌کنیم. $y \underset{\leftarrow}{\overset{\rightarrow}{}} x$

ضریب همبستگی را با r نشان می‌دهیم.

مثلاً با تغییر سن، مهاجرت تغییر می‌کند و برعکس

$$r : \frac{\text{cov}\,xy}{\sqrt{\frac{ssx.ssy}{n}}}$$

* در بحث وابستگی؛ یک متغیر مستقل است و یک متغیر وابسته. برای سنجش وابستگی از آزمون‌های خاص خودش استفاده می‌کنیم. (فرضیه‌های ما ادعاهای ما هستند.)

در وابستگی، واریانس یک طرفه است. ← یک رابطه متقابل داریم.

* در بحث از تفاوت یا رابطه، ما ۲ متغیر را با هم مقایسه می‌کنیم.

مثال: آیا ساعات ورزش زنان و مردان یکسان است (وقتی بگوییم زنان ۲/۵ ساعت ورزش می‌کنند و مردان ۲ ساعت یعنی زنان نسبت به مردان بیشتر ورزش می‌کنند؛ اما آیا این نیم ساعت تفاوت قابل چشم‌پوشی است یا معنادار است و نیز قابل اغماض پس اگر تفاوت‌ها از یک حدی بیشتر شود، می‌گوییم تفاوت آماری معنادار است.)

درجه آزادی df حدی است که براساس آن حد، فرضی را رد یا قبول می‌کنیم. در واقع تفاوت همان رابطه است.

در مثال فوق اگر تفاوت معنادار بود، یعنی زنان و مردان(جنسیت) از نظر ساعات ورزش یکسان نیستند. یا زنان و مردان از نظر میزان مطالعه با یکدیگر تفاوت دارند.

نکته: تفاوت همان رابطه است.

میزان درآمد و میزان مسافرت خارجی ←اگر معنادار شود، یعنی همه گروه‌های درآمدی از نظر میزان مسافرت خارجی تفاوت معنادار دارند. گروه‌های درآمدی از نظر میزان مسافرت خارجی با هم یکسان نیستند.

فرض صفر(H_0) می‌گوید: بین ۲ متغیر رابطه (تفاوت) معنادار وجود ندارد.

فرض مقابل (خلاف) (H_1) می‌گوید: بین ۲ متغیر تفاوت (رابطه) معنادار وجود دارد.

نکته: درجه آزادی $df = n - 1$

نکته: برای جداوّل ۲ بعدی $df = (R-1)(C-1)$

برای پیداکردن معنادار بودن یا معنادار نبودن یک رابطه باید به جدول آماری کتابهای آمار مراجعه کنیم. (برای هر آزمون جداگانه جداولی وجود دارد.)

حداکثر خطای قابل پذیرش ۰/۰۵ است.اگر عدد محاسبه شده ما بزرگتر از عدد جدول بود، بعد می‌رویم سراغ ۰/۰۱ خطا، آنگاه اگر از آن هم بزرگتر بود، با ۹۹ درصد اطمینان و خطای ۱ درصد قضاوت می‌کنیم.

نکته: معمولاً اگر عدد محاسبه شده از عدد جدول بزرگتر باشد ، فرض صفر رد می‌شود یعنی تفاوت معنادار وجود دارد.

نکته: اگر عدد محاسبه شده کوچکتر از عدد جدول بود، تفاوت ندارند و فرض صفر قبول می‌شود.

البته در ۲ آزمون وضعیت عکس است که سپس توضیح داده خواهد شد.

ما در آزمون معناداری مرتکب ۲ نوع خطا می‌شویم:

۱- خطای نوع اوّل α (آلفا)

زمانی است که فرض صفر درست باشد؛ اما ما به اشتباه آن را رد کنیم یا فرض صفر صحیح اشتباهاً رد شود.

۲- خطای نوع دوم β (بتا)

فرض صفر غلط اشتباهاً پذیرفته شود.

کدام‌یک از این خطاها زیان‌بارتر است؟

خطای نوع اوّل ←چون فرض صفر درست است و ما اشتباهاً فرض یک را تأیید می‌کنیم. فرض یک پایه پژوهش ماست. (به آن فرضیه پژوهش گفته می‌شود.)

اصطلاح دیگری داریم به نام توان آزمون

توان آزمون به ما می‌گوید که چگونه احتمال خطای نوع اوّل را کاهش بدهیم.

ضرایب همبستگی:

تعداد ضرایب همبستگی زیاد است که با توجه به سطوح اندازه‌گیری متغیرها مشخص می‌شود. بدین معنا که:

اگر متغیرها در یک سطح باشند، مشکلی نیست اما ما ۴ سطح اندازه‌گیری برای متغیرها داریم:

۱- سطح اسمی: ابتدایی‌ترین سطح است. nominal scale که در آن، بین گزینه‌ها اوّلویتی وجود ندارد: مرد – زن

هر فرد از جامعه مورد مطالعه به یکی از این گزینه‌ها (متغیرها) تعلق دارد: مثل جنسیت ، دین ، گروه خونی

۲- سطح ترتیبی (ordinal): بین گزینه‌ها اوّلویت وجود دارد؛ مانند میزان علاقه به مهاجرت (خیلی کم – کم – متوسط – زیاد – خیلی زیاد)

از کوچک یا بزرگ یا بالعکس قابل مرتب کردن است.

۳- سطح فاصله‌ای (interval) : که در آن صفر قراردادی وجود دارد. نقطه شروع یا صفر را پژوهشگر مشخص می‌کند.

۴- دقیق‌ترین سطح نسبی است (Ratio)؛ یعنی صفر واقعی دارد .

متغیرها از نسبی به فاصله‌ای و ترتیبی و اسمی قابل تغییر هستند.

فاصله‌ای و نسبی کمیت‌پذیر هستند.

مثال: چندتا فرزند دارید؟ ← نسبی اما وقتی بخواهیم ۳ فرزند و بالاتر را بسنجیم ، پس ۳ را صفر در نظر می‌گیریم.

سؤال: کدام مقیاس نسبت به عمل ضرب نامتغیر است؟ اسمی

ضرایب همبستگی:

الف ـ ضریب همبستگی پیرسون:

مفروضه‌های این ضریب ۴ تاست:

۱- سطح اندازه‌گیری، فاصله‌ای و نسبی باشد.

۲- رابطه خطی باشد.

۳- واریانس دو توزیع تقریباً یکسان باشد.

۴- هر کدام از دو توزیع پیش از یک نما نداشته باشد.

$$r=\frac{cov(xy)}{\delta x\delta y}$$

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

مثال:

x	y
۲	۵
۴	۴
۹	۳

اگر y,x هم‌سو باشند یعنی هر دو زیاد شوند یا هر ۲ کم شوند، می‌گوییم همبستگی مثبت / مستقیم است.

اگر جهت تغییرات y,x ناهمسو باشد، همبستگی منفی یا معکوس است. اگر آهنگ تغییرات در y,x با یک فاصله ثابت انجام شود، آن موقع هم‌بستگی کامل است. (برعکس آن همبستگی کامل نیست.)

لزوماً مقدار فاصله ثابت در تغییرات y,x نیاز نیست یکی باشد: مثلاً y می‌تواند ۵۰۰ تا ۵۰۰ تا تغییر کند و x یکی – یکی

اگر همبستگی کامل و مثبت (مستقیم) باشد: $r=1$

اگر همبستگی کامل باشد و منفی (معکوس)، $r=-1$

اگر همبستگی ناقص و مثبت (مستقیم) باشد، $0<r<1$

اگر همبستگی ناقص یا منفی باشد، $-1<r<0$

x تغییر کند، y تغییر نکند. $r=0$

مثال:

$y=$ درآمد ۱۰۰ هزارتومان $x=$ تعداد فرزندان

x	y	xy	x^2	y^2
۲	۵	۱۰	۴	۲۵
۴	۴	۱۶	۱۶	۱۶
۶	۳	۱۸	۳۶	۹
۱۲	۱۲	۴۴	۵۶	۵۰

معکوس است.

در اینجا ما نیاز به محاسبه نداریم. اینجا با یک نگاه درمی‌یابیم که همبستگی کامل و منفی است؛ یعنی $r=-1$ است.

چون x با فاصله ثابت ۲ تا ۲ تا افزایش می‌یابد و y با فاصله ثابت یک کاهش می‌یابد. حال برای اطمینان مقدار ضریب همبستگی را محاسبه می‌نماییم:

$$r=\frac{۴۴-\frac{۱۲\times ۱۲}{۳}}{\sqrt{(۵۶-\frac{(۱۲)^۲}{۳})(۵۰-\frac{(۱۲)^۲}{۳})}}=\frac{-۴}{۴}=-۱$$

ضریب همبستگی کامل و منفی(معکوس) است.

نکته ۱:

اگر تعدادی عدد با فاصله برابر داشته باشیم،

میانگین آنها برابر است با:

الف ـ اگر تعداد فرد باشد، عدد وسط همیشه میانگین است.

ب ـ اگر تعداد زوج باشد، میانگین دو عدد وسط بیانگر میانگین آن اعداد است.

مثال: ۱۰۱—۱۰۲—۱۰۳—۱۰۴—۱۰۵

میانگین $\bar{x}=۱۰۳$

واریانس $s^۲=۲$

نکته ۲: اگر چند گروه داشته باشیم که شبیه هم باشند، واریانس کل برابر چند است؟

پاسخ: وقتی می‌گوییم شبیه هم هستند یعنی واریانس بین گروه‌ها صفر است. پس واریانس کل برابر با واریانس درون گروهها است.

واریانس درون گروه + واریانس بین گروه = واریانس کل

واریانس درون گروه‌ها + ۰ = واریانس کل

موارد استفاده ضرایب همبستگی:

۱-پیرسون ← فاصله‌ای یا نسبی

۲-اسپیرمن ← رتبه‌ای و حجم بالا

۳-کندال ← رتبه‌ای باشد اما حجم نمونه کم باشد.

❋ در منحنی نرمال حداقل نمونه قابل قبول چند است؟ ۳۰ نفر

۴-گاما ← رتبه‌ای باشد. – توزیع متقارن

اینها را اگر برحسب نمره ریاضی‌شان مرتب کنیم:

ریاضی	۱۷	۱۸		۱۵	۱۵	۱۵
آمار	۱۰	۲۰				

متقارن یعنی: رتبه‌هایی که به اینها دادیم از نظر توزیع y مثل x باشد.

۵-سامرز D ← رتبه باشد و توزیع نامتقارن

۶- فی(phi) ← اسمی باشد.

ضرایب وابستگی

برای فرضیه‌هایی وابستگی استفاده می‌شود.

اگر فاصله و نسبی بود ← تحلیل مسیر

اگر رتبه بود ← گودمن – کروسکال (کراسکال)

اگر اسمی بود ← لاندا (لامبدا) lambda

آزمون‌های آماری: براساس سطح اندازه‌گیری متغیرها تقسیم می‌شوند به:

۱-آزمون‌های پارامتریک که برای سطوح فاصله‌ای و نسبی استفاده می‌شود.

۲-آزمون‌های ناپارامتریک که برای سطوح اسمی و ترتیبی استفاده می‌شود.

(بین شاخصهای مرکزی میانگین دقیق‌ترین است.)

نام دیگر آزمون‌های پارامتریک آزمون‌های مقایسه میانگین‌ها است. آزمونهای پارامتریک شامل ۴ آزمون هستند:

آزمون Z (u) برگرفته از بحث منحنی نرمال است. (رجوع به خواص منحنی نرمال)

برای حجم بالای ۳۰ نفر مورد استفاده قرار می‌گیرد. $N > ۳۰$ در این آزمون میانگین به‌دست آمده با میانگین ادعا شده مورد مقایسه قرار می‌گیرد.

اگر بخواهیم با ۹۰ درصد اطمینان فرضیه را آزمون کنیم، مقدار $z = ۱/۶۸$ است.

مقدار z	درصد خطا	درصد اطمینان	مقدار انحراف استاندارد
۱/۶۸	۱۰	۹۰	۱
۱/۹۶	۵	۹۵	۲
۲/۵۸	۱	۹۹	۳

جداوّل آزمون‌های آماری(پارامتریک و ناپارامتریک)

نام آزمون	پایه و اساس	اشکالات	مزایا	کاربرد	فرمول(نحوه قضاوت)	ملاحظات
الف- پارامتریک	اندازه‌گیری مقیاس‌های فاصله‌ای و نسبی	برای همه متغیرها قابل استفاده نیست.	پایدارترین و دقیق‌ترین آزمون‌ها است.	متغیرهای کمی که پارامترهای مرکزی در آن واقعی است.		
۱- آزمون z (u)	مقایسه بین میانگین نمونه و میانگین فرضی $\mu A-\mu n$	وقتی حجم نمونه کمتر از ۳۰نفر باشد، بر منحنی نرمال صدق نمی‌کند.	میزان نرمال‌بودن توزیع متغیرهای تصادفی و کمی را برآورد می‌کند.	آزمون‌های پارامتریک	$Z=u=\dfrac{\lvert m-\mu\rvert}{\dfrac{\sigma}{\sqrt{n}}}$ اگر $U>1/96$ باشد، H_0 رد می‌شود.	
۲- آزمون (t) حالت اوّل	مقایسه بین میانگین نمونه (m) و میانگین جامعه (μ)	اگر بیش از دو جامعه یا نمونه داشته باشیم، جواب نمی‌دهد.	برای نمونه‌های کمتر از ۳۰ نفر کاربرد دارد.	مقایسه یکسانی یا عدم یکسانی دو جامعه	$T=\dfrac{d}{S_d} \quad df=n-1$ $d=\lvert m-\mu\rvert$ $S_z=\dfrac{\sqrt{ssx}}{(n-1)}$ اگر $t_c<t_t$ باشد، H_0 پذیرفته است.	
۳- آزمون (t) حالت دوم	مقایسه دو میانگین از دو جامعه m_1 , m_2	فقط دو جامعه را مقایسه می‌کند نه بیشتر.	برای نمونه‌های کمتر از ۳۰ نفر کاربرد دارد.	بررسی یکسانی دو میانگین $n<30$	$d=\bar{x}_1-\bar{x}_2$ $df=n_1+n_2-2$ $S_d^2=\dfrac{ssx_1+ssx_2}{(n_1+n_2-2)}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)$ $ssx=\sum xi^2 \quad \dfrac{\left(\sum i\right)^2}{N}$	اگر واریانس حقیقی دو جامعه یکسان نباشد، از روش زیر استفاده می‌شود: $s_d^2=\dfrac{ssx_1}{n_1(n_1-1)}+\dfrac{ssx_2}{n_2(n_2-1)}$ اگر آزمون یک دامنه باشد، t ۵ درصد به t ده درصد تبدیل می‌شود. $(t0/1)$

* جدول فوق توسط نگارنده در دوران تحصیل در مقطع دکتری در درس آمار پیشرفته و به توصیه استاد ارجمندم مرحوم دکترکریم منصورفر (که همیشه یادشان را گرامی می‌داریم.) تهیه و تأیید گردید.

جداوّل آزمون‌های آماری (پارامتریک و ناپارامتریک)

نام آزمون	پایه و اساس	اشکالات	مزایا	کاربرد	فرمول(نحوه قضاوت)	ملاحظات
۴- آزمون (t) حالت سوم	مقایسه دو نسبت، یکی نسبت فرض (π) و دیگری نسبت مشاهده شده (P)	فقط نسبت‌ها را مطالعه می‌کند.		صفات کیفی $n<۳۰$	$t=\frac{d}{S_d}$ $Sd=\frac{\sqrt{\pi(I-\pi)}}{n}$ $df=\infty=۱/۹۶$ $d=\vert\pi-p\vert$	
۵- آزمون (t) حالت چهارم	مقایسه دو نسبت مشاهده شده $(p_۱, p_۲)$	فقط نسبت‌ها را مقایسه می‌کند.		صفات کیفی $n<۳۰$	$d=\vert p_۱-p_۲\vert$ $sd=\sqrt{\frac{p_۱(۱-p_۱)}{n_۱}+\frac{p_۲(۱-p_۲)}{n_۲}}$ $df=\infty=۱/۹۶$	
آزمون F (واریانس کل به عوامل اوّلیه تجزیه می‌شود. به آن آزمون آنالیز واریانس گفته می‌شود.)	مقایسه میانگین‌های چند جامعه		بررسی میانگین‌های چند جامعه		$SS_t=\underbrace{SS_a}_{\text{بین گروهها}}+SS_e$ داخل گروهها $df=df_a+df_e$ $df_a=k-۱$ $df_e=df-df$ $F=\frac{MS_a}{MS_e}$	اگر $F<F_۰$ باشد، $H۱$ است؛ یعنی تأئید می‌شود. پس باید طبقه‌بندی شوند. $MS_a=\frac{SS_a}{df_a}$ $MS_e=\frac{SS_e}{df_e}$
L.S.D کوچکترین تفاوت معنی‌دار	مرتب کردن متغیرها برای تعیین چند جامعه که هر یک از نظر متوسط صفت با هم یکسان هستند.		به حداقل رساندن اختلاف گروهها	بررسی میانگین چند جامعه (مقوله‌بندی)	$LSD=t\left(n_۲\ ۰.۰۵\right)\sqrt{\frac{۲MS_e}{n}}$ $t=n^۲=df_e$ متوسط مربعات اشتباهات	اگر میانگین‌ها صعودی مرتب شوند، پس کوچکترین میانگین را با L.S.D جمع می‌کنیم.

مقایسه F و T	F تعمیم یافته T است، پس می‌توان دو جامعه را هم با F محاسبه کرد با این تفاوت که F_c مجذور t_c خواهد بود. $F_c = T_c^2$ اگر برای تمامی داده‌های آماری در آزمون F عددی را کسر یا اضافه یا به یک عددی ضرب یا بر آن تقسیم نماییم، نتیجه تغییری نمی‌کند.

جداوّل آزمون‌های آماری (پارامتریک و ناپارامتریک)

نام آزمون	پایه و اساس	اشکالات	مزایا	کاربرد	فرمول(نحوه قضاوت)	ملاحظات
ب- ناپارامتریک ۱- آزمون خی‌دو X^2	مقایسه فراوانی مشاهده شده و فراوانی مورد انتظار	پیش‌فرض‌های اندازه‌گیری فاصله‌ای و توزیع‌های نرمال را بررسی نمی‌کند. اگر بیش از ۲۰درصد فراوانی‌های تجربی از ۵ کوچکتر باشد، صادق نیست. اگر درجه آزادی بزرگتر از ۳۰ بود باید بر اساس روش میانگیری محاسبه شود.	بررسی فرضیه‌های مربوط به شکل و توزیع جامعه و ارتباط بین دو صفت	بررسی تفاوت دو جامعه	$df = K-1$ $X^2 = \sum_{i=1}^{i=r} \frac{\left(n_i - n_{ith}\right)^2}{n_{ith}}$	اگر جدول ۲×۲ باشد: $X^2 = \frac{N\left(ad-bc\right)^2}{e.f.g.h}$ اگر X^Z محاسبه شده رابطه شدیدی را نشان دهد، $\left(H_1\right)$ را می‌توان از طریق P محاسبه کرد. (ضریب توافقی پیرسون) $P = \frac{\sqrt{X^2}}{N\sqrt{df}}$
۱- کاربرد x^2 در نرمال بودن توزیع	مقایسه‌فراوانی‌ها مشاهده شده و مورد انتظار		براساس‌نرمال محاسبه‌می‌شود. بدست‌آوردن برآوردی از انحراف معیار	در تحقیقات مسائل مربوط به پزشکی، کشاورزی، دامداری و علوم‌اجتماعی	$df = r-3$ $u = \frac{X_i - \bar{X}}{\sigma}$	برای به دست‌آوردن گره اوّل CL عدد سمت راست را به جای Xi قرار می‌دهیم، سپس از ۰/۵ کم و در ۱۰۰ ضرب می‌کنیم. برای گروه آخر جدول CL عدد سمت چپ را به‌جای X قرار داده سپس از۰/۵ کم و در ۱۰۰ ضرب می‌کنیم، برای سایر طبقات عدد U را محاسبه کرده و سپس ساخت زیر منحنی دو حد دو طرف از هم کم شود. در ۱۰۰ ضرب می‌کنیم.
۲- آزمون یک نمونه‌ای سیمرنف- کولموگرف	مقایسه فراوانی تجمعی نسبی مشاهده شده $\frac{FCn_i}{n}$ و مورد انتظار $\frac{F\left(n_{ith}\right)}{n}$			بیش از ۲۰درصد فراوانی‌های مورد انتظار از ۵ کوچکتر باشد. $N < 50$	$Dc = MAX \frac{\left\|F\left(n_i\right) - F\left(n_{ith}\right)\right\|}{n}$ بزرگترین نسبت به دست آمده را با D جدول مقایسه می‌کنیم. اگر $D_t < D_c$ ؛ یعنی $H1$ پذیرفته نیست.	چون در اینجا تجمعی فراوانی‌ها محاسبه می‌شود و طرز قرارگیری فراوانی‌ها در محاسبه اثر می‌گذارد؛ اگر متغیر رتبه‌ای یا کمی باشد بر حسب صعودی مرتب می‌شود.

جداوّل آزمون‌های آماری (پارامتریک و ناپارامتریک)

نام آزمون	پایه و اساس	اشکالات	مزایا	کاربرد	فرمول(نحوه قضاوت)	ملاحظات
۳- من- وایت‌نی (M-W)	رتبه دادن به متغیرها- تفاوت دو جامعه با استفاده از نمونه‌های تصادفی		نشان می‌دهد که آیا دو جامعه از نظر تمرکز با هم فرق دارند یا خیر؟	زمانی که داده‌ها ترتیبی باشند نظیر آزمون T حالت دوم است. جامعه با هم کوچکتر را با اندیس ۱ نشان می‌دهد.	$w = v = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$ $w' = v' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$ اگر $H_0, W_t < W_c$ پذیرفته است.	اگر $n \leq 20$ باشد، تبدیل به U یا Z می‌شود. $z = \frac{\left\lvert W_1 - \frac{n_1 n_2}{2} \right\rvert}{\sqrt{n_1 n_2 \left(\frac{n_1 + n_2 + 1}{12} \right)}}$ $V = \frac{\lvert W - \mu_w \rvert}{\sigma_w}$
۴- آزمون دو نمونه‌ای کولموگرف - لموگرف	مقایسه فراوانی تجمعی نسبی دو جامعه		مقایسه دو جامعه برای یکسانی یا غیر یکسانی	۱- رتبه ای باشند. ۲- تعداد آنها کم باشد. ۳- فراوانی‌هازیاد باشد.	$Dc = MAX \left\lvert \frac{F(n_1)}{n_1} - \frac{F(n^2)}{n_2} \right\rvert$	وقتی تعداد صفحات زیاد باشد، به X^2 نزدیک می‌شود. پس: $X^2 = 4Dc^2 \frac{(n_1 n_2)}{n_1 + n_2}$ $D = D_{\max}$ $df = 2$
۵- آزمون تک نمونه‌ای دوره‌ای Onsample Runs test	ترتیب رویدادها و وقایع مبتنی بر تعداد دورها است.	در آزمونهای پارامتری معادلی برای آزمون دورها وجود ندارد.	ضعف X^2 را نشان می‌دهد چون X^2 فقط برحسب فراوانی‌ها محاسبه می‌شود نه تعداد دورها و تکرارها	برای دو گروه نمونه مستقل	اگر IIC بین دو II جدول باشد، H_0 است وگرنه رد می‌شود. اگر یکی از nها از ۲۰ بزرگتر بود، بر اساس Z محاسبه می‌شود.	اگر تعداد دوره‌ها کم باشد، امکان دارد نوعی روند زمانی یا تجمع وجود داشته باشد و اگر زیاد باشد، احتمالاً نوسانات چرخه‌ای کوتاه‌مدتی به‌طور منظم نمره‌ها را تحت تأثیر قرار داده است.