

آمار زیستی Biostatistics

کارگاه آمار زیستی (مقدمات پروژه کارگاه)

جزوه و کلیپهای آموزشی نرم افزار Excel و SPSS از سایت www.aminsedighi.ir دانلود نمایید
حضور فعال سر کلاس اجباری است

جزوه قسمت اول

سرفصل آمار زیستی

- | | |
|---|---|
| <p>۱۲) منحنی طبیعی و کاربرد آن در آمار</p> <p>۱۳) نمره های استاندارد و انواع آن</p> <p>۱۴) مختصری از احتمالات - تابع احتمال - تابع توزیع احتمال نرمال - تابع توزیع احتمال دوجمله ای - آزمون فرضها - امید ریاضی - کوواریانس</p> <p>۱۵) شاخص های همبستگی و انواع آن با تکیه بر همبستگی پیرسون اسپرمن</p> <p>۱۶) رگرسیون و پیش بینی</p> <p>۱۷) آشنایی با نرم افزارهای آماری از جمله SPSS</p> <p>۱۸) نحوه تکمیل پرسشنامه</p> <p>۱۹) پروژه</p> | <p>۱) تعریف مفاهیم (ریاضی آمار اندازه گیری سنجش)</p> <p>۲) طبقه بندی (آمار توصیفی و استنباطی)</p> <p>۳) جامعه نمونه پارامتر مشخصه آماری</p> <p>۴) مفهوم متعیر و انواع آن</p> <p>۵) طبقه بندی متعیر از لحاظ مقیاس اندازه گیری</p> <p>۶) انواع نمودار و کاربرد آن در آمار</p> <p>۷) طبقه بندی آمار توصیفی (یک متغیری چند متغیری)</p> <p>۸) شاخص های مرکزی (مد میانه میانگین)</p> <p>۹) شاخص های پراکندگی (واریانس انحراف معیار انحراف چارکی و دامنه تغییرات)</p> <p>۱۰) ضریب تغییرات - چولگی - کشیدگی</p> <p>۱۱) نقاط درصدی (چندک ها) و رتبه بندی</p> |
|---|---|

مقدمه

آمار زیستی شاخه‌ای از آمار (کاربردی) است که تمرکز و تأکید آن بر توسعه و استفاده از روشهای آماری است که در راستای حل مسائل و پاسخ به سئوالاتی که در بهداشت، پزشکی و ژنتیک و بیولوژی انسانی مطرح می‌شوند.

تاریخچه رشته و پیشرفتهای جدید

از آغاز قرن هفدهم تا امروز، پدیده‌های اساسی بیولوژیکی (مخصوصاً مرگ و میر و ابتلاء به بیماریها) و جمع‌آوری داده‌ها در این زمینه‌ها بسیار مورد توجه بوده است.

تاریخچه آمار زیستی در مملکت ما به اواسط قرن بیستم برمی‌گردد درحالیکه اولین گروه آمار زیستی در ایران و در دانشکده بهداشت دانشگاه تهران در سال ۱۳۴۵ شمسی میباشد و بمرور زمان و تاکنون به پذیرش و آموزش دانشجویان مقطع کارشناسی و کارشناسی ارشد و در سالهای اخیر در مقطع دکتری تخصصی (Ph.D) ادامه می‌دهد. به کارگیری کامپیوترها با سرعت بسیار بالا مخصوصاً در دهه‌های اخیر تأثیری شگرف بر تمامی عرصه‌های حیات انسان از جمله علوم داشته است.

در این راستا:

- ۱- پیشرفت‌های نظری: توسعه روشهای تئوریک آمار زیستی با اهداف کاربردی آنها در علوم پزشکی و بهداشتی.
- ۲- پیشرفتهای فن آوری: با اهداف روشهای بهینه انتقال علم آمار زیستی به سایر کاربران بهداشتی و علوم زیستی.
- ۳- پیشرفتهای محاسباتی: در زمینه‌های مختلف و متعدد از جمله روشهای آماری، Data Mining, Bootstrapping تحلیل داده‌های سیستم High Throughput و کاربرد آمار در ژنتیک ملکولی.
- ۴- تغییرات فرهنگی - اجتماعی: نیازهای روزافزون به تلخیص و ارائه اطلاعات معتبر.
- ۵- شیوه‌های ارائه خدمات: کامپیوترایزه شدن تبادلات کلیه جنبه‌های اطلاعات آماری.

علم آمار زیستی از (معدود) علمی است که بطور روزافزون در کلیه زمینه‌های پژوهشی بشر از صنعت و کشاورزی، اقتصاد و تجارت گرفته تا بهداشت و بیولوژی و بیوتکنولوژی و پزشکی کاربرد پیدا کرده است.

رسالت برنامه آموزشی در تربیت نیروی انسانی (Mission)

رسالت دوره کارشناسی ارشد در آمار زیستی تربیت متخصصین در زمینه‌های آمار زیستی و مفاهیم اپیدمیولوژی و حتی مفاهیم وسیع‌تر در زمینه‌های بهداشتی است. یک دانش آموخته در این رشته باید قادر باشد درباره موضوعات آماری علوم پزشکی اظهار نظر نماید.

چشم انداز برنامه آموزشی در تربیت نیروی انسانی (Vision)

این برنامه با بکارگیری خلاقیت و تجربیات علمی و کاربردی اساتید، دانش آموختگان و دانشجویان به منظور فراهم نمودن بستر مناسب و تشویق لازم جهت انجام کاربرد و اجرای صحیح علم آمار زیستی در راستای بهینه سازی سلامت عمومی جامعه جهت پیشبرد تحقیقات علمی در زمینه بهداشت عمومی و پزشکی است که بطور پویا همپای دانشگاههای پیشرفته و مطرح جهان قادر به رقابت باشد.

اهداف کلی رشته (Aims)

برنامه کارشناسی آمار زیستی با اهداف کلی زیر طراحی شده است:

- ۱- بکارگیری مهارت‌های ذهنی آمار زیستی در تحقیقات علوم بهداشتی - پزشکی در دانشگاه‌ها.
- ۲- توانمندسازی دانش آموختگان برای ایفای نقش در تحقیقات در زمینه‌های مختلف علوم زیستی.

نقش دانش آموختگان در برنامه آموزشی (Role definition)

- ۱- مشاوره ای
- ۲- آموزشی
- ۳- مدیریتی
- ۴- پژوهشی

وظایف حرفه‌ای دانش آموختگان (Task Analysis)

الف - در زمینه مشاوره‌ای: راهنمایی دانشجویان یا سایر محققین با ارائه مشاوره جهت رسیدن ایشان به اهداف تحقیقاتی موردنظر پژوهش.

ب - در زمینه پژوهشی:

۱- تحقیق و تفحص در زمینه‌های کاربردی مدرن آمار زیستی با هدف بهینه کردن روشهای تحلیل آماری تحقیقات پزشکی و بهداشتی.

۲- کمک به تحقیقات در سایر زمینه‌های علوم زیستی با ارائه روشهای مناسب و جدید و شرکت در تحقیقات سایر متخصصین در پژوهشهای پزشکی - زیستی.

ج - در زمینه‌های آموزشی:

۱- تدریس و آموزش یک یا چند درس در زمینه‌های مشخص آماری در رابطه با آمار زیستی یا آمار پزشکی و در سطوح ابتدائی دانشگاهی یا دیگر مراکز علمی.

۲- پی‌گیری و جستجوی روشهای بهینه جهت آموزش و انتقال مفاهیم آمار زیستی به دانشجویان پزشکی و رشته‌های مشابه.

د - در زمینه مدیریتی

همکاری و مدیریت پروژه‌های تحقیقاتی

استراتژی‌های اجرایی برنامه آموزشی

۱- بهره‌گیری از استراتژی تلفیقی آموزشی (استفاده از استراتژی استاد محوری یا شاگردمحوری برحسب نیاز)

۲- بهره‌گیری از عرصه‌های عملی و واقعی کار و عرصه‌های شبیه‌سازی شده

۳- Community Orientation

۴- استفاده از شیوه‌های مبتنی بر مشکل (Problem Oriented) برحسب مورد

۵- بهره‌گیری از Integration در عرصه‌های کاری

علم آمار

آمار نقشی لاینفک در زندگی روزمره ما بازی می کند. اخبار روزانه رسانه های گروهی با گزارشی از وضع هوا به پایان می رسند و در طول اخبار، به جریانهای بازار بورس و سهام اشاره می شود و روزنامه ها خبر از افزایش نرخ اجناس میدهند و ...

تعریف آمار:

روش و چگونگی جمع آوری اطلاعات و بیان آنها در قالب عدد

تعریف احتمال:

تجزیه و تحلیل اطلاعات عددی جمع آوری شده

استنباط:

قضاوت درباره کل اطلاعات وقتی بخشی از آن رؤیت شده

جمعیت:

مجموعه تمام عناصری که دارای یک یا چند ویژگی مشترک باشند

نمونه:

بخشی از جمعیت میباشد



Biostatistics آمار توصیفی زیستی

آمار توصیفی زیستی، مجموعه ای از روش هایی است که برای سازمان دهی، خلاصه کردن، تهیه جدول، رسم نمودار، توصیف و تفسیر داده های جمع آوری شده از نمونه آماری به کار گرفته می شود. مراحل اساسی توصیف داده ها عبارت است از:

- (۱) خلاصه کردن داده ها و توصیف الگوی کلی
- (۲) فشرده کردن داده ها در قالب جدول های آماری
- (۳) نمایش آن ها به وسیله نمودار
- (۴) محاسبه شاخص های آماری

نقش آمار توصیفی در فرآیند تحلیل آماری بسیار مهم و حیاتی است. آمار توصیفی با خلاصه کردن داده ها، ویژگی های مهم آن ها را نمایان می سازد تا ایده های لازم را در ذهن پژوهش گر برای مرحله دوم تحلیل آماری (آمار استنباطی) ایجاد کند.

👉 آمار استنباطی

آمار استنباطی مشخص می کند که آیا الگوها و فرآیندهای کشف شده در نمونه، در جامعه آماری هم کاربرد دارد یا خیر. بنابراین، آمار استنباطی راجع به ویژگی ها و پارامترهای مربوط به جامعه آماری تحقیق و کیفیت ارتباط بین مفاهیم و متغیرها می باشد. بدین ترتیب، می توان گفت که از آمار استنباطی در تجزیه و تحلیل مقایسه ای و رابطه ای (علی - همبستگی) استفاده می شود.

تفاوت اصلی آمار توصیفی و استنباطی در این است که در آمار توصیفی هیچ گاه نمی توان نتایج به دست آمده از نمونه آماری را به کل جامعه آماری تعمیم داد. چرا که هدف در این نوع آمار، ارائه توصیفی از ویژگی های نمونه آماری تحقیق به همراه شاخص های گرایش به مرکز و یا شاخص های گرایش به پراکندگی می باشد. در حالی که در آمار استنباطی و یا تحلیلی می توان نتایج و یافته های به دست آمده از نمونه آماری را به کل جامعه آماری تحقیق تعمیم داد. به عبارتی، مفهوم کانونی آمار استنباطی، تعمیم پذیری است.

به بیانی روشن تر، تفاوت اصلی یک بررسی توصیفی با یک آمار استنباط آماری این است که نتایج اولی فقط مختص به نمونه مورد بررسی است، در حالی که آمار استنباطی نتایجی را در مورد جامعه بیان خواهد کرد. از این رو آمار توصیفی همراه با عدم قطعیت و آمار استنباطی همواره با قطعیت همراه است.

👉 آمارهای توصیفی عمده :

- ✓ اندازه های گرایش به مرکز (میانگین، میانه، نما)
- ✓ اندازه های پراکندگی (دامنه تغییرات، واریانس، انحراف استاندارد)
- ✓ اندازه های وضعیت نسبی (رتبه درصدی، نمره انحراف متوسط)
- ✓ اندازه های رابطه ای (ضریب همبستگی پیرسون، اسپیرمن)

نرم افزارها در آمار

SPSS (Statistical package for social science)	SAS (Statistical Analysis Software)				
Minitab	Eviews	Statistica	Lisrel	Expert	choice
NCSS	Microfit	GAMS	STATA	SmartPLS	Amos
Statgraph	PASS(NCSS)	R	Excel	Mathlab	

آشنایی با نرم افزار SPSS و آمار در Excel و Mathlab

Statistical Package for the Social Sciences

نرم افزار را از دانشگاه یا از اینترنت یا محل کار خودتان یا فروشگاههای CD سطح شهر تهیه و نصب کنید لیسنس (گواهینامه) آنرا تایید نمایید
به کلیه‌های موجود در سایت www.aminsedighi.ir بخش آمار در خصوص نرم افزار SPSS مراجعه کنید و همچنین آمار در Excel و MathLab

متغیر:

- یک ویژگی که از فردی به فرد دیگر در یک جمعیت (نمونه) تغییر میکند
- ۱- متغیر کمی (مثل وزن یا قد) - صفاتی هستند که قابل اندازه‌گیری یا شمارش هستند مانند سن، طول عمر، وزن، درآمد، هزینه، و...
 - متغیر کمی دو دسته اند: گسسته (مثل تعداد فرزند) - پیوسته (قد یا وزن یا فشارخون)
 - متغیر کمی دو دسته اند: نسبی (Ratio) - فاصله‌ای (Interval)
 - ۲- متغیر توصیفی یا کیفی (مثل مهارت یا هوش) - صفاتی هستند که واحد نداشته و قابل اندازه‌گیری نیستند مانند جنس، مرغوبیت، مهارت، گروه خونی و ...
 - متغیر کیفی دو دسته اند: ترتیبی (Ordinal) - اسمی (Nominal)

مقیاس سازی توصیفی یا کیفی

- ۱- مقیاس اسمی (Nominal) - اعداد برتری نسبت بهم ندارند (نه ارجحیت و برتری دارند و نه ضریب و تناسب) مثال (تهران ۰۲۱ - شیراز ۰۷۱) (کد پستی یا کدملی: بزرگ بودن اعداد کدملی به مفهوم برتری نیست) (گروه خونی: $1=A, 2=B, 3=AB, \dots$ اعداد ارجحیتی و برتری ندارند و تناسب هم ندارند)
- ۲- مقیاس ترتیبی (Ordinal) - اعداد برتری نسبت بهم ارجحیت و برتری دارند ولی ضریب و تناسب ندارند (ضعیف ۰ - متوسط ۱ - خوب ۳ - بسیارخوب ۶) (رتبه اول ۱۹ - رتبه دوم ۱۸,۷۵)

مقیاس سازی کمی

- ۳- مقیاس نسبتی (Ratio) یا (Scale) - $x=at$ - اعداد نسبت داده شده ارجحیت و برتری دارند و هم ضریب تناسب هم دارند (مثل سن و مثل وزن - مثلا وزن احمد ۳۵ و وزن عباس ۷۰ کیلوگرم یعنی عباس بیشتر از احمد است و به نسبت دو برابر هم میباشد) - در این مقیاس صفر بمعنی هیچ است
- ۴- مقیاس فاصله‌ای (Interval) - $x=at+b$ که $b < 0$ مثلاً $f=9/5(c+32)$ - مثلا به یک بیمار بگوییم اگر حداکثر درد مثل یک خط کش عدد ۱۰۰ باشد درد شما چقدر است و بیمار بگوید ۶۵ -- در این مقیاس صفر بمعنی هیچ نیست

انواع مقیاس‌ها

در آمار باید به هر اطلاعات و متغیرهایی، عدد نسبت دهیم که به حالت‌های زیر می‌باشد

متغیرهای کیفی

۱. مقیاس اسمی (NOMINAL)

با کد گذاری عددی - اعداد نسبت داده شده ارجحیت و برتری ندارند (مثل کد شهرها - مثل جنسیت - زن - مرد ۱) (مثل آبی ۱ - زرد ۲ - قرمز ۳ - نارنجی ۴ - ...)

۲. مقیاس رتبه‌ای (ORDINAL) (ترتیبی)

با کد گذاری عددی - اعداد نسبت داده شده ارجحیت دارند ولی تناسب ندارند (شدت و ضعف دارند) (مثل پرسشنامه لینکرت شامل موافق ۳ - بی نظر ۲ - مخالف ۱ یا تحصیلات مثلا دیپلم ۱ فوق دیپلم ۲ لیسانس ۳ و فوق لیسانس ۴ (که ۴ بیشتر از ۱ ولی ۴ چهار برابر ۱ نیست) (مثل ناراضی ۱ - بی نظر ۲ - راضی ۳)

متغیرهای کمی

۳. مقیاس نسبتی (SCALE) (وزنی)

با کد گذاری عددی - اعداد نسبت داده شده ارجحیت دارند و تناسب هم دارند (مثل سن و مثل وزن - مثلا وزن احمد ۳۵ و وزن عباس ۷۰ کیلوگرم یعنی عباس بیشتر از احمد است و به نسبت دو برابر هم می‌باشد)

۴. مقیاس فاصله‌ای

مثلا به یک بیمار بگوییم اگر حداکثر درد مثل یک خط کش عدد ۱۰۰ باشد درد شما چقدر است و بیمار بگوید ۶۵

داده :

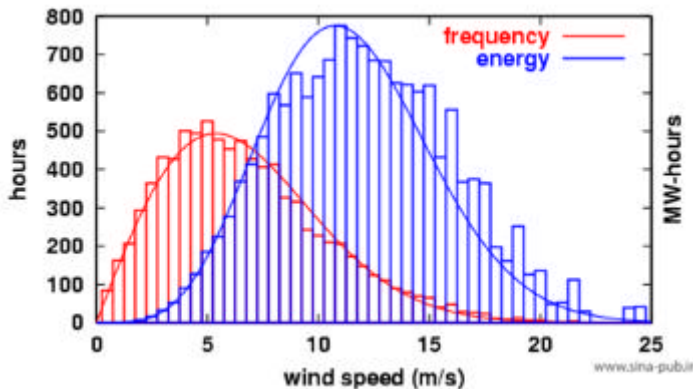
- ۱- داده گسسته (تعداد فرزند) - با شمارش حاصل میشود
- ۲- داده پیوسته (طول قد) - با اندازه گیری حاصل میشود

مشخص کننده مرکزی

- ۱- میانگین (حسابی - وزنی - هندسی - توافقی و ...)
- ۲- میانه (وسط صف منظم داده ها) m
- ۳- نما (داده ای که بیشتر تکرار شده) M
- ۴- میانگین و واریانس (میانگین و میزان انحراف داده ها از میانگین)
- ۵- چارک - دهک - صدک یعنی داده ها را این چندک کوچکترند

فراوانی

- ۱- **فراوانی مطلق**: به تعداد داده‌ها در هر طبقه فراوانی مطلق آن طبقه می‌گویند و آن را با f_i نشان می‌دهند.
- ۲- **فراوانی نسبی**: در صورتی که فراوانی‌های مطلق را بر کل فراوانی‌ها تقسیم کنیم، فراوانی نسبی r_i به دست می‌آید.
- ۳- **فراوانی تجمعی**: به مجموع فراوانی‌های مطلق طبقه‌های قبل و همان طبقه، فراوانی تجمعی آن طبقه می‌گویند و آن را با F_i نمایش می‌دهند.
- ۴- **فراوانی تجمعی نسبی**: می‌توان از تقسیم فراوانی‌های تجمعی بر تعداد داده‌ها، این فراوانی را به دست آورد (R_i).



چندک ها

چارک - دهک - صدک یعنی داده ها از این چندک کوچکترند (ابتدا داده ها را مرتب کنیم سپس تعداد داده ها (n) را در نسبت مربوط به چندک (p) ضرب کنیم n*p (در صورت اعشاری بودن آن را به بالا گرد کنید و داده را استخراج نمایید در صورت مساوی بودن آن P- میانگین آن داده و داده بالاتر را استخراج کنید) تا محل قرار گرفتن چندک معلوم شود.

صدک (P): صدک چهارم است و داده‌ای است که ۴۰ درصد داده‌ها از آن کوچکتر است. بدیهی است که ۶۰ درصد داده‌ها از آن بزرگتر باشد.

دهک ها، حالت خاص صدک ها هستند و به ترتیب به صدک های دهم، بیستم، سی ام و... و نودام، به ترتیب دهک اول، دوم، سوم و ... نهم گفته می‌شود.

چارک Q: چارک سوم است داده‌ای است که سه چهارم (۷۵٪) داده‌ها از آن کوچکتر است. بدیهی است که یک چهارم (۲۵٪) دیگر داده‌ها از آن بزرگتر باشد.

مثال

صدک ۶۵ داده‌های زیر را معلوم کنید. 3.2-3.6-3.5-3.7-3.9-3.8-4.0-4.3-4.4-3.5

حل: ابتدا داده‌ها مرتب میکنیم و یک ردیف بنام ردیف (فراوانی تجمعی) تشکیل میدهم

Data=Xi	3.2	3.5	3.5	3.6	3.7	3.8	3.9	4	4.3	4.4
f	1	1	1	1	1	1	1	1	1	1
F	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰

برای محاسبه صدک ۶۵ باید تعداد داده ها را در ۶۵٪ ضرب کنیم. (۶۵٪ نسبت مربوط به صدک ۶۵ است)

$$Q * \sum f_i = 0.65 * 10 = 6.5 \quad 6.5 + (=7) \quad F=7 \Rightarrow X=3.9$$

صدک ۶۵ ام برابر با ۳,۹ گردید یعنی: $P_{65}=3.9$

یعنی ۶۵٪ داده‌ها مقدار عددیشان از ۳,۹ کمتر است

مثال

دهک چهارم داده‌های زیر را معلوم کنید

7.5- 8.3 -9.8-2.3 - 3- 3.1 -3.2- 3.4 - 3.8- 4 - 4.4- 3.5- 5.3- 6.1- 6.9

داده‌ها را مرتب میکنیم

Xi	2.3	3	3.1	3.2	3.4	3.5	3.8	4	4.4	5.3	6.1	6.9	7.5	8.3	9.8
fi	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Fi	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ri	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15	1/15
Ri	1/15	2/15	3/15	4/15	5/15	6/15	7/15	8/15	9/15	10/15	11/15	12/15	13/15	14/15	15/15

برای محاسبه دهک چهارم باید تعداد داده ها (n=۱۵) را در ۴۰٪ ضرب کنیم

$$Q * \sum f_i = .4 * 15 = 6 \quad 6 + \Rightarrow F \Rightarrow F=7 \quad 7 \Rightarrow X \Rightarrow X=3.8$$

دهک چهارم یعنی ۴ دهم یا ۴۰٪ داده‌ها ۳,۸ یا کمتر از ۳,۸ ام است یعنی: $P_4 = 3.8$

مثال

داده‌هایی بشرح ذیل داریم صدک ۶۰ چه مقدار میشود

13-14-13-15-13-17-14-12-15-17

داده را مرتب مینماییم و یک سطر (ردیف) ایجاد میکنیم

داده Xi	12	13	14	15	17
تعداد f	1	3	2	2	2
تجمع F	۱	۴	۶	۸	۱۰

$Q * \sum f_i = (60/100) * 10 = 6 \Rightarrow$ در جمع $6+ = 7 \Rightarrow x=15$

صدک شصت (شصت درصد) داده ۱۵ یا کمتر از آن میباشد

اگر همان داده را به صورت جدول فراوانی و جمع فراوانی بنویسیم

داده Xi	12	13	14	15	17
تعداد f _i	1	3	2	2	2
F _i	1	4	6	8	10
r _i	1/10	3/10	2/10	2/10	1/10
R _i	1/10	4/10	6/10	8/10	10/10

$n = \sum f_i = 1+3+2+2+2=10$

$Q * \sum f_i = (60/100) * 10 = 6 \Rightarrow 6+ \Rightarrow$ در ردیف $6+ = 8 \Rightarrow$ داده متناظر $x=15$

صدک شصت (شصت درصد) داده ۱۵ یا کمتر از آن میباشد

نکته: در کران بالای هر طبقه عددی ملاحظه میشود که شمول این طبقه بر این داده نیست یعنی آخرین داده این طبقه کمی کمتر از همین داده است و این داده شامل طبقه بعدی است که در کران پایین طبقه بعدی ملاحظه میشود

توجه: عدد چارک دوم = عدد دهک پنجم = عدد صدک پنجاهم = میانه

مشخصه های پراکنندگی:

- ۱- مینیمم داده و ماگزیمم داده
- ۲- دامنه تغییرات (تفاضل بزرگترین داده از کوچکترین داده)
- ۳- انحراف متوسط (میانگین قدر مطلق انحراف ها از میانگین داده ها)
- ۴- انحراف معیار (جذر واریانس) که واریانس میانگین توان دوم انحراف ها از میانگین داده ها

فراوانی:

اگر n شیء متمایز در دسته های مختلفی (n1, n2, ..) را به تعداد مختلف (w1, w2, ..) داشته باشیم w ها را فراوانی گویند

۱- فراوانی مطلق frequency (f) = فراوانی = تعداد هر داده

۲- فراوانی نسبی r_i = نسبت فراوانی هر مجموعه به کل تعداد آن مجموعه $r_i = w_i/n$

$$r_i = \frac{f_i}{\sum f_i}$$

۳- فراوانی انباشته (تجمعی) F = Cumulative frequency = حاصل جمع فراوانی تا موقعیت مورد نظر F_j

$$F_i = \sum_{-\infty}^a f_i$$

۴- فراوانی انباشته نسبی R_i = حاصل جمع W_i ها تقسیم بر n $R_i = \frac{\sum_{-\infty}^a f_i}{\sum f_i}$

میانگین - اگر از هر داده به تعداد یکی داشته باشیم

$$\bar{x} = \mu = \frac{\sum_{i=1}^n x_i}{n}$$

واریانس σ^2 = انحراف معیار σ = معیاری جهت پراکندگی داده ها

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

$$\sigma = \sqrt{\sigma^2}$$

میانگین - اگر از هر داده فراوانی (به تعداد مشخصی) داشته باشیم

$$\bar{x} = \mu = \frac{\sum_{i=1}^n x_i * f_i}{\sum f_i}$$

$$\vartheta(x) = \sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 * f_i}{\sum f_i}$$

$$\sigma = \sqrt{\sigma^2}$$

ثابت شده است که در طبیعت آمار بصورت نرمال توزیع شده است

که با استفاده از فرمول توزیع نرمال ثابت میشود که:

۶۸٪ داده ها بین $\mu \pm \sigma$ خواهد بود و

۹۶٪ داده ها بین $\mu \pm 2\sigma$ خواهد بود

مشخصه ضریب تغییرات = معیاری جهت همگن بودن داده ها

اگر نسبت انحراف استاندارد را به میانگین بدست آوریم. چون صورت و مخرج این کسر هم واحد هستند، حاصل کسر مقداری بدون واحد است که به صورت درصدی نیز می تواند بیان شود. بنابراین ممکن است برای یک سری داده گفته شود که ضریب تغییرات ۱۵٪ است. این امر به معنی آن است که انحراف معیار ۱۵ درصد میانگین است.

$$\rho = \frac{\sigma}{\mu} \times 100$$

ضریب تغییرات کوچکتر یعنی همگن بودن و یکدست بودن داده ها

برای داده های x_i و فراوانی f_i میانگین و واریانس را حساب کنید آیا این داده ها نرمال هستند

$$r_i = \frac{f_i}{\sum f_i} \quad R_i = \sum r_i \quad F_i = \sum f_i$$

	داده x_i	فراوانی f_i	$x_i f_i$	فراوانی نسبی r_i	فراوانی نسبی تجمعی R_i	فراوانی تجمعی F_i
	10.4	1	10.4	0.02	0.02	1
	11.3	6	67.8	0.12	0.14	7
	12.2	7	85.4	0.14	0.28	14
	13.1	12	157.2	0.24	0.52	26
	14	12	168	0.24	0.76	38
	14.9	9	134.1	0.18	0.94	47
	15.8	3	47.4	0.06	1	50
جمع		50	670.3	1		
			13.406			1.80547
			میانگین			واریانس نمونه

$$s^2 = 1.805$$

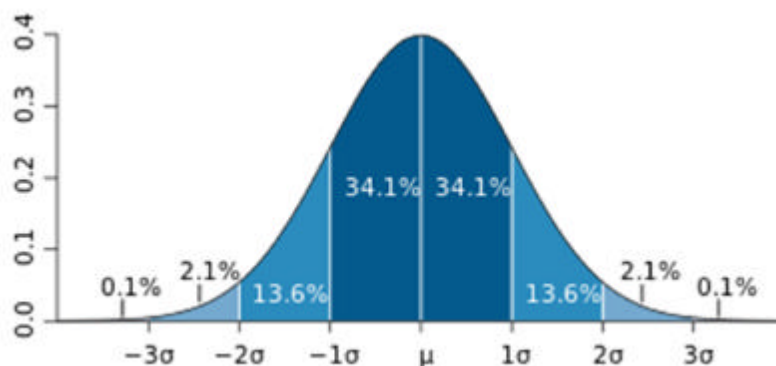
$$s = 1.34$$

$$\mu \pm \sigma = 12.06 \quad 14.74$$

$$\mu \pm 2\sigma = 10.71 \quad 16.09$$

$$\mu \pm 3\sigma = 9.37 \quad 16.43$$

داده بین اعداد	12.06-14.74	10.71-16.09	09.37-16.43
تعداد	31	49	50
درصد	31/50=62%	49/50=98%	50/50=100%



خواص میانگین حسابی

اگر داده های ما x_i باشد میانگین آنها \bar{x} باشد آنگاه جمع داده ها با میانگین با توجه به فراوانی صفر میشود
 $\sum(x_i - \bar{x}) f_i = 0$

اگر داده ها را با یک عدد ثابت جمع یا تفریق کنیم میانگین هم با همین عدد ثابت جمع یا تفریق میشود
 $y_i = x_i \pm a \Rightarrow \bar{y} = \bar{x} \pm a$
 $y_i = x_i * a \Rightarrow \bar{y} = \bar{x} * a$

اگر y_i یکسری اعداد دیگر باشد

$$z_i = x_i + y_i \Rightarrow \bar{z} = \bar{x} + \bar{y}$$

خواص واریانس (پراش)

اگر داده های ما x_i باشد میانگین آنها \bar{x} باشد آنگاه اگر داده ها را با یک عدد ثابت جمع یا تفریق کنیم واریانس و انحراف معیار تغییر نمیکند
 $y_i = x_i \pm a \Rightarrow \sigma_y = \sigma_x$

مثال

قد تعدادی اشخاص بشرح ذیل است انحراف معیار قد را بدست آورید

X	165	168	170	171	173
F	2	4	5	3	1

ضرب اعداد فوق در فراوانی ها وقت زیاد میبرد یک عدد در حدود وسط داده ها انتخاب میکنیم اعداد جدول فوق از سایز عدد ۱۷۰ کم کرده جدول زیر بدست میاید میانگین و واریانس این جدول بدست آورده که ضرب و تقسیم آن ساده تر است بدست آورده و در انتها به میانگین عدد ۱۷۰ اضافه میکنیم و واریانس فرق نمیکند

Y	-5	-2	0	1	3
F	2	4	5	3	1

$$\bar{y} = \frac{\sum y_i f_i}{\sum f_i} = \frac{(-5*2)+(-2*4)+\dots}{2+4+\dots} = \frac{-12}{15} = -0.8 \quad \bar{x} = \bar{y} + 170 = -0.8 + 170 = 169.2$$

$$v(y) = \sigma_y^2 = \frac{\sum (y - \bar{y})^2 f_i}{\sum f_i} = \frac{(-5 - (-0.8))^2 * 2 + (-2 - (-0.8))^2 * 4 + \dots}{2+4+\dots} = 4.56 \quad \sigma_y = \sqrt{4.56} = 2.13$$

$$\sigma_x = \sigma_y = 2.13$$

** اگر داده ها را در یک عدد ضرب کنیم انحراف معیار با مجذور آن عدد نسبت معکوس دارد

$$y_i = x_i * a \Rightarrow \sigma_y = \left(\frac{1}{a^2}\right) * \sigma_x$$

مثال

نمرات 10 واحد درسی دانشجویی به شرح ذیل میباشد

نمره = x	۱۸	۱۲	۱۵	۱۶	۱۰
f = تعداد(واحد)	۱	۴	۲	۱	۲

الف) مد چه عددی است(چرا) - ب) میانه چه عددی است(چرا) ج) صدک ۸۵ داده ها بدست آورید د) میانگین و واریانس و انحراف معیار را بدست آورید
حل: ابتدا داده ها مرتب میکنیم

X_i	۱۰	۱۲	۱۵	۱۶	۱۸	
f_i	۲	۴	۲	۱	۱	
F_i	۲	۶	۸	۹	۱۰	

داد ها گسسته هستند زیرا بین داده ها پیوسته نیست

الف) مد داده با بیشترین تکرار که نمره ۱۲ مود میباشد

ب) برای میانه باید وسط کل داده ها را پیدا کنیم جمع داده ها را بر ۲ تقسیم کنیم(نصف ۱۰ برابر ۵ میشود) و در ستون F_i دنبال خود این جواب و یا کران بالای آن میگردیم که $F_i=5$ نداریم و کران بالا عدد ۶ میشود و داده متناظر آن نمره ۱۲ میانه میشود

$$Q * (\sum f_i) = \left(\frac{1}{2}\right) * (2 + 4 + 2 + 1 + 1) = 5$$

$$5^+ \rightarrow \frac{F}{f} \rightarrow F = 6 \rightarrow \frac{F}{x} \rightarrow x = 12$$

میانه ۱۲ میباشد
(ج)

$$Q * \sum f_i = \left(\frac{85}{100}\right) * (2 + 4 + 2 + 1 + 1) = 8.5$$

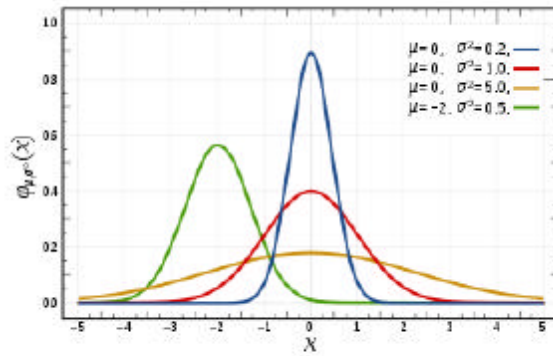
$$\rightarrow 8.5^+ \rightarrow \frac{F}{f} \rightarrow F = 9 \rightarrow \frac{F}{x} \rightarrow x = 16$$

صدک ۸۵ عدد ۱۶ شد یعنی ۸۵٪ داده ها ۱۶ یا کمتر از ۱۶ میباشد

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{(10 * 2) + (12 * 4) + (15 * 2) + (16 * 1) + (18 * 1)}{(2 + 4 + 2 + 1 + 1)} = 13.2$$

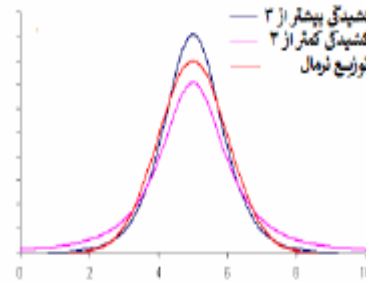
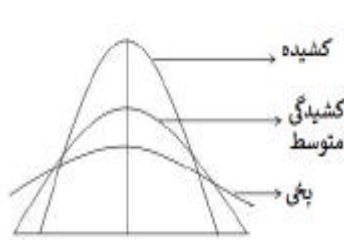
$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} = \frac{(10 - 13.2)^2 * 2 + (12 - 13.2)^2 * 4 + (15 - 13.2)^2 * 2 + (16 - 13.2)^2 * 1 + (18 - 13.2)^2 * 1}{(2 + 4 + 2 + 1 + 1)} = 6.36 = \sigma^2 \quad \sigma = 2.52$$

توزیع نرمال - نرمال استاندارد



منحنی توزیع نرمال

کشیدگی (Kurtosis)



مقدار کشیدگی برای توزیع نرمال برابر ۳ می باشد

چولگی (Skewness)



چنانچه چولگی و کشیدگی در بازه (۲، -۲) نباشند داده‌ها از توزیع نرمال برخوردار نیستند
ضریب چولگی گشتاوری پیرسون
ضریب چولگی اول و دوم پیرسون
چولگی بر مبنای چارک‌ها

رگرسیون و همبستگی خطی

n تا زوج مرتب داریم بهترین رابطه خطی بین دو متغیر زوج پیدا کنیم

$$y = a + bx$$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}, \quad a = \bar{y} - b\bar{x} \Rightarrow y = a + bx$$

در رگرسیون باید ببینیم متغیر مستقل کدام است مثلاً اگر جدولی از طول قامت پدران و پسرانشان داشته باشیم متغیر مستقل پدر است زیرا بتبع قامت پدر قامت پسر تغییر میکند بعبارت دیگر آن بخشی از داده که نداریم و مجهول است y نام میگذاریم نکته مهم: در رابطه فوق در مخرج باید متغیر مستقل باشد

مثال

معادله رگرسیون بر حسب y برای زوجهای مرتب زیر بدست آورید

$$x = 1,3,4,6,8,9,11,14$$

$$y = 1,2,4,4,5,7,8,9$$

$$y = a + bx$$

$$b = 7/11$$

$$a = 5 - (7/11) * 7 = 6/11$$

$$y = 6/11 + (7/11)x$$

اگر بخواهیم معادله خط رگرسیون x بدست آوریم در تمام مخرج رابطه فوق (رابطه b) بجای x میتوان y قرار داد

$$x = a + by$$

$$a = -0.5$$

$$b = 1.5$$

$$x = -0.5 + 1.5y$$

مثال

در جدول زیر بر حسب اینچ طول قد پدر و پسر ملاحظه میکنید. معادله خط رگرسیون را بنویسید؟ اگر قامت پدری ۷۵ شد پیش بینی میکنید فرزند او دارای چه قدی باشد؟

پدر	65	63	67	64	68	62	70	66	68	67	69	71
پسر	68	66	68	65	69	66	68	65	71	67	68	70

حل:

پدر را x و پسر را y مینامیم زیرا قد پسر از پدر تبعیت میکند پس پدر متغیر مستقل است

$$\sum x_i = 65 + 63 + \dots + 71 = 800 \quad \sum y_i = 68 + 66 + \dots + 70 = 811$$

$$\sum x_i^2 = 65^2 + 63^2 + \dots + 71^2 = 53418 \quad \sum x_i y_i = 65 * 68 + 63 * 66 + \dots + 71 * 70 = 54107$$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{54107 - \frac{800 * 811}{12}}{53418 - \frac{(800)^2}{12}} = 0.476 \quad , \bar{x} = 800/12 \quad \bar{y} = 811/12$$

$$a = \bar{y} - b\bar{x} \quad a = (811/12) - (0.476)(800/12) = 35.85 \Rightarrow$$

$$y = a + bx$$

$$y = 35.85 + 0.476x$$

$$y = 35.85 + (0.476 * 75) = 71.55$$

مصرف یک دارو در سه سال گذشته مقادیر زیر بوده در سالهای بعد پیش بینی نمایید

سال = x	۹۳	۹۴	۹۵	۹۶	۹۷
مصرف = y	۱	۲	۴	؟	؟

*** حل : ابتدا در جدول مقادیر X قدیمی را Xm نام گذاشته و آنرا را تغییر میدهیم مثلا همه را از ۹۴ کم میکنیم

سال = Xm	۹۳	۹۴	۹۵
x	-1	0	+1
مصرف = y	۱	۲	۴

$$y = a + bx$$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i * \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{(-1 * 1) + (0 * 2) + (1 * 4) - \frac{(-1 + 0 + 1)(1 + 2 + 4)}{3}}{((-1)^2 + (0)^2 + (1)^2) - \frac{(-1 + 0 + 1)^2}{3}} = \frac{3}{2} = 1.5$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{-1 + 0 + 1}{3} = 0 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{1 + 2 + 4}{3} = \frac{7}{3} = 2.33$$

$$\bar{y} = a + b\bar{x} \quad 2.33 = a + (1.5 * 0) \quad a = 2.33$$

$$y = 2.33 + 1.5x$$

$$Xm = 96 \rightarrow x = 96 - 94 = 2 \rightarrow y = 2.33 + (1.5 * 2) = 5.33$$

$$Xm = 97 \rightarrow x = 97 - 94 = 3 \rightarrow y = 2.33 + (1.5 * 3) = 6.83$$

مثال

در جدول زیر طول قد لوبیا (به سانتیمتر) بر حسب سن (به هفته) ملاحظه میکنید
معادله خط رگرسیون را بنویسید؟

اگر هفته دهم شد پیش بینی میکنید طول لوبیا دارای چه قدی باشد؟

سن	1	2	3	4	5	6	7			10		
طول	5	13	16	23	33	38	40			????		

حل :

سن را X و طول را Y مینامیم زیرا طول از سن تبعیت میکند پس سن متغیر مستقل است

$$\sum x_i = 1+2+\dots+7 = 28 \quad \sum y_i = 5+13+\dots+40 = 168$$

$$\sum x_i^2 = 1^2+2^2+\dots+7^2 = 140 \quad \sum x_i y_i = 1*5+2*13+\dots+7*40 = 844$$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{844 - \frac{28*168}{7}}{140 - \frac{(28)^2}{7}} = 6.143 \quad , \bar{x} = 28/7 \quad \bar{y} = 168/7$$

$$a = \bar{y} - b\bar{x} \quad a = (168/7) - (6.143)(28/7) = -0.572 \Rightarrow$$

$$y = a + bx$$

$$y = -0.572 + 6.143x$$

$$y = -0.572 + (6.143*10) = 60.85$$

تمرین :

طی آماری طول قامت ۵۰ کودک بشرح ذیل است

70(6)	75(9)	80(5)	85(7)	90(4)	95(7)
100(2)	105(2)	115(6)	120(1)	125(1)	

مد، میانه، دهک دوم و دهک هشتم و میانگین و واریانس و انحراف معیار را حساب کنید

احتمالات

تعداد کل حالات مطلوب تقسیم بر تعداد کل حالات ممکن

$$\text{احتمال} = \frac{\text{تعداد حالات خواسته شده}}{\text{تعداد کل حالات}}$$

* همیشه احتمال بین صفر و یک است *

۱) مثال

سکه ای دوبار پرتاب میکنیم احتمال حداقل یک شیر آمدن چقدر است

فضای کلی $S = \{HH, HT, TH, TT\}$

احتمال هر حالت $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$

فضای مورد نظر $A = \{HH, HT, TH\}$

احتمال $P(A) = 3/4$

تعداد شیر X	0	1	2
احتمال شیر $P=f(x)$	$1/4$	$2/4$	$1/4$

۲) مثال

تاسی ناریب که احتمال عدد زوج آن دو برابر فرد است داریم با یک بار پرتاب احتمال اینکه عدد کمتر از ۴ بیاید چقدر است

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

$w \quad 2w \quad w \quad 2w \quad w \quad 2w$

$w+2w+w+2w+w+2w=1$

$9w=1 \quad w = 1/9$

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

$1/9 \quad 2/9 \quad 1/9 \quad 2/9 \quad 1/9 \quad 2/9$

$A = \{1, 2, 3\}$

$1/9 \quad 2/9 \quad 1/9$

$P(A) = 1/9+2/9+1/9=4/9$

تابع احتمال

اگر تمام احتمالات یک رویداد را بنویسیم به آن تابع احتمال گوئیم $f(x)$

توزیع احتمال نرمال X

$$X \sim N(\mu, \sigma^2)$$

توزیع نرمال استاندارد

اگر در توزیع نرمال میانگین مساوی صفر و واریانس یک باشد آنگاه توزیع نرمال استاندارد نامیده میشود

$$Z \sim N(0,1)$$

ثابت شده است که در طبیعت آمار بصورت نرمال توزیع شده است که معمولاً ۶۸٪ بین $\mu \pm \sigma$ و ۹۶٪ بین $\mu \pm 2\sigma$ خواهد بود

توزیع نرمال استاندارد Z

اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد آنگاه متغیر تصادفی $Z = \frac{X - \mu}{\sigma}$ دارای توزیع نرمال استاندارد است با میانگین صفر و واریانس یک $Z \sim N(0,1)$

$$X \sim N(\mu, \sigma^2)$$

$$Z \sim N(0,1)$$

و احتمال آن چنین خواهد بود

$$P(x \leq b)$$

$$p\left(\frac{x - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = p\left(z \leq \frac{b - \mu}{\sigma}\right) \Rightarrow$$

حالا از روی جدول نرمال استاندارد احتمال حاصل میشود

در حالت جمعیت m تایی که نمونه n تایی از داخل m انتخاب کنیم در فرمول بالا بجای σ مقدار $\sqrt{\frac{\sigma^2}{n}}$

طریقه استفاده از جدول نرمال

جدول برای مقادیر کوچکتر یا مساوی است

Z	0	0.01	0.02	0.03	0.04	0.05
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008
-3	0.0012	0.0013	0.0013	0.0012	0.0012	0.0011
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071

در اولین ستون سمت چپ مقدار عدد مربوط به Z که هم مقادیر مثبت و هم منفی دارد و در اولین سطر رقم صدگان بصورت مثبت و دنباله Z است که با هم جمع جبری میشود

در داخل جدول از سمت چپ بالا احتمال صفر است و در سمت راست پایین عدد احتمال یک است
مثلا اگر سوال شود که $P(z \leq -2.73) = ?$

با توجه به جدول فوق

$$P(z \leq -2.73) = 0.0032$$

جدول زیر جدول توزیع احتمال نرمال استاندارد است

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

$Z_{0.05} = a$
 $p(z > a) = 0.05$
 $1 - p(z \leq a) = 0.05$
 $p(z \leq a) = 0.95$
 $z_{\alpha} = -z_{1-\alpha}$

جدول توزیع نرمال استاندارد

۳) مثال

در یک کلاس با ۴۰ دانشجو دارای توزیع نرمال و میانگین نمرات ۱۵ و انحراف معیار ۴ میباشد یک دانشجو انتخاب میکنیم. الف) احتمال اینکه نمره دانشجو حداکثر ۱۰ شود چقدر است؟ ب) احتمال اینکه نمره دانشجو بیشتر از ۱۰ شود چقدر است (جدول پیوست شده) ج) احتمال اینکه نمره دقیقاً ۱۰ شود؟
حل: توجه شود مسئله ذکر کرده توزیع نرمال است و توجه شود که جدول برای مقادیر کوچکتر یا مساوی و همچنین جدول برای نرمال استاندارد است

$$X \approx N(15, 4^2) \quad p(x \leq 10) = ? \quad p\left(\frac{x - \mu}{\sigma} \leq \frac{10 - 15}{4}\right) = ?$$

بدینترتیب میگوییم جمعیت به نرمال استاندارد تغییر یافت

الف) که از جدول نرمال استاندارد $p(z \leq -1.25) = 0.1056$

$$p(x > 10) = 1 - p(z \leq 10) = 1 - p\left(\frac{x - \mu}{\sigma} \leq \frac{10 - 15}{4}\right) = 1 - p(z \leq -1.25) =$$

ب) $1 - 0.1056 = 0.8944$

$$p(x = 10) = p(x \leq 10) - p(x \leq 9) = p\left(\frac{x - \mu}{\sigma} \leq \frac{10 - 15}{4}\right) - p\left(\frac{x - \mu}{\sigma} \leq \frac{9 - 15}{4}\right) =$$

ج) $= p(z \leq -1.25) - p(z \leq -1.5) = 0.1056 - 0.0668 = 0.0388$

توزیع احتمال برنولی (دوجمله ای) (دو وضعیتی)

آزمایشی که دارای دو نتیجه باشد (موفقیت S و شکست F) احتمال موفقیت را p و احتمال شکست $q = 1 - p$ میگردد این آزمایش را برنولی گویند

اگر آزمایش فوق را n بار بطور مستقل تکرار میکنیم آنرا دوجمله ای گویند

میانگین در این آزمایش برابر است با

$$\mu = \bar{x} = n * p$$

واریانس و انحراف معیار در این آزمایش برابر است با

$$\sigma^2 = n * p * q \quad \sigma = \sqrt{n * p * q}$$

اگر تعداد این آزمایش یعنی n کمتر از ۲۰ باشد یک جدول دارد که جزء درس ما نمیباشد

اگر تعداد این آزمایش یعنی n بیشتر از ۲۰ باشد با بدست آوردن میانگین و واریانس و انحراف معیار، توزیع احتمال دوجمله ای را به نرمال تقریب میزنیم و مثل مثال فوق حل میکنیم

۴) مثال

از سوابق یک تیرانداز مشخص میشود که ۸۰٪ تیرهایش به هدف اصابت میکند - این تیر انداز فردا میخواهد ۱۰۰

تیر شلیک کند الف) نوع تابع توزیع احتمال را نام ببرید

ج) احتمال اینکه حداکثر ۷۶ پرتابش به هدف اصابت کند چقدر است؟

د) احتمال اینکه بیش از ۸۶ بار تیرهایش به هدف اصابت کند چقدر است؟

ه) احتمال اینکه دقیقاً ۸۰ بار تیرهایش به هدف بخورد چقدر است

جدول بشرح ذیل در نظر بگیرید

$$p(z \leq 2) = 0.98, \quad p(z \leq 1.5) = 0.93, \quad p(z \leq 1) = 0.85, \quad p(z \leq 0.5) = 0.7, \quad p(z \leq 0) = 0.5$$

$$p(z \leq -2) = 0.02, \quad p(z \leq -1.5) = 0.07, \quad p(z \leq -1) = 0.15, \quad p(z \leq -0.5) = 0.3$$

حل: پرتاب تیر تابع توزیع احتمال دو جمله ای است

چون تعداد بیش از ۲۰ است پس توزیع نرمال است

$$p = 0.8 \quad q = 1 - p = 1 - 0.8 = 0.2$$

در تابع توزیع دو جمله ای میانگین برابر است با

$$\mu = np = 100 * 0.8 = 80$$

در تابع توزیع دو جمله ای انحراف معیار برابر است با

$$\sigma = \sqrt{npq} = \sqrt{100 * 0.8(0.2)} = \sqrt{16} = 4$$

چون در صورت سوال مطرح شده که پرتاب ها نرمال میباشند پس میتوان گفت که میتوان این توزیع دو جمله ای را از طریق جدول توزیع نرمال حل نمود

$$P(x \leq 76) = P\left(\frac{x - \mu}{\sigma} \leq \frac{76 - 80}{4}\right) = P(z \leq -1) = 0.15$$

$$P(x > 86) = 1 - P(x \leq 86) = 1 - P\left(\frac{x - \mu}{\sigma} \leq \frac{86 - 80}{4}\right) = 1 - P(z \leq 1.5) = 1 - 0.93 = 0.07$$

$$P(x = 80) = P(x \leq 80) - P(x \leq 79) = P\left(\frac{x - \mu}{\sigma} \leq \frac{80 - 80}{4}\right) - P\left(\frac{x - \mu}{\sigma} \leq \frac{79 - 80}{4}\right) = P(z \leq 0) - P(z \leq -0.25)$$

$$= 0.5 - 0.4013 = 0.0987 \approx 0.1$$

احتمال توام

اگر دو تابع احتمال به هم ربط داشته باشند، تابع جدید مرتبط با آن دو آنرا تابع احتمال توام گویند

** بر آورد فاصله ای

سطح زیر منحنی ۱ یک است اگر $\alpha/2$ در سمت راست منحنی و $\alpha/2$ در سمت چپ منحنی باشد

بنابراین $1 - \alpha$ درصد مطمئن هستیم که z بین $Z_{\alpha/2}$ و $-Z_{\alpha/2}$ است

$$P(-|Z_{\alpha/2}| < Z < |Z_{\alpha/2}|) = 1 - \alpha$$

$$P(-|Z_{\alpha/2}| < \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} < |Z_{\alpha/2}|) = 1 - \alpha$$

$$P(\bar{x} - |Z_{\alpha/2}| \sqrt{\sigma^2/n} < \mu < \bar{x} + |Z_{\alpha/2}| \sqrt{\sigma^2/n}) = 1 - \alpha$$

بنابراین $1 - \alpha$ درصد مطمئن هستیم میانگین جمعیت بین $\bar{x} - Z_{\alpha/2} \sqrt{\sigma^2/n}$ و $\bar{x} + Z_{\alpha/2} \sqrt{\sigma^2/n}$ است

** آزمون فرضها:

میخواهیم بحث قبول یا رد فرض صفر را بررسی کنیم

فرض H_α مشابه خواسته مسئله در نظر میگیریم و فرض H_0 را مساوی آن مقدار در مسئله قرار میدهم

۱- اگر در فرض H_α علامت کوچکتر بود آنگاه اگر $Z < -|Z_\alpha|$ بود فرض H_0 را رد میکنیم و پذیرای

فرض مقابل یعنی H_α میشویم وگرنه اعلام میکنیم دلیلی بر رد H_0 نداریم

۲- اگر در فرض H_α علامت بزرگتر بود آنگاه اگر $Z > +|Z_\alpha|$ بود فرض H_0 را رد میکنیم و پذیرای

فرض مقابل یعنی H_α میشویم وگرنه اعلام میکنیم دلیلی بر رد H_0 نداریم

حل مسائل آزمون فرضها

نرمال: در مسئله میانگین جمعیت \bar{x} و انحراف معیار جمعیت S و میانگین ادعا μ_0 و یک عدد بعنوان α در نظر گرفته سپس از جدول Z مربوط به α پیدا کرده و یک Z از $\frac{\bar{x} - \mu_0}{\sqrt{s^2/n}}$ در جدول بدست آورده و آنگاه جواب مسئله طبق شرایط فوق مقایسه میکنیم

مثال

لامپهای ساخت کارخانه ای با انحراف معیار ۱۴۲ ساعت در یک نمونه ۱۰۰ تایی میانگین عمرشان ۱۲۸۰ ساعت شد صاحب کارخانه ادعا میکند که با همین اطلاعات میانگین عمر لامپهای ای کارخانه از ۱۲۰۰ ساعت بیشتر است آیا میتوان ادعای او را قبول یا رد کرد؟ (هر وقت آلفا مشخص ندادند مقدارش ۰.۰۵ در نظر میگیریم)

$$n = 100 \quad \bar{x} = 1280 \quad s = 142$$

$$H_a : \mu > 1200$$

$$H_0 : \mu = 1200$$

$$\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{142^2}{100}} \quad Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{1280 - 1200}{\sqrt{\frac{142^2}{100}}} = 5.63$$

$$Z_{\alpha} = Z_{0.05} = m \Leftrightarrow p(z \leq m) = 1 - 0.05 = 0.95$$

$$\text{from table} \Rightarrow m = 1.645 = Z_{0.05} = Z_{\alpha}$$

$$\text{صحیح } (Z = 5.63) > (+Z_{\alpha} = +1.645)$$

چون $Z > +Z_{\alpha}$ میباشد فرض H_0 را رد میکنیم و پذیرای فرض مقابل یعنی H_a میشویم یعنی بله ادعای صاحب کارخانه میتواند درست باشد.

آزمون فرضها: میخواهیم فرض های زیر را تحلیل کنیم

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1: \mu_i \neq \mu_j$$

فرض H_0 چهار میانگین با هم مساوی هستند.

فرض H_1 حداقل بین دو میانگین اختلاف معنی دار وجود دارد.

امید و واریانس و کوواریانس

امید

امید همان معدل همان میانگین (توقع، انتظار) میباشد واریانس هم قبلا عنوان شده است میدانیم که: احتمال p عبارت است: تعداد حالات خواسته شده تقسیم بر تعداد حالات کل تابع احتمال $f(x)$ شامل تمام احتمالات x است بنابراین میتوان گفت $p=f(x)$ میباشد امید = توقع = انتظار = میانگین در احتمال برابر است با

$$\mu = \bar{x} = E(x) = \sum xf(x)$$

واریانس و انحراف معیار

برای تشخیص پراکندگی بین اعداد متغیرها میباشد

$$\sigma^2(x) = E(X^2) - (E(X))^2 \quad \text{واریانس}$$

$$\sigma(x) = \sqrt{\sigma^2(x)} \quad \text{انحراف معیار}$$

کوواریانس

برای تشخیص رابطه بین دو متغیر (دو سری اعداد) میباشد

$$Cov(x, y) = E(x \cdot y) - E(x) \cdot E(y)$$

اگر x و y مستقل باشند آنگاه

$$Cov(x, y) = 0$$

توجه شود کوواریانس صفر الزاما به مفهوم استقلال متغیرهای تصادفی نیست.

اگر x و y مستقل آنگاه $Cov=0$

اگر تغییرات x در خلاف جهت تغییرات y آنگاه $Cov < 0$

اگر تغییرات x در هم جهت تغییرات y آنگاه $Cov > 0$

در حقیقت Cov رابطه بین تغییرات دو متغیر تصادفی را نشان میدهد

ضریب تغییرات

قبلا اشاره شد برای تعیین پراکندگی یک سری داده از واریانس و انحراف معیار استفاده میشود که این شاخصها دارای واحد هستند و هر عدد بسیار کوچک یا بزرگ میتواند بدست آید سپس اشاره به شاخص پراکندگی دیگری به نام ضریب تغییرات (Coefficient of Variation- CV) نمودیم که حاصل تقسیم انحراف معیار بر میانگین بود که این شاخص میزان نسبی پراکندگی را نشان میدهد و بدون واحد میباشد و عدد کوچکی میباشد که به صورت درصدی نیز می تواند بیان شود. بنابراین ممکن است برای یک سری داده گفته شود که ضریب تغییرات ۱۵٪ است. این امر به معنی آن است که انحراف معیار ۱۵ درصد میانگین است.

$$\rho = \frac{\sigma}{\mu} \times 100$$

ضریب تغییرات هر چه کوچکتر (به صفر نزدیکتر) باشد یعنی دادهها یک دست تر میباشد.

برای مقایسه پراکندگی دو گروه داده که با واحدهای اندازه‌گیری متفاوتی جمع‌آوری شده باشند روش دیگری باید تدوین شود

کوواریانس

«کوواریانس» (Covariance)، شاخصی است که میزان هماهنگی یا تغییرات بین دو سری اعداد را نشان می‌دهد. اگر جهت تغییرات بین دو متغیر هم جهت باشد مقدار کوواریانس مثبت و در صورتی که دو متغیر در جهت عکس یکدیگر تغییر کنند، مقدار کوواریانس منفی خواهد بود. و اگر تغییرات اعداد دو متغیر بهم هیچ ربطی نداشته باشد کوواریانس صفر می‌باشد.

محاسبه کوواریانس

مثلا میزان سفره هاب زیر زمین در سالهای مختلف

سال	1394	1395	1396	1397	1398
میزان آب	۱۲	۱۱	۸	۷	۶

در جدول فوق با افزایش سال میزان آب کمتر میشود کوواریانس منفی میشود

مثلا ارزش سهام در شرکتهای مختلف و سودی که به سهامداران میدهند

	A	B	C	D	E	F	G	H	I	G
سهام	۳۰۰	۱۰۰	200	150	400	250	500	450	50	120
سود	۳۰	۸	12	11	40	13	51	45	4	14

در جدول فوق با افزایش سهام میزان سود بیشتر میشود کوواریانس مثبت میشود با افزایش یا کاهش یکی، دیگری هم به همان ترتیب افزایش یا کاهش خواهد داشت.

پس کوواریانس نشان دهنده شدت رابطه بین اعداد دو متغیر می‌باشد

کوواریانس به شکل زیر محاسبه می‌شود.

$$Cov(x, y) = \overline{xy} - \bar{x} \bar{y}$$

محاسبه کوواریانس برای نمایش تغییرات بین دو متغیر خوب است ولی اشکالش این است که کوواریانس بستگی به واحد اعداد متغیرها دارد (واحد اعداد کیلوگرم ، متر ، ریال ، میتواند باشد و عدد کوواریانس متناسب با آن اعداد میباشد)

*ضریب همبستگی (کورولیشن) = ضریب همبستگی پیرسون = ضریب همبستگی

چون کوواریانس وابسته به واحد اعداد است - مقیاس دیگری بنام ضریب همبستگی (با کورولیشن) برای تشخیص رابطه بین دو سری اعداد (متغیرها) معرفی میگردد

$$\rho = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

$$-1 \leq \rho \leq 1$$

اگر x و y رابطه خطی مستقیم داشته باشند $\rho=1$ میباشد
 اگر x و y رابطه معکوس خطی داشته باشند $\rho=-1$ میباشد
 اگر x و y رابطه نداشته باشند $\rho=0$ میباشد

ضریب همبستگی پیرسون

ضریب همبستگی پیرسون که به نام های ضریب همبستگی گشتاوری و یا ضریب همبستگی مرتبه ی صفر نیز نامیده می شود. این ضریب به منظور تعیین میزان رابطه، نوع و جهت رابطه ی بین دو متغیر فاصله ای یا نسبی و یا یک متغیر فاصله ای و یک متغیر نسبی به کار برده می شود. چندین روش محاسباتی معادل می توان برای محاسبه ی این ضریب تعریف نمود.
 روش محاسبه با استفاده از اعداد:

$$\rho(x, y) = Corr(x, y) = r = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{\overline{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y}$$

- ✓ ضریب بین ۰ تا ۰٫۲۹ نشان دهنده همبستگی ضعیف
- ✓ ضریب بین ۰٫۳۰ تا ۰٫۶۹ نشان دهنده همبستگی متوسط
- ✓ ضریب بین ۰٫۷۰ تا ۱ نشان دهنده همبستگی قوی
- ✓ ضریب همبستگی مثبت یعنی دوسری عدد هم جهت و مستقیم هستند
- ✓ ضریب همبستگی منفی یعنی دوسری عدد مخالف جهت و معکوس هستند

ضریب همبستگی پیرسون بین -۱ و ۱ تغییر می کند.

اگر $r=1$ بیانگر رابطه ی مستقیم کامل بین دو متغیر است، رابطه ی مستقیم یا مثبت به این معناست که اگر یکی از متغیرها افزایش (کاهش) یابد، دیگری نیز افزایش (کاهش) می یابد. مانند رابطه ی بین میزان ساعات مطالعه در روز و معدل محصلین.

$r=-1$ نیز وجود یک رابطه ی معکوس کامل بین دو متغیر را نشان می دهد. رابطه ی معکوس یا منفی نشان می دهد که اگر یک متغیر افزایش یابد متغیر دیگر کاهش می یابد و بالعکس.

زمانی که ضریب همبستگی برابر صفر است نشان می دهد که بین دو متغیر رابطه ی خطی وجود ندارد.

نکته:

(۱) صفر بودن ضریب همبستگی تنها عدم وجود رابطه ی خطی بین دو متغیر را نشان می دهد ولی نمی توان مستقل بودن دو متغیر را نیز نتیجه گرفت. هنگامی که ضریب همبستگی پیرسون بین دو متغیر صفر باشد، این متغیرها تنها در صورتی مستقل از یکدیگرند که توزیع متغیرها نرمال باشد.

(۲) همبستگی بین دو متغیر تنها نشان دهنده ی این است که افزایش یا کاهش یک متغیر چه تاثیری بر افزایش یا کاهش متغیر دیگر دارد ولی این همبستگی ضرورتاً دال بر رابطه ی علی بین متغیرها نمی باشد. به طور مثال اگر در یک تحقیق دو متغیر قد و تحصیلات همبستگی مثبت بالایی داشته باشند نمی توانیم نتیجه بگیریم که افراد قد بلندتر دارای تحصیلات بیشتری هستند. بنابراین باید بین مفاهیم همبستگی و رابطه ی علت و معلولی تفاوت قائل شد. به بیان دیگر ممکن است دو متغیر همبستگی داشته

باشند ولی لزومی ندارد که یکی از متغیرها علت و دیگری معلول باشد، علاوه بر این عوامل متعدد دیگری نیز می توانند بر ضریب همبستگی اثرگذار باشند.

ضریب همبستگی

ضریب همبستگی شاخصی است که شدت هماهنگی یا تغییرات بین اعداد دو متغیر را نشان می دهد.

(۱) ضریب همبستگی پیرسون

اگر با فرمول زیر ضریب همبستگی بدست آوریم به نام «ضریب همبستگی پیرسون» (Pearson Correlation Coefficient) معروف است.

$$\rho(x, y) = Corr(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{\bar{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y}$$

مثال

رابطه بین ساعات استفاده از اینترنت با نمره کسب شده توسط دانشجویان

X	۱۰	۱۴	۱۵	۱۵	۱۳	۱۴	۱۳	۱۱	اعتیاد به اینترنت
Y	۲۰	۱۴	۱۲	۱۶	۱۸	۱۵	۱۷	۲۰	نمره

حل: به روش پیرسون

$$\rho(x, y) = Corr(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{\bar{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y}$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{10 + 14 + 15 + 15 + 13 + 14 + 13 + 11}{8} = 13.125$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{20 + 14 + 12 + 16 + 18 + 15 + 17 + 20}{8} = 16.5$$

x*y	200	196	180	240	234	210	221	220	نمره * اعتیاد
-----	-----	-----	-----	-----	-----	-----	-----	-----	---------------

$$\bar{xy} = \frac{\sum x_i y_i}{n} = \frac{200 + 196 + 180 + 240 + 234 + 210 + 221 + 220}{8} = 212.625$$

کوواریانس

$$Cov(x, y) = \bar{xy} - \bar{x} \bar{y} = 212.625 - (13.125 * 16.5) = -3.9375$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 * f_i}{\sum f_i} \quad \sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} = 2.859 \quad \sigma_x = 1.69 \quad \sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = 7 \quad \sigma_y = 2.64$$

$$\rho(x, y) = Corr(x, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{\bar{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y} = \frac{212.625 - (13.125 * 16.5)}{1.69 * 2.64} = \frac{-3.937}{1.69 * 2.64} = -0.88$$

✓ ضریب بین ۰.۷۰ تا ۱ نشان دهنده همبستگی قوی

عدد ۰.۸۸ بیانگر همبستگی قوی بین X و Y است عدم منفی همبستگی در جهت معکوس است یعنی هر چه

اینترنت بیشتر استفاده شود نمره دانشجو کمتر میشود Pearson Correlation = -0.88

اگر دو متغیر از توزیع مشخصی (مثلا: مقیاس نسبتی و وزنی: نرمال - نمایی - کوشی - ..) پیروی کنند برای بررسی رابطه آن از ضریب پیرسون استفاده می کنیم و اگر هر دو ناپارامتری باشند (مثلا برای: مقیاس رتبه‌ای: فوق لیسانس - لیسانس .. و کاملا خوب و خوب و متوسط و ..) از ضریب همبستگی اسپیرمن استفاده خواهیم کرد نکته مهم این است که اگر یکی از آن‌ها پارامتری و دیگری ناپارامتری باشد دیگر این دو روش کاربرد ندارند و برای بررسی ارتباط میان دو متغیر باید از روش رگرسیون استفاده کنیم.

ضریب همبستگی رتبه‌ای اسپیرمن

این ضریب میزان همبستگی رابطه‌ی میان دو متغیر ترتیبی را نشان می دهد و به عبارت دیگر متناظر ناپارامتری ضریب همبستگی پیرسون می باشد. در این ضریب همبستگی به جای استفاده از خود مقادیر متغیرها از رتبه‌های آنان استفاده می شود. رابطه‌ی مربوط به ضریب همبستگی رتبه‌ای اسپیرمن به صورت زیر تعریف می شود.

$$r_s = 1 - \frac{6 \sum_{i=1}^n D^2}{n(n^2 - 1)}$$

که در آن: D تفاوت بین رتبه‌های اعضای متناظر دو گروه مورد بررسی و n حجم هر گروه.

ضریب همبستگی اسپیرمن

$$r_s = 1 - \frac{6 \sum_{i=1}^n D^2}{n(n^2 - 1)}$$

که D اختلاف رتبه‌های بین دو گروه و n تعداد میباشد.

مثال

بین ۷ دانشجو نمرات درس آمار در دانشگاه و نمره ورزش در دبیرستان آنان بشرح زیر است ضریب همبستگی اسپیرمن کدام است؟

نمره ورزش در دبیرستان	نمره درس آمار در دانشگاه
۲۰	۱۰
۱۵	۸
۱۲	۱۲
۱۰	۱۴
۱۵	۴
۱۵	۱۰
۱۶	۵

برای محاسبه ضریب اسپیرمن باید نمره‌ها را رتبه‌بندی کنیم بالاترین نمره رتبه ۱ میشود و بترتیب ادامه میدهم (مثلا در آمار نمره ۱۴ رتبه ۱ میشود و نمره ۱۲ رتبه ۲ میشود و نمره ۱۰ رتبه ۳ و نمره ۱۰ بعدی رتبه ۴ میشود در اینجا باید عدد این دو رتبه با هم جمع تقسیم بر دو نماییم جمع ۳ و ۴ تقسیم بر ۲ که ۳٫۵ میشود و به هر دو رتبه ۳٫۵ میدهم و به هیچکس رتبه ۳ و ۴ تخصیص نمیدهیم و نمره بعدی که ۸ است رتبه ۵ میشود و) (مثلا در درس ورزش نمره ۲۰ رتبه اول میشود و نمره ۱۶ رتبه دوم میشود و نمره ۱۵ سه نفر داریم که رتبه ۳ و ۴ و ۵ میشوند سه عدد ۳ و ۴ و ۵ با هم جمع و تقسیم بر ۳ مینماییم که جواب ۴ میشود و به هر سه نفر رتبه ۴ میدهم و

از این بعد رتبه ۳ و ۴ و ۵ را به هیچ شخصی اختصاص نمیدهیم - نمره بعدی ۱۲ میباشد که رتبه اش ۶ میشود و (...)

A	Ra	rA	M	Rm	rM	D=rA-rM	D ²
۱۰	3	۳,۵	۲۰	1	۱	2.5	6.25
۸	5	۵	۱۵	3	۴	1	1
۱۲	2	۲	۱۲	6	۶	-4	16
۱۴	1	۱	۱۰	7	۷	-6	36
۴	7	۷	۱۵	4	۴	3	9
۱۰	4	۳,۵	۱۵	5	۴	-0.5	0.25
۵	6	۶	۱۶	2	۲	4	16

حال اعداد را در فرمول جایگذاری می کنیم:

$$r_s = 1 - \frac{6 \sum_{i=1}^n D^2}{n(n^2 - 1)} = 1 - \frac{6 * \{2.5^2 + 1^2 + (-4)^2 + (-6)^2 + (3)^2 + (-0.5)^2 + (4)^2\}}{7(7^2 - 1)} = 1 - \frac{6 * 84.5}{335} = -0.51$$

ضریب همبستگی متوسطی در جهت معکوس وجود دارد تقریباً میتوان گفت هرچه نمره ورزش بیشتر بوده نمره آمار کمتر شده است

مثال

رابطه بین ساعات استفاده از اینترنت با نمره کسب شده توسط دانشجویان

X	۱۰	۱۴	۱۵	۱۵	۱۳	۱۴	۱۳	۱۱	اعتیاد به اینترنت
Y	۲۰	۱۴	۱۲	۱۶	۱۸	۱۵	۱۷	۲۰	نمره

حل: به روش اسپیرمن

X	۱۰	۱۴	۱۵	۱۵	۱۳	۱۴	۱۳	۱۱
Rx	1	5	8	7	4	6	3	2
rX	1	5.5	7.5	7.5	3.5	5.5	3.5	2
Y	۲۰	۱۴	۱۲	۱۶	۱۸	۱۵	۱۷	۲۰
Ry	7	2	1	4	6	3	5	8
rY	7.5	2	1	4	6	3	5	7.5
D=rX-rY	-6.5	3.5	6.5	3.5	-2.5	2.5	-1.5	-5.5
D ²	42.25	12.25	42.25	12.25	6.25	6.25	2.25	30.25

$$r_s = 1 - \frac{6 \sum_{i=1}^n D^2}{n(n^2 - 1)} = 1 - \frac{6 * \{(-6.5)^2 + (3.5)^2 + (6.5)^2 + (3.5)^2 + (-2.5)^2 + (2.5)^2 + (-1.5)^2 + (-5.5)^2\}}{8(8^2 - 1)} = 1 - \frac{6 * 154}{504} = -0.83$$

ضریب همبستگی قوی در جهت معکوس وجود دارد یعنی دانشجو هرچه اعتیاد به اینترنت بیشتر داشته باشد نمره کمتری خواهد گرفت

جدول فشار خون افراد در بازه سنین مختلف پس از مصرف سیر

سنین مختلف	فشار خون افراد مختلف بعد از مصرف سیر				
16-25	100	90	80	100	90
26-35	100	120	110	110	100
36-45	130	120	130	135	120
46-55	135	140	140	145	140

تفاوت ضریب همبستگی رتبه ای اسپیرمن و ضریب همبستگی پیرسون

- ضریب همبستگی پیرسون برای محاسبه ی همبستگی دو متغیر فاصله ای یا نسبی به کار برده می شود، ولی ضریب اسپیرمن ، همبستگی موجود بین دو متغیر ترتیبی را نشان می دهد.
- به کمک ضریب همبستگی اسپیرمن روابط غیرخطی بررسی می شود در حالیکه ضریب همبستگی پیرسون به منظور بررسی یک رابطه ی خطی بکار برده می شود.
- کارایی ضریب همبستگی رتبه ای اسپیرمن کمتر از ضریب همبستگی پیرسون است.
- محاسبه ی ضریب همبستگی اسپیرمن ساده تر بوده و نیاز به پیش فرض های کمتری نسبت به ضریب پیرسون دارد.

سایر ضرائب همبستگی

ضریب همبستگی کندال

موریس گریگور کندال به سال ۱۹۳۰ به مطالعه در مورد این ضریب پرداخت. دقت کنید ضریب هماهنگی کندال با ضریب همبستگی تاو کندال تفاوت دارد. کندال در ضریب همبستگی کندال دارای خواصی نظیر ضریب همبستگی ساده است. برای برآورد آن از آماره τ استفاده می‌شود.

ضریب هماهنگی توافقی کندال

ضریب همبستگی کندال که با نماد w نشان داده می‌شود یک آزمون ناپارامتریک است و برای تعیین میزان هماهنگی میان نظرات استفاده می‌شود. ضریب کندال بین ۰ و ۱ متغیر است. اگر ضریب کندال صفر باشد یعنی عدم توافق کامل و اگر یک باشد یعنی توافق کامل وجود دارد. ویژگی‌های ضریب کندال یکی از مهمترین کاربردهای این آزمون را در مدیریت فراهم کرده است. برای پایان راندهای تکنیک دلفی می‌توان از ضریب هماهنگی کندال استفاده کرد.

ضریب همبستگی چوپروف T :

ضریب همبستگی چوپروف به منظور تعیین شدت وابستگی بین متغیرهای مورد مطالعه به کار گرفته می‌شود و مقدار آن همواره بین صفر و یک در نوسان می‌باشد زمانی از آن استفاده کرده که هر دو متغیر اسمی و یا یکی اسمی و دیگری ترتیبی باشد. اما نباید تعداد سطر و ستون با هم برابر باشند. یعنی در جدول توافقی 2×2 نمی‌توان از آن استفاده کرد. در چنین مواردی باید از ضریب فی استفاده کرد.

ضریب همبستگی فی:

به منظور بررسی شدت همبستگی بین دو متغیر اسمی که جدول توافقی 2×2 در 2 می‌باشد مورد استفاده قرار می‌گیرد. حتی دو سطح معنی دار بودن همبستگی بین دو متغیر را تعیین میکند اما ضریب فی شدت همبستگی آنها را نشان می‌دهد. مقدار آن همواره بین صفر و یک در نوسان است.

ضریب گرامر:

این ضریب برای تعیین میزان شدت همبستگی بین دو متغیر اسمی مورد استفاده قرار می‌گیرد و آن را با (V_2) نشان می‌دهند و مقدار آن نیز همواره بین صفر و یک در نوسان است. هم جدول توافقی بیشتر از 2×2 در 2 وهم برای مستطیلی بکار می‌رود

(۱) ضریب همبستگی رتبه‌ای کندال (Kendall coefficient)

ضریب همبستگی کندال برای داده‌های رتبه بندی شده زمانی که تعداد آزمودنی‌ها از ۱۰ نفر کمتر است، به کار می‌رود. کندال از تعداد توافقات و عدم توافقات در رتبه بندی‌ها برای محاسبه ضریب خود استفاده می‌کند. برای مثال فرض کنید استاد روانشناسی بالینی ۱۰ نفر از دانشجویان خود را براساس دانش روانشناسی و شایستگی آن‌ها برای شاغل شدن در این رشته رتبه بندی کرده است. برای بررسی رابطه بین دانش روانشناسی و شایستگی می‌توانیم از ضریب همبستگی کندال استفاده کنیم.

ضریب همبستگی رتبه‌ای کندال به tau-a کندال نیز معروف است، متقارن می‌باشد و مقدار آن بین +۱ و -۱ قرار دارد و مشابه با ضریب همبستگی پیرسون تفسیر می‌شود. فرمول محاسبه ضریب همبستگی رتبه‌ای کندال (tau-a)

$$\tau = \frac{P - Q}{\frac{1}{2}N(N - 1)}$$

P: تعداد توافقات یعنی تعداد اعداد (مواردی) که رتبه بالاتری برای هر رتبه متغیر دوم به دست آورده‌اند.

Q: تعداد عدم توافقات یعنی تعداد اعداد (مواردی) که رتبه پایین‌تری برای هر رتبه متغیر دوم به دست آورده‌اند.

N: تعداد آزمودنی‌ها (حجم نمونه)

داده های پارامتریک

داده های پارامتری به نمونه ای گفته می شود که از توزیع جامعه آماری آن مطلع هستیم معمولاً این توزیع آماری برای داده های کمی، نرمال یک یا چند متغیره در نظر گرفته می شود. در این حالت از آزمون های آماری پارامتری مثل آزمون T، آزمون F و یا آزمون Z استفاده می کنیم. همچنین برای اندازه گیری میزان همبستگی بین متغیرهای دو یا چند بعدی نیز از ضریب همبستگی پیرسون استفاده خواهیم کرد.

داده های ناپارامتریک

اگر توزیع جامعه آماری نامشخص باشد و یا حجم نمونه نیز کوچک باشد بطوری که نتوان از قضیه حد مرکزی برای تعیین توزیع حدی یا جانبی جامعه آماری، استفاده کرد، از تحلیل های ناپارامتری استفاده می شود. مثلاً در متغیرهای رتبه ای ابتدا گروه بندی نموده و سپس برای تحلیل از ضریب همبستگی اسپیرمن استفاده می شود. در آزمونهای ناپارامتریک از آزمونهای - میانه - چندک ها - sign و اسپیرمن و ویلکاکسون و من ویتنی و فریدمن استفاده میشود

آنالیز واریانس ANOVA

آزمون تحلیل واریانس یک طرفه یا آنوا (One-Way Analysis of Variance): در این حالت یک متغیر مستقل کیفی (چند سطحی) و یک متغیر وابسته وجود دارد. یکی از ابزارهای پرکاربرد در آزمون فرض و تحقیقات آماری، «تحلیل واریانس (Analysis of Variance)» است. در این روش سعی بر این است که اختلاف بین چند جامعه آماری، ارزیابی شود. با توجه به پراکندگی کل داده ها، تجزیه واریانس بین گروه های مختلف در این روش امکان پذیر است. به این ترتیب می توان برابر بودن میانگین را بین گروه های مختلف آزمود. همچنین در مدل های رگرسیونی با تجزیه واریانس کل به واریانس مدل و واریانس خطا تشخیص مناسب بودن مدل قابل ارزیابی است.

آنالیز کوواریانس ANCOVA

آزمون تحلیل واریانس یا کوواریانس تک متغیره (Univariate Analysis of Variance): در این حالت دو (یا بیشتر) متغیر مستقل کیفی (چند سطحی) و یک متغیر وابسته وجود دارد. آنالیز کوواریانس ANCOVA یا Analyze of Covariance آنکوا، نوعی آنالیز و تحلیل همانند آنوا می باشد. هرگاه در آنالیز واریانس بخواهیم اثر کمیت های مداخله گر را به روش های آماری حذف کنیم تا نتایج با دقت بیشتری به دست آید از آنالیز کوواریانس استفاده می شود. در این روش هم از کنترل آماری و هم از واریانس استفاده می شود. به عبارت دیگر به جای تحلیل واریانس تحلیل کوواریانس مورد استفاده قرار می گیرد. در واقع آنکوا ANCOVA مدل تعمیم یافته آنوا ANOVA و همچنین مدل های رگرسیونی است. تحلیل کوواریانس مناسبترین آزمون آماری برای طرح پیش آزمون و پس آزمون ۲ گروهی می باشد و تنها عامل تهدید کننده اعتبار درونی تحلیل کوواریانس طرح پیش آزمون است.

➤ آنالیز کوواریانس چندبعدي MANCOVA

آزمون تحلیل واریانس یا کواریانس چند متغیره (Multivariate Analysis of Variance): در این حالت دو (یا بیشتر) متغیر مستقل کیفی (چند سطحی) و دو (یا بیشتر) متغیر وابسته وجود دارد. آنالیز چندمتغیره کوواریانس ها MANCOVA Multivariable Analyze of Covariance یا مانکوا، روش تعمیم یافته آنکوا (آنالیز کوواریانس) است و در مواردی کاربرد دارد که بیش از یک کمیت مستقل وجود داشته باشد و همچنین در مواردی که کمیت‌های وابسته به آسانی نتوانند ترکیب شوند به کار می‌رود MANCOVA همان MANOVA با این تفاوت که به شما امکان می‌دهد اثرات اضافی کمیت‌های مستقل را نیز کنترل کنید. اگر چندین کوواریانس وجود داشته باشد مانکوا (MANCOVA) به جای مانوا (آنالیز واریانس چند متغیره) به کار برده می‌شود.

انتخاب آزمون

توزیع متغیر	نوع آنالیز
نرمال	آنالیز t - آنالیز واریانس
غیر نرمال	آزمون ناپارامتری

انتخاب آزمون یک متغیره و گروهها

یک متغیر			
نوع متغیر	یک گروه	دو گروه	بیش از دو گروه
نرمال	میانگین انحراف معیار	آزمون t	واریانس
غیر نرمال	نسبت	کای دو	ناپارامتری

انتخاب آزمون برای دو سری متغیر (بشکل مختصر)

متغیر اول	متغیر دوم	نوع آزمون
کیفی (اسمی)	کیفی (اسمی)	کای دو
پیوسته (رتبه ای) (ترتیبی)	گسسته (کمی) (فاصله ای - نسبی)	آنالیز واریانس
پیوسته (رتبه ای) (ترتیبی)	پیوسته (رتبه ای) (ترتیبی)	همبستگی

انتخاب آزمونهای آماری برای دو سری متغیر (در حالتهاى مختلف متغیرها)

آزمون های آماری	سطح سنجش متغیرها	
	متغیر اول = متغیر مستقل X	متغیر دوم = متغیر وابسته Y
کای دو - فی - وی کرامر - لاندا	اسمی	اسمی
کای دو - فی - وی کرامر - لاندا	ترتیبی	اسمی
تحلیل واریانس یکطرفه - تی تست - لاندا	فاصله ای یا نسبی	اسمی
کای دو - فی - وی کرامر - لاندا - تتا	اسمی	ترتیبی
ضریب همبستگی رتبه ای اسپیرمن	ترتیبی	ترتیبی
ضریب همبستگی اسپیرمن - کندال	فاصله ای - نسبی	ترتیبی
کای دو - فی - وی کرامر - لاندا - تتا	اسمی	فاصله ای - نسبی
ضریب همبستگی پیرسون ۲	فاصله ای - نسبی	فاصله ای - نسبی
ضریب همبستگی رتبه ای اسپیرمن - کندال	ترتیبی	فاصله ای - نسبی

انتخاب آزمون در انواع متغیرهای دو گروه و بیشتر

بیش از دو گروه		دو گروه		
وابسته	مستقل	وابسته	مستقل	متغیر
واریانس در تکرار مشاهدات	آزمون واریانس یکطرفه	آزمون زوجی t	آزمون مستقل t	کمی
فرید من	کروسکال - والیس	ویلکاکسون - آزمون علامتی	من - ویتنی	رتبه ای
کوکران	کای دو	مک نمار	کای دو	اسمی

انتخاب آزمون برای بیش از چند متغیر

بیشتر از یک گروه	یک گروه
واریانس چند متغیره - تحلیل ممیزی	کوواریانس - واریانس-رگرسیون چندگانه - تحلیل عاملی

انواع نمودارها

((۱) نمودار میله‌ای بارچارت Bar Chart:

برای مشاهده فراوانی داده‌ها و مقایسه داده‌ها نسبت به هم

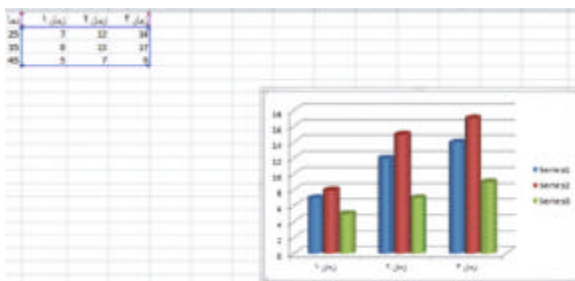


Changes in Consumer Attitudes toward the Gulf Real Estate Industry



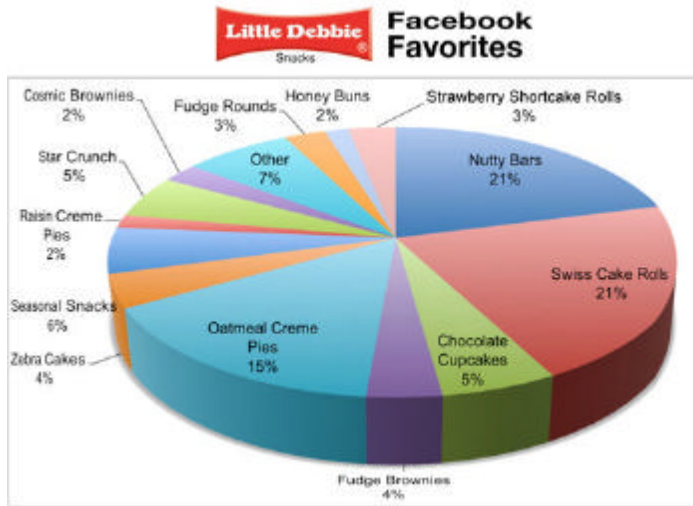
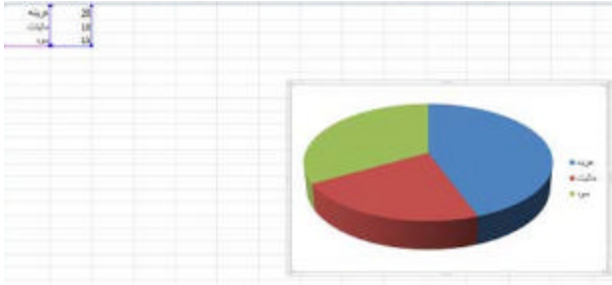
((۲) میله ای ستونی - مشاهده فراوانی و مقایسه

داده‌ها بصورت دسته بندی شده



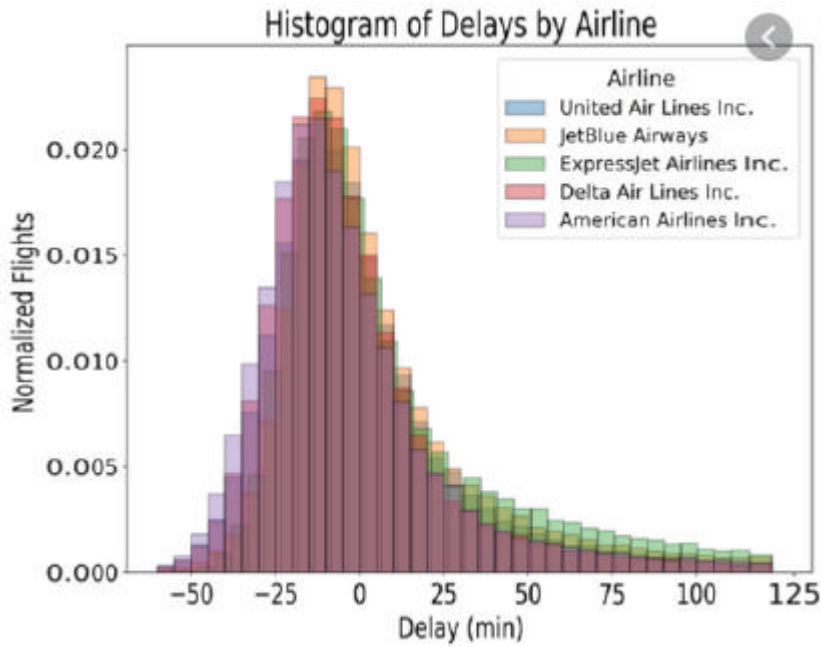
((۳ نمودار دایره ای Pie Chart :

برای مشاهده سهم هر مورد از داده ها

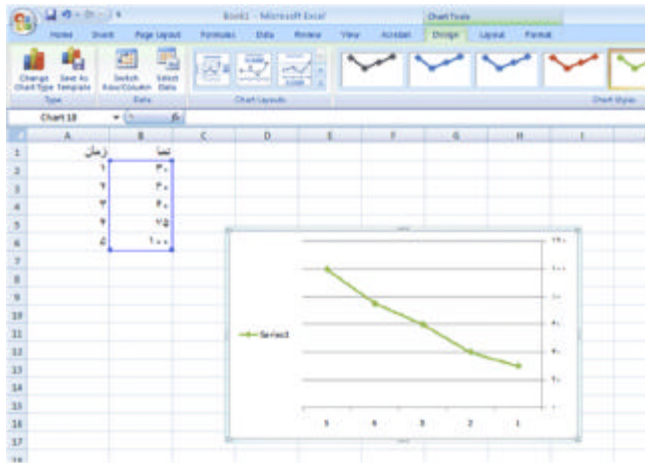


((۴ نمودار هیستوگرام Histogram:

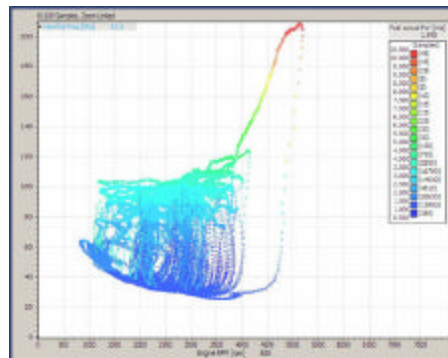
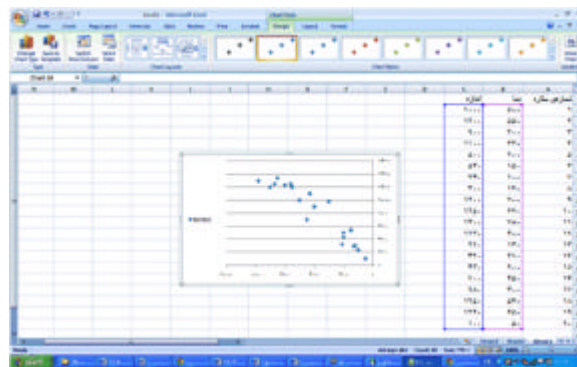
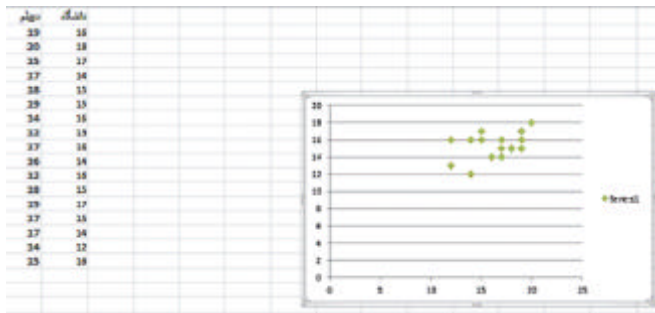
برای مشاهده نحوه توزیع داده ها بصورت پیوسته (مثلا معدل دیپلم و معدل دانشگاه)



نمودار خطی Line Chart: برای نشان دادن رابطه بین دوسری اعداد مثلا خط رگرسیون ((۵



نمودار پراکندگی Scatter Plot: برای نشان دادن پراکندگی بین دوسری اعداد مثلا در رگرسیون ((۶



انواع متغیرها و مقیاس‌های مرتبط

متغیرهای کیفی

۵. مقیاس اسمی (NOMINAL)

با کد گذاری عددی - اعداد نه ضریب و تناسب دارند و ارجحیت و برتری هم ندارند (مثل کد شهرها - مثل جنسیت - زن ۲ مرد ۱) (مثل آبی ۱ - زرد ۲ - قرمز ۳ - نارنجی ۴ - ...)

۶. مقیاس رتبه‌ای (ORDINAL) (ترتیبی)

با کد گذاری عددی - اعداد نسبت داده شده ارجحیت و برتری دارند ولی ضریب و تناسب ندارند (شدت و ضعف دارند) (مثل پرسشنامه لینکرت شامل موافق ۳ - بی نظر ۲ - مخالف ۱ یا تحصیلات مثلا دیپلم ۱ فوق دیپلم ۲ لیسانس ۳ و فوق لیسانس ۴ (که ۴ بیشتر از ۱ ولی ۴ چهار برابر ۱ نیست) (مثل ناراضی ۱ - بی نظر ۲ - راضی ۳)

متغیرهای کمی

۷. مقیاس نسبتی (SCALE) (وزنی)

با کد گذاری عددی - اعداد ارجحیت و برتری دارند و ضریب و تناسب هم دارند (مثل سن و مثل وزن - مثلا وزن احمد ۳۵ و وزن عباس ۷۰ کیلوگرم یعنی عباس بیشتر از احمد است و به نسبت دو برابر هم میباشد)

۸. مقیاس فاصله‌ای

مثلا به یک بیمار بگوییم اگر حداکثر درد مثل یک خط کش عدد ۱۰۰ باشد درد شما چقدر است و بیمار بگوید ۶۵

انواع متغیرها و شاخص‌ها

علامت	نام اختصاری	شاخص پراکندگی	شاخص مرکزی	مقیاس	نوع متغیر
	<i>Nominal</i>	جدول	نما (مد)	اسمی	کیفی (گسسته)
	<i>Ordinal</i>	فراوانی	میانه	ترتیبی	
	<i>Scale</i>	واریانس	میانگین	فاصله‌ای نسبتی	کمی (پیوسته)

روش‌های مقیاس‌گذاری

روش‌های مقیاس‌گذاری به دو دسته کلی غیرمقیاسه‌ای و مقیاسه‌ای تقسیم می‌شود. در ادامه انواع مقیاس‌های مقیاسه‌ای و غیرمقیاسه‌ای مورد بررسی قرار می‌گیرد.

۱. انواع مقیاس‌های مقیاسه‌ای

۱.۱. مقیاس مجموع ثابت:

در این مقیاس‌گذاری پاسخ‌دهنده موظف است از میان مجموعه ثابتی از گزینه‌ها براساس برخی معیارهای خاص، مجموع ثابتی از واحدها نظیر امتیاز یا پول را تخصیص بدهد. مثلاً ممکن است از پاسخ‌دهندگان بخواهند ۱۰۰ امتیاز را به پنج ویژگی اختصاص دهند. نتایج این نوع مقیاس معمولاً به صورت درصدی عنوان می‌شود.

۱.۲. آرایش رتبه‌ای:

در این روش گزینه‌های مختلف همزمان به پاسخ‌دهنده ارائه و از او خواسته می‌شود می‌خواهد گزینه‌ها را براساس برخی معیارها رتبه‌بندی کند. به‌طور مثال ۱۰ کتاب به خواننده داده شده و از وی خواسته می‌شود به کتابی که علاقه بیشتری به آن دارد رتبه یک و به کتابی که اصلاً به آن علاقه ندارد نمره ۱۰ دهد. این روش برای اندازه‌گیری اولویت کتابها و نیز ویژگی‌های کتاب به کار می‌رود.

۱.۳. مقیاسه جفتی:

در این روش یک جفت گزینه به مصرف‌کننده ارائه شده و از او خواسته می‌شود براساس برخی معیارها یکی از آنها را انتخاب کند.

۲. انواع مقیاس‌های غیرمقیاسه‌ای

۲.۱. طیف لیکرت:

این طیف یکی از رایج‌ترین مقیاس‌های اندازه‌گیری در تحقیقاتی است که براساس پرسشنامه انجام می‌شود و توسط رنسیس لیکرت ابداع شده است. در این مقیاس یا طیف، محقق با توجه به موضوع تحقیق خود، تعدادی گویه را در اختیار مصرف‌کنندگان قرار می‌دهد تا براساس گویه‌ها و پاسخ‌های چندگانه، میزان گرایش خود را مشخص کنند.

در این روش پاسخ‌ها به صورت چند گزینه‌ای است که به‌طور مثال در حالت پنج نقطه‌ای گزینه‌ها شامل «کاملاً مخالف، مخالف، نظری ندارم، موافق و کاملاً موافق» است. معمولاً در پرسشنامه‌ها براساس مقیاس لیکرت از حالت پنج‌گانه ذکر شده استفاده می‌شود، اما بسیاری از روان‌سنج‌ها از حالات هفت و نه‌گانه نیز استفاده می‌کنند.

گرچه مطالعات اخیر نشان می‌دهد مقیاس پنج و هفت نقطه‌ای نتایج معتبرتری نسبت به مقیاس ۱۰ نقطه‌ای دارند. سپس هر یک از گویه‌ها از نظر عددی ارزش‌گذاری می‌شوند. حاصل جمع عددی این ارزش‌ها نمره را در این مقیاس به دست می‌دهد که بیانگر گرایش پاسخ‌دهندگان است. به همین دلیل به این مقیاس، مقیاس مجموع نمرات نیز گفته می‌شود.

هدف از طیف لیکرت اندازه گیری گرایش به یک موضوع بر اساس ارزشهای جامعه می باشد و کاربرد این طیف نیز در جهت بررسی گرایشها نسبت به مسئله سیاسی - اجتماعی و اقتصادی می باشد که در سطح ترتیبی نیز مورد سنجش قرار دارد.

گویه ها در این طیف حداقل ۱۵ تا ۳۰ گویه و بیشتر تدوین می شود.

در تدوین گویه ها باید سعی شود از گویه های بی تفاوت، بی ربط و ابهام آور جلوگیری شود تعداد گویه هائی که گرایش مخالف و موافق دارند باید تقریباً به یک اندازه باشد و نیز طیفی که به پاسخگو داده می شود معمولاً از ۵ قسمت تشکیل شده است (کاملاً موافقم - موافقم - تاحدودی - مخالفم - کاملاً مخالفم) که براساس هدف و روش تحقیق می توان کلمات گویه ها را عوض نمود

۲.۲.۲. افتراق معنایی:

مقیاس افتراق معنایی، در مقایسه با مقیاس پنج امتیازی لیکرت، مقیاسی هفت امتیازی و دو قطبی است، در حالی که در مقیاس لیکرت هر عدد آیتم متعلق به مقیاس مشخص شده است در مقیاس افتراق معنایی، نقطه های ابتدا و انتها کاملاً مشخص شده اند. مثلاً می توان از راضی و ناراضی به عنوان نقاط ابتدا و انتها استفاده کرد.

۲.۲.۳. مقیاس استاپل:

مقیاس استاپل از معیاری واحد در میانه بازه مقادیر اعداد زوج از ۵- تا ۵+ و بدون نقطه صفر تشکیل شده است. معمولاً این مقیاس به صورت عمودی است و از پاسخگویان خواسته می شود تا عددی خاص را انتخاب کنند که شیء یا موضوع محرک مربوط به معیار از پیش تعیین شده را توصیف می کند. مزیت مقیاس استاپل این است که این مقیاس مانند مقیاس افتراق معنایی برای دستیابی به ویژگی دو قطبی بودن به هیچ عبارتی نیاز ندارد.

۲.۲.۴. مقیاس رتبه ای پیوسته:

این مقیاس این امکان را برای پاسخ دهنده فراهم می آورد که به جای اینکه از میان پاسخ های از پیش تعیین شده پاسخ مورد نظر را انتخاب کند، در هر نقطه از یک محور علامتی بگذارد. لازم به ذکر است این یادداشت ها جهت آشنایی مخاطبان با تحقیقات بازاریابی است و زمان استفاده از مقیاس های مورد نظر نیاز به مهارت و تجربه دارد. در یادداشت های آتی به بررسی روش های طراحی پرسشنامه پرداخته خواهد شد.

۲.۲.۵. طیف بوگاردوس

هدف از طیف بوگاردوس سنجش میزان فاصله اجتماعی گروهها می باشد.

و کاربرد این طیف در مواقعی است که مثلاً دو گروه مختلف با دو فرهنگ و زبان متفاوت در کنار هم زندگی می کنند و یا در مواقعی که مردمی از ملیتهای گوناگون و با فاصله جغرافیائی در جریان روابطی متقابل قرار میگیرند (مثلاً در سنجش نگرش به شغل / طبقه اجتماعی / گروههای مذهبی) این طیف بکار می رود .

این طیف در یک مقیاس ترتیبی (تمایل کامل/ تمایل متوسط/ تا حدودی/ عدم تمایل) میزان تمایل یک گروه را نسبت به سایر گروهها می سنجد.

آشنایی مقدماتی با نرم افزار SPSS

Statistical Package for the Social Sciences

نرم افزار را از دانشگاه یا از اینترنت یا محل کار خودتان یا فروشگاههای CD سطح شهر تهیه و نصب کنید لیسنس (گواهینامه) آنرا تایید نمایید
به کلیپ موجود در سایت www.aminsedighi.ir بخش آمار توصیفی مراجعه کنید

پروژه

نحوه تهیه و تنظیم پرسشنامه

☞ عنوان و هدف پرسشنامه : (مثلا هدف بررسی میزان محبوبیت سریال تلویزیونی)

☞ یک محل برای درج شماره پرسشنامه

☞ یک جمله خوش آمدگویی

پاراگراف حدودا سه خطی که مضمون خوش آمدگویی برای پاسخگو و علاقمند نمودن پاسخگو در پاسخ به سوالات

☞ جمعیت شناختی (مشخصات پاسخ دهندگان)

مشخصات تکمیل کنندگان فرم که این اطلاعات متغیرهای کمکی CoVariance برای طبقه بندی و مقایسه میباشد.

- مثلا سن: (که عددی و باز است) - جنسیت: (که عددی و بسته است) - تحصیلات : (عددی و بسته است) -

آدرس ایمیل : (عدد یا رشته و باز)

☞ بخش اصلی سوالات (پرسش ها)

با استفاده از روش مقیاس گذاری - مثلا ۱۰ سوال با طیف پنج تایی لیکرت (کاملا موافق - موافق - بی نظر -

مخالف - کاملا مخالف که عددی و بسته است)

یک نمونه پرسشنامه و اعمال نتایج آن در نرم افزار SPSS

نمونه طرح پرسشنامه

عنوان طرح					
طرح تحقیقاتی بررسی بعضی عوامل افسردگی					
شماره					
کد ردیف					
هدف از این طرح، بررسی علل و دلایل رخداد افسردگی میباشد بدیهی است همکاری و حوصله شما در پاسخ به سوالات راهگشای ما خواهد بود. از وقتی که میگذارید کمال امتنان را دارد					
<p>برای پاسخ به سوالات زیر در مقابل آنها پاسخ بنویسید یا تیک بزنید</p> <p>سن جنسیت مرد ● زن ●</p> <p>تحصیلات (زیر دیپلم و دیپلم ● - دانشجو و فوق دیپلم ● - لیسانس ● - فوق لیسانس ● - دکترا ●)</p> <p>آدرس ایمیل (اختیاری): "</p>					
سوالات کمکی: مشخصات پاسخ دهندگان					
برای پاسخ به سوالات زیر در جدول مقابل آنها تیک بزنید					
کام مخالف	مخالف	نظری ندارم	موافق	کاملاً موافق	
					Q1 - هرچه میزان تحصیلات افراد بیشتر باشد افسردگی کمتر رخ میدهد
					Q2 - سن افراد هیچ اثری در افسردگی ندارد
					Q3 - داشتن مال و اموال و پول زیاد در کاهش افسردگی اثر چشمگیری دارد
					Q4 - اگر افراد شاغل باشند کمتر دچار افسردگی میشوند
					Q5 - خانواده و اطرافیان نقش موثری در افزایش افسردگی دارند
سوالات اصلی تحقیق					

Variable Name نام متغیر	Variable Lable توضیح سوال متغیر	Type نوع	Measure مقیاس اندازه گیری	Value Lable توضیح مقادیر
Agge	سن	Numeric	Scale	عددی بین ۱ تا ۱۴۰ عدد منفی ۱ = بدون پاسخ
Gensiatt	جنسیت	Numeric	Nominal	۱ = مرد ۲ = زن ۱- = بدون پاسخ
Tahsilat	تحصیلات	Numeric	Ordinal	۱-دیپلم ۲- دانشجو و فوق دیپلم ۳- لیسانس ۴- ف ل ۵- دکترا
Email_Address	آدرس ایمیل	String	Nominal	Sedighias220@yahoo.com
Q1	سوال ۱ در اینجا کپی شود	Numeric	Ordinal	۵-کاملا موافق ۴-موافق ۳-بدون نظر ۲- مخالف ۱- کاملا مخالف
Q2	سوال ۲ در اینجا کپی شود	Numeric	Ordinal	۵-کاملا موافق ۴-موافق ۳-بدون نظر ۲- مخالف ۱- کاملا مخالف
Q3	سوال ۳ در اینجا کپی شود	Numeric	Ordinal	۵-کاملا موافق ۴-موافق ۳-بدون نظر ۲- مخالف ۱- کاملا مخالف
Q4	سوال ۴ در اینجا کپی شود	Numeric	Ordinal	۵-کاملا موافق ۴-موافق ۳-بدون نظر ۲- مخالف ۱- کاملا مخالف
Q5	سوال ۵ در اینجا کپی شود	Numeric	Ordinal	۵-کاملا موافق ۴-موافق ۳-بدون نظر ۲- مخالف ۱- کاملا مخالف

آمار زیستی Biostatistics

جزوه قسمت اول سرفصل آمار زیستی

- ۱) تعریف مفاهیم (ریاضی آمار اندازه گیری سنجش)
- ۲) طبقه بندی (آمار توصیفی و استنباطی)
- ۳) جامعه نمونه پارامتر مشخصه آماری
- ۴) مفهوم متغیر و انواع آن
- ۵) طبقه بندی متغیر از لحاظ مقیاس اندازه گیری
- ۶) انواع نمودار و کاربرد آن در آمار
- ۷) طبقه بندی آمار توصیفی (یک متغیری چند متغیری)
- ۸) شاخص های مرکزی (مد میانه میانگین)
- ۹) شاخص های پراکندگی (واریانس انحراف معیار انحراف چارکی و دامنه تغییرات)
- ۱۰) ضریب تغییرات - چولگی - کشیدگی
- ۱۱) نقاط درصدی (چندک ها) و رتبه بندی
- ۱۲) منحنی طبیعی و کاربرد آن در آمار
- ۱۳) نمره های استاندارد و انواع آن
- ۱۴) مختصری از احتمالات - تابع احتمال - تابع توزیع احتمال نرمال - تابع توزیع احتمال دوجمله ای - آزمون فرضها - امید ریاضی - کوواریانس
- ۱۵) شاخص های همبستگی و انواع آن با تکیه بر همبستگی پیرسون و اسپرمن
- ۱۶) رگرسیون و پیش بینی
- ۱۷) آشنایی با نرم افزارهای آماری از جمله SPSS
- ۱۸) نحوه تکمیل پرسشنامه
- ۱۹) پروژه

این جزوه با تلاش این حقیر، جهت دانلود مجانی برای همه شما سروران تهیه شده است

بدیهی است خالی از ایراد نیست

هر گونه اشکالی به sedighias220@yahoo.com ارسال نمایید

با تشکر (امین صدیقی)

=====پایان=====

در هر حرفه ای که هستید نه اجازه دهید که به بدبینیهایی بیحاصل آلوده شوید و نه بگذارید که بعضی لحظات تاسف بار که برای هر ملتی پیش می آید شما را به یاس و ناامیدی بکشاند. در آرامش حاکم بر آزمایشگاهها و کتابخانه هایتان زندگی کنید .

نخست از خود بپرسید: " برای یادگیری و خودآموزی چه کرده ام ؟ "

سپس همچنان که پیشتر میروید بپرسید: " من برای کشورم چه کرده ام ؟ "

و این پرسش را آنقدر ادامه دهید تا به این احساس شادبخش و هیجان انگیز برسید که شاید سهم کوچکی در پیشرفت و اعتلای بشریت داشته اید.

اما هر پاداشی که زندگی به تلاشهایمان بدهد یا ندهد هنگامی که به پایان تلاشهایمان نزدیک میشویم هر کدامان باید حق آن را داشته باشیم که با صدای بلند بگوییم

" من آنچه در توان داشته ام انجام داده ام "

لوئی پاستور ۱۸۹۵-۱۸۲۲