آمار استنباطی

آمار : جمع آوری و سازماندهی و خلاصه کردن داده ها در قالب عدد

آمار توصیفی: 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

آمار استنباطی : مشخص می کند که آیا الگوها وفرآیندهای کشف شده در نمونه آمار توصیفی ، در جامعه آماری هم کاربرد دارد

سرفصل آمار استنباطی (با یادآوری آمار توصیفی)

یادآوری آمار توصیفی

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
۲۰)) قوانین احتمالات

آمار استنباطی

۲۱)) نمونه گیری و خطا **Sampling Methods Errors**
۲۲)) روشهای نمونه گیری**Sampling Methods**
۲۳)) برآوردگرها**Estimators**
۲۴)) خطای معیار برآوردگرها - **Standard Errors SE**
۲۵)) ویژگی‌های برآوردگرها
۲۶)) برآوردگر ناریب**UnBiased**
۲۷)) قضیه حد مرکزی
۲۸)) توزیع تی**T-Student Distribution**
۲۹)) برآورد نقطه‌ای **Point Estimation** و فاصله‌ای **Internal Estimation**
۳۰)) برآورد واریانس و انحراف معیار
۳۱)) توزیع کای مربع = خی
۳۲)) برآوردگر باثبات (سازگار)**(Consistent** قانون قوی اعداد بزرگ**(**
۳۳)) برآوردگر کافی (بسنده) **(Sufficent** حداکثر درست‌نمایی**Maximum Liklihood(**
۳۴)) دقت برآوردگر **(** کارائی نسبی برآوردگر **Efficiency)**
۳۵)) آزمون فرض
۳۶)) آزمون فرض میانگین‌ها **Compare Means**

آمار

آمار = جمع آوری و سازماندهی و خلاصه کردن داده ها در قالب عدد

جمعیت = موجودات (جانداران یا موجوداتی که یک یا چند وجه مشترک داشته باشند مثلا وزن و قد و سن و شماره

نمونه = قسمتی از جمعیت

اگر دادها از کل جمعیت بدست آوریم سرشماری میگویند

بررسی داده جمعیت

اگر بخواهیم کل جمعبت را تست کنیم هزینه بر است و وقت گیر است و همه جمعیت در دسترس شاید نباشد امکان از دست رفتن بخشی از جمعیت حین تست انجام کار (و یا در تست کیفیت مقدار ویتامین **C** یک میوه اگر همه میوه ها یکی یکی باز شود و تست کیفیت شود و برای همه تک تک جمعیت میوه ها مقدار ویتامین **C** آن استخراج شود آنگاه آخرش چیزی نمی‌ماند برای فروش)

بهمین دلایل از جمعیت نمونه گیری انجام میشود

آمار توصیفی = روشهای سازماندهی و خلاصه کردن داده موجود **Descriptive Statistics**

آمار استباطی = آیا اطلاعات بدست آمده در نمونه برای جعیت عمومیت دارد به عبارتی روشهای نمونه گیری و بررسی خطاها و تخمین های نقطه ای و فاصله ای پارامترها

آمار استنباطی = **Inferential Statistics**

Estimation / Point / Interval

علم آمار

آمار نقشی لاینفک در زندگی روزمره ما بازی می کند. اخبار روزانه رسانه‌های 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

☜ آمار توصیفی زیستی *Biostatistics*

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 ی 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 *Inferential Statistics*

آمار استنباطی مشخص می کند که آیا الگوها وفرآیندهای 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 رابطه‌ای (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. درحالی 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 :

- ✓ اندازه‌های گرایش به مرکز (میانگین، میانه، نما)
- ✓ اندازه‌های پراکندگی (دامنه تغییرات، واریانس، انحراف استاندارد)
- ✓ اندازه‌های وضعیت نسبی (رتبه درصدی، نمره انحراف متوسط)
- ✓ اندازه‌های رابطه‌ای (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 کلیپهای 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 ه ستند که قابل اندازه‌گیری یا شمارش ه ستند 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**)

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 ۲ 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 موافق۳ – بی نظر۲ – مخالف۱ 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

مشخص کننده مرکزی (شاخص ها)

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 ار این چندک کوچکترند

فراوانی

۱- **فراوانی مطلق** :به تعداد داده‌ها در هر طبقه فراوانی مطلق آن طبقه می‌گویند و آن را با **fi** نشان می‌دهند.

۲- **فراوانی نسبی** :در صورتی که فراوانی‌های مطلق را بر کل فراوانی‌ها تقسیم کنیم، فراوانی نسبی **ri** به دست می‌آید.

۳- **فراوانی تجمعی** :به مجموع فراوانی‌های مطلق طبقه‌های قبل و همان طبقه، فراوانی تجمعی آن طبقه می‌گویند و آن را با **Fi** نمایش می‌دهند.

۴- **فراوانی تجمعی نسبی** :می‌توان از تقسیم فراوانی های تجمعی بر تعداد داده‌ها، این فراوانی را به دست آورد**(Ri)** .

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
تجمیع F	۱	۴	۶	۸	۱۰

x=15 ⇨ داده متناظر با تجمیع مربوطه ⇨ **7 = 6+** در تجمیع ⇨ **$Q * \sum f_i$=(60/100)*10=6**

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

اگر همان داده را به صورت جدول فراوانی و تجمع 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

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

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

مشخصه های پراکندگی :

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)

۵- ضریب تغییرات

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=wi/n**

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

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

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

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

میانگین – اگراز 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

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

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

میانگین – اگراز هر داده فراوانی (به تعداد مشخصی) داشته باشیم

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

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

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

محاسبه واریانس برای نمونه و جمعیت تفاوت دارد

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

میانگین نمونه

$$Variance_{Sample} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

$$Variance_{Population} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$

$$\sigma^2 = \frac{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}{n}$$

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 آوربم. چون صورت و مخرج این کسر هم واحد هستند، حاصل کسر مقداری بدون واحد است که به صورت درصدی نیز می‌تواند 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 داده‌ها

مثال

قد تعداد زیادی اشخاص بشرح ذیل است انحراف معیار قد (جمعیت) 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

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

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

$$\sigma_x = \sigma_y = 2.13$$

فرمولهای فوق برای جمعیت است

برای نمونه در مخرج کسر بجای $\sum f_i$ جمله $(\sum f_i) - 1$ میگذاریم

مثال

نمرات **10** واحد درسی دانشجویی به شرح ذیل میباشد

نمره x =	۱۸	۱۲	۱۵	۱۶	۱۰
تعداد(واحد)=f	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	
Fi	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 ۵میشود) و در ستون **Fi** دنبال خود این جواب و یا کران بالای آن میگردیم که **Fi=5** نداریم و کران بالا عدد۶ میشود و داده متناظر آن نمره ۱۲ میانه میشود

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

$$\boldsymbol{5^+ \rightarrow \frac{\text{در}}{F} \rightarrow F = 6 \rightarrow \frac{\text{در}}{x} \rightarrow x = 12}$$

میانه ۱۲ میباشد

ج)

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

$$\boldsymbol{\rightarrow 8.5^+ \rightarrow \frac{\text{در}}{F} \rightarrow F = 9 \rightarrow \frac{\text{در}}{x} \rightarrow x = 16}$$

صدک ۸۵ عدد ۱۶ شد یعنی ۸۵٪ دادها ۱۶ یا کمتر از ۱۶ میباشند

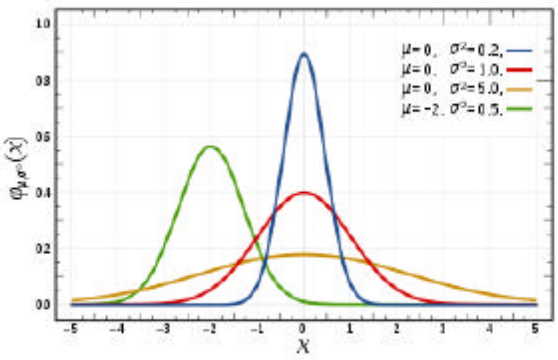
$$\boldsymbol{\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{(10*2)+(12*4)+(15*2)+(16*1)+(18*1)}{(2+4+2+1+1)} = 13.2}$$

$$\boldsymbol{\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}}$$

$$\boldsymbol{= \frac{(10-13.2)^2*2+(12-13.2)^2*4+(15-13.2)^2*2+(16-13.2)^2*1+(18-13.2)^2*1}{(2+4+2+1+1)} = 6.36}$$

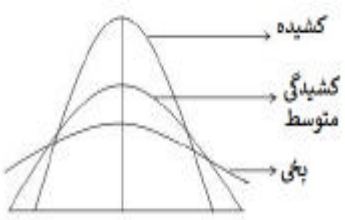
$$\boldsymbol{= \sigma^2 \qquad \sigma = 2.52}$$

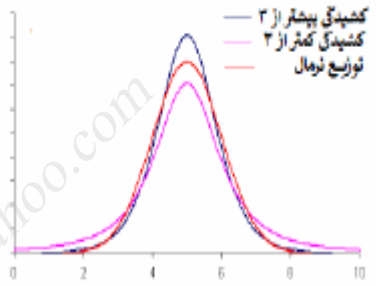
توزیع نرمال – نرمال استاندارد



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 برمبنای چارک‌ها

رگرسیون و همبستگی خطی

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

$$y = a + bx$$

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

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

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

x = سال	۹۳	۹۴	۹۵	۹۶	۹۷
y= مصرف	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

$$y = a + bx$$

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

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

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

$$y = 2.33 + 1.5x$$

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

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

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 **¼ ¼ ¼ ¼**

فضای مورد نظر **A= { HH,HT,TH }**

احتمال **P(A)=3/4**

X تعداد شیر	0	1	2
P=f(x) احتمال شیر	¼	2/4	1/4

۲) مثال

تاسی نااریب 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 2w w 2w w 2w

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

9w=1 w = 1/9

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

1/9 2/9 1/9 2/9 1/9 2/9

A = { 1 , 2 , 3 }

1/9 2/9 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 گوییم $f(x)$

توزیع احتمال نرمال X

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}(\frac{(x-\mu)}{\sigma})^2}$$

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

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 $\mu\pm2\sigma$ 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

$P(x\le b)$

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

حالا از روی جدول نرمال استاندارد احتمال حاصل میشود

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

طریقه استفاده از جدول نرمال استاندارد

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	0.0	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	0.0013	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

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

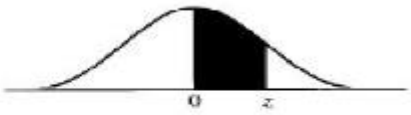
جدول زیر جدول توزیع احتمال نرمال استاندارد است

جدول توزیع نرمال استاندارد (جدول برای مقادیر کوچکتر یا مساوی)

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

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

Normal Curve Areas



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: This table is abridged from Table I of *Statistical Tables and Formulas*, by A. Hald (New York: John Wiley & Sons, Inc., 1952). Reproduced by permission of A. Hald and the publishers, John Wiley & Sons, Inc.

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 دقیقا 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

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

بدینترتیب میگوییم جمعیت به نرمال استاندارد تغییر یافت

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

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

ب) $1 - 0.1056 = 0.8944$

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

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

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

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

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

اگر تعداداین 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 دقیقا ۸۰ بار تیرهایش به هدف بخورد چقدر است

جدول بشرح ذیل در نظر بگیرید

$$p(z \le 2) = 0.98, \quad p(z \le 1.5) = 0.93, \quad p(z \le 1) = 0.85, p(z \le 0.5) = 0.7, p(z \le 0) = 0.5$$
$$p(z \le -2) = 0.02, p(z \le -1.5) = 0.07, \quad p(z \le -1) = 0.15, p(z \le -0.5) = 0.3$$

حل : پرتاب تیر تابع توزیع احتمال دوجمله 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 دوجمله ای میانگین برابر است با

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

در تابع توزیع دوجمله 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 میباشد پس میتوان گفت که که میتوان این توزیع دوجمله ای را از طریق جدول توزیع نرمال حل نمود

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

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

$$p(x = 80) = p(x \le 80) - p(x \le 79) = P\left(\frac{x-\mu}{\sigma} \le \frac{80-80}{4}\right) - P\left(\frac{x-\mu}{\sigma} \le \frac{79-80}{4}\right) = p(z \le 0) - p(z \le -0.25) = 0.5 - 0.4013$$
$$= 0.0987 \approx 0.1$$

** برآورد 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 میدهیم

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 **0.05** در نظر میگیریم)

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

$H_a : \mu > 1200$

$H_0 : \mu = 1200$

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

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

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

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

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

آزمون فرض‌ها : میخواهیم فرض های زیر را تحلیل کنیم

$H0: \mu1= \mu2= \mu3= \mu4$

$H1: \mu i \neq \mu j$

فرض **H0** چهار میانگین با هم مساوی هستند.

فرض **H1** 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

$\sigma^2(x) = E(X^2) - (E(X))^2$ واریانس

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

کوواریانس

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

$$Cov(x,y) = E(x.y) - E(x).E(y)$$

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

کوواریانس به شکل زیر محاسبه می‌شود.

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

محاسبه کوواریانس برای نمایش تغییرات بین دو متغییر 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(یا کورولیشن) برای تشخیص رابطه بین دو سری اعداد (متغیرها) معرفی میگردد

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

$$-1 \le \rho \le 1$$

اگر **x** و **y** رابطه خطی مستقیم داشته باشند **p=1** میباشد

اگر **x** و **y** رابطه معکوس خطی داشته باشند **p= -1** میباشد

اگر **x** و **y** رابطه نداشته باشند **p=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 محاسبه‌ی 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.

☜ ضریب همبستگی

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

۱) ضریب همبستگی پیرسون

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

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

مثال

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

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

حل: به روش پیرسون

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

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

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

*اعتیاد نمره	۲۲۰	221	210	234	240	180	196	200	**x*y**

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

کوواریانس

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

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

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

✓ ضریب بین ۰،۷۰ تا ۱ نشان دهنده همبستگی قوی

عدد ۰،۸۸ بیانگر همبستگی قوی بین **X** و **Y** است عددمنفی همبستگی در جهت معکوس است یعنی هر چه اینترنت بیشتر استفاده شود نمره دانشجو کمتر میشود Pearson Correlation = - 0.88

ضریب همبستگی رتبه ای اسپیرمن

این ضریب میزان همبستگی رابطه ی 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 رتبه های آنان استفاده می شود. رابطه ی مربوط به ضریب همبستگی رتبه ای 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 رتبه های اعضای متناظر دو گروه مورد بررسی و **n** حجم هر گروه.

داده های پارامتریک

داده‌های پارامتری به نمونه‌ای گفته می‌شود که از توزیع جامعه آماری آن مطلع هستیم

معمولا این توزیع آماری برای داده‌های کمی، نرمال یک یا چند متغیره در نظر گرفته می‌شود. در این حالت از آزمون‌های آماری پارامتری مثل آزمون **T**، آزمون **F** و یا آزمون **Z** استفاده می‌کنیم. همچنین برای اندازه‌گیری 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 تحلیل‌های ناپارامتری استفاده می‌شود. مثلا در متغیرهای رتبه ای ابتدا گروه بندی نموده و سپس برای تحلیل از ضریب همبستگی اسپیرمن استفاده میشود. 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 آنووا (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) «اســت. در این روش ســعی بر این اســت که اختلاف بین چند جامعه آماری، ارزیابی شــود. با توجه به پراکندگی کل داده‌ها، تجزیه واریانس بین گروه‌های مختلف در این روش امکان‌پذیر اســت. به این ترتیب می‌توان برابر بودن میانگین را بین گروه‌های مختلف آزمود. همچنین در مدل‌های رگرســیونی با تجزیه واریانس کل به واریانس مدل و واریانس خطا تشــخیص مناســب بودن مدل قابل ارزیابی است.

☜ آنالیز کوواریانس *ANCOVA*

آزمون تحلیل واریانس یا کواریانس 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 Analyze of Covariance یا آنکوا، نوعی آنالیز و تحلیل همانند آنوا می‌باشـــد. هرگاه در آنالیز واریانس بخواهیم اثر کمیت‌های مداخله گر را به روش‌های آماری حذف کنیم تا نتایج با دقت بیشـــتری به دســـت آید از آنالیز کوواریانس ا ستفاده می شود. در این روش هم از کنترل آماری و هم از واریانس ا ستفاده می شود. به‌عبارت دیگر به جای تحلیل واریانس تحلیل کوواریانس مورد استفاده قرار می‌گیرد.
در واقع آنکوا ANCOVA مدل تعمیم یافته آنوا ANOVA و همچنین مدل‌های رگرسیونی است. تحلیل کوواریانس مناسبترین آزمون آماری برای طرح پیش آزمون و پس آزمون ۲ گروهی می‌باشـد و تنها عامل تهدید کننده اعتبار درونی تحلیل کوواریانس طرح پیش آزمون است.

☜ آنالیز کوواریانس چندبعدی *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 کوواریانس‌ها Multivariable Analyze of Covariance MANCOVA یا مانکوا، روش تعمیم‌یافته آنکوا (آنالیز کوواریانس) ا ست و در مواردی کاربرد دارد که بیش از یک کمیت م ستقل وجود دا شته با شد و همچنین در مواردی که کمیت‌های وابســته به آســانی نتوانند ترکیب شـوند به کار می‌رود MANCOVA همان MANOVA با این تفاوت که به شـما 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 – نسبتی)	آنالیز واریانس
پیوسته(رتبه ای)(ترتیبی)	پیوسته(رتبه ای)(ترتیبی)	همبستگی

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

سطح سنجش متغیرها		آزمون های آماری
متغیر اول = متغیر مستقل **x**	متغیر دوم = متغیر وابسته **y**	
اسمی	اسمی	کای دو – فی – وی کرامر – لاندا
اسمی	ترتیبی	کای دو – فی – وی کرامر – لاندا
اسمی	فاصله ای یا نسبی	تحلیل واریانس یکطرفه – تی تست - لاندا
ترتیبی	اسمی	کای دو – فی – وی کرامر – لاندا – تتا
ترتیبی	ترتیبی	ضریب همبستگی رتبه ای اسپیرمن
ترتیبی	فاصله ای – نسبی	ضریب همبستگی اسپیر من – کندال
فاصله ای – نسبی	اسمی	کای دو – فی – وی کرامر – لاندا - تتا
فاصله ای – نسبی	فاصله ای – نسبی	ضریب همبستگی پیرسون **r**
فاصله ای – نسبی	ترتیبی	ضریب همبستگی رتبه ای اسپیرمن - کندال

انتخاب آزمون در انواع متغیرهای دو گروه و بیشتر

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

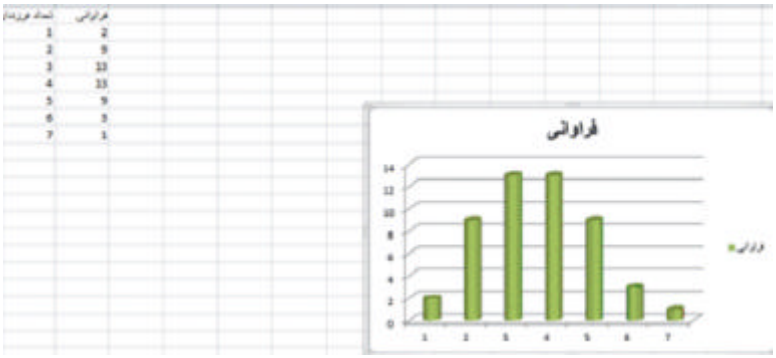
انتخاب آزمون برای بیش از چند متغیر

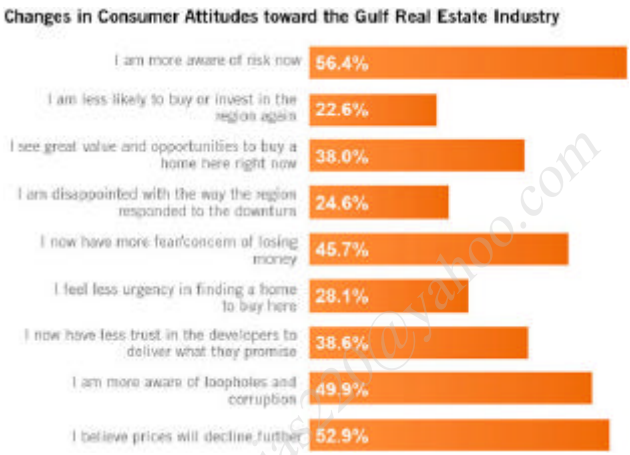
یک گروه	بیشتر از یک گروه
کوواریانس – واریانس-رگرسیون چندگانه – تحلیل عاملی	واریانس چند متغیره – تحلیل ممیزی

☜ انواع نمودارها

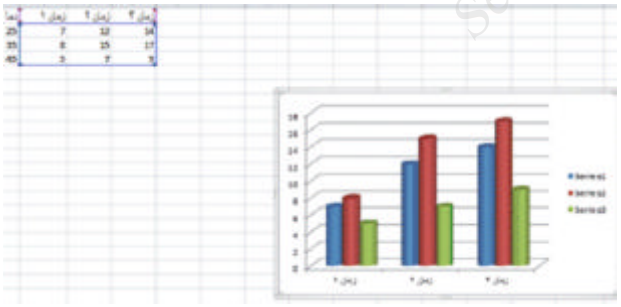
۱)) نمودار میله‌ای بارچارت **Bar Chart**:

برای مشاهده فراوانی داده ها و مقایسه داده ها نسبت به هم



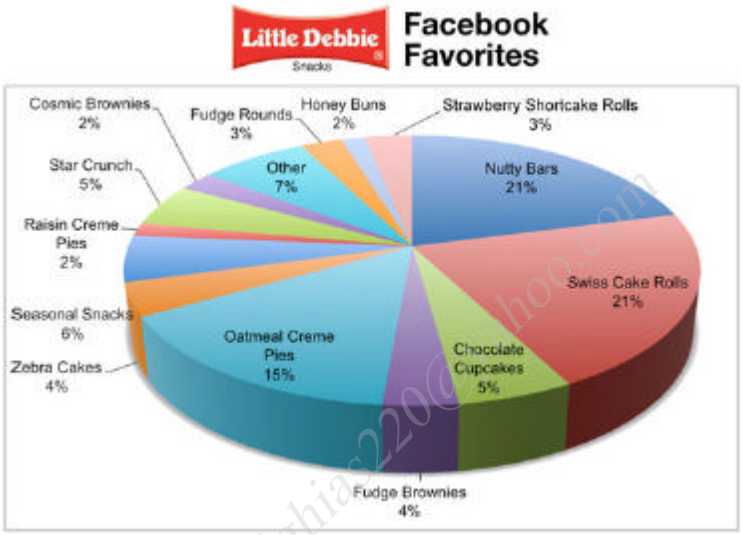


۲)) میله ای ستونی – مشاهده فراوانی و مقایسه داده ها بصورت دسته بندی شده



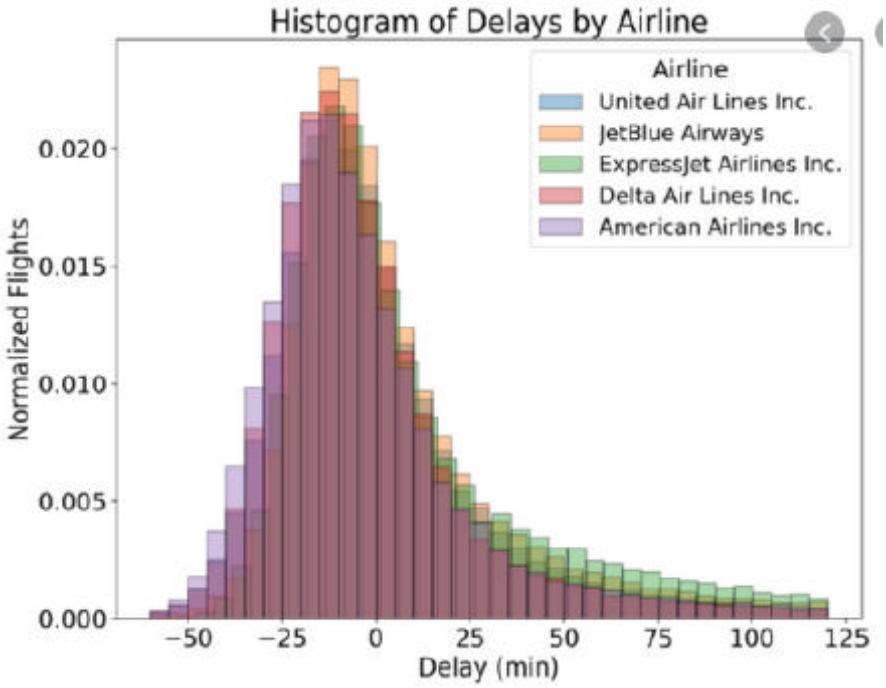
۳)) نمودار دایره ای**Pie Chart** :

برای مشاهده سهم هر مورد از داده ها

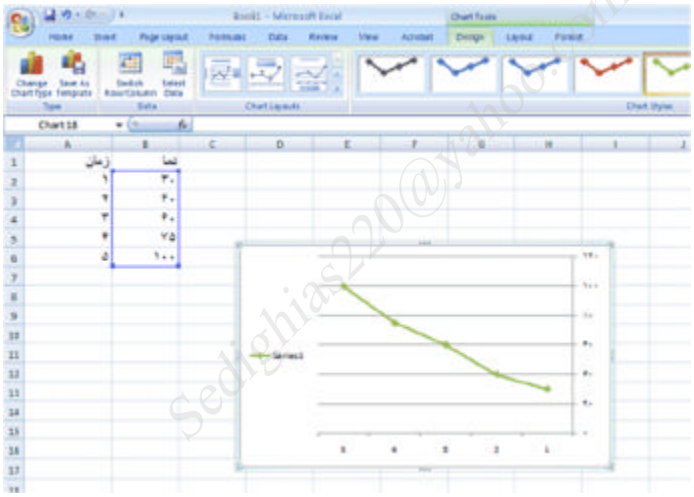


۴)) نمودار هیستوگرام **Histogram**:

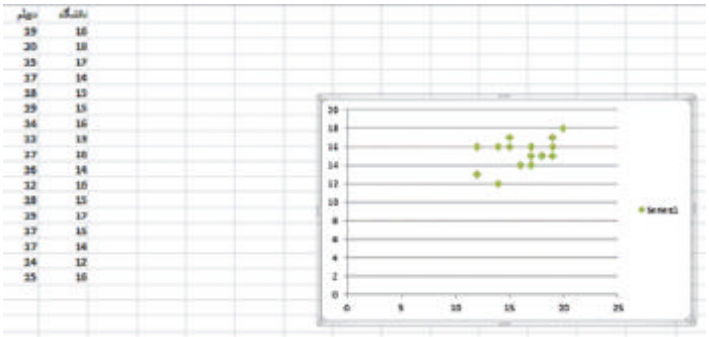
برای مشاهده نحوه توزیع داده ها بصورت پیوسته (مثلا معدل دیپلم و معدل دانشگاه)

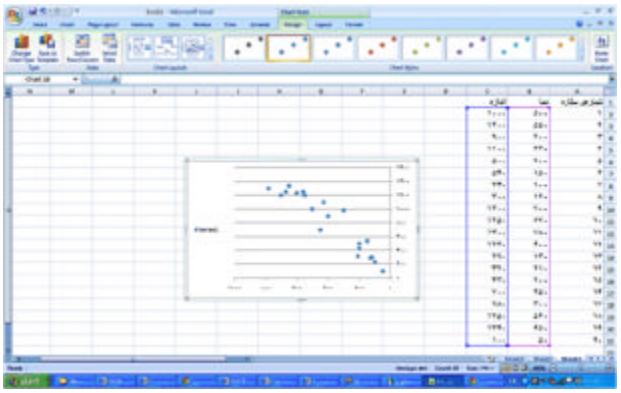


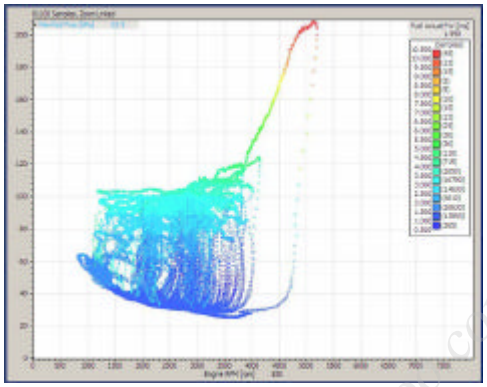
۵)) نمودار خطی **Line Chart**: برای نشان دادن رابطه بین دوسری اعداد مثلا خط رگرسیون



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







انواع متغیرها و شاخص‌ها

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

روش‌های مقیاس‌گذاری

یک نوع از مقیاس های غیرمقایسه‌ای 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 لیکر 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 ۳۰گویه و بیشتر تدوین می شود.

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

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 تلوزیونی)

یک محل برای درج شماره پرسشنامه

یک جمله خوش آمدگویی

پاراگراف حدودا سه خطی که مضمون خوش آمد گویی 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 لینکرت (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**
کد ردیف	شماره
خوش آمد گویی	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
سوالات کمکی: مشخصات پاسخ دهندگان	*برای پاسخ به سوالات زیر در مقابل آنها پاسخ بنویسید یا تیک بزنید* سن جنسیت مرد ○ زن ○ تحصیلات (زیر دیپلم و دیپلم ○ – دانشجو و فوق دیپلم ○ – لیسانس ○ – فوق لیسانس ○ – دکترا ○) آدرس ایمیل (اختیاری) : "

	برای پاسخ به سوالات زیر در جدول مقابل آنها تیک بزنید	کاملا موافق	موافق	نظری ندارم	مخالف	کاملا مخالف
سوالات اصلی تحقیق	**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					

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 استنباطی

نمونه گیری و خطا **Sampling Methods Errors**

روشهای نمونه گیری **Sampling Methods**

برآوردگرها **Estimators**

خطای معیار برآوردگرها **Standard Errors** – **SE**

ویژگی‌های برآوردگرها

برآوردگر نااریب **UnBiased**

قضیه حد مرکزی

توزیع تی **T-Student Distribution**

برآورد نقطه‌ای **Point Estimation** و فاصله‌ای **Internal Estimation**

برآورد واریانس و انحراف معیار

توزیع کای مربع = خی

برآوردگر باثبات (سازگار) **Consistent** (قانون قوی اعداد بزرگ)

برآوردگر کافی (بسنده) **Sufficent** (حداکثر درست‌نمایی **Maximum Liklihood**)

دقت برآوردگر (کارائی نسبی برآوردگر **Efficiency**)

آزمون فرض

آزمون فرض میانگین‌ها **Compare Means**

.....

.....

آیا اطلاعات بدست آمده با روشهای نمونه گیری از نمونه برای جمعیت عمومیت دارد

روشهای نمونه گیری و بررسی خطاها و تخمین های نقطه ای و فاصله ای پارامترها

نمونه گیری و خطا *Sampling Methods Errors*

روشهای نمونه گیری *Sampling Methods*

۱- روش نمونه گیری تصادفی ساده

قرعه کشی = اعطای شانس مساوی برای انتخاب

این روش برای جمعیت های همگن مفید است

مثلا با استفاده از جدول اعداد تصادفی

مثال از جدول که مثلا برای جمعیت ۳۰۰ نفره (به شماره از ۰۰۱ تا ۳۰۰) ، انتخاب نمونه ۵ نفره از داخل اعداد پنج رقمی تصادفی حاصل شده در نرم افزار اکسل

و انتخاب ۵ عدد سه رقمی از وسط جدول به سمت راست و انتخاب اعداد سه رقم وسطی با حذف سه رقمهای بیشتر از ۳۰۰ که نامربوط است و حذف اعداد سه رقمی که قبلا بوجود آمده)

62413 52318 47612 71221 41111 92310 40042 ....

بنابراین ۵ شماره بعنوان نمونه انتخاب شد = نمونه تصادفی ساده

عیب این روش اینکه چون داده ها مرتب نبود برای جمعیت های همگن مفید است و امکان اریبی (**Bias**) (= اشتباهات یک طرفه) وجود دارد.

مثلا اگر از جمعیت ۳۰۰ نفری ، که شامل ۲۰۰ خانم و ۱۰۰ آقا باشد و میخواهیم ۶ نفر انتخاب کنیم وقتی بتصادف ۶ نفر انتخاب کنیم احتمال دارد همه خانم یا همه آقا یا تعداد بیشتری آقا انتخاب شود در حقیقت عمل همگن سازی انجام نشده است

۲- روش نمونه گیری خوشه ای

جمعیت به خوشه هایی تقسیم میشود و کل خوشه در نمونه ظاهر میشود

مثلا نمونه هایی از گروههای دانشجویی رشت های مختلف

مثلا در یک دانشگاه ۴ رشته و ۱۴ کلاس داریم سه کلاس دانشجویی در رشته زیست شناسی و پنج کلاس در رشته میکرو بیولوژی و دو کلاس رشته ریاضی و چهار کلاس در رشته حقوق داریم و در هر کلاس تعدادی دانشجو داریم که تعداد میتواند با هم مساوی نباشد حال میخواهیم یک نمونه خوشه ای از دانشگاه انتخاب کنیم باید از سه کلاس رشته زیست شناسی یک کلاس کل دانشجویان بطور کامل انتخاب کنیم و از پنج کلاس رشته میکروبیولوژی کل دانشجویان یک کلاس و ... انتخاب کنیم حالا این نمونه خوشه ای است در حقیقت ۴ گروه دانشجویی کامل از بین ۱۴ کلاس انتخاب شده است

==========

۳- روش نمونه گیری سیستماتیک(سامان مند)

در این روش داده های جمعیت را بر حسب یک مورد مرتب میکنیم (مثلا مرتب شدن بر حسب جنسیت یا قد یا وزن یا سن یا ....)

مثلا ۳۰۰ نفر را به پنج گروه ۶۰ تایی تقسیم میکنیم و از هر گروه مثلا نفر چهاردهم(بتصادف) را انتخاب میکنیم ، پس اولین نفر، نفر ۱۴ و سپس نفر بعدی ۷۴ و نفر بعد تری ۱۳۴ و .....

این روش از همه طیف ها حضور دارند

ولی باز هم امکان اریبی (**Bias**) (= اشتباهات یک طرفه) وجود دارد ممکن است بعضی گروهها شانس بیشتری برای اظهار نظر داشته باشند (مثلا ممکن است از ۵ نفر انتخاب شده نفر چهارمی(از گروه چهارم) پرسشنامه ما را پاسخ ندهد پس گروههای دیگر شانس‌شان افزایش می‌یابد)

۴- روش نمونه گیری طبقه بندی **Classified**

برای مواردی که میخواهیم نمونه انتخابی، معرف آن جمعیت باشد به تناسب هر جمعیت تعداد نمونه انتخاب میشود

مثلا اگر از جمعیت ۳۰۰ نفری ، که شامل ۲۰۰ خانم و ۱۰۰ آقا باشد و میخواهیم ۶ نفر انتخاب کنبم از ۲۰۰ نفر خانم ۴ نفر را به تصادف انتخاب و از ۱۰۰ آقا ۲ نفر را به تصادف انتخاب میکینم در این روش با منفک کردن خانمها و آقایان، عمل همگن سازی انجام دادیم

۵- روش نمونه گیری با هدف خاص یا در دسترس یا ...

مثلا در یک محله ، عمر خودروهایی که افراد با جقوق متوسط در جامعه دارند مد نظر ما میباشد در این روش نمونه گیری به سراغ ثروتمندان و فقرا نرفتیم و فقط به سراغ افراد با درآمد متوسط رفتیم و عدد سن خودروهایشان برای کار آماری سوال نمودیم.

فرمولها

Statistics Formulas	
Mean	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Variance	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
Standard Deviation	$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$
Mean Absolute Deviation	$MAD = \frac{\sum \lvert x_i - \bar{x} \rvert}{n}$
Standard Error	$SE = \frac{s}{\sqrt{n}}$

برآوردگرها *Estimators*

تعریف پارامتر = شاخص های آماری مرتبط با ویژگیهای جمعیت باشد

تعریف آماره = شاخص های آماری مرتبط با ویژگیهای نمونه باشد

جدول شاخص‌ها

		میانگین	واریانس	انحراف معیار	نسبت	همبستگی	میانه
جمعیت	پارامتر	μ	σ^2	σ	p	ρ	η
نمونه	آماره	$\bar{x}$	S^2	SD	$\hat{p}$	$R\ ,r$	Md

برآوردگر : اگر از آماره برای تخمین پارامتر استفاده شود برآوردگر نامند

برآورد نقطه‌ای : اگر برای تخمین برآورد پارامتر فقط از برآوردگر نمونه استفاده شود آن را برآورد نقطه‌ای گویند

برآورد فاصله‌ای : با در دست داشتن خطای معیار

خطای معیار برآوردگرها *SE - Standard Errors*

۱- در آمار ممکن است شاخص های مرکزی داده‌ها (مکان) **Central Tendency** مد نظر باشد که توجه به مرکز داده ها دارد مثل (میانگین، میانه، مد)

۲- در آمار ممکن است شاخص های پراکندگی داده‌ها **Dispersion** مد نظر باشد که توجه به پراکنده بودن داده ها دارد مثل (مثل واریانس ، انحراف معیار ، ضریب تغییرات، دامنه)

۳- در آمار ممکن است شاخص تعداد **frequency** مد نظر باشد که توجه به مقدار داده ها دارد مثل (فراوانی و فراوانی تجمعی)

۴- در آمار ممکن است شاخص‌های نسبت **Proportion** و یا درصد داده‌ها **Percent** (چندک‌ها) مد نظر باشد که توجه به نسبت داده ها دارد مثل (فراوانی نسبی و فراوانی نسبی تجمعی و چارک و دهک و صدک)

در هر یک از چهار مورد فوق اگر از طریق نمونه شاخص‌ها را بدست آوریم تقریب زدن آن به جمعیت خطا حاصل میکند که به آن خطای معیار گویند

خطای معیار : برای تخمین پارامترجمعیت از آماره نمونه استفاده میکنیم که چون داریم از شاخص نمونه برای شاخص جمعیت تخمین میزنیم پس خطا دارد این را خطای معیار گویند

اگر از جمعیت **N** تایی داده ها یک نمونه **n** تایی را بصورت زیر در نظر بگیریم

$X_1 , X_2 , X_3 , X_4 , ...$

و میانگین نمونه داده ها را $\bar{X}$ بنامیم و انحراف معیار نمونه **Standard Deviation** را **SD** بنامیم آنگاه خطای معیار میانگین

$$SE(\bar{X}) = \frac{SD}{\sqrt{n}}$$

و میانه نمونه داده ها را Md بنامیم و اگر آنگاه خطای معیار میانه

$$SE(Md) = 1.253\frac{SD}{\sqrt{n}} = 1.253 * SE(\bar{X})$$

با همین فرمول متوجه میشویم که خطای معیار میانه از خطای معیار میانگین ۲۵٪ بیشتر است یعنی خطا در میانه بیشتر از خطا میانگین است مگر اینکه تعداد نمونه خیلی زیاد شود

مثال : از نمرات ریاضی یک کلاس ۵۰ نفری یک نمونه ۷ تایی انتخاب میکنیم خطای معیار میانگین و خطای معیار میانه را بدست آورید و مقایسه کنید (نمرات ۱۴ و ۱۲ و ۱۷و ۶ و ۱۷ و ۱۳ و ۱۵)

X_i	6	12	13	14	15	17
f_i	1	1	1	1	1	2

میانگین نمونه

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i} = \frac{(6*1)+(12*1)+(13*1)+(14*1)+(15*1)+(17*1)}{1+1+1+1+1+2} = 13.428$$ میانگین نمونه

واریانس نمونه از یکی از دو فرمول زیر میتوان بدست آورد

$$S^2 = \frac{n\sum x^2-(\sum x)^2}{n(n-1)} = \frac{7(6^2+12^2+13^2+14^2+15^2+17^2+17^2)-(6+12+13+14+15+17+17)^2}{7(7-1)} = 14.29$$

یا به این طریق

$$S^2 = \frac{\sum(x_i-\bar{x})^2 f_i}{(\sum f_i)-1}$$ واریانس نمونه

$$= \frac{(6-13.428)^2*1+(12-13.428)^2*1+(13-13.428)^2*1+(14-13.428)^2*1+(15-13.428)^2*1+(17-13.428)^2*2}{1+1+1+1+1+2-1}$$

$$= 14.29$$

حال انحراف معیار نمونه بدست میاوریم همان **SD** که به **S** نوشتیم

$$S = \sqrt{14.29} = 3.78$$ انحراف معیار نمونه

حال میخواهیم از این نمونه اگر برای جمعیت استفاده کنیم مقدار خطای میانگین چقدر است

$$SE(\bar{X}) = \frac{SD}{\sqrt{n}} = \frac{3.78}{\sqrt{7}} = 1.42$$

و اگر بخواهیم از این نمونه اگر برای جمعیت استفاده کنیم مقدار خطای میانه چقدر است

$$SE(Md) = 1.253\frac{SD}{\sqrt{n}} = 1.253 * SE(\bar{X}) = 1.253 * 1.42 = 1.78$$

اگر در یک آماره نمونه از نسبت ها استفاده کنیم مثلا هرگاه از نمونه **n** تایی $\hat{p}$ نسبت مربوطه به نمونه و p نسبت مربوط به جمعیت باشد اگر $\hat{p}$ درصد مورد نظر (موفقیت) باشد آن آنگاه درصد خلاف آن (شکست) $\hat{q} = 1 - \hat{p}$ خواهد بود و آنگاه مقدار خطا برابر است با

در صد خطا معیار نسبت نمونه $SE(\hat{p}\%) = 100 * \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$ خطای معیار نسبت نمونه $SE(\hat{p}) = \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$

حال اگر بخواهیم درصد این خطا برای فراوانی نمونه f را برای فراوانی تعداد کل جمعیت محاسبه کنیم

$$SE(f) = N * SE(\hat{p}) = N * \sqrt{\frac{\hat{p}\,\hat{q}}{n}}$$ خطای معیار فراوانی

مثال : بتعداد کل ۳۰۰۰ نفر به بیمارستان برای بیماری کرونا مراجعه کردند از بین آنها یک نمونه ۵۰ نفری انتخاب شد و مشخص شد تعداد ۴ نفر از این ۵۰ نفر بیماری کرونا دارند – خطای معیار نسبت و خطای معیار درصد و خطای معیار فراوانی را بدست آورید

$$\hat{p} = \frac{4}{50} = 0.08 \qquad \hat{q} = 1 - \hat{p} = 1 - 0.08 = 0.92$$

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\ \hat{q}}{n}} = \sqrt{\frac{0.08 * 0.92}{50}} = 0.038$$ خطا معیار نمونه

$$SE(\hat{p}\%) = 100 * \sqrt{\frac{0.08 * 0.92}{50}} = \%3.8$$ در صد خطای معیار نمونه

$$SE(f) = N * SE(\hat{p}) = N * \sqrt{\frac{\hat{p}\ \hat{q}}{n}} = 3000 * 0.038 = 114$$

هر گونه تقریب برای ۳۰۰۰ نفر به میزان ۱۱۴ نفر احتمال دارد خطا داشته باشد

مثلا میانگین نمونه بیماران در ۵۰ نفر

$$\bar{x} = n * p * q = 50 * 0.08 * 0.92 = 3.68$$

مثلا میانگین جمعیت ۳۰۰۰ نفر (با تقریب به میانگین نمونه)

$$\bar{x} = n * p * q = 3000 * 0.08 * 0.92 = 220$$

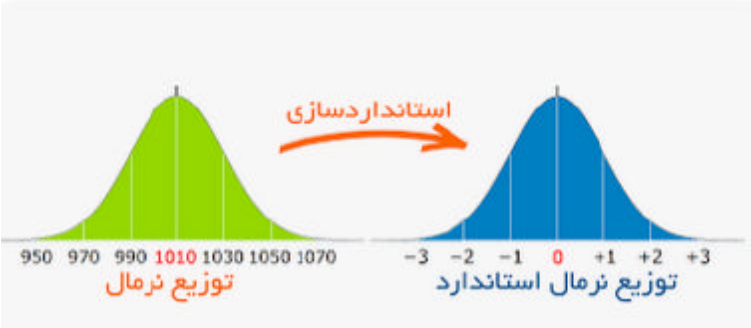
میانگین جمعیت بیماران در بین ۳۰۰۰ نفر 220 ± 114 میباشد یعنی در ۳۰۰۰ نفر بین ۶ نفر تا ۳۳۴ نفر بیمار هستند

اگر بخواهیم میزان خطا کمتر شود باید تعداد نمونه **n** را افزایش دهیم

در خصوص نمونه **n** تایی بیم دو سری عدد و ضریب همبستگی پیرسون **R** بین آن دو سری عدد (که **R** بین -۱ و ۱ است) ، آنگاه خطای معیار برآورد پیرسون برابر است با

$$SE(R) = \frac{1 - R^2}{\sqrt{n - 1}}$$

تمرین ها



ویژگی‌های برآوردگرها

در توزیع‌های متقارن میانگین ، میانه ، مود بر هم منطبق هستند

ابتدا توزیع نرمال تشریح میدهد

برآوردگر ناریب UnBiased

هرگاه امید ریاضی (مقدار انتظاری) مقدار برآوردگر به پارامتر جمعیت نزدیک باشد آنگاه میگوییم برآوردگر ناریب میباشد

قضیه حد مرکزی

اگر تعداد نمونه بیشتر از ۳۰ باشد آنگاه توزع نرمال است میگوییم توزیع نرمال است

اگر میانگین نمونه $\boldsymbol{\theta}$ باشد و $\widehat{\boldsymbol{\theta}}$ میانگین نمونه و توزیع نرمال باشد و خطای معیار نمونه $\boldsymbol{SE(\widehat{\theta})}$ باشد

اگر میانگین نمونه $\boldsymbol{\mu}$ باشد و $\bar{x}$ میانگین نمونه باشد و توزیع نرمال باشد و خطای معیار نمونه $\boldsymbol{SE(\bar{x})}$ باشد

$$\boldsymbol{z=\frac{\widehat{\theta}-\theta}{SE(\widehat{\theta})}} \qquad \boldsymbol{z=\frac{\bar{x}-\mu}{SE(\bar{x})}}$$

قبلا در مورد توزیع نرمال اشاره شد که :

مثال : در کل یک کلاس میانگین ۱۴ و واریانس ۳.۶ میباشد احتمال اینکه نمره یک دانش آموز در این کلاس بیشتز از ۱۷ باشد چقدر است؟ بیشترین نمره برای $X_{0.5}$ بدست آورید

$$X\sim N(\mu=14,\sigma^2=3.6) \qquad \sigma=\sqrt{3.6}=1.89$$

$$\boldsymbol{p(x>17)=1-p(x\le 17)=1-p\left(\frac{x-\mu}{\sigma}\le\frac{17-\mu}{\sigma}\right)=1-p\left(z\le\frac{17-14}{1.89}\right)=1-p(z\le+1.58)}$$

$$\boldsymbol{=1-0.9429=0.057}$$

به احتمال ۵,۷٪ نمره یک دانشجو بیشتر از ۱۷ میشود

$$\boldsymbol{p(x>k)=0.05}$$

$$\boldsymbol{p(x>k)+p(x\le k)=1} \qquad \boldsymbol{p(x>k)=1-p(x\le k)}$$

$$\boldsymbol{0.05=1-p\left(\frac{x-\mu}{\sigma}\le\frac{k-14}{1.89}\right)} \qquad \boldsymbol{0.05=1-p\left(z\le\frac{k-14}{1.89}\right)} \qquad \boldsymbol{p\left(z\le\frac{k-14}{1.89}\right)=0.95}$$

$$\boldsymbol{p\left(z\le\frac{k-14}{1.89}\right)=p(z\le 1.645)}$$

$$\boldsymbol{\frac{k-14}{1.89}=1.645} \qquad \boldsymbol{k-14=1.89*1.645} \qquad \boldsymbol{k=17}$$

یعنی به احتمال ۹۵٪ نمره دانشجویان زیر ۱۷ خواهد بود یا ۵٪ احتمال دارد نمره دانشجویی بیشتر از ۱۷ شود به عبارتی بیشترین نمره ۱۷ خواهد بود

مثال : در یک کلاس میانگین ۱۴ و واریانس ۳,۶ میباشد (انحراف معیار ۱,۸۹) میباشد احتمال اینکه نمره یک دانش آموز در یک کلاس ۲۵ نفری از ۱۷ بیشتر باشد

$$z = \frac{\bar{x} - \mu}{SE(\bar{x})} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} =$$

$$p(x > 17) = 1 - p(x \le 17) = 1 - p\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \le \frac{17 - \mu}{\sigma/\sqrt{n}}\right) = 1 - p\left(z \le \frac{17 - 14}{1.89/\sqrt{25}}\right)$$
$$= 1 - p(z \le +7.936) = 1 - 0.9999 = 0.0001$$

به احتمال خیلی ضعیف یعنی ۰,۰۱٪ در این کلاس نمره دانشجو بیشتر از ۱۷ میشود

توزیع تی T-Student Distribution

در قضیه حد مرکزی توزیع نرمال استاندارد اشاره شد که $z = \frac{\bar{x} - \mu}{SE(\bar{x})} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ میباشد

در خصوص توزیع **T** میتوان نوشت

$$T = \frac{\bar{x} - \mu}{SE(\bar{x})} = \frac{\bar{x} - \mu}{SD/\sqrt{n}}$$

در توزیع **T** اگر حجم نمونه(**n**) زیاد باشد با توزیع نرمال استاندارد **N** مساوی است

دو نمودار نرمال و تی شبیه بهم و هر دو متقارن است اما دامنه توزیع نرمال از -۴ تا +۴ است ولی دامنه **T** بسیار بزرگ است

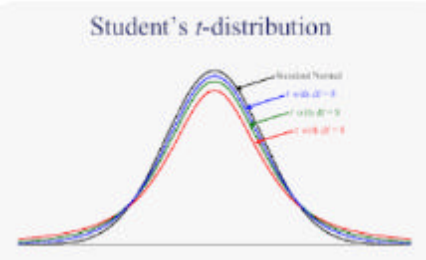
در **T** به شاخصی نیاز داریم بنام درجه آزادی **Df = Degree Of Freedom**

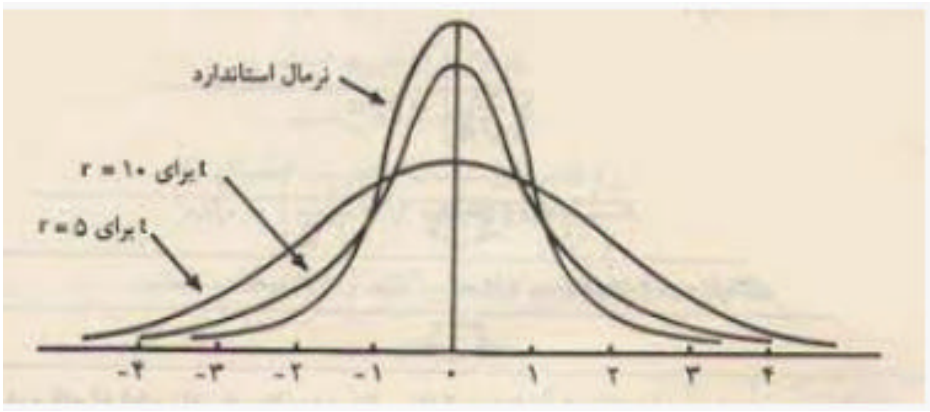
$$df = n - 1$$

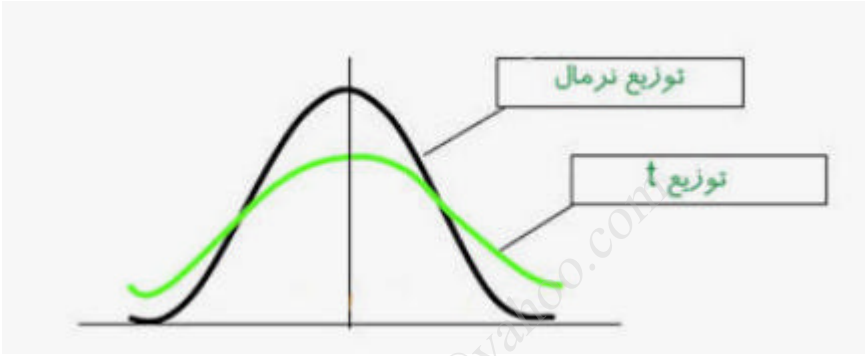
در خصوص میانگین و واریانس قبلا اشاره شد که در نمونه های با حجم کم مخرج کسر واریانس **n-1** میباشد

میانگین $\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i}{n}$ واریانس $S^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{(\sum f_i) - 1} = \frac{\sum (x_i - \bar{x})^2}{n - 1}$

انحراف معیار $SD = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$

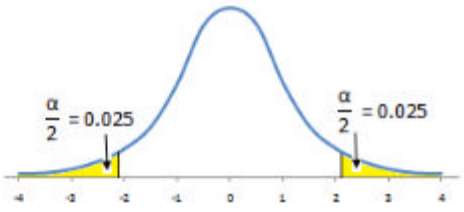






Student's t Distribution Table

For example, the t value for 18 degrees of freedom is 2.101 for 95% confidence interval (2-Tail $\alpha = 0.05$).



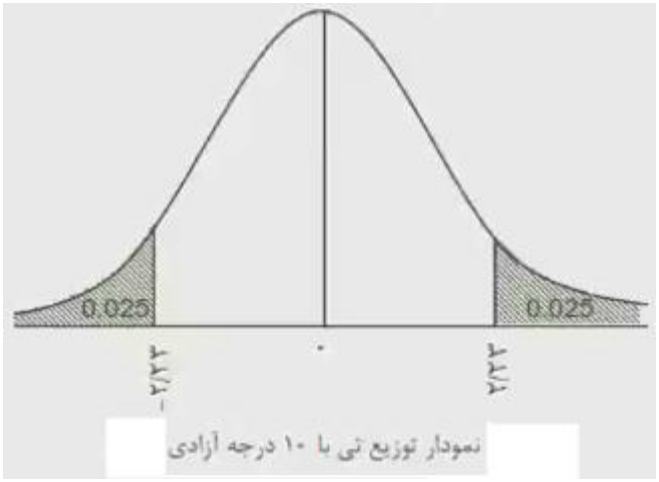
	90%	95%	97.5%	99%	99.5%	99.95%	1-Tail Confidence Level
	80%	90%	95%	98%	99%	99.9%	2-Tail Confidence Level
	0.100	0.050	0.025	0.010	0.005	0.0005	1-Tail Alpha
df	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha
1	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192	
2	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991	
3	1.6377	2.3534	3.1824	4.5407	5.8409	12.9240	
4	1.5332	2.1318	2.7764	3.7469	4.6041	8.6103	
5	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688	
6	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588	
7	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079	
8	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413	
9	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809	
10	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869	
11	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370	
12	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178	
13	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208	
14	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405	
15	1.3406	1.7531	2.1314	2.6025	2.9467	4.0728	
16	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150	
17	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651	
18	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216	
19	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834	
20	1.3253	1.7247	2.0860	2.5280	2.8453	3.8495	
21	1.3232	1.7207	2.0796	2.5176	2.8314	3.8193	
22	1.3212	1.7171	2.0739	2.5083	2.8188	3.7921	
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676	
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.7454	
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.7251	
26	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066	
27	1.3137	1.7033	2.0518	2.4727	2.7707	3.6896	
28	1.3125	1.7011	2.0484	2.4671	2.7633	3.6739	
29	1.3114	1.6991	2.0452	2.4620	2.7564	3.6594	
30	1.3104	1.6973	2.0423	2.4573	2.7500	3.6460	

در این جدول هم نشان میدهد کل مساحت زیر منحنی **1** میباشد و نصف منحنی **0.5** میباشد

که در توزیع **T** در هر سمت **0.025** و در دو سمت با هم **0.05** مساحت زیر منحنی را نشان میدهد

مثلا توزیع **T** با **10** درجه آزادی و سطح اطمینان ۹۷,۵٪ به مقدار **2.228** میباشد

==========



برآورد نقطه‌ای Point Estimation و فاصله‌ای Internal Estimation

آمار توصیفی جمع آوری و سازمان دهی داده ها و بدست آوردن شاخص های آماری نمونه (یا جمعیت) داده ها میباشد

آمار استباطی تحلیل و تخمین از شاخص های نمونه برای جمعیت است

برای تخمین از برآوردگرها استفاده میشود

برآورد نقطه ای : ارائه دادن گزارش دادن شاخص نمونه ، بجای جمعیت (که طبیعتا دقت کمی دارد)

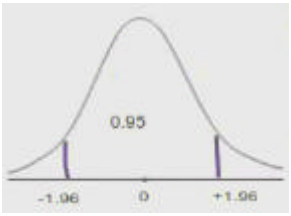
برآورد فاصله‌ای : ارائه دادن گزارش دادن شاخص نمونه ، بجای جمعیت با ارائه ضریب اطمینان

سطح اطمینان را با $\%(\quad -\mathbf{1})$ نشان میدهند

قبلا در توزیع نرمال اشاره شد که سطح زیر منحنی برای ۹۵٪ بین $\boldsymbol{z = \pm 1.96}$ میباشد

$$\boldsymbol{z = \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \qquad -1.96 < \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} < +1.96}$$

$$\boldsymbol{L \;\; w = \hat{\theta} - 1.96 * SE(\hat{\theta}) \qquad U \;\; = \hat{\theta} + 1.96 * SE(\hat{\theta})}$$



یعنی با اطمینان $\boldsymbol{95\% = \%(\quad -1)}$ یعنی ۹۵٪ میتوان گفت که **z** بین ۱,۹۶- تا ۱,۹۶ + میباشد

برای داده با تعداد زیاد در جدول **T** هم همین بدست میاید

یعنی برای تعداد خیلی زیاد، نقطه بحرانی دو دامنه توزیع نرمال استاندارد در سطح ۵٪ از جدول توزیع نرمال با جدول توزیع **T** با هم مساوی میباشد

مثال: از نمونه ۲۰۰ نفری دانشجویان میانگین نمرات ۱۳ و انحراف معیار ۱,۵ و میانه نمرات ۱۲ میباشد فاصله اطمینان ۹۵٪ برای میانگین و میانه را محاسبه کنید و توضیح بنویسید

با توجه به صورت مسئله آماره ها $\boldsymbol{\bar{x} = 13}$ و $\boldsymbol{SD = 1.5}$ و **Md=12** میباشد

برآورد نقطه ای میانگین جمعیت **13** و برآورد نقطه ای میانه جمعیت هم **12** میگردد اما این برآورد از نمونه به جمعیت مقداری خطا دارد

اگر از میانگین نمونه بعنوان میانگین جمعیت $\hat{\mu} = \bar{x} = 13$ استفاده و برآورد نقطه ای را حاصل کنیم میزان خطای معیار میانگین برابر است با

$$SE(\bar{x}) = \frac{SD}{\sqrt{n}} = \frac{1.5}{\sqrt{200}} = 0.11$$

اگر از میانه نمونه بعنوان میانه جمعیت $\hat{\eta} = \bar{x} = 12$ استفاده و برآورد نقطه‌ای شود میزان خطای معیار میانه برابر است با

$$SE(Md) = 1.253 * \frac{SD}{\sqrt{n}} = 1.253 * \frac{1.5}{\sqrt{200}} = 0.13$$

حال چون تعداد نمونه **n=200** عدد بزرگی میباشد میتوان از نرمال استاندارد برای فاصله اطمینان ۹۵٪ استفاده نمود

طبق جدول نرمال استاندارد برای فاصله اطمینان ۹۵٪ از دو طرف مقدار **Z= -1.96** تا **Z= +1.96** میگردد

$$z = \frac{\bar{x} - \mu}{SE(\bar{x})} \qquad -1.96 < \frac{\bar{x} - \mu}{SE(\bar{x})} < +1.96$$

$$L\ \ w = \hat{\mu} - 1.96 * SE(\hat{\mu}) = \bar{x} - 1.96 * SE(\bar{x}) = 13 - 1.96 * 0.11 = 12.78$$
$$U\ \ = \hat{\mu} + 1.96 * SE(\hat{\mu}) = \bar{x} + 1.96 * SE(\bar{x}) = 13 + 1.96 * 0.11 = 13.22$$

یعنی با اطمینان ۹۵٪ میتوان گفت میانگین جمعیت بین ۱۲,۷۸ تا ۱۳,۲۲ میباشد

و برای میانه میتوان نوشت

$$L\ \ w = \hat{\eta} - 1.96 * SE(\hat{\eta}) = Md - 1.96 * SE(Md) = 12 - 1.96 * 0.13 = 11.75$$
$$U\ \ = \hat{\eta} + 1.96 * SE(\hat{\eta}) = Md + 1.96 * SE(Md) = 12 + 1.96 * 0.13 = 12.25$$

یعنی با اطمینان ۹۵٪ میتوان گفت میانه جمعیت بین ۱۱,۷۵ تا ۱۲,۲۵ میباشد

اگر تعداد نمونه **n** کوچک بلشد باید از **T** استیودنت با درجه آزادی $df = n - 1$ و فاصله اطمینان دو دامنه **0.95** از جدول **T** استفاده کنیم و از این دو فرمول برای فاصله اطمینان استفاده میکینم

$$L\ \ w = \hat{\theta} - \ * SE(\hat{\theta})$$
$$U\ \ = \hat{\theta} + t * SE(\hat{\theta})$$

مثال: از نمونه ۲۰ نفری دانشجویان میانگین نمرات ۱۳ و انحراف معیار ۱,۵ میباشد فاصله اطمینان ۹۵٪ برای میانگین را محاسبه کنید و توضیح بنویسید

با توجه به صورت مسئله آماره ها $\bar{x} = 13$ و $SD = 1.5$ میباشد برآورد میانگین جمعیت **13** میگردد اما این برآورد از نمونه به جمعیت مقداری خطا دارد

اگر از میانگین نمونه بعنوان میانگین جمعیت $\hat{\theta} = \bar{x} = 13$ استفاده و برآورد را حاصل کنیم میزان خطای معیار میانگین برابر است با

$$SE(\hat{\theta}) = \frac{SD}{\sqrt{n}} = \frac{1.5}{\sqrt{20}} = 0.33$$

چون **n=20**کوچک میباشد از **T** با درجه آزادی **df=n-1=20-1=19** استفاده میکنیم که برای فاصله اطمینان دو طرفه **0.95** از **-2.09** تا **+2.09** میباشد

$$t = \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} \qquad -2.09 < \frac{\hat{\theta} - \theta}{SE(\hat{\theta})} < +2.09$$

$$L\ \ w = \hat{\theta} - 2.09 * SE(\hat{\theta}) = 13 - (2.09 * 0.33) = 12.31$$
$$U\ \ = \hat{\theta} + 2.09 * SE(\hat{\theta}) = 13 + (2.09 * 0.33) = 13.69$$

یعنی با اطمینان ۹۵٪ میتوان گفت میانگین جمعیت بین ۱۲,۳۱ تا ۱۳,۶۹ میباشد

از جدول **T** با درجه آزادی ۱۹ برای فاصله اطمینان دو طرفه **0.99** از **-2.86** تا **+2.86** میباشد

$$L\ \ w = \hat{\theta} - 2.86 * SE(\hat{\theta}) = 13 - (2.86 * 0.33) = 12.05$$
$$U\ \ = \hat{\theta} + 2.86 * SE(\hat{\theta}) = 13 + (2.86 * 0.33) = 13.94$$

یعنی با اطمینان ۹۹٪ میتوان گفت میانگین جمعیت بین ۱۲,۰۵ تا ۱۳,۹۴ میباشد

مثال : در یک نمونه $n=30$ تایی دانشجویان ضریب همبستگی بین نمرات درس فیزیک و ریاضی $R=0.75$ میباشد یک فاصله اطمینان ۹۹٪ برای ضریب همبستگی بدست آورید

$$SE(R) = \frac{1 - R^2}{\sqrt{n-1}} = \frac{1 - 0.75^2}{\sqrt{30-1}} = \frac{0.437}{5.385} = 0.081$$

از جدول **T** با درجه آزادی **df=n-1=30-1=29** برای فاصله اطمینان دو طرفه **0.99** از **-2.756** تا **+2.756** میباشد

$$L\ \ w = \hat{\rho} - 2.756 * SE(\hat{\rho}) = 0.75 - (2.756 * 0.081) = 0.526$$
$$U\ \ = \hat{\rho} + 2.256 * SE(\hat{\rho}) = 0.75 + (2.756 * 0.081) = 0.973$$

بنابراین با اطمینان ۹۹٪ میتوان گفت که ضریب همبستگی بین نمرات درس فیزیک و ریاضی بین ۰,۵۲ تا ۰,۹۷ میباشد

یعنی نمره یک دانشجو برای نمره درس قیزیک به نمره درس ریاضی هم ربط دارد

برآورد واریانس و انحراف معیار

قبلا فرمول واریانس جمعیت و نمونه ذکر شد

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

$$S^2 = \frac{\sum(x_i - \bar{x})^2 f_i}{(\sum f_i) - 1} = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$$

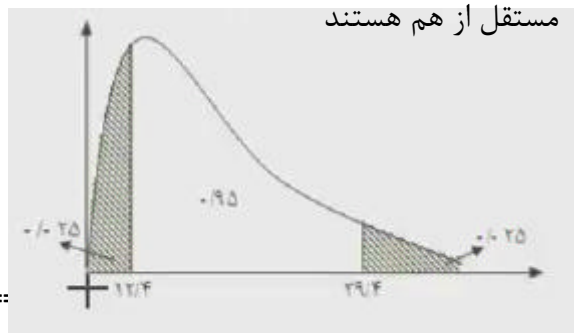
توزیع کای مربع = خی

در یک جمعیت با واریانس σ^2 ، اگر یک نمونه **n** تایی با واریانس S^2 داشته باشیم آنگاه کسر زیر دارای توزیع خاصی میباشد

$$\frac{(n-1)S^2}{\sigma^2}$$

و میتوان نوشت

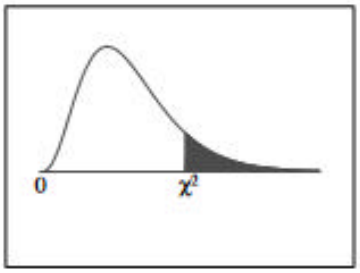
که با درجه آزادی **v = n-1** مینامند و دارای جدول خاصی میباشد نمودار این متقارن نیست و میانگین و واریانس نمونه مستقل از هم هستند



$$\chi_a^2(\nu) < \frac{(n-1)S^2}{\sigma^2} < \chi_b^2(\nu) \qquad \frac{(n-1)S^2}{\chi_b^2(\nu)} < \sigma^2 < \frac{(n-1)S^2}{\chi_a^2(\nu)}$$

بنابراین واریانس با فاصله اطمینان مورد نظر و درجه آزادی **n-1** و از جدول مربوطه ، بین دو مقدار فوق میباشد

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

جدول کای مربع (خی۲)

مثال : در یک کلاس با ۲۵ دانشجو انحراف معیار نمرات ۱,۷۵ گردید اگر توزیع نرمال باشد ضریب اطمینان ۹۵٪ برای انحراف معیار جمعیت را بدست آورید

n = 25 , SD = 1.75

$\hat{\sigma} = SD = 1.75 \qquad \sigma^2 = S^2 = 1.75^2 = 3.06 \qquad \vartheta = n - 1 = 25 - 1 = 24$

از جدول کای مربع (خی) فاصله اطمینان ۹۵٪ (۰,۹۵ یعنی دو تا دنباله ۰,۰۲۵ در چپ و ۰,۰۲۵ در راست) و با ۲۴ درجه آزادی ، برای واریانس دو مقدار زیر حاصل میشود

$\chi^2_{0.025}(24) = 39,4 \qquad \chi^2_{0.975}(24) = 12.4.$

$$L\ w = \frac{(n-1)S^2}{\chi^2_a(v)} = \frac{(25-1)*3.06}{39.4} = 1.86$$

$$U\ = \frac{(n-1)S^2}{\chi^2_b(v)} = \frac{(25-1)*3.06}{12.4} = 5.92$$

از این دو باید جذر گرفت تا حد بالا و پایین برای انحراف معیار حاصل شود

$$L\ w = \sqrt{1.86} = 1.36$$

$$U\ = \sqrt{5.92} = 2.43$$

پس با اطمینان ۹۵٪ میتوان گفت انحراف معیار جمعیت بین ۱,۳۶ تا ۲,۴۳ خواهد بود

برآوردگر باثبات (سازگار) Consistent (قانون قوی اعداد بزرگ)

اگر با افزایش دادن تعداد نمونه آماره نمونه به پارامتر جمعیت نزدیک شویم برآوردگر باثبات گویند

برآوردگر کافی (بسنده) Sufficent (حداکثر درست‌نمایی Maximum Liklihood)

از تمام اطلاعات موجود در نمونه استفاده کنیم

دقت برآوردگر (کارائی نسبی برآوردگر Efficiency)

کارائی برآوردگر **T1** به **T2**

$$Ef\ (T_1 \rightarrow T_2) = (\frac{SE(T_2)}{SE(T_1)})^2$$

کارائی نسبی میانه نسبت به میانگین بدست آورید

$$Ef\ (Md \rightarrow \bar{x}) = \left(\frac{SE(\bar{x})}{SE(Md)}\right)^2 = \left(\frac{\frac{\sigma}{\sqrt{n}}}{1.253\frac{\sigma}{\sqrt{n}}}\right)^2 = 0.637 \cong .64$$

پس میانه کارائی کمتری نسبت به میانگین دارد

در یک نمونه تحقیقاتی ۱۰۰ نفره اگر بخواهیم از شاخص میانه بجای میانگین استفاده کنیم حجم نمونه چقدر باید باشد تا میانه کارائی ، میانگین را داشته باشد.

$$Ef\ (Md \rightarrow \bar{x}) = \left(\frac{SE(\bar{x})}{SE(Md)}\right)^2 \qquad 1 = \left(\frac{\frac{\sigma}{\sqrt{100}}}{1.253\frac{\sigma}{\sqrt{n}}}\right)^2 \qquad 1 = \frac{n}{157} \qquad n = 157$$

اگر حجم نمونه ۱۵۷ شود و میانه را بدست آوریم از این عدد میانه میتوان بجای میانگین استفاده کرد.

بقیه جزوه ادامه دارد

آزمون فرض

آزمون فرض میانگین‌ها *Compare Means*

1- **One Sample**
2- **Two Independent Samples**
3- **Two Paired Samples**

میانگین ، میانه و مود سه شاخص مرکزی (مکان مرکزی) می‌باشند

شاخص میانگین، شاخص مناسب تری برای مرکز داده ها میباشد

اگر جمعیتی دارای میانگین $\boldsymbol{\mu}$ و واریانس $\boldsymbol{\sigma^2}$ باشد

اگر از این جمعیت نمونه ای انتخاب کنیم میانگین $\bar{x}$ و واریانس S^2 باشد

اگر μ_0 یک مقدار مفروض برای میانگین جمعیت باشد سه فرضیه متصور میشود

دو دامنه $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$ یک دامنه $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ یک دامنه $\begin{cases} H_0: \mu \geq \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$

اگر واریانس جمعیت σ^2 باشد و تعداد نمونه **n** باشد (طبیعتا نمونه هم میانگین و واریانس خاص خودش را دارد) برای ای نمونه خطای معیار میانگین $SE(\bar{x}) = \sigma/\sqrt{n}$ میشود و برای هر سه حالت، آماره آزمون برابر است با $\frac{\bar{x}-\mu_0}{SE(\bar{x})}$ که آنگاه آماره آزمون نرمال استاندارد میشود و میتوان نوشت

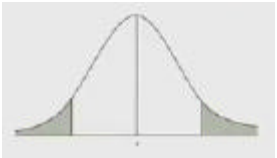
$$z = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

اگر σ نداشتیم آنگاه $SE(\bar{x}) = SD/\sqrt{n}$ و دیگر توزیع نرمال نمیشود و از توزیع **t** با **n-1** درجه آزادی استفاده میکنیم

$$t = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = \frac{\bar{x} - \mu_0}{SD/\sqrt{n}}$$

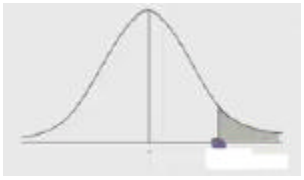
توجه شود در مشاهده نمونه ها $x_1, x_2, x_3, \dots, x_n$ میانگین $\bar{x} = \frac{\sum x_i}{n}$ میباشد که اگر این میانگین در **Z** نرمال قرار دهیم به این Z_{obs} گویند (یعنی **Z** مشاهده شده)

در خصوص آزمون دو دامنه ، ناحیه بحرانی در دو سمت محور قرار میگیرد و سطح معنا داری (**sig**) مساحت زیر منحنی با فرمول زیر حاصل میشود



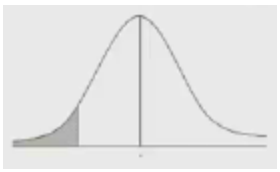
$$Si = 2\,Min\{P(z > z_{obs}), P(z < z_{obs})\}$$

در خصوص آزمون یک دامنه بعدی، که ناحیه بحرانی در سمت راست محور قرار میگیرد و سطح معنا داری (**sig**) مساحت زیر منحنی با فرمول زیر حاصل میشود



$$Si \quad = P(z > z_{obs})$$

در خصوص آزمون یک دامنه بعدی، که ناحیه بحرانی در سمت چپ محور قرار میگیرد و سطح معنا داری (**sig**) مساحت زیر منحنی با فرمول زیر حاصل میشود



$$Si \quad = P(z < z_{obs})$$

اگر **Sig** از مقدار آلفا کوچکتر باشد باید فرض صفر را رد کنیم

اگر واریانس σ^2 نامشخص باشد از t_{obs} جایگزین میشود و از توزیع **t** استفاده میشود (سطح معنا داری **t** خارج از درس میشود)

بجای محاسبه سطح معنا داری میتواند تعیین ناحیه بحرانی مد نظر باشد آنگاه مرز ناحیه بحرانی طوری انتخاب میشود که مساحت فضای هاشور خورده به میزان خطای نوع اول شود

اگر سطح معنا داری از خطای نوع اول α کمتر شود آنگاه با ضریب اطمینان $1 -$ فرض صفر رد میشود

مثال : از تحقیقات قبلی مشخص میشود که در ایران هوش عاطفی زنانی که تحصیلات عالی دارند دارای توزیع نرمال با میانگین ٦٥ و انحراف معیار ١٢ میباشد این تحقیقات ادعا دارد گه در استان آذربایجان میانگین هوش عاطفی زنان بیشتر از دیگر استانها است یک نمونه ١٠٠ نفری در آذربایجان انتخاب کردیم و مشخص شد که میانگین ٦٧ است با سطح اطمینان ٩٥٪ ادعای تحقیقات را آزمون کنید

آزمونی یک دامنه است

$$\begin{Bmatrix} H_0: \mu \le \mu_0 \\ H_1: \mu > \mu_0 \end{Bmatrix} \qquad \begin{Bmatrix} H_0: \mu \le 65 \\ H_1: \mu > 65 \end{Bmatrix}$$

$$E(\overline{x}) = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{100}} = 1.2 \qquad Z_{obs} \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{68 - 65}{1.2} = 2.5$$

$$Si \quad = P(Z > Z_{obs}) = P(Z > 2.5) = 1 - P(Z \le 2.5) = 1 - \ 0.9938 = 0.0062$$

حال چون رابطه زیر برقرار است

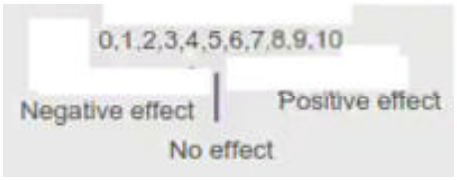
$$Si \quad = 0.0062 \le 0.05$$

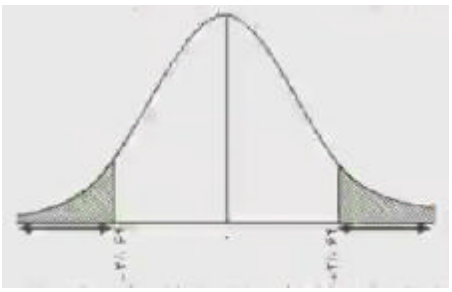
در نتیجه با اطمینان ٩٥٪ میتوان فرض صفر **H0** را رد نمود و پذیرای فزض **H1** شد

در فرض یک ذکر شده که میانگین هوش عاطفی زنان با تحصیلات دانشگاهی از ۶۵ بیشتر است پس ادعا مورد تایید است.

مثال : از ۲۵ نفر از معلمان برای رضایت از اثر بخشی نظام جدید آموزشی، نظر سنجی شد. ۱۰ سوال دو جوابی مطرح شد پاسخها بین صفر تا ۱۰ ده بود بطوریکه عدد صفر بیانگر اثر منفی و عدد پنج بی اثر و عدد ۱۰ اثر مثبت است. میانگین ٦ و انحراف معیار ۳ گردید

ادعا این است که معلمان نظام جدید آموزشی جدید را نسبت به قدیم بدون تاثیر است یعنی تغییری اساسی بوجود نیاورده است. با ضریب اطمینان ۹۵٪ ادعا را بیازمایید





$\begin{Bmatrix} H_0: \mu = \mu_0 \\ H_1: \mu \neq \ _0 \end{Bmatrix}$, $\begin{Bmatrix} H_0: \mu = 5 \\ H_1: \mu \neq \ \ \end{Bmatrix}$ دو دامنه ,

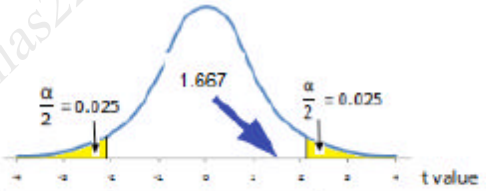
$n = 25$, $\bar{x} = 6$, $SD = 3$

$$t_{ab} = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = \frac{\bar{x} - \mu_0}{SD/\sqrt{n}} = \frac{6-5}{3/\sqrt{25}} = 1.667$$

چون انحراف معیار جمعیت نداریم و انحراف معیار نمونه **SD** داریم آماره آزمون توزیع **t** با **24** درجه آزادی داریم که از جدول **t** برای خطای نوع اول**5%** (ضریب اطمینان **95%**) استفاده میکنیم برای دو دامنه **2.064** میگردد و چون **1.667 < 2.064** میباشد در حقیقت **1.667** در ناحیه بحرانی (هاشور زده) قرار ندارد پس دلیلی بر رد صفر نداریم ($H_0: \mu = \mu_0$) پس با اطمینان **95%** اعلام میکنیم که نظام آموزشی جدید بی تاثیر بوده است (نه منفی است نه مثبت)

Student's t Distribution Table

For example, the t value for 18 degrees of freedom is 2.101 for 95% confidence interval (2-Tail α = 0.05).



	90%	95%	97.5%	99%	99.5%	99.95%	1-Tail Confidence Level
	80%	90%	95%	98%	99%	99.9%	2-Tail Confidence Level
	0.100	0.050	0.025	0.010	0.005	0.0005	1-Tail Alpha
df	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha
22	1.3212	1.7171	2.0739	2.5083	2.8188	3.7921	
23	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676	
24	1.3178	1.7109	2.0639	2.4922	2.7969	3.7454	
25	1.3163	1.7081	2.0595	2.4851	2.7874	3.7251	
26	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066	

اگر محاسبه سطح معنا داری(**sig**) مد نظر نباشد اندیس **obs** در **z** و **t** نمی افزاییم یعنی اگر **z** به یک عدد برسد همان z_{obs} بعنی مشاهده شده است و اگر فرمول باشد اندیس ندارد همان آماره آزمون است ضمنا خطای نوع اول یعنی **5%** یا **1%** است.

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 بدبینیهای 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 شادیبخش 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 کداممان باید حق آن را داشته باشیم که با صدای بلند بگوییم
" من آنچه در توان داشته ام انجام داده‌ام " لوئی پاستور ۱۸۹۵-۱۸۲۲