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)

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 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**)

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

Xiداده	**12**	**13**	**14**	**15**	**17**
fتعداد	**1**	**3**	**2**	**2**	**2**
Fتجمیع	۱	۴	۶	۸	۱۰

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

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

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

Xiداده	**12**	**13**	**14**	**15**	**17**
f_iتعداد	**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** ⇨ **6+** ⇨ در ردیف**Fi** **6+ == 8** ⇨ داده متناظر **x=15**

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** $\qquad R_i = \frac{\sum_{-\infty}^{a} f_i}{\sum f_i}$

میانگین شاخصی جهت مرکز داده ها

واریانس σ^2 و انحراف معیار σ = شاخصی جهت پراکندگی داده ها

میانگین جمعیت و واریانس و انحراف معیار جمعیت – اگراز هر داده فراوانی (به تعداد مشخصی) داشته باشیم و تعداد داده از ۳۰ تا بیشتر باشد

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

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

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

اگراز هر داده به تعداد یکی داشته باشیم میتوان نوشت

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

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

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

فرمول میانگین نمونه با جمعیت فرقی ندارد ولی فرمول واریانس نمونه برابر است با

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

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

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

با استفاده از فرمول توزیع نرمال، ثابت میشود که در طبیعت آمار بصورت نرمال توزیع شده است

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

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

☜ **مشخصه ضریب تغییرات** = **شاخصی جهت میزان پراکندگی 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 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.

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

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	٦	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

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

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

میانه ۱۲ میباشد

ج)

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

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

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

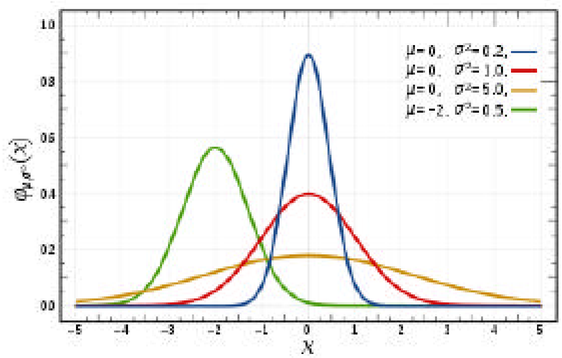
$$\boldsymbol{\overline{x}=\frac{\sum x_i f_i}{\sum f_i}=\frac{(10*2)+(12*4)+(15*2)+(16*1)+(18*1)}{(2+4+2+1+1)}=13.2}$$

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

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

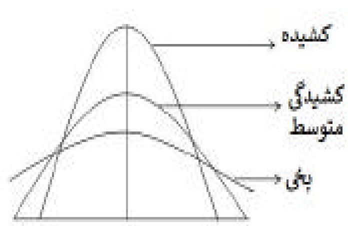
$$\boldsymbol{=\sigma^2 \qquad \sigma=2.52}$$

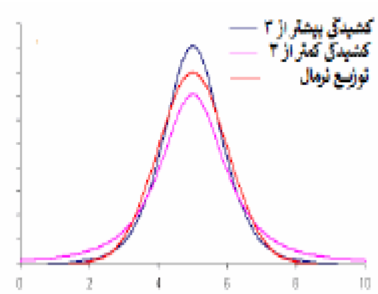
توزیع نرمال – نرمال استاندارد



منحنی توزیع نرمال

کشیدگی (Kurtosis)

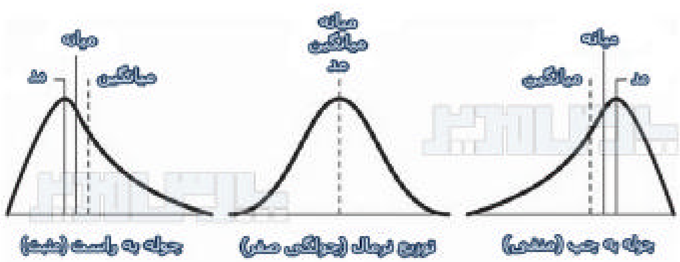




مقدار کشیدگی برای توزیع نرمال برابر ۳ می باشد

****

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

$$y = a + bx$$

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

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

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

$$y = a + bx$$

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

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

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

$$y = 2.33 + 1.5x$$

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

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

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

S= { HH , HT , TH , TT } فضای کلی

احتمال هر حالت **¼ ¼ ¼ ¼**

A= { HH,HT,TH } فضای مورد نظر

P(A)=3/4 احتمال

تعداد شیرX	0	1	2
احتمال شیر P=f(x)	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 $\boldsymbol{f(x)}$

توزیع احتمال نرمال x

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

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

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 $Z=\dfrac{X-\mu}{\sigma}$ 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

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

حالا از روی جدول نرمال استاندارد احتمال حاصل میشود

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

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	[illegible]	0.0003	0.0003	0.0003
-3.3	0.0005	0.0005	[illegible]	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	[illegible]	0.0016	0.0016
-2.8	0.0026	0.0025	0.0024	[illegible]	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 ?= $\boldsymbol{P(z \le -2.73)}$

با توجه به جدول فوق

$$\boldsymbol{P(z \le -2.73) = 0.0032}$$

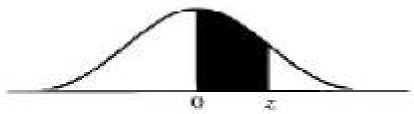
جدول زیر جدول توزیع احتمال نرمال استاندارد است

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

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

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

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 1 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

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

بدینترتیب میگوییم جمعیت به نرمال استاندارد تغییر یافت

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

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

$1 - 0.1056 = 0.8944$ (ب

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

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

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

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

واریانس و انحراف معیار در این آزمایش برابر است با

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

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

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

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

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

ضمنا همانطوریکه قبلا گفته شد میتوان گفت که :

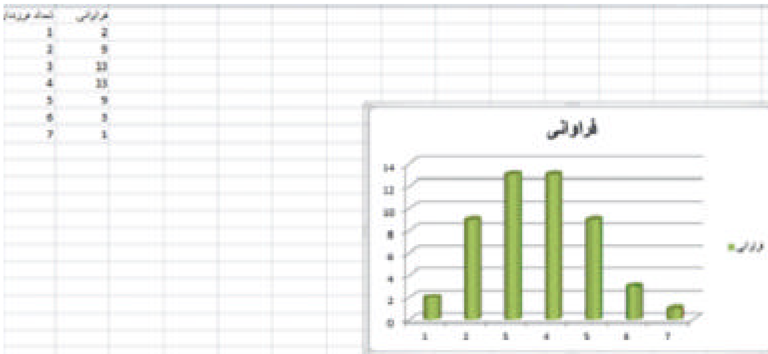
$$\mu \pm \sigma = 80 \pm 4 = 76 \;,\; 84 \qquad\qquad \mu \pm 2\sigma = 80 \pm (2 * 4) = 72 \;,\; 88$$

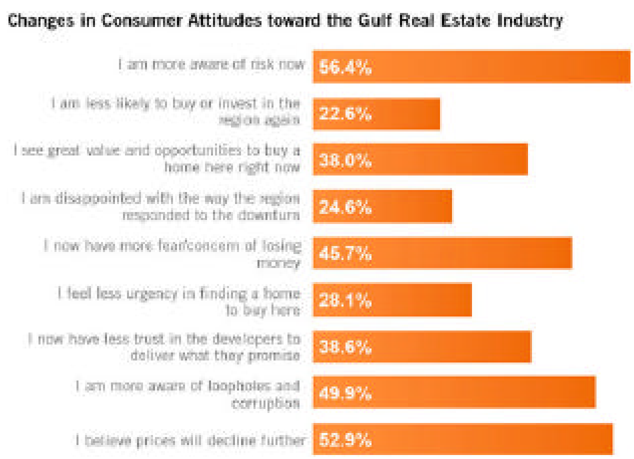
بیش از ۶۸٪ احتمال دارد فردا به تعداد **76** تا **84** تیر به هدف اصابت کند و بیش از ۹۶٪ احتمال دارد فردا به تعداد **72** تا **88** تیر به هدف اصابت کند

☜ انواع نمودارها

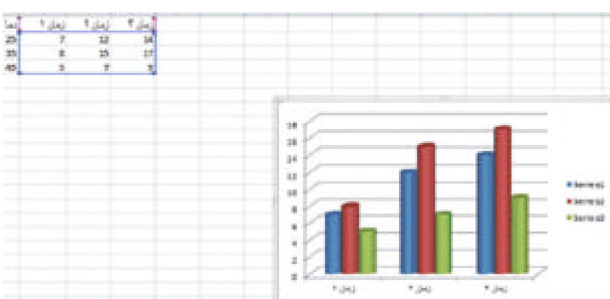
۱)) نمودار میله‌ای بارچارت **Bar Chart**:

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



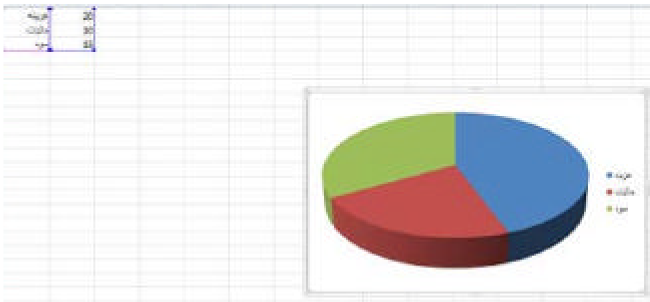


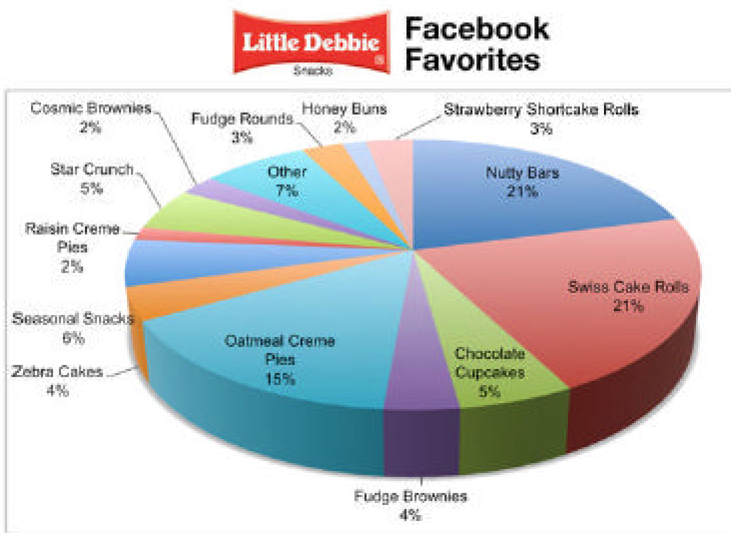
۲)) میله ای ستونی – مشاهده فراوانی و مقایسه داده ها بصورت دسته بندی شده



۳)) نمودار دایره ای**Pie Chart** :

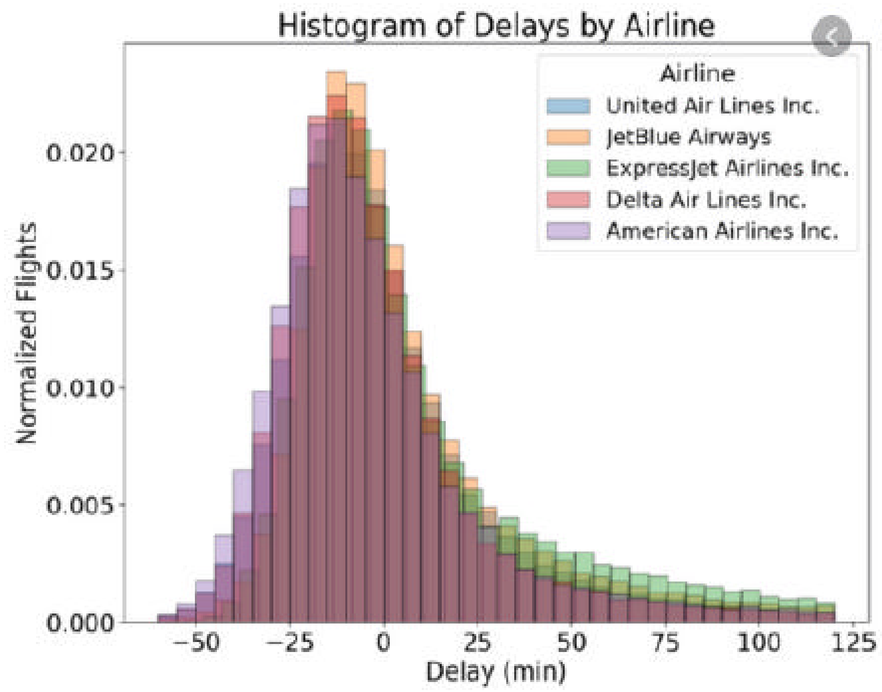
برای مشاهده سهم هر مورد از داده ها



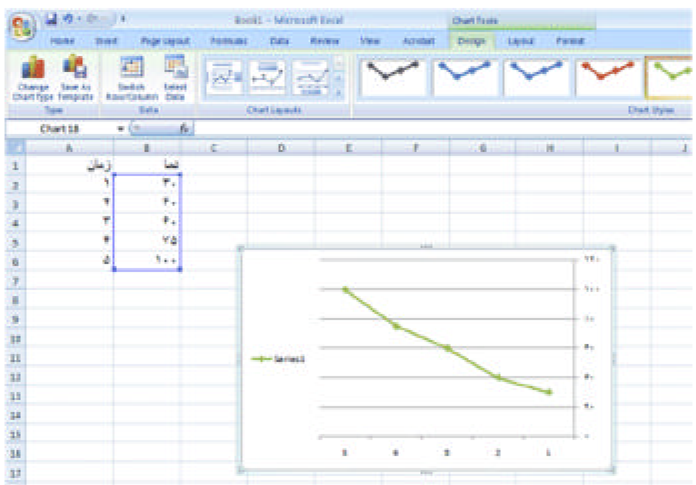


۴)) نمودار هیستوگرام **Histogram**:

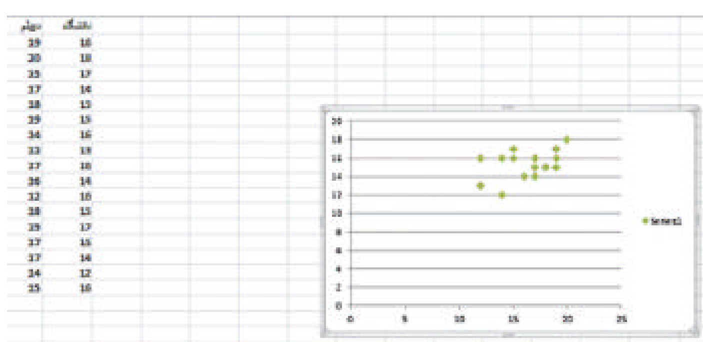
برای مشاهده نحوه توزیع داده ها بصورت پیوسته (مثلا معدل دیپلم و معدل دانشگاه)

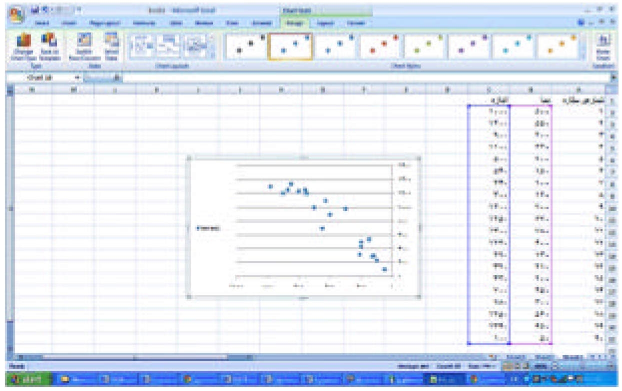


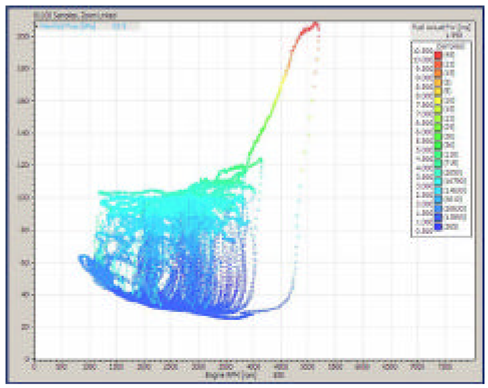
۵)) نمودار خطی **Line Chart**: برای نشان دادن رابطه بین دوسری اعداد مثلا خط رگرسیون



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 *Inferential Statistics*

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

برآوردگرها *Estimators*

تعریف پارامتر = شاخص های آماری مرتبط با ویژگیهای جمعیت باشد

تعریف آماره = شاخص های آماری مرتبط با ویژگیهای نمونه باشد

جدول شاخص‌ها

انحراف معیار	واریانس	میانگین		
$\boldsymbol{\sigma}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{\mu}$	پارامتر	جمعیت
$\boldsymbol{SD}$	$\boldsymbol{S^2}$	$\boldsymbol{\overline{x}}$	آماره	نمونه

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

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

$X_1 ,\ X_2 , X_3 ,\ X_4 , \ldots..$

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

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

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

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

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

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

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 $(1-\alpha)\%$ نشان میدهند

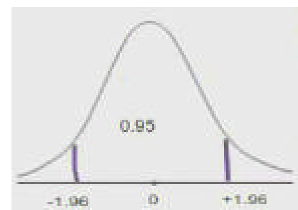
قبلا در توزیع نرمال اشاره شد که سطح زیر منحنی برای ۹۵٪ بین $z = \pm 1.96$ میباشد

یعنی با اطمینان $(1-\alpha)\% = 95\%$ یعنی ۹۵٪ میتوان گفت که **z** بین ۱،۹۶- تا ۱،۹۶ + میباشد

$$z = \frac{x - \bar{x}}{\sigma/\sqrt{n}} \qquad -1.96 < \frac{x - \bar{x}}{\sigma/\sqrt{n}} < +1.96$$

$$Low = \bar{x} - 1.96 * \sigma/\sqrt{n}$$

$$Up = \bar{x} + 1.96 * \sigma/\sqrt{n}$$



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 : در کل یک دانشگاه میانگین ۱۴ و واریانس ۳,۶ میباشد احتمال اینکه نمره یک دانشجو در کل این دانشگاه بیشتر از ۱۶ باشد چقدر است؟ (قبلا در مورد توزیع نرمال اشاره شده بود)

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

چون جمعیت ذکر شده و راجع به جمعیت هم سوال دارد پس مخرج کسر مثل قبل بصورت زیر مینویسیم

$$\boldsymbol{p(x > 16) = 1 - p(x \le 16) = 1 - p\left(\frac{x-\mu}{\sigma} \le \frac{16-\mu}{\sigma}\right) = 1 - p\left(z \le \frac{16-14}{1.89}\right) = 1 - p(z \le +1.058)}$$
$$\boldsymbol{= 1 - 0.8531 = 0.145}$$

به احتمال ۱۴٪ نمره یک دانشجو بیشتر از ۱۶ میشود

مثال : در کل یک دانشگاه میانگین ۱۴ و واریانس ۳,۶ میباشد احتمال اینکه نمره یک دانشجو در یک نمونه کلاس ۳۶ نفره بیشتر از ۱۶ باشد چقدر است؟

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

چون جمعیت ذکر شده ولی راجع به نمونه سوال دارد پس مخرج کسر باید بصورت زیر نوشت

$$\boldsymbol{p(x > 16) = 1 - p(x \le 16) = 1 - p\left(\frac{x-\mu}{\sigma/\sqrt{n}} \le \frac{16-\mu}{1.89/\sqrt{36}}\right) = 1 - p\left(z \le \frac{16-14}{0.315}\right)}$$
$$\boldsymbol{= 1 - p(z \le +6.35) = 1 - 0.9999 = 0.0001}$$

به احتمال ۱٪ نمره یک دانشجو بیشتر از ۱۶ میشود

مثال : در یک کلاس میانگین ۱۴ و واریانس ۳،۶ میباشد با اطمینان ٪۹۵($X_{0.05}$) بیشترین نمره کلاس را بدست آورید

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

چون جمعیت ذکر شده و راجع به جمعیت هم سوال دارد پس مخرج کسر مثل قبل بصورت زیر مینویسیم

$$\boldsymbol{p(x \le k) = 0.95}$$

$$\boldsymbol{p\left(\frac{x-\mu}{\sigma} \le \frac{k-\mu}{\sigma}\right) = 0.95}$$

$$\boldsymbol{p\left(z \le \frac{k-14}{1.89}\right) = 0.95}$$

$$\boldsymbol{p\left(z \le \frac{k-14}{1.89}\right) = 0.95 = \text{از جدول} = 0.95 = p(z \le 1.645)}$$

$$\boldsymbol{\frac{k-14}{1.89} = 1.645 \quad \rightarrow \quad k - 14 = 1.89 * 1.645}$$

$$\boldsymbol{\rightarrow \quad k = 17.1}$$

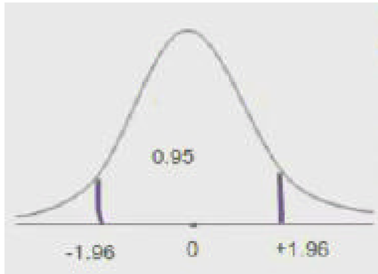
یعنی به احتمال ٪۹۵ نمره دانشجو حداکثر ۱۷،۱ خواهد بود یا ٪۵ احتمال دارد نمره دانشجویی بیشتر از ۱۷،۱ شود

-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	[illegible]	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	9505	.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	[illegible]	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

مثال : در کل یک کلاس میانگین ۱٤ و واریانس ۳,٦ میباشد با اطمینان ۹۵٪ ($X_{0.05}$) محدوده نمرات را بدست آورید

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

برای بدست آوردن اطمینان ۹۵٪ یعنی محدوده بالا و پایین نمرات از هر طرف ۲۵٪ جمعا به احتمال ۹۵٪ میتوان نوشت



حل بطور خلاصه

$$\boldsymbol{Low = k - 1.96 * \sigma = 14 - 1.96 * 1.89 = 14 - 1.96 * 1.89 = 10.29}$$
$$\boldsymbol{Up = k + 1.96 * \sigma = 14 + 1.96 * 1.89 = 14 + 1.96 * 1.89 = 17.7}$$

یعنی با اطمینان ۹۵٪ میتوان گفت میانگین جمعیت بین ۱۰,۲۹ تا ۱۷,۷ میباشد

توضیح کامل

حد نمره بالا

$$\boldsymbol{p(x \le k) = 0.975 \quad \Rightarrow \quad p\left(\frac{x-\mu}{\sigma} \le \frac{k-\mu}{\sigma}\right) = 0.975 \quad \Rightarrow \quad p\left(z \le \frac{k-14}{1.89}\right) = 0.975}$$

$$\boldsymbol{p\left(z \le \frac{k-14}{1.89}\right) = 0.975 \Rightarrow \text{از جدول} \Rightarrow 0.975 = p(z \le 1.96)}$$

$$\boldsymbol{\frac{k-14}{1.89} = 1.96 \quad \Rightarrow \quad k = 14 + (1.96 * 1.89) \quad \Rightarrow k = 17.7}$$

حد نمره پایین

$$\boldsymbol{p(x \le k) = 0.025 \quad \Rightarrow \quad p\left(\frac{x-\mu}{\sigma} \le \frac{k-\mu}{\sigma}\right) = 0.025 \quad \Rightarrow \quad p\left(z \le \frac{k-14}{1.89}\right) = 0.025}$$

$$\boldsymbol{p\left(z \le \frac{k-14}{1.89}\right) = 0.025 \Rightarrow \text{از جدول} \Rightarrow 0.025 = p(z \le -1.96)}$$

$$\boldsymbol{\frac{k-14}{1.89} = -1.96 \quad \Rightarrow \quad k = 14 + (-1.96 * 1.89) \quad \Rightarrow k = 10.29}$$

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	[illegible]	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.025[illegible]	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	[illegible]	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

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	[illegible]	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	[illegible]	0.9808	0.9812	0.9817

مثال: از نمونه ۲۰۰ نفری دانشـــجویان میانگین نمرات ۱۳ و انحراف معیار ۱،۵ میباشـــد فاصـــله اطمینان ۹۵٪ برای میانگین را محاســـبه کنید و توضیح بنویسید

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

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

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

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

1-0.95=0.0.05   0.05/2=0.025   0.95+0.025=0.975 $\boldsymbol{p(z \le 0.975) = 1.96}$

$$Low = \bar{x} - 1.96 * \frac{SD}{\sqrt{n}} = 13 - 1.96 * \frac{1.5}{\sqrt{200}} = 12.78$$

$$Up = \bar{x} + 1.96 * \frac{SD}{\sqrt{n}} = 13 + 1.96 * \frac{1.5}{\sqrt{200}} = 13.22$$

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

یا به شرح زیر میتوان نوشت

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	[illegible]	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	[illegible]	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

$$p = (x \le \hat{x}) = p\left(\frac{x-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{\hat{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = p\left(z \le \frac{\hat{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = p\left(z \le \frac{\hat{x}-13}{\frac{1.5}{\sqrt{200}}}\right)$$

$$p\left(z \le \frac{\hat{x}-13}{0.11}\right) = 0.025 \quad \rightarrow \quad p(z \le -1.96) = 0.025 \quad \rightarrow \quad \frac{\hat{x}-13}{0.11} = -1.96$$

$$\hat{x} = (-1.96 * 0.11) + 13 = 12.78$$

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	[illegible]	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	[illegible]	0.9808	0.9812	0.9817

$$p = (x \leq \hat{x}) = p\left(\frac{x-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\hat{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = p\left(z \leq \frac{\hat{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) = p\left(z \leq \frac{\hat{x}-13}{\frac{1.5}{\sqrt{200}}}\right)$$

$$p\left(z \leq \frac{\hat{x}-13}{0.11}\right) = 0.975 \quad \rightarrow \quad p(z \leq +1.96) = 0.975 \quad \rightarrow \quad \frac{\hat{x}-13}{0.11} = 1.96$$

$\hat{x} = (1.96 * 0.11) + 13 =$**13.22**

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 μ و واریانس σ^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 $\begin{cases} H_0: \mu \leq \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$ *on t i ed* 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

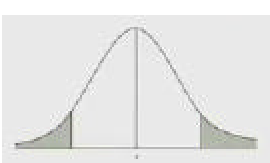
$$z = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

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

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

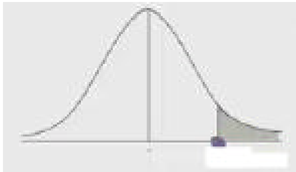
توجه شود در مشاهده نمونه ها $x_1, x_2, x_3, \dots, x_n$ میانگین $\bar{x} = \frac{\sum x_i}{n}$ میباشد که اگر این میانگین در **Z** نرمال قرار دهیم به این Z_{ob} گویند (یعنی **Z** مشاهده شده)

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



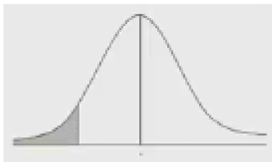
$$Si \quad = 2\, M \;\; n \{P(z > z_{ob}\;), P(z < z_{ob}\;)\}$$

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



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

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



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

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

اگر واریانس σ^2 نامشحص باشد از t_{ob} 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 $1 - \alpha$ فرض صفر رد میشود

مثال : از تحقیقات قبلی مشـــخص میشـــود که در ایران هوش عاطفی زنانی که تحصـــیلات عالی دارند دارای توزیع نرمال با میانگین ٦٥ و انحراف معیار ١٢ میباشد این تحقیقات ادعا دارد گه در استان فارس میانگین هوش عاطفی زنان بیشتر از دیگر استانها است .

یک نمونه ١٠٠ نفری در فارس انتخاب کردیم و مشخص شد که میانگین ٦٨ است

با سطح اطمینان ٩٥٪ ادعای تحقیقات را آزمون کنید یعنی ثابت کنید در استان فارس میانگین هوش عاطفی زنان بیشتر از دیگر استانها است

آزمونی یک دامنه است

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

$$Z_{ob} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{68 - 65}{1.2} = 2.5$$

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

حال چون رابطه زیر برقرار است

$$Si = 0.0062 \le 0.05$$

در نتیجه با اطمینان ۹۵٪ 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.

اینگونه هم میتوان نوشت

$$p(\bar{x} > 65) = 1 - p(\bar{x} \le 65) = 1 - p\left(\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \le \frac{68 - 65}{\frac{12}{\sqrt{100}}}\right) = 1 - p(z \le 2.5) = 1 - 0.9938$$
$$= 0.0062 = s\ g = s\ g \le 0.05$$

پس فرض صفر رد میشود

آزمون فرضها به زبانی دیگر :

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)

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

$H_a:\ \mu>1200$

$H_0:\mu=1200$

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

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

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

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

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