بسم‌الله الرحمن الرحیم

آمار

توصیفی

جزوه قسمت اول

سرفصل آمار توصیفی

1) تعریف مفاهیم (ریاضی آمار اندازه گیری سنجش)
2) طبقه بندی (آمار توصیفی و استنباطی)
3) جامعه نمونه پارامتر مشخصه آماری
4) مفهوم متغیر و انواع آن
5) طبقه بندی متغیر از لحاظ مقیاس اندازه گیری
6) انواع نمودار و کاربرد آن در آمار
7) طبقه بندی آمار توصیفی (یک متغیری چند متغیری)
8) شاخص های مرکزی (مد میانه میانگین)
9) شاخص های پراکندگی (واریانس انحراف معیار انحراف چارکی و دامنه تغییرات)
10) نقاط درصدی (چندک ها) و رتبه بندی
11) منحنی طبیعی و کاربرد آن در آمار
12) نمره های استاندارد و انواع آن
13) شاخص های همبستگی و انواع آن با تکیه بر همبستگی پیرسون و اسپرمن
14) رگرسیون و پیش بینی
15) آشنایی با نرم افزارهای آماری از جمله SPSS

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)

علم آمار

- **آمار :**

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

- **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 Descriptive

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

مراحل اساسی توصیف داده ها عبارت است از:

1) خلاصه کردن داده ها و توصیف الگوی کلی
2) فشرده کردن داده ها در قالب جدول های آماری
3) نمایش آن ها به وسیله ی نمودار
4) محاسبه شاخص های آماری

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

آمار استنباطی

آمار استنباطی مشخص می کند که آیا الگوها 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

نرم افزارهای Lisrel - SPSS - Eviews - SmartPLS - Amos - Minitab - Statgraph - برنامه نویسی GAMS - R - STATA - Statistical Analysis Software (SAS) - PASS(NCSS) -

اس پی اس اس SPSS ((Statistical package for social science))

متغیر:

یک ویژگی که از فردی به فرد دیگر در یک جمعیت (نمونه) تغییر میکند

1- متغیر کمی (وزن یا قد)
2- متغیر توصیفی (مهارت یا هوش)

مقیاس سازی

1- مقیاس اسمی (تهران 021 - شیراز 0711)
2- مقیاس ترتیبی (ضعیف 0 – متوسط 1 – خوب 2)
3- مقیاس فاصله ای (x=at+b که b<>0 مثلاً f=9/5(c+32)) در این مقیاس صفر بمعنی هیچ نیست
4- مقیاس نسبتی (x=at) صفر بمعنی هیچ است این مقیاس خوبی است

داده :

1- داده گسسسته (تعداد فرزند)
2- داده پیوسته (طول قد)

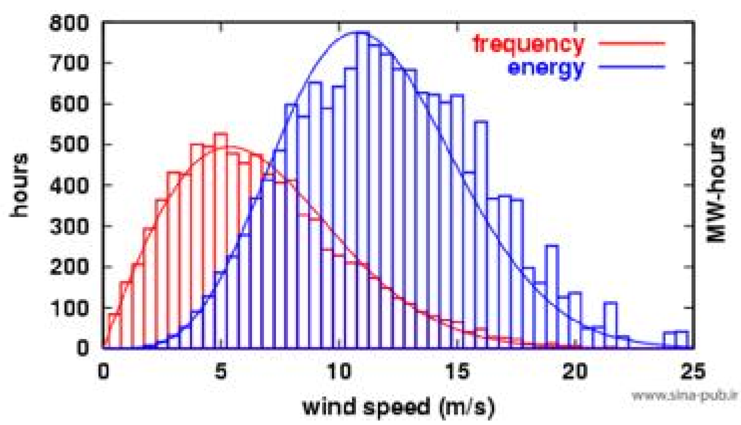
مشخص کننده مرکزی

1- میانگین (حسابی – وزنی – هندسی – توافقی و ...)
2- میانه (وسط صف منظم داده ها) m
3- نما (داده ای که بیشتر تکرار شده) M
4- میانگین و واریانس (میانگین و میزان انحراف داده ها از میانگین)
5- چارک – دهک – صدک یعنی داده ها از این چندک کوچکترند

فراوانی

$p_{\text{ٔ}}$.

1- **فراوانی مطلق :** به تعداد داده‌ها در هر طبقه فراوانی مطلق آن طبقه می‌گویند و آن را با fi نشان می‌دهند.
2- **فراوانی نسبی :** در صورتی که فراوانی‌های مطلق را بر کل فراوانی‌ها تقسیم کنیم، فراوانی نسبی ri به دست می‌آید.
3- **فراوانی تجمعی :** به مجموع فراوانی‌های مطلق طبقه‌های قبل و همان طبقه، فراوانی تجمعی آن طبقه می‌گویند و آن را با Fi نمایش می‌دهند.
4- **فراوانی تجمعی نسبی :** می‌توان از تقسیم فراوانی های تجمعی بر تعداد داده‌ها، این فراوانی را به دست آورد(Ri) .



چندک ها

چارک – دهک – صدک یعنی داده ها ار این چندک کوچکترند (ابتدا داده ها را مرتب کنیم سپس تعداد داده ها (n) را در نسبت مربوط به چندک (p) ضرب کنید n*p (در صورت اعشاری بودن آن را به بالا گرد کنید و داده را استخراج نمایید در صورت مساوی بودن آن - میانگین آن داده و داده بالاتر را استخراج کنید) تا محل قرار گرفتن چندک معلوم شود. P_{40}

صدک (P) : صدک چهلم یا دهک چهارم است و داده‌ای است که 40 درصد داده‌ها از آن کوچکتر است. بدیهی است که 60 درصد داده‌ها از آن بزرگتر باشد.

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

چارک Q: چارک سوم است داده‌ای است که سه چهارم (75٪) داده‌ها از آن کوچکتر است. بدیهی است که یک چهارم (25٪) دیگر داده‌ها از آن بزرگتر باشد.

مثال 1) مثال

صدک 65 داده‌های زیر را معلوم کنید. 3.2-3.6-3.5-3.7-3.9-3.8-4.0-4.3-4.4-3.5

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

Data=Xi	3.2	3.5	3.5	3.6	3.7	3.8	3.9	4	4.3	4.4
F	1	1	1	1	1	1	1	1	1	1
F	1	2	3	4	5	6	7	8	9	10

برای محاسبه صدک 65 باید تعداد داده ها را در 65٪ ضرب کنیم. (65٪ نسبت مربوط به صدک 65 است)

n*p=10*0.65=6.5 6.5+(=7) سطر در n=> n=7 ➔ X=3.9

بنابراین عدد هفتم، صدک 65 ام است یعنی: $P_{65}=3.9$

یعنی 65٪ دادها مقدارعددیشان از 3.9 کمتر است

مثال 2) مثال

دهک چهارم داده‌های زیر را معلوم کنید

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

داده‌ها را مرتب میکنیم

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

برای محاسبه دهک چهارم باید تعداد داده ها (n=15) را در0.4 ضرب کنیم

n*p=15*0.4=6 6+➔ F ➔ F=7 7➔X ➔X=3.8

یعنی 40%داده ها کمتر از 3.8 ام است یعنی: $P_{40}=3.8$

مثال 3) مثال

دادههایی بشرح ذیل داریم صدک 60 چه مقدار میشود

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

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

دادهXi	12	13	13	13	14	14	15	15	17	17
تعدادf	1	1	1	1	1	1	1	1	1	1
تجمیعF	1	2	3	4	5	6	7	8	9	10

n*p=10*(60/100)=6 ⇨ در تجمیع 6+ == 7 ⇨ داده متناظر با تجمیع مربوطه ⇨ x=15

صدک شصت (شصت درصد) داده 15 یا کمتر از آن میباشد

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

Xi داده	12	13	14	15	17
fi تعداد	1	3	2	2	2
Fi	1	4	6	8	10

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

n*p=10*(60/100)=6 ⇨ 6+ ⇨ در ردیفFi 6+ == 8 ⇨ داده متناظر x=15

صدک شصت (شصت درصد) داده 15 یا کمتر از آن میباشد

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

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

مشخصه های پراکندگی :

1- دامنه تغییرات (تفاضل بزرگترین داده از کوچکترین داده)
2- انحراف متوسط (میانگین قدر مطلق انحراف ها از میانگین داده ها)
3- انحراف معیار (جذر واریانس که واریانس میانگین توان دوم انحراف ها از میانگین داده ها)

☜ فراوانی :

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

1- فراوانی نسبی نسبت فراوانی هر مجموعه به کل تعداد آن مجموعه ri=wi/n
2- فراوانی انباشته (تجمعی) حاصل جمع فراوانی تا موقعیت مورد نظر Fj
3- فراوانی انباشته نسبی : حاصل جمع Wi ها تقسیم بر n

☜ میانگین

$$\overline{X} = \frac{1}{n}\sum_{h=1}^{n} x_h$$

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

☜ واریانس σ^2 انحراف معیار σ

$$v(x) = \sigma^2 = \frac{1}{n}(\sum (x_i - \mu)^2$$

یا

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

در حالت داشتن فراوانی

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

$$v(x) = \sigma^2 = \frac{1}{\sum f_i}(\sum f_i (x_i - \mu)^2$$

ثابت شـده اسـت که در طبیعت آمار بصورت نرمال توزیع شده است که معمولاً ٪68 داده ها بین $\mu \pm \sigma$ و ٪96 داده ها بین $\mu \pm 2\sigma$ خواهد بود

مثال 4) مثال

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

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

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

$s^2 = 1.805$

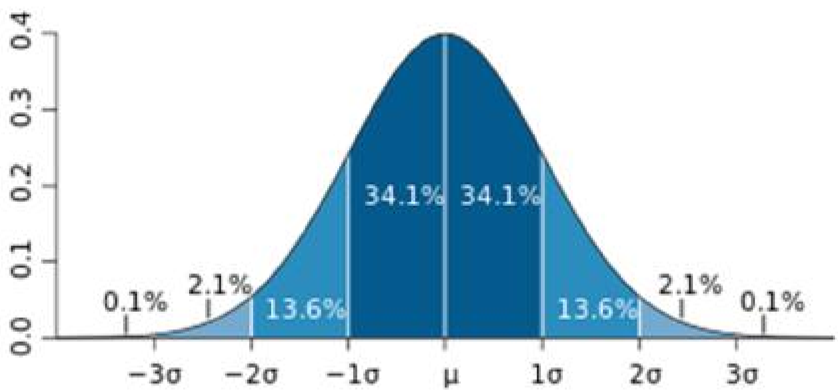
$s = 1.34$

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

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

$\mu \pm 3\sigma = 09.37 \qquad 16.43$

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



خواص میانگین حسابی

اگر داده های ما x_i باشد میانگین آنها $\bar{x}$ باشد آنگاه جمع داده ها با میانگین با توجه به فراوانی صفر میشود

$\sum(x_i-\bar{x})\,f_i=0$

اگر داده ها را با یک عدد ثابت جمع یا تفریق کنیم میانگین هم با همین عدد ثابت جمع یا تفریق میشود

$y_i=x_i\pm a \;\Rightarrow\; \bar{y}=\bar{x}\pm a$

$y_i=x_i*a \;\Rightarrow\; \bar{y}=\bar{x}*a$

اگر y_i یکسری اعداد دیگر باشد

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

خواص واریانس(پراش)

اگر داده های ما x_i باشد میانگین آنها $\bar{x}$ باشد آنگاه

اگر داده ها را با یک عدد ثابت جمع یا تفریق کنیم واریانس و انحراف معیار تغییر نمیکند

$y_i=x_i\pm a \;\Rightarrow\; \boldsymbol{\sigma_y=\sigma_x}$

مثال 5) مثال

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

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

ضرب اعداد فوق در فراوانی ها وقت زیاد میبرد یک عدد در حدود وسط داده ها انتخاب میکنیم اعداد جدول فوق از سایز عدد 170 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 170 اضافه میکنیم و واریانس فرق نمیکند

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

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

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

$\sigma_x=\sigma_y=2.13$

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

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

مثال 6) مثال

نمرات 10 واحد درسی دانشجویی به شرح ذیل میباشد

نمره = x	18	12	15	16	10
تعداد(واحد) =f	1	4	2	1	2

الف) مد چه عددی است(چرا) – ب) میانه چه عددی است(چرا) ج) صدک 85 داده ها بدست آورید د)میانگین و واریانس و انحراف معیار را بدست آورید

حل: ابتدا داده ها مرتب میکنیم

X_i	10	12	15	16	18	
f_i	2	4	2	1	1	
Fi	2	6	8	9	10	

داد ها گسسته هستند زیرا بین داده ها پیوسته نیست

الف) مد داده با بیشترین تکرار که نمره 12 مود میباشد

ب) برای میانه باید وسط کل داده ها را پیدا کنیم جمع داده ها را بر 2 تقسیم کنیم(نصف 10 برابر 5میشود) و در ستون Fi دنبال خود این جواب و یا کران بالای آن میگردیم که Fi=5 نداریم و کران بالا عدد6 میشود و داده متناظر آن نمره 12 میانه میشود

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

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

میانه 12 میباشد

ج)

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

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

صدک 85 عدد 16 شد یعنی 85٪ دادها 16 یا کمتر از 16 میباشند

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

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

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

$$= 6.36 = \sigma^2 \qquad \sigma = 2.52$$

☜ 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 7) مثال

معادله رگرسیون بر حسب y برای زوجهای مرتب زیر بدست آورید

x = 1,3,4,6,8,9,11,14
y = 1,2,4,4,5,7,8,9
y = a+bx
b = 7/11
a = 5-(7/11)*7=6/11
y = 6/11+(7/11)x

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

x = a+by
a = -0.5
b = 1.5
x = -0.5+1.5y

مثال 8) مثال

در جدول زیر بر حسب اینچ طول قد پدر و پسر ملاحظه میکنید . معادله خط رگرسیون را بنویسید؟ اگر قامت پدری 75 شد پیش بینی میکنید فرزند او دارای چه قدی باشد؟

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

حل :

پدر را X و پسر را Y مینامیم زیرا قد پسر از پدر تبعیت میکند پس پدر متغیر مستقل است

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

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

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

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

$y = a + bx$

$$y = 35.85 + 0.476x$$

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

مثال ۹)

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

سال = x	93	94	95	96	97
مصرف =y	1	2	4	؟	؟

*** حل : ابتدا در جدول مقادیر x قدیمی را Xm نام گذاشته و آنرا را تغییر میدهیم مثلا همه را از 94 کم میکنیم

سال = Xm	93	94	95
x	-1	0	+1
مصرف =y	1	2	4

$y = a + bx$

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

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

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

$$y = 2.33 + 1.5x$$

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

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

مثال 10) مثال

در جدول زیر طول قد لوبیا (به سانتیمتر) بر حسب سن (به هفته) ملاحظه میکنید

معادله خط رگرسیون را بنویسید؟

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

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

حل :

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

$\sum x_i = 1+2+.....+7=28$ $\qquad \sum y_i = 5+13+.....+40=168$

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

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

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

$y=a+bx$

$$y=-0.572+6.143x$$

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

تمرین :

طی آماری طول قامت 50 کودک بشرح ذیل است

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

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

احتمالات

تعداد کل حالات مطلوب تقسیم بر تعداد کل حالات ممکن

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

* همیشه احتمال بین صفر و یک است *

مثال 11) 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

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

توزیع‌های مهم

توزیع دوجمله‌ای (برنولی)

توزیع پواسن

توزیع نرمال X توزیع نرمال استاندارد

ثابت شده است که در طبیعت آمار بصورت نرمال توزیع شده است که معمولاً 68٪ بین $\mu \pm \sigma$ و 96٪ بین $\mu \pm 2\sigma$ خواهد بود

توزیع نرمال استاندارد Z

اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد آنگاه متغیر تصادفی $Z = \dfrac{X-\mu}{\sigma}$ 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 n تایی از داخل m انتخاب کنیم در فرمول بالا بجای σ مقدار $\sqrt{\frac{\sigma^2}{n}}$

طریقه استفاده از جدول نرمال

جدول برای مقادیر کوچکتر یا مساوی است در اولین ستون سمت چپ مقدار عدد مربوط بـه Z کـه هم مقـادیر 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

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	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	[illegible]	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

مثلا اگر سوال شود که $P(z \le -2.73) = ?$

با توجه به جدول فوق

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

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

$Z_{0.05} = a$

$p(z > a) = 0.05$

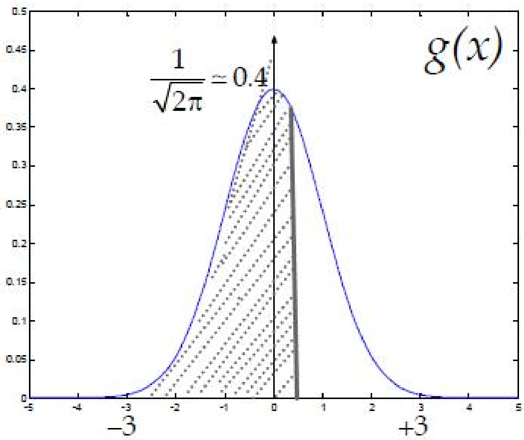
$1 - p(z \leq a) = 0.05$

$p(z \leq a) = 0.95$

$z_{\alpha} = -z_{1-\alpha}$

جدول توزیع نرمال

نرمال استاندارد و منحنی آن



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

$$g(x) = g(-x)$$

$$\int_{-\infty}^{+\infty} g(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

مقدار $g(x)$ برای خارج بازهٔ $(-3, 3)$ بسیار کوچک است و نقاط عطف آن 1 و -1 هستند.

☜ حالت اول توزیع نرمال :

تعداد جمعیت مشخص – میانگین جمعیت مشخص – انحراف معیار جمعیت مشخص – یک سوال میشود که احتمال بدست آمدن a از جمعیت چقدر است

$$X \approx N(\mu , \sigma^2) \qquad p(x < a) = ? \qquad p\left(\frac{x-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = ?$$

بدینترتیب میگوییم جمعیت به نرمال استاندارد تغییر یافت

$p\left(z < \frac{a-\mu}{\sigma}\right) = ?$ حال از جدول نرمال استاندارد

☜ حالت دوم توزیع نرمال :

از جمعیتی - تعداد نمونه مشـخص n – میانگین نمونه مشخص – انحراف معیار نمونه مشخص – یک سوال میشود که احتمال بدست آمدن a از جمعیت چقدر است

$$X \approx N(\mu , \sigma^2) \qquad p(x < a) = ? \qquad p\left(\frac{x-\mu}{\sqrt{\frac{\sigma^2}{n}}} < \frac{a-\mu}{\sqrt{\frac{\sigma^2}{n}}}\right) = ?$$

بدینترتیب میگوییم جمعیت به نرمال استاندارد تغییر یافت

$p\left(z < \frac{a-\mu}{\sqrt{\frac{\sigma^2}{n}}}\right) = ?$ حال از جدول نرمال استاندارد

مثال 12) مثال

در یک کلاس با 40 دانشجو دارای توزیع نرمال و میانگین نمرات 15 و انحراف معیار 4 میباشد یک دانشجو انتخاب میکنیم. الف) احتمال اینکه نمره دانشجو حداکثر 10 شود چقدر است؟ ب) احتمال اینکه نمره دانشجو بیشتر از 10 شود چقدر است(جدول پیوست شده) ج) احتمال اینکه نمره دقیقا 10 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) $= p(z \le -1.25) - p(z \le -1.5) = 0.1056 - 0.0668 = 0.0388$

مثال 13) مثال

از بین 320 نفر دانشجویان دانشگاه، 10 نفر دانشجو انتخاب میکنیم میانگین نمرات این 10 دانشجو برابر با 15 و انحراف معیار 4 میباشد(توزیع نرمال) دانشجویی بتصادف از 320 نفر دانشجویان دانشگاه انتخاب میکنیم

الف) احتمال اینکه نمره دانشجو کمتر از 12 شود چقدر است

ب) احتمال اینکه نمره دانشجو بیشتر از 12 شود چقدر است(جدول پیوست شده)

$$X \approx N(15\,,4^2) \qquad p(x \le 12) = ? \qquad p\left(\frac{x-\mu}{\sqrt{\frac{\sigma^2}{N}}} \le \frac{12-15}{\sqrt{\frac{4^2}{10}}}\right) = ?$$

بدینترتیب میگوییم جمعیت به نرمال استاندارد تغییر یافت

الف) که از جدول نرمال استاندارد $p(z \le -2.37) = 0.0089$

$$p(x > 12) = \; p\left(\frac{x-\mu}{\sqrt{\frac{\sigma^2}{N}}} > \frac{12-15}{\sqrt{\frac{4^2}{10}}}\right) = p(z > -2.37) = 1 - p(z \le -2.37) = 1 - 0.0089 = 0.9911$$

مثال 14) مثال

نمرات امتحان کلاســی دارای توزیع نرمال و میانگین 65 و انحراف معیار 4 میباشــد ، از این کلاس 40 نفری 4 نفر مردود شدند ، کمترین نمره قبولی را مشخص کنید

$$v=\sigma^2=4^2=16$$
$$X\approx N(\mu,\sigma^2)=(65{,}16)$$

برای تقریب به توزیع نرمال چون حدنهایی جمعیت و نمونه یکی است و مشخص است تقسیم بر σ و گرنه تقسیم بر $\sqrt{\frac{\sigma^2}{n}}$

$$p(x<a)=\frac{4}{40}=0.1$$
$$p(\frac{x-\mu}{\sigma}<\frac{a-65}{4})=0.1$$
$$p(z<\frac{a-65}{4})=0.1$$

از روی جدول عدد 1.28 - حاصل میشود

$$\frac{a-65}{4}=-1.28\Rightarrow a=59.88$$

توضیح امید و واریانس و کوواریانس

امید

امید همان معدل همان میانگین میباشد واریانس هم قبلا عنوان شده است p=f(x) میباشد

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

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

$$Cov(x,y)=E(x,y)-E(x).E(y)=E[(x-\bar{x})(y-\bar{y})]$$

اگر x و y مستقل باشند آنگاه

Cov(x,y)=0

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

اگر x و y مستقل آنگاه Cov=0

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

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

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

**ضریب 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

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

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

اگر x و y رابطه خطی 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

مثال در ادامه : *******

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.

الف) روش محاسبه با استفاده از اعداد:

$$p(x,y) = Corr(x,y) = r = \frac{E[(x-\bar{x})(y-\bar{y})]}{\sigma_x \sigma_y} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2} * \sqrt{\sum(y_i-\bar{y})^2}}$$

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

مثال در ادامه : *******

ب) روش محاسبه از طریق نمره های استاندارد شده:

چون $z_x = \frac{x-\bar{x}}{\sigma_x}$ و $z_y = \frac{y-\bar{y}}{\sigma_y}$ که در آن σ_x و σ_y انحراف معیار دو متغیر میباشد بنابراین میتوان نوشت

$$\rho(x,y) = Corr(x,y) = r = \frac{\sum z_x z_y}{n}$$

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

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

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

نکته:

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

2) همبســتگی بین دو متغیر تنها نشان دهنده ی این است که افزایش یا کاهش یک متغیر چه تاثیری بر افزایش یا کاهش متغیر دیگر دارد ولی این همبســتگی ضرورتا دال بر رابطه ی علّی بین متغیرها نمی باشد. به طور مثال اگر در یک تحقیق دو متغیر قد و تحصـیلات همبســتگی مثبت بالایی داشـته باشندنمی توانیم نتیجه بگیریم که افراد قد بلندتر دارای تحصیلات بیشتری هستند. بنابراین باید بین مفاهیم همبستگی و رابطه ی علّت و معلولی تفاوت قائل شـد. به بیان دیگر ممکن است دو متغیر همبستگی داشته باشند ولی لزومی ندارد که یکی از متغیرها علت و دیگری معلول باشـد، علاوه براین عوامل متعدد دیگری نیز می توانند بر ضـریب همبسـتگی اثرگذار باشند.

☜ ضریب همبستگی رتبه ای اسپیرمن

این ضـریب میزان همبســتگی رابطه ی میان دو متغیر ترتیبی را نشـان می دهد و به عبارت دیگر متناظر ناپارامتری ضـریب همبســتگی 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.

مثال در ادامه : ******

تفاوت ضریب همبستگی رتبه ای اسپیرمن و ضریب همبستگی پیرسون

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

سایر ضرائب همبستگی

ضریب همبستگی کندال

موریس گریگور کندال به سال 1930 به مطالعه در مورد این ضریب پرداخت. دقت کنید ضریب هماهنگی کندال با ضریب همبستگی تاو کندال تفاوت دارد. کندال در ضریب همبستگی کندال دارای خواصی نظیر ضریب همبستگی ساده است. برای برآورد آن از آماره τ استفاده می‌شود.

ضریب هماهنگی توافقی کندال

ضریب همبستگی کندال که با نماد w نشان داده می‌شود یک آزمون ناپارامتریک است و برای تعیین میزان هماهنگی میان نظرات استفاده می‌شود. ضریب کندال بین 0 و 1 متغیر است. اگر ضریب کندال صفر باشد یعنی عدم توافق کامل و اگر یک باشد یعنی توافق کامل وجود دارد. ویژگی‌های ضریب کندال یکی از مهمترین کاربردهای این آزمون را در مدیریت فراهم کرده است. برای پایان راندهای تکنیک دلفی می‌توان از ضریب هماهنگی کندال استفاده کرد.

ضریب همبستگی چوپروف *T* :

ضریب هبستگی چوپروف به منظور تعیین شدت وابستگی بین متغیرهای مورد مطالعه به کار گرفته می شود و مقدار آن همواره بین صفر ویک در نوسان می باشد زمانی از آن استفاده کرده که هر دو متغیر اسمی و یا یکی اسمی و دیگری ترتیبی باشد. اما نباید تعداد سطر و ستون با هم برابر باشند.یعنی در جدول توافقی 2در2 نمی توان از آن استفاده کرد. در چنین مواردی باید از ضریب فی استفاده کرد.

ضریب همبستگی فی:

به منظور بررسی شدت همبستگی بین دو متغیر اسمی که جدول توافقی 2 در 2 می باشد مورد استفاده قرار می گیرد.خی دو سطح معنی دار بودن همبستگی بین دو متغیر را تعیین میکند اما ضریب فی شدت همبستگی آنها را نشان می دهد. مقدار آن همواره بین صفر و یک در نوسان است.

ضریب کرامر:

این ضریب برای تغیین میزان شدت همبستگی بین دو متغیر اسمی مورد استفاده قرارمی گیرد و آن را با (V2) نشان می دهند و مقدار آن نیز همواره بین صفر ویک در نوسان است.هم جدول توافقی بیشتر از 2 در 2 وهم برای مستطیلی بکار می رود

☜ ** قضیه حد مرکزی

1- اگر از یک جمعیت n عضوی با میانگین معلوم μ و واریانس معلوم σ^2 یک نمونه تصادفی n تایی $(n \succ 30)$ یکی یکی و با عمل جایگزینی انتخاب کنیم آنگاه میانگین نمونه ای یعنی $\bar{x}$ تقریبا دارای توزیع نرمال با میانگین $\mu_{\bar{x}} = \mu$ و واریانس $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ و انحراف معیار $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$ خواهد بود

و متغیر تصادفی $z = \frac{\bar{x} - \mu_{\bar{x}}}{\sqrt{\sigma^2 / n}}$ دارای توزیع نرمال استاندارد میباشد

2- اگر از یک جمعیت نرمال با واریانس مجهول یک نمونه تصادفی n تایی با n<30 انتخاب کنیم آنگاه متغیر تصادفی

$t = \frac{\bar{x} - \mu}{\sqrt{s^2 / n}}$ دارای توزیع t با n-1 درجه آزادی است که در آن $\bar{x}$ میانگین نمونه و μ میانگین جمعیت و S انحراف معیار نمونه میباشد.

3- اگر در یک جمعیت دوجمله ای با احتمال پیروزی p یک نمونه تصادفی n تایی $(n \succ 30)$ یکی یکی انتخاب کنیم (در دو جمله ای همیشه با عمل جایگزینی است) آنگاه احتمال پیروزی در نمونه یعنی $\hat{p}$ تقریبا دارای توزیع نرمال با میانگین $\mu_{\hat{p}} = E(\hat{p}) = p$ و واریانس $\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}$ خواهد بود

نکته : چون یک توزیع پیوسته به گسسته تقریب زده میشود در موارد کوچکتر نصف واحد یعنی $\frac{1}{2n}$ اضافه و در موارد بزرگتر نصف واحد یعنی $\frac{1}{2n}$ اضافه کم میشود

نکته : آزمایش برنولی مستقل از هم است یعنی با جایگزینی است

☜ حالت دوم توزیع نرمال :

تعداد جمعیت نا مشخص – میانگین جمعت مشخص μ – انحراف معیار جمعیت مشخص σ – تعداد نمونه مشخص n - یک سوال احتمال m از نمونه

$$\bar{x} \sim N(\mu, \sigma^2 / n) \qquad p(\bar{x} < m) = p(\frac{\bar{x} - \mu_{\bar{x}}}{\sqrt{\sigma^2 / n}} < \frac{\bar{x} - \mu_{\bar{x}}}{\sqrt{\sigma^2 / n}}) \qquad p(\bar{x} < m) = p(z < \frac{m - \mu_{\bar{x}}}{\sqrt{\sigma^2 / n}})$$

** 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

1- 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

2- 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

برنولی : در مسئله احتمال نمونه p و میانگین ادعا p_0 و یک عدد بعنوان α در نظر گرفته سپس از جدول Z مربوط به α پیدا کرده و یک Z از $\frac{p-p_0}{\sqrt{p_0(1-p_0)/n}}$ در جدول بدست آورده و آنگاه این دو Z طبق شرایط مسدله و شرایط آزمون در فوق مقایسه میکنیم

مثال 15)

در آزمایش طول قامت 50 کودک میانگین 89.2 سانتیمتر و انحراف معیار 15.5 سانتیمتر میباشد آیا میتوان ادعا کرد طول قامت کودکان کمتر از 90 سانتیمتر است (هر وقت آلفا مشخص ندادند مقدارش 0.05 در نظر میگیریم)

$$n = 50 \qquad \bar{x} = 89.2 \qquad s = 15.5$$

$$\mathrm{H_a}: \ \mu < 90$$

$$\mathrm{H_0}: \mu = 90$$

$$\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{15.5^2}{50}} \qquad\qquad Z = \frac{\bar{x}-\mu_0}{\sigma_{\bar{x}}} = \frac{89.2-90}{\sqrt{\frac{15.5^2}{50}}} = -0.365$$

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

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

$$(Z = -0.365) < (-Z_\alpha = -1.645) \ \text{غلط}$$

چون $Z < -Z_\alpha$ نیست اعلام میکنیم دلیلی بر رد H_0 نداریم یعنی نمیتوان فرض H_α را پذیرفت پس نمیتوان گفت قامت کوچکتر از 90 است

مثال 16)

لامپهای ساخت کارخانه ای با انحراف معیار 142 ساعت در یک نمونه 100 تایی میانگین عمرشان 1280 ساعت شد صاحب کارخانه ادعا میکند که با همین اطلاعات میانگین عمر لامپهای ای کارخانه از 1200 ساعت بیشتر است آیا میتوان ادعای او را قبول یا رد کرد ؟ (هر وقت آلفا مشخص ندادند مقدارش 0.05 در نظر میگیریم)

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

$$\mathrm{H_a}: \ \mu > 1200$$

$$\mathrm{H_0}: \mu = 1200$$

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

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

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

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

مثال 17)

صاحب یک کارخانه داروسازی ادعا میکند که داروی ضد حساسیت این کارخانه در مورد بیش از 90٪ حساس در مدت 8 ساعت پس از مصرف دارو نتیجه مثبت میدهد

یک نمونه 200 نفری آدمهای حساس انتخاب کردیم و دارو را استفاده نمودند و 160 نفر بهبود یافتند آیا این داده ها دلیل کافی ارائه میدهد که ادعای صاحب کارخانه غلط است ؟ ($\alpha = 0.01$)

توجه شود که چون صورت مسئله خلاف نظر صاحب کارخانه خواسته علامت کوچکتر استفاده میشود

$$n = 200 \qquad \hat{p} = 160/200 = 0.8 \qquad s = 200 * 0.8 * (1 - 0.8)$$

$$\mathrm{H_a}: \ \mu < 0.9$$

$$\mathrm{H_0}: \mu = 0.9$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1 - p_0)}{n}} = \sqrt{\frac{0.9(1 - 0.9)}{200}} \qquad\qquad Z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.8 - 0.9}{\sqrt{\frac{0.9(1 - 0.9)}{200}}} = -4.72$$

$$Z_\alpha = Z_{0.01} = m \Leftrightarrow p(z \le m) = 1 - 0.01 = 0.99$$

$$\text{from table} \Rightarrow m = 2.33 = Z_{0.01} = Z_\alpha$$

$(Z = -4.72) < (-Z_\alpha = -2.33)$ 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 غلط باشد

مثال 18)

1. در یک تابع توزیع نرمال با میانگین 10 و واریانس 16 $x \sim N(10,16)$ مطلوبست محاسبه $P(8 \leq x \leq 15)$

$$P(x \leq 15) = p\left(\frac{x-u}{\sigma} \leq \frac{15-10}{4}\right) = P(Z \leq 1.25) = 0.894$$

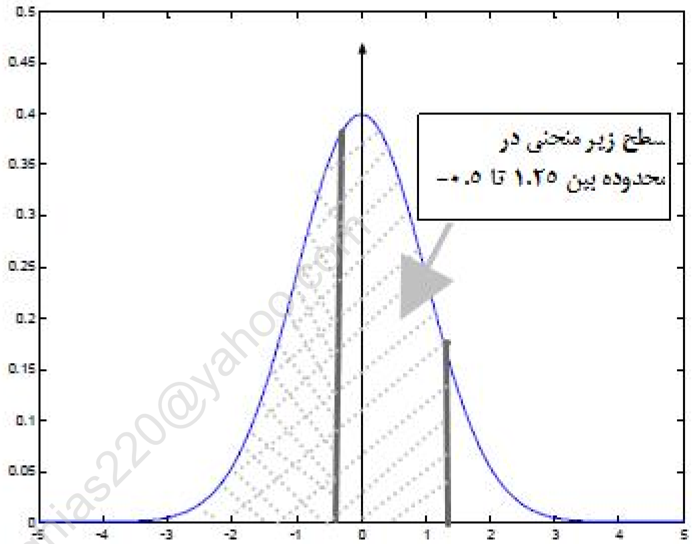
$$P(x \geq 8) = p\left(\frac{x-u}{\sigma} \geq \frac{8-10}{4}\right) = P(Z \geq -0.5) = 1 - P(Z \leq -0.5) = 1 - 0.308 = 0.692$$

$$P(x \leq 8) = p\left(\frac{x-u}{\sigma} \leq \frac{8-10}{4}\right) = P(Z \leq -0.5) = 0.308$$

با توجه به سطح زیر منحنی

$$P(8 \leq x \leq 15) = P(x \leq 15) - P(x \leq 8)$$

$$P(8 \leq x \leq 15) = 0.894 - 0.308 = 0.586$$



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)**

مشـــخصـــات تکمیل کنندگان فرم که این اطلاعات متغیرهای کمکی CoVariance برای طبقه بندی و مقایســـه میباشـد. - 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 اســـتفاده از روش مقیاس گذاری - مثلا 10 ســـوال با طیف پنج تایی لینکرت (کاملا موافق – موافق – بی نظر – مخالف – کاملا مخالف که عددی و بسته است)

انواع مقیاس‌ها

☜ **متغیرهای کیفی**

1. *مقیاس اسمی* (NOMINAL)

با کد گذاری عددی – اعداد نسبت داده شده ارجحیت و برتری ندارند (مثل کد شهرها - مثل جنسیت - زن 2 مرد 1) (مثل آبی 1 – زرد 2 – قرمز 3 – نارنجی 4 - …)

2. *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 موافق3 – بی نظر2 – مخالف1 یا تحصیلات مثلا دیپلم 1 فوق دیپلم 2 لیسانس 3 و فوق لیسانس 4 (که 4 بیشتر از 1 ولی 4 چهار برابر 1 نیست) (مثل ناراضی 1 – بی نظر 2 – راضی 3)

☜ **متغیرهای کمی**

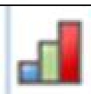
3. *مقیاس نسبتی*(SCALE) *(وزنی)*

با کد گذاری عددی – اعداد نسبت داده شده ارجحیت دارند و تناسب هم دارند (مثل سن و مثل وزن - مثلا وزن احمد 35 و وزن عباس 70 کیلوگرم یعنی عباس بیشتر از احمد است و به نسبت دو برابر هم میباشد)

4. *مقیاس فاصله‌ای*

مثلا به یک بیمار بگوییم اگر حداکثر درد مثل یک خط کش عدد 100 باشد درد شما چقدر است و بیمار بگوید 65

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

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

1. انواع مقیاس های مقایسه‌ای

1.1.مقیاس مجموع ثابت:

در این مقیاس‌گذاری پاسخ‌دهنده موظف است از میان مجموعه ثابتی از گزینه‌ها براساس برخی معیارهای خاص، مجموع ثابتی از واحدها نظیر امتیاز یا پول را تخصیص بدهد. مثلاً ممکن است از پاسخ‌دهندگان بخواهند 100 امتیاز را به پنج ویژگی اختصاص دهند. نتایج این نوع مقیاس معمولاً به‌صورت درصدی عنوان می‌شود.

1.2.آرایش رتبه‌ای:

در این روش گزینه‌های مختلف همزمان به پاسخ‌دهنده ارائه و از او خواسته می‌شود می‌خواهد گزینه‌ها را براساس برخی معیارها رتبه‌بندی کند. به‌طور مثال 10 کتاب به خواننده داده شده و از وی خواسته می‌شود به کتابی که علاقه بیشتری به آن دارد رتبه یک و به کتابی که اصلاً به آن علاقه ندارد نمره 10 دهد. این روش برای اندازه‌گیری اولویت کتابها و نیز ویژگی‌های کتاب به‌کار می‌رود.

1.3.مقایسه جفتی:

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

2. انواع مقیاس های غیرمقایسه‌ای

2.1.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

2.2.افتراق معنایی:

مقیاس افتراق معنایی، در مقایسه با مقیاس پنج امتیازی لیکرت، مقیاسی هفت امتیازی و دو قطبی است، در حالی که در مقیاس لیکرت هر عدد آیتم متعلق به مقیاس مشخص شده است در مقیاس افتراق معنایی، نقطه‌های ابتدا و انتها کاملاً مشخص شده‌اند. مثلاً می‌توان از راضی و ناراضی به‌عنوان نقاط ابتدا و انتها استفاده کرد.

2.3.مقیاس استاپل:

مقیاس استاپل از معیاری واحد در میانه بازه مقادیر اعداد زوج از 5- تا 5+ و بدون نقطه صفر تشکیل شده است. معمولاً این مقیاس به‌صورت عمودی است و از پاسخگویان خواسته می‌شود تا عددی خاص را انتخاب کنند که شیء یا موضوع محرک مربوط به معیار از پیش تعیین شده را توصیف می‌کند. مزیت مقیاس استاپل این است که این مقیاس مانند مقیاس افتراق معنایی برای دستیابی به ویژگی دو قطبی بودن به هیچ عبارتی نیاز ندارد.

2.4.مقیاس رتبه‌ای پیوسته:

این مقیاس این امکان را برای پاسخ‌دهنده فراهم می‌آورد که به جای اینکه از میان پاسخ‌های از پیش تعیین شده پاسخ مورد نظر را انتخاب کند، در هر نقطه از یک محور علامتی بگذارد. لازم به ذکر است این یادداشت‌ها جهت آشنایی مخاطبان با تحقیقات بازاریابی است و زمان استفاده از مقیاس‌های مورد نظر نیاز به مهارت و تجربه دارد. در یادداشت‌های آتی به بررسی روش‌های طراحی پرسشنامه پرداخته خواهد شد.

2.5.طیف بوگاردوس

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

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

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

محاسبه کوواریانس

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

مثلا میزان سفره هآب زیر زمین در سالهای مختلف

سال	1394	1395	1396	1397	1398
میزان آب	12	11	8	7	6

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

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

	A	B	C	D	E	F	G	H	I	G
سهام	300	100	200	150	400	250	500	450	50	120
سود	30	8	12	11	40	13	51	45	4	14

در جدول فوق با افزایش سهام میزان سود بیشتر میشود کوواریانس مثبت میشود

با افزایش یا کاهش یکی، دیگری هم به همان ترتیب افزایش یا کاهش خواهد داشت.

پس کوواریانس نشان دهنده شدت رابطه بین اعداد دو متغیر میباشد

کوواریانس به شکل زیر محاسبه می‌شود.

$Cov(x,y) = \frac{1}{n}\sum(x-\bar{x})(y-\overline{y})$ یا $Cov(y,x) = \frac{1}{n}\sum(x-\bar{x})(y-\overline{y})$

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

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

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

1) ضریب همبستگی پیرسون

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

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

ضریب همبستگی فرمول فوق واحد ندارد عددی بین منفی یک و یک است

ضریب همبستگی پیرسون بین -1 و 1 تغییر می کند.اگر این ضریب برابر یک باشد $r = 1$ بیانگر رابطه ی مستقیم کامل بین دو متغیر است ، رابطه ی مستقیم یا مثبت به این معناست که اگر یکی از متغیرها افزایش

(کاهش) یابد، دیگری نیز افزایش (کاهش) می یابد. مانند رابطه ی بین میزان ساعات مطالعه در روز و معدل محصلین.

اگر ضریب همبستگی منفی یک باشد $r = -1$ نیز وجود یک رابطه ی معکوس کامل بین دو متغیر را نشان میدهد.
رابطه ی معکوس یا منفی نشان می دهد که اگر یک متغیر افزایش یابد متغیردیگر کاهش می یابد و بالعکس.
زمانی که ضریب همبستگی برابر صفر است نشان می دهد که بین دو متغیر رابطه ی خطی وجود ندارد.

- ✓ 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

مثال 19)

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

اعتیاد به اینترنت	11	13	14	13	15	15	14	10	X
نمره	20	17	15	18	16	12	14	20	Y

حل: به روش پیرسون

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

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

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

نمره *اعتیاد	220	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 0.88 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 **- 0.88 = Pearson Correlation**

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

2) ضریب همبستگی اسپیرمن

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

که D اختلاف رتبه های بین دو گروه و n تعداد میباشد.

مثال 20)

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

نمره درس آمار در دانشگاه	نمره ورزش در دبیرستان
10	20
8	15
12	12
14	10
4	15
10	15
5	16

برای محاسبه ضریب اسپیرمن باید نمره ها را رتبه‌بندی کنیم بالاترین نمره رتبه ۱ میشود و بترتیب ادامه میدهیم (مثلا در آمار نمره ۱۴ رتبه 1 میشود و نمره 12 رتبه 2 میشود و نمره 10 رتبه 3 و نمره 10 بعدی رتبه 4 میشود در اینجا باید عدد این دو رتبه با هم جمع تقسیم بر دو نماییم جمع 3 و 4 تقسیم بر 2 که 3.5 میشود و به هر دو رتبه 3.5 میدهیم و به هیچکس رتبه3 و 4 تخصیص نمیدهیم و نمره بعدی که 8 است رتبه 5 میشود و ....) (
مثلا در درس ورزش نمره 20 رتبه اول میشود و نمره 16 رتبه دوم میشود و نمره 15 سه نفر داریم که رتبه 3 و 4 و 5 میشوند سه عدد 3 و 4 و 5 با هم جمع و تقسیم بر 3 مینماییم که جواب 4 میشود و به هر سه نفر رتبه 4 میدهیم و از این ببعد رتبه 3 و 4 و 5 را به هیچ شخصی اختصاص نمیدهیم – نمره بعدی 12 میباشد که رتبه اش 6 میشود و ...)

A	Ra	rA	M	Rm	rM	D=rA-rM	D^2
10	3	3.5	20	1	1	2.5	6.25
8	5	5	15	3	4	1	1
12	2	2	12	6	6	-4	16
14	1	1	10	7	7	-6	36
4	7	7	15	4	4	3	9
10	4	3.5	15	5	4	-0.5	0.25
5	6	6	16	2	2	4	16

حال اعداد را در فرمول جایگذاری می کنیم:

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

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

مثال 21)

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

x	10	14	15	15	13	14	13	11	اعتیاد به اینترنت
Y	20	14	12	16	18	15	17	20	نمره

حل: به روش اسپیرمن

X	10	14	15	15	13	14	13	11
Rx	1	5	8	7	4	6	3	2
rX	1	5.5	7.5	7.5	3.5	5.5	3.5	2
Y	20	14	12	16	18	15	17	20
Ry	7	2	1	4	6	3	5	8
rY	7.5	2	1	4	6	3	5	7.5
D=rX-rY	-6.5	3.5	6.5	3.5	-2.5	2.5	-1.5	-5.5
D^2	42.25	12.25	42.25	12.25	6.25	6.25	2.25	30.25

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

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

مثال 22)

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

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

آزمون فرض‌ها : میخواهیم فرض های زیر را تحلیل کنیم

H0: μ1= μ2= μ3= μ4

H1:μi≠μj

فرض H0 چهار میانگین با هم مساوی هستند.

فرض H1 حداقل بین دو میانگین اختلاف معنی دار وجود دارد.

3) ضریب همبستگی رتبه‌ای کندال (Kendall coefficient)

ضـریب همبســتگی کندال برای داده‌های رتبه بندی شــده زمانی‌که تعداد آزمودنی‌ها از ۱۰ نفر کمتر اســت، به کار می‌رود. کندال از تعداد توافق‌ها و عدم توافق‌ها در رتبه بندی‌ها برای محاسبه ضریب خود استفاده می‌کند. برای مثال فرض کنید اسـتاد روانشناسی بالینی ۱۰ نفر از دانشجویان خود را براساس دانش روانشناسی و شایستگی آن ها برای شـاغل شـدن در این رشته رتبه بندی کرده است. برای بررسی رابطه بین دانش روانشناسی و شایستگی می‌توانیم از ضریب همبستگی کندال استفاده کنیم.

ضـریب همبســتگی رتبه‌ای کندال به tau-a کندال نیز معروف اسـت، متقارن می‌باشد و مقدار آن بین ۱+ و ۱- قرار دارد و مشابه با ضریب همبستگی پیرسون تفسیر می‌شود.

فرمول محاسبه ضریب همبستگی رتبه‌ای کندال (tau-a)

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

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

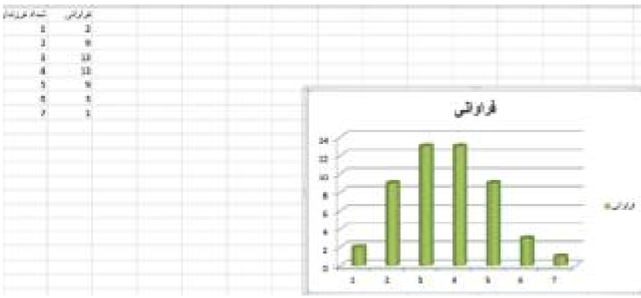
Q: تعداد عدم توافق‌ها یعنی تعداد اعداد (مواردی) که رتبه پایین‌تری برای هر رتبه متغیر دوم به دست آورده‌اند.

N: تعداد آزمودنی‌ها (حجم نمونه)

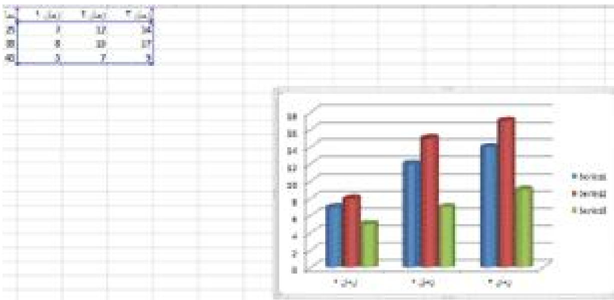
انواع نمودارها

1)) نمودار میله ای Bar Chart:

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

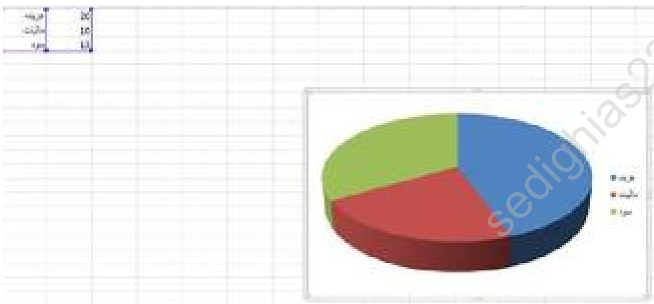


میله ای – مشاهده فراوانی و مقایسه داده ها بصورت دسته بندی شده



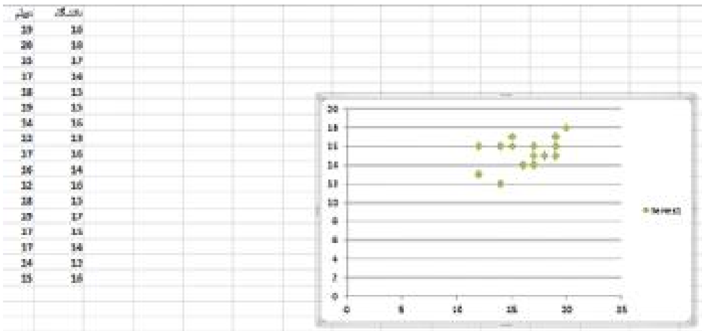
2)) نمودار دایره ای Pie Chart :

برای مشاهده سهم هر مورد از داده ها



3)) نمودار هیستوگرام Histogram:

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



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

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

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)» است. در این روش سعی بر این است که اختلاف بین چند جامعه آماری، ارزیابی شود. با توجه به پراکندگی کل داده‌ها، تجزیه واریانس بین گروه‌های مختلف در این روش امکان‌پذیر است. به این ترتیب می‌توان برابر بودن میانگین را بین گروه‌های مختلف آزمود. همچنین در مدل‌های رگرسیونی با تجزیه واریانس کل به واریانس مدل و واریانس خطا تشخیص مناسب بودن مدل قابل ارزیابی است.

آنالیز کوواریانس ANCOVA

آزمون تحلیل واریانس یا 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 Analyze of Covariance ANCOVA یا آنکوا، نوعی آنالیز و تحلیل همانند آنوا می‌باشد. هرگاه در آنالیز واریانس بخواهیم اثر کمیت‌های مداخله گر را به روش‌های آماری حذف کنیم تا نتایج با دقت بیشتری به دست آید از آنالیز کوواریانس استفاده می‌شود. در این روش هم از کنترل آماری و هم از واریانس استفاده می‌شود. 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 می‌باشد 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) است و در مواردی کاربرد دارد که بیش از یک کمیت مستقل وجود داشته باشد و همچنین در مواردی که کمیت‌های وابسته به آسانی نتوانند ترکیب شوند به کار می‌رود MANCOVA همان MANOVA با این تفاوت که به شما امکان می‌دهد اثرات اضافی کمیت‌های مستقل را نیز کنترل کنید.
اگر چندین کوواریانس وجود داشته باشد مانکوا (MANCOVA) به جای مانوا (آنالیز واریانس چند متغیره) به کار برده می‌شود.

انتخاب آزمون

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

انتخاب آزمون یک متغیره و گروهها

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

انتخاب آزمون برای دو سری متغیر (بشکل مختصر)

متغیر اول	متغیر دوم	نوع آزمون
کیفی(اسمی)	کیفی(اسمی)	کای دو
پیوسته(رتبه ای)(ترتیبی)	گسسته(کمی)(فاصله ای – 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 مستقلx	متغیر دوم = متغیر وابستهy	
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

آشنایی مقدماتی با نرم افزار SPSS

Statistical Package for the Social Sciences

نرم افزار را از دانشـگاه 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

==================== 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 1822-1895

============================ ============================

☜ سرفصل آمار توصیفی

1) تعریف مفاهیم (ریاضی آمار اندازه گیری سنجش)
2) طبقه بندی (آمار توصیفی و استنباطی)
3) جامعه نمونه پارامتر مشخصه آماری
4) مفهوم متغیر و انواع آن
5) طبقه بندی متغیر از لحاظ مقیاس اندازه گیری
6) انواع نمودار و کاربرد آن در آمار
7) طبقه بندی آمار توصیفی (یک متغیری چند متغیری)
8) شاخص های مرکزی (مد میانه میانگین)
9) شاخص های پراکندگی (واریانس انحراف معیار انحراف چارکی و دامنه تغییرات)
10) نقاط درصدی و رتبه بندی
11) منحنی طبیعی و کاربرد آن در آمار
12) نمره های استاندارد و انواع آن
13) شاخص های همبستگی و انواع آن با تکیه بر همبستگی پیرسون و اسپرمن
14) رگرسیون و پیش بینی
15) آشنایی با نرم افزارهای آماری از جمله SPSS