بسم الله الرحمن الرحیم

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

عنوان درس :

آمار مقـــدماتي

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

فصل اول

مفاهيم اساسي علم آمار

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

مقدمه :

واژه **statistics** كه به فارسي آن را آمار ترجمه كرده اند از كلمه لاتين**status** يا كلمه ايتاليايي**statista** ويا كلمه آلماني **statistik** كه همگي به معني دولت هستند استخراج شده است.

آمار را به معني روشهايي براي جمع آوري ،تنظيم و تجزيه و تحليل اطلاعات عددي درباره موضوعي تعريف نموده اند.

www.bookgolden.com

مفهوم آمار توصيفي و آمار استنباطي

توصيفي: روش‌هاي تنظيم ، خلاصه سازي ، رسم نمودار ،

آمار: جدول و پردازش اطلاعات

استنباطي : روش‌هاي جمع آوري ، تجزيه و تحليل اطلاعات ،

تفسير و تعميم اطلاعات

موضوع علم آمار را مي توان از چند ديدگاه مختلف تقسيم بندي كرد . آمار در نخستين مرحله به دو بخش توصـــيفي و امار استنباطي تفكيك مي‌شود .

وقتي با مجموعه اي از اطلاعات روبرو هستيم ابتدا بايد تصـميم بگيريم كه چگونه اين اطلاعات را خلاصه كنيم تا بهتر بتوانيم از آنها استفاده كنيم.

www.bookgolden.com

در واقع آمار توصيفي بخشي از علم آمار است كه به جمع آوري ، خلاصه كردن،نمايش و پردازش اطلاعات مي پردازد.

دراين روش پس از جمع‌آوري اطلاعات،تهيه جدولهاي خلاصه و ر
نمودارهاي اماري به يافته هاي مناسب براي مطالعه تحقيق مي رس

www.bookgolden.com

گرچه آمار توصيفي بخش مهم از آمار است و هنوز هم استـفاده گسترده اي از آن مخصوصاً براي توصيف ويژگيهاي مهم داده‌ها براي به عمل مي آيد

اما در سالهاي اخير توجه بيشتربه آمار استنباطي ، يعني روشهايي اس
كه به كمك آنهامي توان اطلاعات موجوددر مجموعه‌اي محـــدود از
داده‌ها را به مجموعه‌اي بزرگتر كه داده‌ها از آن بدست آمده‌اند تعميم

www.bookgolden.com

مفاهيم بنيادي آمار

انجام پژوهشهاي آماري مستلزم مشخص شدن برخي مفاهيم اساسي مي باشد كه در اين قسمت مطالعه و تشريح اين مفاهيم اوليه و اساسي درآمار مي پردازيم كه عبارتند از:

جمعيت ، نمونه ، متغير و داده

www.bookgolden.com

جمعيت :

مجموعه اي از افراد يااشياء كه مي‌خواهيم يك‌ياچند نفرخصوصيت را در مورد آنها مطالعه كنيم را جمعيت يا جامعه آماري مي نامند.

مثلاً جمعيت دانشجويان دانشگاه پيام نور ،از لحاظ سن،جمعيت كارگران يك كارخانه از نظرميزان دستمزديا جمعيت نوزداني كه سال گذشته بدنيا آمده‌اند از نظر مصرف شير خشك.

www.bookgolden.com

براي انجام هر كارآماري بايد آن جمعيت وخصوصيت مورد مطالعه بدون هر گونه انجام قبلاً مشخص مي شوند .
جمعيت دو نوع است جمعيت متناهي كه تعداد اعضاء تشكيل دهنده معين و محـــدود است

به طوري كه اگر آنها راتك تك شناسايي كنيم و كنار بگذاريم تمام مي‌شوندمانند جمـعيت اتومبيلها شـــهرتهران و جـمعيت نامتناهي عناصرتشكيل دهنده آن نامعين و پايان ناپذير مي باشند.

مانندجمعيت ستارگان آسمان يا جمعيت دانه‌هاي يك مزرعه گندم كه هريك از عناصرجمعيت را يك فرد مي نامند و واژه فرد تنها به انسان اطلاق نمي شود.

www.bookgolden.com

تعداد عناصر جمعيت را اندازه جمعيت گويند و آن را با N نمايش مي دهند.

مطالعه‌تك‌تك افراد جمعيت به علت هزينه زياد، كمي وقت ونداشتن امكانات كافي ،اغلب مقدور نمي باشد .

بنابراين قسمتي از جمعيت را به جاي تمام آن براي مطالعه در نظر مي گيرند.

نمونه :

نمونه بخشي از جامعه آماري است كه طبق اصول و ضوابطي معين انتخاب مي شود و مطالعه نمونه به جاي مطالعه،كل جمعيت توصيه مي شود .

www.bookgolden.com

تعداد عناصر نمونه را اندازه نمونه گويند و با n نمايش مي‌دهند.

انتخاب نمونه طبق روشها ، اصول و ضوابطي معين‌که‌به روشهاي

نمونه‌گيري معروفند صورت مي‌پذيرد .

متغير :

خصوصيت يا صفتي‌مانند گروه خوني،

هوش،قد،وزن،محل‌سکونت،

طول عمر و ... که از هر فرد به فرد ديگر و از هر شخص به شخص ديگر در جمعيت آماري تغييرمي‌کند

رامتغيرمي‌نامند و اغلب باx

نشان داده مي شود .

www.bookgolden.com

متغیرها

الف) متغیرهاي كمي: صفاتي‌هستند كه قابل اندازه گيري يا شمارش هستند مانند سن طول عمر، وزن،درآمد، هزينه،و...

ب) متغيرهاي كيفي :صفاتي هستند كه واحد نداشت و قابل اندازه گيري نيستندمانند جنس، مرغوبيت، مهارت،گروه خوني و...

www.bookgolden.com

همچنين متغيرهارا براساس نقشي‌كه برعهده دارند مي توانند به دودسته زير تقسيم شوند.

الف) متغيرهاي مستقل : اين متغيرها تحت كنترل محقق بوده وتغييرات آنها، دستخوش تغييرات ساير متغيرها نباشد .

متغيرها

ب) متغيرهاي وابسته : تغييرات اين‌گونه‌متغيرهابسته‌به‌تغييرات متغير يا متغيرهاي مستقل مي باشد.

www.bookgolden.com

داده :

در مطالعه خصوصيتي در جمعيت،متغير مورد نظر را با مقياسي مناسب اندازه مي گيرند در اين صورت مجموعه اي از اعداد حاصل‌مي‌شود

كه به آنها داده مي‌گويند.در واقع اطلاعــات عددي درباره موضوعي (خصوصـــيت) را داده مي‌گويند . داده‌ها دو نوع اند .

داده ها

الف) **داده‌هاي‌گسسته**:داده هايي كه ماهيت اعشاري ندارند

مانندتعداد دانشجويان‌يك كلاس،تعدادمدارس‌يك‌شهرو....

ب) **داده هاي پيوسته** :داده هايي كه ماهيت اعشاري دارند

مانند وزن ،قد،دماي هوا،در آمدو...

www.bookgolden.com

پارامتر و آماره

مطالعه جامعه براي‌بدست‌آوردن برخي‌از شاخصهاست. اين شاخصها،چنانچه با اندازه گيري تمامي‌عناصرجامعه‌آماري (سرشماري)بدست‌آمده شده‌باشندآنهارا *پارامتر*و اگر با استفاده از نمونه كه بخشي از جامعه است بدست‌آمده باشند آنها را *آماره* مي‌گويند .

اگرمتوسطدرآمدكاركنان‌دولت‌با استفاده از اندازه‌گيري‌درآمدكليه كاركنان دولت محاسبه‌شودآن‌را پارامتر واگر با استفاده از اندازه‌گيري درآمد نمونه‌اي از كاركنان بدست آيدآن را آماره مي‌گويند .

www.bookgolden.com

مراحل پژوهشي علم آمار

الف) تعيين هـــدف

ب) جمع آوري داده‌ها

ج) تجزيه وتحليل داده‌ها

د) نتيجه گيري

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

اندازه گيري و مقياس سازي

اطلاق يك عدد حقيقي به متغير x ، طبق قاعده‌اي مشخص را اندازه گيري يا مقياس‌سازي x و اين قاعده را مقياس گويند. استـــيونس در مقاله‌اي (به سال1946ميلادي) چهار نوع مختلف از مقياســـهاي اندازه‌گيري را مشخص كرده است . اين مقياسها عبارتند از :

الف) مقياس اسمي :

هر گاه عدد مورد نظر تنها براي تمييز افراد يا اشياء بكار رود و نتوان آن‌را براي مقايسه و چهار عمل اصلي حساب بكار برد از مقياس اسمي استفاده مي‌شود.

www.bookgolden.com

ب) مقياس ترتيبي :

مقياس اسمي هر گاه صعودي باشد را مقياس ترتيبي مي گويند.

ج) مقياس فاصله اي :

وقتي كه مقياس همه خصوصيات يك مقياس ترتيبي را دارا باشد و علاوه بر آن فاصله بين هر دو عدد بر روي مقياس ميزان مشخصي داشته باشد.

درآن صورت به مقياسي دست يافته‌ايم به مراتب قوي‌تر كه آن را مقـ
فاصله‌ايمي نامنددراين مقياس نسبت بي‌معني است و نسبت فواصل
ثابت مي‌ماند دراندازه‌گيري با اين مقياس صفر به معناي هيچ نمي باشـ
وصرفاً جنبه قراردادي و اختياري دارد

www.bookgolden.com

د) مقياس نسبي :

به مقياسي كه علاوه بر همه خصوصيات مقياس فاصله‌اي داراي نقطه صفر واقعي نيز هست مقياس نسبي مي‌گويند در اين مقياس ، نسبت هر دو نقطه روي مقيـــــاس از واحد اندازه گيري مسقل است.

www.bookgolden.com

نماد جمعبندي (Σ)

در مطالعه آمار همه جا با عمل جمع كردن داده ها سر و كار داريم براي اجتناب از نوشتن مكرر علامت(+) شكل بزرگ حرف يوناني سيگما (Σ) را باعنوان نمادي براي عمل جمع كردن به كار مي بريم

انتشارات طلايي
پويندگان دانشگاه
راهنماي طلايي
تست طلايي
پيك طلايي
www.bookgolden.com

مثال :

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$$

$$\sum_{i=2}^{4} x_i = x_2 + x_3 + x_4$$

و

$$\sum_{i=3}^{5} (x_i - 5) = (x_3 - 5) + (x_4 - 5) + (x_5 - 5)$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

برخي از خواص اصلي $\sum$

الف)

$$\sum_{i=1}^{n} bx_i = b\sum_{i=1}^{n} x_i$$

ب)

$$\sum_{i=1}^{n} (bx_i + a) = b\sum_{i=1}^{n} x_i + na$$

ج)

$$\sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} - 2a\sum_{i=1}^{n} x_i + na^2$$

www.bookgolden.com

فصل دوّم

جدولهاي فراواني

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

جدول فراواني :

نمايش داده‌ها را با نظم خاص در چند سطر و ستون يك جدول آماري مي‌گويند يك جدول آماري بايد به گونه‌اي تنظيم شود كه به كمك آن بتوان تا حدودي پاره‌اي اطـــلاعات نهفته در داده‌ها را در آن مشاهده كرد .

www.bookgolden.com

مثال: فرض كنيد پزشكي20 بيمار قلبي دارد كه گروه خوني آنها به صورت زير است.

O,A,B,O,AB,A,A,A,O,O,O,A,A,B,B,AB,O,AB,AB,O

مي توان اطلاعات فوق را در جدول زير خلاصه نمود.

گروه خوني	فراواني (f_i)
A	6
B	3
AB	4

www.bookgolden.com

هر گاه n چيز از k نوع به ترتيب با تعداد $f_1 , f_2 , \dots , f_k$ تشكيل شده باشند

اين تعداد را فراواني ها و $\frac{f_1}{n}, \frac{f_2}{n}, \dots, \frac{f_k}{n}$ را فراواني نسبي گويند و با r_i نمايش مي دهند و

فراواني تجمعي

$$F_i = \sum_{j=1}^{i} f_j$$

فراواني نسبي تجمعي نمايش مي دهند.

$$R_i = \sum_{j=1}^{i} r_j = \frac{F_i}{n}$$

بنابراين مي‌توان جدول فراواني گروه خوني 20 بيمار قلبي را به صورت زيرتكميل نمود .

www.bookgolden.com

X:گروه خوني	f_i	r_i	F_i	R_i
A	6	$\frac{6}{20}$	6	$\frac{6}{20}$
B	3	$\frac{3}{20}$	9	$\frac{9}{20}$
AB	4	$\frac{4}{20}$	13	$\frac{13}{20}$
O	7	$\frac{7}{20}$	20	1

n =20 1

www.bookgolden.com

الگوريتم تنظيم جدولهاي فراواني

براي طبقه‌بندي‌داده‌ها الگوريتم غيرحقيقي در چهار مرحله ارائه مي‌گردد

مرحله اول : داده ها ي جمع آوري شده را از كوچك به بزرگ مرتب و دامنه تغييرات را محاسبه مي كنيم .

دامنه تغييرات = $R = x_{(n)} - x_{(1)} = Max x_i - Min x_i$

www.bookgolden.com

مرحله دوم: تعداد طبقات (k) را مشخص مي‌كنيم دو روش براي تعيين k پيشنهاد مي شود .

الف) فرمول زير توسط استورجس براي تعيين k بدست آورده شده‌است .

كه در آن n تعداد كل داده ها است. $k=[l+3/322\log n]$

ب) k كوچكترين عدد صحيحي است كه به ازاي آن براي اولين بار نامساوي غير اكيد $2^k \geq n$ بر قرار است .

www.bookgolden.com

مرحله سوم :

طول طبقات را از رابطه $\omega = \frac{R}{K}$ محاسبه مي كنيم

.

هر گاه در تعيين طول طبقات حاصل تقسيم دامنه تغييرات بر تعداد طبقات عدد اعشاري بدست آمد عدد مذكور را به نزديكترين واحد به بالا رند مي كنيم .

مرحله چهارم : سازماندهي داده ها را آغاز مي كنيم حد پايين طبقه اول را با كوچكترين مشاهده شروع كرده و حدود طبقات را با فاصله ω تشكيل مي‌هيم

www.bookgolden.com

مثال : مؤسسه‌اي يك نوع باطري براي‌اتومبيلهاي خود استفاده
مي‌كند از مسئول اين مؤسسه طول عمر 40 باطري مستعمل را بر
حسب سال سؤال كرده و داده‌هاي زير را بدست مي‌آوريم .

4/1 3/5 4/5 3/2 3/7 3 2/6 3/4 1/6 3/1 3/2 2/2
3/1 4/7 3/6 2/5 4/3 3/6 3/4 2/9 3/3 3/9 3/1 3/8

3/1 3/7 4/4 3/2 4/1 1/9 3/4 4/7 3/8 3/2 2/6 3/9 3 4/2 3/5 3/3

www.bookgolden.com

در این صورت n =40 و تعداد طبقات از رابطه k=[l+3/322log40]

محاسبه مي شود بنابراين k=[6/322] =6 بنابراين طول هر طب
عبارت است از :

$$\omega = \frac{R}{K} = \frac{\max(x_i) - \min(x_i)}{k} = \frac{4/7 - 1/6}{6} = 0/53 \approx 0/6$$

www.bookgolden.com

حدود واقعي طبقات	حدود طبقات	فراواني(f_i)	مركز (نماينده) طبقات (x_i)
1/55- 2/25	1/6- 2/2	**2**	**1/9**
2/25- 2/85	**2/2**- 2/8	**4**	**2/5**
2/85- 3/45	2/8- 3/4	13	3/1
3/45- 3/95	3/4- 4/0	13	3/7
3/95- 4/55	4/0- 4/6	6	4/3
4/55- 5/15	4/6- 5/2	2	4/9

www.bookgolden.com

فصل سوم

نمودارهاي آماري

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

نمودارهاي آماري

نمايش داده ها را،طبق قراردادهاي خاص به صورت هندسي ، يك نمودار آماري مي‌گويند.

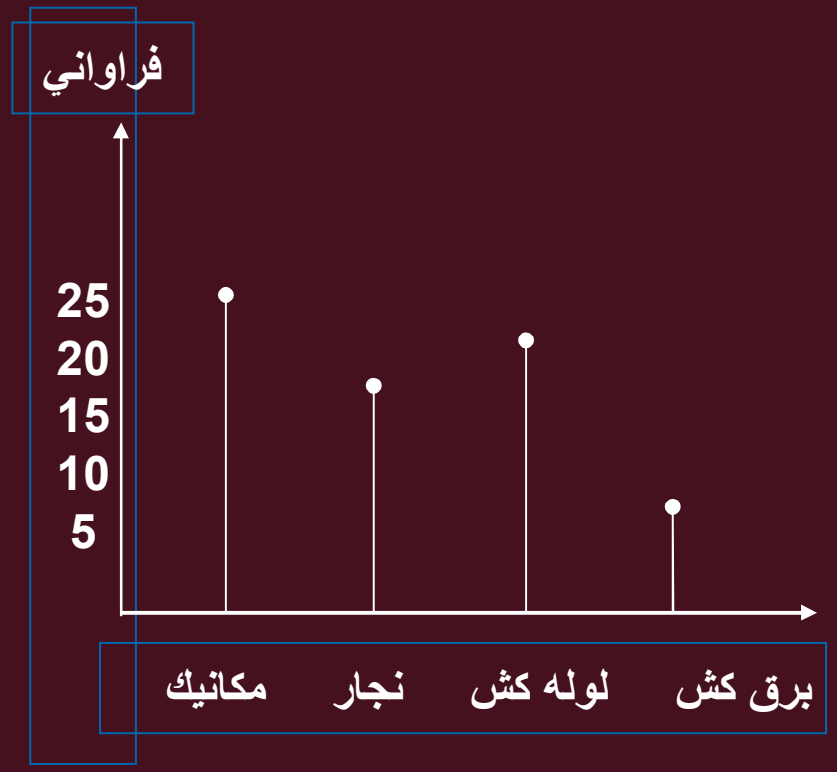
نمودار ميله اي

دراين نمودار صفت كيفي را به صورت نقاطي روي محور (xها)مشخص مي‌كنيم و سپس از نقاط حاصل،خطـهاي عمودي بر محور رسم مي‌كنيم كه ارتفاع آن برابر فراواني يا فراواني نسبي متغير مربوطه مي باشد.

www.bookgolden.com

مثال: توزيع فراواني كارگران يك كارخانه بر حسب تخصص داده شده است نمودارميله اي آن عبارت است از:

تخصيص	فراواني
مكانيك	25
نجار	16
لوله كش	20
برق كش	9



www.bookgolden.com

نمودار ستوني :

در نمودار ميله اي، هرگاه ميله ها را به صورت مستطيل نشان دهيم نمودار ستوني حاصل مي شود.

مثال: در توزيع فراواني تخصص كارگران‌كارخانه نمودار ستوني عبارت است از

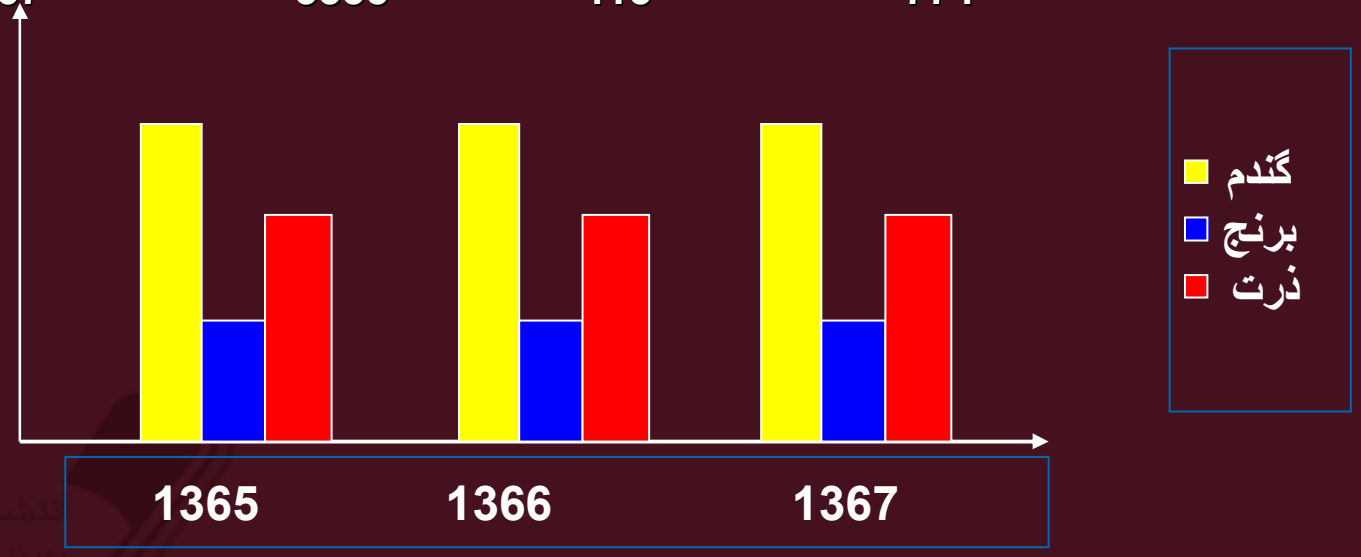
www.bookgolden.com

نمودار ميله اي چند تايي :

اگر داده ها را بر مبناي صفات كيفي طبقه بندي كينم و براي مقايسه آنها از دو يا چند صفت استفاده كنيم در آن صورت از نمودار ميله اي چند تايي استفاده مي كنيم .

مثال : مقدار محصول گندم ، برنج و ذرت بر حسب هزار تن در سالهاي 1365 و 1366 و1367 در يك استان به شرح زير است.

سال	گندم	برنج	ذرت
1365	1916	293	643
1366	2444	338	614
1367	3335	415	774

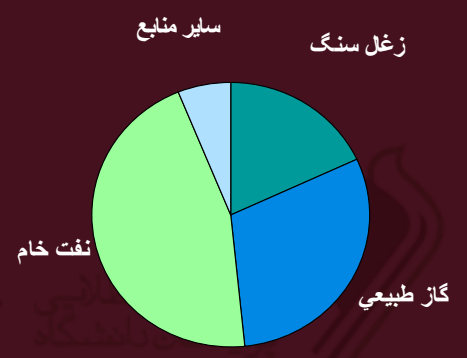


www.bookgolden.com

اين نمودار براي نمايش داده هاي كيفي مورد استفاده قرار مي‌گيرد دايره به قطاعهايي متناسب با فراواني نسبي هر طبقه تقيسم مي شود اندازه هر قطاع از رابطه‌ي $S_i = r_i \times 360^\circ$ برحسب درجه بدست مي‌آيد.

مثال: وضعيت مصرف انرژي بر حسب صد هزار B.T.U از هر نوع در سال 1354داده شده است نمودار دايره اي مربوط به تغييرات نوع انرژي عبارت است از :

سال	گاز طبيعي	زغال سنگ	نفت خام	ساير منابع	مجموع
1354	105	62	158	21	346



گاز طبيعي	$s = \frac{105}{346} \times 360 = 109$
زغال سنگ	$s = \frac{62}{346} \times 360 = 65$
نفت خام	$s = \frac{158}{346} \times 360 = 164$
ساير منابع	$s = \frac{21}{346} \times 360 = 22$

www.bookgolden.com

نمودار شاخه و برگ :

براي تعبيه نمودار شاخه و برگ ارقام مشاهدات را كه دست‌كم دو رقمي هستند را به دو بخش شاخه وبرگ تقسيم مي‌كنيم شاخه شامل يك يا چند رقم اوليه و برگ شامل ارقام باقيمانده ، وقتي‌كه مجموعه شاخه ها انتخاب شد اين اعداد در سمت چپ صفحه نوشته مي شود و در كنار هر شاخه ،تمام برگهاي متناظر با مقادير داده ها به ترتيب مشاهده ثبت مي شود.

مثال: نمودار شاخه و برگ تعداد دانشجويان 15 درس دانشكده عبارت است از :

شاخه	برگ
0	9
1	5
2	0 2 2 4 4 4 8
3	0 0 4 6 7
4	2

www.bookgolden.com

نمودار هيستوگرام :

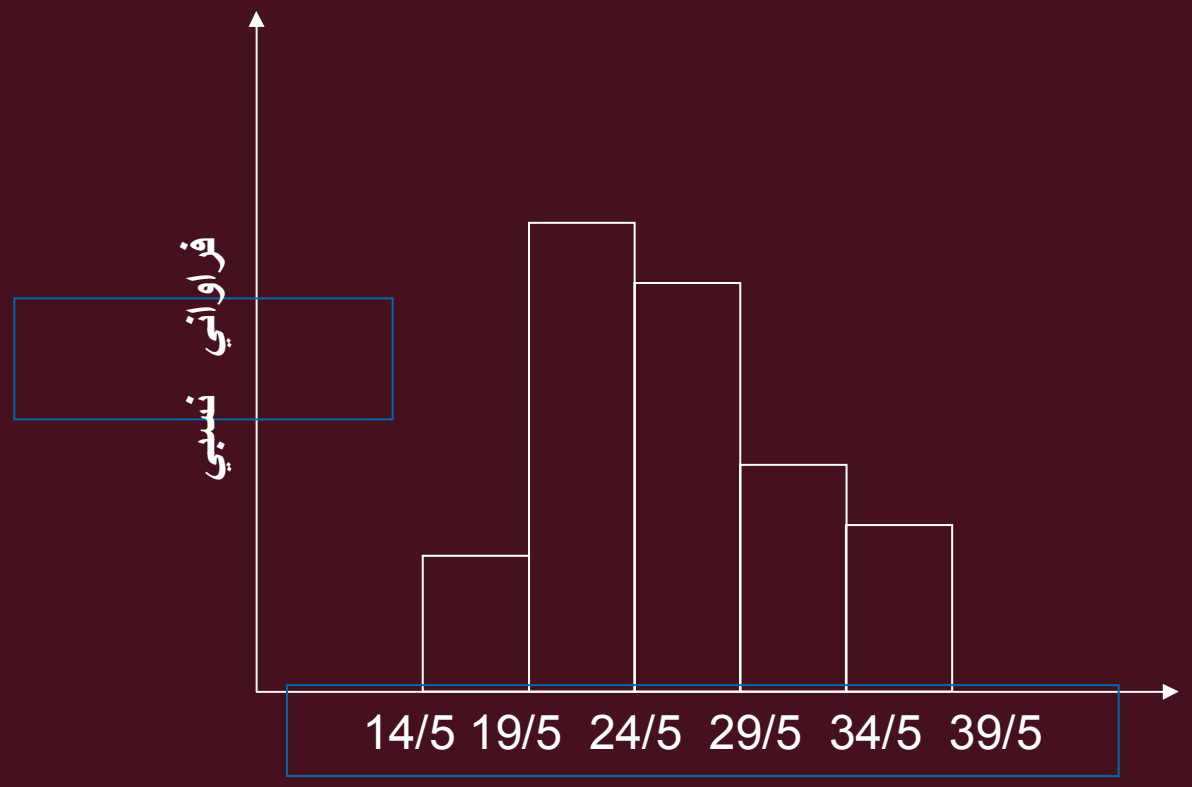
درتوزيعهاي فراواني كه داراي فاصله طبـقات مساوي هستند هيستوگرام نموداري است مركب ازمستطيل هايي كه عرض آن برابر فاصله طبقه و ارتفاع آن برابر فراواني يا فراواني نسبي هر طبقه است. در نمودار هيستوگرام ، محور افقي دستگاه مختصات با حدود طبقات و محور عمودي با فراواني يا فراواني نسبي مدرج مي‌گردد.

www.bookgolden.com

مثال : نمودار هیستوگرام توزیع سن ازدواج عبارت است از :

سن ازدواج	f_i	r_i
14 / 5 – 19 / 5	18	0 / 09
19 / 5 – 24 / 5	74	0 / 37
24 / 5 – 29 / 5	62	0 / 31
29 / 5 – 34 / 5	26	0 / 13
34 / 5 – 39 / 5	20	0 / 1
مجموع	n= 200	

www.bookgolden.com



انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

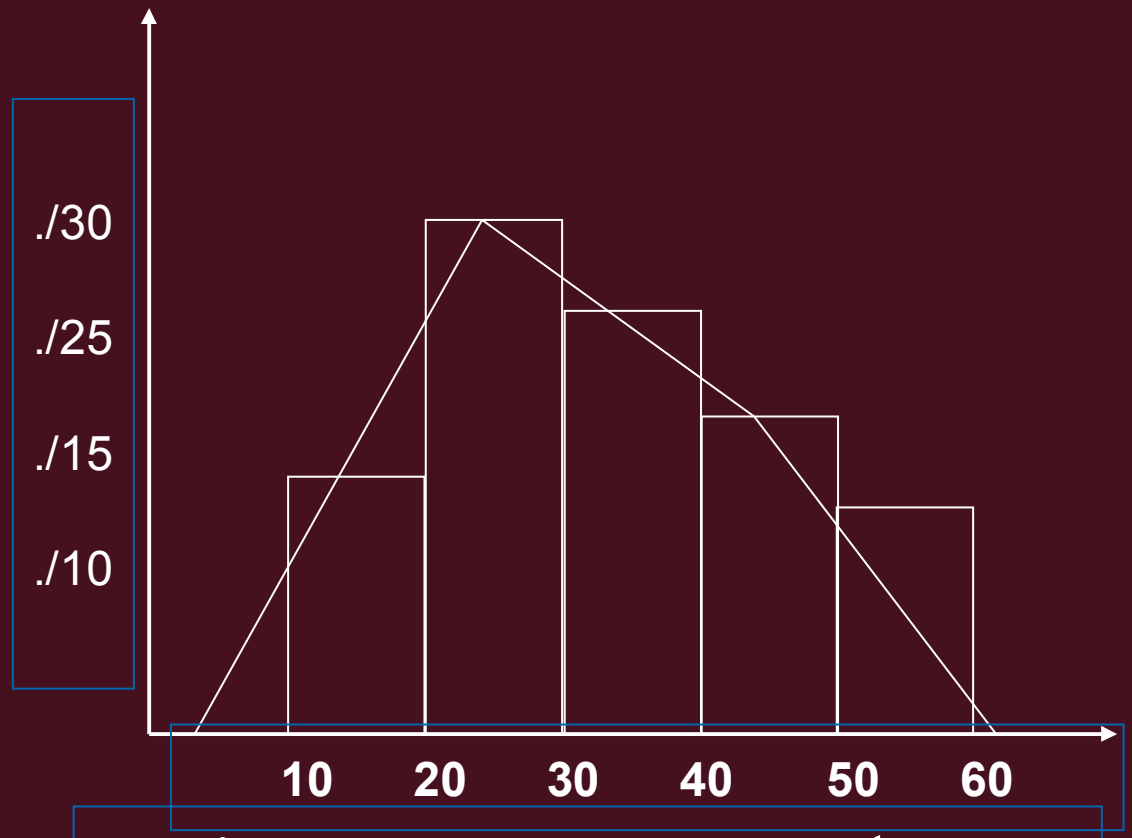
نمودار چند ضلعي (چند برفراواني)

اگر نقطه‌اي وسط قاعده هاي بالايي مستطيل‌هاي هيســـتوگرام ونقطـه‌هاي وسط رده‌هايي را كه بلافاصله در دوانتهاي هيستوگرام بوده و داراي‌فراواني صفر هستند به هم بپيونديم يك خط شكسته به دست مي آيد كه آن را چند بر يا چندضلعي فراواني مي گويند .

www.bookgolden.com

مثال:شرکت ایران دارو داراي 100 کارمند است که در امد ماهیانه آنها در جدول زیر آمده است.

در آمد	f_i
10 – 20	15
20 – 30	30
30 – 40	25
40– 50	20
50 – 60	10
مجموع	n = 100

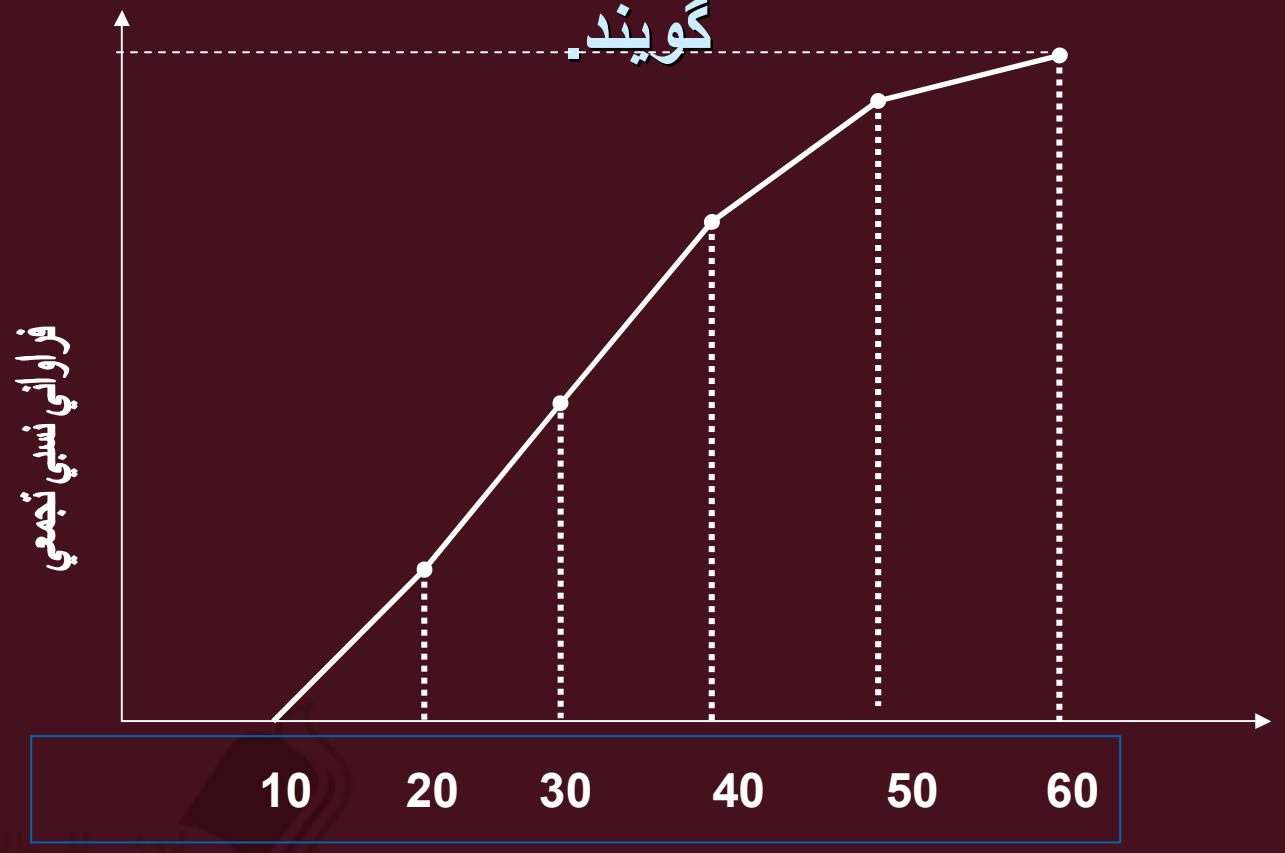


نمودار هیستوگرام و چند ضلعي مربوط به درآمد ماهیانه

www.bookgolden.com

نمودار توزيع نسبي تجمعي

اگر نقاطي را كه طول آنها مرز طبقه‌ها و عرض آنها فراواني نسبي تجمعي تا آن مرز باشد به هم بپيونديم يك خط شكسته بدست مي‌ايد كه آن را نمودار توزيع نسبي تجمعي يا چند ضلعي فراواني نسبي تجمعي گويند.



نمودار توزيع نسبي تجمعي مربوط به در آمد ماهيانه

www.bookgolden.com

منحني‌هاي فراواني و فراواني تجمعي

هرگاه تعداد داده‌ها (n) زياد باشد تعداد طبقات طبق رابطه‌ي

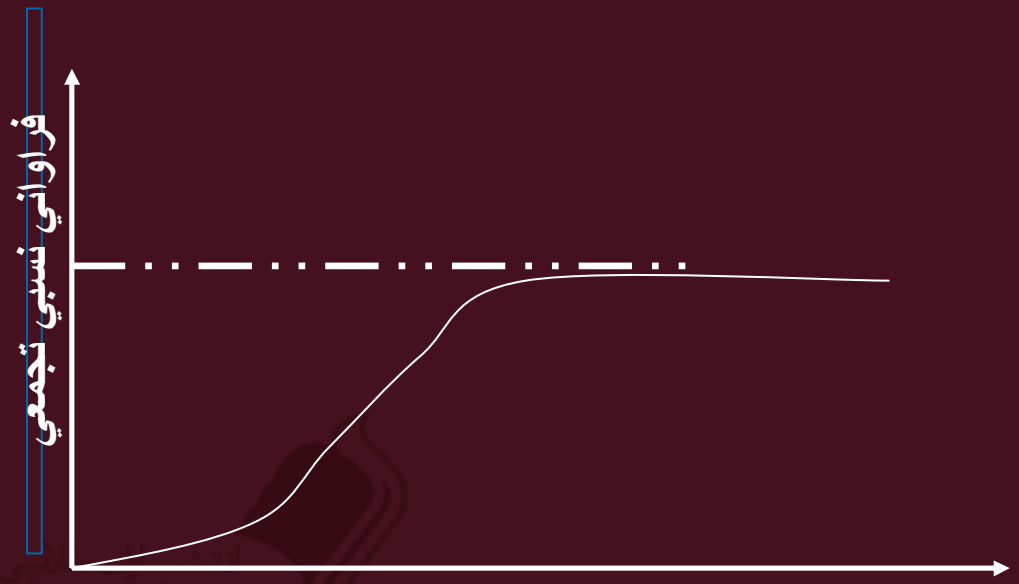
$$k = [1 + 3.3 \log n]$$

$\left(w = \frac{R}{K}\right)$ زياد مي‌شود و در نتيجه طول طبقـــات

كوچك مي‌شود در اين صورت چند ضلعي فراواني و چند ضلعي فراواني تجمعي داراي اضلاع كوچك و زيادي خواهد شد و مي توان بر آنها منحني‌هايي منطبق كرد كه به ترتيب منحني فراواني تجمعي ناميده مي‌شوند .

www.bookgolden.com

منحني فراواني



منحني فراواني نسبي تجمعي

www.bookgolden.com

فصل چهارم

معيارهاي گرايش به مركز

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

معيارهاي گرايش به مركز

در آمار همواره سعي مي‌شود تا اطلاعات نهفته در كل داده‌ها را به صورت يك يا چند عدد معقول در آورد اين اعداد معيارهاي گرايش به مركز ناميده مي‌شود. از مهم‌ترين معيارهاي گرايش به مركز عبارتند از :

الف) ميانگين

ب) نما

ج) ميانه

www.bookgolden.com

الف) ميانگين :

اصلي‌ترين و مورد استفاده‌ترين معيار گرايش به مركز ميانگين است .

فرض كنيد n داده به صورت $x_1, x_2, \ldots, x_k$ با فراواني‌ها $f_1, f_2, \ldots, f_k$ باشند در اين صورت

$$\bar{x} = \frac{\text{تعداد کل داده‌ها}}{\text{مجموع کل داده‌ها}} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$$

را ميانگين داده‌ها مي‌گويند

www.bookgolden.com

مثال : میانگین اعداد 11 ، 9 ،7 ، 5 ، 3 ،1

$$\bar{x} = \frac{1+3+5+7+9+11}{6} = 6$$

مثال : میانگین حقوق کارمندان اداره‌اي بر حسب واحد ده‌هزار تومان در جدول فراواني زیر عبارت است از :

www.bookgolden.com

حقوق	f_i	x_i	$f_i x_i$	$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n} = \frac{360}{40} = 59$
20-40	5	30	150	
40-60	17	50	850	
60-80	13	70	910	
80-100	5	90	450	
	n=40		$\sum f_i x_i = 2360$	

www.bookgolden.com

مثال : ميانگين تعداد فرزندان 20 خانواده در جدول فراواني زير عبارت است از :

x_i	f_i	$f_i x_i$
0	5	0
1	3	3
2	4	8
3	6	18
4	2	8
	$n = 20$	37

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i \times x_i}{n} = \frac{37}{20} = 1/85$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

میانگین جامعه:

هر گاه اطلاعات جمع آوري شده از جامعه اي به حجم N باشد در اين صورت ميانگين بدست آمده ميانگين جامعه بوده و μ با نشان مي دهيم كه در آن :

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

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

ميانگين وزني

در برخي مطالعات يا اندازه گيريها ممكن است كميتهاي بدست آمده داراي وزن يا ضريب خاصي باشند در اين حالت ميانگين وزني محاسبه مي شود . كه فرمول آن عبارت است از:

$$\overline{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

كه در آن w_i وزن داده i ام (x_i) مي باشد

www.bookgolden.com

مثال : دانشجويي دريك درس3 واحدي نمره 14، دردرس2 واحدي12،
در درس 4 واحدي 15 و در 3 واحدي ديگر 11 گرفته است.
معدل وي ويا به بيان آماري ميانگين نمرات وي چقدر است؟

x_i	w_i	$w_i x_i$
14	3	42
12	2	24
15	4	60
11	3	33
مجموع	$\sum w_i = 12$	$\sum w_i x_i = 159$

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{159}{12} = 13/25$$

www.bookgolden.com

مثال : در يك شهر سه روزنامه محلي منتشر مي شود. 18 در صد خانواده‌هاي ساكن اين شهر به هيچيك از روزنامه ها مشترك نيستند . 61 در صد آنها يكي از روزنامه‌ها ، 17 درصد دو روزنامه و4 درصد ديگر هر سه روزنامه مشتركند .

متوسط اشتراك خانواده هاي اين شهر به اين سه روزنامه چقدر است؟

x_i	w_i	$w_i x_i$
0	**18**	**0**
1	**61**	**61**
2	**17**	**34**
3	**4**	**12**
	$\sum w_i = 100$	$\sum w_i x_i = 107$

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i} = \frac{107}{100} = 1/07$$

www.bookgolden.com

ميانگين هندسي

در صورتي كه داده‌ها به صورت نسبت ، نرخ رشد يا از اين قبيل باشند از ميانگين هندسي استفاده مي‌شود كه فرمول آن عبارت است از :

$$G = \sqrt[n]{x_1^{f_1} \times x_2^{f_2} \times \ldots x_k^{f_k}}$$

در كارهاي اقتصادي و جمعيت‌شناسي نيز از اين ميانگين استفاده مي‌كنند.

www.bookgolden.com

مثال : قیمت یك کالا در چهار سال گذشته به ترتیب 100 و 15 و 300 و 800 ریال بوده است میانگین قیمت هر سال به سال قبل را تعیین کنید .

$$x_1 = \frac{150}{100}$$

$$x_2 = \frac{300}{150}$$

$$G = \sqrt[3]{\frac{150}{100} \times \frac{300}{150} \times \frac{800}{300}} = \sqrt[3]{8} = 2$$

$$x_3 = \frac{800}{300}$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

مثال : اگر سرمایه شرکت‌ها پس از یکسال 21 درصد افزایش یابد و سال بعد 44 درصد نسبت به سال قبل افزایش یابد متوسط میزان افزایش سرمایه را محاسبه کنید.

پس از يك سال ، سرمايه نسبت به سال قبل 1/21 برابر و در سال بعد نسبت به سال قبل 1/44 برابر شده است ، لذا يعني ميزان افزايش متوسط سرمايه 32 درصد بوده است

$$G = \sqrt[2]{1/21 \times 1/44} = 1/32$$

يعني ميزان افزايش متوسط سرمايه 32 درصد بوده است

www.bookgolden.com

ميانگين توافقي (هارمونيك)

چنانچه داده‌ها به صورت ميزان تغييرات بيان شده باشند مانند سرعت ، شتاب ، و ... ميانگين توافقي مناسب است كه فرمول آن عبارت است از :

$$H = \frac{n}{\sum_{i=1}^{k} \frac{f_i}{x_i}}$$

اين ميانگين در عينك سنجي و مطالعه برق بكار مي‌رود و گاهي نيز به ميانگين همسازه معروف است .

انتشارات طلايي
پويندگان دانشگاه
راهنماي طلايي
تست طلايي
پيك طلايي
www.bookgolden.com

مثال : اتومبيلي سه قطعه از يك جاده با طول‌هاي يكسان را با سرعت‌هاي 75 و 80 و 95 كيلومتر در ساعت طي مي‌كند متوسط سرعت اين اتومبيل در اين سه قسمت عبارت است از :

n=3

$$x_1 = 75$$

$$x_2 = 80$$

$$x_3 = 95$$

$$H = \frac{3}{\frac{1}{75}+\frac{1}{80}+\frac{1}{95}} = 82/51$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

مثال : جدول توزيع فراواني مصرف برق 100 خانوار به صورت زير داده شده است ميانگين هارمونيك را حساب كنيد .

حدود طبقات	f_i	x_i
$60-62$	5	61
$62-64$	18	63
$64-66$	42	65
$66-68$	27	67
$68-70$	8	69

N = 100

$$H = \frac{n}{\sum \frac{f_i}{x_i}} =$$

$$\frac{5}{\frac{5}{61}+\frac{18}{63}+\frac{42}{65}+\frac{27}{67}+\frac{8}{69}} = 68/03$$

www.bookgolden.com

ميانگين رتبه دو :

اين ميانگين اغلب در محاسبات كارهاي فيزيكي مانند سيستم‌هاي قدرت ولتاژها و جريان‌هاكاربرد دارد و به صورت :

$$M_2 = \left(\frac{1}{n} \sum_{i=l}^{k} f_i x_i^2 \right)^{\frac{1}{2}}$$

تعريف مي‌شود درحقيقت اين

ميانگين جذر ميانگين حسابي $x_1^2, x_2^2, \ldots, x_k^2$ مي‌باشد.

www.bookgolden.com

مثال : براي داده‌هاي

5، 7 ، 7 ، 1 ، 3 ، 7 ، 8 ، 3 ، 12 ، 11 ، 15 ، 11 ، 16 ، 19 ، 15 میانگین رتبه دو عبارت است از :

$$M_2 = \left(\frac{(5)^2 + 3\times(7)^2 + (1)^2 + 2\times(3)^2 + (8)^2 + (12)^2 + 2\times(11)^2 + 2\times(15)^2 + (16)^2 + (19)^2}{15} \right)^{\frac{1}{2}} = 10/6$$

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

میانگین میانگین‌ها

فرض کنید k نمونه با حجم‌هاي $n_1, n_2, \ldots, n_k$ با میانگین‌هاي $x_1, x_2, \ldots, x_k$ باشند در این صورت میانگین میانگین‌ها عبارت است از :

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i}$$

www.bookgolden.com

مثال : میانگین مزد 10 کارگر350 هزار تومان و میانگین مزد 15 کارگر دیگر 250 هزار تومان مي‌باشد میانگین مزد این 25 کارگر را حساب کنید .

$$n_1 = 10$$

$$\bar{x}_1 = \mathbf{350}$$

$$n_2 = 15$$

$$\bar{x}_2 = 250$$

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i} = \frac{10\times350+15\times250}{25} = 290$$

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

ب) نما :

داده‌اي كه فراواني آن نسبت به ساير داده‌ها بيشتر باشد مد يا نما ناميده مي‌شود و با M نشان داده مي‌شود .

مثال : نماي داده‌هاي 3 و 1 و 3 و1 و3 و 3 و 2 و 4 و 5 و 3 و 3 و 1

عبارت است از M=3

مثال : جدول زير تعداد واكسيناسيون انجام شده عليه انواع بيماري‌ها ، در يك سال را نشان مي‌دهد مد يا نما عبارت است از :

نوع بيماري	فلج اطفال	سرخك	سل	ديفتري
فراواني بر ح	11	3	8	4

فلج اطفال = M

www.bookgolden.com

مثال : توزيع فراواني تعداد افراد خانوار به صورت زير داده شده است .

تعداد افراد خانوار	1	2	3	4	5	6	7
فراواني	5	10	20	10	40	20	10

در اين صورت M=5 چون خانواده‌هاي 5 نفري بيشترين فراواني را دارند

مثال : مد يا نما در مجموعه داده‌هاي2 و 1 و 3 و5 و 2 و5 و 3 و 1 وجود ندارد زيرا فراواني تمام داده‌ها يكسان مي‌باشد .

www.bookgolden.com

محاسبه نما براي داده‌هاي طبقه بندي شده :

ابتدا نماينده‌ي طبقه‌اي را كه داراي بيشترين فراواني است و به طبقه‌ي نما معروف است را به عنوان نما اختيار مي‌كنند براي دقت بيشتر مي‌توان نما را از فرمول

$$M = L_M + \frac{d_1}{d_1 + d_2}.w$$

را بدست آورد .

كه در آن :

L_M حد پايين طبقه‌ي نما

d_1 اختلاف فراواني طبقه‌ي نما با طبقه‌ي قبلي

d_2 اختلاف فراواني طبقه‌ي نما با طبقه بعدي

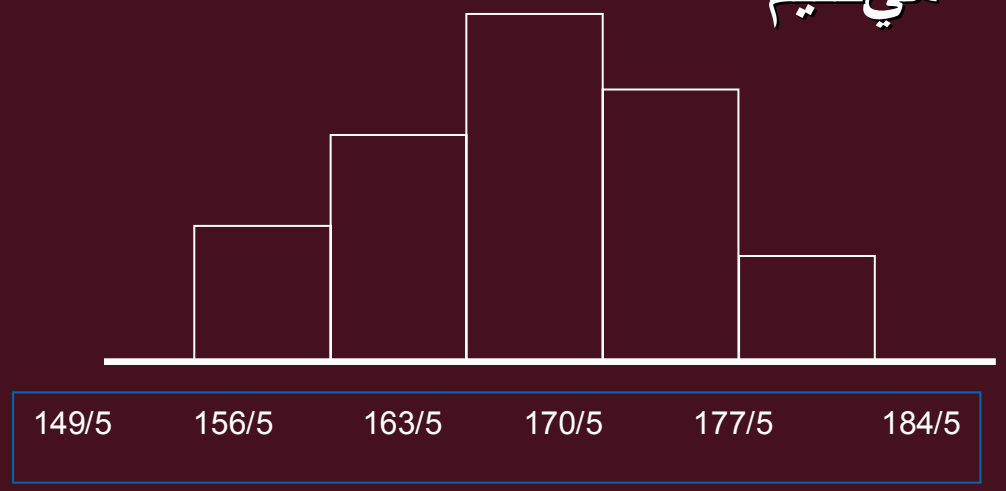
w طول طبقه

www.bookgolden.com

مثال : در جدول فراواني داده شده مقدار نما را بيابيد.

حدود طبقه	فراواني
149/5–156/5	15
156/5–163/5	2
163/5–170/5	30
170/5–177/5	25
177/5–184/5	10

مقداري كه داراي بيشترين فراواني است بعنوان طبقه نما انتخاب مي‌كنيم



www.bookgolden.com

طبقه نما یعني $\frac{163/5+170/5}{2}=167$ بعنوان مقدار تقریبي نما

و

مقدار دقیق آن عبارت است از :

$$M = L_M + \frac{d_1}{d_1+d_2}.w$$

$$M = 163/5 + \frac{10}{10+5} \times 7$$

$$M = 168/17$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

ج) ميانه :

يكي ديگر از معيارهاي گرايش به مركز ميانه مي‌باشد و آن را با m نشان مي‌دهند ميانه داده‌اي است كه تقريباً نصف داده‌ها از آن كوچكتر باشند.

براي بدست آوردن ميانه ابتدا داده‌ها را به صورت غير نزولي از كوچ به بزرگ مرتب مي‌كنند اگر تعداد داده‌ها فرد باشد در اين صورت داده‌اي كه در وسط قرار دارد ميانه محسوب مي‌شود و اگر تعداد داده‌ها زوج باشد نصف مجموع دو داده‌اي كه در وسط قرار دارند ميانه محسوب مي‌شود .

www.bookgolden.com

مثال : براي پيدا كردن ميانه داده‌هاي
20 و 23 و 20 و 19 و 20 و 21 و 27
آنها را به صورت غير نزولي 27 و 23 و 21 و 20 و 20 و 20و 19
مرتب مي‌كنيم لذا ميانه عدد وسطي است m=20

مثال : ميانه داده‌هاي 2 و3 و5 و4 و1 و2 و3 و2 مي‌شود m = 2/5

انتشارات طلايي
پويندگان دانشگاه
راهنماي طلايي
تست طلايي
پيك طلايي
www.bookgolden.com

اگر داده‌ها گسسته همراه با فراواني داده شده باشند در اين صورت با تشكيل ستون فراواني تجمعي ، داده‌اي كه در وسط قرار دارد را بعنوان ميانه انتخاب مي‌كنيم

x_i	12	14	15	17	19
f_i	4	8	6	4	3

مثال : جدول توزيع فراواني نمرات 25 دانشجو به صورت زير داده شده است .

x_i	f_i	F_i
12	4	4
14	8	12
15	6	18
17	4	22
19	3	25

با تشكيل ستون فراواني تجمعي و با توجه به n=24 ملاحظه مي‌شود كه داده‌ي وسطي سيزدهمين داده مرتب شده مي‌باشد

$$M=x_{(13)}=15$$

www.bookgolden.com

محاسبه ميانه براي داده‌هاي طبقه بندي شده :

ابتدا طبقه‌اي را كه نصف داده‌ها در آن قرار دارد و يا به عبارتي اولين طبقه‌اي كه فراواني تجمعي آن برابر نصف يا بيشتر از نصف داده‌ها باشد و به آن طبقه ميانه مي‌گويند را مشخص كرده و سپس از فرمول زير ميانه را محاسبه مي‌كنند.

انتشارات طلايي
پويندگان دانشگاه
راهنماي طلايي
تست طلايي
پيك طلايي
www.bookgolden.com

$$m = L_m + \frac{n \times 0/5 - F_{m-1}}{f_m} \times w$$

که در آن

n : تعداد کل داده‌ها

L_m : حد پاین طبقه میانه

f_m : فراواني طبقه میانه

F_{m-1} : فراواني تجمعي طبقه‌ي ما قبل طبقه میانه

w : طول طبقه

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

مثال : توزيع فراواني وزن 100 نفر دانشجو در جدول زير داده شده است ميانه وزن دانشجويان عبارت است از :

حدود طبقات	فراواني	فراواني تجمعي
59 – 62	5	5
62 – 65	18	23
65 – 68	42	65
68 – 71	27	92
71 – 74	8	100

n = 100

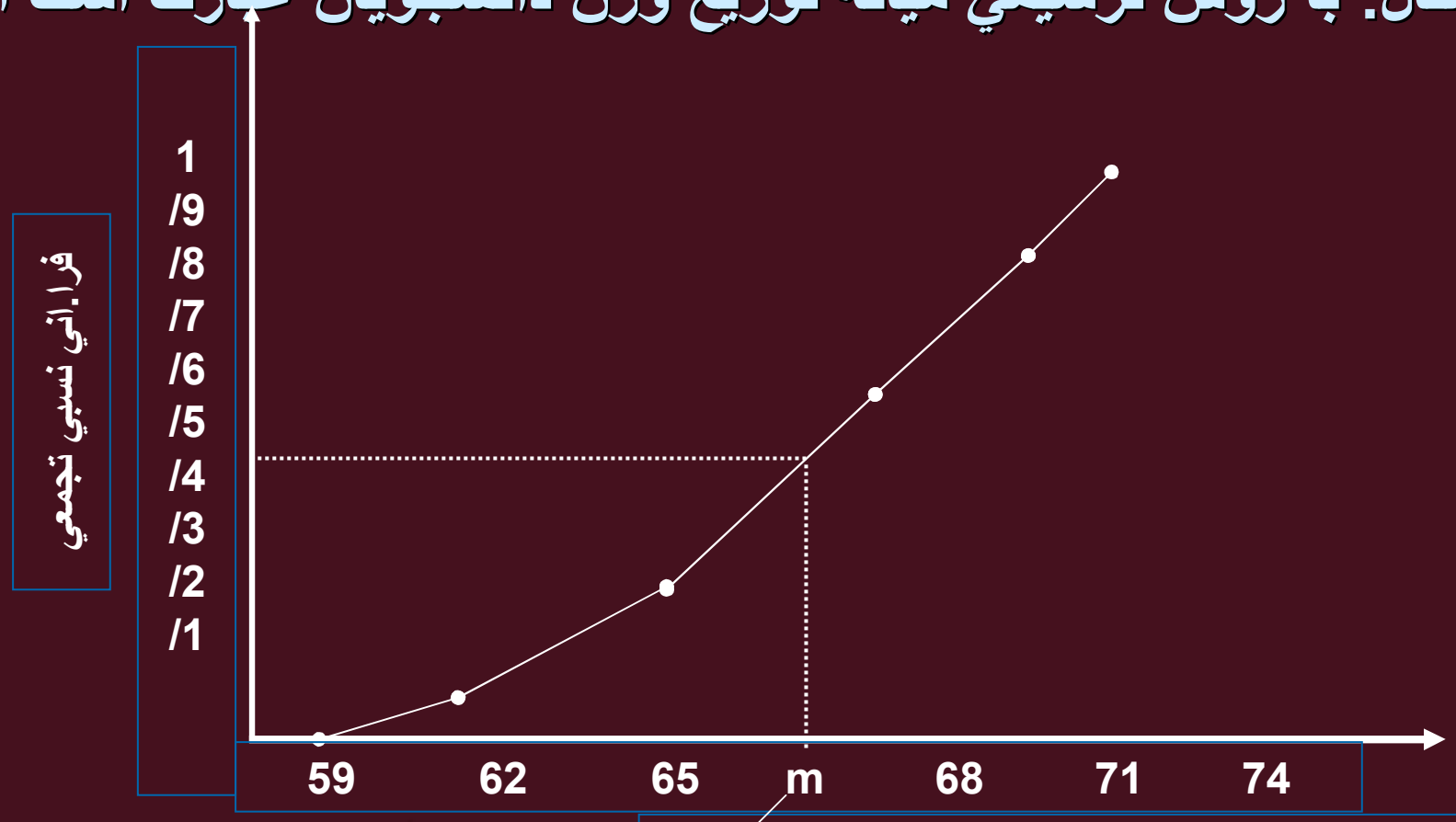
$$m = L_m + \frac{n \times 0/5 - F_{m-1}}{f_m} \times w$$

$$m = 65 + \frac{100 \times 0/5 - 23}{42} \times 3$$

$$m = 66/93$$

www.bookgolden.com

مثال: با روش ترسيمي ميانه توزيع وزن دانشجويان عبارت است از:



نمودار چند ضلعي نسبي تجمعي وزن دانشجويان

m=66/93

www.bookgolden.com

خاصيت مهم ميانه اين است كه حاصل جمع قدر مطلق تفاضلهاي مقاديراز ميانه، از حاصل جمع قدرمطلق تفاضلهاي مقادير از هر عدد ديگر مثل c، كوچكتر است يعني

$$\sum_{i=1}^{k} f_i \left| x_i - m \right| \le \sum_{i=1}^{k} f_i \left| x_i - c \right|$$

به عبارت ديگر ميانه حاصل جمع قدر مطلق تفاضلهاي مقاديري از عدد ديگر را به حداقل مي رساند.

www.bookgolden.com

فصل پنـــجم

چندکها

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

چنـدك ها :

عدد Q_p را كه درآن $0< P<1$ ، چندك مرتبه P مي نامند ، هرگاه 100P درصد داده‌ها كوچكتر ازآن باشند ، مثلاً $Q_{./35}$ ، چندك مرتبه ./35 مي‌باشد ، هرگاه تقريباً 35 درصد داده‌ها كوچكتر از $Q_{./35}$ باشد.

چندك‌هاي معروف عبارتند از :

الف)چارك ها

چارك ها كه به ازاي ./75 و ./5 و ./25 بدست مي‌آيند وآنها رابـ Q_1وQ_2وQ_3 نشان مي‌دهند كه در آن Q_2 همان ميانه مي‌باشد.

www.bookgolden.com

ب)دهك‌ها

دهك‌ها كه به ازاي $P=./1 , ./2 , \ldots , ./9$ بدست مي‌آيند وآن‌ها را با D_1 و D_2 و ... و D_9 نشان مي دهند.

ج)صدك‌ها

صدك‌ها كه بازاي $P=./1 , ./2 , \ldots , ./9$ بدست مي‌آيند و آنها را با P_1 و P_2 و ... و P_9 نشان مي‌دهند.

www.bookgolden.com

محاسبة چندك‌ها در داده هاي طبقه بندي نشده

فرض كنيد n داده به صورت غير نزولي $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ مرتب شده باشد براي بدست آوردن Q_p ابتدا $(n+1)P$ را تشكيل

مي‌دهيم اگر حاصل ،عددي صحيح مانند k شود بنابراين $Q = X_{(k)}$ در غير اينصورت: $(n+p) = k + r$

كه در آن $0 < r < 1$ وقسمت اعشاري $(n+1)P$ مي‌باشد در اين

صورت با يك تناسب فاصله اي ساده داريم:

$$Q_p = X_{(k)} + r\,(X_{(k+1)} - X_{(k)})$$

www.bookholder.com

مثال:درداده‌ها زیر ، چارك اول را محاسبه كنيد.

13و12و9و8و14و16و16و17و14و13و20و16و15و11و10و7و14و6و 12

ابتدا داده ها را به صورت غير نزولي مرتب مي كنيم

20و17و16و16و15و14و14و13و13و12و12و11و10و9و8و7و 6

در اين صورت چارك اول كه همان صدك 25م مي باشد عبارت است از:

$Q_1 = Q_{./25}$

$P = ./25 \quad 19$

$(n+1)P = 20\times ./25 = 4$

n=

بنابراين k =5

و...

$Q_{./25} = X_{(5)} = 10$

www.bookgolden.com

محاسبةچندك ها در داده هاي طبقه بندي شده

روش محاسبه چندك ها كاملاً شبيه محاسبه ميانه است.براي محاسبه چندك Q_p ،ابتدا عدد n.p را با ستــون فراواني تجمــعي مقايسه مي‌كنيم اولين طبقه‌اي كه فراواني تجمعي آن بزرگتر يا مساوي n.p باشد راطبقةچندكpام ناميده و با استفاده از يك تناسب ساده رياضي Q_p را از فرمول زير محاسبه مي كنند.

$$Q_p = L_p + \frac{n.p - F_{(p-1)}}{f_p}.w$$

www.bookgolden.com

كه در آن

L_p : حد پايين طبقةچندك

f_p : فراواني طبقةچندك

F_{p-1} : فراواني تجمعي طبقةماقبل چندك

w : طول طبقه

مثال:در جدول توزيع فراواني سن ازدواج چارك اول و صدك 60ام عبارتنداز:

سن ازدواج	فراواني سن	فراواني تجمعي
14/5 – 19/5	18	18
19/5 – 24/5	74	92
24/5 – 29/5	62	154
29/5 – 34/5	26	180
34/5 – 39/5	20	200

n = 200

www.bookgolden.com

$$Q_1 = Q_{0/25} = 19/5 + \frac{200 \times 0/25 - 18}{74} \times 5 = 21/66$$

$$Q_{60} = Q_{./60} = 24/5 + \frac{20 \times ./6 - 92}{62} \times 5 = 26/76$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

فصل ششم

معيارهاي پراكندگي

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

داده‌ها را معمولاً به صورت يك عدد به نام معيار تمركز خلاصه مي‌كنند وقسمتي از اطلاعات موجود در آنها را در اين عدد منعكس مي‌سازند ولي در تحليلهاي‌آماري ، معيارهاي مركزي به تنهايي نمي‌توانند ميزان تنوع ، تفاوت و يا به عبارتي ديگر ميزان پراكندگي داده‌ها را بيان‌كنند.

فرض كنيد دو مجموعه داده B وA عبارتند از :

A= 1 و 2 و 3 و 4 و4و 6 و 8

B =0 و 1 و 2 و 4 و 4و 8 و 9

ملاحظه مي شود كه ميانگين ،ميانه و نماي هر دو مجموعه ، عدد 4 مي باشددر حالي كه ميزان تجمع و پراكندگي دو مجموعه با هم متفاوت است .

www.bookgolden.com

بنابراين معرفي معيارهايي كه بتواند بيانگر ميزان اختلافها ، تنوعها و ميزان پراكندگي مجموعه داده‌ها باشد ضرورت دارد.اينك به معرفي چند معيار مهم براي سنجش پراكندگي مي‌پردازيم فرض كنيد x_k و... و x_2 و x_1 يك سري داده n تايي با فراواني‌هاي f_k و...وf_2وf_1 با ميانگين $\bar{x}$ باشند .

$$R = \max(x_i) - \min(x_i) = x_{(n)} - x_{(1)}$$

انتشارات طلايي
پويندگان دانشگاه
راهنماي طلايي
تست طلايي
پيك طلايي
www.bookgolden.com

الف) دامنه تغییرات

تفاضل کوچکترین داده از بزرگترین داده را که با R نشان مي‌دهند دامنه تغییرات مي‌گویند .

مثال : حقوق پرداختي ماهیانه کارکنان یك شرکت کوچك به شرح زیر است .

0 6 و 22 و 20 و 17 و 17 و 17 و 16 و 15

دامنه تغییرات R = 60 – 15 = 45 مي باشد .

www.bookgolden.com

ب) ميانگين قدر مطلق انحرافات :

معيار مناسب در سنجش پراكندگي داده‌ها،معياري است كه بتواند ميزان انحراف و پراكندگي كل داده‌ها را در مجموعه داده‌ها اندازه بگيرد . پراكندگي و انحراف زماني مفهوم پيدا مي كند كه داده ها نسبت به يك مبدأ يا مركز مقايسه شوند مناسب ترين مركز براي داده‌ها ، ميانگين آنهاست . بنابراين تعريف مي كنيم :

ميانگين قدر مطلق انحرافات = مجموع كل قدر مطلق انحرافات / تعداد كل انحرافات

$$M.A.D = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n}$$

www.bookgolden.com

مثال : ميزان مهارت مديران يك سازمان در برنامه ريزي به صورت زير است.

45 و 35 و 50 و 65 و 60 و 85 و

$x_i = 80$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{420}{7} = 6$$

در اين صورت

x_i	80	85	60	65	50	35	45	$\sum_{i=1}^{7} x_i$
$\lvert x_i - \bar{x} \rvert$	20	25	0	5	10	25	15	$\sum_{i=1}^{7} \lvert x_i - \bar{x} \rvert = 100$

$$M.A.D = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n} = \frac{100}{7} = 14/28$$

www.bookgolden.com

محاسبه M.A.D در جدولهاي فراواني

اگر جدول فراواني داراي K طبقه با نماينده طبـــقات

x_1 و x_2 و ... و x_k وفراوانـهاي متناظر f_1 و f_2 و ... و f_k باشد در اين صورت

$$M.A.D = \frac{\sum_{i=1}^{k} f_i \left| x_i - \bar{x} \right|}{n}$$

انتشارات طلايي
پويندگان دانشگاه
راهنماي طلايي
تست طلايي
پيک طلايي
www.bookgolden.com

مثال: در جدول توزيع فراواني زير سن 200 مرد زير 40 سال در اولين ازدواج داده شده است.

سن ازدواج	f_i	x_i	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \times \lvert x_i - \bar{x} \rvert$
14/5 - 19/5	18	17	306	8/9	160/2
19/5 - 24/5	74	22	1628	3/9	288/6
24/5 - 29/5	62	27	1674	1/1	682
29/5 - 34/5	26	32	832	6/1	158/6
34/5 - 39/5	20	37	740	11/1	222
مجموع	200		5180		1511/4

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{5180}{200} = 25/9$$

$$M.A.D = \frac{\sum f_i \lvert x_i - \bar{x} \rvert}{n} = \frac{1511/4}{200} = 7/557$$

www.bookgolden.com

ج) واريانس :

در اندازه‌گيري مجموع انحرافات و پراكندگي از مركز داده‌ها، انحرافـات منفي متـأثر از داده‌هاي كوچكتراز ميانگين با انحرافات مثبت خنثي مي‌شوند روشي ديگر براي رفع اين معضل آن است كه انحرافها را مربع كنيم.

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

دراين صورت معيــار ميانگين مربعات انحرافــات از ميانگين به عنوان معيار سنــجش پراكندگي تعريف مي شود. و آن را *واريانس* مي ناميم .

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

با استفاده از جبر مقدماتي مي توان فرمول فوق را به صورت خلاصه تر بازنويسي نمود .

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2$$

www.bookgolden.com

مثال: سود ده ماه يك شركت سرم سازي به اين صورت است.

4 و 6 و 8 و 2 و 3 و 4 و 7 و 5 و 1

و $x_i = 0$

بنابراين

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{60}{10} = 6$$

ونيز به روش خلاصه تر مي توان واريانس را محاسبه نمود .

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2 = \frac{220}{10} - (4)^2 = 6$$

انحراف استاندارد

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

جذر مثبت واريانس را انحراف استاندارد مي گويند

www.bookgolden.com

واريانس جامعه :

$$\delta^2 = \frac{\sum_{i=1}^{N} f_i (x_i - \mu)^2}{N}$$

و گاهـــي با

$$\delta_x^{\ 2} = \frac{\sum_{i=1}^{N} f_i (x_i - \mu_x)^2}{N}$$

نشان مي دهند

انحراف استاندارد جامعه را با جذر مثبت از واريانس جامعه مي‌توان بدست آورد

www.bookgolden.com

محاسبه واريانس در جدولهاي فراواني

اگر جدول فراواني داراي K طبقه باشد نماينده طبــقات را با x_1 و x_2 و...و x_k با فراوانيــهاي f_1 و f_2 و ... و f_k نمايش دهند

در اين صورت فرمول واريانس را مي توان به صورت زير بازنويسي نمود .

$$s^2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{n}$$

و نيز با استفاده از جبر مقدماتي مي توان فرمول فوق را خلاصه تر نمود .

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - \bar{x}^2$$

www.bookgolden.com

مثال:توزيع فراواني سن 200 مرد زير 40 سال در اولين ازدواج در جدول زير آمده است .

سن بر حسب سال	f_i	x_i	$f_i x_i$	$f_i x_i^2$
14/5–19/5	18	17	306	5202
19/5–24/5	74	22	1628	35816
24/5–29/5	62	27	1674	45198
29/5–34/5	26	32	832	26624
34/5–39/5	20	37	740	27380
مجموع	200		5180	140220

$$\bar{x} = \frac{\sum_{i=1}^{x} f_i x_i}{n} = \frac{5180}{200} = 25/9$$ سال

$$s^2 = \frac{\sum_{i=1}^{k} f_i x_i^2}{n} - \bar{x}^2 = \frac{140220}{200} - (25/9)^2 = 30/29$$

$$s = \sqrt{30/29} = 5/5$$ سال

www.bookgolden.com

و) دامنة ميان چاركي :

در توزيع هايي كه داراي تعداد اندكي داده نامعلوم يا پرت ، يعني
خيلي كوچك يا خيلي بزرگ در ابتدا و انتهاي منحني فراواني هستند از انحراف چاركي بعنوان شاخص پراكندگي استفـــاده مي شود.

$$IQR= Q_3 - Q_1$$

گاهي اوقات به جاي دامنة ميان چاركي از نيمه ميان چاركي با عنوان
انحراف چاركي استفاده مي شود كه فرمول آن عبارت است از:

$$SIQR=\frac{Q_3-Q_1}{2}$$

www.bookgolden.com

مثال:براي داده هاي 12و8و7و7و5و4و3و2و1و0داريم:

$$IQR = Q_3 - Q_1 = 7 - 2 = 5$$

$$SIQR = \frac{Q_3 - Q_1}{2} = \frac{7-2}{2} = 2/5$$

$$IQ = \frac{Q_1 + Q_3}{2} = \frac{7+2}{2} = 4/5$$

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

هـ)ضريب تغييرات :

نسبت انحراف استاندارد به ميانگين يعني: $C.V = \frac{s}{\bar{x}}$

را كه اغلب به صورت درصد بيان مي‌شود را ضريب تغييرات

مي‌نامند ، اين معيار كه به واحد اندازه‌گيري بستگي ندارد ؛ در

عمل براي مقايسه بكار مي رود.

www.bookgolden.com

مثال : كارخانه اي دو نوع لاستيك توليد مي كند. براي نوع A ميانگين
عمر 1000 كيلومتر با انحراف استاندارد 200 كيلومتر براي نوع B
ميانگين عمر 1100 كيلومتر با انحراف استاندارد 1000 كيلومتر مي باشد،
كدام نوع لاستيك بهتر است؟

$$CV_A = \frac{S_A}{\overline{X}_A} = \frac{200}{1000} = 0/2 \qquad C.V_B = \frac{S_B}{X_B} = \frac{1000}{1100} = 0/91$$

لاستيك نوع A بهتر است زيرا داراي ضريب تغييرات كمتري است.

www.bookgolden.com

تاثير عدد ثابت بر ميانگين و واريانس :

هرگاه داده هاي $x_1, x_2, \ldots, x_n$ را در عدد حقيقي a ضرب و باعددحقيقي b جمع كنيم،داده اي جديد $ax_1+b, ax_2+b, \ldots, ax_n+b$ حاصل مي شوند كه آنها را با $y_1, y_2, \ldots, y_n$ نشان مي‌دهيم

در اين صورت ميانگين داده هاي جديد عبارت است از:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{n} (ax_i + b)}{n} = a\bar{x} + b$$

www.bookgolden.com

وبه همين شيوه واريانس داده هاي جديد عبارت است از:

$$S_y^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(\left(ax_i + b\right) - \left(a\bar{x} + b\right)\right)^2}{n}$$

$$= a^2 \frac{\sum\left(x_i - \bar{x}\right)^2}{n} = a^2 s^2 x$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

ضريب تغييرات چاركي :

درتوزيعهايي كه داراي اندكي دادة نامعـلوم يا پرت در
ابتدا يا انتهاي منحني فراواني هستند از معيار ضريب تغييرات چاركي استفاده مي شودكه فرمول آن عبارت است از:

$$C.V.Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

www.bookgolden.com

مثال:توزيع فراواني سن افراد جوياي كار در جدول زير آمده است.

	فراواني	فراواني تجمعي
كمتر از12سال	2	2
12-16	18	20
16-20	40	60
20-30	60	120
30-40	50	170
40-60	20	190
بيش از 60سال	10	200

در اين صورت

$$Q_1 = 19$$

$$Q_3 = 36$$

$$C.V.Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{36 - 19}{36 + 19} = \frac{17}{55} = 0/31$$

www.bookgolden.com

گشتاور و گشتاور مرکزي داده ها :

فرض‌کنید n داده به صورت $x_1, x_2, \ldots, x_k$ با فراواني هاي $f_1, f_2, \ldots, f_k$ باشند ، میانگین توان rام x_iها و $(x_i - \bar{x})$ها یعني:

$$m_r^1 = \frac{\sum_{i=1}^{k} f_i x_i^r}{n} \quad , \quad m_r = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^r}{n}$$

را به ترتیب گشتارrام وگشتاور مرکزيrام داده ها مي نامند که در آن r يك عدد طبيعي است واضح است که

$$m_1^1 = \bar{x}$$

$$m_1 = 0$$

$$m_2 = s^2$$

www.bookgolden.com

مثال:گشتاور مرکزي مرتبةاول تا چهارم براي داده‌هاي10و9و8و7و6عبارتند از:

$$\bar{x}=\frac{40}{5}=8$$

$$m_1=\frac{0}{5}=0$$

$$m_2=\frac{10}{5}=2$$

$$m_3=\frac{0}{5}=0$$

$$m_4=\frac{34}{5}=6/8$$

x_i	$(x_i-\bar{x})$	$(x_i-\bar{x})^2$	$(x_i-\bar{x})^3$	$(x_i-\bar{x})^4$
6	-2	4	-8	16
7	-1	1	-1	1
8	0	0	0	0
9	1	1	1	1
10	2	4	8	16
مجموع40	0	10	0	34

www.bookgolden.com

تبديل يا كد گذاري داده ها :

تبـديل يا كد گذاري مجموعـه اي از داده هاي ناجور(خيلي كوچك يا خيلي بزرگ)عبارت از عملياتي است كه طي آن از هرمشاهده ، عدد ثابتي را كم (يا به‌آن اضافه)كرده ونتيجه را

برعدد ثابتي تقسيم(يا درآن ضرب)مي نمايند تا اعداد مناسب جهت محاسبات ساده‌تر حاصل آيد.

www.bookgolden.com

اگر دادهاولیه را با x_i نشان دهیم با انتخاب مقادیر ثابت a وb کد یا داده تبدیلي متناظر به صورت $u_i = \frac{x-a}{b}$ تعریف مي شود.

بنابر این

$$\bar{x} = b\bar{u} + a$$

$$s_x^2 = \frac{\sum\left(x_i - \bar{x}\right)^2}{n} = b^2 \frac{\sum\left(u_i - \bar{u}\right)^2}{n} = b^2 s^2{}_u$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

روش كد گذاري را در مورد داده هاي طبقه بنـــدي شده نيز به

كاربرد . نمايندة طبقه اي كه در وسط جـــدول فراواني قرار داد را به

عنوان مقدار a وطول طبقه را به جاي b فرض مي‌كنيم. درجدول‌هاي

فراواني كه تعداد طبقات زوج مي باشد نمايندة يكي از دو طبقة وسط

(به دلخواه)را بعنوان a برمي گزينيم.

www.bookgolden.com

مثال:در جدول فراواني زير با انتخاب a =17500 و 5000 = b داريم.

حدود طبقات	f_i	x_i	u_i	$f_i u_i$	$f_i u_i^2$
5000-10000	10	7500	-2	-20	40
10000-15000	12	12500	-1	-12	12
15000-20000	35	17500	0	0	0
20000-25000	30	22500	1	30	30
25000-30000	13	27500	2	26	52
	100				

$$\sum f_i u_i = 24 \qquad \sum f_i u_i^2 = 134$$

$$\overline{U} = \frac{\sum f_i u_i}{n} = \frac{24}{100} = 0/24$$

$$S_u^2 = \frac{\sum f_i u_i^2}{n} - \bar{u}^2 = \frac{134}{100} - (0/24)^2$$

$$\bar{x} = b\bar{u} + a = 5000(0/24) + 7500 = 8700$$

$$s_x^2 = b^2 s_u^2 = (5000)^2 = 1/28$$

www.bookgol.com

فصل هفتم

شاخصهاي شكل توزيع

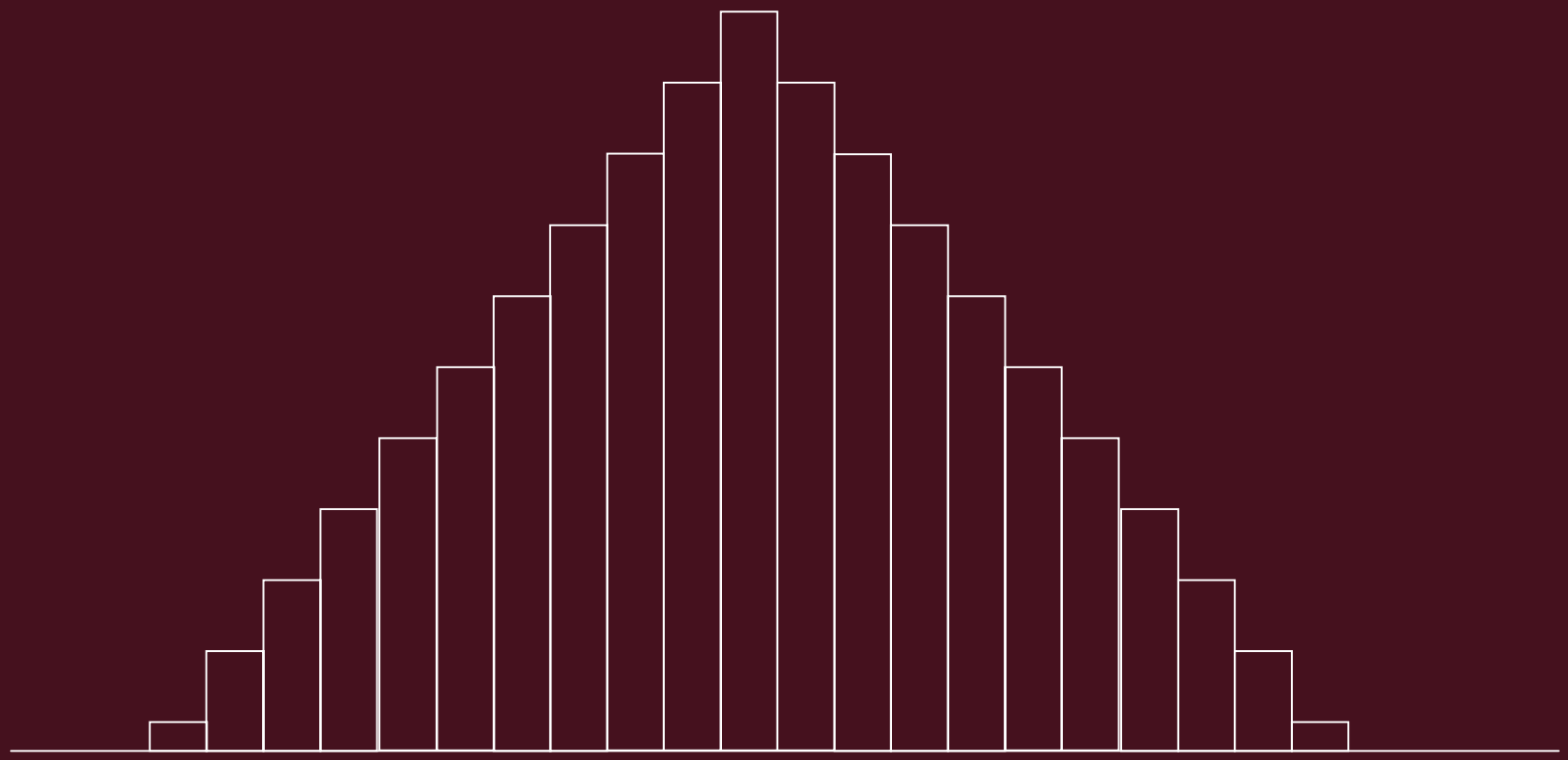
انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

منحني متقارن زنگوله‌اي (نرمال)

اغلب در مواجهه با انبوهي از داده هاي آماري، آنها را در يك جدول فراواني خلاصه كرده و پس از آنكه هيستوگرام مربوطه را رسم كرديم بر آن منحني برازش مي‌دهيم ،

مشاهده خواهيم كرد منحني قرينه زنگوله‌اي شكل بدست خواهد كه به آن منحني طبيعي يا نرمال گفته مي شود . منحني نرمال و مهمي براي توصيف خيلي از مجموعه هاي آماري است.

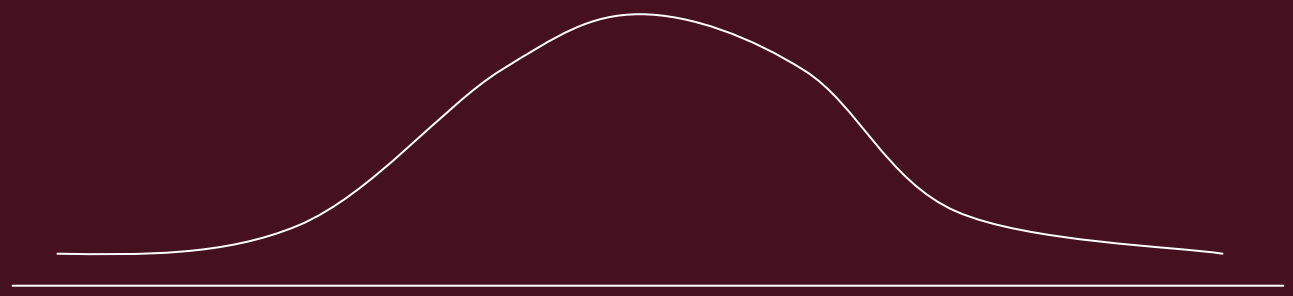
www.bookgolden.com



هیستوگرام متقارن

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

منحني فراواني متناظر آن به شكل زير است .



منحني متقارن زنگوله اي (نرمال)

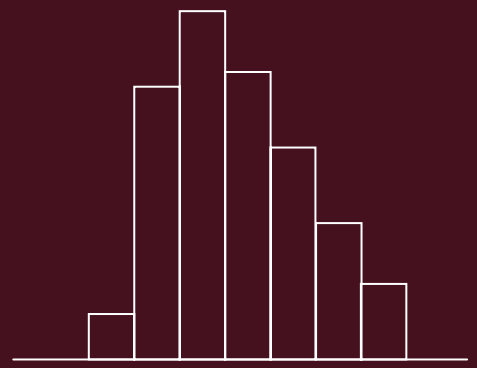
www.bookgolden.com

مفهوم توزيع طبيعي در آمار يك مفهوم اساسي و بنيــادي است. دليل اهميت آن اين است كه توزيع فراواني هيستوگرام بسياري از حوادث طبيعي شكل منحني طبيــعي يا نرمال را دارند يعني فراواني آنها ابتدا به يك مقدار بيشــينه صعود مي‌كند و سپس به طور متقارن كاهش مي‌يابد.

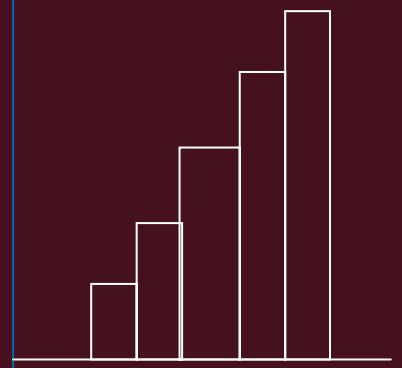
www.bookgolden.com

چولگي:

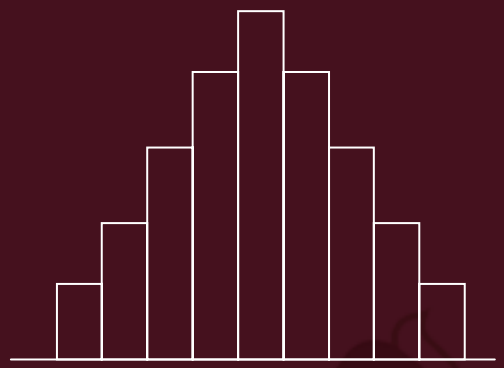
ميزان عدم تقارن منحني را چولگي مي‌گويند. در توزيعهاي چوله معيـــارهاي گرايش مركزي بر هم منطبق نيستند ولي هميشه ميانه بين نما و ميانگين قرار دارد. اگر نما از ميانگين بزرگتر باشد منحني چوله به چپ است يعني دم بلند منحني به طرف چپ مي‌باشد و اگر ميانه از ميانگين كوچكتر باشد منحني چوله به راست مي‌باشد .



هيستوگرام چوله به راست



هيستوگرام چوله به چپ



هيستوگرام متقارن

www.bookgolden.com

تفاوت ميانگين و نما $(\bar{x}-M)$ يا تفاوت آن با ميانه $(\bar{x}-m)$ مي‌تواند مشخص كننده چولگي باشد حال اين اختلافها با تقسيم بريك معيار پراكندگي هم‌بعد مانند s از قيد واحد خارج مي‌شوند .

بااين انگيزه پيرسون براي اندازه گيري ميزان و جهت چولگي ضرائب چولگي زير را تعريف نمود.

www.bookgolden.com

ضريب اول چولگي پيرسون $sk_1 = \frac{\bar{x} - M}{s}$

ضريب اول چولگي پيرسون $sk_2 = \frac{3(\bar{x} - m)}{s}$

ضريب اول چولگي پيرسون $sk_3 = \frac{m_3}{s^3}$

كه درآن m_3 ، گشتاور مركزي مرتبه سوم مي باشد

در توزيعهايي كه چولگي زياد نباشد رابطه تجربي $\bar{x} - m \approx 3(\bar{x} - m)$ كه به رابطه پيرسون شهرت دارد برقرار است.

www.bookgolden.com

تفسير ضرايب چولگي

اگر sk=0 باشد توزيع چولگي ندارد.

اگر sk > 0 باشد توزيع چوله به راست است .

اگر sk < 0 باشد توزيع چوله به چپ است .

در عمل ضرائب چولگي با اعداد 0/1 و 0/5 مقايسه مي شوند بدين ترتيب كه :

اگر |sk| < 0/1 باشد توزيع چوله نيست و مي تواند متقارن باشد.

اگر 0/1< | sk | < 0/5 باشد توزيع چوله است و هرچه sk به 0/1 نزديك باشد چولگي كم و هر چه به 0/5 نزديك باشد چولگي زياد است .

اگر | sk | > 0/5 باشد توزيع خيلي چوله است.

www.bookgolden.com

کرده‌اند و به شرح زیر است.ضریب چولگي اول پیرسون را محاسبه و تفسیر کنید .

x_i	2	3	4	5	6	7
f_i	4	40	24	20	8	4

ابندا معیارهاي گرایش مرکز را محاسبه مي‌کنیم .

x_i	f_i	F_i	$f_i x_i$	$f_i x_i^2$
2	4	4	8	16
3	40	44	120	360
4	24	68	96	384
5	20	88	100	500
6	8	96	48	288
7	4	100	28	196

$n = 100 \quad \sum f_i x_i = 400 \quad \sum f_i x_i^2 = 1744$

www.bookgolden.com

میانگین $= \bar{x} = \frac{\sum f_i x_i}{n} = \frac{400}{100} = 4$

نما $M = 3$

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \bar{x}^2}$$

$$s = \sqrt{\frac{1744}{100} - (4)^2} = 1/2$$

$$sk_1 = \frac{\bar{x} - M}{s} = \frac{4-2}{1/2} = +/83$$

بنابراین چولگي زياد و مثبت است.

نمودار چوله به راست است

www.bookgolden.com

مثال: طول عمر 100 باطري اتومبيل داراي ميانگين،ميانه و انحراف استاندارد 3/5 و 3/48 و 1/65 سال مي باشد ضريب چولگي با توجه به اطلاعات داده شده عبارت است از :

$$\bar{x} = 2/5 \qquad m = 3/48 \qquad s = 1/65$$

$$sk_1 = \frac{3(\bar{x} - m)}{s}$$

$$= \frac{3(3/5 - 3/48)}{1/65} = 0/036$$

www.bookgolden.com

ضريب چولگي باولي :

در توزيع‌هاي كه داراي تعداد اندكي داده نا معلوم يا پرت در ابتدا يا انتهاي منحني فراواني هستند از اين معياركه براساس انحرافات چاركي مي‌باشد براي اندازه‌گيري ميزان‌چولگي استفاده مي‌شود .

كه به آن ضريب چولگي باولي يا ضريب چولگي مي‌گويند و با sk_B نمايش مي‌دهند.

پس :

$$sk_B = \frac{(Q_3 - m) - (m - Q_1)}{(Q_3 - m) + (m - Q_1)}$$

$$sk_B = \frac{Q_3 + Q_1 - 2m}{Q_3 - Q_1}$$

www.bookgolden.com

تفسير چولگي باولي :

اگر $sk_B > 0$ توزيع چوله به راست است.

اگر $sk_B < 0$ توزيع چوله به سمت چپ است.

اگر $sk_B = 0$ توزيع مي‌تواند متقارن باشد يا متقارن فرض شود.

ضريب چولگي باولي در فاصله‌ي (1+ و 1-) تغيير مي‌كند وحال
ن كه ضريب چولگي پيرسون در فاصله‌ي (3+ و3-) تغيير مي‌كند و
از اين جهت است كه ضريب باولي گاهي اوقات ترجيح داده مي‌شود .

www.bookgolden.com

گاهي اوقات نيز از ضريب چولگي صدكي نيز براي اندازه‌گيري ميزان چولگي استفاده مي‌شود كه فرمول آن عبارت است از :

$$sk_P = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

كه در آن P_{10} و P_{50} و P_{90} به ترتيب صدك‌هاي دهم ، پنجاهم و نودم توزيع مي‌باشند

مثال : در P10 و P50 و P90 توزيع مقدار ضريب چولگي صد

برابراست با : P10 = 123/33 و P50 = 162/25 و P90 = 201/5

$$sk_P = \frac{201/5 - 2\times 122/25 + 123/33}{201/5 - 123/33} = 0/004$$

www.bookgolden.com

كشيدگي :

دو توزيع ممكن است داراي ميانگين يكسان باشند و كاملاً نيز قرينه باشند ولي از لحاظ كشيدگي با هم تفاوت زيادي داشته باشند .

ميزان برجستگي يا پخي منحني فراواني نسبت به منحني مقايسه را كشيدگي گويند و منحني مقايسه را محنحني نرمال در نظر مي گيرند.

$$m_4 = \frac{1}{4}\sum f_i\left(x_i - \bar{x}\right)^4$$

باشد . لذا در توزيع نرمال $\frac{m_4}{s^4} - 3 = 0$ و در توزيع‌هاي غير نرمال اين مقدار برابر صفر نيست و آن را ضريب كشيدگي مي‌نامند و با K نشان مي‌دهند.

www.bookgolden.com

تفسير ضريب كشيدگي :

توزيع نرمال است هرگاه $k = 0$ باشد.

توزيع از توزيع نرمال كشيده‌تر است هرگاه $k > 0$ باشد .

توزيع از توزيع نرمال پخ‌تر است هرگاه $k<0$ باشد.

توزيع تقريباً نرمال است هرگاه $|k| < 0/1$ توزيع يا توزيع نرمال تفاوت فاحش دارد هرگاه $0/5 <$ $0/1<|k|$ باشد .

تفاوت توزيع با توزيع نرمال فاحش است هرگاه $0/5 <$ $|k|$

www.bookgolden.com

مثال : براي جدول توزيع فراواني زير ، ضريب كشيدگي را محاسبه كنيد .

1	2	3	4	5
20	10	40	10	20

$$K = \frac{m_4}{s^4} - 3$$

www.bookgolden.com

x_i	f_i	$f_i \times x_i$	$(x_i - \bar{x})$	$f_i(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^4$
1	20	20	–2	80	320
2	10	20	–1	+10	10
3	40	120	0	0	0
4	10	40	1	10	10
5	20	100	2	80	320
	$n = 100$	$\sum f_i x_i = 300$		$\sum f_i(x_i - \bar{x})^4 = 180$	$\sum f_i(x_i - \bar{x})^4 = 660$

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{300}{100} = 3$$

$$s = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{n}} = \sqrt{\frac{180}{100}} = 1/34$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

$$m_4 = \frac{\sum f_i \left(x_i - \bar{x}\right)^4}{n} = \frac{660}{100} = 6/6$$

$$K = \frac{m_4}{s^4} - 3 = \frac{6/4}{(1/34)^4} - 3 = -0/952$$

تفاوت توزيع با توزيع نرمال فاحش است و منحني توزيع از منحني نرمال پخ‌تر(كوتاهتر) است .

ضريب كشيدگي صدكي :

براي تحليل توزيع‌هايي كه با استفاده از گشتاورها قابل توصيف نيستند و با چندك‌ها توصيف مي‌شوند از معيار ضريب كشيدگي چندكي براي اندازه‌گيري ميزان كشيدگي استفاده مي‌شود .

www.bookgolden.com

در توزيع منحني نرمال ثابت مي‌شود كه مقدار $\frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$ كه در آن P_{90} و P_{10} به ترتيب صدك‌هاي نودم و دهم مي‌باشند برابر 0/263 است .

با اين وصف مي‌توان معيار ضريب كشيدگي صدكي را به صورت

$$K_P = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} - 0/263$$

براي اندازه‌گيري ميزان كشيدگي تعريف نمود هر چه اين معيار به صفر نزديك‌تر باشد كشيدگي توزيع به نرمال نزديك‌ترمي‌باشد .

www.bookgolden.com

مثال : ضريب كشيدگي صدكي را براي داده‌ها طبقه‌بندي شده جدول زير:

حدود طبقات	فراواني	فراواني تجمعي
$\langle 5$	2	2
$5-8$	10	12
$8-11$	25	37
$11-14$	23	60
$14-17$	15	75
≥ 17	5	80

$Q_3 = 14$ و $Q_1 = 8/96$

$P_{90}=16/4$ و $P_{10}=6/8$

$$K_P = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} - 0/263$$

$$K_P = \frac{\frac{14-8/96}{2}}{16/4-6/8} - 0/263$$

$$K_P = -/0005$$

ضريب شكستگي نشان مي‌دهد كه توزيع منحني نرمال است.

www.bookgolden.com

فصل هشتم

همبستگي

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

همبستگي:

در بسياري از مطالعات،منظور از بررسي متغيرها،مشخصٍ كردن روابط موجود بين اين متغيرهاست.مثلاً سالها اين‌سوال مطرح بود كه آيا بين استعمال دخانيات وسرطان‌هاي دستگاه تنفس رابطه‌اي وجود دارد يا خير؟

سرانجام با بررسي‌هاي آماري معلوم شد كه رابطه‌اي شديد بين اين دو متغير وجود دارد وتبليغ منع استعمال سيگار شدت بيشتري يافت.مطالعه رابطه بين متغيرها را در اصطلاح آماري مطالعه همبستگي مي گويند.

www.bookgolden.com

ساده‌ترين نوع همبستگي بررسي همبستگي بين دومتغـير مي‌باشد متغيري كه تغييرات آن تحت كنترل نباشد و تغييرات آن دستخوش تغييرات متغير ديگر باشد رامتغيروابسته

يا پاسخ و متغيري كه تغييرات آن توسط آزمايشگر قابل كنترل باشد ودستخوش تغييرات متغير ديگر نباشدرامتغيرمستقل يا كنترل مي نامند.

www.bookgolden.com

دربررسي رابطه بين هزينه تبليغات وافزايش ميزان فروش هزينه تبليغات متغيرمستقل وافزايش ميزان فروش يك متغير وابسته است

و در بررسي مقدار ماده افزودني به بنزين و مقدار كاهش اكسيد ازت، مقدار ماده افزودني متغير كنترل يا مستقل و مقدار كاهش اكسيد ازت يك متغير وابسته يا پاسخ است . متغير كنترل يا مستقل را با x و متغير وابسته يا پاسخ را با y نمايش مي‌دهند.

www.bookgolden.com

انواع همبستگي :

الف) در مواردي كه رابطه بين دو دسته متغيررابطه‌اي مستقيم باشد يعني اين كه افزايش داده ها يك دسته متغير با افزايش داده هاي نظير آنها در متغير ديگر همراه باشد،اصطلاحاً گفته مي شود كه ميان اين دومتغير رابطه مثبت وجود دارد.

www.bookgolden.com

ب)گاهي رابطه بين دو متغير به گونه‌اي است كه تغيير يكي از متغيرها

با تغيير متغير ديگر اما در جهت عكس آن همراه باشد يعني كاهش يكي با افزايش ديگري همراه است. و بالعكس . در چنيــن مواردي گفته مي‌شود بين اين دو دسته متغير رابطه معكوس يا همبستــگي منفي وجود دارد.

ج) بعضي وقتها بين دو دسته متغير هيچ نوع رابطه‌اي خواه مستقيم

و يا خواه غيرمستقيم وجود ندارد. در اين صورت گفته مي‌شود كه همبستگي آنها صفر است.

www.bookgolden.com

جمع آوري داده ها :

در بحث همبســـتگي ساده تنها دو يا چند متغير x (متغير مستقل) و Y

داريم ، براي جمع آوري داده هاي موردنياز اگر متغير xتوسط آزمايشگر قابل كنترل باشد به آزاي مقادير مختلف x مانند $x_1,x_2,\ldots,x_n$ مقاديرy را مشاهده و با $y_1,y_2,\ldots,y_n$ نشان مي دهيم واگر متغير x قابل

كنترل نباشد بايد نمونه اي تصادفي از جامعه مورد مطالعه انتخاب ومقادير x

وyرا براي هر مورد مشاهده وبه صورت زوجهاي

x	x_1	x_2	...	x_n
y	y_1	y_2	...	y_n

ويا $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$

www.bookgolden.com

ثبت مي كنيم.

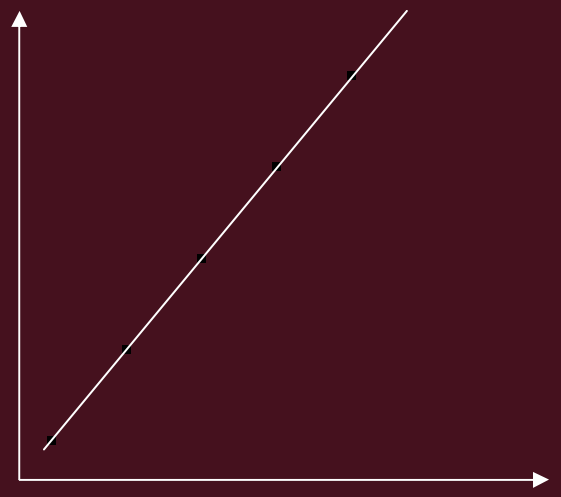
نمودار پراکنش :

در مطالعه رابطه بين دو متغير، اولين قدم منطقي ، رسم داده‌ها به صورت نقاطي بر روي محور مختصـــات دو بعدي دكارتي است. نمودار پراكنش، بينشي درباره ماهيت رابطه‌اي كه به وسيله داده‌ها نمايش داده مي‌شود فراهم مي‌كند

گاهي اوقات اطلاع منعكس در نمودار پراكنش براي جستجوي يك مدل رياضي مناسب (خط راست يا يك نوع منحني مشخص)مفيد است.

در شكلهاي زير نمايش ترسيمي همبستگي ها در حالتهاي مختلف نشان داده شده است.

www.bookgolden.com



همبستگي كامل مثبت

همبستگي ناقص ومثبت

همبستگي كامل ومنفي

همبستگي نا قص ومنفي

www.bookletbank.com

ضرائب همبستگي :

فرض كنيد زوجهاي $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ مربوط به ساعات مطالعه قبل از آزمون ونمرةآزمون باشند.مي خواهيم بررسي كنيم كه آيا بين x وy همبستگي خطي وجود دارد يا خير؟

با رسم نمودار پراكنش تا حدودي مي توان وجود همبستگي ونوع

آن را تشخيص داد اما ما به دنبال معياري عددي هستيم كه بتواند

جهت و ميزان همبستگي بين دو متغير را تعيين كند اين معيار را

www.bookgolden.com

ضريب همبستگي گشتاوري پيرسون:

هنگاميكه داده ها داراي مقياس فاصله اي يا نسبي هستند از اين ضريب استفاده مي كنند كه توسط كارل پيرسون تهيه تنظيم گرديده است وبه نام ضريب همبستگي پيرسون معروف است.

$$r_{x,y} = \frac{s_{xy}}{\sqrt{s_{xx}} \cdot \sqrt{s_{yy}}} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

فرمول فوق را مي توان با محاسبات ساده رياضي به صورت ساده تر باز نويسي نمود

$$r_{x,y} = \frac{\sum_{i=1}^{n} x_i y_i - n\,\overline{x}\,\overline{y}}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2} \times \sqrt{\sum_{i=1}^{n} y_i^2 - n\,\overline{y}^2}}$$

www.bookgolden.com

خواص ضريب همبستگي :

الف) r بايد در فاصله [1+و1-]قرار گيرد.

ب) مقدار عددي r قدرت رابطه خطي وعلامت r نمايانگرجهت اين رابطه است.

ج) اگر مقادير ax+bبا x و مقاديرy با cy + d عوض شوند و در آنها a و c داراي علامت يكسان باشند r تغيير نمي‌كند يا به عبارتي با تغيير مبدأ و واحد اندازه گيري ضريب همبستگي تغيير نمي كند.

www.bookgolden.com

مثال:به عنوان قسمتي از مطالعة همبستگي بين تن وروان در موفقيت ورزشكاران،اندازه هاي زير از اعضاء تيم كشتي المپيك ايران بدست آمده است.

ميزان خشم(x)	6	7	5	21	13	5	13	14
ميزان نيرو مندي(y)	28	23	29	22	20	19	28	19

ضريب همبستگي را بيابيد.

x	y	$x.y$	x^2	y^2
6	28	168	36	784
7	23	161	49	529
5	29	145	25	841
21	22	462	441	484
13	20	260	169	400
5	19	95	25	361
13	28	364	169	784
14	19	266	196	361

$\sum x_i=84 \quad \sum y_i=188 \quad \sum x_i y_i=1921 \quad \sum x_i^2=1110 \quad \sum y_i^2=4544$

www.bookgolden.com

$$s_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 1921 - 8 \times 23/5 = -53$$

$$s_{xx} = \sum x_i^2 - n\bar{x}^2 = 1110 - 8 \times (10/5)^2 = 228$$

$$s_{yy} = \sum y_i^2 - n\bar{x}^2 = 4544 - 8 \times (23/5)^2 = 126$$

$$r_{x.y} = \frac{s_{x.y}}{\sqrt{s_{xx}}\sqrt{s_{yy}}} = \frac{-53}{\sqrt{228}\sqrt{126}} = -0/31$$

همبستگي معكوس ومنفي است.

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

ضـــريب تعيين :

ضريب همبستگي اندازه همبستگي بين دو متغير را نشان مي دهد اما اين شاخص در باره ماهيت اين همبستگي اطلاعات زيادي به ما نمي دهد.آن چه اين روش مشخص مي كند وجود همبستگي و يا بالا و پايين بودن نسبي آن است.

اطلاعات كامل تردر باره همبستگي وقتي امكان پذير است كه شاخص ديگري به نام ضريب تعيين محاسبه مي شود كه فرمول آن عبارت است از: $R^2 = r_{xy}^2 \times 100$

www.bookgolden.com

خواص ضريب تعيين :

الف) $0 \leq R^2 \leq 1$

ب)هرچه R^2 به يك نزديك باشد ميان متغيرها يك رابطه خيلي قوي برقرار است.

هرچه R^2 به يك نزديك باشد بهتر مي توان y را به كمك يك رابطه خطي با داشتن x پيش بيني يا تعيين كرد از اين رو آنرا ضريب تعيين مي‌گويند.

www.bookgolden.com

مثال : براي محاسبه ضريب تعيين بين دو متغير هزينه مسكن (y)ودرآمد(x)، نمونه اي به حجم n=10،انتخاب و براساس داده‌ها ، مقادير زير محاسبه شده است. مقدار R^2 را محاسبه كنيد.

$$\sum x_i y_i = 180 \quad \sum y_i^2 = 210 \quad \sum y_i = 40 \quad \sum x_i^2 = 300 \quad \sum x_i = 50$$

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 50$$

$$S_{yy} = \sum y_i^2 - n\bar{y}^2 = 50$$

$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = -20$$

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{-20}{\sqrt{50}\times\sqrt{50}} = -0/4$$

$$R^2 = r_{x,y}^2 \times 100 = (-0/4)^2 \times 100 = 16$$

يعني 16درصد تغيير پذيري كل به علت رگرسيون روي x و بقيه به علت خطاي پيش بيني(ساير عوامل غير از درآمد) مي باشد.

www.bookgolden.com

ضريب همبستگي رتبه اي اسپرمن :

. ضريب همبستگي پيرسون،براي توصيف همبستگي بين دو متغير كه با استفاده از مقياس فاصله اي يا نسبي اندازه‌گيري شده باشند به كار برده مي شود.

فرض كنيد $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ يك نمونه تصادفي زوجي باشد دراين روش رتبه x_i را با R_i و رتبه y_i را با s_i نشان مي دهيم.اينك به جاي نمونه بالا نمونه زوجي رتبه اي $(R_1, s_1), (R_2, s_2), \ldots, (R_n, s_n)$ بدست مي‌آيد اينك ضريب همبستگي ساده پيرسون را با زوجهاي رتبه اي محاسبه مي كنيم اين ضريب همبستگي را كه اسپرمن درسال 1904معرفي كرده است و با r_{sp} نشان داده ضريب همبستگي اسپرمن يا همبستگي رتبه اي مي ناميم بنابراين:

$$r_{sp} = \frac{\sum_{i=1}^{n}(R_i - \overline{R})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{n}(R_i - \overline{R})^2}\sqrt{\sum_{i=1}^{n}(s_i - \overline{s})^2}}$$

www.bookgolden.com

حال اگر تعريف كنيم $D_i=R_i-S_i$ آنگاه مي توان فرمول ضريب همبستگي را به صورت زير بازنويسي نمود.

$$r_{sp}=1-\frac{6\sum_{i=1}^{n}D_i^2}{n\left(n^2-1\right)}$$

مثال : داده هاي زير تعداد ساعات مطالعه ده دانشجو و نمره هاي آنها در يكي از دروس آمار مي باشد. يافته R_s را پيدا كنيد.

www.bookgolden.com

x:تعداد ساعات	y: نمره	r_i:رتبه x	s_i:رتبهy	D_i	D_i^2
8	56	4 / 5	4	0/ 5	0/ 25
5	44	2 / 5	2	0/ 5	0/ 25
11	79	7	8	– 1	1
13	72	8	7	1	1
10	70	6	6	0	0
5	54	2 / 5	3	– 0/ 5	0/ 25
18	94	10	10	0	0
18	85	9	9	0	0
2	33	1	1	0	0
8	65	4 / 5	5	– 0/ 5	0/ 25

$$r_{sp} = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n\left(n^2 - 1\right)} = 1 - \frac{18}{990} = 0/98$$

www.bookgolden.com

ضريب توافق يا سازگاري چوپروف :

يك نمونه20تايي از كارگران يك كارخانه را انتخاب وآنها را برحسب جنس(x) و به تفكيك مهارت (y)مورد مطالعه قرار داده ايم كه مشاهدات عبارتند از:

(ساده ، زن)(ماهر، مرد)(ماهر، مرد)(نيمه ماهر، زن)(نيمه ماهر، زن)

(نيمه ماهر، مرد)(ساده،مرد)(ساده ، مرد)(ساده ، زن)(ساده ، زن)

(ماهر، مرد)(ماهر، مرد)(ماهر، زن)(ساده ، زن)(ساده ، زن)

(ساده ، زن)(نيمه ماهر ، مرد)(ساده ، مرد)(ساده ، مرد)(ساده ، مرد)

www.bookgolden.com

اطلاعات فوق را مي توان در جدول توزيع فراواني دو بعدي

x \ y	ماهر	نیمه ماهر	ساده
زن	2	2	5
مرد	4	2	5

براي اينكه بدانيم بين دو صفت x وy بستگي ويا به عبارتي ديگر توافق يا سازگاري وجود دارد يا خير، از ضريب توافق يا سازگاري چوپروف استفاده مي كنيم كه فرمول آن عبارت است از:

www.bookgolden.com

$$r_{Tch} = \sqrt{\frac{x^2}{n + x^2}}$$

که در آن

$$x^2 = \sum_{i}^{r} \sum_{j}^{c} \frac{\left(f_{ij} - e_{ij}\right)^2}{e_{ij}}$$

$$e_{ij} = \frac{f_{i_0} \times f_{0j}}{n}$$

که درآن fij فراواني سطر iام و ستون j م مي باشد

با استفاده از محاسبات جبري مي توان عبارت x^2را به صورت زير بازنويسي نمود

$$x^2 = \left(\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{f_{ij}^2}{e_{ij}} \right) - n$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

مثال: جدول زير نتايج يك بررسي در مورد رضـايت از زندگي(x)به تفكيك جنس(y) را نشان مي دهد، ضريب توافق چوپروف را محاسبه كنيد.

x \ y	زن	مرد	f_{i_0}
رضايت مالي	60	80	140
رضايت روحي	70	100	170
رضايت شهرتي	70	120	190
f_{0j}	200	300	$n = 500$

www.bookgolden.com

$$e_{11}=\frac{140\times200}{500}=56 \qquad e_{12}=\frac{140\times300}{500}=84$$

$$e_{21}=\frac{170\times200}{500}=68 \qquad e_{22}=\frac{170\times300}{500}=102$$

$$e_{31}=\frac{190\times200}{500}=76 \qquad e_{32}=\frac{190\times300}{500}=114$$

$$x^2=\left(\sum_{i=1}^{3}\sum_{j=1}^{2}\frac{f_{ij}^2}{e_{ij}}\right)-n=\left(\frac{60^2}{56}+\frac{80^2}{84}+\frac{70^2}{68}+\frac{100^2}{102}+\frac{70^2}{76}+\frac{120^2}{114}\right)-500=1/36$$

$$r_{Tch}=\sqrt{\frac{x^2}{n+x^2}}=\sqrt{\frac{1/36}{500+1/36}}=0/05$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

ضريب توافق پيرسـون :

اندازه درجه رابطه ، بستگي يا ارتباط طبقه‌بنديها در يك جدول توافقي عبارت است از:

$$c = \sqrt{\frac{x^2}{n + x^2}}$$

www.bookgolden.com

ضريب توافق كرامــر :

ضريب توافق پيرسون را معمولاً براي جداول مربعي به كار مي‌برند درصورتي كه جدول توافق به صورت مستطيل باشد آن گاه از ضريب توافق كرامر كه آن را با v نمايش مي‌دهند استفاده مي شود كه فرمول آن عبارت است از:

$$V = \sqrt{\frac{x^2}{n \cdot min.(k-1, l-1)}}$$

كه در آن k و l به ترتيب تعداد سطرها و ستون‌ها مي باشد.

www.bookgolden.com

ضريب كروسكال:

اين ضريب اعم از اينكه متغير كمي يا كيفي باشد به كار برده مي‌شود زيرا فقط از فراوانيهاي جدول تبعيت مي كند و به مقادير متغير ارتباطي ندارد و فرمول آن عبارت است از:

$$\tau_y = \frac{E_1 - E_2}{E_1}$$

كه در آن

$$E_1 = \sum_{j=1}^{l}\left(\frac{N - n_{0j}}{N}\right)n_{0j} \quad , \quad E_2 = \sum_{i=1}^{k}\sum_{j=1}^{l}\left(\frac{n_{i_0} - j_{ij}}{n_{i_0}}\right)n_{ij}$$

لازم به يادآوري است كه n_{i0} فراواني توزيع حاشيه اي xها و n_{0j} فراواني توزيع حاشيه اي yها مي باشد.

www.bookgolden.com

ضريب گاتمن :

اين ضريب نيز از فراواني جدول توافقي تبعيت مي كند واغلب در مواردي به كار برده مي شود كه نتوان صفات مورد مطالعه را به صورت رتبه اي طبقه بندي كرد وفرمول آن عبارت است از:

$$\lambda = \frac{\sum m_y - M_y}{N - M_y}$$

كه درآن M_y عبارت است از $\max n_{.j}$ يعني بزگترين فراواني توزيع حاشيه اي yها و m_y نيز عبارت است از بزرگترين فراواني توزيع هاي شرطي yها و N حجم كل مي باشد.

واضح است كه همواره $0 \le \lambda \le 1$ هرچقدر به يك نزديكتر باشد ميزان وابستگي صفات مورد مطالعه بيشتر است.

www.bookgolden.com

مثال:جدول زير نتيجةآمارگيري نيروي انساني دريك شهر بزرگ را نشان مي دهد مطلوب است ضريب كروسكال وضريب گاتمن وتفسير آنها .

علل بيسوادي

جنس	فقر مالي	مخالفت والدين	نبودن مدرسه	عدم علاقه	n_{i_0}
مرد	57	8	11	12	88
زن	27	48	10	7	92
n_j	84	56	21	19	n=180

www.bookgolden.com

حل:

$$E_1=\sum_{j=1}^{4}\left(\frac{n-n.j}{n}\right)n.j=118/5$$

و به همين شيوه

$$E_2=\sum_{i=1}^{2}\sum_{j=1}^{4}\left(\frac{n_{i_0}-n_{ij}}{n_{i_0}}\right)n_{ij}=104/7$$

$$\tau_y=\frac{E_1-E_2}{E_1}=\frac{118/5-104/7}{118/5}=0/12$$

0/12ميزان شدت وابستگي بين علل بيسوادي وجنس را نشان مي دهد،حال ضريب گاتمن را محاسبه مي كنيم.

$$\lambda=\frac{\sum m_y-M_y}{N-M_y}=\frac{(57+48)-84}{180-84}=0/22$$

عدد 0/22 ميزان شدت وابستگي علل بيسوادي وجنس را نشان مي دهد تا حدودي مي توان گفت ، كه علل بيسوادي به جنس مربوط است.

www.booklanpn.com

فصل نهم

رگرسیون

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

رگرسيـــون

از نظر تاريخي كلمه رگرسيون با زمينه فني كنوني‌اش اثر براي اولين بار توسط فرانسيس گالتن ، كسي كه قد پسران و متوسط قد والدينشان را تجزيه و تحليل كرد بكار برده شد.

گالتن از روي مشاهداتش نتيجه گرفت كه قد پسران متعلق به والدين خيلي بلند (يا كوتاه) به طور كلي بلندتر (كوتاهتر) از قد متوسط بوده است ولي‌البته نه بلندي يا كوتاهي والدينشان. اين نتيجه‌گيري در سال 1855 تحت عنوان « رگرسيون به طرف ميانگين » منتشر شد.

در اين متن اصطلاح رگرسيون اشاره بر اين داشت كه قد پسران به جاي ميل به مقادير فرينتر به سوي مقدار متوسط ميل و بازگشت دارد.

www.bookgolden.com

يك مسئله ساده رگرسيون :

داده‌هاي جدول زير مقدار آب داده شده بر حسب سانتي متر و مقدار محصول يونجه بر حسب تن را نشان مي‌دهد .

مسئله پيش بيني محصول يونجه به عنوان تابعي از مقدار آبياري است.

	12	18	24	30	36	42	48
ميزان محصول (y)	5/27	5/68	6/25	7/21	8/02	8/71	8/42

www.bookgolden.com

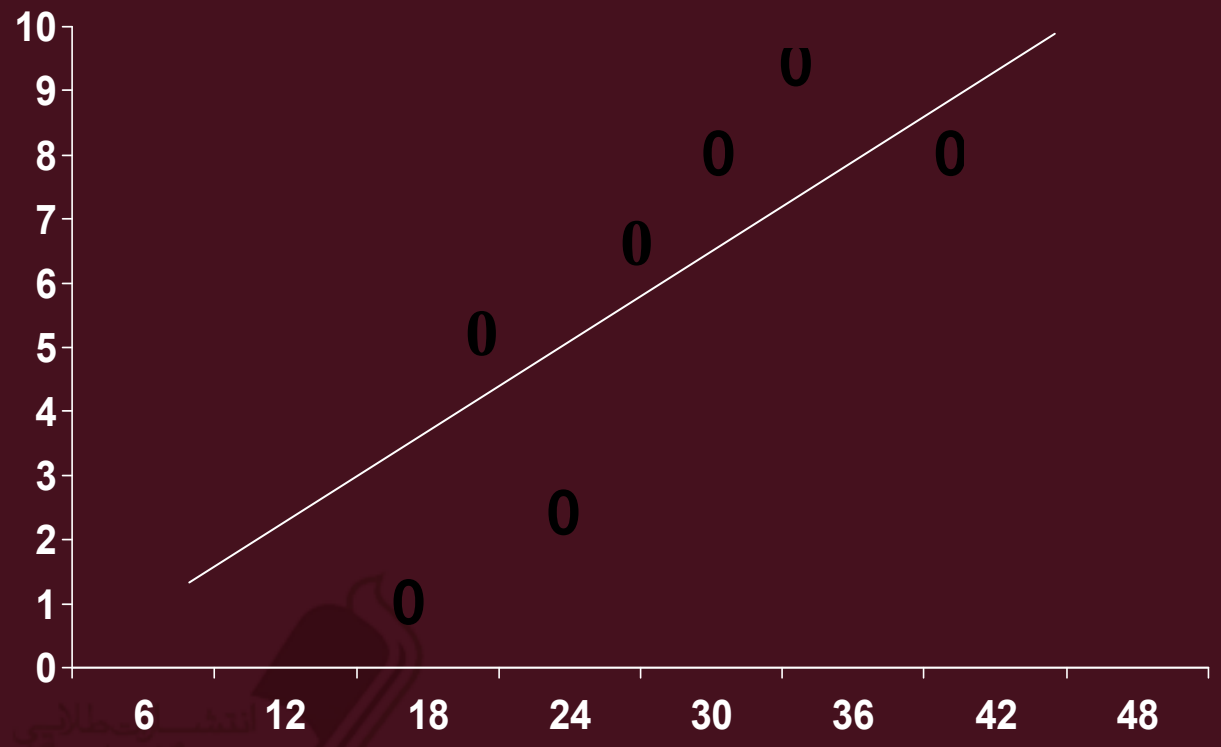
در مطالعه‌ي رابطه بين دو متغير، اولين قدم منطقي رسم داده‌ها به صورت نقاطي بر روي صفحه‌ي مختصات دو بعدي دكارتي مي‌باشد كه شكل حاصل نمودار پراكنش ناميده مي‌شود.

كه چگونگي پراكنش نقاط در اطراف يك خط راست يا يك نوع منحني مشخص را نشان مي‌دهد.

انتشارات طلايي
پويندگان دانشگاه
راهنماي طلايي
تست طلايي
پيك طلايي
www.bookgolden.com

و نيز برداشتن عيني از ميزان پراكندگي داده‌ها پيرامون خط خط يا منحني را فراهم مي‌كند .

بنابراين اطلاع منعكس در نمودار پراكنش براي جستجوي يك مدل رياضي مناسب (خطي يا منحني) مفيد است .



www.bookgolden.com

نمودار پراکنش فوق بازگو کننده‌ي اين موضوع است که نقاط رسم شده روي نمودار پراکنش داده‌ها تقريباً در امتداد يك خط راست پراکنده شده‌اند .

بنابراين مي‌توان به منظور پيشگويي مقدار y از روي مقادير x ، خطي راست به اين مجموعه نقاط برازاند.
چنين خطي در شکل برازش داده شده است ، که مقدار پيشگويي شده‌ي y بعنوان فاصله عمودي نقطه‌ي x تا خط انتخاب مي‌شود.

www.bookgolden.com

اصل حداقل مربعات :

ملاحظه نموديم كه رسم نمودار پراكنش هرگاه نقاط در امتداد يك خط راست پراكنده شده باشند مي‌توان به اين نقاط خط راستي به معادله y=a+bx برازش داد به ازاي

a و b هاي متفاوت مي‌توان بي‌نهايت خط به اين نقاط برازش داد اما درصدد هستيم خط راستي با شيب b و عرض از مبدأ a را با اين ويژگي بيابيم

www.bookgolden.com

$$\left(\sum_{i=1}^{n} \xi_i^2\right)$$

كه به تمام نقاط رسم شده در نمودارپراكنش از هر خط ديگري نزديك‌تر باشد به‌عبارتي مجموع كل مربعات خطاهاي پيش بيني كمترين مقدار را داشته باشد كه به آن اصل حداقل مربعات انحراف مي‌گويند

$$\sum_{i=1}^{n} \xi_i^2$$

شيب خط و عرض از مبدأ چنين خطي كه به ازاي آن مي ني مم شود از رابطه زير محاسبه مي‌شود.

مجموع مربعات خطاهاي پيش بيني

$$= R = \sum_{i=1}^{n} \xi_i^2 = \sum \left(y_i - (a + bx)\right)^2$$

با گرفتن مشتق جزيي نسبت به a و b و حل دستگاه زير

www.bookgolden.com

$$\begin{cases} \dfrac{\delta R}{\delta a} = -2\sum_{i=1}^{n}\left(y_i - \left(a + bx_i\right)\right) = 0 \\ \dfrac{\delta R}{\delta b} = -2\sum_{i=1}^{n} x_i\left(y_i - \left(a + bx_i\right)\right) = 0 \end{cases}$$

داریم

$$\begin{cases} \sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} \quad + b\sum_{i=1}^{n} x_i^2 \end{cases}$$

$$b = \frac{s_{xy}}{s_{xx}}$$

$$a = \overline{y} - b\overline{x}$$

شیب خط

عرض از مبدأ

www.bookgolden.com

که در آن

$$s_{xy} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$s_{xx} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

$$s_{yy} = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

انتشارات طلایی
پویندگان دانشگاه
راهنمای طلایی
تست طلایی
پیک طلایی
www.bookgolden.com

مثال : معادله‌ي خط رگرسيون داده‌هاي مربوط به مقدار آب (x) و ميزان محصول (y) يونجه را تعيين كنيد.

x_i	y_i	x_i^2	$x_i y_i$
12	5 / 27	144	63 / 24
18	5 / 68	324	102 / 24
24	6 / 25	572	150
30	7 / 21	900	216 / 3
36	8 / 02	1296	288 / 72
42	8 / 71	1764	364 / 82
48	8 / 42	2304	404 / 16
مجموع 210	49 / 56	7308	1590 / 48

www.bookgolden.com

$$s_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = \mathbf{1590/48 - 7 \times (30) \times (7/08) = 103/68}$$

$$s_{xx} = \sum x_i^{\,2} - n\bar{x}^{2} = \mathbf{7308 - 7 \times (30)^2 = 1008}$$

$$b = \frac{s_{xy}}{s_{xx}} = \frac{\mathbf{103/68}}{\mathbf{1008}} = \mathbf{0/103}$$

$$a = \bar{y} - b\bar{x} = \mathbf{7/08 - 0/103 \times 30 = 3/99}$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com

بنابراين معادلهي خط رگرسيون عبارت است از :

y=3/99+ 0/103 x

لذا با استفاده از معادله رگرسيون ميتوان به ازاء هر مقدار دلخواه x اندازه متغير وابسته y را پيش بيني كرد.

با استفاده از روابط رياضي ميتوان با استفاده ازضريب همبستگي ، شيب خط رگرسيون را محاسبه نمود .

$$b = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{xx}}}$$

$$b = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}} \cdot \frac{\sqrt{s_{yy}}}{\sqrt{s_{xx}}}$$

$$b = r_{xy} \times \frac{s_y}{s_x}$$

www.bookgolden.com

مثال : فرض كنيد بخواهيم نمره‌هاي دستور زبان انگليسي را از طريق نمره‌هاي درك و فهم پيش بيني كنيم و همبستگي بين اين دو متغير 0/5 و انحراف استانداردهاي نمرات دستور زبان و درك فهم به ترتيب 3 و2 باشد شيب خط رگرسيون را بيابيد.

حل :

$$b_{xy} = r_{x.y} \cdot \frac{s_y}{s_x}$$

$$b_{xy} = \mathbf{0/5} \times \frac{\mathbf{3}}{\mathbf{2}} = \mathbf{0/75}$$

راهنمای طلایی
تست طلایی
پیک طلایی
انتشارات طلایی
پویندگان دانشگاه
www.bookgolden.com