سرفصلهاي آمار و احتمالات

فصل اول: آمار توصيفي

علم آمار: به مجموعه روش هاي علمي اطلاق مي شود، كه براي جمع آوري اطلاعت اوليه، مرتب و خلاصه كردن، طبقه بندي و تجزيه و تحليل اطلاعات اوليه و تفسيي آنها به كار مي‌رود.

جامعه آماري: هر مجموعه از اشياء يا افراد كه لااقل يک صفت مشترک داشته باشند را جامعه آماري گويند.

صفت مشخصه: صفت مشترک بين اعضاي جامعه آماري را صفت مشخصه گويند.

مثال: مجموعه دانشجويان ترم ۵ دانشگاه تهران يک جامعه آماري است كه صفت مشخصه اين جامعه آماري ترم ۵ بودن آنهاست ولي درجامعه آماري هدف بررسي صفات غير مشترک مثل قد، وزن، تاريخ تولد و... مي باشد.

داده هاي آماري: مجموعه مقادير صفت متغير كه بوسيله اعداد يا نشان ها نمايش داده مي شود، را داده هاي آماري مي‌ناميم.

براي مثال اعداد زير با واحد سانتيمتر ميتواند داده‌هاي آماري براي طول قد يک جامعه آماري باشد:

۱۷۰ ۱۶۷ ۱۵۴ ۱۶۱ ۱۵۵ ۱۵۲ ۱۷۸ ۱۶۲ ۱۶۵ ۱۷۱

يا داده هاي زير ميتواند نشانگرگروه خوني جامعه آماري مورد مطالعه باشد.

A B AB O A A AB A

صفات متغير كه در آمارگيري مورد استفاده قرار مي‌گيرند به صورت زير تقسيم بندي مي‌گردند:

- صفت
  - كمي
    واحد اندازه‌گيري دارندو قابل مقايسه و سنجش هستند
    - پيوسته
      مثل: طول، وزن، زمان
    - گسسته
      مثل: تعداد دانشجويان
  - كيفي
    واحد ندارند و قابل مقايسه و سنجش نمي باشند
    - اسمي
      مثل: گروه خوني، اهل يک شهر بودن
    - ترتيبي
      مثل درجه ميوه، پارچه

- آمارگيري
  تمام فعاليتي كه براي جمع آوري داده اي آماري به كار مي رود.
  - سرشماري
    همه افراد جامعه مورد مطالعه قرار مي‌گيرد.
  - نمونه برداري
    بخشي از افراد جامعه طبق اصول خاص* مورد مطالعه قرار مي‌گيرد.

*: بهترين روش نمونه برداري، روش تصادفي است كه بايد به تمام افراد جامعه شانس انتخاب يكسان داد.

آمار توصيفي: آن قسمت از علم آمار كه مشتمل بر خلاصه كردن داده ها در قالب جداول، نمايش ترسيمي آنها بوسيله نمودار و محاسبه شاخصهاي عددي گرايش به مركز، پراكندگي، چولگي و كشيدگي مي باشد را آمار توصيفي نامند.

آمار استنباطي: آن قسمت از علم آمار كه درباره تخمين پارامترهاي جامعه ار روي پارامترهاي نمونه آماري بحث ميكند را آمار استنباطي گويند.

نمايش جدولي داده هاي كيفي:

فرض كنيد گروه خوني ۱۰ دانشجو تعيين و نتايج زير به دست آمده باشد.

A O B AB A B B O O O

نمايش جدولي نتايج به صورت زير مي‌باشد.

گروه خون	فراواني
A	۲
B	۳
O	٤
AB	۱

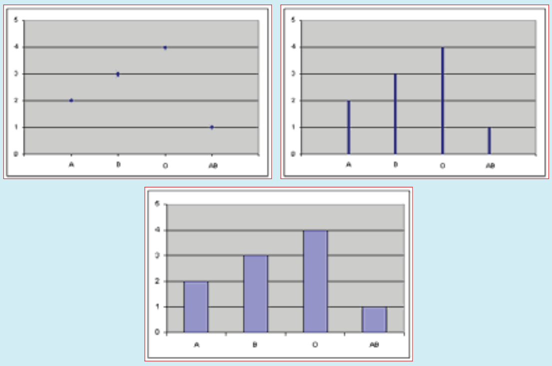
نمودارهاي داده هاي كيفي:

نمودار ميله اي، سوزني، ستوني: اين نمودار در يک دستگاه مختصات كه محور افقي نشان دهنده كيفيت مشاهدات و محور عموديش نشان دهنده فراواني مطلق يا نسبي (فراواني كه در آن نسبت فراواني هر دسته به فراواني كل در نظر گرفته مي‌شود.) هر گروه است ترسيم مي شود. مقدار فرواني را ميتوان با نقطه (سوزني) ميله يا ستون مشخص كرد.

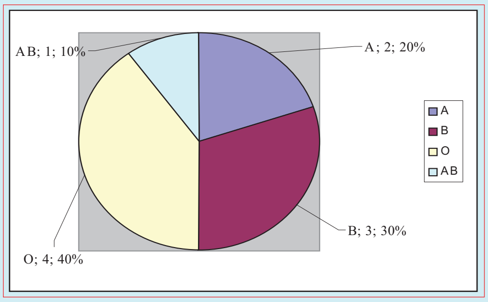
نمودار دايره‌اي: در اين نمودار، دايره را به چند بخش متناسب با فراواني هر دسته تقسيم ميكنيم كه هر بخش براي هر دسته از رابطه زير محاسبه مي گردد.

$$X_i = \frac{360}{\sum f_i} \times f_i$$

نمودار ميله اي، سوزني، ستوني براي مثال گروه خوني



نمودار دايره اي براي مثال گروه خوني



نمايش جدولي داده هاي كمي

اطلاعات به دست آمده از اندازه گيري يا شمارش، همواره به صورت اعداد بيان مي‌شوند كه به آنها **داده هاي خام** مي‌گوييم. داده هاي خام بايد مرتب و طبقه بندي گردند تا قابل تفسير و تجزيه و تحليل آماري شوند.

داده هاي مربوط به نمره درس آمار ۱۰ دانشجو

۵ ۶ ۷ ۱۵ ۱۶ ۱٤ ۱۹ ۱۳ ۱۷ ۱۲

اولين قدم مرتب كردن داده ها از كوچک به بزرگ باشد.

۱۹ ۱۷ ۱۶ ۱۵ ۱۴ ۱۳ ۱۲ ۷ ۶ ۵

حالا با نگاه اجمالي به داده ها مي‌توان اطلاعاتي چون كمترين يا بيشترين نمره را به دست آورد.

در اكثر موارد تعداد داده ها زيادند يا اينكه صفت مورد بررسي از نوع پيوسته اند (قد، وزن، زمان،...) كه در اين صورت داده ها بايد **طبقه بندي** گردند. اولين قدم پيدا كردن پارامترهاي زير است:

• **دامنه:** عبارت است از تفاضل بزرگترين داده از كوچكترين داده كه با R نمايش داده مي شود. در محاسبه پارامترها چون معمولا داده ها گرد ميشوند با توجه به دقت گرد بايد مقادير واقعي مورد استفاده قرار گيرند.

$$\boxed{R = \underset{\forall i}{Max}(X_i) - \underset{\forall i}{Min}(X_i)}$$

تعداد طبقه: با توجه به موضوع مورد بررسي بايد تعداد طبقات تعيين گردد و معمولا بين ۵ تا ۲۰ انتخاب مي‌كنند. تعداد طبقات بايد طوري انتخاب شود كه اطلاعات زياد از دست نرود دو فرمول نيز به صورت زير براي تعيين تعداد طبقات داريم كه اولي به فرمول استورجس معروف است كه در آن N تعداد داده ها و K تعداد طبقات مي‌باشد.

$$K \cong 1 + 3/322 LogN$$
$$K \cong \sqrt{N}$$

فاصله طبقات: از تقسيم دامنه بر تعداد طبقات با تقريب اضافي محاسبه مي شود كه با C نمايش داده مي‌شود.

$$\boxed{C \cong \frac{R}{K}}$$

نمايش جدولي داده‌هاي كمي طبقه بندي شده:

مثال: داده هاي زير طول عمر ۲۵ لامپ بر حسب ساعت مي‌باشد و با دقت كمتر از ۱ گرد شده اند داده ها را طبقه بندي و جدول آن را رسم كنيد؟

۱۰۵ ۹۸ ۱۰۳ ۱۰٤ ۱۱۰ ۱۰٦ ۱۰۳ ۱۱۰ ۱۰۲ ۱۰٤ ۹۷ ۱۰۱ ۱۰۰

۱۰۵ ۱۰٤ ۱۱۰ ۱۰۳ ۱۰۵ ۱۰۹ ۹۸ ۹۹ ۱۰٤ ۱۰۳ ۱۰۹ ۱۰۰

$$R = \underset{\forall i}{Max}(X_i) - \underset{\forall i}{Min}(X_i) = 110.5 - 96.5 = 14$$

$$K \cong \sqrt{N} = \sqrt{25} = 5$$

$$C \cong \frac{R}{K} = \frac{14}{5} \cong 3$$

فراواني تجمعي نسبي	فراواني نسبي	فراواني تجمعي	فراواني مطلق f_i	نماينده دسته	حدود طبقات	حدود واقعي
به درصد نيز بيان ميشود $F_{c_i} = f_{c_i} + F_{c_{i-1}}$	به درصد نيز بيان ميشود $f_{c_i} = \frac{f_i}{N}$	$F_i = f_i + F_{i-1}$	تعداد دفعات تكرار داده ها در هر دسته	ميانگين كران بالا و پايين طبقه $x_i = \frac{L_i + U_i}{2}$	چون داده ها قبلا گرد شده با توجه به دقت گرد حدود واقعي محاسبه ميشود.	
۰,۱۶	۰,۱۶	۴	۴	۹۸	۹۷-۹۹	۹۶,۵-۹۹,۵
۰,۳۲	۰,۱۶	۸	۴	۱۰۱	۱۰۰-۱۰۲	۹۹,۵-۱۰۲,۵
۰,۷۶	۰,۴۴	۱۹	۱۱	۱۰۴	۱۰۳-۱۰۵	....
۰,۸۰	۰,۰۴	۲۰	۱	۱۰۷	۱۰۶-۱۰۸	...
۱	۰,۲	۲۵	۵	۱۱۰	۱۰۹-۱۱۱	....
	$\sum f_{c_i} = 1$		$\sum f_i = 25$			

نمودار هاي داده هاي كمي

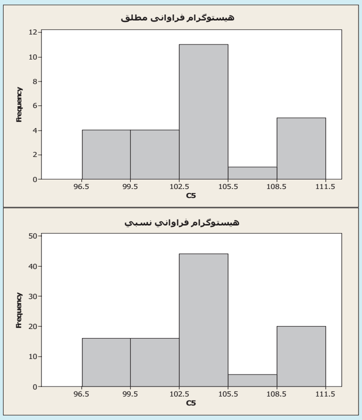
•**هيستوگرام فراواني**: نموداريست در دستگاه مختصات كه محور افقي آن با حدود واقعي طبقات و محور عمودي آن با فراواني مطلق يا فراواني نسبي داده ها مشخص مي شود.

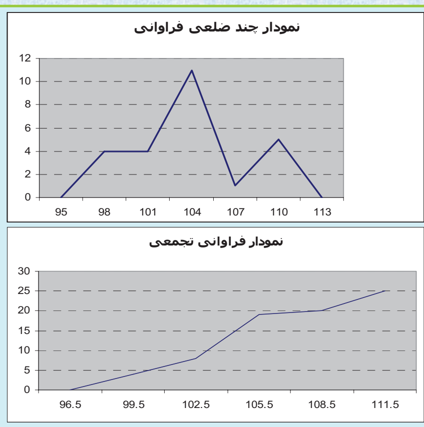
•**چند ضلعي فراواني**: نموداريست كه متناظر با هر نماينده طبقه در محور افقي و فراواني آن در محور عمودي ، يک نقطه در صفحه مختصات ايجاد و به هم وصل مي شوند كه معمولا هيستوگرام فراواني با چند ضلعي فراواني را در يک دستگاه رسم ميكنند..

•**پلي گون فراواني تجمعي**: براي ترسيم اين نمودار ، از نماينده طبقات در محور افقي و فراواني تجمعي در محور عمودي استفاده مي شود ، سپس نقاط ايجاد شده به ترتيب به هم وصل مي شوند.

نمودار هاي داده هاي كمي

نمودار فرواني تجمعي: تنها فرق اين نمودار با نمودار پلي گون فراواني تجمعي در اين است كه در اين نمودار بجاي نماينده طبقات از حدود واقعي استفاده مي شود اين نمودار در محاسبه چندكها (چاركها ، دهكها ، صدكها) و مقايسه پديده هايي (مثل ميزان رشد تورم در كشورها) به كار ميرود.



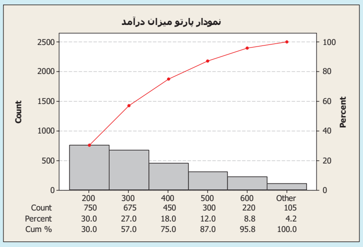


نمودار پارتو: يكي از مهمترين نمودارهاي آماري كه بيشتر براي داده هاي كيفي مي باشد، نمودار پارتو است. اين نمودار داراي سه محور است

۱- محورافقي : نوع موضوعات

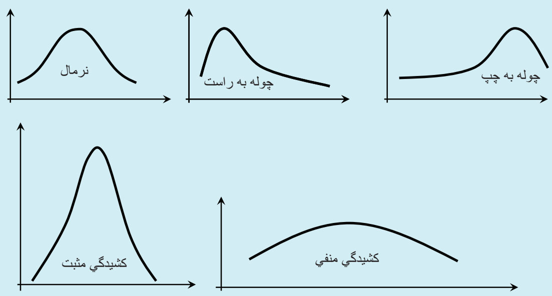
۲- محور عمودي : فراواني مطلق موضوعات

۳- محور سوم (روبروي محور عمودي) : فراواني نسبي تجمعي موضوعات



مثال: اگر ميزان درآمد ۲۵۰۰ نفر از يک نمونه تصادفي با واحد هزار تومان به صورت زير باشد نمودار پارتو را براي آن رسم ميكنيم.

هرگاه تعداد عناصر نمونه زياد يا فاصله طبقات به سمت صفر ميل كند، آنگاه نمودار چند ضلعي فراواني به سمت يک منحني ميل ميکند که آن را منحني فراواني مي نامند و شکل آن بستگي به نوع داده ها دارد كه نمايانگر توزيع جامعه آماري مي‌باشد که به تفصيل در آينده مورد بحث قرار مي‌گيرد.



پارامترهاي آماري داده‌ها:
براي نشان دادن وضعيت داده ها از نقطه نظرات متفاوت به كار مي‌روند

- پارامترهاي مركزي: يک مقدار خاص كه داده‌هاي آماري تجمع و فشردگي در اطراف آن دارند.
- پارامترهاي پراكندگي: پراكندگي داده‌ها را در اطراف پارامترهاي مركزي (معمولا ميانگين) را نشان مي دهند.
- پارامترهاي چولگي: ميزان انحراف دم توزيع جامعه به سمت راست (چوله به راست) يا چپ (چوله به چپ) را نشان مي‌دهند.
- پارامترهاي كشيدگي: اين پارامترها براي مقايسه توزيع جوامع مورد نظر با توزيع جامعه نرمال به لحاظ كشيدگي (كوتاهي و بلندي توزيع) مورد استفاده قرار مي گيرد

پارامترهاي مركزي:

قرارداد:در محاسبه پارامترها دو حالت وجود دارد اگر داده ها طبقه بندي شده باشند كه با B در غير اينصورت با A نشان مي دهيم.

•ميانگين حسابي: در بين پارامترهاي مركزي نوسان كمتري دارد و از ثبات بيشتري برخوردار است.

$$A)\ \mu=\bar{x}=\frac{1}{N}\sum_{i=1}^{N}X_i$$

$$B)\ \bar{x}=\frac{1}{N}\sum_{i=1}^{k}f_i x_i$$

•ميانگين حسابي وزني: اگر داده‌هاي $X_1, X_2, \ldots, X_N$ داراي ضريب وزني $w_1, w_2, \ldots, w_N$ آنگاه

$$\bar{x}=\sum_{i=1}^{N}w_i x_i \Big/ \sum_{i=1}^{N}w_i$$

پارامترهاي مركزي:

•ميانگين هندسي: اين ميانگين حد متوسط شاخصها، نسبتها و درصدها را بيان ميكند مثلا افزايش جمعيت، كشت باكتري، تجزيه راديواكتيو..

براي محاسبه معمولا از log استفاده مي كنيم.

$$A)\ G=\sqrt[N]{x_1\times x_2\times\ldots\times x_n}=\sqrt[N]{\prod_{i=1}^{N}x_i}$$

$$B)\ G=\sqrt[K]{x_1^{f_1}\times x_2^{f_2}\times\ldots\times x_N^{f_k}}=\sqrt[K]{\prod_{i=1}^{K}x_i^{f_i}}$$

پارامترهاي مركزي:

•ميانگين همساز (هارمونيک): اگر هيچكدام از داده ها صفر نباشد ميانگين همساز از رابطه زير محاسبه مي‌گردد.

اين ميانگين براي محاسبه حد متوسط سرعتها، مطالعه در شبكه‌هاي برق و عينک شناسي بكار مي رود.

$$A)\ H=N\Big/\sum_{i=1}^{N}\frac{1}{x_i}$$

$$B)\ H=N\Big/\sum_{i=1}^{k}\frac{f_i}{x_i}$$

پارامترهاي مركزي:

•ميانگين درجه دوم:

$$A)\ Q=\sqrt{\frac{1}{N}\sum_{1}^{N}x_i^2}$$

$$B)\ Q=\sqrt{\frac{1}{N}\sum_{1}^{K}f_i x_i^2}$$

•رابطه بين ميانگينهاي بيان شده به صورت زير است

$$H\le G\le\bar{X}\le Q$$

•ميانگينهاي پيراسته و وينزوري بعد از معرفي چاركها بيان مي‌شود.

پارامترهاي مركزي:

•میانه: ابتدا داده ها را بـه طـور غیرنزولي مرتـب مـيكنیم اگر تعداد دادهها فرد باشد میانه عدد وسطي است در غیر اینصورت میانگین حسابي دو عدد وسطي معرف میانه میباشد. در دادههاي طبقه بندي شده از رابطه زير بدست مي آيد.

i: شماره طبقه میانه دار و عبارتست از اولـین طبقـه اي کـه فراوانـي تجمعـي آن بزرگتـر یـا مساوي $N/2$ باشد.

$$M_d = L_i + \frac{\frac{N}{2} - F_{i-1}}{f_i} \times C$$

میانـه مشاهدات را بـه دو بخش مساوي تقسیم مي کند و تحت تأثیر داده هاي پرت قرار نميگیرد.

پارامترهاي مركزي:

•مد یا نما: اندازه اي از متغیر است که فراواني آن ماکسیمم باشد که در داده هاي کیفي مهمترین شاخص مرکزي مد ميباشد.

i: شـماره طبقـه مـد دار و عبارتـست از طبقه اي که فراواني آن ماکسیمم است.

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \times C$$

رابطه تجربي بین میانگین، میانه و مد:

$\bar{X} \le M_d \le M_o$ $M_o \le M_d \le \bar{X}$ $\bar{X} = M_o = M_d$

در توزیعهایي که چولگي زیاد نباشد رابطه تجربي زیر که به رابطه پیرسن معروف است، برقرار مي باشد.

{میانگین – مد = ۳ (میانگین – میانه)}

پارامترهاي مركزي:

•چندکها: هرگاه داده ها را به طور غیر نزولي مرتب کنیم. عددي را که لااقل p درصد دادهها کوچکتر از آن و (p-۱۰۰) درصد دادها بزرگتر از آن باشند، صدک Pام نامند. صدک ۵۰ام، میانه است حال صدک هاي معروف را نام مي بریم.

•چارک ها: آن ها را Q_1، Q_2، Q_3 نشان مي دهند که بـه ترتیب برابر است با P_{25}، P_{50}، P_{75}. در اینجا نیز Q_2 همان میانه است.

•دهکها: آن را با D_1، D_2، ... ، D_9 نشان میدهیم که که بـه ترتیب برابر است با P_{10}، P_{20}، ...، P_{90}. در اینجا نیز D_5 همان میانه است.

محاسبه صدکها: داده ها را بهطور غیر نزولي مرتب ميكنیم اگر N تعداد کل داده ها باشد براي محاسبه صدک pام ابتدا $(N+1)p$ را محاسبه میکنیم.

$$[(N+1)p] = K$$
$$\text{if } K \in \mathbf{Z} \Rightarrow \mathbf{x}_K = P_p \qquad \text{else}$$
$$K = k + r \qquad 0 < r < 1$$
$$P_p = (1-r)x_k + r x_{k+1}$$

اگر داده ها طبقه بندي باشند:

$$P_p = L_i + \frac{(N \times p/100) - F_{i-1}}{f_i} \times C$$

که در آن i طبقه صدک دار مي باشد و عبارتست از اولین طبقهاي که فراواني تجمعي آن بزرگتر یا مساوي $NP/100$ میباشد.

پارامترهاي مركزي:

•میانگین پیراسته: براي محاسبه آن تمام داده هاي بزرگتر از چارک سوم و کوچکتر از چارک اول را کنار میگذاریم سپس میانگین حسابي باقي دادهها را حساب میکنیم.

•میانگین وینزوري: به جاي تمام داده هاي کوچکتر از چارک اول، مقدار چارک اول و بجاي تمام داده هاي بزرگتر از چارک سوم مقدار چارک سوم را قرار مي دهیم سپس میانگین حسابي مجموعه دادههاي جدید را حساب میکنیم.

پارامترهاي پراكندگي

•دامنه تغييرات: عبارتست از اختلاف بزرگترين و كوچكترين داده كه از اهميت كمتري برخوردار است چون تحت تاثير دو داده از مجموعه داده‌ها قرار دارد.

$$R = \underset{\forall i}{Max}(X_i) - \underset{\forall i}{Min}(X_i)$$

•انحراف چاركها: به صورت زير تعريف مي‌شود. كه اين پارامتر تحت تاثير ۵۰ درصد داده‌هاست

$$Q = \frac{1}{2}(Q_3 - Q_1)$$

پارامترهاي پراكندگي

•انحراف متوسط يا انحراف از ميانگين: به صورت زير تعريف مي‌شود كه تحت تاثير تمام داده هاست. مشكلي كه اين پارامتر دارد وجود عبارت قدر مطلق مي‌باشد كه محاسبات را پيچيده ميكند.

$$A)\ A.D = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - \bar{X}\right|$$

$$B)\ A.D = \frac{1}{N}\sum_{i=1}^{K} f_i \left|x_i - \bar{X}\right|$$

پارامترهاي پراكندگي

•واريانس: در يک نمونه به صورت زير تعريف مي‌شود كه تحت تاثير تمام داده هاست و پيچيدگي محاسبات قدر مطلق نيز در آن نيست.

$$A)\ S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$$

$$B)\ S^2 = \frac{1}{N-1}\sum_{i=1}^{K} f_i (x_i - \bar{X})^2$$

حال فرض كنيد داده ها مربوط به طول قد افراد برحسب سانتي‌متر باشد ولي واحد $(x_i - \bar{X})^2$ سانتي‌مترمربع خواهد شد كه ايراد واريانس نيز در اين مورد مي باشد.

پارامترهاي پراكندگي

•چنانچه مطالعات روي جامعه باشد و يا تعداد نمونه زياد گردد در اين صورت ميانگين را با μ و واريانس با σ^2 نمايش داده و به شكل زير محاسبه مي گردد.

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

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

پارامترهاي پراكندگي

•انحراف معيار: جذر مثبت واريانس را انحراف معيار مي‌ناميم كه مشكل مربع شدن واحد در واريانس را نيز ندارد و به عنوان بهترين پارامتر پراكندگي مي‌باشد.

انحراف معيار نمونه $\longleftarrow S = \sqrt{S^2}$

انحراف معيار جامعه $\longleftarrow \sigma = \sqrt{\sigma^2}$

•ضريب تغييرات: براي مقايسه پراكندگي دو جمعيت استفاده مي گردد و مزيت آن در بدون واحد بودن آن است.

$$C = \frac{\sigma}{\bar{X}} \times 100\%$$

پارامترهاي پراكندگي

•گشتاور: يكي ديگر از پارامترهاي پراكندگي مي باشد. گشتاور مرتبه rام پيرامون نقطه ثابت a را براي جامعه به شكل زير نمايش و تعريف مي‌كنند.

$$A)\ m_r^a = \frac{1}{N}\sum_{i=1}^{N}(x_i - a)^r$$

$$B)\ m_r^a = \frac{1}{N}\sum_{i=1}^{K} f_i (x_i - a)^r$$

•گشتاور حول ميانگين جامعه را گشتاور مركزي نامند و آن را با μ_r نشان مي‌دهند

$$A)\ m_r^{\mu} = \mu_r = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^r$$

$$B)\ m_r^{\mu} = \mu_r = \frac{1}{N}\sum_{i=1}^{K} f_i (x_i - \mu)^r$$

پارامترهاي پراكندگي

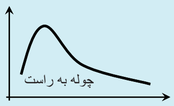
•واضح است كه واريانس جامعه برابر است با گشتاور مركزي مرتبه دوم.

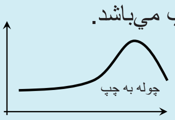
$$\mu_2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \sigma^2$$

•گشتاور حول مبدا را نيز به اختصار با μ'_r نشان مي‌دهند در نتيجه رابطه گشتاور به صورت زير در مي‌آيد.

$$\mu'_r = \frac{1}{N}\sum_{i=1}^{N}(x_i)^r$$

پارمترهاي چولگي: آن را با SK نشان ميدهند و اگر دم توزيع جامعه به سمت راست باشد، توزيع را چوله به راست و در صورت عكس، آن را چوله به چپ مي‌نامند. اين مقدار طوري تعريف گرديده است كه در صورت صفر بودن گويند توزيع بدون چولگي است و نرمال مي‌باشد در غير اينصورت با مثبت بودن مقدار چولگي بسته به عدد چولگي، توزيع چوله به راست مي‌باشد و در صورت منفي بودن، باز بسته به مقدار آن توزيع چوله به چپ مي‌باشد.





پارامترهاي چولگي:

•ضريب چولگي گشتاوري:

$$SK = \frac{\mu_3}{\sigma^3}$$

•ضريب هاي چولگي پيرسون:

$$1)\ SK = \frac{\mu - M_o}{\sigma}$$

$$2)\ SK = \frac{3(\mu - M_d)}{\sigma}$$

•ضريب هاي چولگي چندكي:

$$1)\ SK = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

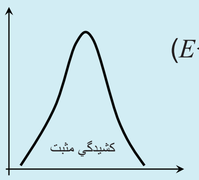
$$2)\ SK = \frac{P90 - 2P50 + P10}{P90 - P10}$$

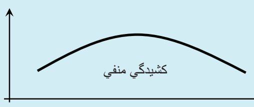
پارامترهاي كشيدگي: اين پارامترها براي مقايسه توزيع جوامع مورد نظر با توزيع جامعه نرمال به لحاظ كشيدگي (كوتاهي و بلندي توزيع) مورد استفاده قرار مي‌گيرد كه آن را با E نشان مي‌دهيم.كه باز سه حالت زير پيش مي‌آيد:

•مساوي توزيع نرمال $(E=0)$

• بلندتر از توزيع نرمال $(E>0)$

•كوتاه تر (پهن‌تر) از توزيع نرمال $(E<0)$





پارامترهاي كشيدگي:

• ضريب كشيدگي گشتاوري

$$E = \frac{\mu_4}{\sigma^4} - 3$$

روش كد گذاري در محاسبه پارامترها:

در بعضي مواقع در محاسبه پارامترها راحتتر اين است كه مشاهدات را تغيير دهيم و به جاي محاسبه پارامترها با داده هاي اصلي، با داده‌هاي جديد حساب كنيم ولي اين تغيير بايد طوري گردد كه بتوان با پارامتر محاسبه شده به پارامتر اصلي برسيم.

يكي از اين تغييرات به شكل زير است

اگر x_i ها مجموعه داده هاي ما در جامعه آماري باشند (در داده‌هاي طبقه بندي شده نماينده طبقات در نظر گرفته مي‌شود) داده‌هاي جديد را به صورت زير در نظر مي‌گيريم:

$$u_i = \frac{x_i - A}{C}$$

روش كد گذاري در محاسبه پارامترها:

كه در آن A و C اعداد ثابتي هستند. در دادههاي طبقه بندي شده C را فاصله طبقات A را x_i مربوط به نماينده طبقه وسط يا طبقه مد دار در نظر مي گيرند. حال داريم:

$$u_i = \frac{x_i - A}{C} \Rightarrow x_i = C u_i + A$$
$$\bar{X} = C \bar{U} + A$$
$$S_x^2 = C^2 S_u^2$$
$$S_x = C^2 S_u$$
$$m_r^a(x) = C^r m_r^a(u)$$

تمرين: جدول زير مربوط به آب مصرفي ۲۰ خانوار ميباشد ميباشد مطلوبست:

- تكميل جدول با پارامترهاي نماينده ، فراواني تجمعي، فراواني نسبي و فراواني تجمعي نسبي هر طبقه.
- رسم نمودار هيستوگرام و چندضلعي فراواني
- رسم نمودار دايره اي بر اساس درجه ميزان مصرف خانوار
- محاسبه ميانگين، مد و ميانه دادها
- محاسبه انحراف چارك ها، واريانس و ضريب تغيرات
- محاسبه ضريب چولگي پيرسون

استفاده از روش كد گذاري در محاسبه پارامترها مجاز مي باشد.

حدود طبقات	فراواني	درجه ميزان مصرف
۱۰ - ۱۲	۲	بسيار كم مصرف
۱۳ – ۱۵	۴	كم مصرف
۱۶ – ۱۸	۶	مصرف متوسط
۱۹ – ۲۱	۳	پر مصرف
۲۲ - ۲۴	۵	بسيار پر مصرف

تمرين: جدول زير مربوط به طول قد ۱۰۰ دانشجو ميباشد مطلوبست:

- تكميل جدول با پارامترهاي نماينده ، فراواني تجمعي، فراواني نسبي و فراواني تجمعي نسبي هر طبقه.
- رسم نمودار هاي هيستوگرام و چند ضلعي فراواني
- قد چند درصد از دانشجويان بالاي ۱۶۸ سانتي متر مي باشد.
- محاسبه ميانگين، مد و ميانه قد دانشجويان
- محاسبه انحراف چارك ها، انحراف از ميانگين واريانس و انحراف معيار
- محاسبه ضريب چولگي

استفاده از روش كد گذاري در محاسبه پارامترها مجاز مي باشد.

حدود طبقات	فراواني
۱۵۰ - ۱۵۶	۱۵
۱۵۷ – ۱۶۳	۲۰
۱۶۴ – ۱۷۰	۳۰
۱۷۱ – ۱۷۷	۲۵
۱۷۸ - ۱۸۴	۱۰

تمرين: فرض كنيد فاصله سه شهر A, B, C از يكديگر برابر باشد، اتومبيلي فاصله بين A تا B را باسرعت ۳۰ كيلومتر در ساعت و فاصله B تا C را با سرعت ۴۰ كيلومتر در ساعت و فاصله C تا A را با سرعت ۵۰ كيلومتر در ساعت پيموده است، سرعت متوسط حركت اين اتومبيل را حساب نماييد.

تمرين: فرض كنيد ميزان توليد كارخانه اي در چهار سال متوالي ۲، ۴، ۶ و ۲۷ برابر نسبت به سال قبل باشد. مطلوبست ميزان افزايش متوسط توليد كارخانه

تمرين: ميانگين درجه دوم داده هاي زير را حساب كنيد.

۳	۵	۶	۶	۷	۱۰	۱۲

فصل دوم: نظريه مجموعهها،

آناليز تركيبي

مجموعه: عبارتست از يک دسته از اشياء كه كاملا مشخص باشند و بطوري كه هر شئي مفروض، يا متعلق (عضو) به مجموعه هست و يا نيست.

•بحثهاي نمودار ون، اشتراك، اجتماع، زير مجموعه، متمم مجموعه، مكمل مجموعه، مجموعه مرجع، دو مجموعه جدا از هم، تفاضل دو مجموعهاز مفاهيم اصلي نظريه مجموعه ها هستند.

•خاصيتهاي جاجايي، شركت پذيري، توزيع پذيري و قوانين دمرگان نيز از خواص مجموعه ها ميباشند.

1) $A \cup B = B \cup A$

2) $A \cup (B \cup C) = (A \cup B) \cup C \qquad A \cap (B \cap C) = (A \cap B) \cap C$

3) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

4) $(\bigcup_{i=1}^{n} A_i)' = \bigcap_{i=1}^{n} A_i' \qquad (\bigcap_{i=1}^{n} A_i)' = \bigcup_{i=1}^{n} A_i'$

فضاي نمونه:

آزمايش تصادفي: فعاليتي که نتيجه آن از قبل مشخص نيست ولـي کل حالات ممکن آن معلوم است، مثل پرتاب يک سکه، که معلوم نيست شير خواهد آمد يا خط. چون در نظريه احتمال فقط آزمايشهاي تصادفي مورد نظر مي‌باشد لذا براي سادگي کلمه تصادفي ذکر نمي گردد.

پيشامد تصادفي: پيش آمدي که در اثر يک آزمايش تصادفي مـي تواند رخ دهد يا رخ ندهد.

فضاي نمونه: مجموعه پيامدهاي ممکن يک آزمايش تصادفي را فضاي نمونه آن آزمايش مي گويند که آن را با S نشان مي‌دهند.

•تعداد اعضاي فضاي نمونه را با $n(S)$ نشان مي‌دهيم.

فضاي نمونه محدود: يعني اين تعداد اعضاي فضاي نمونه آزمايش متناهي باشد.

فضاي نمونه نامحدود: يعني اينکه تعداد اعضاي فضاي نمونه آزمايش نامتناهي است.

•هر عضو فضاي نمونه را نقطه نمونه مي ناميم.

•پيشامدي که شامل يک نقطه نمونه باشد را پيشامد ساده و اگر بيش از يک نقطه نمونه داشته باشد را پيشامد مرکب مي ناميم.

•پيشامد $A=S$ را پيشامد حتمي و پيشامد تهي را پيشامد غير ممکن ناميم.

تمرين: سکه را سه بار پرتاپ مي کنيم، فضاي نمونه آن را مشخص کنيد؟

تمرين: يک سکه را با تاس همزمان پرتاپ مي کنيم فضاي نمونه آزمايش را مشخص کنيد.

آناليز ترکيبي:

کاربردهاي قواعد شمارش: از اين قواعد در وضعيت هايي استفاده مي شود که فهرست نمودن تمام حالات ممکن آزمايش مقدور نمي‌باشد، يا نيازي به فهرست نمودن آنها نمي‌باشد، لذا فقط به ذکر تعداد حالات ممکن و مختلف اکتفا مي شود.

اصل اساسي شمارش: اگر عملي به m_1 طريق و براي هر کدام از آنها عمل ديگري را به m_2 طريق و براي هر يک از اين دو عمل سومي را به m_3 طريق و ... و عمل K امي را به m_k طريق بتوان انجام داد، آنگاه اين $m_1 \times m_2 \times \dots \times m_k$ عمل را باهم به طريق ميتوان انجان داد.

مثال ۱: سه سکه را پرتاپ ميکنيم مطلوبست تعداد اعضاي فضاي نمونه؟

مثال ۲: دو تاس را با چهار سکه پرتاپ مي کنيممطلوبست تعداد اعضاي فضاي نمونه؟

مثال ۲: چند عدد سه رقمي از ارقام ۴، ۳، ۲، ۱، ۰ ميتوان نوشت بطوريکه تکرار ارقام مجاز باشد؟

تمام ترتيب هاي ممکن دسته اي از اشياء و يا قسمتي از آن را تبديل يا جايگشت گوييم.

قضيه: تعداد تبديلهاي n شئي متمايز برابر با $n!$

مثال: به چند طريق مي توان يک صف ۵ نفري براي سوار شدن به اتوبوس تشکيل داد

مثال: چند عدد سه رقمي از ارقام ۴، ۳، ۲، ۱، ۰ ميتوان نوشت بطوريکه تکرار ارقام مجاز نباشد؟

قضيه: تعداد تبديلهاي rتايي از n شئي متمايز برابر است با

$$P_n^r = \frac{n!}{(n-r)!} = n\,(n-1)\,(n-2)\dots(n-r+1)$$

قضيه: تعداد تبديلهاي دوري (تبديلهايي که بوسيله يک دسته از اشيا روي محيط دايره مرتب مي‌شوند)، n شئي متمايز برابر است با $(n-1)!$

قضيه: تعداد تبديلهاي مختلف n شئي که n_1 شئي آن از نوع اول، n_2 شئي آن از نوع دوم و .... و n_r شئي آن از نوع r باشد، برابر است با

$$\binom{n}{n_1,n_2,\dots,n_r} = \frac{n!}{n_1!n_2!\dots n_r!} \quad , \quad n_1+n_2+\dots+n_r=n$$

مثال ۱: به چند طريق ميتوان ۳ لامپ قرمز، ۲ لامپ سبز و ۴ لامپ آبي را روي يک صفحه نصب نمود.

مثال ۲: به چند طريق ۵ نفر مي توانند در اطراف يک ميز دايره اي بنشينند؟

مثال ۳: ۶ زوج مختلف با هم دور يک ميز دايره اي مي نشينند تعداد حالت آن را در صورتيکه هر زوج کنار هم بنشينند را بيابيد؟

مثال ۴: به چند طريق مي توان ۷ نفر با يک وسيله نقليه که به ترتيب گنجايش ۲، ۳ و ۷ نفر را دارند، از محلي به محل ديگر منتقل نمود؟

چگونگي انتخاب r شئي از n شئي بدون در نظر گرفتن ترتيب، ترکيب ناميده مي‌شود.

قضيه: تعداد ترکيبهاي r تايي از n شئي متماير برابر با:

$$C_r^n = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

مثال: از يک گروه مرکب از ۵ پزشک و ۳ پرستار، چند کميته ۳ نفره مي توان تشکيل داد؟

$$\binom{8}{3} = \frac{8!}{3!\,(5)!} = 56$$

تمرين: از ۱۰ دستگاه تلويزيون موجود در يک فروشگاه ۳ تلويزيون نقص فني دارند تعداد حالات انتخاب ۴ دستگاه از اين تلويزيون ها بطوريکه حداقل ۲ تلويزيون نقص فني داشته باشند؟

تمرين: به چند طريق ۳ مرد و ۲ زن ميتوانند در يک صف قرار گيرند، اگر

- محدوديتي نداشته باشيم.
- زن ها کنار هم باشند.
- مردها کنار هم و زن ها کنار هم باشند.

تمرين: فرض کنيد ۱۱ دوست صميمي داريد، به چند طريق ميتوان ۵ نفر از آن‌ها را به مهماني دعوت کرد اگر

- محدوديتي نداشته باشيم.
- دو نفر از آنها بخواهند با هم در مهماني شرکت کنند.
- دو نفر از آنها نمي خواهند با هم در مهماني شرکت کنند.

فصل سوم: احتمال

مفهوم احتمال: احتمال يعني شانس وقوع يک پيشامد خاص و احتمال وقوع يک پيشامد برابر است با نسبت دفعاتي که پيشامد خاصي در تکرارهاي زياد رخ مي دهد.

در نظريه احتمالات به هر نقطه از فضاي نمونه متناهي عددي نسبت داده مي شود که «وزن» آن نقطه ناميده مي‌شود.

مجموع وزن‌هاي تمام نقاط موجود در پيشامد A را «احتمال پيشامد A» گوييم و آن را با $P(A)$ نشان مي دهيم. با توجه به تعريف پيشامد A، داريم

1) $P(\Phi)=0$

2) $P(S)=1$

3) $\forall A \subseteq S \Rightarrow 0 \le P(A) \le 1$

پیشامدهاي A_1، A_2، ... ، A_n را پیشامدهاي ناسازگار گوییم اگر وقوع همزماني هر دو پیشامدي، غیر ممکن باشد.

$$A_i \cap A_j = \Phi \quad \forall i \neq j, 1 \leq i, j \leq n \Rightarrow$$
$$P(A_i \cap A_j) = 0$$

اگر پیشامدهاي A_1، A_2، ... ، A_n همه نتایج ممکن در اثر آزمایش را در برگیرند، در اینصورت پیشامدها را فرسا گوییم.

اگر شرایط آزمایش به گونه اي باشد که احتمال وقوع هر یک از پیشامدهاي A_1، A_2، ... ، A_n برابر باشند، در اینصورت پیشامدها را همتراز یا همشانس گوییم.

اگر پیشامدهاي A_1، A_2، ... ، A_n هر سه خاصیت ناسازگار، فرسا و همترازي را داشته باشند در نظریه احتمال آنها را حالتها یا شانسها نامیده و ميگوییم توسط مدل کلاسیک بیان ميشوند.

اگر یک آزمایش را بتوان توسط مدل کلاسیک بیان نمود، در اینصورت احتمال هر پیشامد A در این آزمایش برابر است با

$$\boxed{P(A) = \frac{n(A)}{n(S)}}$$

اگر آزمایش تصادفي، یک آزمایش با فضاي نمونه پیوسته باشد تعریف احتمال براي یک پیشامد دلخواه به صورت زیر در ميآید.

اگر آزمایش مربوط به طول باشد: $\longrightarrow P(A) = \frac{L(A)}{L(S)}$

اگر آزمایش مربوط به طول باشد: $\longrightarrow P(A) = \frac{S(A)}{S(S)}$

اگر آزمایش مربوط به حجم باشد: $\longrightarrow P(A) = \frac{V(A)}{V(S)}$

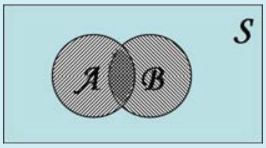
مثال: سکه سالمي را سه با پرتاپ مي کنیم مطلوبست احتمال اینکه دقیقا یکبار شیر بیاید

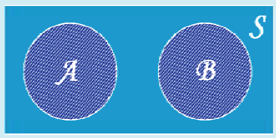
مثال: جعبه اي محتوي ۶ توپ آبي و ۱۲ توپ زرد مي باشد، سه توپ بدون جایگذاري به تصادف برمي داریم مطلوبست:

- احتمال اینکه هر سه توپ آبي باشند
- دو توپ آبي و دیگري زرد باشد
- حداقل دو توپ آبي باشد.

قضیه: اگر A و B دو پیشامد باشند آنگاه

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

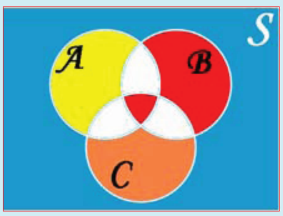




$$if \ A \cap B = \Phi \Rightarrow P(A \cup B) = P(A) + P(B)$$

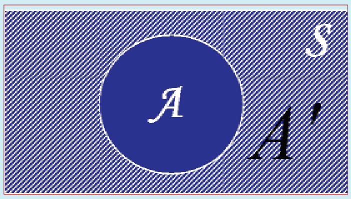
قضیه: اگر A و B و C سه پیشامد باشند آنگاه

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$
$$- P(A \cap B) - P(A \cap C) - P(B \cap C)$$
$$+ P(A \cap B \cap C)$$



قضيه: اگر A و A' دو پيشامد متمم باشند آنگاه

$$P(A')=1-P(A)$$



تمرين: ظرفي محتوي ۲۰ كارت در چهار رنگ مختلف است. بطوريكه از هر رنگ، ۵ كارت و كارتهاي هم رنگ از ۱ تا ۵ شماره گذاري شده‌اند. ۲ كارت به تصادف بر مي داريم مطلوبست احتمال اينكه ۲ كارت داراي يک شماره باشند؟

تمرين: از ظرفي محتوي ۳ مهره سفيد و ۴ مهره سياه، ۳ مهره به تصادف خارج مي كنيم، مطلوبست احتمال اينكه:

أ- ۳ مهره همرنگ باشند

ب- ۳ مهره همرنگ نباشند

احتمال شرطي: احتمال رخ دادن پيشامد B، مشروط بر آنكه بدانيم A رخ داده است را احتمال B به شرط A گويند و به صورت زير تعريف مي گردد.

$$P(A|B)=\frac{P(A\cap B)}{P(A)} \qquad P(A)\neq 0$$

قضيه: اگر وقوع پيشامدهاي A و B بطور همزمان امكان‌پذير باشد، آنگاه

$$P(A\cap B)=P(A|B)P(B)$$
$$P(A\cap B)=P(B|A)P(A)$$

اگر رخ دادن پيشامد A، اثري در رخ دادن B نداشته باشد در اينصورت دو پيشامد را مستقل گوييم.

دو پيشامد A و B را مستقل گوييم، اگر و تنها اگر

$$P(A\cap B)=P(A)P(B)$$

پيشامدهاي A_1، A_2، ... ،A_n مستقل گوييم اگر و تنها اگر

$$P(A_i\cap A_j)=P(A_i)P(A_j) \qquad \forall\, i\neq j$$
$$P(A_i\cap A_j\cap A_k)=P(A_i)P(A_j)P(A_k) \qquad \forall\, i\neq j\neq k$$
$$P(\bigcap_{i=1}^{n}A_i)=\prod_{i=1}^{n}P(A_i)$$

مثال: دو تاس را پرتاپ مي كنيم، در صورتيكه بدانيم مجموع اعداد ظاهر شده برابر ۶ است مطلوبست احتمال اينكه يكي از تاسها عدد ۲ را نشان دهند؟

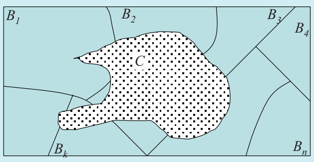
مثال: از جعبه اي محتوي ۴ مهره سفيد و ۳ مهره سياه، ۲ مهره را با جايگذاري خارج مي كنيم (يک مهره را برداشته پس از مشاهده رنگ آن، مهره را به جعبه برمي گردانيم) مطلوبست محاسبه احتمال اينكه:

أ- هر دو مهره سفيد باشد.

ب- مهره اول سفيد و مهره دوم سياه باشد.

قضيه (بيز): فرض كنيد مجموعه $\{B_1, B_2, \dots, B_n\}$ يک افراز از فضاي نمونه S باشد. اگر C پيشامدي دلخواه از S باشد و $P(C)$ مخالف صفر باشد آنگاه به ازاي $K= 1, 2, \dots, n$ داريم:

$$P(B_k|C)=\frac{P(B_k\cap C)}{\sum_{i=1}^{n}P(B_i\cap C)}=\frac{P(B_k)P(C|B_k)}{\sum_{i=1}^{n}P(B_i)P(C|B_i)}$$



مثال: کارخانه اي داراي سه ماشين است که ۵۰%، ۳۰% و ۲۰% محصول آن کارخانه را توليد مي کنند و ميدانيم درصد کالاهاي معيوب اين سه ماشين به ترتيب ۳%، ۴% و ۵% است. مطلوبست محاسبه احتمال اينکه اگر کالاي انتخاب شده معيوب باشد، اين کالا توسط ماشين اول توليد شده باشد:

تمرين: از شش زوج خواهر و برادر متفاوت، ۲ نفر را به تصادف انتخاب ميکنيم، مطلوبست محاسبه احتمال اينکه اين دو نفر

الف: با هم خواهر و برادر باشند.

ب: يکي مرد و يکي زن باشد.

تمرين: در کلاسي که ۱۵ دانشجو دارد در مورد موضوع خاصي، ۹ نفر موافق، ۴ نفر موافق و ۲ نفر ممتنع هستند، ۳ نفر را به تصادف انتخاب مي کنيم و نظرشان را ميپرسيم مطلوبست محاسبه احتمال اينکه:

الف: حداقل دو نفر موافق باشند.

ب: دونفر اول موافق و نفر سوم مخالف باشد.

تمرين: از جعبه اي که محتوي ۵ مهره سفيد و ۲ مهره سياه است، يک مهره را به تصادف خارج کرده و بدون آنکه رنگ آن را مشاهده کنيم آنرا کنار ميگذاريم و سپس مهره ديگري را خارج ميکنيم، مطلوبست محاسبه احتمال اينکه اين مهره سفيد باشد.

تمرين: جعبه A محتوي ۶ توپ سفيد و ۴ توپ سياه است و جعبه B محتوي ۲ توپ سفيد و ۲ توپ سياه ميباشد. از جعبه A، ۲ توپ به تصادف برداشته و در جعبه B ميگذاريم، سپس دو توپ از جعبه B بدون جايگذاري خارج ميکنيم. مطلوبست محاسبه احتمال اينکه فقط يکي از دو توپ سفيد باشد.

تمرين: در يک مهماني سه زوج ازدواج کرده باهم دور يک ميز گرد مي‌نشينند مطلوبست محاسبه احتمال اينکه

الف: سه زن کنار هم نشسته باشند.

ب: سه زوج کنار هم بنشينند.

تمرين: فرض کنيد ۴ لامپ از ۱۲ لامپ موجود در جعبه اي سوخته باشند ۲ لامپ را به تصادف و بدون جايگذاري برميداريم مطلوبست محاسبه احتمال اينکه:

الف: هر دو لامپ سالم باشد.

ب: حداقل يک لامپ سوخته باشد.

تمرين: سه ظرف کاملا مشابه با محتويات زير مفروض هستند. ظرفي را به تصادف انتخاب و مهره اي را از آن خارج ميکنيم، اگر اين مهره سفيد باشد، مطلوبست محاسبه احتمال اينکه ظرف اول را انتخاب کرده باشيم.

	ظرف اول	ظرف دوم	ظرف سوم
مهره سفيد	۲	۳	۵
مهره سياه	۴	۵	۷

فصل چهارم: متغيرهاي تصادفي

با فرض اينکه هر آزمايش تصادفي داراي فضاي نمونه S باشد متغير تصادفي تابعي است که به وسيله آن به هر نقطه از فضاي نمونه اي يک عدد حقيقي نسبت ميدهد پس متغير تصادفي X تابعي است از فضاي نمونه به مجموعه اعداد حقيقي:

$$X : S \longrightarrow R$$

براي مثال دو سکه را همزمان پرتاپ مي کنيم،

ميدانيم S={HH, TH, HT, TT}

حال ميتوانيم X را يک متغير تصادفي که نشان دهنده تعداد شيرها در آزمايش باشد در نظر بگيريم. که به ترتيب مقادير ۰، ۱، ۱، ۲ را به اعضاي فضاي نمونه نسبت مي دهد.

فضاي نمونه S را گسسته گوييم، اگر تعداد عضوهاي آن متناهي يا نامتناهي شمارا باشد. مثل فضاي نمونه پرتاپ يک سکه يا پرتاپ يک سکه تا شير بيايد.

فضاي نمونه S را پيوسته گوييم، اگر تعداد عضوهاي آن نامتناهي ناشمارا باشد مثل فضاي نمونه آزمايش طول قد يا وزن افراد يک شهر.

اگر متغير تصادفي X روي فضاي نمونه گسسته تعريف گردد متغير تصادفي گسسته و اگر روي فضاي نمونه پيوسته تعريف گردد متغير تصادفي پيوسته ناميم.

توزيع احتمال گسسته: يک متغير تصادفي گسسته هر يک از مقادير خود را با احتمالي معين اختيار مي کند حال، جدول يا فرمولي که تمام مقادير متغير تصادفي را همراه با احتمالهاي مربوطه نشان دهد تابع احتمال ناميده و آن را با $f(x)$، $g(x)$ و ... نشان ميدهيم.

•در حالتي که متغير تصادفي گسسته باشد، تابع احتمال را توزيع احتمال مي‌ناميم.

•نوشتن توزيع احتمال هم به صورت جدول و هم فرمول براي يک متغير تصادفي ممکن است امکام پذير نباشد و بايد به يکي اکتفا کرد.

مثال: سکه اي را سه مرتبه پرتاب مي‌کنيم، اگر متغير تصادفي X نشان دهنده شير در اين آزمايش باشد توزيع احتمال X را به صورت جدول و فرمول بيان ميکنيم؟

S={ TTT, TTH, THT, HTT, THH, HTH, HHT, HHH }

X	0	1	2	3
$f(x)=P(X=x)$	1/8	3/8	3/8	1/8

$$f(x)=\frac{n(A)}{n(S)}=\frac{\binom{3}{x}}{2^3} \quad ,x=0,\ 1,\ 2,\ 3$$

تابع $f(x)$ را يک تابع توزيع احتمال براي متغير تصادفي X گوييم اگر:

$$\boxed{\begin{array}{l} 1)\ \forall x_i\ \ f(x_i)\geq 0 \\ 2)\ \sum_i f(x_i)=1 \\ 3)\ P(X=x_i)=f(x_i) \end{array}}$$

مثال: اگر $f(x)$ يک تابع توزيع احتمال مطلوبست مقدار k؟

$$f(x)=\frac{k+1}{2^x} \qquad x=0,1,2,\ldots$$

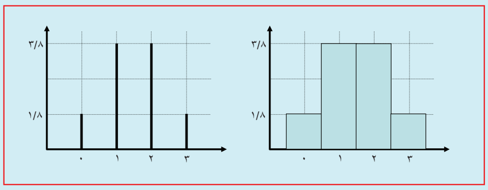
$$f(x)>0: 2^x>0 \Rightarrow k+1>0 \Rightarrow k>-1$$

$$\sum f(x)=1 \ \Rightarrow\ (k+1)\sum_{x=0}^{\infty}2^{-x}=1 \ \Rightarrow (k+1)(1+\frac{1}{2}+\frac{1}{4}+\cdots)=1$$

$$\Rightarrow\ (k+1)(\frac{1}{1-\frac{1}{2}})=1 \ \Rightarrow\ 2k+2=1 \ \Rightarrow\ k=-\frac{1}{2}$$

نمايش ترسيمي توزيع احتمال معمولا درک مطلب را ساده تر مي نمايد دو نمودار ميله‌اي و هيستوگرام براي نمايش توزيع احتمال به صورت نمودار به کار مي رود. که براي مثال نمودار ميله‌اي و هيستوگرام پرتاب سه سکه را رسم مي‌کنيم.

X	0	1	2	3
$f(x)=P(X=x)$	1/8	3/8	3/8	1/8



اگر X يک متغير تصادفي گسسته با توزيع احتمال $f(x)$ باشد تابع توزيع تجمعي X به شکل زير نمايش و تعريف مي شود.

$$F_x: R \longrightarrow R$$

$$F_x(x)=P(X\leq x)=\sum_{t\leq x}f(t)$$

مثال: جدول زير، توزيع احتمال متغير تصادفي X را نشان مي‌دهد، تابع توزيع تجمعي X را تعيين و نمودار آن را رسم کنيد:

X	0	1	2	3
$f(x)=P(X=x)$	1/8	3/8	3/8	1/8

$$x<0 \quad \Rightarrow \quad F(x)=0$$

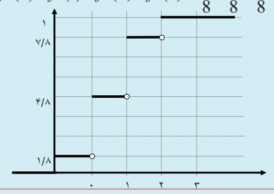
$$0\le x<1 \quad \Rightarrow \quad F(x)=\sum_{t<1} f(t)=f(0)=\frac{1}{8}$$

$$1\le x<2 \quad \Rightarrow \quad F(x)=\sum_{t<2} f(t)=f(0)+f(1)=\frac{1}{8}+\frac{3}{8}=\frac{4}{8}$$

$$2\le x<3 \quad \Rightarrow \quad F(x)=\sum_{t<3} f(t)=f(0)+f(1)+f(2)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}=\frac{7}{8}$$

$$3\le x<4 \quad \Rightarrow \quad F(x)=\sum_{t<3} f(t)=f(0)+f(1)+f(2)+f(3)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1$$

$$F(x)=\begin{cases} 0 & x<0 \\ 1/8 & 0\le x<1 \\ 4/8 & 1\le x<2 \\ 7/8 & 2\le x<3 \\ 1 & x\ge 3 \end{cases}$$



تابع توزيع تجمعي داراي خواص زير مي‌باشد:

۱) بازاي هر x: $0 \le F_x(x) \le 1$

۲) $F(x)$ تابعي غير نزولي است.

۳) $F(-\infty)=\lim_{x\to-\infty} F(x)=0 \quad , \quad F(+\infty)=\lim_{x\to+\infty} F(x)=1$

۴) تابع $F(x)$ در تمام نقاط از سمت راست پيوسته است.

۵) به ازاي a و b ثابت داريم

$$P(a<X\le b)=F(b)-F(a)$$
$$P(X=a)=F(a)-F(a-)$$

تمرين: يک جفت تاس را پرتاپ مي‌کنيم، اگر متغير تصادفي X نشان دهنده مجموع اعداد ظاهر شده باشد، مطلوبست تعيين:

- توزيع احتمال X
- نمودار توزيع احتمال
- تابع توزيع تجمعي
- نمودار توزيع تجمعي X
- احتمالهاي زير:

$$P(2<X\le 5) \quad , \quad P(2\le X\le 5)$$
$$P(X\ge 4) \quad , \quad P(X<5)$$

همانطور که ديديم به تابع احتمال يک متغير تصادفي گسسته تابع توزيع احتمال گفتيم حال اگر متغير تصادفي پيوسته باشد به تابع احتمال آن تابع چگالي احتمال مي گوييم و فرق اساسي آن با حالت گسسته در اين است که احتمال اينکه متغير تصادفي پيوسته دقيقاً يکي از مقادير خود را اختيار کند برابر صفر مي‌باشد بعبارت ديگر اگر X يک متغير تصادفي پيوسته باشد آنگاه

$$P(X=a)=0$$

يعني در تابع چگالي بحث روي بازه هست نه نقطه.

تمامي خواصي که براي تابع توزيع احتمال گفته شد در تابع چگالي احتمال نيز صادق است که به آنها اشاره مي کنيم.

تابع $f(x)$ را يک تابع چگالي گوييم هرگاه:

1) $\forall x\in R \; , \; f(x)\ge 0$

2) $\int_{-\infty}^{+\infty} f(x)\,dx=1$

3) $P(a\le X\le b)=P(a<X<b)=\int_a^b f(x)\,dx$

مثال: در تابع زير پارامتر a را طوري بيابيد که تابع $f(x)$ يک تابع چگالي احتمال باشد:

$$f(x)=\frac{1}{a(1+x^2)} \quad , \quad x\ge 0$$

$$\frac{1}{a}\int_0^{+\infty}\frac{dx}{1+x^2}=1 \quad\Rightarrow\quad \frac{1}{a}\tan^{-1}x\Big|_0^{+\infty}=1 \quad\Rightarrow$$

$$\frac{1}{a}(\tan^{-1}x-\tan^{-1}0)=1 \quad\Rightarrow\quad \frac{1}{a}\left(\frac{\pi}{2}-0\right)=1 \quad\Rightarrow$$

$$\frac{\pi}{2a}=1 \quad\Rightarrow\quad a=\frac{\pi}{2}$$

اگر X یک متغیر تصادفي پیوسته با تابع چگالي احتمال $f(x)$ باشد، تابع توزیع تجمعي X را با $F(x)$ نشان داده و به صورت زیر تعریف مي کنیم:

$$F(x)=P(X\le x)=\int_{-\infty}^{x} f(t)dt$$

با توجه به تعریف بالا داریم:

$$P(a<X<b)=P(X<b)-P(X<a)=F(b)-F(a)$$

از تعریف اتگرال معین نیز میتوان نوشت:

$$P(a<X<b)=\int_a^b f(x)dx=F(b)-F(a)\Rightarrow \boxed{F'(x)=f(x)}$$

تمامي خواص تابع توزیع تجمعي در حالت گسسته در حالت پیوسته نیز برقرار است:

۱) بازاي هر x : $0\le F_x(x)\le 1$

۲) $F(x)$ تابعي غیر نزولي است.

۳) $F(-\infty)=\lim_{x\to-\infty}F(x)=0\ \ ,\ \ F(+\infty)=\lim_{x\to+\infty}F(x)=1$

تمرین: اگر نقطه اي به تصادف در داخل دایره اي به شعاع r انتخاب کنیم و اگر متغیر تصادفي X نشان دهنده فاصله این نقطه از مرکز دایره باشد مطلوبست محاسبه $F(x)$ و $f(x)$ و $P(a<X<b)$ (a و b داخل دایره اند).

تمرین: اگر متغیر تصادفي X داراي تابع چگالي احتمال زیر باشد مقدار a و $F(x)$ را بیابید.

$$f(x)=\begin{cases}\dfrac{a}{x^2} & ,\ x>1\\ 0 & ,\ x<1\end{cases}$$

احتمال

تا حالا با فضاهاي نمونه یک بعدي و متغیرهاي تصادفي مربوط به این فضاها آشنا شدیم ولي آزمایشهاي زیادي وجود دارند که بطور همزمان دو یا چندین نتیجه خواهند داشت.

بعنوان مثال اگر یک اسید را با فلزي ترکیب کنیم بطور همزمان میخواهیم مقدار رسوب حاصل و مقدار گاز متصاعد شده را مورد بررسي کنیم.

یا اگر کیسه محتوي ۵ مهره قرمز ۴ مهره آبي و ۳ مهره زرد باشد ۳ توپ به تصادف خارج کنیم X میتواند تعداد توپ قرمز، Y تعداد توپ آبي و Z تعداد توپ زرد باشد که هر کدام از این متغیرها مي توانند اعدد ۰، ۱، ۲ و ۳ را بگیرند.

احتمال

اگر X و Y دو متغیر تصادفي باشند، احتمال وقوع همزماني آن‌ها با علامت تابعي $f(x, y)$ نشان داده و آن را تابع احتمال X و Y مي نامیم.

در حالتي که X و Y هر دو گسسته باشند $f(x, y)$ را توزیع احتمال و در حالتي که X و Y هر دو پیوسته باشند $f(x, y)$ را تابع چگالي احتمال مي نامیم.

خصوصیات تابع احتمال :

1) $f(x,y)\ge 0\quad ,\quad \forall(x,y)$

2) $\sum_x\sum_y f(x,y)=1\longleftarrow$ اگر X و Y گسسته باشند.

$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dy\,dx=1\longleftarrow$ اگر X و Y پیوسته باشند.

3) $P[(X,Y)\in A]=\sum_A\sum f(x,y)\longleftarrow$ اگر X و Y گسسته باشند.

$P[(X,Y)\in A]=\int_A\int f(x,y)dy\,dx\leftarrow$ اگر X و Y پیوسته باشند.

حال اين سئوال مطرح است اگر توزيع احتمال توأم دو يا چند متغير تصادفي معلوم باشد آيا مي توان توزيع هر كدام از اين متغيرها را به صورت جداگانه حساب كرد يا بالعكس، اگر توزيع هاي چند متغير معلوم باشد آيا امكان محاسبه توزيع احتمال توأم آنها (در صورت امكان وقوع همزمانشان) وجود دارد.

جواب: پاسخ قسمت اول همواره مثبت است ولي قسمت دوم فقط در مواردي امكان پذير است.

اگر X و Y دو متغير تصادفي با توزيع احتمال توأم (تابع چگالي توأم)، $f(x,y)$ باشند، توزيع احتمال هر يک از متغيرهاي تصادفي X و Y به صورت زير نمايش و محاسبه مي كنيم:

$$g(x)=\sum_{y} f(x,y)$$
$$h(y)=\sum_{x} f(x,y)$$
در حالت گسسته

$$g(x)=\int_{-\infty}^{+\infty} f(x,y)dy$$
$$h(y)=\int_{-\infty}^{+\infty} f(x,y)dx$$
در حالت پيوسته

$g(x)$ و $h(y)$ را توزيعهاي حاشيه اي X و Y مي ناميم.

مثال: شيشه اي حاوي ۳ قرص آسپرين، ۲ قرص خواب آور و ۲ قرص مسكن مفروض است . شخصي به تصادف دو قرص از اين ظرف خارج مي كند . اگر فرض كنيم X معرف قرص آسپرين و Y معرف قرص خواب آور باشد .مطلوبست:

- توزيع احتمال توأم را براي دو متغير تصادفي X و Y هم به صورت جدول و هم به صورت فرمول.
- توزيعهاي حاشيه اي X و Y

$$f(x,y)=P(X=x,Y=y)$$
$$=\frac{\binom{3}{x}\binom{2}{y}\binom{2}{2-x-y}}{\binom{7}{2}} \qquad x=0,1,2\ ,y=0,1,2$$

ادامه مثال:

Y \ X	0	1	2	$h(y)$
0	1/21	6/21	3/21	10/21
1	4/21	6/21	0	10/21
2	1/21	0	0	1/21
$g(x)$	6/21	12/21	3/21	

$$g(x)=\sum_{y=1}^{2} f(x,y)=f(x,0)+f(x,1)+f(x,2) \quad ,x=0,1,2$$
$$h(y)=\sum_{x=1}^{2} f(x,y)=f(0,y)+f(1,y)+f(2,y) \quad ,y=0,1,2$$

تمرين: تابع چگالي احتمال توأم دو متغير تصادفي X و Y به صورت زير تعريف شده است. مطلوبست:

- تعيين مقدار K
- توزيع هاي حاشيه اي X و Y
- محاسبه $P(X>3\ ,\ Y<2)$

$$f(x,y)=\begin{cases} Kxy & 0<x<4\ ,\ 1<y<5 \\ 0 & \text{other points} \end{cases}$$

اگر X و Y دومتغير تصادفي با تابع احتمال توأم $f(x,y)$ و توزيعهاي حاشيه اي $g(x)$ و $h(y)$ باشند، تابع احتمال شرطي متغير تصادفي Y در صورتيكه $X=x$ داده شده باشد، و تابع احتمال شرطي متغير تصادفي X در صورتيكه $Y=y$ داده شده باشد، عبارتست از:

$$k(x\mid y)=\frac{f(x,y)}{h(y)} \quad ,h(y)>0$$
$$k(y\mid x)=\frac{f(x,y)}{g(x)} \quad ,g(x)>0$$

مثال: اگر X و Y داراي تابع چگالي احتمال توأم زير باشند مطلوبست محاسبه $k(y \mid x)$ و $k(x \mid y)$

$$f(x,y)=\begin{cases}8xy & ,\ 0\le x\le 1\ ,\ 0\le y\le x\\ 0 & \text{ohetr points}\end{cases}$$

$$g(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy=\int_0^x 8xy\,dy=\begin{cases}4x^2 & ,\ 0\le x\le 1\\ 0 & \text{ohetr points}\end{cases}$$

$$h(y)=\int_{-\infty}^{+\infty} f(x,y)\,dx=\int_y^1 8xy\,dx=\begin{cases}4y(1-y^2) & ,\ 0\le y\le 1\\ 0 & \text{ohetr points}\end{cases}$$

$$k(y\mid x)=\frac{f(x,y)}{g(x)}=\frac{8xy}{4x^2}=\begin{cases}\dfrac{2y}{x^2} & ,\ 0\le y\le x\\ 0 & \text{ohetr points}\end{cases}$$

$$k(x\mid y)=\frac{f(x,y)}{h(y)}=\frac{8xy}{4y(1-y^2)}=\begin{cases}\dfrac{2x}{1-y^2} & ,\ y\le x\le 1\\ 0 & \text{ohetr points}\end{cases}$$

اگر X و Y دومتغير تصادفي با تابع احتمال توأم $f(x,y)$ و توزيع‌هاي حاشيه اي $g(x)$ و $h(y)$ باشند، X و Y را از لحاظ آماري مستقل گوييم اگر و تنها اگر براي هر (x,y) داشته باشيم:

$$f(x,y)=g(x)h(y)$$

پس فقط در حالت استقلال دو متغير تصادفي است كه مي‌توان از توزيع‌هاي حاشيه اي به تابع توزيع احتمال توأم رسيد.

تمرين: تابع چگالي احتمال توأم دو متغير تصادفي X و Y به صورت زير مفروض است مطلوبست:

$$f(x,y)=\begin{cases}K(x^2+y^2) & ,\quad 0\le x\le 1\quad ,\quad 0\le y\le 1\\ 0 & \text{other points}\end{cases}$$

- مقدار ثابت K
- توزيع هاي حاشيه اي X و Y
- $K(y|x)$ و $K(x|y)$
- بررسي استقلال دو متغير تصادفي
- محاسبه $P(X<1/2\ ,\ Y>1/2)$ و $P(0<X<1/2|Y=1/2)$

فصل پنجم: اميد رياضي و گشتاورها

اميد رياضي يا ميانگين يا متوسط يک متغير تصادفي مانند X، پارامتري است كه نشان دهنده مقدار مورد انتظار براي آن متغير تصادفي در اثر تكرار آن آزمايش به دفعات زياد، مي‌باشد كه به صورت زير محاسبه مي گردد:

اگر X گسسته باشد $\longleftarrow$ $E(X)=\sum_x xf(x)$

اگر X پيوسته باشد $\longleftarrow$ $=\int_{-\infty}^{+\infty} xf(x)\,dx$

مثال: متغير تصادفي X با تابع چگالي احتمال زير مفروض است اميد رياضي آن را به دست آوريد؟

$$f(x)=\begin{cases}\dfrac{x}{8} & ,\ 0<x<4\\ 0 & ,\ \text{other points}\end{cases}$$

$$E(X)=\int_{-\infty}^{+\infty} xf(x)=\int_0^4 x\,\frac{x}{8}\,dx=\frac{1}{24}x^3\Big|_0^4=\frac{8}{3}$$

قضيه: اگر متغير تصادفي X داراي تابع احتمال $f(x)$ باشد، اميد رياضي هر تابعي از X مانند $h(x)$ عبارتست از:

اگر X گسسته باشد $\longleftarrow$ $E(h(x)) = \sum_x h(x) f(x)$

اگر X پيوسته باشد $\longleftarrow$ $= \int_{-\infty}^{+\infty} h(x) f(x) dx$

اگر X و Y دو متتغر تصادفي با تابع احتمال $f(x,y)$ باشند، اميد رياضي هر تابعي از X و Y مانند $h(x,y)$ عبارتست از:

اگر X و Y هر دو گسسته باشد

$$E(h(x,y)) = \sum_x h(x,y) f(x,y)$$

اگر X و Y هر دو پيوسته باشد

$$E(h(x,y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x,y) f(x,y) dx dy$$

مثال: فرض كنيد X و Y دو متغير تصادفي مستقل با توابع چگالي زير باشند مطلوبست اميد رياضي $h(x,y)=xy$

$$g(x) = \begin{cases} \frac{8}{x^3} & , x > 2 \\ 0 & , \text{other points} \end{cases} \quad , \quad h(y) = \begin{cases} 2y & , 0 < y < 1 \\ 0 & , \text{other points} \end{cases}$$

چون دو متغير تصادفي مستقلند پس چگالي توأم آنها برابر است با حاصل ضرب چگالي هاي حاشيه اي:

$$f(x,y) = \begin{cases} \frac{16y}{x^3} & , x > 2 \ , 0 < y < 1 \\ 0 & , \text{other points} \end{cases}$$

$$E(xy) = \int_2^{\infty} \int_0^1 xy \frac{16y}{x^3} dx dy = \frac{8}{3}$$

قوانين اميد رياضي: اگر X و Y دو متغير تصادفي، a و b دو عدد ثابت و توابع g و h توابعي از X و Y باشند داريم:

$E(b) = b$

$E(aX + b) = aE(x) + b$

$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)]$

$E[g(X,Y) \pm h(X,Y)] = E[g(X,Y)] \pm E[h(X,Y)]$

اگر X و Y مستقل باشند: $E(XY) = E(X)E(Y)$

تمرين: اگر متغير تصادفي X داراي توزيع احتمال زير باشد مطلوبست محاسبه

x	-2	-1	1	2
$f(x)$	¼	¼	¼	¼

• $E(3x^2-4x+1/2)$

• اگر $Y=X^2$ توزيع احتمال Y و E(Y)

تمرين: اگر متغير تصادفي X داراي تابع چگالي احتمال زير باشد مطلوبست محاسبه E(X)

$$f(x) = \begin{cases} 2e^{-2x} & , x > 0 \\ 0 & , \text{other point} \end{cases}$$

گشتاورها: اگر X يک متغير تصادفي با تابع احتمال $f(x)$ باشد گشتاورهاي مختلف آن را به صورت زير نشان داده و تعريف مي‌كنيم:

گشتاور مرتبه rام پيرامون نقطه a $\longleftarrow$ $\mu_r^a = E(X-a)^r$

گشتاور مرتبه rام پيرامون مبدأ $\leftarrow$ $if \ (a=0) \Rightarrow \mu_r^a = \mu_r' = E(X^r)$

گشتاور مركزي مرتبه rام $\leftarrow$ $if \ (a=\mu) \Rightarrow \mu_r^\mu = \mu_r = E(X-\mu)^r$

واريانس X $\longleftarrow$ $if \begin{Bmatrix} a=\mu \\ r=2 \end{Bmatrix} \Rightarrow \mu_2 = V(X) = \sigma_X^2 = E(X-\mu)^2$

جذر مثبت واريانس را انحراف معيار گوييم.

واريانس متغير تصادفي به صورت زير راحتتر محاسبه مي‌گردد.

$$\sigma_X^2 = E(X-\mu)^2 = E(X^2) - \mu^2$$

كوواريانس X و Y را به صورت زير تعريف مي كنيم

$$\sigma_{XY} = \mathrm{cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

كوواريانس دو متغير تصادفي به صورت زير راحتتر محاسبه مي‌گردد.

$$\sigma_{XY} = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E(XY) - \mu_X \mu_Y$$

• براي بررسي چگونگي همبستگي خطي بين دو متغير تصادفي X و Y از كوواريانس آنها استفاده مي كنيم اگر تغييرات X و Y همسو باشند كوواريانس مثبت، فاقد تغييرات مشترک باشند كوواريانس صفر و تغييرات مخالف يكديگر باشند كوواريانس منفي خواهد بود.

• اگر X و Y مستقل باشند كوواريانس آنها صفر ولي عكس مطلب صحيح نيست.

مثال: براي جدول توزيع احتمال زير واريانس X وY و كواواريانس آنها را پيدا كرده و تفسير نماييد:

$$\sigma_X^2 = E(X^2) - \mu_x^2 = \sum_{x=0}^{2} x^2 g(x) - \left[\sum_{x=0}^{2} x g(x)\right]^2 =$$

$$(0\times 6/21 + 1\times 12/21 + 4\times 3/21) - (0\times 6/21 + 1\times 12/21 + 2\times 3/21)^2 = 20/49$$

$$\sigma_Y^2 = E(Y^2) - \mu_Y^2$$

$$\sigma_{XY} = \mathrm{cov}(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E(XY) - \mu_X \mu_Y =$$

$$E(XY) = \sum_y \sum_x xyf(x,y) = \frac{6}{21} = \frac{2}{7}$$

$$E(X) = \sum_x x g(x) = \frac{12}{21} + \frac{6}{21} = \frac{6}{7}$$

$$E(Y) = \sum_y y h(y) = \frac{10}{21} + \frac{2}{21} = \frac{4}{7}$$

$$\sigma_{XY} = \frac{2}{7} - \frac{6}{7}\times\frac{4}{7} = \frac{-10}{49} < 0$$

Y \ X	0	1	2	$h(y)$
0	1/21	6/21	3/21	10/21
1	4/21	6/21	0	10/21
2	1/21	0	0	1/21
$g(x)$	6/21	12/21	3/21	

اگر X و Y دو متغير تصادفي باشند. ضريب همبستگي خطي بين X و Y به صورت زير نشان داده و تعريف مي كنيم:

$$\rho = \rho(X,Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

• براي بررسي چگونگي همبستگي خطي بين دو متغير تصادفي X و Y استفاده مي كنيم.

• برتري ضريب همبستگي به كوواريانس در آنست كه ضريب همبستگي بستگي به واحد اندازه گيري ندارد

• اگر X و Y مستقل باشند ضريب همبستگي بين آنها صفر است.

• ضريب همبستگي همواره در فاصله [-1 , 1] قرار دارد.

• كوواريانس و ضريب همبستگي هميشه هم علامتند.

مثال: اگر X و Y دو متغير تصادفي با توزيع احتمال زير باشند، ضريب همبستگي بين آنها را حساب نماييد؟

$$f(x,y) = \begin{cases} \frac{1}{18}(x+2y) & , \ x=1,2 \ , \ y=1,2 \\ 0 & , \ \text{other points} \end{cases}$$

$$\mu_X = \frac{14}{9} , \ \mu_y = \frac{29}{18} , \ \sigma_X^2 = \ , \ \sigma_y^2 =$$

$$E(XY) = \sum_y \sum_x xy \frac{x+2y}{18} = \frac{45}{18}$$

$$\sigma_{xy} = \frac{45}{18} - \left(\frac{14}{9}\times\frac{29}{18}\right) =$$

$$\rho = \rho(X,Y) = \frac{\sigma_{Xy}}{\sigma_X \sigma y} = \frac{-\frac{1}{162}}{\sqrt{\frac{20}{81}\times\frac{77}{324}}} = -0.025$$

خواص واريانس و كوواريانس و ضريب همبستگي: اگر a، b، c و d اعداد ثابت و X و Y دو متغير تصادفي با توابع احتمال حاشيه‌اي $g(x)$ و $h(y)$ و تابع احتمال توأم $f(x,y)$، و $z(h)$ تابعي دلخواه از X باشد داريم:

1) $\sigma_{h(x)}^2 = E\left\{\left[h(x) - \mu_{h(x)}\right]^2\right\}$ 2) $\sigma_b^2 = 0$

3) $\sigma_{ax+b}^2 = a^2 \sigma_X^2$ 4) $\sigma_{ax+by+c}^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_{xy}$

5) $Cov(aX+b, dY+c) = adCov(X,Y)$

6) $\rho(aX+b, dY+c) = \rho(X,Y)$

اگر X و Y دو متغیر تصادفي با تابع احتمال توأم $f(x,y)$ باشند **امید رياضي Y به شرط آنکه $X=x$** باشد، عبارتست از:

اگر X و Y پیوسته باشند: $E(Y \mid X=x)=\int_{-\infty}^{+\infty} y\, k(y \mid x)\,dy$

اگر X و Y گسسته باشند: $=\sum_{y} y\, k(y \mid x)$

گشتاور مشروط Y بشرط آنکه $X=x$ باشد حول نقطه دلخواه a عبارتست از:

اگر X و Y پیوسته باشند: $E((Y-a)^r \mid X=x)=\int_{-\infty}^{+\infty}(y-a)^r\, k(y \mid x)\,dy$

اگر X و Y گسسته باشند: $=\sum_{y}(y-a)^r\, k(y \mid x)$

تمرین: اگر X و Y دو متغیر تصادفي با تابع چگالي احتمال توأم زیر باشند. مطلوبست

$$f(x,y)=\begin{cases} x+y & , \ 0\le x\le 1 \ , \ 0\le y\le 1 \\ 0 & , \ \text{other points} \end{cases}$$

- ضریب همبستگي بین X و Y
- محاسبه $E(Y \mid X=x)$
- محاسبه $V(Y \mid X=x)$
- محاسبه V(4X+3Y-2)

تابع مولد گشتاور: تابع مولد گشتاور X با تابع احتمال $f(x)$ را با نماد $M_x(t)$ یا $M(t)$ نشان داده و به صورت زیر تعریف مي‌کنیم:

$$M_x(t)=E(e^{tx}) \ , \ t\in R$$

$$=\begin{cases} \sum_{x} e^{tx} f(x) & \text{اگر X گسسته باشند:} \\ \int_{-\infty}^{+\infty} e^{tx} f(x)\,dx & \text{اگر X پیوسته باشند:} \end{cases}$$

یکي از کاربردهاي این تابع محاسبه میانگین و واریانس متغیرهاي تصادفي است.

قضیه: اگر X یک متغیر تصادفي با تابع مولد گشتاور $M(t)$ باشد آنگاه:

$$\left.\frac{d^r M(t)}{dt^r}\right|_{t=0}=\mu_r'$$

قضیه: اگر X و Y دو متغیر تصادفي با توابع مولد گشتاور $M_x(t)$ و $M_y(t)$ باشند، آنگاه X و Y داراي **توابع توزیع یکسان** هستند اگر و فقط اگر **$M_x(t) = M_y(t)$**

قضیه: اگر X و Y دو متغیر تصادفي مستقل با توابع مولد گشتاور $M_x(t)$ و $M_x(t)$ باشند، آنگاه

$$M_{x+y}(t)=M_x(t)M_y(t)$$

مثال: اگر متغیر تصادفي X داراي توزیع احتمال زیر باشد مطلوبست تعیین $M(t)$، میانگین و واریانس X:

$$f(x)=2\left(\frac{1}{3}\right)^x \quad , x=1,2,3,\ldots$$

$$M(t)=E(e^{tx})=\sum_{x=1}^{\infty}2\left(\frac{1}{3}\right)^x e^{tx}=2\sum_{x=1}^{\infty}\left(\frac{e^t}{3}\right)^x$$

$$if \ \frac{e^t}{3}<1 \Rightarrow M(t)=2\frac{e^t}{3}\times\frac{1}{1+e^t/3}=\frac{2e^t}{3-e^t}$$

$$\mu_X=\mu_1'=M'(t)\big|_{t=0}$$

$$\sigma_X^2=\mu_2'-(\mu')^2=\left[M''(t)-(M'(t))^2\right]\Big|_{t=0}$$

تمرین: اگر متغیر تصادفي X داراي تابع چگالي احتمال زیر باشد، مطلوبست محاسبه $M(t)$ و میانگین و واریانس X از روي آن؟

$$f(x)=\begin{cases} xe^{-x} & , \ x\ge 0 \\ 0 & ,\text{other points} \end{cases}$$

قضیه چیبیشف:

طبق این قضیه اطلاعاتي از میزان پراکندگي داده ها در اطراف میانگین بدست خواهیم آورد. اگر متغیر تصادفي X داراي واریانس کوچکي باشد انتظار مي رود که مشاهدات در اطراف میانگین متمرکزتر باشند.

قضیه: اگر X یک متغیر تصادفي و h(X) یک تابع نامنفي از X باشد آنگاه:

$$P(h(X)\ge c)\le \frac{1}{c}E(h(X)) \quad ,c>0$$

قضیه چیبیشف: اگر X یک متغیر تصادفي با واریانس محدود باشد آنگاه:

$$1)\ P\left[|X-\mu|\ge k\sigma\right]=P\left[(X-\mu)^2\ge k^2\sigma^2\right]\le \frac{1}{k^2} \quad ,\ k>0$$

$$P\left[|X-\mu|<k\sigma\right]\ge 1-\frac{1}{k^2}$$

مثال: اگر متغیر تصادفي X داراي میانگین ۲۵ و واریانس ۱۶ باشد، با استفاده از قضیه چیبیشف احتمال زیر را تعیین کنید؟

$$P(17<X<33)=P(-8<X-25<8)=P(|X-25|<2\times 4)\ge 1-\frac{1}{4}=3/4$$

$$P(|X-25|\ge 12)=P(|X-25|\ge 3\times 4)\le 1/9$$

فصل ششم: بررسي چند توزیع متغیرهاي تصادفي گسسته

میدانیم هر آزمایش تصادفي فضاي نمونه اي را ایجاد خواهد که مي توان متغیر تصادفي را روي این فضا تعریف نمود. همچنین هر متغیر تصادفي داراي تابع احتمالي است که میتوان آن را مشخص نمود حال این سؤال مطرح است:

آیا براي هر آزمایش آماري باید یک تابع احتمال تعیین کرد یا اینکه مي توان آزمایشهاي آماري را به دسته هاي مختلفي را تقسیم بندي کرد که هر دسته داراي خواص مشترکي باشند؟

دو فصل آینده سعي در دسته بندي آزمایشهاي آماري است به طوري که مطالعه این آزمایشات را اصولي تر کند.

توزیع یکنواخت: اگر متغیر تصادفي X مقادیر x_1، x_2، ...، x_n را با احتمال مساوي قبول کند، توزیع احتمال این متغیر تصادفي را توزیع یکنواخت مي‌نامیم و آن را به صورت زیر نمایش و تعریف مي‌کنیم.

$$f(x_i;n)=\frac{1}{n} \quad , \quad X=x_1,x_2,\dots,x_n$$

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

مثال: جعبه اي شامل صفحه کلید با شماره هاي ۱ تا k است. که شانس انتخاب شماره‌هاي این صفحه کلید یکسان مي باشد. اگر شماره‌اي از این صفحه کلید به تصادف انتخاب شود و X را شماره کلید انتخابي در نظر بگیریم آنگاه X داراي تابع توزیع احتمال یکنواخت است

$$f(x_i;k)=\frac{1}{k} \quad , \quad X=1,2,\dots,k$$

$$\mu=\frac{1}{k}\sum_{i=1}^{k}x_i=\frac{1}{k}\times\frac{k(k+1)}{2}=\frac{(k+1)}{2}$$

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2=E(X^2)-\mu^2=\frac{(k+1)(2k+1)}{6}-\frac{(k+1)^2}{4}$$

$$\sigma^2=\frac{(k+1)(k-1)}{12}$$

$$E(X^2)=\frac{1}{k}\sum_{i=1}^{k}x_i^2=\frac{1}{k}\frac{k(k+1)(2k+1)}{6}=\frac{(k+1)(2k+1)}{6}$$

توزيع برنولي: اگر نتيجه آزمايشي فقط دو وضعيت را معرفي كند (پيروزي يا شكست) آن آزمايش را آزمايش برنولي مي‌گويند. توزيع منتسب به اين آزمايش، توزيع برنولي نام دارد. اگر احتمال موفقيت P و احتمال شكست q=1-P فرض شود. در اين صورت توزيع مورد نظر به صورت زير خواهد بود:

$$f(x)=P(X=x)=\begin{cases} p^x q^{1-x} & ,x=0,1 \\ 0 & ,\text{other points} \end{cases}$$

$$\mu=E(X)=p$$

$$\sigma^2=E(X^2)-(E(X))^2=pq=p(1-p)$$

مثال: احتمال موفقيت براي يك داوطلب در جهت اخذ گواهينامه ۷۰% مي باشد، توزيع احتمال براي اين شخص كه در آزمون رانندگي شركت مي كند بنويسيد .

$$f(x)=\begin{cases} (0.7)^x(0.3)^{1-x} & ,x=0,1 \\ 0 & ,\text{other points} \end{cases}$$

$$\mu=p=0.7$$

$$\sigma^2==pq=0.7\times 0.3=0.21$$

$$M(t)=E(e^{tx})=\sum_x e^{tx}(p)^x(q)^{1-x}=q+e^t p=0.3+0.7e^t$$

توزيع دو جمله‌اي: اگر يك آزمايش برنولي n بار به صورت مستقل تكرار شود و در اين آزمايش متغير تصادفي X نشان دهنده تعداد پيروزي باشد، توزيع منتسب به اين آزمايش، توزيع دوجمله‌اي نام دارد و به اين شكل تعريف و نمايش داده مي‌شود:

$$f(x)=b(x;n,p)=\binom{n}{x}p^x q^{n-x} \quad ,x=0,1,2,\dots,n$$

$$M(t)=E(e^{tx})=\sum_x e^{tx}\binom{n}{x}p^x q^{n-x}=$$

$$\sum_x e^{tx}\binom{n}{x}(e^t p)^x q^{n-x}=(e^t p+q)^n$$

$$\mu=np \quad , \quad \sigma^2=npq$$

مثال: يك فوتباليست با احتمال ۸۰% توپي را با موفقيت به درون دروازه مي فرستد . اين فوتباليست ۶ توپ را در اختيار دارد . مطلوبست:

- معرفي توزيع احتمال مربوطه
- احتمال اينكه درست ۲ توپ را به درون دروازه بفرستد .
- احتمال اينكه حداكثر ۲ توپ را به درون دروازه بفرستد .
- احتمال اينكه حداقل ۲ توپ را به درون دروازه بفرستد.

$$f(x)=P(X=x)=\binom{6}{x}(0.8)^x(0.2)^{6-x}$$

$$f(2)=P(X=2)=\binom{6}{2}(0.8)^2(0.2)^4=15\times 0.64\times 0.0016$$

$$P(X\le 2)=P(X=0)+P(X=1)+P(X=2)$$

$$P(X\ge 2)=P(X=2)+\cdots+P(X=6)=1-P(X<2)=$$

$$1-\left[P(X=0)+P(X=1)\right]$$

براي محاسبه راحت تر احتمالهايي $P(X\le r)$ به صورت با توزيع دوجمله اي، جدولهايي آماده براي n هاي مختلف و به ازاي P هاي متفاوت با r هايي مشخص احتمال آن محاسبه شده است كه در محاسبات مي توان به آنها مراجعه نمود.

تمرين: يك صفحه هدف زني دايره اي شكل به ۱۵ قطاع مساوي تقسيم شده و با شماره هاي ۱ تا ۱۵ متمايز گرديده است، فرض كنيد متغير تصادفي X برابر عددي باشد كه سوزن در قطاع مربوط به آن اصابت ميكند، توزيع احتمال X را مشخص كرده و ميانگين و واريانس آن را بدست آوريد؟

تمرين: احتمال زدن تير به هدف ۱/۳ است، اگر ۵ با تير شليك شود مطلوبست محاسبه احتمال اينكه:

الف: حداقل دو بار تير به هدف بخورد

ب: لااقل چند بار تير را شليك كرد تا با احتمال بيش از ۹۰% تير به هدف بخورد.

توزيع چند جمله اي: آزمايش برنولي را تعميم مي دهيم، به اين صورت كه نتيجه آزمايش به يكي از پيشامدهاي E_1، E_2، ... ، E_k منجر شود به طوري كه اين پيشامدها ناسازگار و فرسا (در جريان آزمايش الزاماً يكي از آنها رخ دهد) باشد.

حال آزمايش را n بار به طور مستقل انجام مي‌دهيم اگر احتمال وقوع پيشامدهاي E_1، E_2، ... ، E_k به ترتيب برابر با P_1، P_2، ... ، P_k در نظر بگيريم آنگاه توزيع احتمال توأم متغيرهاي تصادفي X_1، X_2، ... ، X_k كه بترتيب نشان دهنده تعداد نتايج پيشامدهاي E_1، E_2، ... ، E_k در اين n آزمايش مستقل را **توزيع چند جمله اي** مي ناميم.

توزيع چند جمله اي:

$$f(x_1,x_2,\dots,x_k,p_1,p_2,\dots,p_k,n)=\binom{n}{x_1,x_2,\dots,x_k}p_1^{x_1},p_2^{x_2},\dots,p_k^{x_k}$$

$$\sum_{i=1}^{k}x_i=n\ ,\quad \sum_{i=1}^{k}p_i=1$$

مثال: در يكي از شهرهاي كشور، ۴ شبكه تلويزيون قابل استفاده است. بدين ترتيب شهروندان از شبكه یک به ميزان ۳۰%، از شبكه دو به ميزان ۱۰% و از شبكه سه به ميزان ۴۰% و بقيه از شبكه چهار مي توانند استفاده مطلوب ببرند. اگر ۱۰ نفر از جمعيت اين شهرستان به طور تصادفي به مصاحبه دعوت شوند ، احتمال آن پيشامدي را بيابيد كه ۳ نفر از شبكه يك و ۲ نفر از شبكه دو و ۴ نفر از شبكه سه و بقيه از شبكه چهار استفاده نموده اند:

ج: طبق توزيحاتي كه داده شد توزيع، یک توزيع چند جمله اي است پس:

$$P(X_1=3,X_2=2,X_3=4,X_4=1)=\binom{10}{3,2,4,1}(0.3)^1(0.1)^2(0.4)^4(0.2)^1$$

توزيع فوق هندسي: اگر آزمايشي دو شرط زير را داشته باشد آن را **فوق هندسي** مي‌ناميم:

۱- از جمعيتي با N عضو، يك نمونه تصادفي n تايي (بدون جايگذاري) انتخاب كنيم.

۲- k عضو از N عضو بنام موفقيت و $N-k$ عضو ديگر بنام عدم موفقيت باشند.

توزيع احتمال آن به صورت زير نمايش و تعريف مي گردد.

$$h(x;N,n,k)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}\ ,\quad \begin{array}{l} x=0,1,2,\dots,n\quad (n<k)\\ x=0,1,2,\dots,k\quad (k<n)\end{array}$$

مثال: محموله اي شامل ۱۰ تلويزيون كه ۳ تاي آن معيوب است به فروشگاهي ارسال مي‌شود. شخصي به تصادف ۴ تلويزيون را مي‌خرد اگر X تعداد تلويزيونهاي خرابي باشد كه در خريد اين شخص باشد توزيع احتمال آن را به دست آوريد.

$$h(x;10,4,3)=\frac{\binom{3}{x}\binom{10-3}{4-x}}{\binom{10}{4}}\ ,\quad x=0,1,2,3$$

اگر در توزيع فوق هندسي N خيلي بزرگ باشد به طوري كه نسبت n/N خيلي به چشم نياد اين توزيع به سمت توزيع دو جمله زير ميل خواهد كرد

$$h(x;N,n,k)\approx b(x;n,\frac{k}{N})$$

براي مثال فرض كنيد در يك شهر ۱۰۰۰۰ نفري نسبت رأي موافق به مخالف در مورد موضوع خاصي ۶۰ به ۴۰ است. اگر ۱۵ نفر از اين افرار را به تصادف انتخاب كنيم:

اگر اين انتخاب با جايگذاري باشد يعني هر شخص بعد از انتخاب شانس دوباره انتخاب شدن را داشته باشد توزيع دو جمله اي است ولي اگر بدون جايگذاري باشد توزيع فوق هندسي است، از آنجايي كه N بزرگ است و نسبت n/N كوچك مي توان با توزيع دو جمله اي تقريب زد.

تعميم توزيع فوق هندسي: در توزيع فوق هندسي فقط دو حالت داشتيم موفقيت و عدم موفقيت.

حال اگر جمعيت N عضوي به k دسته A_1، A_2، ... ، A_k افراز شود به طوري كه از N عضو جمعيت، a_1 عضو در A_1، a_2 عضو در A_2، ... ، و a_k عضو در A_k باشد، از اين جمعيت يك نمونه تصادفي n تايي بدون جايگذاري برداريم ميخواهيم احتمال توأم آن را حساب كنيم كه x_1 عضو از A_1، x_2 عضو از A_2، ... ، و x_k عضو از A_k انتخاب شود كه به صورت زير است.

$$f(x_1,x_2,\dots,x_k;a_1,a_2,\dots,a_k,N,n)=\frac{\binom{a_1}{x_1}\binom{a}{x_1}\cdots\binom{a_k}{x_k}}{\binom{N}{n}},\qquad \sum_{i=1}^{k}a_i=N,\quad \sum_{i=1}^{k}x_i=n$$

مثال: از گروهي متشكل از ۷۵ شيميدان، ۱۵ پزشك و ۳۵ رياضيدان يك كميته ۶ نفره به تصادف انتخاب مي كنيم مطلوبست محاسبه احتمال اينكه ۲ شيميدان، ۳ پزشك و ۱ رياضيدان شركت كنند:

$$f(2,3,1;25,15,35,75,6)=\frac{\binom{25}{2}\binom{15}{3}\binom{35}{1}}{\binom{75}{6}}$$

توزيع دو جمله اي منفي: اگر آزمايش برنولي را آنقدر ادامه دهيم تا r موفقيت داشته باشيم و پس از وقوع rامين موفقيت آزمايش متوقف گردد، اين آزمايش را دو جمله اي منفي مي‌ناميم حال اگر متغير تصادفي X نشان دهنده تعداد دفعات آزمايش دو جمله اي منفي باشد توزيع آن به صورت زير نمايش و تعريف مي‌گردد.

$$b^*(x;r,p)=\binom{x-1}{r-1}p^r q^{x-r}\quad ,x=r,r+1,\dots$$

$$M(t)=E(e^{tx})=p^r e^{tr}(1-qe^t)^{-r}\quad ,t<-\ln q$$

$$\mu=\frac{r}{p}\qquad ,\qquad \sigma^2=\frac{rq}{p^2}$$

مثال: احتمال اينكه راننده اي از يك چراغ قرمز عبور نمايد و پليس آنرا متوقف كند ۴۰% مي‌باشد، احتمال آن پيشامد را بيابيد كه در حين عبور از چراغ قرمز چهارم پليس او را متوقف كند؟

$$b^*(4;1,0.4)=f(4)=\binom{3}{0}(0.4)^1(0.6)^3=0.09$$

مثال: احتمال اينكه اگر فردي شايعه اي را بشنود آنرا باور كند ۰,۷ مي‌باشد. احتمال آن پيشامدي را بيابيد كه ۵ امين فردي كه اين شايعه را مي شنود سومين فردي باشد كه آنرا باور مي كند؟

$$b^*(5;3,0.7)=f(5)=\binom{4}{2}(0.7)^3(0.3)^2$$

توزيع هندسي: اگر در آزمايش دو جمله اي منفي $r=1$ يعني آزمايش را تا حصول اولين پيروزي ادامه دهيم اين آزمايش را آزمايش هندسي مي‌ناميم و X را كه نشان دهنده تعداد دفعات آزمايش هندسي است متغير تصادفي هندسي ناميم و توزيع آن را به صورت زير نمايش و تعريف مي‌كنيم.

$$g(x;p)=pq^{x-1}\quad x=1,2,\dots$$

مثال: اگر موفقيت جهت اخذ گواهينامه داوطلب ۸۰% باشد. اولاً توزيع احتمال منتسب به اين آزمايش را معرفي كنيد. ثانياً احتمال آن پيشامدي را بيابيد كه اين داوطلب در مرحله سوم موفق به اخذ گواهينامه شده است .

$$g(x;p)=g(x;0.8)=0.8\times 0.2^{x-1}\quad x=1,2,\dots$$

$$g(3;0.8)=f(3)=(0.8)(0.2)^2=0.032$$

تمرين: از جعبه اي شامل ۵ توپ قرمز و ۴ توپ سبز يك نمونه تصادفي ۶ تايي انتخاب مي‌كنيم مطلوبست محاسبه اينكه ۴ توپ قرمز و ۲ توپ سبز باشد؟

تمرين: احتمال اينكه در آزمايش پرتاب سكه، سومين شير در هفتمين بار پرتاب بدست آيد را محاسبه كنيد؟

تمرين: احتمال اينكه تيراندازي به هدف بزند برابر ۰٫۸ است مطلوبست محاسبه احتمال اينكه:

الف: كمتر از ۵ بار لازم باشد تا اولين تير به هدف بخورد؟

ب: تعداد زوج بار لازم باشد تا اولين تير به هدف بخورد؟

توزيع پواسن: برخي آزمايشها به گونه اي هستند كه نتايج حاصل از آنها، تعداد وقايعي است كه در فواصل زماني يا در ناحيه مكاني مشخص به وقوع مي‌پيوندند، چنين آزمايشهايي به آزمايش پواسن معروفند. در واقع بررسي پيشامدهاي جدا در در يك فاصله زماني يا يك مكان مشخص فرايند پواسن ناميده مي‌شود.

مثل تعداد تلفنهايي كه در ساعات مشخص به يك مركز زده مي‌شود. يا تعداد تصادفاتي كه در يك چهارراه مشخص رخ مي‌دهد. يا تعداد اشتباهاتي كه يك تايپيست در يك صفحه دارد و ...

توزيع پواسن: متغير تصادفي X كه نشان دهنده تعداد موفقيت در يك آزمايش پواسن باشد را متغير تصادفي پواسن ناميده و توزيع احتمال آن در صورتيكه ميانگين تعداد موفقيت در فاصله زماني يا ناحيه مكاني مشخص (λ) معلوم باشد، به صورت زير نمايش و تعريف مي‌گردد

$$P(x;\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}\quad ,x=0,1,2,....$$

$$M(t)=E(e^{tx})=e^{\lambda(e^t-1)}$$

$$\mu=\lambda \quad , \quad \sigma^2=\lambda$$

مثال: پزشكي به طور متوسط در هر ساعت ۴ بيمار را ويزيت مي كند . احتمال پيشامدهاي زير را براي او محاسبه كنيد .

۱) در يك ساعت خاص پزشك بيكار باشد .

۲) در يك ساعت خاص درست دو بيمار را ويزيت كند .

۳) در يك ساعت خاص حداكثر دو بيمار را ويزيت كند .

۴) حداقل دو بيمار را ويزيت كند .

۵) در ۱۵ دقيقه اول درست سه بيمار را ويزيت كند .

جواب:

$$P(x,\lambda)=P(x,4)=\frac{e^{-4}\cdot 4^x}{x!}\qquad ,x=0,1,\cdots$$

$$1)\,P(0,4)=f(0)=e^{-4}$$

$$2)\,P(2,4)=f(2)=\frac{e^{-4}\cdot 4^2}{2!}=8e^{-4}$$

$$3)\,P(x\le 2,4)=f(x\le 2)=P(X=0)+P(X=1)+$$
$$P(X=2)=e^{-4}+4e^{-4}+8e^{-4}$$

$$4)\,P(x\ge 2,4)=f(x\ge 2)=1-P(X<2)=$$
$$1-(e^{-4}+4e^{-4})=1-5e^{-4}$$

$$5)\lambda=\lambda\cdot t=4\times\frac{1}{4}=1 \quad\Rightarrow\quad P(x,1)=\frac{e^{-1}\cdot 1^x}{x!}$$

$$P(3,1)=f(3)=\frac{e^{-1}}{3!}$$

قضيه: اگر متغير تصادفي X داراي توزيع دو جمله‌اي با توزيع احتمال b(x ; n,p) باشد، آنگاه شكل حدي اين توزيع وقتي كه $n \to \infty$ و $p \to 0$ يا $p \to 1$، توزيع پواسن با پارامتر $\lambda=np$ مي‌باشد بعبارت ديگر:

$$b(x;n,p)\approx P(x;np)$$
$$n\to\infty$$
$$p\to 0 \text{ or } p\to 1$$

مثال: در يك استوديوم ۱۰۰۰۰ نفري احتمال اينكه فردي در اثر گرمازدگي بيمار شود ۰٫۰۰۰۳ مي باشد . احتمال آن پيشامدي را بيابيد كه در اين جمع ۴ نفر مبتلا به گرمازدگي شوند؟

$$b(4;10000,0.0003)=f(4)=\binom{10000}{4}(0.0003)^4(0.997)^{9996}$$

$$b(x;n,p)\approx P(x;np)$$
$$n\to\infty$$
$$p\to 0 \text{ or } p\to 1$$

$$b(4;10000,0.0003)\approx P(4;3)=f(4)=\frac{e^{-3}\cdot 3^4}{4!}=\frac{81}{24}e^{-3}$$

تمرين: فرض كه در هر سال، در بين هر ۵۰۰۰۰ نفر، ۲ نفر خودكشي ميكنند، احتمال اينكه در يك شهر، ۱۰۰۰۰۰ نفري در يك سال معين،

الف: صفر خود كشي

ب: يك خودكشي

ت: ۲ خودكشي يا بيشتر تنجام گيرد؟

تمرين: در يك جاده به خصوص به طور متوسط ۶ تصادف در سال رخ مي دهد، مطلوبست محاسبه احتمال اينكه در يك سال مشخص

الف: كمتر از ۴ تصادف رخ دهد؟

ب: حداقل ۴ تصادف رخ دهد؟

فصل هفتم: بررسي چند توزيع پيوسته

توزيع يكنواخت پيوسته: متغير تصادفي X در فاصله [a , b] داراي توزيع يكنواخت است اگر تابع چگالي آن به صورت زير باشد كه به صورت U [a , b] نشان مي‌دهند.

$$f(x)=\begin{cases}\dfrac{1}{b-a} & , a\le x\le b\\ 0 & , \text{other points}\end{cases}$$

$$\mu=\frac{a+b}{2}\ ,\ \sigma^2=\frac{(b-a)^2}{12}$$

بعبارت ديگر، در توزيع يكنواخت احتمال قرار گرفتن يك نقطه در قسمتي از اين فاصله متناسب با طول آن قسمت بوده و قرار گرفتن آن نقطه در درون اين فاصله حتمي است.

مثال: اگر X ، U[3 , 10] باشد مطلوبست محاسبه

الف: ميانگين و واريانس توزيع

ب: P(4<X<8)

$$f(x)=\begin{cases}\dfrac{1}{b-a}=\dfrac{1}{7} & , 3\le x\le 10\\ 0 & , \text{other points}\end{cases}$$

$$\mu=\frac{a+b}{2}=\frac{3+10}{2}=6.5$$

$$\sigma^2=\frac{(b-a)^2}{2}=\frac{(10-3)^2}{2}=24.5$$

$$P(4<X<8)=\int_4^8\frac{1}{7}dx=\frac{4}{7}$$

توزيع نمايي: گفتيم توزيع پواسن تعداد وقايع در يك ناحيه پيوسته يا يك فاصله زماني است حال اگر متغير تصادفي X زمان اولين اتفاق يا زمان بين دو اتفاق متوالي در توزيع پواسن، باشد X يك توزيع نمايي است با پارامتر θ كه در آن θ /1تعداد وقايع در واحد زمان مي‌باشد.

$$f(x)=\frac{1}{\theta}e^{-\frac{1}{\theta}x}\quad x\ge 0\ ,\ \theta>0$$

$$M_X(t)=\frac{1}{1-t\theta}\quad ,t<\frac{1}{\theta}$$

$$\mu_X=E(X)=\theta\quad ,\ \sigma_X^2=\theta^2$$

مثال: به طور متوسط در هر ساعت تعداد ۲۰ اتومبيل وارد يك پاركينگ مي‌شوند اگر ساعت ۶ صبح پاركينگ باز شود مطلوبست محاسبه احتمال اينكه

الف: حداقل ۵ دقيقه طول بكشد تا اولين اتومبيل وارد پاركينگ شود؟

ب: زمان بين ورود دو ماشين به پاركينگ حداقل ۵ دقيقه باشد؟

ب: تا ساعت ۶:۱۰ صبح اولين اتومبيل وارد پاركينگ شود؟

پ: اگر تا ساعت ۶:۲۰ اتومبيلي وارد پاركينگ نشده باشد بعد از ۶:۲۵ وارد شود؟

ج: اگر تا ساعت ۶:۲۰ اتومبيلي وارد پاركينگ نشده باشد تا ۶:۲۵ وارد شود؟

جواب:

$$\frac{1}{\theta} = 20/60 = 1/3$$

$$f(x) = \frac{1}{3} e^{-\frac{1}{3}x} \quad x > 0$$

$$P(X > 5) = \int_5^\infty \frac{1}{3} e^{-\frac{1}{3}x} dx = e^{-\frac{5}{3}}$$

$$P(X < 10) = \int_0^{10} \frac{1}{3} e^{-\frac{1}{3}x} dx = \frac{1}{3}(1 - e^{-\frac{10}{3}})$$

$$\begin{Bmatrix} P(X > t_1 + t_2 \mid X > t_1) = P(X > t_2) \\ P(X < t_1 + t_2 \mid X > t_1) = P(X < t_2) \end{Bmatrix} \longrightarrow$$ توزيع نمايي فاقد حافظه است

$$P(X > 25 \mid X > 25) = P(X > 5)$$
$$P(X < 25 \mid X > 20) = P(X < 5)$$

توزيع گاما

تابع گاما: به صورت زير تعريف مي‌گردد

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

كه داراي خواص مقابل مي‌باشد.

$$\Gamma(1) = 1$$
$$\Gamma(1/2) = \sqrt{\pi}$$
$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$
$$\Gamma(\alpha+1) = \alpha!$$

توزيع گاما: در بخش قبل زمان لازم براي اولين رخداد توزيع پواسن مطرح گرديد كه گفتيم توزيع نمايي است با پارامتر θ. حال مطلب فوق را تعميم مي‌دهيم و زمان لازم براي α امين رخداد را بررسي مي‌كنيم.

چنين توزيعي را توزيع گاما با پارامترهاي α و θ مي‌ناميم كه تابع چگالي آن به شكل زير مي‌باشد

$$f(x) = \begin{cases} \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\theta}} & 0 \le x < \infty \\ 0 & x < 0 \end{cases}$$

$$M_x(t) = \left(\frac{1}{1-\theta t}\right)^\alpha = (1-\theta t)^{-\alpha} \quad , t < \frac{1}{\theta}$$

$$\mu_X = \alpha\theta \quad , \quad \sigma_X^2 = \alpha\theta^2$$

مثال: فرض كنيد در هر ساعت به طور متوسط ۳۰ اتومبيل وارد پاركينگ ميشوند، مطلوبست محاسبه

الف: احتمال اينكه مأمور پاركينگ لااقل ۵ دقيقه منتظر بماند تا دومين ماشين وارد پاركينگ شود؟

ب: ميانگين و واريانس متغير تصادفي X

$$\frac{1}{\theta} = 30/60 = 1/2 \quad \Rightarrow \theta = 2$$

$$\alpha = 2$$

$$f(x) = \begin{cases} \frac{1}{2^2 \Gamma(2)} x^{2-1} e^{-\frac{x}{2}} & 0 \le x < \infty \\ 0 & x < 0 \end{cases}$$

$$P(X > 5) = \int_5^\infty \frac{1}{4} x e^{-\frac{x}{2}} dx = \frac{7}{2} e^{-\frac{5}{2}}$$

$$\mu_X = \alpha\theta = 2 \times 2 = 4 \quad , \quad \sigma_X^2 = 2 \times 2^2 = 8$$

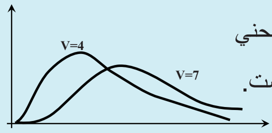
توزيع توان دوم كاي (خي دو): اگر در توزيع گاما θ=2 و α=v/2 (v عدد صحيح و مثبت) باشد مي‌گوييم متغير تصادفي X داراي توزيع X^2 با درجه آزادي v است.

$$f(x) = \begin{cases} \frac{1}{2^{(\frac{v}{2})} \Gamma(\frac{v}{2})} x^{\frac{v}{2}-1} e^{-\frac{x}{2}} & 0 \le x < \infty \\ 0 & x < 0 \end{cases}$$

$$M_x(t) = (1-2t)^{(-\frac{v}{2})}$$

$$\mu_X = v \quad , \quad \sigma_X^2 = 2v$$

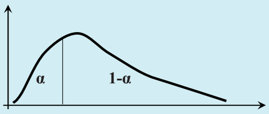
مقادير X^2 منفي نمي باشند و منحني توزيع آن حول x=0 متقارن نيست.



چون اين توزيع در عمل مورد استفاده فراوان قرار مي‌گيرد سطح زير منحني اين توزيع به ازاي مقادير مختلف X^2 و درجات آزادي مختلف محاسبه و جدول آن تهيه شده است. چگونگي استفاده از اين جدول را با مثالي شرح مي‌دهيم:

مثال: اگر متغير تصادفي X داراي توزيع توان دوم كاي با v=10 باشد مطلوبست محاسبه P(3.5<X<20.5) و P(X>3.5؟

$$P(2.5 < X < 9.34) = P(X < 9.34) - P(X < 2.5) = 0.5 - 0.01 = 0.49$$
$$P(X > 2.5) = 1 - P(X < 2.5) = 1 - 0.01 = 0.99$$



v	$X^2_{0.005}$	$X^2_{0.01}$	$X^2_{0.25}$	$X^2_{0.05}$	$X^2_{0.1}$	$X^2_{0.25}$	$X^2_{0.5}$
1							
2							
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34

تابع بتا: به صورت زير تعريف مي گردد و داراي خواص زير مي باشد:

$$\beta(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$\beta(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

توزيع بتا: متغير تصادفي X داراي توزيع بتا است هرگاه تابع چگالي احتمال آن به صورت زير باشد :

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}.(1-x)^{\beta-1} & , 0 < x < 1 \, , \ \alpha,\beta > 0 \\ 0 & , \text{other points} \end{cases}$$

$$\mu = \frac{\alpha}{\alpha+\beta} \quad , \quad \sigma^2 = \frac{\alpha\beta^2}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

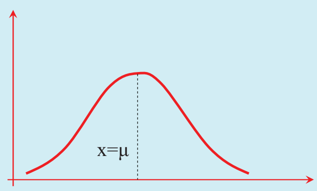
در سالهاي اخير توزيع بتا كاربردهاي مهمي در استنباطهاي بيزي پيدا كرده است كه در آن پارامترها به عنوان متغير تصادفي در نظر گرفته مي‌شوند كه با تغيير پارامترها تابع چگالي شكلهاي گوناگوني به خود مي‌گيرد.

در حالتي كه α=β=1 داريم

$$f(x) = \begin{cases} \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} x^0.(1-x)^0 = 1 & , 0 < x < 1 \\ 0 & , \text{other points} \end{cases}$$

كه همان تابع چگالي يكنواخت مي‌باشد.

توزيع نرمال: نمودار اين توزيع به منحني نرمال معروف است و زنگي شكل بوده و بيشتر وقايعي كه در طبيعت و تحقيقات علمي بوقوع مي‌پيوندد از اين منحني پيروي ميكند. متغير تصادفي X كه منحني آن زنگي شكل باشد را متغير تصادفي نرمال مي‌ناميم. اين منحني نسبت به خط متقارن x=μ است .

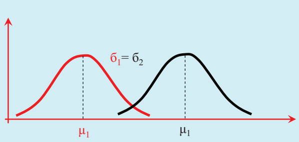


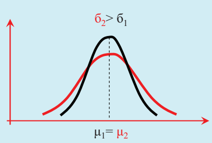
توزيع نرمال: متغير تصادفي X داراي توزيع نرمال است هرگاه تابع چگالي احتمال آن به صورت زير باشد :

$$n(x;\mu,\sigma) = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \ : x \in \mathbb{R}$$

$$M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$$

$$\mu_X = \mu \quad , \quad \sigma_X^2 = \sigma^2$$





خواص توزيع نرمال:

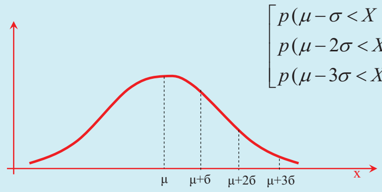
اين منحني نسبت به خط متقارن $x=\mu$ است .

ميانه، ميانگين و مد بر هم منطبقند.

منحني داراي دو نقطه عطف در نقاط $x=\mu \pm \sigma$

محور xها مجانب افقي منحني است يعني $\lim f(x)=0$ وقتي $x\to\pm\infty$

$$\begin{bmatrix} p(\mu-\sigma<X<\mu+\sigma)=0.68 \\ p(\mu-2\sigma<X<\mu+2\sigma)=0.95 \\ p(\mu-3\sigma<X<\mu+3\sigma)=0.997 \end{bmatrix}$$



متغير تصادفي نرمال استاندارد:

متغير تصادفي با ميانگين $\mu=0$ و واريانس $\sigma^2=1$ را متغير تصادفي نرمال استاندارد مي‌گوييم و آن را با Z نشان مي‌دهيم.

• اگر X يك متغير تصادفي با ميانگين μ و واريانس σ^2 باشد آنگاه متغير تصادفي $Z=\frac{X-\mu}{\sigma}$ داراي توزيع نرمال استاندارد است .

براي محاسبه احتمال $P(x_1<X<x_2)$ كه در آن X نرمال با ميانگين و واريانس مشخص باشد به صورت زير عمل مي‌كنيم:

$$\mu_X=\mu \quad , \quad \sigma_X^2=\sigma^2$$

$$p(x_1<X<x_2)=p\left(\frac{x_1-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{x_2-\mu}{\sigma}\right)=$$

$$P(z_1<Z<z_2)=P(Z<z_2)-P(Z<z_1)=\phi(z_2)-\phi(z_1)$$

مقدار تابع Φ كه مقادير متغير تصادفي نرمال استاندارد براي احتمال‌هايي به صورت $P(Z<z)$ مي‌باشد در جداولي به صورت آماده محاسبه مي‌گردد كه چگونگي استفاده از اين جدول با مثالي شرح داده مي‌شود.

مثال: اگر X داراي توزيع نرمال با $\mu=1$ و $\sigma^2=4$ مطلوبست محاسبه احتمالهاي $P(0<X<3)$ و $P(X<2.3)$

$$P(0<X<3.5)=P\left(\frac{0-1}{2}<Z<\frac{3.5-1}{2}\right)=P\left(\frac{-1}{2}<Z<1.25\right)$$

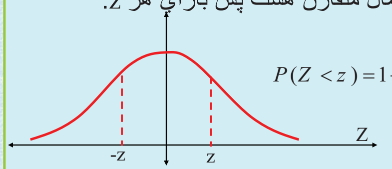
$$=\phi(1.25)-\phi(-0.5)=0.8944-0.3085=0.5859$$

$$P(X>4)=P\left(Z>\frac{4-1}{2}\right)=P(Z>1.5)=1-P(Z<1.5)$$

$$=1-\phi(1.5)=1-0.9332=0.0668$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
-0.5	0.3085						
1.2						0.8944	
1.5	0.9332						

چون تابع توزيع نرمال متقارن هست پس بازاي هر z:



$$P(Z<z)=1-P(Z<-z)$$

$$P(Z<-1)=1-P(Z<1)=1-0.8413=0.1587$$

$$P(-0.5<Z<0.5)=P(Z<0.5)-P(Z<-0.5)=$$

$$P(Z<0.5)-(1-P(Z<0.5)=2P(Z<0.5)-1=$$

$$(2\times 0.6915)-1=0.383$$

مثال: اگر X داراي توزيع نرمال با $\mu=25$ و $\sigma=6$ باشد مطلوبست تعيين ثابت c در احتمال زير؟

$$P(|X-25|\le c)=0.9544$$

$$P(|X-25|\le c)=0.9544=P(-c\le X-25\le c)=$$

$$P\left(\frac{-c}{6}\le\frac{X-25}{6}\le\frac{c}{6}\right)=P\left(\frac{-c}{6}\le Z\le\frac{c}{6}\right)=$$

$$2P\left(Z\le\frac{c}{6}\right)-1=0.9544 \Rightarrow P\left(Z\le\frac{c}{6}\right)=0.9772$$

$$\frac{c}{6}=2 \Rightarrow c=12$$

تقريب توزيع نرمال براي توزيع دو جمله اي:

زماني كه n در توزيع دوجمله اي بزرگ باشد، عملا محاسبه احتمالات دو جمله اي ممكن نيست در اين حالت مي توانيم از توزيع نرمال براي تقريب دوجمله اي استفاده كنيم . وقتي$np>5$، تقريب نرمال براي توزيع دوجمله‌اي بسيار خوب است.

براي آن كه بتوانيم توزيع دوجمله اي را به كمک توزيع نرمال تقريب بزنيم بايد توجه داشته باشيم كه چون دوجمله اي يک توزيع گسسته مي‌باشد و آن را به كمک توزيع نرمال كه يک توزيع پيوسته مي باشد تقريب مي زنيم بايد از تصحيح پيوستگي به صورت زير استفاده كنيم :

$$p(X=k)=p(k-0.5\le X\le k+0.5)$$

تقريب توزيع نرمال براي توزيع دو جمله اي:

اگر X يك متغير تصادفي گسسته از نوع دو جمله‌اي با ميانگين $\mu=np$ و واريانس $\sigma^2=npq$ باشد، شكل حدي توزيع متغير تصادفي زير وقتي n به سمت بينهايت ميل مي‌كند توزيع نرمال استاندارد مي‌باشد.

$$Z=\frac{X-np}{\sqrt{npq}}$$

مثال : فرض مي كنيم X داراي توزيع دوجمله اي به صورت $b(x;15,0.4)$ باشد، مطلوب است محاسبه $P(7<X<10)$؟

حل :

$$np=6>5$$
$$p(7\le X\le 10)=p(6.5\le X\le 10.5)$$
$$p\left(\frac{6.5-6}{\sqrt{6(0.6)}}\le\frac{X-np}{\sqrt{np(1-p)}}\le\frac{10.5-6}{\sqrt{6(1-0.4)}}\right)=$$
$$p(0.26<Z<2.37)=\Phi(2.37)-\Phi(0.26)=$$
$$0.9911-0.6026=0.3885$$

قضيه: اگر X يك متغير تصادفي نرمال با ميانگين μ و واريانس σ^2 باشد، آنگاه متغير تصادفي V داراي توزيع توان دوم كاي با يك درجه آزادي است.

$$V=\left(\frac{X-\mu}{\sigma}\right)^2$$
$$P(|Z|<1.96)=P(Z^2<3.842)P(X^2<3.84)=0.95$$

قضيه: اگر X_1، X_2، ...، X_n، n متغير تصادفي مستقل نرمال با ميانگين هاي μ_1، μ_2، ...، μ_n، و واريانس‌هاي σ_1^2، σ_2^2، ...، σ_n^2 باشند، آنگاه U داراي توزيع توان دوم كاي با n درجه آزادي است.

$$U=\sum_{i=1}^{n}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2$$

تمرين: معدل نمره ۳۰۰ دانشجوي يك دانشكده تقريبا داراي توزيع نرمال با ميانگين ۲/۱ و انحراف معيار ۱/۲ است. در صورتيكه معدل‌ها تا يكدهم تقريب محاسبه شده باشند، مطلوبست

الف: معدل چند نفر از دانشجويان در فاصله [۲/۵ ، ۳/۵] قرار دارد؟

الف: احتمال اينكه معدل يك دانشجو كه به تصادف انتخاب مي‌شود دقيقا ۲٫۵ باشد؟

تمرين: تاس همگني را ۷۲۰ بار پرتاب مي كنيم اگر متغير تصادفي X نشان دهنده تعداد ۶ هاي ظاهر شده باشد، مطلوبست محاسبه $P(100<X<125)$

تمرين: اگر متغير تصادفي X داراي توزيع توان دوم كاي با $v=4$ باشد، مقدار x را در حالات زير محاسبه كنيد؟

الف: $P(X<x)=0.99$

ب: $P(X>x)=0.75$

تمرين: اگر متغير تصادفي X داراي توان دوم كاي با $v=18$ باشد مطلوبست:

الف: محاسبه ميانگين و واريانس توزيع X

ب: $P(X>17.3)$

تمرين: اگر متغير تصادفي X داراي توزيع نمايي با ميانگين ۰/۵ باشد مطلوبست محاسبه $P(X>4)$

تمرين: اگر متغير تصادفي X داراي توزيع نرمال با ميانگين ۳ و واريانس ۱۶مطلوبست محاسبه $P(4<X<8)$

تمرين: طول عمر يك نوع وسيله الكتريكي داراي توزيع نرمال با ميانگين ۲ و انحراف معيار ۰/۳ سال است. احتمال اينكه يك وسيله از اين نوع داراي طول عمر كمتر از ۲/۳ سال باشد، چقدر است؟

توزیع t student: فرض کنید Z يك متغير تصادفي نرمال استاندارد و χ^2 يك متغير تصادفي توان دوم كاي با درجه آزادي v باشد و همچنين Z و χ^2 مستقل باشند، آنگاه

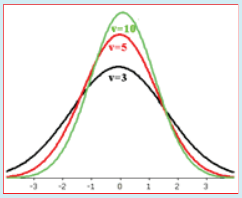
$$T = \frac{Z}{\sqrt{\chi^2 / v}} =$$

داراي تابع چگالي به صورت زير مي‌باشد كه به توزيع t با v درجه آزادي مي‌ناميم.

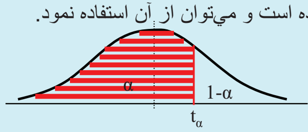
$$f(t) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{-\frac{1}{2}(v+1)} \qquad ,t \in \mathbb{R}$$

$$\mu = 0 \ , \ \sigma^2 = \frac{v}{v-2} \ , v > 2$$

توزيع t مانند توزيع نرمال زنگي شكل است و همانند توزيع نرمال استاندارد نسبت به خط x=0 متقارن است و هر چه درجه آزادي بزرگتر باشد توزيع t بسمت نرمال استاندارد ميل خواهد كرد. و وقتي $v \to +\infty$ منحني توزيع t همانند توزيع نرمال استاندارد است.



محاسبه احتمال در توزيع t: چون توزيع t، داراي تابع چگالي احتمال پيچيده‌اي است و انتگرال گيري از آن مشكل مي‌باشد، احتمالهاي اين توزيع نيز مانند توزيع نرمال در جداولي براي P(T<t) با درجات آزادي مختلف محاسبه شده است و مي‌توان از آن استفاده نمود.



$P(T < t_\alpha) = \alpha$	$P(Z < z_\alpha) = \alpha$
$P(T < -t_\alpha) = 1-\alpha$	$P(Z < -z_\alpha) = 1-\alpha$
$P(T < t_{1-\alpha}) = P(T < -t_\alpha)$	$P(Z < z_{1-\alpha}) = P(Z < -z_\alpha)$
$t_{1-\alpha} = -t_\alpha$	$z_{1-\alpha} = -z_\alpha$

مثال:

$$P(T < 1.75) \quad , v = 16 \Rightarrow = 0.95$$
$$P(T > 1.75) \quad , v = 16 = 1 - P(T < 1.75) = 1 - 0.95 = 0.05$$
$$P(T < t) = 0.95 \ , v = 24 \Rightarrow t = 1.71$$
$$P(T > t) = 0.65 \ , v = 24 \Rightarrow 1 - P(T < t) = 0.65$$
$$P(T < t) = 0.35 \quad \Rightarrow \quad t = 0.65$$

v	$t_{0.55}$	$t_{0.6}$	$t_{0.7}$	$t_{0.75}$	$t_{0.8}$	$t_{0.9}$	$t_{0.95}$
16	0.128					1.34	
24		0.256	0.531	0.685			1.71

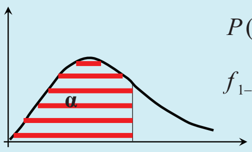
توزيع F: اگر U و V دو متغير تصادفي مستقل توان دوم كاي با درجات آزادي v_1 و v_1 باشند آنگاه متغير تصادفي F را توزيع F با درجه آزادي v_1 و v_1 مي‌ناميم كه ترتيب درجات آزادي با توجه به شكل تابع توزيع مهم مي‌باشد. كه تابع چگالي آن به شكل زير است:

$$\boxed{F = \frac{U/v_1}{V/v_2}}$$

$$f(x) = \begin{cases} \dfrac{\Gamma((v_1+v_2)/2)(v_1/v_2)^{v_1/2}}{\Gamma(v_1/2)\Gamma(v_1/2)} \cdot \dfrac{x^{\frac{v_1}{2}-1}}{(1-v_1 x/v_2)^{(v_1+v_2)/2}} & , x > 0 \\ 0 \ , \text{other points} & \end{cases}$$

$$\mu_X = \frac{v_2}{v_2-2} \ , \sigma_X^2 = \frac{2v_2^2(v_1+v_2-2)}{v_1(v_2-4)(v_2-2)^2} \ , Mo = \left(\frac{v_1-2}{v_1}\right)\left(\frac{v_2}{v_2+2}\right)$$

محاسبه احتمال در توزيع F: چون توزيع F تابع چگالي پيچيده‌اي دارد و محاسبه انتگرال ساده نمي‌باشد احتمالهاي اين توزيع نيز بر حسب v_1 و v_2 هاي مختلف محاسبه مي‌گردد. كه $F_\alpha(v_1 , v_2)$ نقطه‌اي از توزيع F با درجات v_1 و v_2 مي‌باشد كه مساحت α در سمت چپ آن واقع شده است.



$$P(F(v_1,v_2) \le f_\alpha(v_1,v_2)) = \alpha$$

$$f_{1-\alpha}(v_1,v_2) = \frac{1}{f_\alpha(v_2,v_1)}$$

مثال:

$$P(F(7,4) \le 6.09) = 0.95$$
$$P(F(3,5) > a) = 0.05 \Rightarrow 1 - P(F(3,5) < a) = 0.05$$
$$P(F(3,5) < a) = 0.95 \Rightarrow a = 5.41$$

$F_{0.95}$							
$v_2 \backslash v_1$	1	2	3	4	5	6	7
4							6.09
5			5.41				

رابطه زير در محاسبه احتمالهاي مختلف در سه تابع توزيع نرمال، t و F رابطه اي مفيد مي‌باشد.

$$P(z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}) = 1-\alpha$$
$$P(t_{\frac{\alpha}{2}} < T < t_{1-\frac{\alpha}{2}}) = 1-\alpha$$
$$P\left(f_{\alpha/2}(v_1,v_2) < F < f_{1-\alpha/2}(v_1,v_2)\right) = 1-\alpha$$

فصل هشتم: تابع‌هاي متغيرهاي تصادفي

مقدمه: در اين فصل هدف پيدا كردن تابع توزيع يا چگالي احتمال توابعي مانند $y=u(x_1, x_2, \dots, x_n)$ با فرض معلوم بودن توابع احتمال متغيرهاي تصادفي x_1، x_2، ...، x_n است.

براي اين كار چندين روش موجود است:

۱- تكنيك تابع توزيع

۲- تكنيك تبديل متغيرها

۳- تكنيك تابع مولد گشتاورها

كه بسته به نوع مسأله مي‌توان از يك يا چند تكنيك مختلف براي پيدا كردن توابع احتمال استفاده كرد.

تكنيك تابع توزيع:

روشي براي به دست آوردن چگالي احتمال تابعي از متغيرهاي تصادفي پيوسته است كه اگر x_1، x_2، ... ، x_n متغيرهاي تصادفي پيوسته با چگالي احتمال مفروضي باشند آنگاه:

$$Y = u(x_1, x_2, \dots, x_n)$$
$$F(y) = P(Y \le y) = P[u(x_1, x_2, \dots, x_n) \le y] \Rightarrow$$
$$f(y) = \frac{dF(y)}{dy}$$

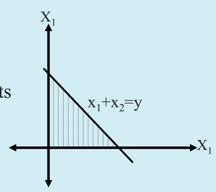
مثال: اگر توزيع توأم x_1 و x_2 به صورت زير باشد مطلوبست چگالي احتمال توأم $y=x_1+x_2$

$$f(x_1,x_2) = \begin{cases} 6e^{-3x_1-2x_2} & ; x_1,x_2 > 0 \\ 0 & ; \text{other points} \end{cases}$$

$$y = x_1 + x_2 :$$

$$F(y) = P(Y < y) = P(x_1 + x_2 < y) = \int_0^y \int_0^{y-x_2} 6e^{-3x_1-2x_2} dx_1 dx_2 =$$
$$= 1 + 2e^{-3y} - 3e^{-2y}$$

$$f(y) = \frac{dF(y)}{dy} \begin{cases} 6(e^{-2y} - e^{-3y}) & ; y > 0 \\ 0 & ; \text{other points} \end{cases}$$



تکنيك تبديل متغير:

اين تکنيك در هر دو حالت گسسته و پيوسته كاربرد دارد:

الف: در حالت گسسته مادامي كه رابطه بين مقادير x (متغير تصادفي با تابع توزيع معلوم) و تابع y=u(x) (متغير تصادفي با تابع توزيع مجهول) يك رابطه يك به يك است آنچه بايد انجام دهيم فقط يك جايگذاري است ولي اگر اين رابطه يك به يك نباشد بايد دقت شود اين جايگذاري به ازاي تمام روابط جايگزين شود.

مثال: اگر x تعداد شيرهايي باشد كه در چهار پرتاپ يك سكه همگن به دست مي‌آيد توزيع احتمال y=1/(1+x) و $z=(x-2)^2$ را بيابيد؟

$$X: b(x;4,0.5) \Rightarrow f(x)=\binom{4}{x}\cdot 0.5^x\cdot 0.5^{4-x}=\binom{4}{x}\cdot 0.5^4 \;;x=0,1,2,3,4$$

$$y=\frac{1}{1+x}\Rightarrow x=\frac{1}{y}-1\Rightarrow g(y)=f\left(\frac{1}{y}-1\right)=\binom{4}{\frac{1}{y}-1}\cdot 0.5^4 \;;y=1,1,2,3,4$$

$$z=(x-2)^2=\begin{cases}4 & ;x=0\\ 1 & ;x=1\\ 0 & ;x=2\\ 1 & ;x=3\\ 4 & ;x=4\end{cases}\Rightarrow\begin{cases}0 & ;x=2\\ 1 & ;x=1,3\\ 4 & ;x=0,4\end{cases}\Rightarrow h(z)=\begin{cases}f(2) & ;z=0\\ f(1)+f(3) & ;z=1,3\\ f(0)+f(4) & ;z=0\end{cases}$$

تکنيك تبديل متغير:

ب: در حالت پيوسته، فرض كنيم f(x) مقدار چگالي احتمال متغير تصادفي x باشد اگر تابعي كه به صورت y=u(x) داده شده است مشتق‌پذير و به ازاي تمامي مقادير برد x كه براي آنها $f(x)\neq 0$، صعودي يا نزولي باشد آنگاه، براي اين مقادير x معادله y=u(x) را مي‌توان به صورت يكتا بر حسب x حل كرد تا x=w(y) به دست آيد و چگالي احتمال y به صورت زير است:

$$g(y)=\begin{cases}f[w(y)].|w'(y)| & ;u'(x)\neq 0\\ 0 & ;\text{other points}\end{cases}$$

مثال: اگر x داراي توزيع نمايي به صورت زير باشد مطلوبست چگالي احتمال متغير تصادفي y؟

$$f(x)=\begin{cases}e^{-x} & ;x>0\\ 0 & ;\text{other points}\end{cases}$$

$$y=\sqrt{x}\quad\Rightarrow f(y)?$$

$$y^2=x\quad\Rightarrow x'=2y$$

$$f(y)=f(y^2)\cdot|2y|=\begin{cases}2ye^{-y^2} & ;y>0\\ 0 & ;\text{other points}\end{cases}$$

تکنيك تبديل متغير:

اين تکنيك را مي‌توان براي پيدا كردن توزيع متغير تصادفي كه تابعي از دو يا چند متغير تصادفي است به كار برد به اين صورت كه اگر تابع احتمال توأم x_1، x_2، ...، x_n مفروض باشد و بخواهيم تابع احتمال $y=u(x_1,x_2,\ldots,x_n)$ را به دست آوريم رابطه يكي از x_i ها و y را با ثابت نگه داشتن ساير x_i ها حساب مي‌كنيم:

$$\boxed{y=u(x_1,x_2,\ldots,x_n)\;:\exists x_k:x_k=v(y,x_1,\ldots,x_{k-1},x_{k+1},\ldots x_n)}$$

حال در حالت گسسته توزيع توأم y و ساير x_i ها را با جايگذاري به دست مي‌آوريم سپس توزيع حاشيه‌اي y را مي‌توان از توزيع توأم بدست آورد ولي در حالت پيوسته از رابطه زير استفاده مي‌كنيم:

$$\boxed{g(y,x_1,\ldots,x_{k-1},x_{k+1},\ldots x_n)=f(x_1,x_2,\ldots,x_n).\left|\frac{\partial x_k}{\partial y}\right|}$$

حال مي‌توان چگالي احتمال حاشيه‌اي y را از روي چگالي توأم موجود به دست آورد.

مثال: اگر x_1 و x_2 متغيرهاي تصادفي مستقلي باشند كه توزيع‌هاي پواسن با پارامترهاي λ_1 و λ_2 دارند مطلوبست توزيع احتمال متغير تصادفي y؟

$$y=x_1+x_2$$

$$f(x_1,x_2)=\frac{e^{-\lambda_1}\lambda_1^{x_1}}{x_1!}\cdot\frac{e^{-\lambda_2}\lambda_2^{x_2}}{x_2!}=\frac{e^{-\lambda_1+\lambda_2}\lambda_1^{x_1}\lambda_2^{x_2}}{x_1!x_2!}\;;x_1,x_2=0,1,\ldots$$

$$y=x_1+x_2\;\Rightarrow\;x_1=y-x_2\Rightarrow$$

$$g(y,x_2)=\frac{e^{-\lambda_1+\lambda_2}\lambda_1^{x_2}\lambda_2^{y-x_2}}{x_2!(y-x_2)!}\Rightarrow h(y)=\sum_{x_2=0}^{y}\frac{e^{-\lambda_1+\lambda_2}\lambda_1^{x_2}\lambda_2^{y-x_2}}{x_2!(y-x_2)!}=$$

$$=\frac{e^{-\lambda_1+\lambda_2}}{y!}\sum_{x_2=0}^{y}\frac{y!}{x_2!(y-x_2)!}\lambda_1^{x_2}\lambda_2^{y-x_2}=\frac{e^{-\lambda_1+\lambda_2}}{y!}(\lambda+\lambda_2)^y\quad y=0,1,\ldots$$

مثال: اگر چگالي توأم x_1، x_2 و x_3 به صورت زير باشد مطلوبست چگالي احتمال متغير تصادفي y؟

$$y=x_1+x_2+x_3$$

$$f(x_1,x_2,x_3)=\begin{cases} e^{-(x_1+x_2+x_3)} & ;x_1,x_2,x_3>0 \\ 0 & ;\text{other points}\end{cases}$$

$$x_1=y-x_2-x_3 \Rightarrow \frac{\partial x_1}{\partial y}=1$$

$$g(y,x_2,x_3)=e^{-y}.|1|=e^{-y} \quad ;x_2,x_3>0 \ \& \ y>x_2+x_3 \Rightarrow$$

$$h(y)=\int_0^y\int_0^{y-x_3} e^{-y}dx_2dx_3=\frac{1}{2}y^2e^{-y}$$

تكنيك تابع مولد گشتاورها:

اين تكنيك در تعيين توزيع احتمال يا چگالي احتمال تابعي از متغيرهاي تصادفي، وقتي تابع مزبور تركيب خطي از متغير هاي تصادفي مستقل است نقش مهمي دارد به اين صورت كه اگر x_1، x_2، ...، x_n متغيرهاي تصادفي مستقل و $M_{x_i}(t)$ مقدار تابع مولد گشتاورهاي x_i به ازاي t باشند و $y=x_1+x_2+\dots+x_n$، آنگاه

$$M_y(t)=\prod_{i=1}^n M_{x_i}(t)$$

مثال: اگر x_1، x_2، ...، x_n متغيرهاي تصادفي مستقل از توزيع نمايي با پارامتر θ باشند مطلوبست چگالي احتمال y؟

$$y=x_1+x_2+..x_n:$$

$$M_{x_i}(t)=(1-\theta t)^{-1}\Rightarrow M_y(t)=\prod_{i=1}^n(1-\theta t)^{-1}=(1-\theta t)^{-n}$$

پس y يك توزيع گاما با $\alpha=n$ و $\beta=\theta$ است

مثال: اگر x_1، x_2، ...، x_n متغيرهاي تصادفي مستقل از توزيع پواسن با پارامتر λ باشند مطلوبست چگالي احتمالy؟.

$$y=x_1+x_2+..x_n:$$

$$M_{x_i}(t)=e^{\lambda_i(e^t-1)})\Rightarrow M_y(t)=\prod_{i=1}^n(e^{\lambda_i(e^t-1)})=e^{(\lambda_1+\dots+\lambda_n)(e^t-1)}$$

پس y يك توزيع پواسن با پارامتر $\lambda=\lambda_1+\lambda_2+\dots+\lambda_n$ است

تمرين: اگر چگالي احتمال x به صورت زير باشد مطلوبست چگالي احتمال $y=x^3$

$$f(x)=\begin{cases} 6x(1-x) & ;0<x<1 \\ 0 & ;\text{other points}\end{cases}$$

تمرين: اگر $y=|x|$ مطلوبست تابع احتمال y ؟

تمرين: اگر x داراي توزيع نرمال استاندارد باشد مطلوبست چگالي احتمال $z=x^2$

تمرين: اگر F(x) تابع توزيع تجمعي متغير تصادفي x باشد مطلوبست چگالي احتمال $y=F(x)$؟

تمرين: اگر چگالي توأم دو متغير تصادفي به صورت زير باشد مطلوبست تابع چگالي y؟

$$f(x_1,x_2)=\begin{cases} e^{-(x_1+x_2)} & ;x_1,x_2>0 \\ 0 & ;\text{other points}\end{cases}$$

$$y=\frac{x_1}{x_1+x_2};\ h(y)?$$

فصل نهم: جامعه و نمونه آماري

جامعه آماري:

- به مجموعه كل مشاهداتي كه مورد بررسي هستند، جامعه آماري اطلاق مي‌شود.
- تعداد عضوهاي جامعه را حجم جامعه مي‌ناميم.
- مشاهدات آماري مقاديري از يك متغير تصادفي هستند پس به هر جامعه آماري مي توان يك توزيع احتمال متناسب با متغير تصادفي آن، نسبت داد.
- جامعه مي‌تواند طبق حجم آن، متناهي يا نامتناهي باشد.
- در جوامع نامتناهي از ميانگين و واريانس آن به دليل عدم امكان دسترسي به تمام اعضاي آن، اطلاعي نداريم از اين رو ميانگين و واريانس جامعه را پارامترهاي جامعه مي‌ناميم.
- براي تخمين (برآورد) پارامترهاي جامعه نامتناهي، از نمونه (تصادفي) استفاده مي‌كنيم
- نمونه بايد معرف جامعه باشد (مشت نمونه خروار باشد).

نمونه تصادفي: يك نمونه تصادفي به حجم n، نمونه‌ايست كه هر زير مجموعه n عضوي از جامعه داراي شانس انتخاب يكسان باشند بعبارت ديگر:

اگر X_1، X_2، ...، X_n متغير تصادفي با تابع احتمال يكسان $f(x)$ باشند، آنگاه x_1، x_2، ...، x_n را يك نمونه تصادفي با حجم n از جامعه مي‌ناميم و توزيع احتمال آن عبارتست از:

$$f(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_1)$$

هدف تخمين پارامترهاي جامعه از روي نمونه تصادفي است (آمار استنباطي)

آماره: هر تابعي از عضوهاي نمونه تصادفي كه شامل پارامترهاي مجهول نباشد را يك آماره مي‌ناميم اگر X_1، X_2، ...، X_n يك نمونه تصادفي از متغير تصادفي X باشند آنگاه توابع زير يك آماره هستند:

$$X_1 + 3X_2 - 1 \;,\; \frac{\sum_{i=1}^{n} X_i}{n} \;,\; \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n} \;,\; \frac{X_1}{X_4}$$

آماره‌ها را با يكي از حروف لاتين نشان داده و روي آن علامتي به صورت (— ، ^ يا ~) قرار مي‌دهيم. هدف تعيين توابع توزيع آماره‌ها براي بررسي آنها و تخمين پارامترهاي جامعه از روي آنهاست.

بررسي چند آماره مفيد

يادآوري:

گشتاور: يکي از پارامترهاي پراكندگي جامعه مي باشد. گشتاور مرتبه rام پيرامون نقطه ثابت a را براي يک جامعه به شکل زير نمايش و تعريف مي‌كنند.

$$\boxed{m_r^a = \frac{1}{N}\sum_{i=1}^{N}(x_i - a)^r}$$

گشتاور حول ميانگين جامعه را گشتاور مركزي نامند و آن را با μ_r نشان مي‌دهند

$$\boxed{m_r^\mu = \mu_r = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^r}$$

گشتاور حول مبدا را نيز به اختصار به صورت زير نمايش داده مي‌شود

$$\boxed{m_r^0 = \mu_r' = \frac{1}{N}\sum_{i=1}^{N}(x_i)^r}$$

گشتاور نمونه اي : اگر X_1، X_2، ...، X_n يك نمونه تصادفي از متغير تصادفي X باشد آنگاه روابط زير به ترتيب معرف گشتاور نمونه مرتبه r ام حول نقطه a، حول ميانگين نمونه و حول مبدأ تعريف ميگردد.

$$M_r^a = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^r \;, r = 1, 2, 3, \dots$$

$$a = \bar{X} \Rightarrow M_r^{\bar{X}} = M_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^r \;, r = 1, 2, 3, \dots$$

$$a = 0 \Rightarrow M_r^0 = M_r' = \frac{1}{n}\sum_{i=1}^{n}(X_i)^r \;, r = 1, 2, 3, \dots$$

در گشتاور نمونه مرتبه r ام حول مبدأ اگر r=1 ميانگين نمونه بدست مي آيد

$$M_1' = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \bar{X}$$

قضيه: اگر X_1، X_2، ...، X_n يك نمونه تصادفي از جامعه اي با متغير تصادفي X و تابع احتمال f(x) باشند آنگاه:

$$\text{if } (\mu_r' \text{ exist}) \Rightarrow$$
$$E(M_r') = \mu_r'$$
$$V(M_r') = \frac{1}{n}\left[\mu_{2r}' - (\mu_r')^2\right]$$

در حالت خاص اگر r=1:

$$** \ E(\bar{X}) = \mu_X$$
$$** \ V(\bar{X}) = \frac{\sigma_X^2}{n}$$

واريانس نمونه : اگر X_1، X_2، ...، X_n يك نمونه تصادفي از جامعه‌اي با متغير تصادفي X و تابع احتمال f(x) باشند، واريانس نمونه به صورت زير تعريف مي‌شود:

$$S_n^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \ , n > 1$$

قضيه: اگر X_1، X_2، ...، X_n يك نمونه تصادفي از جامعه‌اي نامتناهي با متغير تصادفي X و تابع احتمال f(x) باشند آنگاه:

$$E(S^2) = \sigma_X^2$$
$$\text{var}(S^2) = \frac{2\sigma^4}{n-1}$$

توزيع نمونه آماره S^2: مي دانيم كه آماره S^2 واريانس نمونه است و به صورت زير تعريف مي‌شود:

$$S_n^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \ , n > 1$$

قضيه: توزيع آماره S^2 مشخص نيست ولي اگر S^2 واريانس يك نمونه تصادفي با حجم n از يك جامعه نرمال با واريانس σ^2 باشد آنگاه توزيع آماره

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

توزيع توان دوم كاي با v=n-1 درجه آزادي است.

مثال: يك نمونه تصادفي ۲۵ تايي از يك جامعه نرمال با واريانس ۶ انتخاب مي‌كنيم مطلوبست احتمال زير؟

$$P(S^2 > 9.1)?$$

$$P(S^2 > 9.1) = 1 - P(S^2 < 9.1) =$$
$$1 - P\left(\frac{(25-1)S^2}{6} < \frac{(25-1)9.1}{6}\right) = 1 - P(\chi^2 < 36.4) = 0.05$$

توزيع آماره S_1^2/S_2^2: در بسياري از مواقع لازم است كه واريانس هاي دو جامعه با يكديگر مقايسه شوند براي اين منظور از قضيه زير استفاده مي‌كنيم:

قضيه: توزيع آماره S_1^2/S_2^2 مشخص نيست اما اگر n_1 و n_2 دو حجم نمونه مستقل از دو جامعه نرمال با واريانس‌هاي σ_1^2 و σ_2^2 باشد آنگاه توزيع آماره

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}$$

توزيع F با $v_1=n_1-1$ و $v_2=n_2-1$ درجه آزادي است.

مثال: اگر S_1^2 و S_2^2 واريانس‌هاي دو نمونه تصادفي با حجم‌هاي $n_1=25$ و $n_2=31$ از دو جامعه نرمال با واريانس‌هاي $\sigma_1^2=10$ و $\sigma_2^2=15$ باشند مطلوبست محاسبه احتمال

$$P\left(\frac{S_1^2}{S_2^2} > 1.26\right)?$$

$$P\left(\frac{S_1^2}{S_2^2} > 1.26\right) = P\left(\frac{S_1^2}{S_2^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} > \frac{15}{10}\times 1.26\right) =$$
$$P(F(24,30) > 1.89) = 1 - P(F(24,30) < 1.89)$$
$$= 1 - 0.95 = 0.05$$

توزيع نمونه ميانگين

قضيه : اگر X_1، X_2، ...، X_n يك نمونه تصادفي از يك توزيع نرمال با ميانگين µ و واريانس متناهي $σ^2$ باشد آنگاه توزيع نمونه $\bar{X}$ نرمال با ميانگين µ و واريانس $σ^2/n$ مي‌باشد.

توزيع نمونه ميانگين

قضيه: اگر X_1، X_2، ...، X_n يك نمونه تصادفي "با جايگذاري" از جامعه‌اي متناهي با ميانگين µ و واريانس متناهي $σ^2$ باشد آنگاه توزيع نمونه $\bar{X}$:

وقتي n≥30 باشد، بدون توجه به شكل توزيع جامعه، نرمال با ميانگين µ و واريانس $σ^2/n$ مي‌باشد يعني $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ نرمال استاندارد است.

ولي اگر n<30 باشد، در صورت نرمال بودن توزيع جامعه، نرمال با ميانگين µ و واريانس $σ^2/n$ مي‌باشد.

توزيع نمونه ميانگين

قضيه : اگر X_1، X_2، ...، X_n يك نمونه تصادفي "بدون جايگذاري" از جامعه اي متناهي به حجم N با ميانگين µ و واريانس متناهي $σ^2$ باشد آنگاه توزيع نمونه $\bar{X}$:

وقتي n≥30 و N≥2n باشد، بدون توجه به شكل توزيع جامعه، توزيع نمونه نرمال با ميانگين µ و واريانس $(σ^2/n)[(N-n)(N-1)]$ مي‌باشد.

ولي اگر n<30 باشد، در صورت نرمال بودن توزيع جامعه، نرمال با ميانگين µ واريانس $(σ^2/n)[(N-n)(N-1)]$ مي‌باشد.

توزيع نمونه ميانگين

قضيه حد مركزي : اگر X_1، X_2، ...، X_n يك نمونه تصادفي (چه با جايگذاري و چه بدون جايگذاري) از جامعه اي نامتناهي (بزرگ) با ميانگين µ و واريانس متناهي $σ^2$ باشد آنگاه شكل حدي توزيع :

$$Z_n=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

وقتي n→∞ (n≥30) باشد، بدون توجه به شكل توزيع جامعه، نرمال استاندارد است يعني توزيع نمونه $\bar{X}$ نرمال با ميانگين µ و واريانس $σ^2/n$ مي‌باشد.

مثال: ميانگين طول عمر يك نوع لامپ الكتريكي ۷۸۰ ساعت و انحراف معيار آن ۳۶ ساعت است. يك نمونه تصادفي ۳۶ تايي از اين نوع لامپ خريداري مي‌شود، احتمال اينكه ميانگين طول عمر اين نمونه، از ۷۷۴ ساعت بيشتر باشد، چقدر است.

حل: طبق قضيه حد مركزي حل مي‌كنيم

$$\mu_{\bar{X}}=\mu=780\ ,\ \sigma_{\bar{X}}^2=\frac{\sigma^2}{n}=\frac{(36)^2}{36}\Rightarrow\sigma_{\bar{X}}=6$$

$$P(\bar{X}>774)=1-P(\bar{X}<774)=$$

$$1-P(\frac{\bar{X}-780}{6}<\frac{774-780}{6})=1-P(Z<1)=1-0.1587=0.8413$$

مثال: يك جامعه متناهي به حجم ۸۵ با ميانگين ۱,۵ و واريانس ۱۶ مفروض است يك نمونه تصادفي ۶۴ تايي از اين جامعه، {الف:با جايگذاري ب: بدون جايگذاري} انتخاب ميكنيم مطلوبست $P(0.5<\bar{X}<2)$ در هر دو حالت؟

حل: با توجه به حجم نمونه (n>30) توزيع نمونه $\bar{X}$ توزيع نرمال است اما در حالت الف توزيع آن نرمال با ميانگين µ و واريانس $σ^2/n$ مي‌باشد ولي در حالت ب توزيع آن نرمال با ميانگين µ و واريانس $(σ^2/n)(N-n/N-1)$ مي‌باشد پس

$$\mu=1.5\ ,\ n=64\ ,\ N=124\ ,\ \sigma^2=16$$

$$A:\bar{X}:n(\mu,\sigma/\sqrt{n})=n(1.5,\ 0.5)$$

$$P(0.5<\bar{X}<2)=P(\frac{0.5-1.5}{0.5}<\frac{\bar{X}-1.5}{4/8}<\frac{2-1.5}{0.5})=P(-2<Z<1)=?$$

$$B)\bar{X}:n(\mu,\frac{\sigma}{\sqrt{n}}\sqrt{(N-n)/(N-1)})=n(1.5,\ 0.5\times0.5)$$

$$P(0.5<\bar{X}<2)=P(\frac{0.5-1.5}{0.25}<\frac{\bar{X}-1.5}{0.25}<\frac{2-1.5}{0.25})$$

توزیع نمونه میانگین

هنگام استفاده از قضیه حد مرکزي هرگاه واریانس جامعه معلوم نباشد میتوان به جاي واریانس جامعه (σ^2) از واریانس نمونه (S^2) استفاده کرد اما در این حالت توزیع $\left[\frac{\bar{X}-\mu}{S/\sqrt{n}}\right]$

اگر $n\geq 30$ باشد، توزیع نرمال استاندارد است.

ولي اگر $n<30$ باشد، داراي توزیع t با n-1 درجه آزادي است..

مثال: تولید کنندهاي ادعا ميکند که محصولات کارخانهاش داراي حد متوسط ۱۸,۳ کیلوگرم است، يك نمونه ۸ تایي از محصولات کارخانه را وزن کردیم که میانگین و واریانس نمونه آن به ترتیب، ۱۹,۵ و ۴,۲۸ شد، آیا با ادعاي کارخانهدار موافق هستید؟

حل: چون $n<30$ و واریانس جامع معلوم نمي باشد پس براي استفاده از واریانس نمونه به جاي واریانس جامعه باید از توزیع t استفاده کنیم یعني توزیع نمونه، توزیع t با ۷ درجه آزادي است

$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}=\frac{19.5-18.3}{0.73}=1.64$$

$$\text{From table:}(t_{0.975},\nu=7)=2.36 \Rightarrow$$

$$P(-2.36<T<-2.36)=2P(T<2.36)-1=0.95$$

$$\boxed{-2.36<1.64<2.36}$$

پس با احتمال ۹۵% ادعاي کارخانه دار درست است.

آماره ($\bar{X}_1-\bar{X}_2$): در بسیاري از مواقع لازم است که میانگینهاي دو جامعه با یکدیگر مقایسه شوند براي این منظور قضیه زیر بیان ميگردد.

قضیه: اگر نمونه هاي تصادفي با حجمهاي n_1 و n_2 از دو جامعه با میانگینهاي μ_1 و μ_2 و واریانسهاي σ^2_1 و σ^2_2 انتخاب شوند آنگاه توزیع نمونه $\bar{X}_1-\bar{X}_2$

اگر n_1 , $n_2\geq 30$، بدون توجه به توزیع دو جامعه، نرمال با میانگین $\mu_1-\mu_2$ و واریانس $(\sigma^2_1/n_1+\sigma^2_2/n_2)$ خواهد بود وگرنه در صورت نرمال بودن دو جامعه، نرمال با میانگین و واریانس ذکر شده خواهد بود.

مثال: دو نمونه گیري از دو جامعه نرمال طبق جدول زیر انجام ميدهیم مطلوبست

جامعه اول	جامعه دوم
$\mu_1=80$	$\mu_2=75$
$\sigma_1=5$	$\sigma_2=3$
$n_1=25$	$n_2=36$

$$P(3.4<\bar{X}_1-\bar{X}_2<8.9)=?$$

$$\bar{X}_1-\bar{X}_2 \text{ is Normal } \& \ \mu_{\bar{X}_1-\bar{X}}=80-75\ ,$$

$$\sigma_{\bar{X}_1-\bar{X}}=\sqrt{25/25+9/36}=\sqrt{5/4}$$

$$P(3.4<\bar{X}_1-\bar{X}_2<8.9)=P(\frac{3.4-5}{\sqrt{5/4}}<Z<\frac{8.9-5}{\sqrt{5/4}})=$$

$$P(-1.34<3.48)=?$$

آماره ترتیب : اگر X_1، X_2، ...، X_n يك نمونه تصادفي از متغیر تصادفي X باشند و آنها را به طور غیر نزولي مرتب کنیم

$$Y_1\leq Y_2\leq\ldots\leq Y_n$$

(که Y_iها مرتب شده X_iها هستند) آنگاه Y_i را iامین آماره ترتیب مينامیم که میانه نمونه اگر داده ها فرد باشند برابر است با داده وسطي و اگر زوج باشند برابر با میانگین حسابي دو داده وسطي است. و برد نمونه نیز برابر است با $R=Y_n-Y_1$

قضیه : به ازاي nهاي بزرگ، توزیع نمونهاي میانه نمونه تصادفي به اندازه 2n+1، تقریباً نرمال با میانگین μ و واریانس $\pi\sigma^2/4n$ ميباشد

اگر واریانس میانه را براي نمونههایي به اندازه 2n+1 از يك جامعه نامتناهي با واریانس میانگین نمونه یعني ($\sigma^2/2n+1$) مقایسه کنیم در ميیابیم که در نمونههاي بزرگ از يك جامعه نرمال میانگین قابل اعتمادتر ازمیانه است.

تمرين:ميانگين نمرات تست هوش دانشجويان سال اول يك دانشكده ۵۴۰ و انحراف معيار آن ۵۰ است دو نمونه تصادفي با حجم $n_1=32$ و $n_2=50$ انتخاب مي‌كنيم مطلوبست احتمال اينكه تفاضل ميانگين نمرات اين دو نمونه

الف: بيش از ۲۰ باشد

ب: بين ۵ و ۱۰ باشد

تمرين: از يك جامعه نرمال با ادعاي ميانگين ۲۰ و واريانس مجهول، يك نمونه تصادفي ۹ تايي با ميانگين و انحراف معيار ۲۴ و ۴,۱ انتخاب ميكنيم آيا ادعاي ميانگين ۲۰ مورد قبول است؟

تمرين: از جامعه اي نرمال با ميانگين ۴۳/۲ و واريانس ۳۹/۶۹ يك نمونه تصادفي ۳۶ تايي انتخاب مي‌كنيم، احتمال اينكه ميانگين اين نمونه بين ۴۲ و ۴۵ باشد چقدر است؟

تمرين: طول قد ۳۰۰۰ كودك داراي توزيع نرمال با ميانگين ۶۸ و انحراف معيار ۳ مي‌باشد، اگر ۸۰ نمونه تصادفي ۲۵ تايي از اين جامعه انتخاب كنيم، ميانگين و واريانس $\bar{X}$ را حساب كنيد در صورتيكه نمونه برداري

الف با جايگذاري باشد   ب: بدون جايگذاري باشد

پ: چند ميانگين نمونه در فاصله ۶۸/۸-۶۶/۳ قرار مي‌گيرند

ت:چند ميانگين نمونه كمتر از ۶۶/۴ قرار گيرند؟

تمرين: اگر S_1^2 و S_2^2 به ترتيب واريانس‌هاي دو نمونه تصادفي مستقل از دو جامعه نرمال باشند، در صورتيكه واريانسهاي دو جامعه مساوي بوده و نمونه هاي با حجم $n_1=8$ و $n_2=12$ انتخاب شده باشند، مطلوبست محاسبه

$P(S_1^2>4.89\ S_2^2)$

فصل دهم: نظريه برآورد كردن
(تئوري تخمين)

در اين فصل هدف برآورد پارامترهاي جامعه از روي آماره‌ها و قضاياي معرفي شده در فصل قبل مي‌باشد.

نظريه تصميمگيري: به روشهايي اطلاق مي‌شود كه به كمك آن از نمونه منتخب جامعه، بتوان استنباطهايي را در مورد پارامترهاي جامعه به دست آورد. كه داراي دو شاخه اصلي مي‌باشد:

- نظريه برآورد كردن (تئوري تخمين)
- آزمون فرض‌ها

پارامترهاي جامعه را به دو طريق مي‌توان برآورد كرد (تخمين زد):

الف: برآورد نقطه‌اي: مقدار يك آماره را براي پارامتر يك جامعه به كار مي‌بريم.

ب:برآورد فاصله‌اي: پارمترهاي جامعه را به صورت يك بازه قابل اعتماد برآورد مي‌كنيم.

برآورد نقطه‌اي

به طور كلي پارامتر جامعه را با θ، يك برآورد نقطه‌اي آن را با $\hat{\theta}$ و آماره‌اي كه براي بررسي اين پارامتر استفاده خواهد شد را با $\hat{\Theta}$ نشان مي‌دهيم به عنوان مثال اگر μ ميانگين جامعه مورد بررسي باشد:

$$\theta=\mu \quad , \quad \hat{\theta}=\bar{X} \quad , \quad \hat{\Theta}=\bar{X}$$

برآوردگر: آماره‌اي را كه براي تخمين نقطه‌اي پارامتري استفاده شود، برآوردگر مي‌ناميم يك پارامتر ممكن است داراي چند برآوردگر باشد براي مثال براي θ=μ سه برآوردگر شناخته شده ميانگين، ميانه و مد نمونه، وجود دارد حال اين سئوال مطرح است كه بين برآوردگرهاي يك پارامتر، كداميك بهتر است؟

براي جواب دادن به سئوال بالا تعاريف زير را انجام مي‌دهيم

تعريف: برآوردگر $\hat{\Theta}$ را يك برآوردگر ناريب پارامتر θ گوييم اگر و فقط اگر $E(\hat{\Theta})=\theta$

براي مثال: $\bar{X}$ و S^2 به ترتيب برآورد كننده‌هاي ناريب براي ميانگين و واريانس جامعه خواهد بود زيرا:

$$E(\bar{X})=\mu \quad , \quad E(S^2)=\sigma^2$$

تعريف: از ميان برآوردگرهاي ناريب پارامتر θ، برآوردي كه كمترين واريانس را داشته باشد، كاراترين برآوردگر ناميده مي‌شود.

مثال: در برآورد ميانگين يك جامعه نرمال برمبناي يك نمونه تصادفي به اندازه 2n+1، كارايي ميانه نسبت به ميانگين چيست؟

حل: مي‌دانيم كه هر دو برآورد ناريب است ولي

$$Var(\bar{X})=\frac{\sigma^2}{2n+1} \quad , \quad Var(\tilde{X})=\frac{\pi\sigma^2}{4n}$$

$$\frac{Var(\bar{X})}{Var(\tilde{X})}=\frac{\frac{\sigma^2}{2n+1}}{\frac{\pi\sigma^2}{4n}}= \Rightarrow$$

$$\lim_{n\to\infty}\frac{4n}{\pi(2n+1)}=\frac{2}{\pi}=0.64<1$$

قضيه: اگر $\hat{\Theta}$ يك برآوردگر ناريب پارامتر θ باشد و

$$\operatorname{var}(\hat{\Theta})=\frac{1}{nE\left[\left(\frac{\partial \ln f(x)}{\partial\theta}\right)^2\right]}$$

آنگاه $\hat{\Theta}$ يك برآوردگر ناريب با كمترين واريانس از پارامتر θ است.

مثال: نشان دهيد كه ميانگين نمونه يك برآورد كننده ناريب با كمترين واريانس ميانگين جامعه نرمال است؟

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad : x\in\mathbb{R}$$

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

$$E\left(\left(\frac{\partial \ln f(x)}{\partial\mu}\right)^2\right)=\frac{1}{\sigma^2}E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]=\frac{1}{\sigma^2}.1=\frac{1}{\sigma^2}$$

$$\frac{1}{nE\left[\left(\frac{\partial \ln f(x)}{\partial\theta}\right)^2\right]}=\frac{1}{n\frac{1}{\sigma^2}}=\frac{\sigma^2}{n}$$

$$\bar{X}\ \text{is unbiased}\ \&\ V(\bar{X})=\frac{\sigma^2}{n}$$

تعريف: آماره $\hat{\Theta}$ يك برآوردكننده سازگار پارامتر θ است اگر و فقط اگر به ازاي هر ثابت c،

$$\lim_{n\to\infty}P(|\hat{\Theta}-\theta|\ge c)=0 \quad OR \quad \lim_{n\to\infty}P(|\hat{\Theta}-\theta|<c)=1$$

قضيه: آماره $\hat{\Theta}$ يك برآوردكننده سازگار پارامتر θ است اگر

الف: $\hat{\Theta}$ ناريب باشد

ب: $\lim_{n\to\infty}\operatorname{var}(\hat{\Theta})=0$

مثال: واريانس نمونه (S^2) يك برآوردگر سازگار است زيرا:

$$E(S^2)=\sigma^2 \quad \Rightarrow S^2 \text{ is unbiased}$$

$$\lim_{n\to\infty} Var(S^2)=\lim_{n\to\infty}\frac{2\sigma^4}{n-1}=0$$

شرط گفته شده براي سازگاراي برآوردگر يك شرط كافي است نه لازم يعني برآورد كننده اي مي‌تواند سازگار باشد بدون اينكه ناريب باشد. ولي در اين حالت نيز بايد مجانباً ناريب باشد. براي مثال برآوردكننده $(X+1)/(n+2)$ يك برآوردكننده پارامتر دو جمله اي است ولي

$$E\left[\frac{(X+1)}{(n+2)}\right]=\frac{1}{(n+2)}E(X+1)=\frac{1}{(n+2)}(E(X)+1)\neq\theta$$

$$\lim_{n\to\infty} E\left[\frac{(X+1)}{(n+2)}\right]=E\lim_{n\to\infty}\frac{(X+1)}{(n+2)}=\theta$$

تعريف: آماره $\hat\Theta$ يك برآوردگر كافي پارامتر θ است اگر و تنها اگر با ازاي هر مقدار $\hat\Theta$ ، توزيع شرطي نمونه تصادفي X_1، X_2، ...، X_n به شرط $\hat\Theta=\hat\theta$ مستقل از θ باشد.

قضيه: آماره $\hat\Theta$ يك برآوردگر كافي پارامتر θ است اگر و تنها اگر، بتوان تابع احتمال توأم نمونه تصادفي را تجزيه كرد به طوري كه

$$f(x_1,x_2,\ldots,x_n\,|\,\theta)=g(\hat\theta,\theta)\cdot h(x_1,x_2,\ldots,x_n)$$

كه در آن g تنها به θ و $\hat\theta$ بستگي دارد و h به θ بستگي ندارد.

مثال: نشان دهيدكه ميانگين نمونه يك برآوردگر كافي ميانگين جامعه نرمال با واريانس معلوم σ^2 است؟

$$f(x_1,x_2,\ldots,x_n\,|\,\mu)=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\cdot e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}$$

$$\left\{\sum_{i=1}^{n}(X_i-\mu)^2=\sum_{i=1}^{n}(X_i-\bar X)^2+n(\bar X-\mu)^2\right\}$$

$$f(x_1,x_2,\ldots,x_n\,|\,\mu)=\left[\frac{\sqrt n}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\bar X-\mu}{\sigma/\sqrt n}\right)^2}\;e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{X_i-\bar X}{\sigma}\right)^2}\right]$$

براي برآورد پارامترهاي جامعه به روش برآورد نقطه‌اي روشهاي متعددي وجود دارد كه به دو روش مهم آن اشاره مي‌كنيم:

روش گشتاورها: عبارتست از حل دستگاه معادلات زير:

$$M'_k=\mu'_k \qquad k=1,2,\ldots,p$$

كه در آن p تعداد پارامتر جامعه است.

مثال: نمونه‌اي تصادفي به اندازه n از جامعه گاما داريم، براي برآورد كردن پارامترهاي α و β روش گشتاورها را به كار بريد؟

$$M'_1=\mu'_1 \quad , \quad M'_2=\mu'_2$$

$$\{\mu'_1=\alpha\beta \quad , \quad \mu'_2=\alpha(\alpha+1)\beta^2\}$$

$$\hat\alpha=\frac{(M'_1)^2}{M'_2-(M'_1)^2} \qquad \hat\beta=\frac{M'_2-(M'_1)^2}{M'_1}$$

روش درستنمايي ماكسيمم: در اين روش مقادير يك نمونه تصادفي را در نظر مي‌گيريم، سپس به عنوان برآورد خود از پارامتر مجهول جامعه، مقداري را بر مي‌گزينيم كه به ازاي آن احتمال به دست آوردن داده‌هاي مشاهده شده، ماكسيمم گردد يعني ماكسيمم تابع

$$L(\theta)=f(x_1,x_2,\ldots,x_n;\theta)$$

را نسبت به θ كه تابع درستنمايي ناميده مي‌شود، برآورد درستنمايي ماكسيمم θ مي‌ناميم.

مثال: با مفروض بودن x پيروزي در n امتحان برآوردگر درستنمايي ماكسيمم پارامتر θ را در توزيع دو جمله‌اي به دست آوريد؟

$$L(\theta)=b(x;n,\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

$$\ln L(\theta)=\ln\binom{n}{x}+x\ln\theta+(n-x)\ln(1-\theta)$$

$$\frac{d\ln L(\theta)}{d\theta}=\frac{x}{\theta}-\frac{n-x}{1-\theta}=0\Rightarrow\theta=\frac{x}{n}$$

$$\hat{\Theta}=\frac{x}{n}$$

برآورد فاصله اي

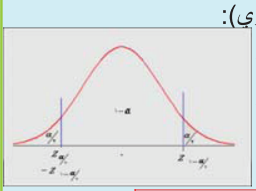
برآورد فاصله اي: فرض كنيم $1-\alpha$ احتمال مشخص ولي بزرگي باشد (۰/۹۵ يا ۰/۹۹ يا ...)$(0<\alpha<1)$، يك برآورد فاصله اي براي پارامتري مانند θ عبارتست از فاصله‌ي بازي مانند (L,U)، تابعي از نمونه تصادفي X_1، X_2، ... ، X_n، كه با احتمال $1-\alpha$ داشته باشيم:

$$P(L<\theta<U)=1-\alpha$$

$1-\alpha$را سطح اطمينان (ضريب اطمينان)مي ناميم و معمولا با درصد بيان مي شود .پس در برآورد فاصله اي بايد فاصله ي بازي مانند (L,U) را بيابيم كه با احتمال مشخص $1-\alpha$، تعريف بالا برقرار باشد.

برآورد فاصله‌اي ميانگين جامعه (μ):

وقتي نمونه تصادفي با حجم n از جامعه‌اي با واريانس معلوم σ^2 انتخاب شود فاصله اطمينان براي μ به شكل زير محاسبه مي‌گردد (استفاده از قضيه حد مركزي):



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

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

$$\boxed{\bar{X}-\frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}<\mu<\bar{X}+\frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}}$$

نكته‌هاي كاربردي:

- اگر نمونه تصادفي داراي حجم $n\geq30$ باشد، بدون توجه به نوع توزيع جامعه مي‌توان از فرمول بالا استفاده نمود.
- اگر نمونه تصادفي داراي حجم $n<30$ باشد، بايد توزيع جامعه نرمال باشد تا بتوان از فرمول بالا استفاده نمود.
- اگر نمونه تصادفي داراي حجم $n\geq30$ باشد، در صورت نامعلوم بودن واريانس جامعه مي‌توان به جاي σ^2 از واريانس نمونه (S^2) استفاده نمود.

نكته‌هاي كاربردي:

- با توجه به رابطه گفته شده طول فاصله اطمينان برابر است با:

$$L=2\frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}$$

پس براي اينكه فاصله اطمينان خوبي داشته باشيم بايد حجم نمونه را زياد كنيم.

- محاسبه خطا در برآورد μ:

$$-\frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}<\mu-\bar{X}<\frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}\Rightarrow|\mu-\bar{X}|<\frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}$$

نكته‌هاي كاربردي:

• حداكثر خطا در برآورد μ برابر است با:

$$e = \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2}$$

• مي‌توان گفت خطا كمتر از مقدار e است اگر حجم نمونه از رابطه زير به دست آيد:

$$n = \left[\frac{\sigma}{\sqrt{e}} Z_{1-\alpha/2}\right]^2$$

مثال: فرض مي كنيم عمر كامپيوترهاي توليدي يک موسسه از توزيع نرمال با ميانگين μ و واريانس ۰/۲۵ سال پيروي مي‌كند. بر اساس نمونه ي تصادفي از اين جامعه مقادير زير حاصل شده است :

۷ - ۶ - ۹- ۵ - ۸ - ۵ - ۶ - ۷ - ۱۰

يک فاصله اطمينان ۰/۹۵ براي μ بيابيد . مقدار حداكثر خطا را براي اين فاصله اطمينان محاسبه نماييداگر بخواهيم با اطمينان ۹۵% خطايي كمتر از ۲۵/. داشته باشيم حجم نمونه را چه تعداد بايد انتخاب كنيم.

$$\bar{X} = \frac{\sum x_i}{n} = \frac{63}{9} = 7$$

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow Z_{1-\alpha/2} = Z_{0.975} = 1.96$$

$$\mu : 7 \pm (1.96)(\frac{0.5}{\sqrt{9}}) \Rightarrow \mu : (6.67, 7.33)$$

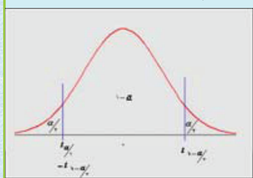
$$7 - (1.96)(\frac{0.5}{\sqrt{9}}) < \mu < 7 + (1.96)(\frac{0.5}{\sqrt{9}}) \Rightarrow 6.67 < \mu < 7.33$$

$$e = \frac{\sigma}{\sqrt{n}} Z_{1-\alpha/2} = \frac{0.5}{3} \times 1.96 = 0.64$$

$$n = \left[\frac{\sigma}{\sqrt{e}} Z_{1-\alpha/2}\right]^2 = (\frac{0.5}{0.25} \times 1.96)^2 = (3.92)^2 = 15.36 \approx 16$$

برآورد فاصله‌اي ميانگين جامعه (μ):

وقتي نمونه تصادفي با حجم n<30 از جامعه‌اي نرمال با واريانس نامعلوم σ^2 انتخاب شود فاصله اطمينان براي به شكل زير محاسبه مي‌گردد :



$$P(-t_{1-\alpha/2,n-1} < T < t_{1-\alpha/2,n-1}) = 1-\alpha \Rightarrow$$

$$P(-t_{1-\alpha/2,n-1} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{1-\alpha/2,n-1}) = 1-\alpha \Rightarrow$$

$$\boxed{\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1} < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1}}$$

$$e = \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1} \quad , n = \left[\frac{S}{\sqrt{e}} t_{1-\alpha/2,n-1}\right]^2$$

مثال: يك نمونه تصادفي ۱۰ تايي با ميانگين نمونه ۴/۳۸ و انحراف معيار نمونه ۰/۰۶ از يك جامعه نرمال با ميانگين μ و واريانس نامعلوم σ^2 انتخاب مي‌كنيم مطلوبست يك فاصله اطمينان ۹۵ % براي ميانگين جامعه:

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.5$$

$$t_{1-\alpha/2,n-1} = t_{0.975,9} = 2.26$$

$$\bar{X} - \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1} < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{1-\alpha/2,n-1}$$

$$4.38 - \frac{0.06}{\sqrt{10}} \times 2.26 < \mu < 4.38 + \frac{0.06}{\sqrt{10}} \times 2.26$$

$$4.338 < \mu < 4.22$$

برآورد فاصله‌اي تفاضل ميانگين دو جامعه (μ_1-μ_2):

حالت اول: فرض مي كنيم دو نمونه ي تصادفي مستقل از هم به حجم n_1 و n_2 از دو جامعه‌ي نرمال با ميانگين‌هاي μ_1 و μ_2 و واريانس‌هاي معلوم σ_1^2 و σ_2^2 استخراج كرده باشيم يک برآورد فاصله‌اي براي تفاضل ميانگين دو جامعه (μ_1-μ_2) با ضريب اطمينان 1-α به صورت زير محاسبه مي‌گردد:

$$(\bar{X}_1 - \bar{X}_2) - \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} Z_{1-\alpha/2} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} Z_{1-\alpha/2}$$

$$\mu_1 - \mu_2 : (\bar{X}_1 - \bar{X}_2) \pm \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} Z_{1-\alpha/2}$$

نكته‌هاي كاربردي:

• اگر نمونه‌هاي تصادفي داراي حجم $n_1, n_2 \geq 30$ باشد، بدون توجه به توزيع دو جامعه مي‌توان از فرمول بالا استفاده نمود.

• اگر نمونه‌هاي تصادفي داراي حجم $n_1, n_2 < 30$ باشد، بايد توزيع دوجامعه نرمال باشد تا بتوان از فرمول بالا استفاده نمود.

• اگر نمونه تصادفي داراي حجم $n_1, n_2 \geq 30$ باشد، در صورت نامعلوم بودن واريانس‌هاي جامعه مي‌توان به جاي σ_1^2 و σ_1^2 از واريانس‌هاي نمونه (S_1^2 و S_2^2) استفاده نمود.

حالت دوم:

•اگر نمونه‌هاي تصادفي داراي حجم $n_1, n_2<30$ از دو جامعه نرمال باشد در صورت نامعلوم بودن واريانس‌هاي جامعه ولي با فرض برابري واريانس‌ها $(\sigma_1^2=\sigma_1^2)$ برآورد تفاضل ميانگين‌ها به شكل زير در مي‌آيد:

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \longleftarrow \text{واريانس ادغام شده}$$

$$\mu_1-\mu_2:(\bar{X}_1-\bar{X}_2)\pm S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,t_{1-\alpha/2,n_1+n_2-2}$$

نكته‌هاي كاربردي

•چنانچه واريانس‌هاي دو جامعه مساوي نباشند ولي $n_1= n_2$ باز مي‌توان از رابطه بالا استفاده نمود.

•چنانچه واريانس‌هاي دو جامعه مساوي نباشند و امكان انتخاب دو نمونه با حجم‌هاي مساوي نباشند مي‌توانيم از رابطه زير استفاده كنيم:

$$v\approx\frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\left[\left(\frac{S_1^2}{n_1}\right)^2/n_1-1\right]\left[\left(\frac{S_2^2}{n_2}\right)^2/n_2-1\right]}$$

$$\mu_1-\mu_2:(\bar{X}_1-\bar{X}_2)\pm\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\,t_{1-\alpha/2,v}$$

مثال: از دو جامعه نرمال با ميانگين‌هاي μ_1 و μ_2 و واريانس‌هاي $\sigma_1^2=25$ و $\sigma_2^2=9$ دو نمونه تصادفي مستقل با حجم‌هاي $n_1=25$ و $n_2=36$ انتخاب مي‌كنيم اگر ميانگين نمونه اول ۸۰ و نمونه دوم ۷۵ باشد مطلوبست فاصله اطمينان ۹۴% براي $\mu_1-\mu_2$ ؟

$$1-\alpha=0.94\Rightarrow 1-\alpha/2=0.97\quad,\quad Z_{0.97}=1.88$$

$$\mu_1-\mu_2:(\bar{X}_1-\bar{X}_2)\pm\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\,Z_{1-\alpha/2}$$

$$\mu_1-\mu_2:(80-75)\pm\sqrt{\frac{25}{25}+\frac{9}{36}}\times1.88$$

$$2.9<\mu_1-\mu_2<7.1$$

مثال: از دو جامعه نرمال با واريانس‌هاي مساوي $\sigma_1^2=\sigma_2^2$، دو نمونه تصادفي مستقل $n_1=9$ و $n_2=16$ انتخاب كرده‌ايم و نتايج زير به دست آمده است.

$$\bar{X}_1=64\ ,\ S_1^2=36\ ,\ \bar{X}_2=59\ ,\ S_2^2=25$$

۹۵% فاصله اطمينان را براي $\mu_1-\mu_2$ بدست آوريد.

$$1-\alpha=0.95\ \Rightarrow\ 1-\alpha/2=0.975$$

$$v=n_1+n_2-2=23\ ,\ t_{0.975,23}=2.07$$

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=\frac{1}{23}(8\times36+15\times25)=28.83$$

$$\mu_1-\mu_2:(\bar{X}_1-\bar{X}_2)\pm S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\,t_{1-\alpha/2,n_1+n_2-2}\Rightarrow$$

$$\mu_1-\mu_2:(64-59)\pm5.37\sqrt{\frac{1}{9}+\frac{1}{25}}\times2.07\Rightarrow$$

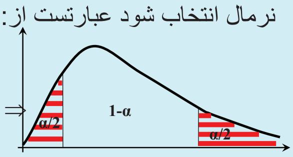
$$0.68<\mu_1-\mu_2<9.32$$

برآورد فاصله‌اي واريانس جامعه:

برآورد فاصله‌اي با ضريب اطمينان $(1-\alpha)$ براي واريانس جامعه (σ^2) هنگاميكه نمونه تصادفي با حجم n از يك جامعه نرمال انتخاب شود عبارتست از:

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

$$P(\chi_{\alpha/2}^2<X^2<\chi_{1-\alpha/2}^2)=1-\alpha\Rightarrow$$



$$\boxed{\frac{(n-1)S^2}{\chi_{1-\alpha/2,(n-1)}^2}<\sigma^2<\frac{(n-1)S^2}{\chi_{\alpha/2,(n-1)}^2}}$$

مثال: يك نمونه تصادفي با حجم $n=20$ از يك جامعه نرمال داراي ميانگين نمونه ۳۲/۸ و انحراف معيار ۴/۵۱ انتخاب مي‌كنيم. يك فاصله اطمينان ۹۵% براي σ^2 و σ به دست آوريد؟

$$1-\alpha=0.95\Rightarrow\alpha=0.05\ ,\ \alpha/2=0.025\ ,\ 1-\alpha/2=0.975$$

$$\chi_{0.975,19}^2=32.9\ ,\ \chi_{0.025,19}^2=8.91$$

$$\frac{(n-1)S^2}{\chi_{1-\alpha/2,(n-1)}^2}<\sigma^2<\frac{(n-1)S^2}{\chi_{\alpha/2,(n-1)}^2}$$

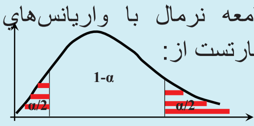
$$\frac{19\times(4.51)^2}{32.9}<\sigma^2<\frac{19\times(4.51)^2}{8.91}\Rightarrow$$

$$11.75<\sigma^2<43.37$$

$$3.42<\sigma<6.58$$

برآورد فاصله‌اي نسبت واريانس‌هاي دو جامعه:

برآورد فاصله‌اي با ضريب اطمينان (1-α) براي (σ_1^2/σ_2^2)، وقتي دو نمونه تصادفي مستقل با حجم‌هاي n_1 و n_2 از دو جامعه نرمال با واريانس‌هاي σ_1^2 و σ_2^2 انتخاب شوند، عبارتست از:



$$F=\frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

$$P(f_{\alpha/2}(n_1-1,n_2-1)<F<f_{1-\alpha/2}(n_1-1,n_2-1))=1-\alpha \Rightarrow$$

$$\boxed{\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{1-\alpha/2}(n_1-1,n_2-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}f_{1-\alpha/2}(n_2-1,n_1-1)}$$

مثال: يك فاصله اطمينان ٩٨% براي نسبت (σ_1^2/σ_2^2) و (σ_1/σ_2) از دو جامعه نرمال تعيين كنيد در صورتيكه واريانس‌هاي دو نمونه تصادفي با حجم‌هاي $n_1=9$ و $n_2=16$ از اين دو جامعه به ترتيب $S_1^2=36$ و $S_2^2=25$ باشد؟

$$1-\alpha=0.98 \Rightarrow \alpha=0.02\ ,1-\alpha/2=0.99$$

$$f_{0.99}(8,15)=4\ \ ,\ f_{0.99}(15,8)=5.52$$

$$\frac{S_1^2}{S_2^2}\cdot\frac{1}{f_{1-\alpha/2}(n_1-1,n_2-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2}{S_2^2}f_{1-\alpha/2}(n_2-1,n_1-1)$$

$$\frac{36}{25}\times\frac{1}{4}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{36}{25}\times 5.52$$

$$0.36<\frac{\sigma_1^2}{\sigma_2^2}<7.59$$

$$0.6<\frac{\sigma_1^2}{\sigma_2^2}<2.81$$

برآورد فاصله‌اي نسبت (p): يكي از پارامترهايي كه معمولاً مورد بررسي قرار مي‌گيرد پارامتر p، نسبت يا احتمال موفقيت در آزمايش دو جمله‌اي است. يك برآوردگر ناريب براي اين پارامتر برآوردگر $\hat{P}=\frac{X}{n}$ است. (متغير تصادفي X تعداد موفقيت در آزمايش دو جمله‌اي است) بنابراين اين برآوردگر يك برآورد نقطه‌اي خوب براي نسبت p است مي‌دانيم وقتي n بزرگ باشد اين برآوردگر توزيع نرمال با ميانگين و واريانس زير مي‌باشد

$$\mu_{\hat{P}}=E(\hat{P})=E(\frac{X}{n})=\frac{np}{n}=p$$

$$\sigma_{\hat{P}}^2=\mathrm{var}(\frac{X}{n})=\frac{1}{n^2}\sigma_X^2=\frac{npq}{n^2}=\frac{pq}{n}$$

برآورد فاصله‌اي نسبت p (نسبت در توزيع دوجمله‌اي) :

برآورد فاصله‌اي براي نسبت p با ضريب اطمينان (1-α) وقتي حجم نمونه تصادفي n≥30 باشد عبارتست:

$$Z=\frac{\hat{P}-p}{\sqrt{\hat{p}\hat{q}/n}}\sim n(0,1)\ \Rightarrow P(-Z_{1-\alpha/2}<Z<Z_{1-\alpha/2})=1-\alpha$$

$$\boxed{\hat{P}-\sqrt{\hat{p}\hat{q}/n}\,Z_{1-\alpha/2}<p<\hat{P}+\sqrt{\hat{p}\hat{q}/n}\,Z_{1-\alpha/2}}$$

در رابطه بالا حداكثر خطا و مقدار حجم نمونه براي اينكه خطايي كمتر از e داشته باشيم از رابطه زير به دست مي‌آيد:

$$e=\sqrt{\frac{\hat{p}\hat{q}}{n}}Z_{1-\alpha/2}\qquad n=\frac{1}{e^2}\hat{p}\hat{q}\,(Z_{1-\alpha/2})^2$$

مثال: در يك نمونه تصادفي با حجم n=640 مشخص شده است كه x=160 خانواده داراي كامپيوتر شخصي هستند، يك فاصله اطمينان ٩٥% براي نسبت واقعي تمام خانواده‌هاي كامپيوتردار تعيين كنيد، حداكثر خطا را تعيين كنيد و مشخص كنيد چقدر بايد نمونه داشته باشيم تا خطايي كمتر از ١% داشته باشيم؟

$$1-\alpha=0.95 \Rightarrow 1-\alpha/2=0.975$$

$$\hat{p}=\frac{X}{n}=\frac{160}{640}=0.25\ \ ,\ \ \hat{q}=1-\hat{p}=0.75$$

$$\hat{P}-\sqrt{\hat{p}\hat{q}/n}\,Z_{1-\alpha/2}<p<\hat{P}+\sqrt{\hat{p}\hat{q}/n}\,Z_{1-\alpha/2}$$

$$p:0.25\pm\sqrt{0.25\times0.75/640}\times1.96 \Rightarrow \boxed{0.21<p<0.28}$$

$$e=\sqrt{\hat{p}\hat{q}/n}\,Z_{1-\alpha/2}=\sqrt{0.25\times0.75/640}\times1.96=0.3$$

$$n=\frac{1}{e^2}\hat{p}\hat{q}\,(Z_{1-\alpha/2})^2=\frac{1}{(0.01)^2}\times0.25\times0.75\times(1.96)^2=7203$$

برآورد تفاضل دو نسبت (p_1-p_2) در توزيع دوجمله‌اي:

برآورد فاصله‌اي براي تفاضل پارامترهاي دو جامعه دوجمله‌اي (p_1-p_2) هنگاميكه دو نمونه تصادفي مستقل با حجم‌هاي $n_1,n_2\geq30$ انتخاب شوند عبارتست از:

$$\hat{p}_1=\frac{X_1}{n_1}\ \ ,\ \ \hat{q}_1=1-\hat{p}_1\ \ ,\ \ \hat{p}_2=\frac{X_2}{n_2}\ \ ,\ \ \hat{q}_2=1-\hat{p}_2$$

$$Z=\frac{(\hat{P}_1-\hat{P}_2)-(p_1-p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}}\sim n(0,1)\ \Rightarrow P(-Z_{1-\alpha/2}<Z<Z_{1-\alpha/2})=1-\alpha$$

$$\boxed{p_1-p_2:(\hat{P}_1-\hat{P}_2)\pm\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}\,Z_{1-\alpha/2}}$$

تمرين: اگر X_1، X_2، ... ، X_n مقادير يك نمونه تصادفي از جامعه‌اي نمايي باشند، برآورد درستنمايي پارامتر θي جامعه را پيدا كنيد؟

تمرين: فرض مي كنيم X_1، X_2، ... ، X_n مقادير يك نمونه تصادفي از توزيع زير باشد:

$$\boxed{f(x,\theta)=\theta^{x}(1-\theta)^{1-x}\ :x=0,1}$$

θ را به روش MLE برآورد كنيد

تمرين: فرض مي كنيم تابع چگالي احتمال X عبارتست از :

$$f(x,\theta)=\theta e^{-\theta x}\ :x>0$$

θ را به روش گشتاورها برآورد كنيد

تمرين: از جامعه‌اي نرمال، نمونه تصادفي با حجم n=5 انتخاب و نتايج زير به دست آمده است.

$X_1=3$	$X_2=1$	$X_3=4$	$X_4=3$	$X_5=4$

فاصله اطمينان ۹۹% براي ميانگين اين جامعه را به دست آوريد؟

تمرين: در صورتيكه ميانگين نمونه يك نمونه تصادفي با حجم n از يك جامعه نرمال با انحراف معيار ۳/. مفروض باشد، تعداد n را طوري تعيين كنيد كه ۹۹% فاصله اطمينان براي ميانگين جامعه به صورت $(\bar{X}-0.1,\bar{X}+0.1)$ باشد؟

تمرين: يك نمونه تصادفي $n_1=15$ تايي از يك جامعه با ميانگين نمونه ۶۰ و انحراف معيار نمونه ۳ است و نمونه تصادفي ديگري با حجم $n_2=21$ تايي از يك جامعه ديگر داراي ميانگين نمونه ۵۸ و انحراف معيار نمونه ۲ است، فرض كنيد دو جامعه نرمال با واريانس‌هاي مساوي باشند مطلوبست

الف: فاصله اطمينان ۹۵% براي تفاضل ميانگين دو جامعه؟

ب: اگر واريانس‌ها نامساوي ولي $n_1=n_2=10$ و داراي همان ميانگين نمونه باشد فاصله اطمينان ۹۵% براي تفاضل ميانگين دو جامعه؟

ج: اگر واريانس‌ها نامساوي باشند فاصله اطمينان ۹۵% براي تفاضل ميانگين دو جامعه؟

تمرين: يك نمونه تصادفي $n_1=150$ تايي از يك جامعه نرمال داراي ميانگين نمونه ۱۴۰۰ و انحراف معيار نمونه ۱۲۰ است و نمونه تصادفي ديگري با حجم $n_2=200$ از يك جامعه نرمال ديگر داراي ميانگين نمونه ۱۲۰۰ و انحراف معيار نمونه ۸۰ است، مطلوبست فاصله اطمينان ۹۵% براي تفاضل ميانگين دو جامعه؟

تمرين: يك نمونه ۱۶ تايي از محصلان يك مدرسه انتخاب و انحراف معيار را اندازه گرفتيم، S=2.4 سانتي‌متر است. فاصله اطمينان ۹۵% براي انحراف معيار دانش‌آموزان را در صورتيكه كه جامعه نرمال فرض گردد به دست آوريد؟

تمرين: توزيع جامعه معدل دانشجويان دو دانشگاه نرمال فرض مي‌شوند مطلوبست فاصله اطمينان ۹۸% براي نسبت واريانس‌ها و انحراف معيارهاي دو جامعه در صورتيكه دو نمونه منتخب تصادفي مستقل از اين دو جامعه به شكل زير باشد؟

دانشگاه A	3	3.2	2.9	3.4	2.5	2.8	3	3.3	2.7	3.1
دانشگاه A	3	2.6	3.1	3.5	3.2	3.9	2.7			

تمرين: سكه‌اي را ۴۰ بار پرتاپ مي‌كنيم و ۲۴ بار شير آمده است مطلوبست يك فاصله اطمينان ۹۵% براي نسبت واقعي شيرها؟

تمرين: مدير كارخانه‌اي كه دو نوع لامپ با ماركهاي A و B توليد مي‌كند، مدعي است كه لامپ‌هاي مارك A، ۸% بيشتر از مارك B فروش دارد. يك نمونه تصادفي با حجم $n_1=200$ نفر از خريداران انتخاب شده كه ۴۲ نفر از اين افراد از لامپ نوع A استفاده مي‌كنند. نمونه تصادفي ديگري با حجم $n_2=150$ نفر از خريداران انتخاب شده كه مشخص شده ۱۸ نفر از لامپ نوع B استفاده مي‌كنند يك فاصله اطمينان ۹۴% براي تفاضل p_1-p_2 تعيين كنيد، آيا با ادعاي مدير كارخانه موافقيد؟

فصل يازدهم: آزمون فرض‌ها

همان‌طور كه اشاره شد نظريه تصميم‌گيري (روش‌هاي رسيدن از نمونه به جامعه) به دو دسته تقسيم مي‌شود، يكي نظريه برآورد كردن (تئوري تخمين) و ديگري آزمون فرض‌ها.

آزمون فرض‌ها يكي از مهمترين شاخه‌هاي نظريه تصميم‌گيري است.

فرض آماري: فرض، بيان يا حدسي است درباره ي توزيع جامعه يا پارامتر جامعه كه درست بودن فرض آماري را بايد از نتايج به دست آمده از نمونه‌گيري بررسي نمود (آزمون فرض).

مثال : ادعا شده است متوسط عمر تراشه هاي توليدي يک كارخانه بزرگ برابر ۱۰۰۰ساعت است . بنابراين فرض آماري عبارتست از :

$$\theta=\mu=1000$$

فرض آماري: فرض ها به دو دسته تقسيم مي شوند :

يا ساده هستند يعني مقدار دقيقي براي پارامتر جامعه مشخص مي‌گردد مثل $(\theta=\theta_0)$

يا مرکب هستند که خود سه دسته‌اند.

يک طرفه از چپ $(\theta<\theta_0)$

يک طرفه از راست $(\theta>\theta_0)$

دو طرفه $(\theta\neq\theta_0)$

يك فرض آماري بايد چگونه بيان شود؟

همان طور که بيان شد، فرض يا ادعاهايي در مورد پارامتر يك جامعه بيان مي‌شود، از آنجا كه اين ادعا ممكن است درست يا نادرست باشد، لذا دو فرض مكمل مطرح است يكي آنكه (فرض: ادعا درست است) و ديگري آنكه (فرض: ادعا نادرست است). پس شروع يك آزمون فرض با دو فرض آماري در مقابل هم مي‌باشد فرض آماري كه براي رد شدن تنظيم مي‌شود را فرض صفر (خنثي) ناميده و آن را با H_0 نشان مي‌دهيم كه هميشه شرايط موجود را در نظر مي‌گيريم و فرض آماري را كه در مقابل فرض صفر قرار مي‌گيرد، فرض مقابل ناميده و با H_1 نشان مي‌دهيم كه اين فرض همواره طوري انتخاب مي‌شود كه مي‌خواهد شرايط موجود (فرض صفر) را تغيير دهد.

مثال:

$$H_0 : \mu = \theta_0$$
$$H_1 : \mu \neq \theta_0 \quad OR \quad \mu < \theta_0 \quad OR \quad \mu > \theta_0$$

$$H_0 : \mu_1 - \mu_2 = 0 \Leftrightarrow \mu_1 = \mu_2$$
$$H_1 : \mu_1 - \mu_2 \neq 0 \quad OR \quad \mu_1 - \mu_2 < \theta_0 \quad OR \quad \mu_1 - \mu_2 > \theta_0$$

$$H_0 : P = \theta_0$$
$$H_1 : P \neq \theta_0 \quad OR \quad P < \theta_0 \quad OR \quad P > \theta_0$$

براي پاسخ دقيق به آزمون آماري نياز است كه كل جامعه مورد بررسي قرار گيرد ولي مي‌دانيم كه اين كار ميسر نيست و پاسخ ما از روي نمونه جامعه خواهد بود پس آزمون آماري حتماً داراي خطا خواهد بود و سعي ما در حداقل ساختن اين خطاست.

براي حداقل ساختن اين خطاها بايد اول اين خطاها شناسايي گردند تا بتوان در مورد آنها بحث نمود. دو نوع خطا داريم

رد فرض صفر در حاليكه درست باشد، خطاي نوع اول و قبول كردن فرض صفر در حاليكه نادرست باشد را خطاي نوع دوم مي‌ناميم.

احتمال خطاي نوع اول را با α و احتمال خطاي نوع دوم را با β نشان مي‌دهيم:

α = P (رد فرض H_0 | درست باشد H_0) = P (RH_0 | H_0)

β = P (قبول فرض H_0 | درست باشد H_1) = P (AH_0 | H_1)

α را ميزان معني دار بودن يا سطح تشخيص مي‌ناميم.

ناحيه‌اي كه باعث رد فرض صفر مي‌شود ناحيه بحراني يا ناحيه رد و ناحيه‌اي كه باعث قبول فرض صفر مي‌شود را ناحيه قبولي و مرز بين اين دو ناحيه را نقطه بحراني مي‌ناميم. پس آزمون فرض عبارتست از افراز فضاي نمونه به دو قسمت ناحيه بحراني و ناحيه قبول.

- يك آزمون خوب آزموني است كه هر دو نوع خطا در آن كم باشند، بعبارت ديگر نه فرضي را بي‌مورد رد و نه بي‌مورد قبول كنيم.
- خطاي نوع اول و دوم به يكديگر وابسته‌اند و افزايش يكي باعث كاهش ديگري مي‌شود.
- با تغيير ناحيه بحراني مي‌توان احتمال خطاها را تغيير داد.
- با افزايش مقدار حجم نمونه، مي‌توان هر دو خطا را به طور همزمان كاهش داد.

مثال: فرض كنيد يك نوع كپسول سرماخوردگي (A) در كشور توليد مي‌شود كه پس از ۵ روز ۷۰% مؤثر است. كارخانه‌اي مدعي است كپسولي مشابه با نام B توليد مي‌كند كه اثر آن پس از ۵ روز بيشتر از ۷۰% است. فرض‌هاي آماري را براي بررسي ادعاي كارخانه جديد تشكيل دهيد:

$$H_0 : p = 0.7$$
$$H_1 : p > 0.7$$

يك نمونه ۲۰ تايي انتخاب مي‌كنيم آماره‌اي كه تصميم‌گيري بر اساس آن صورت مي‌گيرد تعداد افرادي است كه اثر كپسول B در آنها موثر بوده است اگر X را تعداد افرادي فرض كنيم كه كپسول B در آنها موثر بوده است اگر $X \geq 16$ را ناحيه بحراني (ناحيه رد) و $X<16$ ناحيه قبول در نظر مي‌گيريم. خطاهاي آزمون فرض را محاسبه مي‌كنيم:

$$\alpha = P(RH_0 \mid H_0) = P(X > 15.5 \mid p = 0.7) = \sum_{x=16}^{20} b(x;20,0.7) = 0.2375$$
$$\beta = P(AH_0 \mid H_1) = P(X < 15.5 \mid p > 0.7) = ?????$$

ملاحضه مي‌كنيد كه β قابل محاسبه نمي‌باشد زيرا فرض مقابل بطور مقابل مشخص نشده ولي اگر فرض مقابل را تغيير داده و به صورت (H_1: p=۰/۸) در نظر بگيريم، β قابل محاسبه است:

$$\beta = P(AH_0 \mid H_1) = P(X < 15.5 \mid p = 0.8) = \sum_{x=0}^{15} b(x;20,0.8) = 0.3704$$

با تغيير نقطه بحراني خطاها تغيير مي‌كنند در مثال بالا اگر نقطه بحراني را از ۱۵,۵ به ۱۴,۵، تغيير دهيم:

$$\alpha = P(RH_0 \mid H_0) = P(X > 14.5 \mid p = 0.7) = \sum_{x=15}^{20} b(x;20,0.7) = 0.4164$$
$$\beta = P(AH_0 \mid H_1) = P(X < 14.5 \mid p = 0.8) = \sum_{x=0}^{14} b(x;20,0.8) = 0.1958$$

گفتيم تصميم‌گيري خوب وقتي انجام مي‌پذيرد كه هر دو نوع خطا كم باشد براي كاهش همزمان هر دو نوع خطا بايد حجم نمونه را افزايش دهيم در مثال بالا اگر حجم نمونه را ۱۰۰ انتخاب كنيم و نقطه بحراني را به دلخواه ۷۳/۵ اختيار كنيم خطاها به صورت زير در مي‌آيد:

$$\alpha = P(RH_0 \mid H_0) = P(X > 73.5 \mid p = 0.7) = ???$$
$$np > 5 \ \& \ n \to \infty \Rightarrow b(x;n,p) \approx n(np,\sqrt{npq}) \Rightarrow$$
$$\mu = np = 100 \times 0.7 = 70 \ \ \& \ \sigma^2 = npq = 21 \Rightarrow b(x;100,0.7) \approx n(70,\sqrt{21})$$
$$\alpha \approx P(Z > \frac{73.5-70}{4.58}) = 1 - P(Z < 0.76) = 0.2236$$

$$\beta = P(AH_0 \mid H_1) = P(X < 73.5 \mid p = 0.8) = ???$$
$$np > 5 \ \& \ n \to \infty \Rightarrow b(x;n,p) \approx n(np,\sqrt{npq}) \Rightarrow$$
$$\mu = np = 100 \times 0.8 = 80 \ \ \& \ \sigma^2 = npq = 16 \Rightarrow b(x;100,0.7) \approx n(80,4)$$
$$\beta \approx P(Z < \frac{73.5-80}{4}) = P(Z < -1.625) = 0.0521$$

پس اگر آزمون فرض مربوط به توزيع گسسته باشد براي بررسي آن بايد نقطه بحراني را به طور دلخواه انتخاب و به كمك آن خطاهاي نوع اول و دوم را به دست آوريم و اگر اين خطاها بزرگ نباشند ناحيه بحراني را مشخص و چنانچه نتيجه آزمايش در ناحيه بحراني قرار گرفت فرض صفر را رد و در غير اينصورت فرض صفر پذيرفته مي‌شود. اما اگر احتمال خطاهاي نوع اول و دوم زياد شد با تغيير نقطه بحراني (با ثابت نگه داشتن حجم نمونه) و در صورت نرسيدن به مقدار مطلوب خطاها، در صورت امكان حجم نمونه را افزايش مي‌دهيم.

ولي اگر آماره‌اي كه تصميم‌گيري بر اساس آن صورت مي‌پذيرد داراي توزيع پيوسته باشد مي‌توان ابتدا خطاي نوع اول را اختيار كرد و به كمك آن ناحيه بحراني را تعيين نمود و نيازي نيست كه مانند گسسته‌ها ابتدا ناحيه بحراني را انتخاب و α محاسبه كنيم و در صورت قابل قبول نبودن، ناحيه بحراني را تغيير دهيم.

در مثال قبل براي تعيين احتمال خطاها يك فرض ساده را در برابر يك فرض ساده ديگر مي‌آزموديم ولي در عمل چنين نيست و در بيشتر مواقع با فرض‌هاي مركب سرو كار داريم و ناچار به تعيين احتمال خطاها در اين حالت هستيم براي اينكار از تابع توان آزمون فرض استفاده مي‌كنيم.

تابع توان يك آزمون فرض آماري H_0 در برابر فرض مقابل H_1 به صورت زير است:

$$\pi(\theta) = \begin{cases} \alpha(\theta) & \text{براي مقادير } \theta \text{ كه تحت } H_0 \text{ اختيار مي‌شوند} \\ 1-\beta(\theta) & \text{براي مقادير } \theta \text{ كه تحت } H_1 \text{ اختيار مي‌شوند} \end{cases}$$

بنابراين تابع توان براي مقادير θ تحت H_0، احتمال ارتكاب خطاي نوع اول و براي مقادير θ تحت H_1، احتمال مرتكب نشدن خطاي نوع دوم مي‌باشد.

تابع‌هاي توان نقش بسيار مهمي در ارزيابي آزمون‌هاي آماري، بويژه در مقايسه چندين ناحيه بحراني براي يك آزمون را دارند.

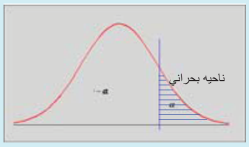
در صنعت تابع (1-π(θ)) به عنوان منحني مشخصه عمل كاربرد دارد كه اين تابع طبق تعريف احتمال‌هاي قبول H_0 به جاي رد آن است.

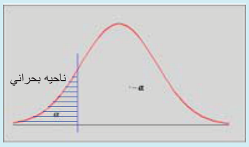
آزمون‌هاي يك طرفه و دو طرفه:

آزمون فرض را **يكطرفه** مي‌ناميم اگر فرض مقابل آن يكطرفه باشد:

$H_0: \theta = \theta_0$
$H_1: \theta > \theta_0$

$H_0: \theta = \theta_0$
$H_1: \theta < \theta_0$

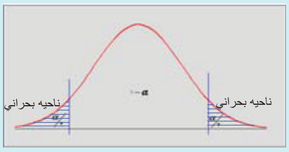




آزمون‌هاي يك طرفه و دو طرفه:

آزمون فرض را **دو طرفه** گوييم اگر فرض مقابل دو طرفه باشد.

$H_0: \theta = \theta_0$
$H_1: \theta \neq \theta_0$



معمولاً در آزمون فرض‌ها α را ۰/۰۵ يا ۰/۰۱ اختيار مي‌كنيم، اگر α=۰/۰۵ آزمون را **معني دار** و اگر α=۰/۰۱ آزمون را **بسيار معني دار** مي‌ناميم. اگر مقدار α مشخص نباشد آن را ۰/۰۵ يعني معني‌دار در نظر مي‌گيريم.

مراحل آزمون فرض‌ها:

۱- تشكيل فرض صفر $H_0: \theta = \theta_0$

۲- تشكيل فرض مقابل $H_1: \theta \neq \theta_0 \quad OR \quad \theta < \theta_0 \quad OR \quad H_1: \theta > \theta_0$

۳- انتخاب ميزان معني‌دار بودن

۴- انتخاب آماره مناسب براي پارمتر θ و تشخيص نوع توزيع آن طبق قضايا و مطالب فصل قبل، و سپس تشكيل ناحيه بحراني

۵- محاسبه مقدار آماره انتخابي به كمك نمونه تصادفي انتخاب شده

۶- نتيجه‌گيري: اگر مقدار آماره در ناحيه بحراني قرار گيرد فرض H_0 رد، در غير اينصورت فرض H_0 پذيرفته مي‌شود.

H_0	آماره آزمون	H_1	ناحيه بحراني
$\mu = \mu_0$ مثال	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ (σ معلوم، جامعه نرمال و n<30) (σ معلوم، هر جامعه‌اي و n≥30) (σ نامعلوم (استفاده از واريانس نمونه)، هر جامعه‌اي و n≥30)	$\mu < \mu_0$	$Z<-z_{1-\alpha}$
		$\mu > \mu_0$	$Z>z_{1-\alpha}$
		$\mu \neq \mu_0$	$Z<-z_{1-\alpha/2}$ $Z>z_{1-\alpha/2}$
$\mu = \mu_0$ مثال	$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}; v = n-1$ (σ نامعلوم (استفاده از S به جاي σ)، جامعه نرمال و n<30)	$\mu < \mu_0$	$T<-t_{1-\alpha}$
		$\mu > \mu_0$	$T>t_{1-\alpha}$
		$\mu \neq \mu_0$	$T<-t_{1-\alpha/2}$ $T>t_{1-\alpha/2}$

H_0	آماره آزمون	H_1	ناحيه بحراني
$\mu_1-\mu_2 = d_0$ مثال	$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$ (σ₁ و σ₂ معلوم، جوامع نرمال و $n_1,n_2<30$) (σ₁ و σ₂ معلوم، هر جامعه‌اي و $n_1,n_2\geq30$) (σ₁ و σ₂ نامعلوم (استفاده از واريانس نمونه)، هر جامعه‌اي و $n_1,n_2\geq30$)	$\mu_1- \mu_2 < d_0$	$Z<-z_{1-\alpha}$
		$\mu_1- \mu_2 < d_0$	$Z>z_{1-\alpha}$
		$\mu_1- \mu_2 \neq d_0$	$Z<-z_{1-\alpha/2}$ $Z>z_{1-\alpha/2}$

H_0	آماره آزمون	H_1	ناحيه بحراني
$\mu_1-\mu_2 = d_0$ مثال	$T = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{S_p\sqrt{(1/n_1) + (1/n_2)}}$ $v = n_1 + n_2 - 2$ $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$ (σ₁ و σ₂ نامعلوم ولي مساوي، جوامع نرمال و $n_1,n_2<30$) (σ₁ و σ₂ نامعلوم و نامساوي ولي $n_1=n_2$، جوامع نرمال و $n_1,n_2<30$)	$\mu_1- \mu_2 < d_0$	$T<-t_{1-\alpha}$
		$\mu_1- \mu_2 < d_0$	$T>t_{1-\alpha}$
		$\mu_1- \mu_2 \neq d_0$	$T<-t_{1-\alpha/2}$ $T>t_{1-\alpha/2}$

H_0	آماره آزمون	H_1	ناحيه بحراني
$\sigma^2=\sigma_0^2$ مثال	$X^2=\frac{(n-1)S^2}{\sigma_0^2}$ $v=n-1$ جامعه نرمال	$\sigma^2<\sigma_0^2$	$X^2<\chi^2_{\alpha}$
		$\sigma^2>\sigma_0^2$	$X^2>\chi^2_{1-\alpha}$
		$\sigma^2 \neq \sigma_0^2$	$X^2<\chi^2_{\alpha/2}$ $X^2>\chi^2_{1-\alpha/2}$
$\sigma_1^2=\sigma_2^2$ مثال	$F=\frac{S_1^2}{S_1^2}$ $v_1=n_1-1$ $v_1=n_1-1$ جوامع نرمال	$\sigma_1^2<\sigma_2^2$	$F<f_{\alpha}(v_1,v_2)$
		$\sigma_1^2>\sigma_0^2$	$F<f_{1-\alpha}(v_1,v_2)$
		$\sigma_1^2 \neq \sigma_2^2$	$F<f_{\alpha/2}(v_1,v_2)$ $F<f_{1-\alpha/2}(v_1,v_2)$

H_0	آماره آزمون	H_1	ناحيه بحراني
$p=p_0$ مثال	$Z=\frac{\hat{P}-p_0}{\sqrt{p_0q_0/n}}=\frac{X-np_0}{\sqrt{np_0q_0}}$ $\hat{P}=\frac{X}{n}\ ,q_0=1-p_0$ np>5 و n →∞	$p<p_0$	$Z<-z_{1-\alpha}$
		$p>p_0$	$Z>z_{1-\alpha}$
		$p \neq p_0$	$Z<-z_{1-\alpha/2}$ $Z>z_{1-\alpha/2}$
$p_1=p_2=p$ مثال	$Z=\frac{\hat{P}_1-\hat{P}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_2}+\frac{1}{n_2})}},\ \hat{P}_1=\frac{x_1}{n_1}$ $,\hat{P}_2=\frac{x_2}{n_2},\ \text{if}\ (p\ ?)\ \hat{p}=\frac{x_1+x_2}{n_1+n_2}$ $(n_1,n_2\geq30)\ n_1,n_2\rightarrow\infty$	$p_1<p_2$	$Z<-z_{1-\alpha}$
		$p_1>p_2$	$Z>z_{1-\alpha}$
		$p_1 \neq p_2$	$Z<-z_{1-\alpha/2}$ $Z>z_{1-\alpha/2}$

در آزمون نسبت فرض شد كه (n→∞) و آزمون مورد بررسي قرار گرفت حال اگر n كوچك باشد به صورت زير (طبق روش تابع توزيع گسسته) عمل مي‌كنيم

۱- تشكيل فرض صفر $H_0: p=p_0$

۲- تشكيل فرض مقابل $H_1: p<p_0\ \ OR\ \ p>p_0\ \ OR\ \ p\neq p_0$

۳- انتخاب α و تشكيل ناحيه بحراني به صورت زير

$$\text{if}\ (H_1: p<p_0)\ \Rightarrow P(X\leq x\mid p=p_0)<\alpha$$
$$\text{if}\ (H_1: p>p_0)\ \Rightarrow P(X\geq x\mid p=p_0)<\alpha$$
$$\text{if}\ (H_1: p\neq p_0)\ \Rightarrow \begin{cases} if\ (x<np_0)\Rightarrow P(X\leq x\mid p=p_0)<\alpha/2 \\ if\ (x>np_0)\Rightarrow P(X\geq x\mid p=p_0)<\alpha/2 \end{cases}$$

۴- محاسبات: به كمك نمونه منتخب با حجم n، مقدار x را تعيين و ناحيه بحراني را تعيين مي‌كنيم.

۵- نتيجه‌گيري: اگر x در ناحيه بحراني قرار گرفت، فرض صفر را رد مي‌كنيم.

مثال

آزمون‌هاي مربوط به تفاضل‌هاي بين k نسبت: براي تفاضل دو نسبت روابط آزمون مورد بررسي قرار گرفت حال اگر آزمون فرض، مربوط به تفاضل‌هاي k نسبت باشد تكليف چيست؟

براي اين منظور از روش كلي زير استفاده مي‌كنيم:

۱- تشكيل فرض صفر $H_0: p_1=p_2=\dots=p_k=p$

۲- تشكيل فرض مقابل حداقل يكي از نسبتها (p_iها) برابر با p نيست H_1:

۳- انتخاب α

۴- آماره مناسب براي اين آزمون به صورت زير محاسبه مي‌گردد:

آزمون‌هاي مربوط به تفاضل‌هاي بين k نسبت:

داده‌ها را به صورت جدول زير مرتب مي‌كنيم:

	پيروزي‌ها	شكست‌ها
نمونه اول	x_1	n_1-x_1
نمونه دوم	x_2	n_2-x_2
⋮	...	...
نمونه k	x_k	n_k-x_k

كه درايه‌هاي آن را فراواني‌هاي خانه‌اي مشاهده شده f_{ij} مي‌ناميم كه در آن اولين انديس نشانه سطر و دومين انديس نشانه ستون اين جدول 2*k است.

اميد فراواني‌هاي خانه‌اي مشاهده شده را با e_{ij} نشان داده و به صورت زير محاسبه مي‌كنيم.

$$e_{ij}=\begin{cases} n_i p & j=1 \\ n_i(1-p) & j=2 \end{cases}\qquad if\ (p\ ?)\Rightarrow \hat{p}=\frac{x_1+x_2+\dots+x_k}{n_1+n_2+\dots+n_k}$$

آزمون‌هاي مربوط به تفاضل‌هاي بين k نسبت:

آماره X^2 از رابطه زير به دست مي‌آيد:

$$X^2=\sum_{i=1}^{k}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

۵- ناحيه بحراني به صورت $X^2 \geq \chi^2_{(1-\alpha)}$ با v=k-1 درجه آزادي مشخص مي‌شود.

۶- نتيجه‌گيري: اگر مقدار آماره در ناحيه بحراني قرار گيرد فرض H_0 رد، در غير اينصورت فرض H_0 پذيرفته مي‌شود.

مثال

آزمون جدول‌هاي توافقي:

يك جدول توافقي $r*c$ داراي c ستون معرف رسته‌هاي مختلف A_1، A_2، ... ، A_c از يك متغير و r سطر معرف رسته‌هاي مختلف B_1، B_2، ... ، B_r از متغير ديگر است و f_{ij} فراواني خانه‌اي مشاهده شده براي خانه‌اي است كه به سطر i ام و ستون j ام تعلق دارد.

	A_1	A_2	…	A_c	
B_1	f_{11}	f_{12}	…	f_{1c}	f_{10}
B_2	f_{21}	f_{22}		f_{2c}	f_{20}
⋮	…	…	…	…	⋮
B_r	f_{r1}	f_{r2}	…	f_{rc}	f_{r0}
	f_{01}	f_{02}	…	f_{0c}	f

$$f_{i0}=\sum_{j=1}^{c} f_{ij}$$

$$f_{0j}=\sum_{i=1}^{r} f_{ij}$$

$$f=\sum_{i=1}^{r}\sum_{j=1}^{c} f_{ij}$$

آزمون جدول‌هاي توافقي:

فرض صفري كه مي‌خواهيم آزمون كنيم مستقل بودن دو متغير است يعني اگر θ_{ij} احتمال آن باشد كه فقره‌اي در خانه‌اي قرار خواهد گرفت كه به سطر i ام و ستون j ام تعلق دارد و θ_{i0} احتمال آن باشد كه فقره‌اي به سطر i ام و θ_{j0} احتمال آن باشد كه فقره‌اي به ستون j ام تعلق دارد. پس فرض صفر به صورت زير در مي‌آيد:

$$\theta_{ij}=(\theta_{i0})(\theta_{0j}) \quad i=1,2,\dots,r \;\;\&\;\; j=1,2,\dots,c$$

و فرض مقابل هم به اين شكل است كه بازاي حداقل يك زوج i و j مقدار احتمال θ_{ij} برابر حاصلضرب θ_{i0} و θ_{j0} نيست.

آزمون جدول‌هاي توافقي:

فراواني‌هاي خانه‌اي مورد انتظار به صورت زير محاسبه مي‌شود:

$$\hat{\theta}_{i0}=\frac{f_{i0}}{f} \quad , \quad \hat{\theta}_{0j}=\frac{f_{0j}}{f}$$

$$e_{ij}=(\hat{\theta}_{i0})(\hat{\theta}_{0j}).f=\frac{f_{i0}}{f}.\frac{f_{0j}}{f}.f=\frac{(f_{i0})(f_{0j})}{f}$$

آماره اين آزمون نيز به شكل زير محاسبه مي‌گردد:

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

ناحيه بحراني براي اين آزمون به صورت $X^2 \geq \chi^2_{1-\alpha}$ با $v=(r-1)(k-1)$ درجه آزادي مشخص مي‌شود

آزمون جدول‌هاي توافقي:

ناحيه بحراني براي اين آزمون به صورت $X^2 \geq \chi^2_{1-\alpha}$ با $v=(r-1)(c-1)$ درجه آزادي مشخص مي‌شود كه اگر آماره آزمون در اين محدوده قرار گرفت فرض صفر رد مي‌شود.

اين آزمون را وقتي به كار مي‌بريم كه هيچ يك از e_{ij} ها كمتر از ۵ نباشند اين امر مستلزم اين است كه برخي خانه‌ها را باهم ادغام كنيم كه در نتيجه از درجه آزادي متناظر كاسته مي‌شود.

مثال

آزمون نيكويي برازش:

اين آزمون مواقعي به كار مي‌رود كه مي‌خواهيم تعيين كنيم آيا مي‌توان مجموعه‌اي از داده‌ها را به عنوان نمونه‌هايي تصادفي از جامعه‌اي با توزيع مفروض تلقي كرد يا نه؟

براي اين منظور داده‌ها را طبق اصول آمار توصيفي طبقه‌بندي مي‌كنيم كه معمولاً اين طبقه‌بندي در داده‌هاي گسسته صورت نمي‌گيرد ولي اگر دامنه تغييرات داده‌ها زياد باشد در اين حالت نيز مي‌توان داده‌ها را به چند طبقه بر اساس اصول آمار توصيفي طبقه‌بندي كرد سپس جدول فراواني داده‌ها تكميل مي‌گردد.

آزمون نيكويي برازش:

با توجه به قضاياي گفته شده پارامترهاي مورد نياز توزيع مفروض از روي نمونه، تخمين مي‌گردد تا احتمالهاي مورد نظر محاسبه شوند.

فراواني‌هاي مورد انتظار $e_{ij}=p_i*f$	احتمال هر رده با توجه به توزيع مفروض p_i	فراواني مشاهده شده f_i	داده‌ها يا نماينده هر دسته در حالت دسته‌بندي	حدود طبقات در حالت دسته‌بندي
…	…	…	…	…
		$\sum f_i=f$		

آزمون نيكويي برازش:

فرض صفر در اين آزمون به اين صورت است كه مجموعه‌اي از داده‌هاي مشاهده شده، داراي توزيع مفروض مي‌باشند و فرض مقابل هم اين است كه داده‌ها داراي توزيع ديگري است.

آماره اين آزمون به شكل زير محاسبه مي‌گردد:

$$X^2=\sum_{j=1}^{m}\frac{(f_i-e_i)^2}{e_i}$$

ناحيه بحراني براي اين آزمون به صورت $X^2 \geq \chi_{1-\alpha}^2$ با $v=s+t-1$ درجه آزادي مي‌باشد كه با توجه به آماره محاسبه شده نتيجه لازم مشخص مي‌شود.

آزمون نيكويي برازش:

در اين آزمون درجه آزادي $v=s-t-1$ مشخص شد. كه s تعداد جملات و t تعداد پارامترهاي مستقلي است كه به جاي آنها برآوردهايشان گذاشته مي‌شود.

• رابطه $v=s-t-1$ يك رابطه كلي است هرگاه فراواني‌هاي خانه‌اي مورد انتظار در فرمول‌هاي خي‌دو بر مبناي داده‌هاي شمارشي نمونه‌اي برآورد شوند. براي مثال در جدول توافقي $s=rc$ و $t=(r+c-2)$ در نتيجه:

$$s-t-1=rc-(r+c-2)-1=(r-1)(c-1)$$

و در آزمون تفاضل k نسبت، $s=2k$ و $t=k$ در نتيجه:

$$s-t-1=2k-(k)-1=k-1$$

مثال

مثال: اگر ميانگين يك نمونه تصادفي با حجم $n=64$ از يك جامعه، ۷۸/۸ و انحراف معيار جامعه $\sigma=12.8$ باشد آيا در ميزان معني دار بودن ۰/۰۲ مي‌توان گفت كه ميانگين جامعه از ۷۵ بزرگتر است؟

1) $H_0:\mu=75$

2) $H_1:\mu>75$

3) $\alpha=0.02$

4) $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ & Rejection region: $Z>z_{1-\alpha}=z_{0.98}=2.054$

5) $Z=\dfrac{78.8-75}{12.8/\sqrt{64}}=2.375$

۶- چون Z در ناحيه بحراني قرار مي‌گيرد پس فرض H_0 رد مي‌شود.

مثال: فرض كنيد X داراي توزيع نرمال با ميانگين μ و واريانس ۳۶ است در آزموني به صورت

H_0: μ=50 , H_1: μ=55

مي‌باشد ناحيه بحراني و n را طوري تعيين كنيد كه به ازاي آن‌ها خطاي نوع اول ۰/۰۵ و خطاي دوم ۰/۱ باشد؟

حل: فرض كنيد c نقطه بحراني باشد:

$$\alpha=P(\bar{X}>c\mid\mu=50)\Rightarrow 0.05=P(Z>\frac{c-50}{6/\sqrt{n}})\Rightarrow$$

$$P(Z<\frac{c-50}{6/\sqrt{n}})=0.95\Rightarrow \boxed{1.645=\frac{c-50}{6/\sqrt{n}}}\ (1)$$

$$\beta=P(\bar{X}<c\mid\mu=55)\Rightarrow 0.1=P(Z<\frac{c-55}{6/\sqrt{n}})\Rightarrow \boxed{-1.288=\frac{c-55}{6/\sqrt{n}}}\ (2)$$

$$(1)\&(2)\Rightarrow c=52.8\ ,\ n=13$$

برگشت

مثال: يك نمونه ۵ تايي از جامعه‌اي نرمال انتخاب كرده‌ايم، كه ميانگين نمونه و انحراف معيار نمونه آن به ترتيب ۳ و ۰/۹ است. آيا مي‌توان گفت ميانگين جامعه از ۳/۵ كمتر است؟

1) $H_0:\mu=3.5$

2) $H_1:\mu<3.5$

3) $\alpha=0.5$

4) $T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}$ & Rejection region: $T<-t_{1-\alpha,n-1}=-t_{0.95,4}=-2.132$

5) $T=\dfrac{3-3.5}{0.9\sqrt{5}}=-1.238$

۶- چون T در ناحيه بحراني قرار نمي‌گيرد پس H_0 رد نمي‌شود.

برگشت

مثال: از دو كلاس ۴۰ و ۵۰ نفري امتحان مشابهي را گرفتيم، نتايج ميانگين و انحراف معيار نمرات اين دو گروه به صورت زير است:

$\mu_1=74$, $s_1=8$, $\mu_2=78$, $s_2=7$

آيا مي‌توان گفت اين دو كلاس با هم اختلاف دارند؟

1) $H_0:\mu_1-\mu_2=0$

2) $H_1:\mu_1-\mu_2\neq 0$

3) $\alpha=0.5$

4) $Z=\dfrac{(\bar{X}_1-\bar{X}_2)-0}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}$ & Rejection region: $\begin{cases} Z<-z_{1-\alpha/2}=-z_{0.975}=-1.96 \\ Z>z_{1-\alpha/2}=z_{0.975}=1.96 \end{cases}$

5) $Z=\dfrac{74-78}{\sqrt{\dfrac{64}{40}+\dfrac{49}{50}}}=-2.49$

۶- چون Z در ناحيه بحراني قرار مي‌گيرد پس H_0 رد مي‌شود.

برگشت

مثال: در مثال قبل اگر $n_1=16$ و $n_2=25$ با فرض برابري واريانس‌هاي دو جامعه آزمون را دو باره انجام دهيد:

1) $H_0: \mu_1 - \mu_2 = 0$

2) $H_1: \mu_1 - \mu_2 \neq 0$

3) $\alpha = 0.5$

4) $T = \dfrac{(\bar{X}_1 - \bar{X}_2) - 0}{S_p\sqrt{(1/n_1)+(1/n_2)}}$ & Rejection region: $\begin{cases} v = n_1 + n_2 - 2 = 23 \\ T < -t_{1-\alpha/2} = -t_{0.975} = -2.07 \\ T > t_{1-\alpha/2} = t_{0.975} = 2.07 \end{cases}$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(16-1)\cdot 64 + (8-1)\cdot 49}{23} = 56.65$$

5) $T = \dfrac{74-78}{7.53\times\sqrt{\dfrac{1}{16}+\dfrac{1}{8}}} = -1.23$

۶- چون T در ناحيه بحراني قرار نمي‌گيرد پس H_0 رد نمي‌شود.

برگشت

مثال: اگر انحراف معيار نمونه تصادفي با حجم $n=25$ از يك جامعه نرمال $s=0.28$ باشد آيا مي‌توان ادعا كرد كه انحراف معيار اين جامعه كمتر از ۰/۳ است؟

1) $H_0: \sigma = 0.3$

2) $H_1: \sigma < 0.3$

3) $\alpha = 0.5$

4) $X^2 = \dfrac{(n-1)S^2}{\sigma^2}$ & Rejection region: $X^2 < \chi^2_{\alpha,n-1} = 13.8$

5) $X^2 = \dfrac{(25-1)(0.28)^2}{(0.3)^2} = 20.907$

۶- چون X^2 در ناحيه بحراني نيست پس فرض صفر رد نمي‌شود.

برگشت

مثال: فرض كنيد معدل دانشجويان دانشگاه A و دانشگاه B داراي توزيع نرمال با واريانس‌هاي σ^2_A و σ^2_A باشند دو نمونه تصادفي به حجمهاي $n_1=10$ و $n_2=7$ از دو دانشگاه انتخاب مي‌كنيم واريانس‌هاي نمونه آن به ترتيب برابر با ۰/۰۷۶۵ و ۰/۲۰۲۸ مي‌باشد آيا مي‌توان ادعا كرد كه دو دانشگاه در ميزان معني دار بودن ۰/۰۲ داراي واريانس برابر هستند؟

1) $H_0: \sigma_A = \sigma_B$

2) $H_1: \sigma_A \neq \sigma_B$

3) $\alpha = 0.02$

4) $F^2 = \dfrac{S_A^2}{S_B^2}$ & Rejection region: $\begin{cases} F < f_{\alpha/2}(n_1-1, n_2-1) = 0.172 \\ F < f_{1-\alpha/2}(n_1-1, n_2-1) = 7.98 \end{cases}$

5) $F = \dfrac{0.0765}{0.2028} = 0.3772$

۶- چون f در ناحيه بحراني قرار نمي‌گيرد، لذا فرض صفر را رد نمي‌كنيم.

برگشت

مثال: تيراندازي مدعي است كه ۶۰% تيرهايي را كه شليك مي‌كند به هدف مي‌خورد اگر ۵۵ تير از ۱۰۰ تيري كه شليك كرده به هدف خورده باشد، در مورد اين ادعا چه اظهار نظري مي‌توان كرد؟

1) $H_0: p = 0.6$

2) $H_1: p \neq 0.6$

3) $\alpha = 0.05$

4) $Z = \dfrac{x - np_0}{\sqrt{np_0q_0}}$ & Rejection region: $\begin{cases} Z > z_{1-\alpha/2} = z_{0.975} = 1.96 \\ Z < -z_{1-\alpha/2} = -z_{0.975} = -1.96 \end{cases}$

5) $Z = \dfrac{55 - 100\times 0.6}{\sqrt{100\times 0.6\times 0.4}} = -1.02$

۶- چون Z در ناحيه بحراني قرار نمي‌گيرد پس فرض صفر رد نمي‌شود.

برگشت

مثال: ۵۶ نفر از ۲۰۰ نفر كه مصرف كننده نوشابه هستند نوع A را ترجيح مي‌دهند و ۲۹ نفر از ۱۵۰ نفر ديگر كه مصرف كننده نوشابه هستند نوع B را ترجيح مي‌دهند، آيا در سطح تشخيص ۰/۰۶ مي‌توان نتيجه گرفت كه نوع A از نوع B بهتر است؟

1) $H_0: p_A = p_A$

2) $H_1: p_A > p_A$

3) $\alpha = 0.06$

4) $Z = \dfrac{\hat{P}_A - \hat{P}_B}{\sqrt{\hat{p}\hat{q}(\dfrac{1}{n_2}+\dfrac{1}{n_2})}}$ & Rejection region: $Z > z_{1-\alpha/2} = z_{0.94} = 1.555$

5) $\hat{P}_1 = \dfrac{x_A}{n_A} = \dfrac{56}{200} = 0.28$, $\hat{P}_2 = \dfrac{x_B}{n_B} = \dfrac{29}{150} = 0.19$, $\hat{p} = \dfrac{x_A + x_B}{n_A + n_B} = \dfrac{56+29}{200+150} = 0.24$

$$Z = \frac{0.28 - 0.19}{\sqrt{0.27\times 0.76\times(\frac{1}{200}+\frac{1}{150})}} = 1.95$$

۶- پس فرض صفر رد مي‌شود يعني جواب مثبت است.

برگشت

مثال: بازيكني مدعي است كه ۶۰% از توپ‌هايي را كه پرتاپ مي‌كند وارد سبد مي‌شوند، از ۱۵ توپي كه پرتاپ كرده ۱۰ توپ وارد سبد شده است، آيا در سطح تشخيص ۵% مي‌توان اين ادعا را پذيرفت؟

1) $H_0: p = 0.6$

2) $H_1: p \neq 0.6$

3) $\alpha = 0.05$

Rejection region: $\begin{cases} x=10, np_0 = 15\times 0.6 = 9 \\ x > np_0[10>9] \Rightarrow P(X > x \mid p=0.6) < (\alpha/2) = 0.025 \end{cases}$

4) $P(X > x \mid p = 0.6) = \sum_{x=10}^{15} b(x;15,0.6) =$

$$1 - \sum_{x=0}^{9} b(x;15,0.6) = 1 - 0.5968 = 0.4032$$

۵- چون احتمال محاسبه شده كمتر از ۰/۰۲۵ نيست پس فرض صفر را رد نمي‌كنيم بعبارت ديگر دليلي براي شك كردن به ادعاي بازيكن نيست.

برگشت

مثال: بر مبناي دادههاي نمونهاي كه در جدول زير نشان داده، تعيين كنيد آيا نسبت واقعي مشترياني كه ماده پاك كننده A را به ماده پاك كننده B ترجيح ميدهند، در هر سه شهر يكسان است يا نه؟

	ماده A	ماده B	
شهر الف	232	168	n_1=400
شهر ب	260	240	n_1=500
شهر ج	197	203	n_1=400

$$\begin{cases} H_0: \theta_1 = \theta_2 = \theta_3 \\ H_1: \text{سه } \theta \text{ همه با هم برابر نيستند} \end{cases}$$

e_{ij}	1	2
۱	212	188
۲	265	235
۳	212	188

$$e_{ij} = \begin{cases} n_i p & j=1 \\ n_i(1-p) & j=2 \end{cases}, \hat{p} = \frac{x_1+x_2+x_3}{n_1+n_2+n_3}$$

$$\hat{p} = \frac{232+260+197}{400+500+400} = 0.53$$

ناحيه بحراني: $X^2 \geq \chi^2_{0.95,2} = 5.991$

$$X^2 = \sum_{i=1}^{3}\sum_{j=1}^{2}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(232-212)^2}{212} + \ldots \frac{(203-188)^2}{188} = 6.48$$

چون X^2 در ناحيه بحراني قرار دارد پس فرض صفر بايد رد شود، به عبارت ديگر، نسبتهاي واقعي مشترياني كه ماده پاك كننده A را بر ماده پاك كننده B در سه شهر ترجيح ميدهند، يكسان نيستند.

برگشت

مثال: جدول توافقي زير از مطالعه بستگي بين استعداد شخصي در رياضيات و علاقه او به آمار به دست آمده است، مستقل بودن استعداد رياضي شخص و علاقه او را در سطح معنيدار بودن ۰/۰۱ آزمون كنيد؟

		استعداد رياضي			
		ضعيف	متوسط	عالي	f_{i0}
علاقه به آمار	ضعيف	۶۳	۴۲	۱۵	۱۲۰
	متوسط	۵۸	۶۱	۳۱	۱۵۰
	عالي	۱۴	۴۷	۲۹	۹۰
	f_{0j}	۱۳۵	۱۵۰	۷۵	۳۶۰

e_{ij}	۱	۲	۳
۱	۴۵	۵۰	۲۵
۲	۵۶	۶۲	۳۱
۳	۳۴	۳۸	۱۹

$$f_{i0} = \sum_{j=1}^{3} f_{ij}$$

$$f_{0j} = \sum_{i=1}^{3} f_{ij}$$

$$f = \sum_{i=1}^{3}\sum_{j=1}^{3} f_{ij}$$

$$e_{ij} = \frac{(f_{i0})(f_{0j})}{f}$$

استعداد رياضي و علاقه به آمار مستقلند: H_0

ايـــن دو متغيـــر مستقـــل نيســـتند: H_1

ناحيه بحراني $X^2 \geq \chi^2_{0.99,4} = 13.277$ (v=(3-1)(3-1)=4)

$$X^2 = \sum_{i=1}^{3}\sum_{j=1}^{3}\frac{(f_{ij}-e_{ij})^2}{e_{ij}} = \frac{(63-45)^2}{45} + \ldots + \frac{(29-19)^2}{19} = 32.51$$

چون $X^2 \geq 13.277$ پس فرض صفر (مستقل بودن استعداد رياضي و علاقه به آمار) رد ميشود.

برگشت

مثال: فرض كنيد جدول زير تعداد خطاهاي يك تاپيپست در هرصفحه از يك متن ۴۴۰ صفحهاي است، آيا تعداد خطاها در تايپ متغير تصادفي با توزيع پواسن است؟

تعداد خطاها	فراواني مشاهده شده (f_i)	احتمالهاي پواسن با $\lambda=3$	فراوانيهاي مورد انتظار $e_i=p_i f$
۰	۱۸	۰/۰۴۹۸	۲۱,۹
۱	۵۳	۰/۱۴۹۲	۶۵/۷
۲	۱۰۳	۰/۲۲۴۰	۹۸/۶
۳	۱۰۷	۰/۲۲۴۰	۹۸/۶
۴	۸۲	۰/۱۶۸۰	۷۳/۹
۵	۴۶	۰/۱۰۰۸	۴۴/۰
۶	۱۸	۰/۰۵۰۴	۲۲/۲
۷	۱۰	۰/۰۲۱۶	۹/۵
۸	۲ } ۱۳	۰/۰۰۸۱	ادغام ۳/۶ } ۱۴/۳
۹	۱	۰/۰۰۲۷	۱/۲
	$\sum f_i=f$		

فرض صفر H_0: جامعه داراي توزيع پواسن است.

فرض مقابل H_1: جامعه داراي توزيع پواسن نيست.

$$\hat{\lambda} = \bar{X} = \frac{\sum_{i=1}^{k} x_i f_i}{N} = \frac{(0\times18)(1\times53)\ldots(8\times2)(9\times1)}{440} = \frac{1341}{440} = 3.05$$

$$X^2 = \sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i} = \frac{(18-21.9)^2}{21.9} + \ldots + \frac{(13-14.3)^2}{14.3} = 5.95$$

$$\begin{cases} \alpha=0.05 \\ \text{v=s-t-1=8-1-1=6} \qquad \Rightarrow 5.95 \text{ not} \geq 12.6 \\ \text{Rejection region:} X^2 \geq \Rightarrow X^2 \geq \chi^2_{0.95} = 12.6 \end{cases}$$

نتيجهگيري: پس نميتوان فرض صفر را كه دارا بودن توزيع پواسن جامعه است، رد نمود

برگشت

فصل دوازدهم: رگرسيون و همبستگي

فرض كنيد X و Y دو متغير تصادفي از دو جامعه آماري مختلف باشند هر مشاهده از اين دو متغير را به صورت زوج مرتب (x , y) نشان مي‌دهيم در اين فصل دو مسأله زير مورد بررسي قرار مي‌گيرد:

۱- آيا بين دو متغير X و Y همبستگي (رابطه‌اي) وجود دارد؟

۲- آيا اين همبستگي يا رابطه را مي‌توان به صورت معادله بيان نمود؟

براي مثال مي‌خواهيم بررسي كنيم كه آيا بين سن و وزن افراد رابطه‌اي وجود دارد و اگر رابطه‌اي وجود دارد معادله اي را كه اين رابطه را بيان مي‌كند پيدا كنيم. معمولا متغيري كه از روي مقادير آن پيشگويي انجام مي‌شود را با X و متغيري كه پيشگويي مي‌شود را با Y نشان مي‌دهيم. X را متغير مستقل و Y را متغير وابسته مي‌ناميم.

براي مثال اگر جدول زير دو نمونه از دو متغير تصادفي x و y باشند:

سن (X)	۵۶	۴۸	۴۲	۳۵	۵۸	۴۰	۳۹	۵۰
وزن (Y)	۴۵ ۶۴/۰ ۶۵/۳	۴۲ ۹۵/۲	۷۵/۰	۶۹/۹	۵۹/۰	۹۴/۹	۶۲/۰	۶۵/۵ ۸۷/۲

هدف بررسي اين موضوع است كه آيا بين سن و وزن افراد رابطه‌اي وجود دارد؟

اگر x را متغير مستقل و y را متغير وابسته در نظر بگيريم هدف تعيين

• متوسط مقدار y به ازاي مقدار مفروضي از x يا ميانگين شرطي $\mu_{y|x}$ مي‌باشد.

كه اگر f(x,y) تابع احتمال توأم دو متغير تصادفي x و y باشند

$$\mu_{y|x} = E(y\,|\,x) = \int_y y.k(y\,|\,x)dy$$

$$\mu_{y|x} = E(y\,|\,x) = \sum_y y.k(y\,|\,x)$$

معادله رگرسيون y به روي x ناميده ميشود.

مثال: با مفروض بودن متغيرهاي تصادفي x و y با چگالي توأم زير معادله رگرسيون y روي x را بيابيد؟

$$f(x,y) = \begin{cases} xe^{-x(1+y)} & ; x,y > 0 \\ 0 & ; \text{other point} \end{cases}$$

$$g(x) = \int_y xe^{-x(1+y)}dy = \begin{cases} e^{-x} & ; x > 0 \\ 0 & ; \text{other point} \end{cases}$$

$$k(y\,|\,x) = \frac{f(x,y)}{g(x)} = \begin{cases} xe^{-xy} & ; y > 0 \\ 0 & ; \text{other point} \end{cases}$$

$$\mu_{y|x} = E(y\,|\,x) = \int_y y.k(y\,|\,x)dy = \int_y y.xe^{-xy}dy = \begin{cases} \frac{1}{x} & ; x > 0 \\ 0 & ; \text{other point} \end{cases}$$

رگرسيون خطي: اگر منحني رگرسيون y نسبت به x خطي باشد معادله آن را به صورت

$$\boxed{\mu_{y|x} = \alpha + \beta x}$$

در نظر مي‌گيريم كه در آن α و β ضرايب رگرسيون ناميده مي‌شوند. به دلايل زير رگرسيون خطي مورد توجه مي‌باشد:

• اعمال رياضي بر اين معادلات به سادگي صورت مي‌گيرد.
• اغلب تقريب بسيار خوبي براي معادلات رگرسيون پيچيده‌تر هستند.
• در حالت توزيع نرمال دو متغيره معادلات رگرسيون خطي‌اند.

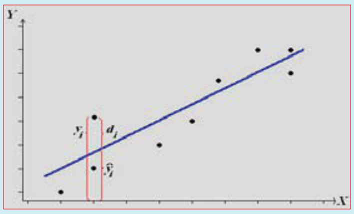
قضيه: اگر رگرسيون y روي x خطي باشد آنگاه

$$\mu_{y|x} = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x) \Rightarrow \begin{cases} \beta = \rho\frac{\sigma_y}{\sigma_x} = \frac{\sigma_{xy}}{\sigma_x^2} \\ \alpha = \mu_y - \rho\frac{\sigma_y}{\sigma_x}\mu_x \end{cases}$$

ولي در اغلب مواقع اطلاعي از توزيع توأم دو متغير تصادفي براي محاسبه رگرسيون خطي نداريم در اين حالت از روي نمونه تصادفي مقادير α و β را تخمين مي‌زنيم كه معادله برآورد (تخمين) شده را به صورت زير نشان مي‌دهيم:

$$\boxed{\hat{y} = \hat{\alpha} + \hat{\beta}x}$$

برآورد پارامترها به روش کمترین مربعات: فرض کنید n نمونه از دو متغیر تصادفي X و Y مشاهده شده است که هر زوج از دو متغیر تصادفي به صورت $(x_i\ ,\ y_i)$ مسأله ما پیدا کردن مناسب‌ترین خط به طوري که n نقطه تا جايي که امکان دارد به خط رگرسيون نزديك باشد



برآورد پارامترها به روش کمترین مربعات:

• مدل مورد نظر براي n مشاهده به صورت زير در نظر گرفته مي‌شود:

$$Y_i = \alpha + \beta X_i + E_i$$

که در آن E_i ها متغير تصادفي نرمال با ميانگين صفر و واريانس σ^2 مي‌باشد.

در اين روش براي تخمين ضرايب رگرسيون عبارت زير را که مجموع مانده‌ها ناميده مي‌شود، مينيمم مي‌گردد:

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

برآورد پارامترها به روش کمترین مربعات:

$$\boxed{\hat{y} = \hat{\alpha} + \hat{\beta} x}$$

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \Rightarrow$$

$$\begin{cases} \dfrac{\partial SSE}{\partial \hat{\alpha}} = -2\sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \\ \dfrac{\partial SSE}{\partial \hat{\beta}} = -2\sum_{i=1}^{n} x_i (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0 \end{cases}$$

$$\begin{cases} n\hat{\alpha} + \hat{\beta}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ \hat{\alpha}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \Longrightarrow$$

اين معادلات را معادلات نــــرمال مي ناميم که از حــل اين دستگاه ضرايب رگرسيون بدست مي‌آيد.

برآورد پارامترها به روش کمترین مربعات:

پس از حل دستگاه معادلات نرمال ضرايب رگرسيون از روابط زير به دست مي‌آيد:

$$\begin{cases} 1)\bar{x} = \dfrac{\sum x_i}{n} \quad ; \quad \bar{y} = \dfrac{\sum y_i}{n} \\ 2)S_{xx}\,(SS_x) = \sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2 \\ 3)S_{yy}\,(SS_y) = \sum (y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2 \\ 4)S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - n\bar{x}\bar{y} \\ 5)SSE = S_{yy} - \hat{\beta}^2 S_{xx} = S_{yy} - \dfrac{(S_{xy})^2}{S_{xx}} = S_{yy} - \hat{\beta} S_{xy} \end{cases}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad \& \quad \hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

مثال: مطلوبست خط رگرسیون و مجموع مانده‌ها براي داده‌هاي جدول زير؟

	x_i	y_i	x_iy_i	x_i^2	y_i^2
	4	31	124	16	961
	9	58	522	81	3364
	10	65	650	100	4225
	14	73	1022	196	5329
	4	37	148	16	1369
	7	44	308	49	1936
	12	60	720	144	3600
	22	91	2002	484	8281
	1	21	21	1	441
	17	84	1428	289	7056
$\sum$	100	564	6945	1376	36562

$$\bar{x} = \frac{\sum x_i}{n} = \frac{100}{10} = 10$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{564}{10} = 56.4$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{6945 - 10\times 10\times 56.4}{1376 - 10\times 10^2} = 3.471$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 56.4 - 3.471\times 10$$

$$= 21.69$$

$$SSE = S_{yy} - \hat{\beta} S_{xy} = 222.745$$

تمرين: مطلوبست معادله رگرسيوني خطي و مجموع مانده‌ها بين دو متغير Y و X بر اساس نمونه تصادفي به اندازه ۱۵ طبق نتايج حاصله زير از نمونه :

$$\sum x_i = 162 \quad \sum y_i = 1840.5 \quad \sum x_i y_i = 19945.7$$

$$\sum x_i^2 = 1820.2 \quad \sum y_i^2 = 1348$$

تمرين: براي تعيين رابطه هزينه حمل يک کالا که آن را با y نشان مي دهيم, با فاصله فروشگاه که آن را با x نشان مي دهيم, نمونه اي تصادفي شامل ۸ فروشگاه را انتخاب و براساس آن نتايج زير را بدست آورده ايم:

(X)	۸	۱۰	۲۱	۱۵	۶	۱۲	۲۱	۱۴
(Y)	۱۱۳	۶۸	۹۱	۱۳۹	۹۵	۴۹	۵۶	۱۶۹

معادله خط رگرسيون و مجموع مانده‌ها را حساب کنيد؟

تحليل رگرسيوني نرمال: در اين تحليل فرض مي‌شود كه به ازاي هر x_i ثابت، چگالي شرطي متغيرهاي تصادفي y_i، چگالي نرمال

$$w(y_i \mid x_i) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left[\frac{y_i-(\alpha+\beta x_i)}{\sigma}\right]^2} \;,\; y_i \in \mathbb{R}$$

است.

با استفاده از برآورد درست نمايي براي n مشاهده مي‌توان α، β و σ را تخمين زد كه با مقايسه نتايج آن مشاهده مي‌گردد كه همان نتايج روش كمترين مربعات براي تخمين پارامترها به دست مي‌آيد.

روابط زير در تحليل رگرسيوني نرمال صادقند:

$$1)\ \hat{\alpha} \text{ is normal } \&\ \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$E(\hat{\alpha}) = \alpha \ \&\ \operatorname{var}(\hat{\alpha}) = \sigma^2\left(\frac{\sum x_i^2}{nS_{xx}}\right)$$

$$2)\ \hat{\beta} \text{ is normal } \&\ \hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$E(\hat{\beta}) = \beta \ \&\ \operatorname{var}(\hat{\alpha}) = \frac{\sigma^2}{S_{xx}}$$

$$3)\ S^2 = \frac{SSE}{n-2} \ \&\ E(S^2) = \sigma^2$$

$$4)\ (n-2)\frac{S^2}{\sigma^2} \sim X^2_{n-2}$$

يك **فاصله اطمينان** $100(1-\alpha)\%$ براي ضرايب رگرسيون و خط رگرسيون به ازاي x مفروض به شكل زير محاسبه مي‌گردد:

$$T = \frac{\hat{\alpha}-\alpha}{S\sqrt{\frac{\sum x_i^2}{nS_{xx}}}} \sim t_{(n-2)} \quad ; \quad \alpha : \hat{\alpha} \pm t_{1-\alpha/2,(n-2)} . S\sqrt{\frac{\sum x_i^2}{nS_{xx}}}$$

$$T = \frac{\hat{\beta}-\beta}{S/\sqrt{S_{xx}}} \sim t_{(n-2)} \quad ; \quad \beta : \hat{\beta} \pm t_{1-\alpha/2,(n-2)} . S/\sqrt{S_{xx}}$$

$$T = \frac{\hat{y}_0 - y_0}{S\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}} ; \quad y_0 : \hat{y}_0 \pm t_{1-\alpha/2,(n-2)} . S\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}$$

آزمون فرض براي ضرايب رگرسيون:

H_0	آماره آزمون	H_1	ناحيه بحراني
$\alpha = \alpha_0$	$T = \frac{\hat{\alpha}-\alpha_0}{S\sqrt{\frac{\sum x_i^2}{nS_{xx}}}}$, $v = n-2$	$\alpha < \alpha_0$	$T < -t_{1-\alpha}$
		$\alpha > \alpha_0$	$T > t_{1-\alpha}$
		$\alpha \neq \alpha_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$
$\beta = \beta_0$	$T = \frac{\hat{\beta}-\beta_0}{S/\sqrt{S_{xx}}}$, $v = n-2$	$\beta < \beta_0$	$T < -t_{1-\alpha}$
		$\beta > \beta_0$	$T > t_{1-\alpha}$
		$\beta \neq \beta_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$

آزمون فرض براي خط رگرسيون به ازاي x_0 مفروض :

H_0	آماره آزمون	H_1	ناحيه بحراني
$y = y_0$	$T = \frac{\hat{y}-y_0}{S\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}}}$, $v = n-2$	$y < y_0$	$T < -t_{1-\alpha}$
		$y > y_0$	$T > t_{1-\alpha}$
		$y \neq y_0$	$T < -t_{1-\alpha/2}$ $T > t_{1-\alpha/2}$

مثال: جدول زير تعداد ساعات مطالعه در برابر نمره دريافتي براي درس زبان براي ۱۰ نفر مي باشد. با توجه به اين داده ها مطلوبست

- معادله خط رگرسيون y روي x
- يك فاصله اطمينان ۹۵% براي β
- آزمون فرض $\beta=3$ در برابر $\beta>3$ در سطح ۰/۰۱

(X)	۴	۹	۱۰	۱۴	۴	۷	۱۲	۲۲	۱	۱۷
(Y)	۳۱	۵۸	۶۵	۷۳	۳۷	۴۴	۶۰	۹۱	۲۱	۸۴

x_i	y_i	x_iy_i	x_i^2	y_i^2
4	31	124	16	961
9	58	522	81	3364
10	65	650	100	4225
14	73	1022	196	5329
4	37	148	16	1369
7	44	308	49	1936
12	60	720	144	3600
22	91	2002	484	8281
1	21	21	1	441
17	84	1428	289	7056
$\sum$ 100	564	6945	1376	36562

$$\bar{x}=\frac{\sum x_i}{n}=\frac{100}{10}=10$$

$$\bar{y}=\frac{\sum y_i}{n}=\frac{564}{10}=56.4$$

$$\hat{\beta}=\frac{S_{xy}}{S_{xx}}=\frac{\sum x_i y_i-n\bar{x}\bar{y}}{\sum x_i^2-n\bar{x}^2}=\frac{6945-10\times10\times56.4}{1376-10\times10^2}=3.471$$

$$\hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}=56.4-3.471\times10=21.69$$

$$\boxed{\hat{y}=\hat{\alpha}+\hat{\beta}x=21.7+3.47x}$$

$$\begin{cases} S_{xx}=\sum x_i^2-n\bar{x}^2=1376-10\times10^2=376 \\ S_{yy}=\sum y_i^2-n\bar{y}^2=36562-10\times(56.4)^2=4752.4 \\ S_{xy}=\sum x_i y_i-n\bar{x}\bar{y}=6945-10\times10\times56.4= \\ SSE=S_{yy}-\hat{\beta}S_{xy}=4752.4-3.47\times1305=224.05 \\ S^2=\frac{SSE}{n-2}=\frac{224.05}{8}=28 \end{cases}$$

$$\hat{\beta}-t_{1-\alpha/2,(n-2)}.\frac{S}{\sqrt{S_{xx}}}<\beta<\hat{\beta}+t_{1-\alpha/2,(n-2)}.\frac{S}{\sqrt{S_{xx}}}$$

$$1-\alpha=0.90\Rightarrow1-\alpha/2=0.975\Rightarrow t_{0.975,(8)}=2.31$$

$$3.47-2.31.\sqrt{\frac{28}{376}}<\beta<3.47+2.31.\sqrt{\frac{28}{376}}\Rightarrow2.84<\beta<4.1$$

1) $H_0:\beta=3$

2) $H_1:\beta>3$

3) $\alpha=0.01\Rightarrow1-\alpha=0.99$

4) $T=\dfrac{\hat{\beta}-\beta_0}{S/\sqrt{S_{xx}}}$ & Rejection region:$T>t_{0.99,8}=2.90$

5) $T=\dfrac{3.47-3}{\sqrt{\frac{28}{376}}}=1.722$

۶-چون T در ناحيه بحراني نيست پس فرض H_0 رد نمي‌شود.

تمرين: براي جدول مقادير زير معادله خط رگرسيون را بنويسيد.

X	۵۲	۷۵	۳۴	۴۷	۵۷	۲۸	۳۹	۲۱	۴۳	۶۴
Y	۷۵	۹۸	۵۶	۸۹	۹۲	۷۳	۶۵	۵۲	۷۸	۸۲

- يك فاصله اطمينان ۹۵% براي ضرايب رگرسيون به دست آوريد.
- يك فاصله اطمينان ۹۵% براي y_0 پيدا كنيد وقتي $x_0=47$
- آزمون فرض $\beta=0$ را در مقابل $\beta\neq0$ در سطح ۰/۰۵ انجام دهيد؟
- آزمون فرض $\alpha=40$ را در مقابل $\alpha<40$ در سطح ۰/۰۵ انجام دهيد؟
- آزمون فرض $y_0=75$ را در مقابل $y_0>75$ وقتي $x_0=47$ در سطح ۰/۰۵ انجام دهيد؟

$$\bar{x}=\frac{\sum x_i}{n}=\frac{460}{10}=46 \qquad \bar{y}=\frac{\sum y_i}{n}=\frac{760}{10}=76$$

$$S_{xx}=\sum x_i^2-n\bar{x}^2=2472 \quad S_{yy}=\sum y_i^2-n\bar{y}^2=2056$$

$$S_{xy}=\sum x_i y_i-n\bar{x}\bar{y}=1894$$

$$\hat{\beta}=\frac{S_{xy}}{S_{xx}}=\frac{1894}{2474}=0/76 \quad \hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}=40/8$$

$$\hat{y}=\hat{\alpha}+\hat{\beta}x=40/81+0/76x$$

$$SSE=S_{yy}-\hat{\beta}S_{xy}=2056-40/81\times0/76=627/0$$

$$S^2=\frac{SSE}{n-2}=\frac{627/01}{8}=78/37$$

$$T=\frac{0/76-0}{\sqrt{\frac{78/37}{2474}}}=4/47 \qquad t_{0/025,df=8}=2/306$$

$$\Rightarrow T=4/47>2/306=t_{0/025,df=8}\Rightarrow RH_0$$

همبستگي:

يادآوري: اگر X و Y دو متغير تصادفي با تابع احتمال توأم f(x,y) باشند. ضريب همبستگي خطي بين X و Y به صورت زير نشان داده و تعريف مي كنيم:

$$\rho=\rho(X,Y)=\frac{\sigma_{xy}}{\sigma_x\sigma_y}$$

- براي بررسي چگونگي همبستگي خطي بين دو متغير تصادفي X و Y استفاده مي‌كنيم.
- اگر X و Y مستقل باشند ضريب همبستگي بين آنها صفر است.
- ضريب همبستگي همواره در فاصله [-1 , 1] قرار دارد.
- ضريب همبستگي بستگي به واحد اندازه گيري ندارد.
- همبستگي مستقيم: در اين حالت اندازه هاي دو متغير با هم تغيير مي‌كنند يعني اگر يكي زياد شود ديگري هم زياد مي‌شود و بر عكس
- همبستگي معكوس: در اين حالت دو متغير در جهت مخالف هم مي‌باشند.

همبستگي:

حال اگر تابع احتمال توأم را نداشته باشيم و هدف تعيين چگونگي همبستگي بين دو متغير تصادفي باشد براي تعيين آن دو نمونه تصادفي از دو متغير تصادفي انتخاب و از رابطه زير استفاده مي‌كنيم.

$$\rho = \rho(X, Y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$r = \frac{S_{xy}}{\sqrt{S_{yy} S_{xx}}} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum y_i^2 - n\bar{y}^2)(\sum x_i^2 - n\bar{x}^2)}}$$

آزمون فرض براي معني دار بودن همبستگي با توجه به r :

ناحيه بحراني	H_1	آماره آزمون	H_0
$T<-t_{1-\alpha}$	$\rho < \rho_0$	$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, $v = n-2$	$\rho = \rho_0$
$T>t_{1-\alpha}$	$\rho > \rho_0$		
$T<-t_{1-\alpha/2}$ $T>t_{1-\alpha/2}$	$\rho \neq \rho_0$		
$Z<-t_{1-\alpha}$	$\rho < \rho_0$	$Z = \frac{\sqrt{n-3}}{2} \cdot \ln\frac{(1+r)(1-\rho)}{(1-r)(1+\rho)}$	$\rho = \rho_0$
$Z>t_{1-\alpha}$	$\rho > \rho_0$		
$Z<-t_{1-\alpha/2}$ $Z>t_{1-\alpha/2}$	$\rho \neq \rho_0$		

مثال: فرض كنيد يك نمونه ۱۰ تايي انتخاب و r=0.966 محاسبه شده است ميزان معني دار بودن همبستگي را در سطح ۰/۰۵ آزمون كنيد؟

1) $H_0: \rho = 0$

2) $H_1: \rho \neq 0$

3) $\alpha = 0.05 \Rightarrow 1-\alpha/2 = 0.975$

$$\begin{cases} 4)\ T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \ \ \& \ \text{Rejection region:} \begin{cases} T<-t_{0.975,8} = -2.31 \\ T>t_{0.975,8} = 2.31 \end{cases} \\ 5)\ T = \frac{0.966\sqrt{8}}{\sqrt{1-(0.966)^2}} = 10.565 \end{cases}$$

$$\begin{cases} 4)\ Z = \frac{\sqrt{n-3}}{2} \cdot \ln\frac{(1+r)(1-\rho)}{(1-r)(1+\rho)} \ \ \& \ \text{Rejection region:} \begin{cases} Z<-Z_{0.975} = -1.96 \\ Z>Z_{0.975} = 1.96 \end{cases} \\ 5)\ Z = \frac{\sqrt{10-3}}{2} . \ln\frac{(1+0.966)(1-0)}{(1-0.966)(1+0)} = 5.37 \end{cases}$$

۶- در هر دو آزمون فرض صفر رد مي‌شود، يعني وجود همبستگي معني‌دار است.

فصل سيزدهم: تحليل واريانس

مقدمه:

در مسائل تحليل واريانس بررسي مي‌كنيم كه:

آيا تفاوتهاي مشاهده شده بين بيش از دو ميانگين نمونه‌اي را مي‌توان معلول تصادف دانست و يا بين ميانگين‌هاي جامعه‌هاي مورد نمونه‌گيري تفاوت‌هاي واقعي وجود دارد؟

مثلاً ممكن است بخواهيم بر مبناي داده‌هاي نمونه‌اي تصميم بگيريم آيا واقعاً تفاوتي بين ميزان مؤثر بودن سه روش تدريس يك درس وجود دارد يا آيا واقعاً تفاوتي بين متوسط مصرف بنزين چند ماشين وجود دارد يا نه؟ يا آيا تفاوتي بين ميزان مؤثر بودن چند ماده شيميايي شوينده وجود دارد؟

تحليل واريانس يكطرفه:

در حالت كلي در چنين مسائلي، k نمونه تصادفي به اندازه n از k جامعه داريم كه مقدار j ام از جامعه i ام را با x_{ij} نشان مي‌دهند.

فرض مي‌كنيم كه x_{ij}ها مستقلند و داراي توزيع‌هاي نرمال با ميانگين‌هاي μ_i و واريانس مشترك σ^2 باشند. پس مي‌توان نوشت:

جامعه ۱	x_{11}	x_{12}	…	x_{1n}
جامعه ۲	x_{21}	x_{22}	…	x_{2n}
⋮ .	…	…	…	…
جامعه k	x_{k1}	x_{k2}	…	x_{kn}

$$x_{ij} = \mu_i + e_{ij} \quad ; e_{ij} : n(0, \sigma) \Rightarrow$$

$$\boxed{x_{ij} = \mu + \alpha_i + e_{ij}}$$

كه در اينجا به μ ميانگين كل اطلاق مي‌شود و α_i ها اثرهاي تيماري ناميده مي‌شوند كه $\sum_{i=1}^{k} \alpha_i = 0$

تحليل واريانس يكطرفه:

آزمون فرض ما عبارتست از فرض برابري ميانگين‌هاي جوامع $(\mu_1=\mu_2=\ldots=\mu_k)$ در برابر نابرابري آنها يا

$$\begin{cases} H_0: \alpha_i = 0 \\ H_1: \exists i \;;\; \alpha_i \neq 0 \end{cases} \quad ; i = 1, 2, \ldots, k$$

قضيه:

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{..})^2 = n\sum_{i=1}^{k}(\bar{x}_{i.}-\bar{x}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i.})^2$$

$$SST = SS(Tr) + SSE$$

$\bar{x}_{i.}$: ميانگين مشاهدات جامعه i ام	$\bar{x}_{..}$: ميانگين همه nk مشاهده
SS(Tr): مجموع مربعات تيمارها	SSE: مجموع مربعات خطا
SST: مجموع كل مربعات	

تحليل واريانس يكطرفه:

قضيه:

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}x_{ij}^2 - \frac{1}{kn}T_{..}^2 \quad \& \quad SS(Tr) = \frac{1}{n}\sum_{i=1}^{k}T_{i.}^2 - \frac{1}{kn}T_{..}^2$$

$T_{i.}$: مجموع مشاهدات براي تيمار i ام	$T_{..}$: مجموع همه nk مشاهده

$$\frac{1}{\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i.})^2 : \chi^2_{k(n-1)} \Rightarrow \frac{SSE}{k(n-1)} \approx \sigma^2$$

$$\frac{n}{\sigma^2}\sum_{i=1}^{k}(\bar{x}_{i.}-\bar{x}_{..})^2 : \chi^2_{k-1} \Rightarrow \frac{SS(Tr)}{k-1} \approx \sigma^2$$

$$F = \frac{\dfrac{SS(Tr)}{k-1}}{\dfrac{SSE}{k(n-1)}} = \frac{k(n-1)SS(Tr)}{(k-1)SSE} : F(k-1, k(n-1))$$

تحليل واريانس يكطرفه:

نتيجه آزمون:

منبع تغيير	درجه آزادي	مجموع مربعات	ميانگين مربعات	F
تيمارها	K-1	SS(Tr)	$MS(Tr) = \frac{SS(Tr)}{k-1}$	$\frac{MS(Tr)}{MSE}$
خطا	K(n-1)	SSE	$MSE = \frac{SSE}{k(n-1)}$	
جمع	Kn-1	SST		

هرگاه مقدار حاصل F از $F_{1-\alpha}(k-1,k(n-1))$ بيشتر شود فرض صفر در سطح α رد مي‌شود.

مثال: فرض كنيد كه بخواهيم عمل پاك كنندگي سه ماده پاك‌كننده را بر مبناي درجه سفيدي ۱۵ قواره پارچه سفيد كه ابتدا به مركب آلوده شده و سپس در يك ماشين لباس‌شويي با اين سه ماده پاك كننده شسته شده‌اند مقايسه كنيم:

جامعه (تيمار) ۱	۸۰	۷۶	۷۱	۸۱	۷۷
جامعه (تيمار) ۲	۷۰	۶۶	۷۴	۵۸	۷۲
جامعه (تيمار) k	۷۷	۸۰	۸۲	۸۵	۷۶

در سطح معني‌دار بودن ۰/۰۱ آزمون برابري ميانگين‌ها را انجام دهيد؟

$$\begin{cases} 1) H_0: \alpha_i = 0 \\ 2) H_1: \exists i \;;\; \alpha_i \neq 0 \end{cases} \quad ; i = 1, 2, \ldots, k$$

$3)\alpha = 0.01 \Rightarrow 1-\alpha = 0.99$

$4) F = \frac{MS(Tr)}{MSE}$ & Rejection region:

$F > F_{1-\alpha}(k-1, k(n-1)); F > F_{0.99}(2,12) = 6.93$

۵) محاسبه آماره

$T_{1.}=385$	$T_{2.}=340$	$T_{3.}=400$	$T_{..}=1125$	$\sum\sum x_i^2=85041$

$SST=85041-1/15(1125)^2=666$

$SS(Tr)=1/5(385^2+340^2+400^2)-1/15(1125)^2=390$

$SSE=SSt-SS(Tr)=666-390=276$

منبع تغيير	درجه آزادي	مجموع مربعات	ميانگين مربعات	F
تيمارها	2	390	390/2=195	195/23=8.48
خطا	12	276	276/12=23	
جمع	14	66		

۶) چون مقدار F در ناحيه بحراني قرار مي‌گيرد پس فرض صفر رد مي‌گردد.

تحليل واريانس دو طرفه:

داده‌هاي زير مربوط به زمان لازم (برحسب دقيقه) براي شخصي است كه با اتومبيل خود از شنبه تا چهارشنبه با استفاده از چهار مسير مختلف به سر كار خود مي‌رسد:

مسير ۱	۳۱	۲۵	۲۵	۲۶	۲۲
مسير ۲	۲۹	۲۶	۲۸	۲۷	۲۵
مسير ۳	۳۳	۳۰	۳۳	۲۹	۲۹
مسير ۴	۳۰	۳۰	۲۷	۲۸	۲۶

ميانگين‌هاي اين چهار نمونه عبارتند از ۲۵/۸، ۲۷/۰، ۳۰/۲ و ۲۸/۲.چون اختلاف بين ميانگين‌ها نسبتا بزرگ است اين استنتاج معقول خواهد بود كه اختلاف بين متوسط زمان لازم براي رسيدن شخص به محل كار خود از چهار مسير مختلف وجود دارد ولي اين امر از تحليل واريانس يك‌طرفه نتيجه نمي‌شود زيرا با استفاده از محاسبات لازم داريم

$$F = 2.80 \;\; \& \; F_{0.95,3,16} = 3.24$$

تحليل واريانس دو طرفه:

پس اين اختلاف‌هاي نسبتاً بزرگ بين ميانگين‌ها و مقادير بين نمونه‌ها را در هر دسته معلول كدام علت مي‌توان دانست؟
اختلاف و تغيير در داخل نمونه‌ها مي‌تواند معلول شرايط رانندگي در روزهاي مختلف هفته دانست كه اگر چنين باشد، اين تغييرات، مجموع مربعات خطا را در تحليل واريانس يك‌طرفه تحت تاثير خود قرار مي‌دهد كه اين اثر موجب تغيير آماره F (بزرگتر شدن مخرج آن و در نهايت كوچك شدن خود F) مي‌شود كه اين تغيير خود باعث بي‌معني شدن نتيجه آزمون مي‌گردد.

تكليف چيست؟

يك راه حل اين است كه عامل غير مربوط (شرايط رانندگي در روزهاي مختلف) را ثابت بگيريم براي مثال نتايج را براي يك روز در نظر بگيريم. ولي با اين كار به ندرت به اطلاعاتي كه لازم داريم، مي‌رسيم.

تحليل واريانس دو طرفه:

امكان ديگر اين است كه عامل غير مربوط را در روي بردي به وسعت لازم تغيير دهيم به طوري كه به توان تغيير ناشي از آن را اندازه گرفته و بنابراين از مجموع مربعات خطا حذف نمود. يعني يك تحليل واريانس دو طرفه انجام داد كه در آن، تغييرات كل داده‌ها به سه جزء تيمارها (در مثال ما چهار مسير) عامل غير مربوط (در مثال ما شرايط رانندگي در روزهاي مختلف) و خطاي آزمايش افراز مي‌شود.
روش پيشنهادي بلوك‌بندي نام دارد و به اجزاي عامل غير مربوط (در مثال ما شرايط رانندگي در روزهاي مختلف) بلوك اطلاق مي‌شود.
پس بلوك‌ها سطوحي هستند كه در آن سطحها عامل غير مربوط ثابت در نظر گرفته مي‌شود و اگر هر تيمار در هر بلوك به تعداد دفعات مساوي ظاهر شود گوييم آزمايش يك طرح بلوكي كامل است.

تحليل واريانس دو طرفه:

فرض مي‌كنيم كه x_{ij} ها مستقلند و داراي توزيع‌هاي نرمال با ميانگين‌هاي μ_{ij} و واريانس مشترك σ^2 باشند. پس مي‌توان نوشت

	بلوك ۱	بلوك ۲	...	بلوك n
تيمار ۱	x_{11}	x_{12}	...	x_{1n}
تيمار 2	x_{21}	x_{22}	...	x_{2n}
$\vdots$.	...	...	...	...
تيمار k	X_{k1}	X_{k2}	...	X_{kn}

$$x_{ij}=\mu+\alpha_i+\beta_j+e_{ij}$$
$$;e_{ij}:n(0,\sigma)$$

كه در اينجا به μ ميانگين كل اطلاق مي‌شود و α_i ها اثرهاي تيماري و β_i ها اثرهاي بلوكي ناميده مي‌شوند كه

$$\sum_{i=1}^{k}\alpha_i=0 \quad \& \quad \sum_{j=1}^{n}\beta_j=0$$

تحليل واريانس دو طرفه:

آزمون فرضي كه انجام خواهيم داد عبارت است از دو فرض صفر "اثرهاي تيماري و بلوكي صفرند" در برابر دو فرض مقابل "اثرهاي تيماري و بلوكي همه برابر صفرنيستند"

$$\begin{cases} H_0:\alpha_i=0 & ;i=1,2,\dots,k \\ H_0':\beta_j=0 & ;j=1,2,\dots,n \end{cases}$$

$$\begin{cases} H_1:\exists i\ ;\ \alpha_i\neq 0 \\ H_1':\exists j\ ;\ \beta_j\neq 0 \end{cases}$$

تحليل واريانس دو طرفه:

قضيه:

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{\cdot\cdot})^2=n\sum_{i=1}^{k}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2+kn\sum_{j=1}^{k}(\bar{x}_{\cdot j}-\bar{x}_{\cdot\cdot})^2$$
$$+\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\bar{x}_{i\cdot}-\bar{x}_{\cdot j}+\bar{x}_{\cdot\cdot})^2$$

$$SST=SS(Tr)+SSB+SSE$$

$\bar{x}_{i\cdot}$ ميانگين مشاهدات تيمار i ام	$\bar{x}_{\cdot\cdot}$ ميانگين همه nk مشاهده
$\bar{x}_{\cdot j}$ ميانگين مشاهدات بلوك j ام	SST: مجموع كل مربعات
SS(Tr): مجموع مربعات تيمارها	SSE: مجموع مربعات خطا
SSB: مجموع مربعات تيمارها	

تحليل واريانس دو طرفه:

قضيه:

$$SSB=\frac{1}{k}\sum_{j=1}^{n}T_{\cdot j}^2-\frac{1}{kn}T_{\cdot\cdot}^2$$

$T_{..}$: مجموع كل تمام مشاهدات	$T_{..}$: مجموع مقادير حاصل براي بلوك j ام

تحليل واريانس دو طرفه: نتيجه آزمون

منبع تغيير	درجه آزادي	مجموع مربعات	ميانگين مربعات	F
تيمارها	k-1	SS(Tr)	$MS(Tr)=\frac{SS(Tr)}{k-1}$	$F_{Tr}=\frac{MS(Tr)}{MSE}$
بلوك‌ها	n-1	SSB	$MSB=\frac{SSB}{n-1}$	$F_B=\frac{MSB}{MSE}$
خطا	(K-1)(n-1)	SSE	$MSE=\frac{SSE}{(k-1)(n-1)}$	
جمع	Kn-1	SST		

فرض هاي صفر رد مي‌شود اگر:

$$F_{Tr}\ge F_{1-\alpha}(k-1,(k-1)(n-1))\ \ \&\ \ F_B\ge F_{1-\alpha}(n-1,(k-1)(n-1))$$

مثال: براي جدول داده‌هاي زير در سطح معني‌دار بودن ۰/۰۵ آزمون كنيد كه آيا اختلاف‌هاي بين ميانگين‌هاي حاصل براي مسيرهاي مختلف (تيمارها) معني‌دار هستند يا نه، و نيز آيا اختلاف‌هاي بين ميانگين‌هاي حاصل براي روزهاي مختلف هفته (بلوك‌ها) معني دارند يا نه؟

مسير ۱	۲۲	۲۶	۲۵	۲۵	۳۱
مسير ۲	۲۵	۲۷	۲۸	۲۶	۲۹
مسير ۳	۲۹	۲۹	۳۳	۳۰	۳۳
مسير ۴	۲۶	۲۸	۲۷	۳۰	۳۰

$$1)\begin{cases}H_0:\alpha_i=0 & ;i=1,2,\dots,k\\ H_0':\beta_j=0 & ;j=1,2,\dots,n\end{cases}\qquad 2)\begin{cases}H_1:\exists i\ ;\ \alpha_i\ne 0\\ H_1':\exists j\ ;\ \beta_j\ne 0\end{cases}$$

$$3)\ \alpha=0.05\Rightarrow 1-\alpha=0.95$$

$$4)\begin{cases}F_{Tr}=\frac{MS(Tr)}{MSE}\\ F_B=\frac{MSB}{MSE}\end{cases}\Rightarrow \text{Rejection region:}\begin{cases}\begin{cases}F_{Tr}>F_{1-\alpha}(k-1,(k-1)(n-1))\\ ;F>F_{0.95}(3,12)=3.49\end{cases}\\ \begin{cases}F_B>F_{1-\alpha}(n-1,(k-1)(n-1))\\ ;F>F_{0.95}(4,12)=3.26\end{cases}\end{cases}$$

۵- محاسبه آماره

$T_{.1}=99$	$T_{.2}=110$	$T_{.3}=113$	$T_{.4}=111$	$T_{.5}=123$
$T_{1.}=129$	$T_{2.}=135$	$T_{3.}=151$	$T_{4.}=141$	$T_{..}=556$
$\sum x^2=15610$				
	$SST=15610-1/20(556)^2=153.20$			
	$SS(Tr)=1/5(129^2+\dots+141^2)-1/20(556)^2=52.8$			
	$SSB=1/4(99^2+\dots+123^2)-1/20(556)^2=73.2$			
	$SSE=153.2-52.8-73.2=27.2$			

منبع تغيير	درجه آزادي	مجموع مربعات	ميانگين مربعات	F
تيمارها	3	52.8	52.8/3=17.6	17.6/2.27=7.75
بلوك‌ها	4	73.2	73.2/4=18.3	18.3/2.27=8.06
خطا	3*4=12	27.2	27.2/12=2.27	
جمع	20-1=19	153.2		

۶) نتيجه آزمون:

چون $F_{Tr}=7.75$ از $F_{0.95,3,12}=3.49$ بيشتر است و همچنين $F_B=8.06$ از $F_{0.95,4,12}=3.26$ بيشتر است نتيجه مي‌گيريم كه هر دو فرض صفر بايد رد گردد به عبارت ديگر، اختلاف‌هاي بين ميانگين هاي حاصل براي چهار مسير مختلف و همچنين اختلاف‌هاي بين ميانگين‌هاي حاصل براي روزهاي مختلف هفته معني‌دار است.

پايان