به نام یگانه هستی بخش

راهنمای مطالعه جزوه

بخش اول: براساس تقسیم‌بندی در ۹ فصل، خلاصه درس به همراه مثال‌هایی آورده شده‌اند که مطالعه آن‌ها باعث یادگیری بهتر مطالب می‌گردد.

بخش دوم: جداول توزیع موضوعات هر فصل در سؤالات ۵ سال اخیر تنظیم شده‌اند و در انتها توصیه‌های کلیدی در مورد هر فصل و مطالب آن برای مطالعه بهتر در اختیار داوطلبین قرار گرفته است.

بخش سوم: تعداد ۲۰ سؤال به عنوان نمونه سؤالات شبیه‌سازی شده امتحان تهیه شده است که به بررسی ۲۰ مطلب مهم درس می‌پردازد. داوطلبین عزیز پس از مطالعه دقیق جزوه به حل این سؤال‌ها پرداخته و پس از بررسی حل تشریحی نقاط قوت و ضعف خود را پیدا می‌کنند. برای داوطلبین رشته ریاضی ۲۵ سوال در نظر گرفته شده است.

امیدوارم این گام کوچک بتواند در موفقیت شما مؤثر باشد، سعادت و بهروزی شما را از درگاه ایزد منان خواستارم.

با تشکر ــ گائینی
دی ماه ۸۹

در جدول ذیل دروس به سرفصلهای مهم آن طبقه بندی شده و مشخص شده است که در هر سال از هر مبحث چند تست سوال شده است و دانشجوی محترم می تواند زمان باقیمانده تا کنکور را با توجه به اهمیت مباحث مدیریت نماید.

رشته: مهندسی صنایع درس: آمار و احتمال مهندسی

ردیف	مبحث	۱۳۸۵ تعداد تست	۱۳۸۶ تعداد تست	۱۳۸۷ تعداد تست	۱۳۸۸ تعداد تست	۱۳۸۹ تعداد تست	مجموع ۵ سال	نسبت از کل
1	آنالیز ترکیبی و ترکیب ها	0	0	0	0	0	0	0%
2	مدل توپ و جعبه	0	0	0	1	1	2	2%
3	خواص تابع احتمال	1	0	1	1	0	3	3%
4	محاسبه احتمال یکنواخت	2	0	0	0	0	2	2%
5	احتمال شرطی و قانون بیز	3	2	0	1	0	6	5%
6	استقلال پیشامد ها	1	2	2	0	2	7	6%
7	تابع احتمال و تابع چگالی	0	0	0	0	0	0	0%
8	تابع توزیع تجمعی و خواص	2	2	0	0	0	4	3%
9	امید ریاضی و واریانس	2	0	0	2	1	5	4%
10	متغیر توام گسسته و پیوسته	2	0	0	2	0	4	3%
11	توابع احتمال شرطی و امید شرطی	0	0	2	1	0	3	3%
12	کواریانس و ضریب همبستگی	0	1	0	0	0	1	1%
13	تابع مولد گشتاور	0	0	0	0	0	0	0%
14	نامساوی مارکوف و چبی شف	0	0	0	1	0	1	1%
15	توزیع یکتواخت گسسته	0	0	0	0	0	0	0%
16	توزیع دو جمله ای و برنولی و چند جمله ای	0	1	2	2	4	9	8%
17	توریع پواسن	0	1	1	0	1	3	3%
18	توزیع هندسی و دوجمله ای منفی	1	0	0	0	0	1	1%
19	توزیع فوق هندسی	0	0	0	0	1	1	1%
20	توزیع یکتواخت پیوسته	0	1	0	0	0	1	1%
21	توریع نرمال	0	2	0	0	1	3	3%
22	توزیع گاما و مربع کای	0	1	0	0	2	3	3%
23	توزیع نمایی	0	1	1	1	1	4	3%
24	توزیع کوشی و بتا	1	0	0	0	0	1	1%
25	روابط بین توزیع ها	0	2	1	0	1	4	3%
26	قضیه حد مرکزی و واریانس میانگین	0	0	1	0	0	1	1%
27	توزیع T و ساختار آن	0	0	0	0	0	0	0%
28	توزیع مربع کای و F و ساختار آن	0	0	0	0	0	0	0%
29	روش گشتاورها MME	0	1	1	0	0	2	2%
30	روش MLE	2	1	1	2	0	6	5%
31	نا اریبی و سازگاری و MSE	1	1	0	1	0	3	3%
32	برآورد فاصله ای μ و σ و ρ	1	1	1	1	1	5	4%
33	تعریف α و β و توان	2	0	1	1	1	5	4%
34	لم نیمن پیرسون و نسبت درستنمایی	1	1	0	0	0	2	2%
35	انواع آزمون μ و σ و ρ	0	0	1	0	0	1	1%
36	آزمون انطباق و استقلال	1	0	0	0	0	1	1%
37	روش P-value	1	1	0	1	1	4	3%
38	برآورد n	0	0	1	0	0	1	1%
39	رگرسیون و همبستگی	3	7	1	1	2	14	12%
40	آنالیز واریانس	2	1	1	1	0	5	4%
41	آمار توصیفی	1	0	1	0	0	2	2%
	جمع	**30**	**30**	**20**	**20**	**20**	**120**	**100%**

بخش اول

خلاصه درس آمار و احتمال

فصل اول: آنالیز ترکیبی

ترتیب

تعداد گروه‌های r تایی متمایز و با رعایت ترتیب را که می‌توان از میان n شیء متمایز انتخاب کرد با P_n^r نشان می‌دهند. برای به دست آوردن P_n^r با استفاده از اصل ضرب داریم:

$$P_n^r = n(n-1)\dots(n-r+1) = \frac{n!}{(n-r)!} \qquad \begin{matrix} r \le n \\ r,n \in N \cup \{0\} \end{matrix}$$

به طور کلی هر وقت بتوان لااقل دو گروه متمایز تشکیل داد که اعضای دو گروه یکسان باشند، می‌گوییم ترتیب مهم است یا رعایت شده است.

n شیء متمایز دارای $n!$ جایگشت هستند.

مثال : به چند طریق می‌توان **4** قطعه متمایز را در دستگاهی که **10** مکان در آن برای قطعات وجود دارد قرار دارد.

حل: باید 4 مکان از 10 مکان را با ترتیب انتخاب کنیم که می‌شود

$$P_{10}^4 = \frac{10!}{6!}$$

جایگشت تکراری

هرگاه n شیء که به k نوع متمایز تقسیم‌بندی شده‌اند موجود باشد و نوع i ام شامل n_i عضو نامتمایز باشد به طوری که $n_1 + n_2 + \dots + n_k = n$، تعداد جایگشت‌های این n شیء برابر است با:

$$\frac{n!}{n_1! \dots n_k!}$$

در واقع $n!$ را بر فاکتوریل تکراری‌ها تقسیم می‌کنیم.

مثال : فرض کنید ده توپ سفید و نامتمایز و **5** توپ قرمز متمایز را در کنار هم قرار می‌دهیم در چند حالت توپ‌های اول و آخر قرمز هستند.

حل: توپ اول به 5 و آخر به 4 طریق انتخاب می‌شود 13 توپ باقی می‌ماند که ده عدد نامتمایزند پس جواب می‌شود.

$$5(4)\frac{13!}{10!} = 20(13)(12)(11)$$

جایگشت دوری

تعداد جایگشت‌های دوری n شیء متمایز برابر $(n-1)!$ است.

مثال : به چند طریق 6 زن و 6 مرد می‌توانند دور یک میز بنشینند طوری که هیچ دو نفری از یک جنسیت کنار هم نباشند.

حل: ابتدا زن‌ها به $5!$ طریق می‌نشینند سپس مردها به $6!$ بین آن‌ها می‌نشینند پس جواب می‌شود $5!6!$

ترکیب

تعداد گروه‌های r تایی متمایز بدون رعایت ترتیب را که می‌توان از میان n شیء متمایز انتخاب کرد با C_n^r یا $\binom{n}{r}$ نشان می‌دهند و به صورت زیر حساب می‌کنند:

$$C_n^r = \frac{P_n^r}{r!}$$

یا

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \qquad r \le n \text{ و } r,n \in N \cup \{0\}$$

لازم به ذکر است در ترکیب، دو گروه تنها در صورتی متمایزند که لااقل یکی از اعضای دو گروه متفاوت باشند.

مثال : به چند طریق می‌توان 20 دانشجو را در 5 اتاق 4 نفره نامتمایز اسکان داد؟

حل: توجه شود که:

$$\binom{20}{4}\binom{16}{4}\binom{12}{4}\binom{8}{4}\binom{4}{4} = \binom{20}{4,4,4,4,4}$$

تعداد طرق تقسیم 20 نفر است به 5 گروه 4 نفره که گروه‌ها متمایز باشند حال که اتاق‌ها نامتمایزند جواب می‌شود.

$$\frac{\binom{20}{4,4,4,4,4}}{5!} = \frac{20!}{(4!)^5\,5!}$$

نکته : موضوعات زیر به طور یکسان محاسبه می‌شوند.

$$\binom{r+n-1}{n-1} = \begin{cases} \text{تعداد مشتق جزئی مرتبه } r \text{ از تابع } n \text{ متغیره } f(x_1,\dots,x_n) \\ \text{تعداد جواب معادله } x_1+\dots+x_n=r \text{ با متغیرهای صحیح نامنفی} \\ \text{تعداد جملات بسط } (x_1+\dots+x_n)^r \\ \text{تعداد طرق توزیع } r \text{ توپ نامتمایز در } n \text{ جعبه} \end{cases}$$

مثال : الف) از تابع $h(x_1,\dots,x_{10})$ چند مشتق جزئی مرتبه 8 می‌توان گرفت که نسبت به مشتق اول 4 بار مشتق بگیریم.

ب) تعداد جواب‌های معادله $x_1+\dots+x_{10}=8$ وقتی $x_1=4$ چند است.

ج) بسط $(x_1+\dots+x_n)^8$ چند جمله دارد که توان x_1 برابر 4 است.

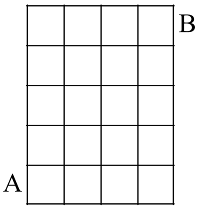
د) به چند طریق می‌توان 8 توپ نامتمایز را در ده جعبه توزیع کرد که جعبه اول 4 توپ داشته باشد.

حل: جواب همه طبق نکته بالا می‌شود $\binom{12}{8}=\binom{4+9-1}{9-1}$ چون یکی از متغیرها برابر 4 است یکی از ده کم می‌شود و 4 واحد از 8

نکته : اگر شبکه‌ای m سطر و n ستون داشته باشد تعداد طرق انتقال بدون برگشت از یک سر قطر تا سر دیگر می‌شود.
$$\frac{(m+n)!}{m!n!}$$

مثال : به چند طریق می‌توان از نقطه **A** به نقطه **B** رفت. (هربار یک قدم به راست یا بالا برداشته می‌شود.)

حل: طبق نکته بالا:



$$\frac{(5+4)!}{5!4!}$$

تعدادی از اتحادهای مهم آنالیز ترکیبی

$\binom{n}{r}=\binom{n}{n-r}$	$\sum_{r=0}^{n}\binom{n}{r}^2=\binom{2n}{n}$	$\binom{m+n}{r}=\sum_{k=0}^{r}\binom{m}{k}\binom{n}{r-k}$
$\binom{n}{r}=\binom{n-r}{r}+\binom{n-r}{r-1}$	$r\binom{n}{r}=n\binom{n-1}{r-1}$	$\sum_{i=1}^{n}i\binom{n}{i}=n2^{n-1}$
$\binom{n}{r}=\binom{n-r}{r}+\binom{n-2}{r-2}+2\binom{n-2}{r-1}$	$\sum_{i=0}^{n}\binom{n}{i}=2^n$	$\sum_{r=0}^{n}(-1)^r\binom{n}{r}=0$

مثال : حاصل $\binom{8}{6}+\binom{8}{4}+\binom{8}{2}$ چقدر است؟

حل: بنا به آخرین اتحاد جدول فوق:

$$\binom{8}{0}+\binom{8}{2}+\binom{8}{4}+\binom{8}{6}+\binom{8}{8}=\binom{8}{1}+\binom{8}{3}+\binom{8}{5}+\binom{8}{7}=2^7$$

پس:

$$\binom{8}{2}+\binom{8}{4}+\binom{8}{6}=2^7-2$$

مثال : **عدد** $7200=2^5.3^2.5^2$ **چند مقسوم‌علیه دارد؟**

حل: کافی است به توان یک واحد اضافه کرده و حاصل را در هم حساب کنیم.

$6(3\times3)=54=$ تعداد مقسوم‌علیه

فصل دوم: اصول احتمال و احتمال شرطی

اصول احتمال

به پیشامدهای تک‌عضوی **پیشامد ساده** گفته می‌شود.

اگر A و B دو پیشامد دلخواه باشند آن‌گاه $P(B-A)=P(B)-P(A\cap B)$.

اگر A و B دو پیشامد دلخواه باشند:

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

تفاضل متقارن دو پیشامد

اگر A و B دو پیشامد از فضای نمونه‌ای S باشند $A\Delta B$ تفاضل متقارن A و B نامیده می‌شود و زمانی رخ می‌دهد که فقط یکی از A و B رخ دهند.

به عبارت دیگر:

$$A\Delta B=(A\cup B)-(A\cap B)=(A-B)\cup(B-A)$$

پس

$$P(A\Delta B)=P(A)+P(B)-2P(A\cap B)$$

مثال : اگر **A** و **B** با احتمال **0.3** با هم رخ دهند و با احتمال **0.4** فقط یکی از آن‌ها رخ دهد چقدر احتمال دارد هیچ‌یک از آن‌ها رخ ندهند؟

حل:

$$P(A'\cap B')=1-P(A\cup B)$$

اما

$$P(A\cup B)=P(A\Delta B)+P(A\cap B)=0.4+3=0.7 \quad \Rightarrow \quad P(A'\cap B')=1-0.7=0.3$$

مثال : فرض کنید $P(A\cap B)=0.1 \quad , \quad P(A\cup B)=0.4$ حاصل $P(A\Delta B\,|\,A\cup B)$ چقدر است؟

حل:

$$P(A\Delta B\,|\,A\cup B)=\frac{P(A\Delta B)}{P(A\cup B)}=\frac{P(A\cup B)-P(A\cap B)}{P(A\cup B)}=\frac{0.4-0.1}{0.4}=0.75$$

کران‌های احتمال

بر اساس نامساوی‌های زیر می‌توان حداقل و حداکثر احتمال را برای اجتماع و اشتراک پیشامدها یافت:

$$\max_i \{P(A_i)\} \le P\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} P(A_i)$$

$$\underbrace{\sum_{i=1}^{n} P(A_i) - n + 1 \le P\left(\bigcap_{i=1}^{n} A_i\right)}_{\text{نامساوی سمت چپ به نامساوی بن فرونی شهرت دارد.}} \le \min_i \{P(A_i)\}$$

در حالت خاص:

$$\max\{P(A),P(B)\} \le P(A \cup B) \le P(A)+P(B)$$

$$P(A)+P(B)-1 \le P(A \cap B) \le \min\{P(A),P(B)\}$$

مثال : اگر $P(B)=0.8$, $P(A)=0.4$ **کران پایینی** $P(B|A)$ **چقدر است؟**

$$P(B|A)=\frac{P(A \cap B)}{P(A)} \ge \frac{P(A)+P(B)-1}{P(A)} = \frac{0.2}{0.4} = \frac{1}{2}$$

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

مثال : از میان اعداد $40 , \dots , 2 , 1$ تعداد **3** عدد را بدون جاگذاری برمی‌داریم چقدر احتمال دارد اعداد انتخاب شده متوالی باشند؟

حل: کل حالات عبارتست از $\binom{n}{3}$ و حالات مطلوب عبارتست از:

$$(1,2,3)(2,3,4)(3,4,5)\dots(n-2,n-1,n)$$

که تعداد آن‌ها می‌شود $n-2$ حالت پس جواب می‌شود:

$$\frac{n-2}{\binom{n}{3}}$$

مثال : در مثال بالا اگر انتخاب با جاگذاری باشد جواب چه می‌شود؟

حل: مخرج می‌شود $n.n.n=n^3$ و صورت همان $(n-2)$ حالت است که چون ترتیب در مخرج مهم است در صورت هم جواب در $3!$ ضرب می‌شود پس جواب می‌شود:

$$\frac{3!(n-2)}{n^3}$$

(مسئله روز تولد) فرض کنید N نفر به طور تصادفی یک جا جمع شده‌اند. چقدر احتمال دارد روز تولد همهٔ آن‌ها تفاوت داشته باشد؟

$$\frac{(365)(364)\dots(365-N+1)}{(365)(365)\dots(365)} = \frac{P_{365}^{N}}{(365)^N} = H(N)$$

حالت کلی مسئله روز تولد احتمال آن است که در توزیع r توپ متمایز درون n جعبه، در هر جعبه حداکثر یکی برود یعنی تکرار مجاز نباشد، مثال زیر یک نوع بیان دیگر از این مسئله است.

اتوبوسی با r مسافر در n ایستگاه توقف دارد چقدر احتمال دارد هیچ دو نفری در یک ایستگاه پیاده نشوند؟ (به عبارت دیگر در هر ایستگاه حداکثر یکی پیاده شود)

جواب عبارت است از $\frac{P_n^r}{n^r}$ که $r \leq n$

(مسئله جورها) فرض کنید n نفر کلاه خود را در یک مهمانی به شخصی می‌دهند که نگهداری کند. او کلاه‌ها را به هم ریخته و به هر نفر به طور تصادفی کلاهی می‌دهد. اگر کسی کلاه خود را دریافت کند می‌گویند یک **جور** رخ داده است. حساب کنید احتمال آنکه:

(الف) لااقل یک جور رخ دهد.

$$P\left(\bigcup_{i=1}^{n} A_i\right)=1-\frac{1}{2!}+\frac{1}{3!}-\ldots+(-1)^{k+1}\frac{1}{k!}+\ldots+(-1)^{n+1}\frac{1}{n!}=\sum_{i=1}^{n}(-1)^{i+1}\frac{1}{i!}\to 1-e^{-1}$$

(ب) هیچ جوری رخ ندهد.

$$\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^n}{n!}=\sum_{i=0}^{n}\frac{(-1)^i}{i!}\to e^{-1}$$

$$P\ (\text{جور } k)=\frac{\sum_{i=0}^{n-k}\frac{(-1)^i}{i!}}{k!}\longrightarrow\frac{e^{-1}}{k!}$$

نکته : الف) تعداد جورها نمی‌تواند $n-1$ باشد چون در این صورت فقط باید یک نفر کلاه خود را نگیرد. (ب) تعداد حالاتی که در جایگشت n شیء متمایز هیچ‌کدام در جای خود نیستند عبارت است از:

$$n!\sum_{i=0}^{n}\frac{(-1)^i}{i!}$$

(مسئله کفش‌ها)

$$P\ (\text{جفت } n \text{ از لنگه } m \text{ انتخاب در میان جفت } k)=\frac{\binom{n}{k}\binom{n-k}{m-2k}2^{m-2k}}{\binom{2n}{m}}$$

مثال : اگر **8** توپ یکسان به تصادف روی خانه‌های شطرنج قرار داده شوند احتمال آن‌که روی یک خط راست باشند چقدر است؟

حل: کل حالات $\binom{64}{8}$ است و حالات مطلوب 8 سطر و 8 ستون و 2 قطر است که می‌شود 18 تا پس جواب عبارت است از $\frac{18}{\binom{64}{8}}$

مثال : سه قفل بر یک در زده شده و کلید هر قفل در بین **10** کلیدی است که همراه ماست اگر دو کلید از این ده کلید گم شود چقدر احتمال دارد در را بتوانیم باز کنیم اگر دو کلید از این ده کلید گم شود چقدر احتمال دارد در را بتوانیم باز کنیم؟

حل: باید این دو کلید گم شده از 7 کلیدی باشد که در را باز نمی‌کند یعنی $\dfrac{\binom{7}{2}}{\binom{10}{2}}$

احتمال شرطی و فرمول بیز

(قانون ضرب احتمال) $P\left(A_1 \cap A_2 \cap A_3\right)=P\left(A_1\right) P\left(A_2 | A_1\right) P\left(A_3 | A_1 \cap A_2\right) \ldots P\left(A_n \,\middle|\, \bigcap_{i=1}^{n-1} A_i\right)$

(قاعدهٔ پواسون) جعبه‌ای a توپ سفید و b توپ قرمز دارد. اگر از رنگ توپ‌های استخراج‌شده قبلی اطلاع نداشته باشیم احتمال سفید بودن توپ در هر مرحله‌ای $\frac{a}{a+b}$ است که همان احتمال سفید بودن اولی است. این نتیجه برای تعداد مراحل بیشتر هم درست است مثلاً احتمال سفید بودن توپ در هر دو مرحله برابر احتمال سفید بودن اولی و دومی است.

فرمول بیز $P\left(A_k | E\right)=\dfrac{P\left(A_k\right) P\left(E A_k\right)}{\sum_{i=1}^{n} P\left(A_i\right) P\left(E A_i\right)}$

اساساً در مسایل فرمول بیز بایستی فضای نمونه‌ای توسط چند پیشامد افراز شود و سؤال مربوط به احتمال یکی از این پیشامدهای افراز به شرط رخ دادن پیشامدی است که با عناصر افراز اشتراک دارد.

مثال : فرض کنید احتمال آنکه یک بیمار به سرطان مبتلا باشد **0.6** است. این فرد را با یک آزمایش سرطان که **2%** افراد سالم را مبتلا و **1%** افراد مبتلا را سالم تشخیص می‌دهد مورد آزمایش قرار می‌دهیم. اگر این آزمایش فرد مورد نظر را مبتلا تشخیص داده باشد، چقدر احتمال دارد او واقعاً مبتلا باشد؟

حل: قرار دهید:

پیشامد تشخیص مبتلا بودن به سرطان: E

پیشامد واقعاً سالم بودن: A

پیشامد واقعاً مبتلا بودن: A'

پس بنا به فرمول بیز:

$$P\left(A' | E\right)=\frac{P\left(A'\right) P\left(E | A'\right)}{P(A) P(E | A)+P\left(A'\right) P\left(E | A'\right)}=\frac{(0.6)(0.99)}{(0.4)(0.02)+(0.6)(0.99)}=0.98$$

مثال : فرض کنید شخصی جواب درست یک تست **k** گزینه‌ای را با احتمال **p** می‌داند. اگر نداند به شانس متوسل می‌شود. اگر او به سؤالی جواب درست داده باشد چقدر احتمال دارد او واقعاً جواب را می‌دانسته است؟

حل: قرار دهید:

پیشامد جواب درست دادن: E

پیشامد جواب درست را دانستن: A

پیشامد جواب درست را ندانستن: A'

بنا به فرمول بیز:

$$P(A|E)=\frac{P(A)P(E|A)}{P(A)\,P(E|A)+P(A')\,P(E|A')}=\frac{p.1}{p.1+(1-p)\,\frac{1}{k}}=\frac{kp}{kp+1-p}$$

مثال : از میان **10** مرد و **5** زن به ترتیب بدون جاگذاری **5** نفر را انتخاب می‌کنیم چقدر احتمال دارد نفرات دوم و چهارم زن باشند؟

حل: بنا به قاعده پواسون احتمال آن را می‌یابیم که اولی و دومی زن باشند که بنا به قانون ضرب می‌شود:

$$\frac{5}{15}\;\frac{4}{14}$$

مثال : فرض کنید حسن و حسین از جعبه‌ای که **4** سفید و **2** قرمز دارد به ترتیب و متوالیاً بدون جایگذاری توپ خارج کنند چقدر احتمال دارد حسن زودتر از حسین خارج کند.

حل: حسن یا بار اول موفق می‌شود یا بار دوم

$$\frac{4}{6}+\frac{2}{6}\frac{1}{5}\frac{4}{4}=\frac{4}{6}+\frac{2}{30}=\frac{20+2}{30}=\frac{22}{30}=\frac{11}{15}$$

استقلال پیشامدها

نکته‌ها:

۱ ـ اگر A , B مستقل باشند آن‌گاه A' , B نیز مستقل‌اند.

۲ ـ اگر A و B مستقل باشند آن‌گاه A' و B' نیز مستقل‌اند.

۳ ـ اگر A , B غیرتهی باشند

A , B مستقل نیستند $\Rightarrow$ A , B ناسازگار

A , B ناسازگار نیستند $\Rightarrow$ A , B مستقل

۴ ـ برای تحقیق مستقل بودن n پیشامد باید به تعداد 2^n-n-1 تساوی مورد بررسی قرار گیرند.

۵ ـ اگر A و B و C مستقل باشند آن‌گاه:

الف) A و $B\cup C$ ب) A و $B\Delta C$

مستقل‌اند.

۶ ـ فرض کنید یک آزمایش را به طور متوالی و مستقل از هم تکرار می‌کنیم. اگر A و B دو پیشامد ممکن برای این آزمایش باشند **احتمال وقوع** A **قبل از** B را پیدا کنید.

$$P\left(\text{وقوع } A \text{ قبل از } B\right)=\frac{P(A)}{P(A\cup B)}$$

حالت خاص که A و B ناسازگار باشند:

$$P\left(\text{وقوع } A \text{ قبل از } B\right)=\frac{P(A)}{P(A)+P(B)}$$

مثال : حسن و حسین هر کدام مستقلاً یک جفت سکه را پرتاب می‌کنند چقدر احتمال دارد حسن قبل از حسین جفت شیر بیاورد؟

حل: بنا به نکته 6 اگر A موفقیت حسن و B موفقیت حسین باشد

$$P\,(B \text{ قبل از } A)=\frac{P(A)}{P(A\cup B)}=\frac{P(A)}{1-P(A'\cap B')}$$

$$=\frac{P(A)}{1-P(A')P(B')}=\frac{\frac{1}{4}}{1-\frac{3}{4}\,\frac{3}{4}}=\frac{4}{7}$$

مثال : (استقلال مشروط)

فرض کنید با احتمال $\frac{1}{4}$از کارخانه A و با احتمال $\frac{3}{4}$ از B خرید می‌کنیم. تولیدات A با احتمال $\frac{2}{100}$ و تولیدات B با احتمال $\frac{3}{100}$ خراب هستند. دو تولید از یک کارخانه خریده‌ایم اگر اولی خراب باشد چقدر احتمال دارد و دومی خراب باشد؟

حل:

$$P(F_2|F_1)=\frac{P(F_1\cap F_2)}{P(F_1)}$$

حال صورت و مخرج را به کارخانه‌ها مشروط می‌کنیم.

$$=\frac{P(A)P(F_1|A)P(F_2|A)+P(B)P(F_1|A)P(F_2|B)}{P(A)P(F_1|A)+P(B)P(F_1|B)}=\frac{\frac{1}{4}\left(\frac{2}{100}\right)^2+\frac{3}{4}\left(\frac{3}{100}\right)^2}{\frac{1}{4}\left(\frac{2}{100}\right)+\frac{3}{4}\left(\frac{3}{100}\right)}$$

مثال : اگر **C,B,A** مستقل باشند و $P(C)=P(A)=P(B)=0.4$ ، $P\left[(A\Delta B)\cap C\right]$ چقدر است؟

حل: می‌دانیم وقتی C,B,A مستقلند $A\Delta B$ با C مستقل از هم هستند پس

پس

$$P((A\Delta B)\Delta C)=P(A\Delta B)P(C)=(0.48)(0.4)=0.192$$

چون

$$P(A\Delta B)=P(A)+P(B)-2P(A\cap B)=0.8-2(0.16)=0.8-0.32=0.48$$

فصل سوم: متغیرهای تصادفی و امید ریاضی

تابع توزیع تجمعی

برای هر متغیر تصادفی X (اعم از گسسته یا پیوسته) تابع توزیع تجمعی X عبارت است از:

$$F_X : R \longrightarrow [0,1]$$

$$F_X(x) = P(X \le x)$$

نمودار تابع توزیع متغیرهای گسسته همواره به صورت **پله‌ای** است و نمودار تابع توزیع متغیرهای پیوسته همواره پیوسته است.

$$P(a < X \le b) = F_x(b) - F_x(a) \quad , \quad P(a \le X \le b) = F_X(b) - F_X(a^-)$$

$$P(a \le X < b) = F_X(b^-) - F_X(a^-) \quad , \quad P(a < X < b) = F_X(b^-) - F_X(a)$$

مثال : اگر $F(x) = \begin{cases} 0 & x < -2 \\ \dfrac{7-x^2}{10} & -2 \le x < 0 \\ 0.8 & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$ تابع توزیع **X** باشد $P(-2 \le X < 1)$ چقدر است؟

حل: بنابه نکته بالا:

$$P(-2 \le X < 1) = F(1^-) - F(-2^-) = 0.8 - 0 = 0.8$$

امید ریاضی

امید ریاضی یا میانگین متغیر تصادفی X را با $E(X)$ یا μ نشان می‌دهیم و به صورت زیر تعریف می‌کنیم:

$$E(X) = \mu = \begin{cases} \sum_x x P(X = x) & \text{اگر X گسسته باشد} \\ \int_{-\infty}^{+\infty} x f(x) dx & \text{اگر X پیوسته باشد} \end{cases}$$

دقت شود $E(X)$ در صورت همگرای مطلق بودن سری یا انتگرال بالا تعریف می‌شود و وجود دارد. در این کتاب $E(X)$ متناهی فرض می‌شود.

اگر X آمیخته بود (یعنی بعضی جاها گسسته و بعضی جاها پیوسته بود) باید امید ریاضی بخش پیوسته و بخش گسسته را جداگانه حساب کرده، باهم جمع کنیم.

مثال : اگر $x \ge 0 \quad f(x) = xe^{-x}$ تابع چگالی **X** باشد امید ریاضی **X** را بیابید؟

حل:

$$E(X) = \int_0^\infty x f(x) dx = \int_0^\infty x^2 e^{-x} dx = -x^2 e^{-x} - 2xe^{-x} - 2e^{-x} \Big]_0^\infty = 2$$

x^2	e^{-x}
$2x$	$-e^{-x}$
2	$+e^{-x}$
0	$-e^{-x}$

برای انتگرال‌گیری جزء به جزء از روش جدولی روبه‌رو استفاده شده است.

مثال : اگر **X** دارای تابع احتمال $x=1,2,\cdots$ $\quad f(x)=\frac{-q^x}{xLnp}$ باشد $E(X)$ چقدر است؟

$$E(X)=\sum_{x=1}^{\infty}x\left(\frac{-q^x}{xLnp}\right)=-\frac{1}{Lnp}\sum_{x=1}^{\infty}q^x=\frac{-1}{Lnp}\frac{q}{1-q}=\frac{-q}{pLnp}$$

مثال : اگر تابع توزیع **X** به صورت زیر باشد امید ریاضی **X** چقدر است؟

$$F_X(x)=\begin{cases}0 & x<1\\ \frac{x^2+6}{10} & 1\le x<2\\ 1 & x\ge 2\end{cases}$$

حل: X آمیخته است یعنی در یک نقطه گسسته و در یک فاصله چگالی پیوسته دارد.

بخش گسسته : $P(X=1)=F(1)-F(\bar{1})=\frac{7}{10}-0=\frac{7}{10}\rightarrow E(X)=1\left(\frac{7}{10}\right)=\frac{7}{10}$

بخش پیوسته $f(x)=F'(x)=\frac{2x}{10}=\frac{x}{5} \qquad 1<x<2$

$$E(X)=\int_1^2\frac{x^2}{5}dx=\frac{x^3}{15}\Big]_1^2=\frac{8-1}{15}=\frac{7}{15}$$

$$E(X)=\frac{7}{10}+\frac{7}{15}=\frac{21+14}{30}=\frac{35}{30}=\frac{7}{5}$$

مد و میانه

نکته : $g(a)=E(X-a)^2$ به ازای $a=\mu$ مینیمم می‌شود یعنی:

$$\forall a;\quad E(X-a)^2\ge E(X-\mu)^2=Var(X)$$

$g(a)=E|X-a|$ به ازای $a=m$ مینیمم می‌شود یعنی:

$$\forall a;\quad E|X-a|\ge E|X-m|$$

که **m میانه توزیع X** است.

طبق تعریف عدد m میانه توزیع پیوسته X است اگر:

$$F_X(m)=\frac{1}{2}$$

البته در حالت کلی عدد m را میانه توزیع X گویند اگر:

$$P(X\ge m)\ge\frac{1}{2}\quad,\quad P(X\le m)\ge\frac{1}{2}$$

در متغیرهای پیوسته a مد است اگر تابع چگالی را ماکزیمم کند یعنی:

$$f(x)\le f(a)$$

و در متغیرهای گسسته a مد است اگر احتمال a از بقیه بیشتر باشد.

مثال : مد توزیع با تابع چگالی $f(x)=\frac{1}{\pi(1+x^2)}$ کدام است؟

حل: با توجه به آنکه $f(x)<f(0)=\frac{1}{\pi}$ پس توزیع در $x=0$ مد دارد.

مثال : میانه توزیع با تابع چگالی $x\in R$ $\quad f(x)=\frac{1}{\pi(1+x^2)}$ کدام است؟

حل: باید $F(m)=\frac{1}{2}$ اما $F(m)=\int_{-\infty}^{m}\frac{1}{\pi(1+x^2)}dx=\frac{\tan^{-1}x}{\pi}\Big]_{-\infty}^{m}=\frac{\tan^{-1}m+\frac{\pi}{2}}{\pi}=\tan^{-1}m+\frac{1}{2}=\frac{1}{2}$ پس $\tan^{-1}m=0$

یعنی $m=0$ میانه توزیع است.

نامساوی مارکوف ـ نامساوی چه‌بی‌شف

به $P(g(X)\ge c)\le\frac{E(g(X))}{c}$ نامساوی مارکوف گفته می‌شود و به $P(|X-\mu|\ge k\sigma)\le\frac{1}{k^2}$ یا $P(|X-\mu|<k\sigma)\ge 1-\frac{1}{k^2}$ که k عدد مثبت دلخواه است نامساوی چه‌بی‌شف گفته می‌شود. کافی است k را بیابید تا سمت راست نامساوی را پیدا کنید.
درمورد نامساوی‌ها دقت کنید احتمال با این نامساوی ها پیدا نمی‌شود بلکه کران احتمال را می‌توان به‌دست آورد. پس موقعی باید از این نامساوی‌ها کمک گرفت که درصورت سؤال حداقل یا حداکثر احتمال را بخواهند.

مثال : اگر $E(X)=8$ و $Var(X)=9$ حداقل $P(2<X<14)$ کدام است؟

حل:

$$P(2<X<14)=P(2-8<X-8<14-8)=P(-6<X-8<6)=P(|X-8|<6)$$

$6=k\sigma=k(3)$ پس $k=2$ و طبق نامساوی چه‌بی‌شف $P(|X-8|<6)\ge 1-\frac{1}{4}=\frac{3}{4}$

مثال : **X** دارای میانگین **1** و واریانس **4** است کران بالایی برای $P(|X|\ge 3)$ چقدر است؟

حل:

$$P(|X|\ge 3)=P(X^2\ge 9)\le\frac{E(X^2)}{9}=\frac{5}{9}$$

زیرا

$$E(X^2)=\mu^2+\sigma^2=1+4=5$$

نکته : وقتی حدود تغییرات دو متغیر به هم وابسته باشد آن دو متغیر وابسته‌اند، اگر حدود تغییرات Y,X به هم وابسته نباشند (ناحیه مستطیلی باشد) و تابع چگالی توأم به صورت دو تابع مجزا از Y,X تجزیه شود، آن‌گاه این دو متغیر مستقل هستند.

مثال : **X** و **Y** با تابع چگالی توأم $f(x,y)=kx^2y$ که $0<x<1$ و $0<y<1-x$ وابسته‌اند.

حل:

اما در تابع چگالی $0<y<2$ و $0<x<1$ $\quad f(x,y)=kx^2y$ $\quad$ X و Y مستقلند چون $f(x,y)=(kx^2)y=g_1(x)g_2(y)$ و حدود x و y به هم ربطی ندارد.

مثال : اگر امید ریاضی و واریانس X به ترتیب **2** و **5** باشد و $E(Y|X)=1$, $Var(Y|X)=2$ واریانس **XY** چقدر است؟

حل:

$$Var(XY)=E\left(Var(XY|X)+Var\left(E(XY|X)\right)\right)$$
$$=E\left(X^2Var(Y|X)+Var(XE(Y|X))\right)$$
$$=E\left(2X^2\right)+Var(X)=2(4+5)+5=23$$

محاسبه احتمال توام ـ تابع توزیع توأم و خواص آن

تابع توزیع توأم X و Y عبارتست از:

$$F_{x,y}(x,y)=P(X\le x, Y\le y)$$

که دارای خواص زیر است:

(a) $F_{X,Y}(+\infty,y)=F_Y(y)$

(b) $F_{X,Y}(x,+\infty)=F_X(x)$

(c) X,Y مستقل‌اند $\Leftrightarrow F_{X,Y}(x,y)=F_X(x)F_Y(y)$

(d) $f_{X,Y}(x,y)=\dfrac{\partial^2 F_{X,Y}(x,y)}{\partial x\,\partial y}$

(e) $F_X(x)+F_Y(y)-1\le F_{X,Y}(x,y)\le\sqrt{F_X(x)F_Y(y)}$

توابع توزیع کناری X و Y از روابط a و b پیدا می‌شوند و رابطه c تعریف استقلال X و Y را نشان می‌دهد رابطه d چگونگی یافتن تابع چگالی توأم از روی تابع توزیع توأم است و رابطه e کران‌هایی برای تابع توزیع نشان می‌دهد.

مثال : اگر $P(X\le 1.2)=0.4$ و $P(Y>-2)=0.1$، آنگاه حداقل و حداکثر $F_{x,y}(1.2,-2)$ چقدر است؟

حل: طبق رابطه e چون $F_y(-2)=0.9$ پس: $0.3\le F_{x,y}(1.2,-2)\le\sqrt{0.36}=0.6$

امید ریاضی شرطی ـ امید ریاضی مضاعف ـ واریانس شرطی

فرمول‌های f و g امید ریاضی شرطی را پیدا می‌کنند. معمولاً در صورت مسئله واضح است که امید ریاضی یک متغیر را به شرط آنکه متغیر دیگر مقدار معلومی بگیرد می‌خواهند.

(f) $E[X|Y=y]=\int_{-\infty}^{+\infty}xf(x|y)dx=\int_{-\infty}^{+\infty}x\dfrac{f(x,y)}{f(y)}dx$

(g) $E[Y|X=x]=\sum_y yP(Y=y|X=x)=\sum_y y\dfrac{P(X=x\,,\,Y=y)}{P(X=x)}$

(h) $\begin{cases}E(E(Y|X))=E(Y)\\E(E(X|Y))=E(X)\end{cases}$

به (h) امید ریاضی مضاعف و به (k) فرمول واریانس شرطی گویند.

(k) $Var(X)=E[Var(X|Y)]+Var[E(X|Y)]$

مثال : اگر تابع احتمال توأم **X** و **Y** به‌صورت زیر باشد. امید ریاضی **X** را وقتی $Y=0$ پیدا کنید.

Y \ X	-1	2	3	
0	0.1	0.2	0	
1	0.3	0.2	0.2	

حل: ابتدا احتمال Y در نقطه 0 را می‌یابیم:

$$f_y(0)=P(Y=0)=0.1+0.2=0.3$$

$$f(X|y=0)=\frac{f(x,0)}{f_y(0)}=\frac{f(x,0)}{0.3} \longrightarrow E(X|y=0)=\sum_{x=-1,2,3} x\frac{f(x,0)}{0.3}=(-1)\frac{0.1}{0.3}+2\frac{0.2}{0.3}+3\frac{0}{0.3}=1$$

نکته : امید ریاضی و واریانس **تعدادی تصادفی** از متغیرهای تصادفی

$$E\left(\sum_{i=1}^{N} X_i\right)=E(X_i)E(N)$$

$$Var\left(\sum_{i=1}^{N} X_i\right)=Var(X_i)E(N)+Var(N)\left(E(X_i)\right)^2$$

مثال : فرض کنید روزانه تعدادی مشتری با میانگین و واریانس برابر 200 نفر وارد یک فروشگاه می‌شوند هر مشتری دارای خرید با میانگین 180 دلار و انحراف معیار 10 دلاردارد میانگین و واریانس فروش روزانه فروشگاه چقدر است؟

حل:

چون

$$E(X_i)=180 \quad , \quad Var(X_i)=100 \quad , \quad E(N)=Var(N)=200$$

پس اگر $Y=\sum_{i=1}^{N} x_i$ فروش روزانه باشد:

$$E(Y)=180(200)=36000 \quad , \quad Var(Y)=100(200)+200(180)^2=6500000$$

کوواریانس و ضریب همبستگی

کوواریانس دو متغیر تصادفی Y,X عبارت است از:

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

منفی بودن کوواریانس نشان دهنده آن است که دو متغیر بر روی هم **اثر معکوس** دارند، یعنی با افزایش یکی دیگری کاهش پیدا می‌کند و برعکس.

$$Cov(X,Y)=Cov(Y,X)$$

$$Cov(X,X)=Var(X)$$

$$Cov(a,X)=0$$

$$Cov(aX,bY)=ab\,Cov(X,Y)$$

$$Cov\left(\sum_{i=1}^{n} X_i,\sum_{j=1}^{m} Y_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i,Y_j)$$

$$Var(X+Y)=Var(X)+Var(Y)+2\,Cov(X,Y)$$

مستقل $X,Y \Rightarrow Cov(X,Y)=0$

ضریب همبستگی $\rho_{X,Y}=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$ در نامساوی $|\rho|\le 1$، یعنی $-1\le\rho\le 1$ صدق می‌کند.

نکته : شرط لازم و کافی برای آنکه ضریب همبستگی بین دو متغیر برابر 1 یا 1- شود، آن است که دو متغیر رابطهٔ خطی داشته باشند.

$$Y=aX+b \quad \Leftrightarrow \quad \rho=\frac{a}{|a|} \quad ; \quad a\ne 0$$

اگر X و Y مستقل باشند، ضریب همبستگی آن‌ها صفر است اما اگر ضریب همبستگی صفر باشد، استقلال دو متغیر نتیجه نمی‌شود. در این حالت گویند X و Y ناهمبسته‌اند، یعنی رابطهٔ خطی ندارند.

اگر $U=aX+b$, $W=cY+d$ دو ترکیب خطی از X و Y باشند، ضریب همبستگی U و W از نظر مقداری با ضریب همبستگی X و Y برابر است.

یعنی:

$$\rho_{aX+b,cY+d}=\frac{ac}{|ac|}\rho_{x,y}=\begin{cases}\rho_{X,Y} & \text{a و c هم‌علامت باشند}\\ -\rho_{X,Y} & \text{A و c هم‌علامت نباشند}\end{cases}$$

نکته : **واریانس XY** (وقتی X و Y مستقل هستند)

$$Var(XY)=E\left(X^2Y^2\right)-\left(E(X)^2(E(Y))\right)^2=\left(\mu_X^2+\sigma_X^2\right)\left(\mu_Y^2+\sigma_Y^2\right)-\sigma_X^2\sigma_Y^2$$

مثال : اگر $\frac{4X}{Y}+\frac{Y}{4X}=-2$ آنگاه ضریب همبستگی **X** و **Y** کدام است؟

حل: داریم $16X^2+Y^2=-8XY$ یا $(4X-Y)^2=0$ پس $Y=4X$ و لذا X با Y رابطه خطی دارد و ضریب همبستگی یک می‌شود.

مثال : اگر ضریب همبستگی T_1 و T_2 برابر -0.4 باشد ضریب همبستگی $X=2T_1-4$ و $Y=3-2T_2$ چقدر است؟

حل: بنا به نکته بالا ضریب همبستگی 0.4 می‌شود.

مثال : اگر $Var(Y)=5$, $Var(X)=3$, $E(Y)=-2$, $E(X)=4$ و **X** و **Y** مستقل باشند واریانس XY کدام است؟

حل:

$$Var(XY)=16(5)+3(-2)^2+3(5)=107$$

تابع مولد گشتاورها

به تابع $M_X(t)=E\left(e^{tX}\right)$ تابع مولد گشتاورها گفته می‌شود. برای یافتن گشتاور مرتبه kام یعنی $E\left(X^k\right)$ باید k بار از $M_X(t)$ مشتق گرفت و بازای t عدد صفر را قرار داد. یعنی:

$$M_X^{(k)}(0)=E\left(X^k\right)$$

نکته :

$$M_{aX+b}(t)=e^{bt}M_X(at)$$

مثال : اگر $M_X(t)=\frac{\left(1+2e^{t}\right)^4}{3^4}$ و $Y=3X-2$ تابع مولد گشتاور **Y** کدام است.

حل:

$$M_Y(t)=e^{-2t}M_X(3t)=e^{-2t}\frac{\left(1+2e^{3t}\right)^4}{3^4}$$

نکته : با داشتن تابع مولد می‌توان به توزیع اصلی پی برد.

مثال : اگر $M_X(t)=0.2e^{-3t}+0.3e^{4t}+0.5$ تابع احتمال **X** کدام است؟

حل: باتوجه به آنکه اگر

X	-3	4	0
	0.2	0.3	0.5

پس $M_X(t)=0.2e^{-3t}+0.3e^{4t}+0.5$ پس توزیع x پیدا شد.

فصل چهارم : توزیع‌ها

توزیع یکنواخت گسسته

اگر X مقادیر n ,....., 2 , 1 را با احتمالات $P(X=x)=\frac{1}{n}$ اختیار کند، می‌گویند X توزیع یکنواخت گسسته بر $\{1,2,...,n\}$ دارد.

تابع مولد گشتاورها عبارت است از:

$$M_x(t)=\sum_{x=1}^{n}\frac{e^{tx}}{n}=\frac{e^t+e^{2t}+...+e^{tn}}{n}=\frac{e^t-e^{t(n+1)}}{n\left(1-e^t\right)}$$

و امید ریاضی برابر است با: $E(X)=\sum_{x=1}^{n}\frac{x}{n}=\frac{n+1}{2}$ و واریانس این توزیع، $Var(X)=\frac{n^2-1}{12}$ است.

مثال : در پرتاب یک تاس واریانس شماره ظاهر شده چقدر است؟

حل: اگر x شماره ظاهر شده باشد توزیع یکنواخت بر $\{1,2,..,6\}$ دارد و $Var(x)=\frac{6^2-1}{12}=\frac{35}{12}$

توزیع برنولی

به هر آزمایشی که بتوان نتیجه آن را به دو صورت پیروزی و شکست بیان کرد **آزمایش برنولی** گویند.

$$P(X=x)=p^x(1-p)^{1-x} \qquad x=0,1$$

تابع توزیع تجمعی X عبارت است از:

$$F_X(x)=\begin{cases}0 & x<0\\ q & 0\le x<1\\ 1 & 1\le x\end{cases}$$

تابع مولد گشتاورهای X را می‌توان به صورت زیر یافت:

$$M_X(t)=E\left(e^{tX}\right)=\sum_{x=0}^{1}e^{tx}\;P(X=x)=P(X=0)+e^tP(X=1)=q+pe^t$$

$$E(X)=q(0)+p(1)=p \text{ و } E\left(X^2\right)=q(0)^2+p(1)^2=p \text{ و } Var(X)=E\left(X^2\right)-(E(X))^2=pq$$

نکات:

۱ـ اگر X توزیع برنولی با پارامتر p داشته باشد $Y=X^k$ نیز همان توزیع را دارد یعنی:

$$X\sim b(p) \;\Rightarrow\; Y=\underset{k>0}{X}^k\sim b(p)$$

۲ـ اگر X توزیع برنولی با پارامتر p داشته باشد $Y=1-X$ توزیع برنولی با پارامتر q دارد یعنی:

$$X\sim b(p) \;\Rightarrow\; Y=1-X\sim b(q)$$

۳ـ حاصل‌ضرب n متغیر i.i.d برنولی با پارامتر p توزیع برنولی با پارامتر p^n دارد یعنی:

$$X_1,\dots,X_n \overset{i.i.d}{\sim} b(p) \;\Rightarrow\; Y=\prod_{i=1}^{n}X_i\sim b\left(p^n\right)$$

۴ ـ مینیمم n متغیر i.i.d برنولی با پارامتر p توزیع برنولی با پارامتر p^n دارد یعنی:

$$X_1,\dots,X_n \overset{i.i.d}{\sim} b(p) \quad \Rightarrow \quad Y=\min_i\{X_i\}\sim b\left(p^n\right)$$

۵ ـ ماکزیموم n متغیر i.i.d برنولی با پارامتر p توزیع برنولی با پارامتر $1-q^n$ دارد یعنی:

$$X_1,\dots,X_n \overset{i.i.d}{\sim} b(p) \quad \Rightarrow \quad Y=\max_i\{X_i\}\sim b\left(1-q^n\right)$$

مثال : اگر $X_1,\dots,X_5$ مستقلاً توزیع برنولی با پارامتر $p=0.2$ داشته باشند تابع مولد گشتاورهای $X_{(5)}=\max\{X_i\}$ کدام است؟

حل: بنا به نکته ۵ می‌دانیم:

$$X_{(5)}\sim b\left(p=1-(0.8)^5\right)$$

$$M_y(t)=\left(q+pe^t\right)=(0.8)^5+\left(1-(0.8)^5\right)e^t$$

مثال : اگر $X_1,\dots,X_n \overset{i.i.d}{\sim} b(p)$ و $X_{(1)}=\min\{X_i\},X_{(n)}=\max\{X_i\}$ و $E\left(X_{(1)}X_{(n)}\right)$ را پیدا کنید.

حل:

$$X_{(1)}X_{(n)}=\begin{cases}1 & X_{(1)}=X_{(n)}=1\\ 0 & \text{در غیر این صورت}\end{cases}$$

$$\begin{aligned}E\left(X_{(1)}X_{(n)}\right)&=P\left(X_{(1)}=X_{(n)}=1\right)\\&=P(X_1=X_2=\cdots=X_n=1)=p^n\end{aligned}$$

دقت شود $X_{(1)},X_{(n)}$ وابسته‌اند و $E\left(X_{(1)}X_{(n)}\right)\neq E\left(X_{(1)}\right)E\left(X_{(n)}\right)$

مثال: اگر $X\sim b(n_1,p),Y\sim b(n_2,q)$ که $p+q=1,n_1+n_2=n$ چقدر احتمال دارد $X=Y$

حل:

$$\begin{aligned}P(X=Y)&=\sum_{x=0}^{n_1}P(X=x,Y=x)=\sum_{x=0}^{n_1}P(X=x)P(Y=x)\\&=\sum_{x=0}^{n_1}\binom{n_1}{x}p^xq^{n_1-x}\binom{n_2}{x}q^xp^{n_2-x}=p^{n_2}q^{n_1}\sum_{x=0}^{n_1}\binom{n_1}{x}\binom{n_2}{x}\\&=p^{n_2}q^{n_1}\binom{n_1+n_2}{n_2}=p^{n_2}q^{n_1}\binom{n}{n_2}\end{aligned}$$

مثال : **5** توپ از ظرفی که **3** سیاه و **6** سفید دارد با جاگذاری بر می‌داریم چقدر احتمال دارد تعداد توپ سفید خارج شده زوج باشد

حل: اگر X تعداد سفید خارج شده باشد $X\sim b\left(5,\frac{2}{3}\right)$, $P(X)=\frac{1+(q-p)^n}{2}=\frac{1+\left(-\frac{1}{3}\right)^5}{2}=\frac{121}{243}$

توزیع دوجمله‌ای

فرض کنید یک آزمایش برنولی را n بار به طور متوالی و مستقل از هم تکرار کنیم و متغیر تصادفی X تعداد پیروزی را بشمارد، پس:

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$$

$x=0,1,2,\dots,n$

$M_X(t)=\left(q+pe^t\right)^n \quad , \quad E(X)=np \quad , \quad Var(X)=npq$

دقت شود $q=\frac{Var(X)}{E(X)}$ و با داشتن امید ریاضی و واریانس می‌توان n و p را یافت.

نکات:

۱ـ تعداد شکست‌ها توزیع برنولی با پارامتر n و q دارد یعنی:

$X\sim(n,p) \Rightarrow Y=n-X\sim b(n,q)$

$\rho_{X,Y}=-1$

ضریب همبستگی تعداد پیروزی‌ها و تعداد شکست‌ها برابر -1 است.

$Cov(X,Y)=Cov(X,n-X)=-Var(X)=-npq$

۲ـ اگر $X\sim b(n,p)$، تابع احتمال X یعنی $P(X=x)$ بیشترین مقدار خود را در نقطه $x_0=[(n+1)p]$ اختیار می‌کند یا به عبارتی:

$\forall x; P(X=x_0)\geq P(X=x)$

اگر $x_0=(n+1)p$ عدد صحیح بشود، تابع احتمال در x_0 و در x_0-1 دارای بیشترین احتمال است. یعنی در این حال برای هر x داریم:

$P(X=x_0)=P(X=x_0-1)\geq P(X=x)$

۳ ـ هر متغیر دوجمله‌ای از مجموع n متغیر i.i.d برنولی تشکیل شده است.

مثال : اگر در یک توزیع دوجمله‌ای امید ریاضی **4** برابر واریانس باشد چقدر احتمال دارد تمام آزمایش‌ها پیروزی یا تمام آن‌ها شکست باشد.

حل: ازآنجاکه $p=\frac{3}{4}$ $\frac{Var(X)}{E(X)}=\frac{1}{4}=q\rightarrow$ پس $P(X=0)+P(X=n)=q^n+p^n=\left(\frac{1}{4}\right)^n+\left(\frac{3}{4}\right)^n$

مثال : اگر $X\sim b\left(10,\frac{1}{3}\right)$ و $Y\sim b\left(11,\frac{2}{3}\right)$ مستقل باشند $Y=X-Y+11$ چه توزیعی دارد.

حل: اولاً $11-Y\sim b\left(11,\frac{1}{3}\right)$ پس $T=11-Y+X\sim b\left(21,\frac{1}{3}\right)$

مثال : اگر **X** توزیع دوجمله‌ای با پارامترهای $n=59$ و $p=\frac{2}{3}$ داشته باشد کدام مقدار **x** دارای احتمال ماکزیمم است.

۱) فقط 40 ۲) فقط 39 ۳) فقط 20 ۴) فقط 39 و 40

حل: بنا به نکته بالا گزینه ۴ درست است.

چون $(n+1)p=40$ عدد صحیح می‌باشد.

توزیع پواسون

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

گویند X توزیع پواسون با پارامتر $\lambda>0$ دارد و می‌نویسند $X\sim P(\lambda)$.

$$M_X(t)=e^{\lambda(e^t-1)} \quad \text{و} \quad E(X)=Var(X)=\lambda$$

نکات:

۱ـ تابع احتمال پواسون در $x_0=[\lambda]$ بیشترین احتمال را دارد یعنی اگر X توزیع پواسون با پارامتر λ داشته باشد آن‌گاه:

$$\forall_x \ ; P(X=x)\le P(X=x_0)$$

و نیز اگر $x_0=\lambda$ عددی طبیعی باشد،

$$\forall x; P(X=x)\le P(X=\lambda)=P(X=\lambda-1)$$

۲ـ اگر متغیر تصادفی X مقادیر صحیح نامنفی بگیرد و $\frac{P(X=x+1)}{P(X=x)}=\frac{\lambda}{x+1}$ آن‌گاه X توزیع پواسون با پارامتر $\lambda>0$ دارد.

متغیرهای شرطی پواسون به شرط معلوم بودن مجموع

۳ـ اگر X و Y مستقل باشند و $X\sim P(\lambda_1), Y\sim P(\lambda_2)$ آن‌گاه:

$$P(X=x|X+Y=n)=\binom{n}{x}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-x}$$

اگر مجموع دو متغیر مستقل پواسون معلوم باشد هر یک از آن‌ها توزیع دوجمله‌ای دارند.

مثال : اگر تعداد زنان و مردان ورودی به یک فروشگاه در روز مستقل از هم توزیع پواسون به‌ترتیب با $\lambda_1=100$ و $\lambda_2=200$ داشته باشد کدام عدد برای تعداد زنان احتمال ماکزیمم دارد اگر بدانیم تعداد کل مشتری‌ها 599 نفر بوده است.

حل: می‌دانیم $X|X+Y=n\sim b\left(n,\frac{\lambda_1}{\lambda_1+\lambda_2}\right)$ پس $X\sim b\left(599,\frac{1}{3}\right)$ تعداد زنان

و $n_0=[(n+1)p]=200$ دارای بیشترین احتمال است ضمناً $P(X=199)=P(X=200)$

مثال : اگر در یک توزیع پواسون $P(X=4)=P(X=5)$ پس $P(X\le 0)$ چقدر است؟

حل: بنا به نکته ۲ داریم:

$$P(X\le 0)=P(X=0)=e^{-5} \qquad , \qquad \lambda=5$$

مثال : اگر **X** دارای توزیع پواسون با $P(X=0)=a$ باشد $E(X!)$ چقدر است؟

حل: می‌دانیم $E(X!)=\sum_{x=0}^{\infty}e^{-\lambda}\lambda^x=\frac{e^{-\lambda}}{1-\lambda}$ اما $a=e^{-\lambda}$ پس $\lambda=-Lna$

لذا جواب می‌شود $\frac{a}{1+Lna}$

مثال : اگر **X** تعداد تصادف روزانه چهارراهی با $\lambda = 2$ باشد چقدر احتمال دارد در **5** روز تعداد تصادف‌ها زوج باشد

حل: اگر $X \sim P(\lambda)$ پس $P(\text{زوج } X) = \frac{1+e^{-2\lambda}}{2}$ اما در 5 روز $\lambda = 10$ لذا $P(X\text{زوج}) = \frac{1+e^{-20}}{2}$

توزیع هندسی

یک آزمایش برنولی را آنقدر به طور متوالی و مستقل از هم تکرار کنیم تا به نخستین پیروزی برسیم و X تعداد تکرار لازم را بشمارد. پس:

$$P(X = x) = pq^{x-1} \qquad x = 1, 2, 3, \ldots.$$

$$M_X(t) = \frac{pe^t}{1 - qe^t} \quad , \quad t < -Ln\,q \quad , \quad E(X) = \frac{1}{p} \quad , \quad Var(X) = \frac{q}{p^2}$$

نکته‌ها:

۱ـ بی‌حافظه‌بودن

اگر X توزیع هندسی با پارامتر p داشته باشد:

$$P(X = a + b \mid X > a) = P(X = b) \quad ; \quad a, b \in \mathbb{N}$$

$$P(X > m + n \mid X > m) = P(X > n)$$

۲ـ مینیمم هندسی‌ها

اگر $X_1, \ldots, X_n \overset{i.i.d}{\sim} G(p)$ آن‌گاه:

$$Y = \min_i \{X_i\} \sim G(1 - q^n)$$

۳ـ مد در توزیع هندسی

با توجه به آنکه در توزیع هندسی $P(X=1) \geq P(X=2) \geq P(X=3) \geq \cdots$ نتیجه می‌شود برای هر $x \in N$؛ $P(X=1) \geq P(X=x)$ یعنی توزیع هندسی در $x = 1$ مد دارد.

۴ـ متغیرهای شرطی هندسی به شرط معلوم بودن مجموع

اگر $X, Y \overset{i.i.d}{\sim} G(p)$ آن‌گاه $P(X = x \mid X + Y = n) = \frac{1}{n-1} \quad , \quad x = 1, 2, \cdots, n-1$

در واقع وقتی $X + Y = n$، X دارای توزیع یکنواخت گسسته بر مجموعه $\{1, 2, \cdots, n-1\}$ است.

مثال : اگر **X** توزیع هندسی با میانگین **5** داشته باشد انحراف معیار **X** کدام است؟

حل:

$$E(X) = \frac{1}{p} = 5 \Rightarrow p = \frac{1}{5} \Rightarrow \sigma^2 = Var(X) = \frac{q}{p^2} = \frac{\frac{4}{5}}{\frac{1}{25}} = 20 \Rightarrow \sigma = \sqrt{20}$$

مثال : اگر **Y,X** مستقل هندسی با $P=\frac{2}{5}$ باشند چقدر احتمال دارد $X=Y$

حل:

$$P(X=Y)=\frac{p}{2-p}=\frac{p}{1+q}=\frac{\frac{2}{5}}{1+\frac{3}{5}}=\frac{2}{8}=\frac{1}{4}$$

مثال : اگر **A** و **B** مستقلاً تاسی را پرتاب کنند تا **6** بیابید و در مجموع با **20** پرتاب هردو موفق شده باشند چقدر احتمال دارد **A** بیشتر از **B** پرتاب کرده باشد.

حل: بنا به نکته بالا:

$$P(X=x \mid X+Y=20)=\frac{1}{19}$$

$$x=1,2,\ldots,19$$

$$P(X>Y)=\sum_{x=10}^{19}\frac{1}{19}=\frac{10}{19}$$

توزیع دوجمله‌ای منفی (پاسکال)

فرض کنید یک آزمایش برنولی را آن‌قدر به طور متوالی و مستقل از هم تکرار کنیم تا به rامین پیروزی برسیم و متغیر تصادفی X تعداد تکرار لازم را بشمارد. پس:

$$P(X=x)=\binom{x-1}{r-1}p^r q^{x-r} \qquad x=r,r+1,\ldots$$

در این حالت می‌گویند X توزیع دوجمله‌ای منفی با پارامترهای p,r دارد و می‌نویسند $X\sim Nb(r,p)$

$$M_X(t)=\left[\frac{pe^t}{1-qe^t}\right]^r \quad \text{و} \quad E(X)=\frac{r}{p} \quad \text{و} \quad Var(X)=\frac{rq}{p^2}$$

مثال : یک فوتبالیست که با احتمال **0.6** پنالتی را گل می‌کند به‌طور متوسط چند پنالتی بزند تا **7** عدد گل داشته باشد.

$$E(x)=\frac{r}{p}=\frac{7}{0.6}=\frac{70}{6}$$

مثال : یک عدد از بازده $(0,1)$ انتخاب می‌کنیم در بسط اعشاری این عدد به طور متوسط چند رقم قبل از **4** امین **6** وجود دارد؟

حل: اگر X تعداد رقم‌ها تا رسیدن به 4امین 6 باشد پس

$$X\sim Nb(r=4,p=0.1)$$

$$E(X)=\frac{r}{p}=\frac{4}{\frac{1}{10}}=40$$

چون گفته قبل از چهارمین 6 جواب 39 می‌شود.

توزیع فوق هندسی

فرض کنید جعبه‌ای N توپ دارد که a عدد آن سفیدند. از این جعبه بدون جایگذاری n توپ خارج می‌کنیم و X تعداد توپ سفید خارج‌شده را می‌شماریم. پس:

$$P(X=x)=\frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}} \qquad x=0,1,2,....,n$$

$$E(X)=n\frac{a}{N}$$

$$Var(X)=n\frac{a}{N}\left(1-\frac{a}{N}\right)\left(\frac{N-n}{N-1}\right)$$

نکته : متغیرهای شرطی دوجمله‌ای به شرط معلوم بودن مجموع آن‌ها

اگر $X\sim b(n_1,p)$, $Y\sim b(n_2,p)$ مستقل باشند آن‌گاه:

$$P(X=x|X+Y=m)=\frac{\binom{n_1}{x}\binom{n_2}{m-x}}{\binom{n_1+n_2}{m}}$$

(در واقع به شرط آنکه $X_1+X_2=m$، X_1 توزیع فوق هندسی با پارامترهای $N=n_1+n_2$ و $n=m$ و $a=n_1$ دارد)

مثال : فرض کنید تعداد زنان و مردان دارای تحصیلات عالیه یک شرکت مستقل از هم توزیع دوجمله‌ای به‌ترتیب با $n_1=100,p=\frac{1}{4}$ و $n_2=200,p=\frac{1}{4}$ داشته باشد. اگر بدانیم **80** نفر تحصیلات عالیه داشته‌اند چقدر احتمال دارد 20 نفر زن باشند.

حل: بنا به نکته بالا:

$$P(X=20|X+Y=80)=\frac{\binom{100}{20}\binom{200}{60}}{\binom{300}{80}}$$

توزیع یکنواخت پیوسته

فرض کنید تابع چگالی متغیر پیوسته X به صورت زیر باشد:

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & \text{در غیر این صورت}\end{cases}$$

می‌گویند X توزیع یکنواخت بر بازه $[a\ ,\ b]$ دارد و می‌نویسند $X\sim U_{[a,b]}$.

$$F_X(x)=\frac{x-a}{b-a}\quad,\quad E(X)=\frac{a+b}{2}\quad,\quad V_{an}(X)=\frac{(b-a)^2}{12}$$

نکته : اگر X متغیری پیوسته باشد که $F_x(t)$ تابعی اکیداً صعودی و تابع توزیع آن است، متغیر $Y=F(X)$ دارای توزیع یکنواخت بر بازه $[0,1]$ است.

مثال : اگر $F_X(t)$ تابع توزیع متغیر و پیوسته **X** باشد واریانس $Y=F(X)$ کدام است؟

حل: بنا به نکته بالا $Y\sim U(0,1)$ و $Var(Y)=\dfrac{(b-a)^2}{12}=\dfrac{1}{12}$

توزیع نرمال

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

X توزیع نرمال با پارامترهای $\mu\in R$, σ^2 دارد.

تابع توزیع تجمعی نرمال استاندارد:

$$\Phi(x)=P(Z\le x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz$$

$$\Phi(-x)=1-\Phi(x)\quad \text{و}\quad \Phi(0)=\frac{1}{2}$$

$$X\sim N(\mu,\sigma^2)\quad\Leftrightarrow\quad Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

$$M_X(t)=e^{\mu t+\frac{\sigma^2t^2}{2}}\quad,\quad E(X)=\mu\quad,\quad Var(X)=\sigma^2$$

نکته : گشتاورهای مرکزی مرتبه فرد توزیع نرمال همواره صفر می‌شوند.

و گشتاورهای زوج عبارتند از:

$$\mu_{2k}=E(X-\mu)^{2k}=\frac{(2k)!\sigma^{2k}}{k!2^k}$$

نکته : نرمال دومتغیره:

توزیع‌های کناری X و Y نیز نرمال هستند و توزیع‌های شرطی نیز نرمال هستند و X و Y مستقلند اگر و تنها اگر $\rho=0$

نکته : هر ترکیب خطی از متغیرهای تصادفی مستقل نرمال دارای توزیع نرمال است.

$$Y=\sum_{i=1}^{n} c_i X_i \sim N\left(\sum_{i=1}^{n} c_i \mu_i \,,\, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$$

مثال : اگر **X** دارای تابع مولد گشتاور $M_X(t)=e^{t(t-1)}$ باشد $P(X<1)$ کدام است.

حل: بنا بر تطبیق با تابع مولد گشتاور نرمال داریم:

$$M_X(t)=e^{-t+t^2}=e^{\mu t+\frac{\sigma^2 t^2}{2}}$$

پس $\sigma^2=2$ و $\mu=-1$ و لذا:

$$P(X<1)=P\left(Z<\frac{1+1}{\sqrt{2}}\right)=P\left(Z<\sqrt{2}\right)=\Phi\left(\sqrt{2}\right)$$

مثال : اگر $X_i \sim N\left(i,i^2\right)$ که $i=1,2,3$ مستقل باشند آنگاه وقتی $X=\sum_{i=1}^{3} iX_i$، $P(X<10)$ چقدر است؟

حل:

درواقع:

$$X=X_1+2X_2+3X_3 \sim N(1+2(2)+3(3),1+4(4)+9(9)) \approx N(14,98)$$

مثال : اگر **X** توزیع نرمال با $\mu=2$, $\sigma^2=1$ داشته باشد امید ریاضی e^{-4x} می‌شود.

حل:

$$E\left(e^{-4x}\right)=M_X(-4)=e^{\mu t+\frac{\sigma^2+2}{2}}=e^{2(-4)+\frac{16(1)}{2}}=e^{-8+8}=e^0=1$$

تابع گاما

تعریف: تابع $\Gamma(\alpha)=\int_0^{\infty} x^{\alpha-1}e^{-x}dx$ را تابع گاما می‌گویند.

(الف) تابع $\Gamma(\alpha)$ به ازای هر $\alpha>0$ همگرا است.

(ب) برای هر $\alpha>1$ داریم $\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$

(ج) مقدر تابع گاما در نقطه 1 برابر یک و مقدار تابع گاما در نقطه $\frac{1}{2}$ برابر $\sqrt{\pi}$ است ضمناً وقتی $n\in N$ داریم: $\Gamma(n)=(n-1)!$

توزیع گاما

$$f(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} \qquad x>0$$

$$X \sim G(\alpha,\beta)$$

$$M_X(t)=E\left(e^{tX}\right)=\left(\frac{\beta}{\beta-t}\right)^{\alpha} \qquad \beta>t \quad , \quad E(X)=\frac{\alpha}{\beta} \quad , \quad Var(X)=\frac{\alpha}{\beta^2}$$

نکته : اگر $X\sim G(\alpha\,,\beta)$ و $k>0$ و $Y=kx$ پس $Y\sim G\left(\alpha\,,\frac{\beta}{k}\right)$

اگر متغیری را که دارای توزیع گاما است در عددی مثبت ضرب کنیم باز هم توزیع گاما برقرار می‌شود.

گشتاورهای گاما $\quad \alpha+k>0 \quad E\left(X^k\right)=\dfrac{\Gamma(\alpha+k)}{\beta^k\Gamma(\alpha)}$

مثال: اگر از جامعه با چگالی $f(x)=2e^{-2x}$ نمونه **2** تایی X_1, X_2 را بگیریم $E\left(\dfrac{1}{X_1+X_2}\right)$ را بیابید.

حل: چون X_1+X_2 توزیع گاما با $\beta=2, \alpha=2$ دارد پس $E\left(\dfrac{1}{X_1+X_2}\right)=\dfrac{\beta}{\alpha-1}=2$

مثال: اگر **X** توزیع گاما با پارامترهای **4** و **4** داشته باشد $Y=4X$ چه توزیعی دارد؟

حل: بنا به نکته بالا $Y\sim G(4,1)$ که چگالی Y عبارتست از: $\quad f_Y(y)=\dfrac{y^3e^{-y}}{6} \quad y>0$

مثال: در مثال بالا امید ریاضی $Y=X^4$ چقدر است؟

حل:

$$E\left(X^4\right)=\frac{\Gamma(8)}{4^4\Gamma(4)}=\frac{7!}{4^4(3!)}=\frac{210}{64}$$

توزیع نمایی (حالت خاص گاما)

اگر در توزیع گاما قرار دهیم $\alpha=1$ و $\beta=\lambda$ پس:

$$f(x)=\lambda e^{-\lambda x} \qquad x>0$$

X توزیع نمایی (منفی) با پارامتر $\lambda>0$ دارد بدیهی است با توجه به ویژگی‌های توزیع گاما:

$$M_X(t)=\frac{\lambda}{\lambda-t} \quad \lambda>t \quad , \quad E(X)=\frac{1}{\lambda} \quad , \quad Var(X)=\frac{1}{\lambda^2}$$

اگر X توزیع نمایی با پارامتر $\lambda>0$ داشته باشد $E\left(X^k\right)$ عبارت است از (**گشتاورهای مرتبه k‌ام توزیع نمایی**)

$$E\left(X^k\right)=\frac{\Gamma(k+1)}{\lambda^k\Gamma(1)}=\frac{k!}{\lambda^k} \qquad k\in N$$

خاصیت فاقد حافظه بودن $P(X>a+b|X>a)=P(X>b)$ **در میان پیوسته‌ها فقط برای نمایی برقرار است.**

نکته: اگر $X\sim U_{[0,1]}$ متغیر $Y=\dfrac{-\operatorname{Ln}X}{\lambda}$ که $\lambda>0$ است دارای توزیع نمایی با پارامتر λ است.

نکته: اگر X_iها مستقل و $X_i\sim Exp(\lambda_i)$ پس:

$$X_{(1)}=Y_1=\min_i\{X_i\}\sim E_{xp}\left(\sum_{i=1}^n\lambda_i\right)$$

نکته: توزیع ارلنگ $G(n,\lambda)$ توزیع زمان انتظار برای وقوع nامین پیروزی در یک فرایند پواسون است.

مثال: اگر $X_1\sim E_{Xp}(2)$ و $X_2\sim E_{Xp}(3)$ و $X_3\sim E_{Xp}(4)$ مستقل از هم باشد چقدر احتمال آن متغیری که از بقیه کوچکتر است (**Min**) از **3** کوچکتر باشد.

حل: بنا به نکته بالا $X_{(1)} = \min_i \{X_i\} \sim E_{Xp}(2+3+4)$ و $P\left(X_{(1)} < 3\right) = \int_0^3 9e^{-9x}\, dX = 1 - e^{-2v}$

مثال : فرض کنید عمر لامپ‌ها توزیع نمایی با میانگین **2000** ساعت دارد چقدر احتمال دارد یک لامپ که **500** ساعت عمر کرده حداقل **1500** ساعت عمر کند.

حل:

$$P(X > 1500 | X > 500) = P(X > 1000) = e^{-1000\left(\frac{1}{2000}\right)} = e^{-\frac{1}{2}}$$

توزیع مربع کای (حالت خاص گاما)

اگر در توزیع گاما قرار دهیم:

$$\alpha = \frac{v}{2}\ ,\ \beta = \frac{1}{2}$$

$$M_X(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2} - t}\right)^{\frac{v}{2}} = \frac{1}{(1-2t)^{\frac{v}{2}}} \quad t < \frac{1}{2} \qquad , \qquad E(X) = v \qquad , \qquad Var(X) = 2v$$

نکته : اگر X توزیع نرمال استاندارد داشته باشد $Y = X^2$ دارای توزیع مربع کای با $v = 1$ درجه آزادی است.

نکته : اگر $X \sim G(\alpha, \beta)$ پس $2\beta X \sim G\left(\alpha, \frac{1}{2}\right) \equiv \chi^2_{2\alpha}$ بر این اساس هر توزیع گاما را می‌توان به مربع کای تبدیل کرد.

رابطه مربع کای و نمایی

$$\chi^2_{(2)} \equiv E_{xp}\left(\frac{1}{2}\right)$$

توزیع مربع کای با 2 درجه آزادی همان نمایی با پارامتر $\frac{1}{2}$ است.

مثال : اگر **x** توزیع یکنواخت بر بازه $(0,1)$ داشته باشد $Y = -2L_n X$ چه توزیعی دارد.

۱) نمایی با میانگین $\frac{1}{2}$ ۲) نمایی با میانگین 4 ۳) مربع کای با 2 درجه آزادی ۴) یکنواخت بر $(0,\infty)$

حل: گزینه ۲ درست است.

$$Y = -2\operatorname{Ln} X = \frac{-\operatorname{Ln} X}{\frac{1}{2}} \sim Exp\left(\frac{1}{2}\right) \equiv X^2_{(2)}$$

توزیع بتا

$$f(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1} \qquad 0<x<1$$

که در آن:

$$B(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$E(X)=\frac{a}{a+b} \quad , \quad Var(X)=\frac{ab}{(a+b)^2(a+b+1)}$$

مثال : اگر $f(x)=Kx^2(1-x)^3 \,, 0<x<1$ تابع چگالی باشد امید ریاضی و واریانس X چقدر است.

حل: دقت شود $X \sim B(a=3,b=4)$ لذا $E(X)=\frac{a}{a+b}=\frac{3}{7}$ و $Var(x)=\frac{ab}{(a+b)^2(a+b+1)}=\frac{12}{7^2(8)}=\frac{3}{98}$

توزیع کوشی

اگر تابع چگالی متغیر پیوسته X به صورت زیر باشد:

$$f(x)=\frac{1}{\pi\sigma}\frac{1}{\left[1+\left(\frac{x-\mu}{\sigma}\right)^2\right]} \qquad x\in R$$

می‌گویند X توزیع کوشی با پارامترهای $\mu\in R$ ، $\sigma>0$ دارد. در حالت خاص که $\mu=0$ و $\sigma=1$ داریم:

$$f(x)=\frac{1}{\pi}\frac{1}{1+x^2} \qquad x\in R$$

می‌گویند X توزیع **کوشی استاندارد** دارد.

نکته : اگر $X\sim U_{\left(-\frac{\pi}{2},\frac{\pi}{2}\right)}$ و $Y=tg X$ آنگاه Y توزیع کوشی استاندارد دارد.

نکته : اگر X کوشی استاندارد داشته باشد، آنگاه $Y=\frac{1}{X}$ نیز همان توزیع را دارد.

نکته : اگر $N(0,1) \overset{i.i.d}{\sim} X,Y$ آنگاه $\frac{X}{Y},\frac{X}{|Y|},\frac{|X|}{Y}$ دارای توزیع کوشی استاندارد هستند.

مثال : اگر **x** توزیع نرمال استاندارد و **Y** توزیع نرمال با میانگین **2** و واریانس **1** داشته باشد **X**و **Y** مستقل باشند $P\left(\frac{X}{\sqrt{(Y-2)^2}}<1\right)$

چقدر است؟

حل: چون $\sqrt{(Y-2)^2}=|Y-2|$ پس:

$$P\left(\frac{X}{|Y-2|}<1\right)=\int_{-\infty}^{1}\frac{dt}{\pi(1+t^2)}=\frac{\tan^{-1}t}{\pi}\Big|_{-\infty}^{1}=\frac{1}{4}$$

نکته : روش یافتن تابع چگالی Y از روی تابع چگالی X

اگر X پیوسته و g تابعی اکیداً یکنوا باشد، آن‌گاه:

$$f_Y(y)=f_X\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right|$$

مثال : اگر **X** دارای توزیع نمایی با پارامتر $\lambda = 2$ داشته باشد $Y = e^{2X}$ چه توزیعی دارد.

حل: بنا به نکته بالا ابتدا x را برحسب y پیدا می‌کنیم.

$$2x = \operatorname{Ln} y \quad \Rightarrow \quad x = \frac{1}{2}\operatorname{Ln} y$$

سپس از آن مشتق می‌گیریم $\frac{dx}{dy}=\frac{1}{2y}$ حال در تابع چگالی نمایی $f(x)=2e^{-2x}$ به جای x می‌گذاریم $\frac{1}{2}\operatorname{Ln} y$ و حاصل را در مشتق ضرب می‌کنیم. $y>1$

$$f(y)=f\left(\frac{1}{2}\operatorname{Ln} y\right)\left|\frac{1}{2y}\right|=2e^{-\operatorname{Ln} y}=\frac{1}{y^2}$$

قضیهٔ حد مرکزی

اگر X_iها مستقل و دارای توزیع یکسان باشند و $E(X_i)=\mu$ و $V_{ar}(X_i)=\sigma^2<\infty$ و $\overline{X}=\frac{\sum_{i=1}^{n}X_i}{n}$ پس:

$$P\left(\frac{\sum_{i=1}^{n}X_i-n\mu}{\sigma\sqrt{n}}\le a\right)=P\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\le a\right)\xrightarrow{n\to\infty}\Phi(a)=\int_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{\frac{-z^2}{2}}dz$$

قضیهٔ حد مرکزی یک بیان احتمالی دارد و آن بیان این است که **مجموع n متغیر مستقل از هر توزیعی، دارای توزیع حدی نرمال است.** بنابراین توزیع‌های دوجمله‌ای، پواسون، گاما، دوجمله‌ای منفی و مربع کای دارای توزیع تقریبی نرمال هستند. چون متغیر دوجمله‌ای را می‌توان به صورت مجموع متغیرهای برنولی نوشت، می‌توان متغیر پواسون را به صورت مجموع متغیرهای پواسون و متغیر گاما را به صورت مجموع متغیرهای نمایی و متغیر دوجمله‌ای منفی را به صورت مجموع متغیرهای هندسی و متغیر مربع کای را به صورت مجموع متغیرهای مربع کای نوشت.

مثال : فرض کنید زمان لازم برای پاسخگویی به هر سؤال توزیع نمایی با میانگین یک دقیقه دارد چقدر احتمال دارد مجموع زمان لازم برای امتحان ۱۰۰ سؤالی کمتر از **90** دقیقه بشود.

حل: بنا به قضیه حد مرکزی:

$X=\sum_{i=1}^{100}X_i$ توزیع حدی نرمال دارد

$$P(X<90)=P\left(Z<\frac{90-100(1)}{1\sqrt{100}}\right)=P(Z<-1)=0.1587$$

نکته : قطع‌کردن توزیع‌ها

بعضی مواقع در توزیع‌های معروف بخشی از مقادیر ممکن x را حذف می‌کنند و بخشی از مقادیر ممکن x را نگه می‌دارند که به آن A می‌گوییم. تابع احتمال جدید همان تابع احمال قبلی است که بر $P(A)$ تقسیم می‌شود مثلاً:

اگر توزیع پواسون را برای $X\neq 0$ در نظر بگیریم تابع احتمال می‌شود:

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

یا در توزیع نمایی برای $x>a$ تابع احتمال می‌شود:

$$\frac{\lambda e^{-\lambda x}}{P(x>a)}=\frac{\lambda e^{-\lambda x}}{e^{-\lambda a}}\qquad x>a$$

$$=\lambda e^{-\lambda(x-a)}\qquad x>a$$

امید ریاضی و واریانس هم براین اساس گرفته می‌شود. مثلاً $E(X\,|\,X>a)=a+\frac{1}{\lambda}\quad\Rightarrow\quad X\sim Exp(\lambda)$

جدول خلاصه توزیع‌ها

ردیف	نام توزیع	تابع احتمال	تابع مولد گشتاورها	امیدریاضی	واریانس	توضیحات (حالت‌های خاص)
۱	دوجمله‌ای $b(n,p)$	$\binom{n}{x}p^x q^{n-x}$ $x=0,1,2,\dots,n$	$(q+pe^t)^n$	np	npq	اگر $n=1$ توزیع برنولی بوجود می‌آید.
۲	پواسون $P(\lambda)$	$\frac{e^{-\lambda}\lambda^x}{x!}$	$e^{\lambda(e^t-1)}$	λ	λ	-
۳	دوجمله‌ای منفی $Nb(r,p)$	$\binom{x-1}{r-1}p^r q^{x-r}$	$\left(\frac{pe^t}{1-qe^t}\right)^r$	$\frac{r}{p}$	$\frac{rq}{p}$	اگر $r=1$ توزیع هندسی بوجود می‌آید.
۴	فوق هندسی $HG(N,n,a)$	$\frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$	-	$n\frac{a}{N}$	$n\frac{a}{N}\left(1-\frac{a}{N}\right)\left(\frac{N-n}{N-1}\right)$	-
۵	یکنواخت پیوسته $U[a,b]$	$x=0,1,\dots,n$ $\frac{1}{b-a}\quad a\le x\le b$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	-
۶	گاما $G(\alpha,\beta)$	$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}\quad x>0$	$\left(\frac{\beta}{\beta-t}\right)^\alpha$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	اگر $\alpha=1$ توزیع نمایی بوجود می‌آید و اگر $\alpha=\frac{v}{2}$ و $\beta=\frac{1}{2}$ توزیع مربع کای بوجود می‌آید.
۷	نرمال $N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$e^{\mu t+\frac{\sigma^2t^2}{2}}$	μ	σ^2	
۸	بتا $B(a,b)$	$\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}$ $B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$	-	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	

فصل پنجم: برآوردکننده‌ها و توزیع‌های نمونه‌ای

روش گشتاوری MME

در روش گشتاوری اگر توزیع، یک پارامتر مجهول داشته باشد کافی است $E(X)$ (میانگین جامعه) را با $\bar{X}$ (میانگین نمونه) برابر قرار داده پارامتر را برحسب $\bar{X}$ پیدا کنید.

مثال : اگر **X** دارای چگالی $f(x)=\theta(\theta+1)x(1-x)^{\theta-1}$ باشد. برآورد روش گشتاوری θ را بیابید.
$0<x<1$

حل:

هر چند با انتگرال‌گیری هم جواب پیدا می‌شود اما X توزیع بتا با $\alpha=2$ و $\beta=\theta$ دارد پس $E(X)=\frac{2}{2+\theta}=\bar{X}$ و

$$MME(\theta)=\frac{2-2\bar{X}}{\bar{X}}$$

روش ماکزیمم درستنمایی MLE

در روش ماکزیموم درستنمایی باید تابع $L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$ را به عنوان تابعی از θ ماکزیموم کنید.

ـ در روش ماکزیموم درستنمایی داریم: $MLE(g(\theta))=g(MLE(\theta))$

یعنی برای یافتن $MLE(g(\theta))$ باید $MLE(\theta)$ را درون تابع g قرار داد. به همین دلیل باید اغلب MLE های معروف را حفظ باشید.

ـ اگر x به θ وابسته نباشد مشتق $L_nL(\theta)$ را گرفته صفر قرار می‌دهید θ را می‌یابید.

ـ اگر x به θ وابسته باشد برحسب آنکه $f(x,\theta)$ تابعی صعودی یا نزولی از θ است و با توجه به $\theta\le x$ یا $\theta\ge x$ به ترتیب $X_{(1)}=\min X_i$ و $X_{(n)}=\max X_i$ مفید خواهند بود.

مثال : براساس نمونه **n** تایی از چگالی $0<x<1$ $\quad f(x,\theta)=\theta x^{\theta-1}$ برای $e^{-\theta}$ برآورد ماکزیمم درستنمایی بسازید.

حل:

$$L(\theta)=\theta^n\left(\prod_{i=1}^{n} xi\right)^{\theta-1}\to L_nL(\theta)=nL_n\theta+(\theta-1)L_n\prod_{i=1}^{n} x_i$$

$$\to\frac{dL_nL(\theta)}{d\theta}=\frac{n}{\theta}+L_n\prod_{i=1}^{n} x_i=0\to\hat{\theta}=MLE(\theta)=\frac{-n}{L_n\prod_{i=1}^{n} x_i}$$

پس:

$$MLE(e^{-\theta})=e^{\frac{-n}{L_n\prod_{i=1}^{n} x_i}}$$

مثال : اگر نمونه **n** تایی از جامعه با تابع احتمال $P(X=x)=\left(\frac{\theta}{2-\theta}\right)\left(\frac{2(1-\theta)}{2-\theta}\right)^{x}$، $x=0,1,2,\cdots,0<\theta<1$ بگیریم $MME(\theta)$ را بیابید.

حل: می‌دانیم $X\sim G\left(p=\frac{\theta}{2-\theta}\right)$ چون

$$q=1-p=1-\frac{\theta}{2-\theta}=\frac{2-\theta-\theta}{2-\theta}=\frac{2(1-\theta)}{2-\theta}$$

در توزیع هندسی $E(X)=\frac{1}{p}$ و $\frac{1}{p}=\bar{X}$ پس $MME(p)=\frac{1}{\bar{X}}$ پس کافی است.

$$\frac{2-\theta}{\theta}=\frac{1}{\bar{X}}$$

را بر حسب θ حل کنیم که می‌شود

$$MME(\theta)=\frac{2}{2+\bar{X}}$$

مثال : اگر نمونه‌ای **n** تایی از جامعه با چگالی $x>0,\theta>0$, $f(x)=\frac{1-\theta}{1+\theta}e^{-\frac{1-\theta}{1+\theta}x}$ باشد $MLE(\theta)$ را بیابید.

حل: می‌دانیم $X\sim Exp\left(\frac{1-\theta}{1+\theta}\right)$ و در توزیع نمایی $MLE(\lambda)=\frac{1}{\bar{X}}$ پس $\frac{1-MLE(\theta)}{1+MLE(\theta)}=\frac{1}{\bar{X}}$ که داریم $MLE(\theta)=\frac{\bar{X}-1}{\bar{X}+1}$ البته اگر $\bar{X}\le 1$ بایستی $MLE(\theta)=0$

مثال : اگر $X_1,\ldots,X_n \underset{\sim}{i.i.d} U_{[0,\theta]}$ مطلوبست: $MLE(E(X))$

حل:

$$f(x_i)=\frac{1}{\theta}\qquad 0\le x_i\le\theta$$

چون حدود x به θ وابسته است و تابعی نزولی از θ است پس باید θ کمترین مقدار ممکن باشد که بزرگ‌ترین مقدار نمونه است پس $MLE(\theta)=X_{(n)}$ لذا چون $E(X)=\frac{\theta}{2}$ پس $MLE\left(\frac{\theta}{2}\right)=\frac{X_{(n)}}{2}$

جدول MME و MLE های معروف

توزیع	پارامتر	MME	MLE
برنولی	p	$\hat{p}=\frac{X}{n}$	$\hat{p}=\frac{X}{n}$
پواسون	λ	$\bar{X}$	$\bar{X}$
هندسی (مدل تکرار)	p	$\frac{1}{\bar{X}}$	$\frac{1}{\bar{X}}$
هندسی (مدل شکست)	p	$\frac{1}{1+\bar{X}}$	$\frac{1}{1+\bar{X}}$
نمایی (نوع اول)	λ	$\frac{1}{\bar{X}}$	$\frac{1}{\bar{X}}$
نمایی (نوع دوم)	λ	$\bar{X}$	$\bar{X}$
یکنواخت $[a,\theta]$	θ	$2\bar{X}-a$	$X_{(n)}=\max_i\{X_i\}$

خواص برآوردکننده‌ها

ـ اگر از ناریبی سؤال آمد، امید ریاضی برآورد مشترک در گزینه‌ها را یافته با سعی و خطا گزینه صحیح را می‌یابید.

ضمناً توجه کنید $\bar{X}$ برای $\mu=E(X)$ و $\hat{p}=\frac{X}{n}$ برای p و $S^2=\frac{\sum(X_i-\bar{X})^2}{n-1}$ برای $\sigma^2=Var(X)$ همواره ناریب هستند.

مثال: اگر از جامعه یکنواخت بر $(0,\theta)$ نمونه‌ای **n** تایی بگیریم و $X_{(n)}$ ماکزیمم نمونه‌ها باشد بازای کدام مقدار **C** آماره $CX_{(n)}$ برای θ ناریب است.

حل: می‌دانیم چگالی ماکزیمم نمونه‌ها به صورت زیر است

$$f_{X_{(n)}}(t)=n(F(t))^{n-1}f(t)=n\left(\frac{t}{\theta}\right)^{n-1}\frac{1}{\theta}\quad 0<t<\theta$$

$$E(X_{(n)})=\int_0^\theta \frac{nt^n}{\theta^n}dt=\frac{n}{n+1}\theta\rightarrow E\left(\frac{n+1}{n}X_{(n)}\right)=\theta\rightarrow C=\frac{n+1}{n}$$

ـ معیار میانگین مجذور خطا (MSE)

$$MSE(\hat{\theta})=E(\hat{\theta}-E(\hat{\theta}))^2+(E(\hat{\theta})-\theta)^2=Van(\hat{\theta})+b^2(\hat{\theta})$$

MSE می‌تواند برآوردکننده‌های اریب و ناریب را با هم مقایسه کند، ضمناً:

MSE = واریانس + (اریبی) ^{2}

در توزیع‌های برنولی ـ نرمال ـ پواسون ـ نمایی (با میانگین λ) $\bar{X}$ کاراترین برآوردکنندهٔ میانگین جامعه است. اما این مطلب عمومیت ندارد.

نکته: در نمونه‌گیری از جامعهٔ نرمال کمترین مقدار MSE در میان برآوردکننده‌های به فرم:

$$\hat{\sigma}^2=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{k}$$

برای σ^2 زمانی است که $k=n+1$.

نکته : از جامعه نمایی با میانگین θ، $\frac{n\bar{X}}{n+1}$ دارای کمترین مقدار MSE در میان برآوردکننده‌های به صورت $C\bar{X}$ است.

مثال : در نمونه‌گیری از جامعه نرمال $MSE\left(MLE\left(\sigma^2\right)\right)$ را بیابید.

حل:

می‌دانیم $\frac{n-1}{n}S^2=\frac{\sum(X_i-\bar{X})^2}{n}=\hat{\sigma}^2=MLE\left(\sigma^2\right)$ و پس:

$$V_{an}\left(\hat{\sigma}^2\right)=\frac{(n-1)^2}{n^2}V_{ar}\left(S^2\right)=\frac{(n-1)^2}{n^2}\,\frac{2\sigma^{-4}}{n-1}\qquad,\qquad b\left(\hat{\sigma}^2\right)=\frac{n-1}{n}\sigma^2-\sigma^2=-\frac{1}{n}\sigma^2$$

پس:

$$MSE\left(\hat{\sigma}^2\right)=MSE\left(MLE\left(\sigma s^2\right)\right)=\frac{1}{n^2}\sigma^4+\frac{(n-1)^2}{n^2}\,\frac{2\sigma^4}{n-1}=\frac{2n-1}{n^2}\sigma^4$$

مثال : نمونه **5** تایی از جامعه نمایی با پارامتر λ گرفته‌ایم مطلوب است

$$MSE\left(\frac{1}{15}\sum_{i=1}^{5}iX_i\right)$$

حل:

$$E\left(\frac{1}{15}\sum_{i=1}^{5}iX_i\right)=\frac{1}{15}\sum_{i=1}^{5}iE(X_i)=\frac{\lambda}{15}\sum_{i=1}^{5}i=\frac{\lambda}{15}(15)=\lambda$$ (پس این برآورد ناریب است)

در نتیجه:

$$MSE\left(\frac{1}{15}\sum_{i=1}^{5}iX_i\right)=Var\left(\frac{1}{15}\sum_{i=1}^{5}iX_i\right)=\frac{1}{225}\sum_{i=1}^{5}i^2\,Var(X_i)=\frac{\lambda^2}{225}\sum_{i=1}^{5}i^2=\frac{\lambda^2}{225}=\frac{55\lambda^2}{225}$$

میانگین و واریانس $\bar{X}$

در هر حال $E\left(\bar{X}\right)=\mu$ و اگر جامعه نامتناهی و نمونه‌گیری با جایگذاری $(i.i.d)$ داریم $Var\left(\bar{X}\right)=\frac{\sigma^2}{n}=\frac{Var(X)}{n}$

اگر جامعه متناهی N تایی با وایانس σ^2 و نمونه‌گیری n تایی بدون جایگذاری باشد:

$$Var\left(\bar{X}\right)=\frac{\sigma^2}{n}\,\frac{N-n}{N-1}$$

مثال : در نمونه‌گیری **6** تایی از جامعه‌ای با چگالی $f(x)=kx^3(1-x)^2$ واریانس میانگین نمونه چقدر است؟
$0<x<1$

حل: در نمونه i.i.d

$$Var(\bar{X})=\frac{Var(X)}{n}=\frac{\sigma^2}{n}$$

اما $X\sim B(a=4,b=3)$ پس $Var(\bar{X})=\frac{12}{6(7)^2(8)}=\frac{1}{196}$

مثال : در نمونه‌گیری دوتایی و بدون جایگذاری از جامعه‌ای شامل **6** و **5** و **4** واریانس میانگین نمونه چقدر است؟

حل:

طبق نکته بالا $\sigma^2=\bar{X^2}-\bar{X}^2=\frac{16+25+36}{3}-25=\frac{77-75}{3}=\frac{2}{3}$ اما $Var(\bar{X})=\frac{\sigma^2}{n}\frac{N-n}{N-1}=\frac{\sigma^2}{2}\frac{3-2}{3-1}=\frac{\sigma^2}{4}=\frac{1}{6}$

قضیه حد مرکزی

ـ در قضیه حد مرکزی توزیع $\bar{X}$ از هر جامعه‌ای به طور تقریبی نرمال می‌شود $Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{\bar{X}-E(\bar{X})}{\sqrt{Var(\bar{X})}}$ توزیع حدی نرمال استاندارد دارد.

ضمناً قضیه حد مرکزی در نمونه‌گیری از جامعه نرمال لازم نیست و حکم برای هر n برقرار است.

ـ در نمونه‌گیری از جامعه نرمال:

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

پس توزیع S^2 عبارت است از:

$$S^2\sim G\left(\frac{n-1}{2},\frac{n-1}{2\sigma^2}\right)\ \Rightarrow\ Var(S^2)=\frac{2\sigma^4}{n-1}\ \text{و}\ E(S^2)=\sigma^2$$

اگر μ در جامعه نرمال معلوم باشد به جای S^2 در روابط بالا می‌گذاریم $S^{*2}=\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$ و هر جا $n-1$ بود می‌شود n

مثال : احتمال آن‌که میانگین **100** نقطه که به تصادف از بازه $\left(0,\frac{1}{\sqrt{3}}\right)$ انتخاب شود در فاصله 0.032 از نقطه میانی بازه قرار گیرد تقریباً چقدر است؟

حل: می‌دانیم

$$E(X)=\frac{\frac{1}{\sqrt{3}}+0}{2}=\frac{1}{2\sqrt{3}}$$

$$Var(X)=\frac{\left(\frac{1}{\sqrt{3}}-0\right)^2}{12}=\frac{1}{36}$$

$$P\left(\left|\bar{X}-\frac{1}{2\sqrt{3}}\right|<0.032\right)=P\left(\left|\frac{\bar{X}-\frac{1}{2\sqrt{3}}}{\frac{\frac{1}{6}}{\sqrt{100}}}\right|<\frac{0.032}{\frac{1}{60}}\right)=P\left(|Z|<1.92\right)=2(1.92)-1=0.9452$$

توزیع T:

برای ساختن توزیع T داریم:

$$\left.\begin{array}{l} Z\sim N(0,1) \\ Y\sim \chi^2_{(v)} \\ Y,Z \text{ مستقل} \end{array}\right\} \Rightarrow T=\frac{Z}{\sqrt{\frac{Y}{v}}}\sim t_{(v)}$$

اگر از جامعه نرمال نمونه‌ای n تایی $(n<30)$ بگیریم آنگاه $T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}\sim t_{(n-1)}$ ضمناً برای $V>1$ $E(T)=0$ و برای

$$\mathrm{Var}(T)=\frac{V}{V-2} \qquad V>2$$

درجه آزادی توزیع T را از روی درجه آزادی مربع کای مخرج کسر پیدا می‌کنیم. خصوصاً وقتی نمونه‌های تشکیل‌دهنده $\bar{X}$ و S^2 به ترتیب n و m تایی و مستقل از هم باشند.

$$T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}\sim t_{m-1}$$

ضمناً $t_{(1)}\equiv C(0,1)$ یعنی برای توزیع t با یک درجه آزادی جدول لازم نیست و می‌توان از توزیع کوشی استاندارد انتگرال گرفت.

مثال: اگر $X_1,\dots,X_{10} \underset{\sim}{\text{i.i.d}} N(0,\sigma^2)$ واریانس $W=\frac{X_1-2X_2+X_3}{\sqrt{\sum_{i=4}^{10}X_i^2}}$ چقدر است؟

حل:

$$X_1-2X_2+X_3\sim N\left(0,6\sigma^2\right) \Rightarrow Z=\frac{X_1-2X_2+X_3}{\sqrt{6}\sigma}\sim N(0,1)$$

$$Y=\frac{\sum_{i=4}^{10}\left(\frac{X_i}{\sigma}\right)^2}{7}=X^2_{(7)}\rightarrow T=\frac{Z}{\sqrt{\frac{6}{7}}}=\sqrt{\frac{6}{7}}W\sim t_{(7)}$$

پس:

$$V_{ar}(W)=\frac{6}{5}$$

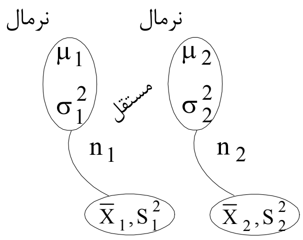
مثال : اگر از جامعه نرمال با میانگین صفر و واریانس σ^2 نمونه دوتایی X_1, X_2 را بگیریم توزیع $Y=\frac{X_1+X_2}{|X_1-X_2|}$ را تعیین کنید.

حل:

$$Y=\frac{X_1+X_2}{|X_1-X_2|}=\frac{\frac{X_1+X_2}{\sqrt{2}}}{\left|\frac{X_1-X_2}{\sqrt{2}}\right|}=\frac{Z}{|Z|}=C(0,1)\equiv T_{(1)}$$

توزیع F :

برای نسبت واریانس‌های نمونه‌ای از دو جامعه مستقل و نرمال توزیع F را به‌کار می‌بریم.



$$\frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}\sim F_{n_1-1,n_2-1}$$

و در حالت کلی:

$$\left.\begin{array}{l} X_1\sim\chi^2_{v_1} \\ X_2\sim\chi^2_{v_2}\end{array}\right\}\Rightarrow \frac{X_1 v_2}{X_2 v_1}\sim F_{v_1,v_2}$$

نکته : اگر $X\sim F_{a,b}$ آن‌گاه $Y=\frac{1}{X}\sim F_{b,a}$

نتیجه: اگر $f_{\alpha,a,b}$ عددی باشد که مساحت زیر منحنی $F_{a,b}$ در سمت راست آن α بشود، آن‌گاه:

$$f_{1-\alpha,b,a}=\frac{1}{f_{\alpha,a,b}}$$

نتیجه: میانه توزیع F با درجه آزادی a و a برابر یک است.

نکته : اگر $X\sim t_v$ پس $Y=X^2\sim F_{1,v}$ (رابطه F با T)

نتیجه (الف) $X\sim t_v \Rightarrow Y=\frac{1}{X^2}\sim F_{v,1}$

اگر $t_{\alpha,v}$ عددی باشد که مساحت زیر منحنی T در سمت راست آن α شود، آن‌گاه $t^2_{\frac{\alpha}{2},v}=f_{\alpha,1,v}$

مثال : اگر $\bar{X}$ و S^2 میانگین و واریانس نمونه‌ای $n=2$ تایی از جامعه نرمال باشد $W=\frac{S}{\sqrt{n}(\bar{X}-\mu)}$ چه توزیعی دارد؟

حل: قرار دهید $W=\frac{1}{T}$ پس:

$$T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{2}}}\sim t_{(1)}\equiv C(0,1)\rightarrow W\sim C(0,1)$$

چون می‌دانیم معکوس کوشی توزیع کوشی دارد.

مثال : فرض کنید **X** توزیع **t** با $V = 8$ درجه آزادی داشته باشد $Y = \frac{1}{X^2}$ چه توزیعی دارد؟

حل:

$$X^2 \sim F_{1,8} \quad \to \quad \frac{1}{X^2} \sim F_{8,1}$$

رابطهٔ گاما با F :

ازآنجاکه:

$$X \sim G(\alpha,\beta) \Rightarrow 2\beta X \sim \chi^2_{(2\alpha)}$$

پس می‌توان نسبت هر دو متغیر i.i.d گاما را به صورت F نشان داد. یعنی اگر $X, Y \overset{i.i.d}{\sim} G(\alpha,\beta)$ پس:

$$\frac{X}{Y} \sim F_{2\alpha,2\alpha}$$

در حالت خاص اگر X و Y به طور مستقل دارای توزیع نمایی با پارامتر یکسان λ باشند

$$\frac{X}{Y} \sim F_{2,2}$$

مثال : در نمونه‌گیری از جامعه نرمال با میانگین صفر اگر $P(4\bar{X} > 2.131S) = 0.025$ ، $F_{0.95,1}$ $F_{0.05,1,15}$ چقدر است؟

حل:

$$P(4\bar{X} > 2.131S) = P\left(\frac{\bar{X}-0}{\frac{S}{\sqrt{16}}} > 2.131\right) = P(T_{15} > 2.131) = 0.025$$

پس $t_{0.025,15} = 2.131$ لذا $f_{0.05,1,15} = (2.131)^2$

فصل ششم: برآوردهای فاصله‌ای و آزمون فرض

فاصله اطمینان برای μ

ردیف	تعداد جامعه	توزیع جامعه	واریانس	تعداد نمونه	پارامتر	برآورد فاصله‌ای دو طرفه
۱	یک	نرمال	معلوم	دلخواه	μ	$\bar{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$
۲	یک	دلخواه	معلوم	بزرگ	μ	$\bar{X} \pm \frac{\sigma}{\sqrt{n}} Z_{\frac{\alpha}{2}}$
۳	یک	نرمال	مجهول	کوچک	μ	$\bar{X} \pm \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}, n-1}$
۴	یک	نرمال	مجهول	بزرگ	μ	$\bar{X} \pm \frac{S}{\sqrt{n}} Z_{\frac{\alpha}{2}}$
۵	دو	نرمال و مستقل	معلوم	دلخواه	$\mu_1 - \mu_2$	$\left(\bar{X}_1 - \bar{X}_2\right) \pm \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\; Z_{\frac{\alpha}{2}}$
۶	دو	نرمال و مستقل	مجهول	هر دو نمونهٔ بزرگ	$\mu_1 - \mu_2$	$\left(\bar{X}_1 - \bar{X}_2\right) \pm \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\; Z_{\frac{\alpha}{2}}$
۷	دو	نرمال و مستقل	مجهول اما برابر	کوچک	$\mu_1 - \mu_2$	$\left(\bar{X}_1 - \bar{X}_2\right) \pm S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\; t_{\frac{\alpha}{2}, n_1 + n_2 - 2}$ واریانس آمیخته $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

نکته : برای یک طرفه نمودن آن‌ها کافی است $Z_{\frac{\alpha}{2}}$ به Z_α تبدیل شود و همان جهت که لازم است بماند مثلاً یک طرفه از پایین برای ردیف ۱ در جدول بالا می‌شود:

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} Z_\alpha \;, +\infty\right)$$

مثال : در نمونه **n** تایی از جامعه نرمال $N\left(\theta, \theta^2\right)$ برآورد فاصله‌ای θ در سطوح $100(1-\alpha)$ ؟؟

حل:

$$Z = \frac{\bar{X} - \theta}{\frac{\theta}{\sqrt{n}}} = \frac{\sqrt{n}\,\bar{X}}{\theta} - \sqrt{n} \quad \to \quad -Z_{\frac{\alpha}{2}} < \frac{\sqrt{n}\,\bar{X}}{\theta} - \sqrt{n} < Z_{\frac{\alpha}{2}}$$

که پس از ساده کردن داریم
$$\frac{\bar{X}}{1 + \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}} < \theta < \frac{\bar{X}}{1 - \frac{Z_{\frac{\alpha}{2}}}{\sqrt{n}}}$$

فاصله اطمینان برای p

برآورد فاصله‌ای در سطح $100(1-\alpha)\%$ برای p :

$$\hat{p} \pm \sqrt{\frac{\hat{p}\hat{q}}{n}}\ Z_{\frac{\alpha}{2}}$$

و برآورد مذکور برای $p_1 - p_2$ عبارت است از:

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}\ Z_{\frac{\alpha}{2}}$$

در حالت کلی می‌توان برای θ برآورد فاصله‌ای تقریبی به صورت $\hat{\theta} \pm \sqrt{g(\hat{\theta})}\, Z_{\frac{\alpha}{2}}$ تشکیل داد که $g(\theta)$ واریانس $\hat{\theta}$ می‌باشد.

مثال : اگر از **100** دانشجو **20** نفر سیگاری باشند برآورد فاصله‌ای **99%** یک طرفه از پایین برای نسبت واقعی دانشجویان سیگاری بسازید.

حل: چون که $\hat{p} = \frac{X}{n} = \frac{20}{100} = 0.2$ پس داریم $\left(\hat{p} - \sqrt{\frac{\hat{p}\hat{q}}{n}} Z\alpha, 1\right)$ یا $\left(0.2 - \sqrt{\frac{(0.2)(0.8)}{100}} 2.33, +\infty\right)$ و یا $\left(0.2 - \frac{0.4}{10} 2.33, +\infty\right)$ لذا داریم $(0.2932, +1)$

مثال : اگر در نمونه‌ای **100** تایی از توزیع پواسون میانگین **4** باشد برآورد فاصله‌ای برای λ در سطح $95\ \%$ و یک طرفه از بالا را بیابید.

حل: چون $\hat{\lambda} - \bar{X} = 4$ و $V_{ar}\left(\bar{X}\right) = \frac{\lambda}{n}$ پس برآورد مذکور می‌شود:

$$\left(0, \bar{X} + \sqrt{\frac{\bar{X}}{n}}\ Z_{\frac{\alpha}{2}}\right)$$

یا:

$$\lambda \in \left(0, 4 + \sqrt{\frac{4}{100}}\ \ 1.96\right) = (0, 0.392)$$

فاصله اطمینان برای واریانس

برای σ^2 $\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}}\ \text{و}\ \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}\right)$ و برای $\frac{\sigma_1^2}{\sigma_2^2}$ $\left(\frac{S_1^2}{S_2^2 f_{1-\frac{\alpha}{2}, n_1-1, n_2-1}}\ ,\ \frac{S_1^2}{S_2^2 f_{\frac{\alpha}{2}, n_1-1, n_2-1}}\right)$

فرمول تعداد نمونه (فاصله اطمینان p)

$$n=\begin{cases}\hat{p}\hat{q}\left(\dfrac{Z_{\frac{\alpha}{2}}}{e}\right)^2 & \text{با نمونهٔ مقدماتی}\\[2ex] \dfrac{1}{4}\left(\dfrac{Z_{\frac{\alpha}{2}}}{e}\right)^2 & \text{بدون نمونهٔ مقدماتی}\end{cases}$$

تعداد نمونهٔ لازم برای آنکه در سطح اطمینان $100(1-\alpha)\%$ حداکثر خطای ناشی از جایگزینی $\hat{p}$ به‌جای p برابر e شود.

مثال : اگر بخواهیم در سطح اطمینان 98% حداکثر خطای ناشی از برآورد نسبت بیکاران کشور 3% باشد به تعداد نمونه لازم چقدر است؟

حل:

$$n=\frac{1}{4}\left(\frac{2.33}{0.03}\right)^2=1509$$

فرمول تعداد نمونه (فاصله اطمینان p_1-p_2)

با نمونه‌گیری‌های مقدماتی: $n=\left(\hat{p}_1\hat{q}_1+\hat{p}_2\hat{q}_2\right)\left(\dfrac{Z_{\frac{\alpha}{2}}}{e}\right)^2$ وقتی $n_1=n_2=n$ داریم و بدون نمونه‌گیری‌های مقدماتی: $n=\dfrac{1}{2}\left(\dfrac{Z_{\frac{\alpha}{2}}}{e}\right)^2$

فرمول تعداد نمونه (فاصله اطمینان μ)

$$n=\left(\frac{\sigma}{e}Z_{\frac{\alpha}{2}}\right)^2$$

مثال : محققی می‌خواهد از جامعه‌ای نرمال با $\sigma^2=25$ نمونه‌ای انتخاب کند که با 95% اطمینان حداکثر خطای برآورد یک شود چه تعداد نمونه لازم دارد.

حل: $n=\left(\dfrac{\sigma}{e}Z_{\frac{\alpha}{2}}\right)^2=\left(\dfrac{5}{1}1.96\right)^2\cong 97$

فرمول تعداد نمونه (فاصله اطمینان $(\mu_1-\mu_2)$)

$$n=\left(\sigma_1^2+\sigma_2^2\right)\left(\frac{Z_{\frac{\alpha}{2}}}{e}\right)^2$$

مثال : اگر بدانیم واریانس‌های دو جامعه طول قد زنان و مردان کشور به ترتیب $\sigma_1^2 = 1.8$ و $\sigma_2^2 = 4.2$ می‌باشد و بخواهیم در سطح اطمینان 95% و با حداکثر خطای 2% تفاضل میانگین‌های دو جامعه را برآورد کنیم و از هر جامعه تعداد مساوی نمونه بگیریم تعداد نمونه هر کدام چقدر است؟

حل: بنا به نکته بالا:

$$n = (1.8+4.2)\left(\frac{1.96}{0.02}\right)^2 = 57624$$

α,β خطاهای آماری و توان آزمون:

هر آزمون دو نوع خطای آماری دارد؛ خطای نوع اول (I)، رد به ناحق H_0 و خطای نوع دوم (II) قبول به ناحق H_0 است.

این خطاها پیشامدهای تصادفی ناشی از انتخاب نمونه‌های تصادفی هستند و می‌توان احتمال آن‌ها را یافت:

$$\alpha = P(\text{خطای نوع } I) = P(H_0 \text{ رد} \mid H_0 \text{ درستی})$$

$$\beta = P(\text{خطای نوع } II) = P(H_0 \text{ قبول} \mid H_0 \text{ نادرستی})$$

هرچند α و β متمم نیستند اما به دلیل مکمل بودن نواحی رد و قبول در حجم نمونهٔ ثابت، کاهش یکی از این دو مقدار موجب افزایش دیگری می‌شود. **ثابت می‌شود همواره:** $\alpha+\beta<1$

تابع توان ــ توان آزمون: در آزمون $\begin{cases} H_0 : \theta \in \omega_0 \\ H_1 : \theta \in \omega_1 \end{cases}$ به طوری که $(\omega_0 \cap \omega_1 = \varphi)$، تابع توان عبارت است از:

$$\Pi(\theta) = P(H_0 \text{ رد}) = \begin{cases} \alpha(\theta) & \theta \in \omega_0 \\ 1-\beta(\theta) & \theta \in \omega_1 \end{cases}$$

مقدار تابع توان به ازای $\theta_1 \in \omega_1$ را **توان آزمون در** θ_1 می‌نامند. در حقیقت توان آزمون به ازای $\theta \in \omega_1$. احتمال رد صحیح H_0 است.

مثال : جعبه‌ای **N** توپ دارد که **4** عدد از آن‌ها سفیدند **2** توپ با جاگذاری بر می‌داریم اگر هر دو سفید بود فرض $H: N=8$ را به نفع $H_1 : N=10$ رد می‌کنیم α و β و توان آزمون چقدر است؟

حل:

$$\alpha = P(\text{هر دو سفید} \mid N=4) = \left(\frac{4}{8}\right)^2 = \frac{1}{4} = 0.25$$

$$\beta = 1-P(\text{هر دو سفید} \mid N=10) = 1-\left(\frac{4}{10}\right)^2 = 0.84 \quad \rightarrow \quad 1-\beta = 0.16 \quad \text{توان}$$

مثال : اگر **X** دارای توزیع نمایی با پارامتر $\lambda>0$ باشد $\left(f(x)=\lambda e^{-\lambda x} \quad x>0\right)$ برای آزمون $H_0 : \lambda=1$ در برابر $H_1 : \lambda=2$ در سطوح α احتمال خطای نوع دوم یعنی β چقدر است؟

حل: جهت ناحیه رد برخلاف آن‌که λ بزرگتر می‌شود باید به صورت $X<2$ باشد چون $E(X)=\frac{1}{\lambda}$ (X با λ رابطه عکس دارد)

حال C را بر حسب α پیدا می‌کنیم.

$$\alpha = P(X<C \mid \lambda=1) = 1-e^{-c} \quad \rightarrow \quad C=-Ln(1-\alpha)$$

حال

$$\beta = P\left(X \geq -Ln(1-\alpha) \mid \lambda = 2\right) = (1-a)^2$$

مثال: اگر نمونه‌ای **n** تایی از جامعه برنولی با پارامتر p بگیریم و $H_0 : p = \frac{1}{3}$ را در برابر $H_1 : p = \frac{2}{3}$ آزمون کنیم و ناحیه رد به صورت $\sum_{i=1}^{10} X_i \geq 9$ باشد توان ازمون چقدر است؟

حل:

$$\text{توان} = 1-\beta = P\left(\sum X_i \geq 9 \middle| P = \frac{2}{3}\right) = \left(\frac{2}{3}\right)^{10} + \binom{10}{9}\left(\frac{2}{3}\right)^9\left(\frac{1}{3}\right) = \frac{2^{11}}{3^9}$$

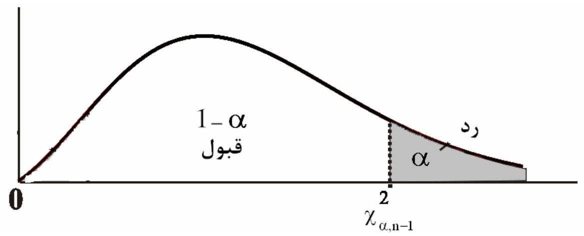
نکته : در مورد انجام آزمون فرض‌ها دو مطلب اساسی وجود دارد.

اول یافتن عبارتی که براساس H_0 توزیع معلوم دارد.

دوم یافتن ناحیه رد که با توجه به H_1 تعیین می‌شود.

به طور کلی ناحیه رد باید دارای مساحتی اندازه α بوده و کوچک باشد ناحیه قبول که متمم رد و بزرگ‌تر است و دارای اندازه $1-\alpha$ است مانند فاصله اطمینان‌ها می‌باشد.

مثلاً اگر $H_0 : \sigma^2 = 4$ در برابر $H_1 : \sigma^2 > 4$ و $n = 11$ و $S = 3$ باید ابتدا $X^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{10(9)}{4} = 22.5$ را به دست آوریم چون می‌دانیم درصورت درستی H_0 داریم $X^2 \sim \chi^2_{(n-1)}$ پس با توجه به جهت H_1 در توزیع مربع کای ناحیه رد از نقطه‌ای به بعد است که مساحت آن α شود:



پس H_0 رد می‌شود اگر $X^2 > \chi^2_{\alpha,n-1} = \chi^2_{0.05,10}$ و چون $22.5 > \chi^2_{0.05,10} = 18.3$ پس H_0 رد می‌شود.

به جداول زیر که نواحی رد و آماره آزمون را تنظیم می‌کند توجه کنید. البته اگر به توزیع‌های نمونه‌ای نرمال و مربع کای و t و F تسلط داشته باشید و روش بالا را به‌کار برد حفظ کردن این فرمول‌هایی که در جداول بعدی می‌آیند ضرورت ندارد.

انواع آزمون برای میانگین جامعه نرمال

ردیف	تعداد جامعه	واریانس	تعداد نمونه	H_0	H_1	آمارهٔ آزمون	ناحیهٔ بحرانی: H_0 رد می‌شود اگر
۱	یک	معلوم	دلخواه	$\mu=\mu_0$	$\mu<\mu_0$ $\mu>\mu_0$ $\mu\neq\mu_0$	$Z=\dfrac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$Z<-z_\alpha$ $Z>z_\alpha$ $\lvert Z\rvert>z_{\frac{\alpha}{2}}$
۲	یک	مجهول	کوچک	$\mu=\mu_0$	$\mu<\mu_0$ $\mu>\mu_0$ $\mu\neq\mu_0$	$T=\dfrac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}$ $v=n-1$	$T<-t_\alpha$ $T>t_\alpha$ $\lvert T\rvert>t_{\frac{\alpha}{2}}$
۳	دو جامعهٔ مستقل	معلوم	دلخواه	$\mu_1-\mu_2=d_0$	$\mu_1-\mu_2<d_0$ $\mu_1-\mu_2>d_0$ $\mu_1-\mu_2\neq d_0$	$Z=\dfrac{\bar{X}_1-\bar{X}_2-d_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$Z<-z_\alpha$ $Z>z_\alpha$ $\lvert Z\rvert>z_{\frac{\alpha}{2}}$
۴	دو جامعهٔ مستقل	مجهول اما برابر	کوچک	$\mu_1-\mu_2=d_0$	$\mu_1-\mu_2<d_0$ $\mu_1-\mu_2>d_0$ $\mu_1-\mu_2\neq d_0$	$T=\dfrac{\bar{X}_1-\bar{X}_2-d_0}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_1}}}$ $S_P^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ $v=n_1+n_2-2$	$T<-t_\alpha$ $T>t_\alpha$ $\lvert T\rvert>t_{\frac{\alpha}{2}}$
۵	دو جامعهٔ وابسته (مشاهدات زوجی)	مجهول	کوچک	$\mu_d=d_0$	$\mu_d<d_0$ $\mu_d>d_0$ $\mu_d\neq d_0$	$T=\dfrac{\bar{d}-d_0}{\frac{S_d}{\sqrt{n}}}$ $v=n-1$	$T<-t_\alpha$ $T>t_\alpha$ $\lvert T\rvert>t_{\frac{\alpha}{2}}$

تذکر: آزمون مشاهدات زوجی غالباً برای مقایسه قبل و بعد از یک رخداد به‌کار می‌رود با این حال به مثال زیر توجه کنید.

مثال : **5** قطعه یکبار با ترازوی **A** و همان **5** قطعه یکبار دیگر با ترازوی **B** وزن شده‌اند مقدار آماره آزمون برای مقایسه میانگین وزنی که دو ترازو نشان می‌دهند پیدا کنید.

A	110	99	112	85	99
B	112	101	113	88	101

حل: با توجه به وابسته بودن مشاهدات به طور زوجی داریم:

$d_i=2 \quad 2 \quad 1 \quad 3 \quad 2$

$$\bar{d}=2\ ,\ S_x^2=\frac{\sum(d_i-2)^2}{n-1}=\frac{1}{2}$$

و لذا $T=\dfrac{2-0}{\dfrac{\sqrt{\frac{1}{2}}}{\sqrt{5}}}=2\sqrt{10}$ یعنی فرض برابری دقت دو ترازو رد می‌شود.

انواع آزمون برای واریانس

ردیف	تعداد جامعه	توزیع جامعه	میانگین	H_0	H_1	آمارۀ آزمون	ناحیۀ بحرانی: H_0 رد می‌شود اگر
۱	یک	نرمال	مجهول	$\sigma^2=\sigma_0^2$	$\sigma^2<\sigma_0^2$ $\sigma^2>\sigma_0^2$ $\sigma^2\neq\sigma_0^2$	$X^2=\dfrac{(n-1)S^2}{\sigma_0^2}$ $S^2=\dfrac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$ $v=n-1$	$X^2<\chi^2_{1-\alpha,n-1}$ $X^2>\chi^2_{\alpha,n-1}$ $X^2<\chi^2_{1-\frac{\alpha}{2},n-1}$ یا $X^2>\chi^2_{\frac{\alpha}{2},n-1}$
۲	یک	نرمال	معلوم	$\sigma^2=\sigma_0^2$	$\sigma^2<\sigma_0^2$ $\sigma^2>\sigma_0^2$ $\sigma^2\neq\sigma_0^2$	$X^2=\dfrac{nS^{*2}}{\sigma_0^2}$ $S^{*2}=\dfrac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$ $v=n$	$X^2<\chi^2_{1-\alpha,n}$ $X^2>\chi^2_{\alpha,n}$ $X^2<\chi^2_{\frac{\alpha}{2},n}$ یا $X^2<\chi^2_{1-\frac{\alpha}{2},n}$
۳	دو جامعۀ مستقل	نرمال	مجهول	$\sigma_1^2=\sigma_2^2$	$\sigma_1^2<\sigma_2^2$ $\sigma_1^2>\sigma_2^2$ $\sigma_1^2\neq\sigma_2^2$	$F=\dfrac{S_1^2}{S_2^2}$ $v_1=n_1-1$, $v_2=n_2-1$	$F<F_{1-\alpha,n_1-1,n_2-1}$ $F>F_{\alpha,n_1-1,n_2-1}$ $F>F_{\frac{\alpha}{2},n_1-1,n_2-1}$ یا $F<F_{1-\frac{\alpha}{2},n_1-1,n_2-1}$

مثال : واریانس نمونه **8** تایی از جامعه نرمال **12** بوده است واریانس نمونه‌ای **6** تایی از جامعه نرمال و مستقل از اولی حداقل چقدر باشد تا فرض $H_0:\sigma_1^2=2\sigma_2^2$ در برابر $H_1:\sigma_1^2<2\sigma_2^2$ در سطح $\alpha=0.01$ رد می‌شود؟

حل: ناحیه رد عبارت است از:

$$\frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}=\frac{S_1^2}{2S_2^2}<f_{1-\alpha,n_1-1,n_2-1} \quad \text{یا} \quad \frac{12}{2S_2^2}<f_{0.99,7,5}=\frac{1}{7.46}$$

در نتیجه:

$$S_2^2>6(7.46)$$

$$S_2^2>44.76$$

انواع آزمون برای نسبت پیروزی‌ها درجامعه برنولی

ردیف	تعداد جامعه	توزیع جامعه	H_0	H_1	آمارهٔ آزمون (تقریبی)	ناحیهٔ بحرانی: H_0 رد می‌شود اگر
۱	یک	برنولی	$p=p_0$	$p<p_0$ $p>p_0$ $p\neq p_0$	$Z=\frac{X-np_0}{\sqrt{np_0q_0}}=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0q_0}{n}}}$ X تعداد پیروزی در نمونه n تایی است.	$Z<-Z_\alpha$ $Z>Z_\alpha$ $\lvert Z\rvert>Z_{\frac{\alpha}{2}}$
۲	دو جامعهٔ مستقل	برنولی	$p_1=p_2$	$p_1<p_2$ $p_1>p_2$ $p_1\neq p_2$	$Z=\frac{\hat{p}_1-\hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ $\hat{p}_1=\frac{X_1}{n_1}\quad \hat{p}_2=\frac{X_2}{n_2}$ $\hat{p}=\frac{X_1+X_2}{n_1+n_2}$	$Z<-Z_\alpha$ $Z>Z_\alpha$ $\lvert Z\rvert>Z_{\frac{\alpha}{2}}$
۳	k جامعهٔ مستقل	برنولی	$p_1=\cdots=p_k=p_0$	حداقل یک جامعه نسبت نابرابر با p_0 دارد.	$X^2=\frac{\sum_{i=1}^{k}\left(X_i-n_ip_0\right)^2}{n_ip_0q_0}$	$X^2>\chi^2_{\alpha,k}$
۴	k جامعهٔ مستقل	برنولی	$p_1=\cdots=p_k$	حداقل دو جامعه نسبت نابرابر دارند.	$X^2=\frac{\sum_{i=1}^{k}\left(X_i-n_i\hat{p}\right)^2}{n_i\hat{p}\hat{q}}$ که $\hat{p}=\frac{\sum_{i=1}^{k}X_i}{\sum_{i=1}^{k}n_i}$	$X^2>\chi^2_{\alpha,k-1}$

مثال : برای آزمون این‌که آیا تماس **18** شهر یک استان نرخ بی‌کاری یکسان و برابر **14%** دارند یا نه چه توزیعی مناسب است.

حل: توزیع مربع کای با 18 درجه آزادی بر اساس پندسوم جدول بالا.

مثال : حسن مدعی است حداکثر **20%** از پنالتی‌ها را گل نمی‌کند از **100** پنالتی حداکثر چند عدد را گل کند که ادعایش در سطح $\alpha=0.05$ رد بشود.

حل: گل نکردن پنالتی را پیروزی می‌گیریم ناحیه رد می‌شود

$H_0: p\leq 0.20$

$H_1: p>0.20$

$$\frac{X-np_0}{\sqrt{np_0g_0}}=\frac{X-20}{\sqrt{16}}>Z_{0.05}=1.645$$ در $X>20+6.580$

$X\geq 27$ پس باید $100-X\leq 73$ عدد را گل کند.

ـ روش P – value

در روش P مقدار ابتدا مشاهده انجام می‌گیرد و سپس P – value به صورت زیر تعریف و محاسبه می‌شود:

$$P-\text{value} = P\left(\text{مشاهده مجدد مشاهده اولیه یا موارد بدتر از آن برای } H_0 \;\middle|\; \text{درستی } H_0\right)$$

و H_0 رد می‌شود اگر $P-\text{value} < \alpha$.

در واقع اگر α معلوم باشد درصورتی‌که $P-\text{value}\in(0,\alpha)$، H_0 رد می‌شود؛ به عبارت دیگر **P - value حداکثر مقدار ممکن α است که H_0 پذیرفته شود.**

جهت نامساوی در محاسبه P – value معمولاً در جهت H_1 می‌باشد و در فرض‌های دو طرفه $(\neq)$ به صورت دو برابر احتمال کمتر محاسبه می‌شود.

مثال : از جامعه‌ای نرمال با انحراف معیار 10 نمونه‌ای 100تایی گرفته‌ایم و میانگین نمونه 12.5 به دست آمده است برای آزمون $H_\circ:\mu=14$ در برابر $H_1:\mu<14$ مقدار احتمال P – value چقدر است؟

حل:

$$P-\text{value} = P\left(\bar{X}\le 12.5 \middle| \mu=14\right) = P\left(Z\le \frac{12.5-14}{\frac{10}{10}}\right) = \phi(-1.5) = 0.0668$$

اگر $H_1:\mu\neq 14$ بود بایستی $P\left(\bar{X}\le 12.5\right)$ و $P\left(\bar{X}\ge 12.5\right)$ را محاسبه و آن یکی را که کوچک‌تر بود دو برابر می‌کردیم.

مثال : از **6** موشک یکی به هدف خورد آیا $H_0:p=\frac{3}{4}$ در برابر $H_1:p<\frac{3}{4}$ در سطح $\alpha=0.027$ رد می‌شود

حل: با روش P – Value داریم:

$$P-\text{Value} = P\left(X\le 1 \middle| p=\frac{3}{4}\right) = P\left(X=0 \middle| p=\frac{3}{4}\right) + P\left(X=1 \middle| p=\frac{3}{4}\right)$$

$$= \left(\frac{1}{4}\right)^6 + \binom{6}{1}\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right) = \frac{1+18}{2048} = \frac{19}{2048} = 0.00928 < \alpha = 0.027$$

پس H_0 رد می‌شود.

تذکر: می‌توان نشان داد ناحیهٔ رد آزمون‌های تواناترین مطرح شدهٔ بالا به صورت $\begin{cases} H_0:\theta=\theta_0 \\ H_1:\theta>\theta_0 \end{cases}$ با ناحیهٔ رد آزمون‌های $\begin{cases} H_0:\theta\le\theta_0 \\ H_1:\theta>\theta_0 \end{cases}$ معادل است پس در **تمام H_0 های ساده به صورت** $H_0:\theta=\theta_0$ **می‌توان** H_0 **مرکب یک‌طرفه مثل $H_0:\theta\le\theta_0$ در نظر گرفت.**

ـ تعداد نمونه n وقتی α و β را داده‌اند و σ^2 معلوم است.

$$n = \frac{\sigma^2\left(Z_\alpha + Z_\beta\right)^2}{\left(\mu_1-\mu_0\right)^2}$$

که $\begin{matrix} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{matrix}$ ضمناً این فرمول برای $H_1 : \mu < \mu_0$ و $H_1 : \mu > \mu_0$ نیز برقرار است.
به علاوه وقتی $H_1 : \mu \neq \mu_0$ تنها تغییر در فرمول Z_α به $Z_{\frac{\alpha}{2}}$ تبدیل می‌شود در همه این موارد وقتی β را می‌دهند باید μ_1 را نیز اعلام کنند. جدول زیر اطلاعات فوق را تنظیم کرده است.

جدول فرمول‌های تعداد نمونه در آزمون‌های میانگین با α و β معلوم

ردیف	H_0	H_1	α	β	واریانس	n
۱	$\mu=\mu_0$	$\mu=\mu_1$	معلوم	معلوم	معلوم	$n=\frac{\sigma^2(Z_\alpha+Z_\beta)^2}{(\mu_0-\mu_1)^2}$
۲	$\mu=\mu_0$	$\mu<\mu_0$	معلوم	به ازای $\mu_1<\mu_0$، β معلوم است	معلوم	$n=\frac{\sigma^2(Z_\alpha+Z_\beta)^2}{(\mu_0-\mu_1)^2}$
۳	$\mu=\mu_0$	$\mu\neq\mu_0$	معلوم	به ازای $\mu_1<\mu_0$، β معلوم است	معلوم	$n=\frac{\sigma^2\left(Z_{\frac{\alpha}{2}}+Z_\beta\right)^2}{(\mu_0-\mu_1)^2}$
۴	$\mu_1-\mu_2=d_0$	$\mu_1-\mu_2<d_0$	معلوم	به ازای $d_1<d_0$، β معلوم است	در دو جامعهٔ معلوم	$n=\frac{(\sigma_1^2+\sigma_2^2)(Z_\alpha+Z_\beta)^2}{(d_0-d_1)^2}$ به طور مساوی در هر جامعه
۵	$\mu_1-\mu_2=d_0$	$\mu_1-\mu_2\neq d_0$	معلوم	به ازای $d_1<d_0$، βمعلوم است	در دو جامعهٔ معلوم	$n=\frac{(\sigma_1^2+\sigma_2^2)\left(Z_{\frac{\alpha}{2}}+Z_\beta\right)^2}{(d_0-d_1)^2}$ به طور مساوی در هر جامعه

مثال : در آزمون $H_0 : \mu = 4$ در برابر $H_1 : \mu \neq 4$ اگر بخواهیم $\alpha = 0.05$ و مقدار β بازاری $\mu = 7$ برابر 0.01 باشد و بدانیم واریانس جامعه برابر **25** است چه مقدار نمونه لازم است؟

حل:

$$n = \frac{\sigma^2\left(Z_{\frac{\alpha}{2}}+Z_\beta\right)^2}{(\mu_1-\mu_0)} = \frac{25(1.96+2.33)^2}{(7-4)^2} = 52$$

آزمون نیکویی برازش (تست انطباق)

در این آزمون، انطباق یک توزیع ادعاشده بر جامعه را تست می‌کنیم. به طور بدیهی این یک روش ناپارامتری است، چون از توزیع جامعه بی‌اطلاع هستیم.

$$\begin{cases} H_0 : \text{جامعه دارای توزیع ادعاشده است} \\ H_1 : \text{جامعه دارای توزیع ادعاشده نیست} \end{cases}$$

بر اساس نمونهٔ n تایی از جامعه، جامعه را به k طبقه تقسیم می‌کنیم. سپس قرار می‌دهیم:

تعداد مشاهده از طبقهٔ i ام : O_i

احتمال آنکه نمونه مشاهده‌شده از طبقهٔ iام باشد طبق فرض H_0 : p_i

تعداد مشاهده مورد انتظار طبق فرض H_0 : $e_i = np_i$

می‌توان نشان داد وقتی تعداد نمونه‌ها بزرگ باشد $X^2 = \sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i}$ دارای توزیع تقریبی مربع کای با $k-1$ درجهٔ آزادی است.

پس H_0 رد می‌شود اگر $X^2 > \chi^2_{\alpha, k-1}$.

وقتی تعداد مشاهدات یک طبقه از 5 کمتر باشد برای دقت بیشتر توصیه می‌شود آن طبقه در طبقات مجاور ادغام و یک واحد از درجهٔ آزادی کسر شود.

در این آزمون اگر لازم باشد تعدادی پارامتر برآورد شوند، باید به تعداد پارامتر برآوردشده از درجهٔ آزادی کاسته شود.

مثال : اگر بخواهیم و بر اساس نمونه‌ای 200 تایی که در 7 دسته تنظیم شده‌اند نرمال بودن جامعه را آزمون کنیم توزیع مناسب کدام است؟

حل: توزیع مربع کای می‌باشد که

$$V = k - t - 1 = 7 - 2 - 1 = 4$$

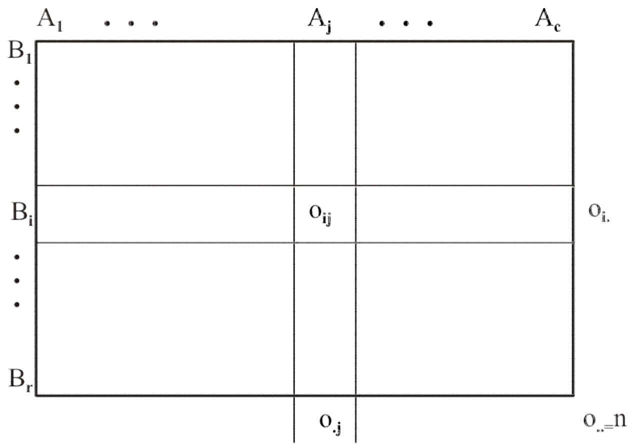
چون توزیع نرمال 2 پارامتر σ, μ دارد و برآورد آن‌ها لازم است.

آزمون استقلال

آزمون زیر **یک روش ناپارامتری برای بررسی استقلال دو متغیر تصادفی که توزیعشان معلوم نیست** ارائه می‌دهد.

$$\begin{cases} H_0 : \text{دو متغیر } X \text{ و } Y \text{ مستقل‌اند} \\ H_1 : \text{دو متغیر } X \text{ و } Y \text{ وابسته‌اند} \end{cases}$$

اگر بر اساس نمونهٔ n تایی جامعه را برحسب X به c طبقه و برحسب Y به r طبقه تقسیم کنیم، جدول زیر را می‌توانیم تشکیل دهیم:



که در این جدول:

تعداد مشاهدات خانه سطر i ستون j : O_{ij}

تعداد مشاهدات سطر i : $\sum_{j=1}^{c} O_{ij} = O_{i\bullet}$

تعداد مشاهدات ستون j : $\sum_{i=1}^{r} O_{ij} = O_{\bullet j}$

تعداد کل مشاهدات : $O_{\bullet\bullet} = n$

با درستی H_0 که سطرها و ستون‌ها را مستقل می‌داند تعداد مشاهدات خانه سطر i و ستون j عبارت است از:

$$e_{ij} = \frac{O_{i\bullet}\, O_{\bullet j}}{n} = n\frac{O_{i\bullet}}{n}\frac{O_{\bullet j}}{n} = n\hat{p}_i\hat{p}_j$$

حال آمارهٔ آزمون را به صورت زیر تعریف می‌کنیم:

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

و H_0 رد می‌شود.

$$X^2 > \chi^2_{\alpha\, ,\ (r-1)(c-1)}$$

مثال : اعداد زیر مربوط به استفاده از وسیله شحصی و وضع جسمانی **200** نفر از ساکنان یک شهر است.

آماره ازمون استقلال دو متغیر را پیدا کنید.

استفاده

وضع جسمانی		می‌کنند	نمی‌کنند	
	لاغر (مناسب)	20 $\boxed{40}$	80 $\boxed{60}$	100
	چاق	60 $\boxed{40}$	40 $\boxed{60}$	100
		80	120	n=200

$$X^2=\frac{(20-40)^2}{40}+\frac{(80-60)^2}{60}+\frac{(60-40)^2}{40}+\frac{(40-60)^2}{60}$$

$$=10+\frac{20}{3}+10+\frac{20}{3}=33.33$$

فصل هفتم: رگرسیون

روابط A و B در روش کم‌ترین مربعات:

ضرایب خط رگرسیون عبارتند از:

$$B=\frac{\sum_{i=1}^{n} x_iY_i-n\bar{x}\bar{Y}}{\sum_{i=1}^{n} x_i^2-n\bar{x}^2}=\frac{S_{xy}}{S_{xx}} \quad \text{و} \quad A=\bar{Y}-B\bar{x}$$

دقت شود عامل خطا در اینجا تفاضل $\hat{Y}$ با مقدار y به ازای هر x مشاهده شده است، یعنی $e_i=y_i-\hat{Y}_i$.

نکته‌ها:

۱ـ همواره $\sum_{i=1}^{n} e_i=0$ در نتیجه: همواره $\sum_{i=1}^{n} Y_i=\sum_{i=1}^{n} \hat{Y}_i$ چون $e_i=Y_i-\hat{Y}_i$

۲ـ همواره $\sum_{i=1}^{n} x_ie_i=0$ در نتیجه: همواره $\sum_{i=1}^{n} x_iY_i=\sum_{i=1}^{n} x_i\hat{Y}_i$

۳ـ وقتی خط رگرسیون X روی Y را بخواهند $B^*=\frac{S_{xy}}{S_{yy}}, A^*=\bar{x}-B^*\bar{Y}$ را به‌کار ببرید.

۴ـ خطوط برآورد رگرسیونی از نقطهٔ $(\bar{x},\bar{Y})$ عبور می‌کند.

۵ـ در حالت خاص که خط از مبدأ می‌گذرد.

$$\hat{\beta}=B=\frac{\sum_{i=1}^{n} x_iY_i}{\sum_{i=1}^{n} x_i^2}$$

۵ـ محاسبه SSE از روی سایر SS ها:

$$SSE=\sum_{i=1}^{n}\left[Y_i-\left(A+Bx_i\right)\right]^2=S_{YY}-BS_{xY}=S_{YY}-B^2S_{xx}=S_{YY}-\frac{S_{xY}^2}{S_{xx}}$$

مثال : براساس مشاهدات

x	-2	-1	0	1	2
y	4	3	0	-1	-1

ضرایب خط رگرسونی را برآورد کنید:

حل: $\sum x_iy_i=-8-3-1-2=-14$ و $\sum x_i^2=10$ و $\bar{y}=1$ و $\bar{x}=0$ و $n=5$ پس:

$$B=\frac{-14-0}{10-0}=-1.4$$

پس خط برآورد رگرسیونی می‌شود: $y=A+BX=1-1.4x$ و $A=1$

مثال : حاصل $\sum e_i\hat{Y}_i$ چقدر است؟

حل: با توجه به تعریف: $\hat{Y}_i = A + Bx_i$ پس $\sum e_i (A + Bx_i) = A\sum e_i + B\sum x_i e_i$

$= A(0) + B(0) = 0$

مسایل ۱ و ۲ صفحه ۵۶۹ را ببینید.

ضریب همبستگی ـ ضریب تعیین و رابطه آن‌ها با B

$$MME(\rho) = MLE(\rho) = r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]\left[\sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2\right]}} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}}$$

پس $r = B\sqrt{\frac{S_{xx}}{S_{YY}}} = B\frac{S_x}{S_y}$ نشان می‌دهد r و B هم‌علامت هستند.

$$BB^* = \frac{S_{xY}}{S_{xx}} \cdot \frac{S_{xY}}{S_{YY}} = \frac{S_{xY}^2}{S_{xx} . S_{YY}} = r^2$$

به $R^2 = r^2(100)$ **ضریب تعیین** می‌گویند. R^2 نشان‌دهندهٔ درصدی از تغییرات Y است که به واسطهٔ داشتن رابطه با X قابل توضیح است.

نکته : R^2 می‌تواند شدت همبستگی‌ها را مقایسه کند.

نکته : فرض کنیم ضریب همبستگی نمونه‌ای مربوط به (X_i, Y_i)ها برابر $r_{X,Y}$ باشد. حال اگر $T_i = aX_i + b$, $W_i = a' Y_i + b'$ ضریب همبستگی نمونه‌ای مربوط به (T_i, W_i) عبارت است از:

$$r_{T,W} = \frac{aa'}{|aa'|} r_{X,Y}$$

مثال : اگر ضریب همبستگی نمونه‌ای -0.4 و انحراف معیار مشاهدات **x** برابر **2** و انحراف معیار مشاهدات **Y** برابر **3** باشد شیب خط رگرسیونی **Y** روی **x** کدام است؟

حل: می‌دانیم

$$B = r\frac{Sy}{Sx} = -0.4\,\frac{3}{2} = -.6$$

مثال : اگر $r_{X,Y} = 0.4$ و $r_{T,W} = 0.8$ رابطه **T,W** چند برابر قوی‌تر از رابطه X و Y است؟

حل: با توجه به آنکه $4 = \frac{0.64}{0.16} = \frac{R_{T,W}^2}{R_{X,Y}^2}$ پس رابطه W و T، 4 برابر قوی‌تر از رابطه X و Y است.

مثال : اگر ضریب همبستگی T_1 و T_2 برابر -0.3 باشد ضریب همبستگی $X = 2T_1 - 1$ و $Y = 3.2T_2$ چقدر است؟

حل: بنا به نکته بالا

$$r_{x,y} = +0.32$$

آزمون‌های مربوط به A و B و r در جامعه نرمال

نکته : B , A برای β, α نااریب هستند

$$B \sim N\left(\beta , \frac{\sigma^2}{S_{xx}}\right) \qquad \text{و} \qquad A \sim N\left(\alpha , \frac{\sum_{i=1}^{n} x_i^2}{n S_{xx}} \sigma^2\right)$$

نکته : در مدل رگرسیونی نرمال بر اساس نمونهٔ n تایی $\frac{SSE}{\sigma^2} \sim \chi^2_{(n-2)}$.

به‌علاوه SSE از A و از B مستقل است. بر این اساس $\hat{\sigma}^2 = \frac{SSE}{n-1}$ **برآورد نااریب برای** σ^2 **است.**

توزیع نمونه‌ای تقریبی r

$$\frac{1}{2} Ln \frac{1+r}{1-r} \sim N\left(\frac{1}{2} Ln \frac{1+\rho}{1-\rho} , \frac{1}{n-3}\right)$$

حال با توجه به اینکه $tgh^{-1} x = \frac{1}{2} Ln \frac{1+x}{1-x}$ می‌توان نتیجه گرفت:

$$Z = \frac{tgh^{-1} r - tgh^{-1} \rho}{\frac{1}{\sqrt{n-3}}} \overset{n \to \infty}{\sim} N(0,1)$$

نکته : اگر بخواهیم صفر بودن یا نبودن ضریب همبستگی جامعه را تست کنیم، به آزمون مربوط **آزمون معنی داری** ρ گویند.

$$\begin{cases} H_0 : \rho = 0 & (\text{معنی‌دار نیست } \rho) \\ H_1 : \rho \neq 0 & (\text{معنی‌دار است } \rho) \end{cases}$$

برای این آزمون در نمونه‌های کوچک می‌توان از توزیع t – استیودنت استفاده کرد. می‌توان نشان داد اگر جامعه نرمال باشد، H_0 رد می‌شود اگر:

$$|T| = \left| \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right| > t_{\frac{\alpha}{2}, n-2}$$

مثال : اگر ضریب همبستگی نمونه‌ای **18** تایی **0.6 –** شود آماره آزمون معنی‌دار بودن ρ چقدر است؟

حل: بنا به نکته بالا

$$T = \frac{-0.6\sqrt{18-2}}{\sqrt{1-0.36}} = -3$$

مثال : اگر **r** ضریب همبستگی نمونه‌ای **53** تایی از جامعه نرمال باشد $Var\left(Ln \frac{1+r}{1-r}\right)$ تقریباً چقدر است؟

حل: $Var\left(\frac{1}{2} Ln \frac{1+r}{1-r}\right) \approx \frac{1}{n-3}$ پس $Var\left(Ln \frac{1+r}{1-r}\right) = \frac{4}{n-3} = \frac{4}{50} = 0.08$

فصل هشتم: آنالیز واریانس

نکته ۱: مشاهدات را می‌دهند و آماره آزمون یا یکی از اجزاء آن را می‌خواهند که فرمول‌های زیر راهگشا می‌باشد.

	μ_1, σ^2	μ_2, σ^2	$\cdots$	μ_k, σ^2
نمونه‌ها	X_{11}	X_{12}	$\cdots$	X_{1k}
	X_{21}	X_{22}	$\cdots$	X_{2k}
	$\vdots$	$\vdots$		$\vdots$
	X_{n1}	X_{n2}	$\cdots$	X_{nk}
جمع	$X_{\cdot 1}$	$X_{\cdot 2}$	$\cdots$	$X_{\cdot k}$
میانگین	$\bar{X}_{\cdot 1}$	$\bar{X}_{\cdot 2}$	$\cdots$	$\bar{X}_{\cdot k}$

ANOVA

S.V	S.S	D.F	M.S	T.S	C.R
منابع تغییر	مجموع مربعات	درجهٔ آزادی	میانگین مربعات	آمارهٔ آزمون	ناحیهٔ رد (بحرانی)
Tr تیمار	SST_r	$k-1$	$MST_r=\frac{SST_r}{k-1}$	$F=\frac{MST_r}{MSE}$	H_0 رد می‌شود اگر $F>f_{\alpha,k-1,N-k}$
E خطا	SSE	$N-k$	$MSE=\frac{SSE}{N-k}$		
جمع	SS	$N-1$	-	-	-

ضریب تصحیح $Cf=\frac{X_{..}^2}{kn}$ و $SS=\sum_{j=1}^{k}\sum_{i=1}^{n}X_{ij}^2-Cf$ و $SST_r=\sum_{j=1}^{k}\frac{X_{.j}^2}{n}-Cf$ و $SSE=SS-SST_r$

ضمناً توجه کنید MSE همواره برای σ^2 نااریب است اما MST_r وقتی H_0 درست باشد برای σ^2 نااریب است.

مثال: اعداد زیر مربوط به کارکرد کالا با **4** نوع ماده اولیه است.

$A \rightarrow$	B	C	D
4	3	2	5
6	3	2	5
	4	6	

جدول آنالیز واریانس را رسم کنید:

حل: می‌دانیم $X_{..}\equiv 10+10+10+10=40$ و $N=2+3+3+2=10$ پس $Cf=\frac{X_{..}^2}{N}=160$ و:

$$SS=\left[4^2+6^2+3^2+3^2+4^2+2^2+2^2+6^2+5^2+5^2\right]-Cf=180-160=20$$

$$SST_r=\left(\frac{10^2}{2}+\frac{10^2}{3}+\frac{10^2}{3}+\frac{10^2}{2}\right)-Cf=166.6-160=6.6 \Rightarrow SSE=20-6.6=13.4$$

ANOVA

S.V	S.S	d-f	M . S	T . S	C , R
تیمار	6.6	3	2.2	$F=1$	$1 \le f_{0.05\,,\,3.6}$
خطا	13.4	6	2.23		پس H_0 رد نمی‌شود و اختلاف میانگین‌ها معنی‌دار نیست
کل	20	9	—		

نکته ۲: نوع دوم تعداد نمونه‌ها و میانگین نمونه‌ها و واریانس نمونه‌ها را می‌دهند و آماره آزمون با یکی از اجزاء آن را می‌خواهند که فرمول‌های زیر راهگشا می‌باشد:

فرض شود برای k جامعه به ترتیب به تعداد $n_1,\cdots,n_k$ نمونه گرفته‌ایم که میانگین‌های نمونه‌ای به ترتیب $\bar{X}_1,\cdots,\bar{X}_k$ و واریانس‌های نمونه‌ای به ترتیب $S_1^2,\cdots,S_k^2$ هستند.

یعنی جدول اطلاعات زیر را داریم:

تعداد نمونه n_j	n_1	. . .	n_k
میانگین نمونه $\bar{X}_{\bullet j}$	$\bar{X}_{\bullet 1}$	. . .	$\bar{X}_{\bullet k}$
واریانس نمونه S_j^2	S_1^2	. . .	S_k^2

$$\bar{X}_{\bullet\bullet}=\frac{\sum_{j=1}^{k} n_j\bar{X}_{\bullet j}}{\sum_{j=1}^{k} n_j}$$ میانگین کل و $$SST_r=\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\bar{X}_{\bullet j}-\bar{X}_{\bullet\bullet}\right)^2=\sum_{j=1}^{k} n_j\left(\bar{X}_{\bullet j}-\bar{X}_{\bullet\bullet}\right)^2$$ و $$SSE=\sum_{j=1}^{k}\sum_{i=1}^{n}\left(X_{ij}-\bar{X}_{\bullet j}\right)^2=\sum_{j=1}^{k}\left(n_j-1\right)S_j^2$$

$N=\sum_{j=1}^{k} n_J$ و $d.f_{SST_r}=k-1$ و $d.f_{SS}=N-10$ و $d.f_{SSE}=N-K$

مثال : اگر اطلاعات زیر موجود باشد آماره آزمون آنالیز واریانس را پیدا کنید.

n_i	4	3	8	5
$\bar{X}_i$	4	1	2	5
S_i^2	1	1	2	3

حل:

$$\bar{X}_{\bullet\bullet}=\frac{4(4)+3(1)+8(2)+5(5)}{4+3+8+5}=3$$

$$SST_r=\sum n_i\left(\bar{X}_i-\bar{X}_{\bullet\bullet}\right)^2=4(4-3)^2+3(1-3)^2+8(2-3)^2+5(5-3)^2=44$$

$$SSE=\sum\left(n_i-1\right)S_i^2=3(1)+2(1)+7(2)+4(3)=31 \quad , \quad N=20 \quad , \quad K=4$$

$$F=\frac{\frac{SST_r}{k-1}}{\frac{SSE}{N-K}}=\frac{\frac{44}{3}}{\frac{31}{16}}=7.57$$

فصل نهم: آمار توصیفی

میانگین هندسی و نکته‌های آن

۱ـ به طور کلی اگر $X_1,\dots,X_n$ مقادیر مثبت و متوالی باشند، چون G برای نسبت‌ها محاسبه می‌شود به‌کمک فرمول زیر پیدا می‌شود:

$$G = \sqrt[n-1]{\frac{X_n}{X_1}} = \sqrt[n-1]{\frac{X_2}{X_1}\frac{X_3}{X_2}\cdots\frac{X_n}{X_{n-1}}}$$

۲ـ اگر مشاهدات به صورت درصد رشد مطرح شوند، برای محاسبهٔ میانگین هندسی باید درصدها را ابتدا بر صد تقسیم کنیم سپس یک واحد به آن‌ها اضافه کنیم و در نهایت میانگین هندسی بگیریم. یعنی اگر $a_1,\cdots,a_n$ درصدهای رشد باشند و بخواهیم میانگین بگیریم باید از فرمول زیر استفاده کنیم:

$$G = \sqrt[n]{\prod_{i=1}^{n}\left(\frac{a_i}{100}+1\right)}$$

نامساوی $H \le G \le \bar{X} \le Q_2 \le Q_r$ که $r \ge 2$ بین میانگین‌ها برقرار است.

مثال : اگر تولید یک کارخانه در **5** سال بر حسب صد هزار تن به صورت زیر باشد متوسط رشد سالانه چقدر است؟

سال	۸۳	۸۴	۸۵	۸۶	۸۷
تولید	۶	۱۰	۸	۱۴	۱۸

حل: بر اساس نکته بالا

$G = \sqrt[4]{\frac{18}{6}} = \sqrt[4]{3} = 1.31 \rightarrow G-1 = 0.31$ متوسط رشد سالانه $\rightarrow$ %31 درصد متوسط رشد سالانه

مثال : اگر درصدهای رشد سالانه تورم در **2** سال متوالی 18%و 4%− بوده باشد متوسط رشد سالانه تورم در این دو سال چند درصد است؟

حل: بنا به نکته بالا

$G = \sqrt{(1.18)(0.96)} = 1.06 \rightarrow G-1 = 0.06$ متوسط رشد سالانه $\rightarrow$ %6 درصد رشد سالانه

روش یافتن میانه برای داده‌های دسته‌بندی‌شده

دستهٔ شامل میانه، به نخستین دسته‌ای گفته می‌شود که فراوانی تجمعی آن از نصف تعداد داده‌ها بیشتر شود؛ پس میانه را از فرمول زیر حساب می‌کنیم:

$$M_e = L_m + \frac{\left(\frac{n}{2} - F_{m-1}\right)}{f_m} w_m$$

که در آن L_m، کران پایین دسته شامل میانه و F_{m-1}، فراوانی تجمعی دسته قبل از دسته شامل میانه و f_m، فراوانی مطلق دسته شامل میانه و w_m طول دسته میانه است.

مثال : در جدول زیر میانه را پیدا کنید.

حل:

دسته‌ها	f_i	F_i
$2-5$	8	8
$5-10$	12	20
$10-20$	10	30
شامل میانه $20-25$	70	100

حل:

$$n=100 \to \frac{n}{2}=50$$

$$M_e=20+\frac{50-30}{70}=21.42$$

روش یافتن P_k (صدک kام) در داده‌های دسته‌بندی‌شده

ابتدا دسته شامل P_k را تعیین می‌کنیم. دسته شامل P_k به نخستین دسته‌ای گویند که فراوانی تجمعی آن از $\frac{nk}{100}$ بیشتر شود پس:

$$P_k=L_k+\frac{\left(\frac{nk}{100}-F_{k-1}\right)}{f_k}w_k$$

که در آن:

L_k،کران پایین دسته شامل P_k و F_{k-1}، فراوانی تجمعی دسته قبل از دسته شامل P_k و f_k، فراوانی مطلق دسته شامل P_k و w_k، طول دسته شامل P_k است.

تذکر: اگر k مضرب 10 باشد اعداد حاصل را **دهک** و اگر مضرب 25 باشد اعداد حاصل را **چارک** گویند و از این‌رو به میانه "صدک پنجاهم" و "دهک پنجم " و " چارک دوم" نیز گفته می‌شود.

روش یافتن مد در داده‌های دسته‌بندی‌شده

ابتدا دستهٔ مددار، یعنی دسته‌ای را که فراوانی مطلق ماکزیمم دارد، تعیین می‌کنیم. سپس با فرمول:

$$Mo=L_m+\frac{d_1}{d_1+d_2}W_m$$

مد را حساب می‌کنیم که در آن:

L_M کران پایین دستهٔ مددار و d_1 اختلاف فراوانی مطلق دستهٔ مددار از دسته قبلی $\left(d_1=f_m-f_{m-1}\right)$ و d_2 اختلاف فراوانی مطلق دستهٔ مددار از دستهٔ بعدی $\left(d_2=f_m-f_{m+1}\right)$ و W_m طول دستهٔ مددار است.

مثال : در جدول زیر مد را بیابید.

دسته‌ها	f_i
2-5	8
5-10	12
10-20	10
20-25	70

شامل مد

حل:

$$Mo=20+\frac{60}{60+70}5=22.3$$

مقایسه پارامترهای مرکزی

میانگین

۱ـ ازآنجاکه $\bar{X}$ همه داده‌ها را مشارکت می‌دهد اعتبار بیشتری از M_e و M_o دارد.

۲ـ مرکز ثقل داده‌هاست چون مجموع انحراف‌ها از میانگین همواره صفر است. یعنی $\sum_{i=1}^{n}\left(X_i-\bar{X}\right)=0$

۳ـ در داده‌های دسته‌بندی‌شده اگر ابتدا یا انتهای داده‌ها معین نباشد میانگین قابل محاسبه نیست.

۴ـ از داده‌های پرت خیلی تأثیر می‌گیرد.

۵ـ همواره وجود دارد و منحصربه‌فرد است.

میانه

۱ـ فقط چند داده را مورد توجه قرار می‌دهد و به بقیه کاری ندارد.

۲ـ همواره وجود دارد و منحصربه‌فرد است.

۳ـ تحت تأثیر داده‌های پرت نیست.

مد

۱ـ از تمام داده‌ها استفاده نمی‌کند.

۲ـ ممکن است وجود نداشته باشد و یا منحصربه‌فرد نباشد.

۳ـ تحت تأثیر داده‌های پرت نیست.

۴ ـ برای داده‌های کیفی به‌کار می‌رود.

پارامترهای پراکندگی

میان‌دامنهٔ چارکی (IQR)

به تفاضل چارک اول از چارک سوم **میان‌دامنهٔ چارکی** می‌گویند و آن را با IQR نشان می‌دهند؛ یعنی $IQR=Q_3-Q_1$. این معیار پراکندگی از دامنه تغییرات بهتر است چون تأثیر داده‌های پرت بر آن کمتر است. از روی میان‌دامنهٔ چارکی می‌توان انحراف چارکی را حساب کرد؛ به $SIQR=\frac{Q_3-Q_1}{2}$ **انحراف چارکی** می‌گویند.

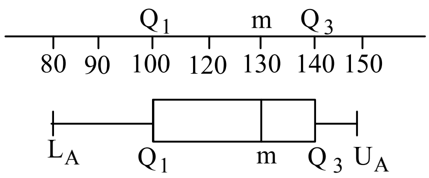
ضریب تغییرات

به دلیل وابسته بودن مقدار σ^2 به واحد اندازه‌گیری، واریانس به تنهایی نمی‌تواند گویای پراکندگی باشد. مثلاً واریانس یک سری عدد بر حسب کیلومتر به مراتب از واریانس همان اعداد بر حسب متر بزرگ‌تر است؛ پس بهتر است معیاری ارائه شود که به طور مطلق یعنی **بدون واحد اندازه‌گیری** پراکندگی را توصیف کند. ضریب تغییرات عبارت است از:

$$C.V=\frac{\sigma}{\bar{X}}$$

C.V واحد ندارد چون از تقسیم دو کمیت با واحد یکسان پدید آمده است.

مثال : اگر نمودار جعبه‌ای داده‌ها به شکل زیر باشد دامنه چارک‌ها **(IQR)** کدام است؟ (سیستم ۸۸)



۱) 30 ۲) 50

۳) 40 ۴) 60

حل: گزینه ۳ درست است.

در نمودار جعبه‌ای ۵ نقطه مهم وجود دارد کران پایین L_A چارک اول Q_1 میانه $M_e = m$ چارک سوم Q_3 و کران بالا U_A که در شکل پیداست ضمناً $IQR = Q_3 - Q_1$ پس جواب می‌شود. $IQR = 140 - 100 = 40$

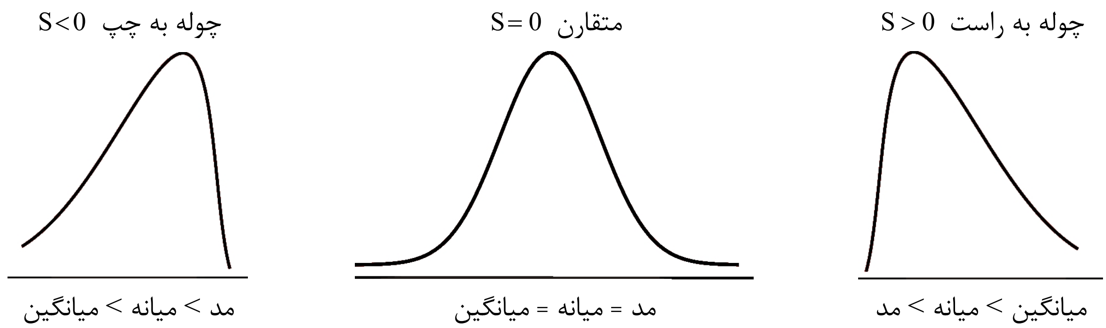
ضریب چولگی

برای داده‌های خام $X_1, \dots, X_n$ ضریب چولگی گشتاوری عبارت است از:

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

که $\mu_3 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$ (گشتاور مرکزی مرتبه سوم) و $\sigma^3 = (\sigma^2)^{\frac{3}{2}} = \left(\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}\right)^{\frac{3}{2}}$ است.

در مورد نمودارهای پیوسته تک‌مدی معمولاً یکی از سه حالت زیر اتفاق می‌افتد:



پس در این حالت وقتی $S > 0$ نمودار چوله به راست است و $Mo < M_e < \bar{X}$ و وقتی $S < 0$ نمودار چوله به چپ است و در این حالت $Mo > M_e > \bar{X}$.

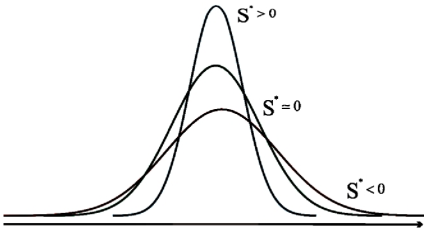
ضریب کشیدگی

برای داده‌های خام $X_1,\dots,X_n$ ضریب کشیدگی عبارت است از:

$$K=\frac{\mu_4}{\sigma^4}$$

$$\sigma^4=\left(\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}{2}\right)^2 \quad \text{و} \quad \mu_4=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^4}{n}$$ گشتاور مرکزی مرتبه چهارم هستند

معمولاً به‌جای K از $K^*=K-3$ که 3 مقدار K برای نمودار نرمال استاندارد است استفاده می‌شود. این ضریب از نمودارهای متقارن تک‌مدی نشان‌دهندهٔ بالاتر یا پایین‌تر بودن منحنی نسبت به منحنی نرمال استاندارد است. اگر $|K^*|\le 0.1$، کشیدگی نمودار تقریباً هم‌اندازه نرمال می‌شود و اگر $|K^*|>0.5$، نمودار با نرمال اختلاف کشیدگی فاحشی خواهد داشت.



تبدیل خطی بر داده‌ها

اگرپارامترهای داده‌های X_i را داشته باشیم و تبدیل خطی $Y_i=aX_i+b$ را به‌کار بریم پارامترها براساس جدول زیر تغییر می‌کند.

پارامتر	X_i	$Y_i=aX_i+b$
میانگین	$\bar{X}$	$\bar{Y}=a\bar{X}+b$
میانه	$\tilde{X}$	$\tilde{Y}=a\tilde{X}+b$
مد	$\hat{X}$	$\hat{Y}=a\hat{X}+b$
واریانس	σ_x^2	$\sigma_Y^2=a^2\sigma_x^2$
انحراف معیار	σ_x	$\sigma_Y=\lvert a\rvert\sigma_x$
ضریب تغییرات	$C.V_x$	$C.V_Y=\frac{\bar{X}\lvert a\rvert C.V_x}{a\bar{X}+b}$
ضریب چولگی	S_x	$S_Y=\frac{a}{\lvert a\rvert}S_x$
ضریب کشیدگی	K_x	$K_Y=K_x$

بخش دوم

توصیه‌های کلیدی

فصل اول: آنالیز ترکیبی

در مجموع، احتمال آمدن سؤال مستقیم از این فصل زیاد نیست و مهم‌ترین بحث این فصل جایگشت تکراری و استفاده از فرمول $\binom{r+n-1}{n-1}$ و کاربردهای آن است.

جدول توزیع موضوعات فصل ۱ در سؤالات ۵ سال اخیر

سال / رشته	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	-	-	-	-	بند VI مدل توپ و جعبه	$n(A\cup B)$
سیستم	ترکیب	-	-	-	-	-
	جایگشت تکراری					-
ریاضی	-	-	-	-	-	-

فصل دوم: اصول احتمال و احتمال شرطی

از این فصل به طور متوسط ۲ یا ۳ سؤال می‌آید که مهم‌ترین مطالب این فصل احتمال شرطی و فرمول بیز ـ استقلال پیشامدها و کران‌های احتمال می‌باشد. مطالعه مسایل معروف احتمال مثل مسئله جورها ـ روز تولد ـ کفش‌ها مفید می‌باشد.

جدول توزیع موضوعات فصل ۲ در سؤالات ۵ سال اخیر

سال / رشته	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	فرمول بیز	استقلال مشروط	استقلال	استقلال	احتمال یکنواخت	استقلال
	فرمول بیز	احتمال یکنواخت	استقلال و تفاضل متقارن	احتمال یکنواخت	احتمال و تفاضل متقارن	استقلال
		قاعده پواسون	فرمول بیز	استقلال مشروط		مشروط کردن
		قانون ضرب احتمال	فرمول بیز			
		قانون احتمال کل				
سیستم	قانون ضرب	نامساوی بن فرونی	استقلال	احتمال یکنواخت	احتمال یکنواخت	فرمول بیز
	احتمال شرطی	احتمال شرطی	استقلال و تفاضل متقارن	فرمول بینر	نامساوی بن فرونی	استقلال
	قانون احتمال کل	احتمال شرطی	احتمال و تفاضل متقارن	تفاضل متقارن	نامساوی بن فرونی	
	فرمول بیز	احتمال یکنواخت	احتمال شرطی		فرمول بیز	
ریاضی			فرمول بیز			احتمال یکنواخت
	تفاضل متقارن	احتمال یکنواخت	فرمول بیز	احتمال یکنواخت	تفاضل متقارن	استقلال
	احتمال شرطی	فرمول بیز	استقلال و احتمال A قبل از B	احتمال یکنواخت	فرمول بیز	قانون احتمال کل
		احتمال یکنواخت			احتمال یکنواخت	احتمال یکنواخت
					احتمال یکنواخت	فرمول بیز

فصل سوم: متغیرهای تصادفی و امید ریاضی

از این فصل به طور متوسط ۳ یا ۴ سؤال می‌آید که مهم‌ترین مطالب این فصل محاسبه امید ریاضی ـ نامساوی چه‌بی‌شف و مارکوف ـ خواص کوواریانس و امید ریاضی مضاعف و واریانس شرطی می‌باشد. محاسبه احتمال توأم از روی تابع چگالی توأم و مطالعه خواص تابع توزیع مفید می‌باشد.

جدول توزیع موضوعات فصل ۳ در سؤالات ۵ سال اخیر

رشته / سال	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	امید ریاضی	نرخ خرابی	تابع توزیع	کوواریانس	امید ریاضی	-
	ضریب همبستگی	امید شرطی	میانه و مد	واریانس شرطی	امید ریاضی	-
		ماکزیمم و مینیمم			واریانس شرطی	-
		تابع توزیع			نامساوی چه‌بی‌شف	-
		احتمال توأم			احتمال توأم	-
		امید ریاضی توأم			احتمال توأم	-
سیستم	امید ریاضی	استقلال متغیرها	تابع احتمال		نامساوی مارکوف	نامساوی مارکوف
	ضریب همبستگی	نرخ خرابی	مد و میانه		احتمال توأم	کوواریانس
		تابع احتمالی				
ریاضی	واریانس	نامساوی مارکوف	احتمال توأم	استقلال متغیرها	تابع احتمال	تابع احتمال
	امید ریاضی	کوواریانس	میانه و مد	واریانس XY	واریانس مجموع	تابع احتمال
		واریانس مجموع	چگالی شرطی	نامساوی چه‌بی‌شف	امید توأم	نامساوی مارکوف
		امید شرطی	احتمال توأم		واریانس مجموع	احتمال توام
		امید ریاضی	احتمال یک متغیره			امید توام
		واریانس XY	احتمال توأم			استقلال
			واریانس			
			امید شرطی			
			نامساوی مارکوف			

فصل چهارم : توزیع‌ها

از این فصل به طور متوسط ۳ تا ۵ سؤال می‌آید که توزیع‌های مهم‌تر نمایی ـ نرمال ـ دوجمله ـ پواسون و هندسی می‌باشند. رابطه‌های پواسون مشروط با دوجمله‌ای ـ هندسی مشروط با یکنواخت و دوجمله‌ای مشروط با فوق هندسی مهم هستند. $P(X=Y)$ در توزیع‌های هندسی و دوجمله مهم هستند. مینیمم هندسی و نمایی‌ها و برنولی‌ها مهم هستند. دانستن محتوای جدول خلاصه توزیع‌ها ضروری است.

جدول توزیع موضوعات فصل ۴ در سؤالات ۵ سال اخیر

سال / رشته	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	رابطه پواسون با گاما	دوجمله‌ای منفی	پواسون	جمع پواسون‌ها	$P(X=Y)$ دوجمله‌ای	ماکزیموم و مینیموم برنولی
	$P(X=Y)$ دوجمله‌ای	فاقد حافظه نمایی	تقریب دوجمله‌ای به پواسون	دوجمله‌ای	گاما	$P(X=Y)$ دوجمله‌ای مینیموم
	رابطه مربع کای با نمایی	امید ریاضی نرمال	نمایی	امید نمایی بریده	تعداد شکست (دوجمله‌ای)	مینیموم نمایی
	ضریب همبستگی پیروزی و شکست		بتا (یکنواخت)	$P(X=Y)$ دوجمله‌ای		پواسون
	قضیه حد مرکزی		دوجمله‌ای ترکیبی	شرطی گاما به پواسون		رابطه نمایی و پواسون
	ضریب همبستگی پیروزی و شکست		امید مینیمم و ماکزیمم			تقریب فوق هندسی به دوجمله‌ای
	گشتاورهای نرمال		دوجمله‌ای (برنولی)			تابع مولد احتمال
						تابع احتمال شرطی و توام
						خواص توزیع گاما
						امید بتا
						دوجمله‌ای
سیستم	پواسون	نرمال	امید نمایی	نرمال و مربع کای	احتمال شرطی گاما	دوجمله‌ای
	پواسون	یکنواخت پیوسته	توزیع $F(X)$	یکنواخت بریده	برنولی شرطی	تقریب فوق هندسی به دوجمله‌ای
	دوجمله‌ای	رابطه والد	دوجمله‌ای مشروط و فوق هندسی		بتا	تابع مولد پواسون
	امید یکنواخت گسسته	مجموع نمایی‌ها	امید هندسی		احتمال در نمایی	احتمال زوج بودن دوجمله‌ای
	امید دوجمله‌ای	یکنواخت پیوسته			پواسون بریده	دوجمله‌ای
	پواسون	نرمال				
	برنولی					
	مولد گشتاور نرمال					
	هندسی					

رشته / سال	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
ریاضی	پواسون	نرمال	هندسی	دوجمله‌ای	گاما و مربع کای	یکنواخت
	توزیع $F(X)$	مولد گشتاور نرمال	دوجمله‌ای مشروط و فوق هندسی	پواسون	مولد گشتاور نمایی	امید مضاعف
	یکنواخت	مینیمم هندسی‌ها	پواسون	هندسی	$P(X=Y)$ هندسی	فرمول واریانس شرطی
	ضریب همبستگی پیروزی و شکست	مینیمم نمایی‌ها	قضیه حد مرکزی	مجموع هندسی	$P(X<Y)$ نمایی	توزیع $F(X)$
	رابطه پواسون و نمایی	فاقد حافظه نمایی	دوجمله‌ای منفی	یکنواخت	هندسی مشروط و یکنواخت گسسته	
	قضیه حد مرکزی	هندسی (شکست)	تقریب دوجمله پواسون	نمایی		
		یکنواخت	دوجمله‌ای	نرمال		
		برنولی		نرمال و دوجمله‌ای		
		پواسون		پواسون		
		نمایی		نرمال		

فصل پنجم: برآوردکننده‌ها و توزیع‌های نمونه‌ای

از این فصل به طور متوسط ۲ یا ۳ سؤال می‌آید که معمولاً از MME و MLE می‌باشد. جدول MMEها و MLEهای معروف و خواص این دو مهم می‌باشند. دانستن واریانس $\overline{X}$ در دو حالت ضروری است. روابط توزیع‌های T و F مفید می‌باشند.

جدول توزیع موضوعات فصل ۵ در سؤالات ۵ سال اخیر

رشته / سال	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	توزیع T	MLE	ناریبی	قضیه حد مرکزی	MLE هندسیِ شکست	
	واریانس $\bar{X}(\text{iid})$	MLE	MME دوجمله‌ای	MME فوق هندسی	MLE یکنواخت	
	واریانس $\bar{X}(\text{iid})$	MSE	MLE پواسون	MLE پواسون	-	
	MME	-	-	-	-	
	MME	-	-	-	-	
	ناریبی	-	-	-	-	
	واریانس $\bar{X}$ (بدون جایگذاری)	-	-	-	-	
سیستم	واریانس $\bar{X}$	MLEیکنواخت	واریانس $\hat{p}$	جدول نرمال MLE (پواسون)	واریانس $\bar{X}$ (بدون‌جاگذاری)	واریانس نمونه
	واریانس $\bar{X}$ (i.i.d)	MME	MLE	MLE	MME نمایی	توزیع $\bar{X}$
	MME یکنواخت	MLE	-	-	MLE	حد مرکزی دوجمله‌ای
	MLE نمایی	MLE	توزیع مربع کای	-	-	استاندارد کردن
	MLE برنولی	ناریبی		-	-	
ریاضی	ناریبی	MLE گاما	MLE	MLE نرمال	قضیه حد مرکزی	قضیه حد مرکزی
	MME (بتا)	ناریب	MME	MME	MME هندسی	MME هندسی
	توزیع $\bar{X}$	-	-	MME گاما	MLE نرمال	MLE
	-	-	-	ناریبی	MLE نمایی	ناریبی
	-	-	-	MLE	رابطه توزیع t و کوشی	

فصل ششم: برآوردهای فاصله‌ای و آزمون فرض

از این فصل به طور متوسط ۳ یا ۴ سؤال می‌آید که دانستن و حل کردن چند مثال از $P-\text{Value}$ و β و α و توان و تعداد نمونه براساس فاصله اطمینان و براساس α و β ضروری می‌باشد.

تشکیل ناحیه رد براساس H_1 در توزیع‌های مهم مفید است، یعنی در توزیع‌های نرمال و برنولی جهت رد H_0 همان جهت مطرح شده در H_1 است.

جدول توزیع موضوعات فصل ۶ در سؤالات ۵ سال اخیر

رشته / سال	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	فاصله اطمینان نرمال (θ,θ^2)	فاصله اطمینان نرمال (θ,θ^2)	تعداد نمونه (فاصله اطمینان p)	تعداد نمونه (آزمون با α,β)	P-value (دوجمله‌ای)	فاصلمه اطمینان تقریبی σ
	α,β برنولی	تعداد نمونه (فاصله اطمینان μ)	آزمون p	β (نمایی)	لم نیمن پیرسون	α در توزیع هندسی
	α برنولی	توان آزمون	فاصله اطمینان μ دوطرفه	فاصله اطمینان p	α در آزمون بتا	$P-\text{Value}$ دوجمله‌ای
	آزمون برابری دو میانگین	$P-\text{value}$ دوطرفه	-	نیکویی برازش	-	
	آزمون σ^2	آزمون میانگین ۲ جامعه	-	-	-	
	آزمون	آزمون یک جامعه	-	-	-	
سیستم	رابطه β,α	تعداد نمونه (با β,α)	تعداد نمونه (فاصله اطمینان p)	فاصله اطمینان یک طرفه (یکنواخت)	فاصله اطمینان μ دوطرفه	فاصله اطمینان μ
	آزمون t	$P-\text{value}$ (ساده)	$P-\text{value}$ دوطرفه	فاصله اطمینان یک طرفه (نمایی)	$P-\text{value}$ دوجمله‌ای	فاصله اطمینان $\mu_1-\mu_2$
	α (برنولی)	β (برای p)	فاصله اطمینان μ دوطرفه	فاصله اطمینان تقریبی	-	فاصله اطمینان μ
	آزمون زوجی	لم نیمن پیرسون	فاصله اطمینان نرمال (θ,θ^2)	فاصله اطمینان $\frac{\sigma_1^2}{\sigma_2^2}$	-	آزمون یک‌طرفه μ
	آماره آزمون واریانس	-	تعداد نمونه (با β,α)	آزمون $\theta_1=\theta_2$ نرمال	-	α در توزیع نمایی
	آزمون میانگین	-	آزمون استقلال		-	
ریاضی	α	آزمون μ	α (برنولی)	β یکنواخت	توان (برنولی)	فاصله اطمینان σ
	β برنولی	فاصله اطمینان (یکنواخت)	فاصله اطمینان یک طرفه (نمایی)	α,β دوجمله‌ای	فاصله اطمینان σ^2	α در توزیع فوق هندسی
	-	α (یکنواخت)	-	$P-\text{value}$ نمایی	-	تابع توان در برنولی
	-	-	-	α,β	-	$P-\text{Value}$ نرمال

فصل هفتم: رگرسیون

از این فصل به طور متوسط ۱ یا ۲ سؤال می‌آید که دانستن روابط A و B و r و حل چند مثال از آن‌ها ضروری است، همچنین مطالعه و کار با ضریب تعیین و آزمون معنی‌داری ρ مفید می‌باشد.

جدول توزیع موضوعات فصل ۷ در سؤالات ۵ سال اخیر

سال / رشته	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	B,A	رابطه B با r	$r_{ax+b,cy+d}$	R^2 ضریب تعیین	A (خط از مبدأ)	B,A آزمون ρ
	B	B , A	رابطه r با B			
		فاصله پیشگویی	B (خط از مبدأ)			
			$\hat{\sigma}^2$			
			B			
			R^2 ضریب تعیین			
			SSE			
			B			
سیستم	-	$r_{x,y}$	B	رابطه r با B	-	$r_{x,y}$
		A , B	A	آزمون ρ		A,B
		ضریب تعیین R^2	SSE_r			
ریاضی	-	-	-	-	A , B	A,B

فصل هشتم: آنالیز واریانس

از این فصل به طور متوسط ۱ یا ۲ سؤال می‌آید که از نکته‌های ۱ و ۲ مطرح شده در خلاصه درس‌ها می‌باشد.

جدول توزیع موضوعات فصل ۸ در سؤالات ۵ سال اخیر

سال / رشته	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	F با نکته ۱	F با نکته ۲	F با نکته ۲	F با نکته ۲	F با نکته ۱	-
		$\bar{X}..$ با نکته ۲				-
سیستم	F با نکته ۲	$\bar{X}..$ با نکته ۲	$\bar{X}..$ با نکته ۲	رابطه SS ها (نکته ۱)	SST_r با نکته ۱	-
			SSE با نکته ۲	F با نکته ۲		-
			SST_r با نکته ۲			-
			F با نکته ۲	P-value		-
ریاضی	-	-	-	-	SSE با نکته ۲	-

فصل نهم: آمار توصیفی

از این فصل سؤال زیاد نمی‌آید اما دانستن ضرایب تغییرات و چولگی و کشیدگی و میانگین هندسی و تبدیل خطی و صدک ها مفید می‌باشد. ضمناً نمودارهای شاخه و برگ ـ جعبه‌ای و هیستوگرام را ملاحظه بفرمایید.

جدول توزیع موضوعات فصل ۹ در سؤالات ۵ سال اخیر

رشته / سال	۸۴	۸۵	۸۶	۸۷	۸۸	۸۹
صنایع	ـ	چارک‌ها (نمودار شاخه و برگ)	ـ	صدک بیستم	ـ	ـ
سیستم	میانگین آمیخته	رابطه میانگین میانه و مد (ضریب چولگی)	ضریب تغییرات	میانگین هندسی	نمودار جعبه‌ای و IQR میانگین انحراف از میانه	ـ
ریاضی	ضریب تغییرات	ـ	ـ	چارک سوم	واریانس	چارک‌ها در نمودار ساقه و برگ

تذکر:

تعداد سؤالاتی که برای هر فصل پیش‌بینی شده مربوط به رشته‌های صنایع و سیستم می‌باشد. برای دانشجویان رشته ریاضی چون تعداد سؤالات ۲۵ عدد می‌باشد معمولاً تعداد سؤالات فصل‌های سوم و چهارم از تعداد مطرح شده بیشتر می‌آید. توزیع تعداد سؤالات از فصل‌های اول تا نهم برای دانشجویان رشته ریاضی عبارت است از:

فصل اول یا نهم	۰ یا ۱ سؤال
فصل دوم	۳ یا ۴ سؤال
فصل سوم	۶ یا ۷ سؤال
فصل چهارم	۵ یا ۶ سؤال
فصل پنجم	۳ یا ۴ سؤال
فصل ششم	۳ یا ۴ سؤال
فصل هفتم	۰ یا ۱ سؤال
فصل هشتم	۰ یا ۱ سؤال

بخش سوم

نمونه سؤالات شبیه‌سازی شده امتحان

۱ ـ فرض کنید می‌خواهیم از تابع $y=\tan(x_1+x_2+\ldots+x_k)$ مشتق جزئی مرتبه r بگیریم طوری که نسبت به x_k حداقل k بار مشتق گرفته باشیم. به چند طریق این کار امکان‌پذیر است؟

۱) $\binom{r-1}{k-1}$ ۲) $\binom{r-2}{k-2}$ ۳) $\binom{r-3}{k-2}$ ۴) $\binom{k-2}{r-1}$.

۲ ـ اگر $P(A)=0.4$ و $P(B)=0.3$ و $P(A\cap B)=0.1$ حاصل $P(A\Delta B|A\cap B)$ چقدر است؟

۱) 0.1 ۲) 0.3 ۳) 0.7 ۴) صفر

۳ ـ فرض کنید عدد A را از مجموعه $\{1,2,3,4\}$ انتخاب نموده عدد B را از مجموعه $\{1,2,\ldots,A\}$ انتخاب می‌کنیم احتمال آن‌که $A=2$ به شرط آن‌که بدانیم $B=2$ چقدر است؟

۱) $\frac{5}{12}$ ۲) $\frac{1}{4}$ ۳) $\frac{6}{13}$ ۴) $\frac{7}{13}$

۴ ـ فرض کنید زمان تولید یک کالا دارای تابع چگالی $x\geq 0$ $f(x)=0.1e^{-0.1x}$ بر حسب ساعت می‌باشد. دستمزد تا a ساعت بر حسب هر ساعت یک تومان پرداخت می‌شود و برای زمان بیشتر از a دستمزد a تومان پرداخت می‌گردد متوسط دستمزد پرداختی چقدر است؟

۱) $10-e^{-10a}$ ۲) $10-10e^{-10a}$ ۳) e^{-10a} ۴) $\frac{a}{2}$

۵ ـ اگر تابع احتمال توام X و Y به صورت

Y \ X	−1	0	4
2	0.2	0.1	0.3
5	0.4	0	0

باشد امید ریاضی $|X-a|$ به ازای کدام مقدار a مینیمم می‌گردد؟

۱) 1 ۲) 0 ۳) -1 ۴) -0.5

۶ ـ اگر $X_1,\ldots,X_{10}$ مستقل از هم دارای امید ریاضی 4 و واریانس 3 باشند ضریب همبستگی $T_1=X_1+\ldots+X_5+X_6$ و $T_2=X_3+\ldots+X_{10}$ کدام است؟

۱) $\sqrt{3}$ ۲) $\frac{1.2}{\sqrt{3}}$ ۳) $\frac{\sqrt{3}}{3}$ ۴) $\frac{1}{\sqrt{6}}$

۷ ـ اگر X دارای توزیع پواسون با انحراف معیار 2 باشد $E|X-2|$ چقدر است؟

۱) 2 ۲) $12e^{-4}-2$ ۳) $12e^{-4}+2$ ۴) $e^{-4}+1$

۸ ـ فرض کنید زمان انتظار برای ورود مشتری در یک باجه پستی توزیع نمایی با میانگین 2 دقیقه دارد. چقدر احتمال دارد در 5 دقیقه حداقل 2 مشتری وارد شود؟

۱) $1-1.5e^{-2.5}$ ۲) $1-3.5e^{-2.5}$ ۳) $3.5e^{-6}$ ۴) $1-11e^{-10}$

۹ ـ اگر تابع مولد گشتاورهای X به صورت $M_X(t)=2^te^{t^2}$ باشد میانه توزیع X کدام است؟

۱) صفر ۲) 2 ۳) L_n2 ۴) $\sqrt{2}$

۱۰ـ در نمونه n تایی از جامعه هندسی با مدل شکست برآورد روش ماکزیمم درستنمایی ضریب تغییرات $(C.V)$ کدام است؟

۱) $\sqrt{1+\bar{X}}$ ۲) $\sqrt{\frac{1}{1+\bar{X}}}$ ۳) $\sqrt{\frac{\bar{X}}{1+\bar{X}}}$ ۴) $\sqrt{\frac{1+\bar{X}}{\bar{X}}}$

۱۱ـ فرض کنید از جامعه با چگالی $f(x,\theta)=\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}$ $x \in R$ نمونه‌ای n تایی گرفته‌ایم برآورد کننده روش گشتاوری θ کدام است؟

۱) $\sqrt{\frac{\sum_{i=1}^{n} X_i^2}{2n}}$ ۲) $\sqrt{\frac{\sum_{i=1}^{n} X_i^2}{n}}$ ۳) $\bar{X}$ ۴) $\frac{\sum_{i=1}^{n}|X_i|}{n}$

۱۲ـ اگر $X_1,\dots,X_5$ نمونه‌ای تصادفی از جامعه نرمال با میانگین صفر و واریانس σ^2 باشد C چه باشد تا $T=C\frac{X_5-3X_1}{\sqrt{X_2^2+X_3^2+X_4^2}}$ دارای توزیع $t-$ استیودنت گردد؟

۱) 1 ۲) $\sqrt{\frac{2}{3}}$ ۳) $\sqrt{\frac{10}{3}}$ ۴) $\sqrt{\frac{3}{10}}$

۱۳ـ بر اساس نمونه‌ای 11 تایی از جامعه نرمال انحراف معیار نمونه‌ای برابر 4 به دست آمده است برآورد فاصله‌ای $99\ \%$ اطمینان برای واریانس جامعه یک طرفه از بالا کدام است؟

۱) $(0, 6.9)$ ۲) $(0, (6.9)^2)$ ۳) $(0, \sqrt{6.9})$ ۴) $(0, 16)$

۱۴ـ فرض کنید میانگین نمونه‌ای 25 تایی از جامعه نرمال برابر 14 به دست آمد. اگر واریانس جامعه 9 باشد و بخواهیم $H_0 : \mu = 15$ را در برابر $H_1 : \mu \neq 15$ آزمون کنیم $P-Value$ کدام است؟

۱) 0.095 ۲) 0.0475 ۳) 0.05 ۴) 0.1

۱۵ـ با اطلاعات سؤال 14 اگر بخواهیم احتمال خطای نوع اول 0.05 و احتمال خطای نوع دوم به ازای $\mu = 17$ برابر 0.01 باشد چه تعداد نمونه باید بگیریم؟

۱) 23 ۲) 41 ۳) 42 ۴) 24

۱۶ـ اگر براساس نمونه X_2 و X_1 تایی از جامعه دارای توزیع یکنواخت بر بازه $(0,\theta)$ بخواهیم $H_0 : \theta = 2$ را در برابر $H_1 : \theta = 1$ موقعی قبول کنیم که $X_1 + X_2 < 1.5$ توان آزمون چقدر است؟

۱) $\frac{1}{2}$ ۲) $\frac{1}{4}$ ۳) $\frac{1}{8}$ ۴) $\frac{1}{16}$

۱۷ـ اگر در مدل رگرسیون نرمال با واریانس $\sigma^2 = 4$ نمونه‌ای 5 تایی به صورت $\begin{array}{c|ccccc} x & -2 & 2 & 4 & 3 & 3 \\ \hline y & 1 & 2 & -1 & 3 & 0 \end{array}$ داشته باشیم مقدار کوورایانس ضرایب خط برآورد رگرسیونی چقدر است؟

۱) $\frac{2}{15}$ ۲) $\frac{4}{11}$ ۳) $-\frac{4}{11}$ ۴) $-\frac{2}{15}$

۱۸ـ اگر با نمونه‌ای 18 تایی $Y = A + 04X$ و $X = A^* + 0.9Y$ خطوط برآورد رگرسیونی باشند آماره آزمون معنی داری ضریب همبستگی جامعه کدام است؟

۱) -3 ۲) 3 ۳) $\frac{3}{4}\sqrt{17}$ ۴) $-\frac{3}{4}\sqrt{17}$

۱۹ـ اگر در یک مدل آنالیز واریانس یک طرفه برای نمونه‌های $n_3 = 5$ و $n_2 = 8$ و $n_1 = 7$ تایی داشته باشیم $\bar{X}_3 = 1$ و $\bar{X}_2 = 3$ و $\bar{X}_1 = 3$ مجموع مربعات تیمار کدام است؟

۱) 12 ۲) 12.25 ۳) 15 ۴) 14.25

۲۰ـ اگر $X_1 \sim b\left(10, \frac{1}{3}\right)$ و $X_2 \sim b\left(20, \frac{2}{3}\right)$ مستقل باشند و بدانیم $X_2 - X_1 = 7$ انتظار دارید X_1 چه مقداری باشد؟

۱) $\frac{10}{3}$ ۲) $\frac{17}{3}$ ۳) $\frac{13}{3}$ ۴) $\frac{16}{3}$

سوالات زیر برای دانشجویان ریاضی در نظر گرفته شده است.

۲۱ـ تابع چگالی توأم متغیرهای تصادفی X و Y به صورت زیر است. امید ریاضی X کدام است؟

$$f(x,y)=\begin{cases}\binom{x}{y}\left(\frac{1}{2}\right)^x\left(\frac{x}{15}\right) & ;x=1,2,3,4,5 \text{ و } y=0,1,\dots,x \\ 0 & \text{در غیر این صورت}\end{cases}$$

۱) $\frac{3}{11}$ ۲) 3 ۳) $\frac{11}{3}$ ۴) 11

۲۲ـ اگر $X_1, X_2, \dots, X_n$، n متغیر تصادفی باشند، به طوری که $i = 1, 2, \dots, n$ و $P\left(X_i = e^{-1}\right) = \frac{1}{2}$ و $P\left(X_i = e^2\right) = \frac{1}{2}$ مقدار $E\left(\mathrm{Ln}\prod_{i=1}^{n} X_i\right)$ کدام است؟

۱) $\frac{1}{2^n}$ ۲) $\frac{\left(e^3 - e\right)^n}{2^n e^n}$

۳) $n - n\ln 2 - n\ln\left(e^2 - 1\right)$ ۴) $\frac{n}{2}$

۲۳ـ متغیر تصادفی X دارای تابع توزیع احتمال به صورت زیر است:

$$F_X(x)=\begin{cases}0 & x<1 \\ \frac{x^2}{5} & 1\le x<2 \\ 1 & x\ge 2\end{cases}$$

واریانس متغیر تصادفی X برابر است با:

۱) $\frac{140}{450}$ ۲) $\frac{67}{450}$ ۳) $\frac{67}{225}$ ۴) $\frac{32}{225}$

۲۴ ـ اگر تا لحظه t تعداد مشتریان مرد و زن یک فروشگاه مستقلاً دارای توزیع پواسون به ترتیب با پارامترهای λt و μt باشد ، احتمال اینکه اولین مشتری فروشگاه مرد باشد برابر است با:

۱) $1-e^{\frac{-\lambda}{\mu}}$ ۲) $\frac{\lambda}{\lambda+\mu}$ ۳) $\frac{\lambda!}{(\lambda+\mu)!}$ ۴) $\frac{\lambda!}{\mu!}e^{-\frac{\lambda}{\mu}}$

۲۵ ـ فرض کنید $X \sim N\left(0,\sigma^2\right)$. ضریب اطمینان فاصله $\left(|X|,10|X|\right)$ برای σ کدام است؟

۱) $\Phi(1.1)$ ۲) $\Phi(1)-\Phi(0.1)$ ۳) $\Phi(0.1)+\Phi(1)$ ۴) $2\left(\Phi(1)-\Phi(0.1)\right)$

حل تشریحی

۱ ـ گزینه ۱ درست است.

در واقع باید تعداد جواب معادله $x_1+\ldots+x_k=r-k$ را بیابیم که می‌شود:

$$\binom{r-k+k-1}{k-1}=\binom{r-1}{k-1}$$

۲ ـ گزینه ۴ درست است.

طبق تعریف

$$P(A\Delta B|A\cap B)=\frac{P(A\Delta B)\cap(A\cap B)}{P(A\cap B)}=\frac{P(\phi)}{P(A\cap B)}=0$$

زیرا:

$$A\,\Delta\,B=(A\cup B)-(A\cap B)$$

۳ ـ گزینه ۳ درست است.

بنا به فرمول بیز:

$$P(A=2|B=2)=\frac{P(A=2)P(B=2)(A=2)}{\sum_{i=1}^{4}P(A=i)P(B=2|A=i)}=\frac{\frac{1}{4}\ \frac{1}{2}}{\frac{1}{4}(0)+\frac{1}{4}\ \frac{1}{2}+\frac{1}{4}\ \frac{1}{3}+\frac{1}{4}\ \frac{1}{4}}=\frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}}=\frac{6}{13}$$

۴ ـ گزینه ۳ درست است.

در واقع اگر Y دستمزد باشد پس Y مینیموم X و a است چون:

$$Y=\begin{cases}X & X\le a\\ a & X>a\end{cases}$$

$$\begin{aligned}E(Y)&=E(x)P(X\le a)+E(a)P(X>a)\\&=\int_0^a x\lambda e^{-\lambda x}dx+a\int_a^\infty \lambda e^{-\lambda x}dx\\&=-xe^{-\lambda x}-\frac{1}{\lambda}e^{-\lambda x}\Big]_0^a+ae^{-\lambda a}=\frac{1}{\lambda}-\frac{1}{\lambda}e^{-\lambda a}=\frac{1-e^{-\lambda a}}{\lambda}\end{aligned}$$

۵ ـ گزینه ۳ درست است.

زیرا تابع احتمال کناری X عبارت است از $\begin{array}{c|ccc}X & -1 & 0 & 4\\ \hline & 0.6 & 0.1 & 0.3\end{array}$ و $E|X-a|$ به ازای میانه توزیع مینیموم می‌گردد و میانه این توزیع -1 است چون $P(X\ge -1)\ge\frac{1}{2}$ و $P(X\le -1)\ge\frac{1}{2}$

۶ ـ گزینه ۳ درست است.

$$\begin{aligned}Cov(T_1,T_2)&=Cov(X_1+X_2+X_3+X_4+X_5+X_6 \text{ و } X_3+X_4+X_5+X_6+X_7+X_8+X_9+X_{10})\\&=Var(X_3)+Var(X_4)+Var(X_5)+Var(X_6)=4(3)=12\end{aligned}$$

$$Var(T_1)=6(3)=18\quad,\quad Var(T_2)=8(3)=24$$

پس:

$$\rho=\frac{12}{\sqrt{18(24)}}=\frac{1}{\sqrt{3}}$$

۷ـ گزینه ۳ درست است.

$$E|X-2|=\sum_{x=0}^{\infty}|x-2|\frac{e^{-4}4^x}{x!}=2e^{-4}+\frac{e^{-4}4^1}{1!}+0+\sum_{x=3}^{\infty}(x-2)\frac{e^{-4}4^x}{x!}$$

$$=2e^{-4}+4e^{-4}+2e^{-4}+4e^{-4}+\sum_{x=0}^{\infty}(x-2)\frac{e^{-4}4^x}{x!}=12+E(X-2)=12e^{-4}+4-2=12e^{-4}+2$$

۸ـ گزینه ۲ درست است.

بنا به رابطه پواسون با نمایی تعداد مشتری ورودی در هر دقیقه توزیع پواسون با $\lambda=\frac{1}{2}$ دارد و در 5دقیقه $\lambda=\frac{5}{2}$ حال:

$$P(X\ge 2)=1-P(X=0)-P(X=1)=1-e^{-2.5}-2.5e^{-2.5}=1-3.5e^{-2.5}$$

۹ـ گزینه ۳ درست است.

$$M_X(t)=2^te^{t^2}=e^{tL_n2+t^2}$$

که با تطبیق مولد گشتاورهای نرمال داریم:

$$\mu=L_n2 \quad , \quad \sigma^2=2$$

و در توزیع نرمال میانه و میانگین برابرند.

۱۰ـ گزینه ۴ درست است.

می‌دانیم $C.V=\frac{\sigma}{\mu}$ و در توزیع هندسی با مدل شکست:

$$E(X)=\frac{q}{p} \quad , \quad V_{an}(X)=\frac{q}{p^2}$$

$$C.V=\frac{\sqrt{\frac{q}{P^2}}}{\frac{q}{p}}=\frac{1}{\sqrt{q}}=\frac{1}{\sqrt{1-p}}$$

اما $MLE(p)=\frac{1}{1+\bar{X}}$ پس $MLE\left(\frac{\sigma}{\mu}\right)=\frac{1}{\sqrt{1-\frac{1}{1+\bar{X}}}}=\sqrt{\frac{1+\bar{X}}{\bar{X}}}$

۱۱ـ گزینه ۱ درست است.

با توجه به آن‌که $E(X)=0$ پس:

$$E(X^2)=\int_{-\infty}^{+\infty}\frac{x^2}{2\theta}e^{-\frac{|x|}{\theta}}dx=\int_0^{\infty}\frac{x^2}{\theta}e^{\frac{-x}{\theta}}dx=2\theta^2$$

حال از $2\theta^2 = \overline{X^2}$

داریم:

$$MME(\theta) = \hat{\theta} = \sqrt{\frac{\overline{X^2}}{2}} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{2n}}$$

۱۲ ـ گزینه ۴ درست است.

چون $X_5 - 3X_1 \sim N\left(0, 10\sigma^2\right)$, $Z = \frac{X_5 - 3X_1}{\sqrt{10}\sigma} \sim N(0,1)$ و $Y = \frac{X_2^2 + X_3^2 + X_4^2}{\sigma^2} \sim \chi^2_{(3)}$ لذا:

$$T = \frac{Z}{\sqrt{\frac{Y}{3}}} = \sqrt{\frac{3}{10}} \frac{X_5 - 3X_1}{\sqrt{X_2^2 + X_3^2 + X_4^2}} \sim t_{(3)}$$

و درنتیجه $C = \sqrt{\frac{3}{10}}$

۱۳ ـ گزینه ۱ درست است.

برآورد مذکور عبارتست از $\left(0, \frac{(n-1)S^2}{\chi^2_{1-\alpha, n-1}}\right)$ که چون $\chi^2_{0.99, 10} = 23.2$ پس جواب می‌شود $\left(0, \frac{10(4)^2}{23.2}\right)$ یا $(0, 6.9)$

۱۴ ـ گزینه ۱ درست است.

$$P-value = 2\min\begin{Bmatrix} P\left(\bar{X} \le 14 | \mu = 15\right) \\ P\left(\bar{X} \ge 14 | \mu = 15\right) \end{Bmatrix} = 2P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{14 - 15}{\frac{3}{5}}\right) = 2P\left(Z \le \frac{-5}{3}\right) = 2\Phi(-1.67) = 2(0.0478) = 0.0950$$

۱۵ ـ گزینه ۳ درست است.

$\mu = 17$ نقش $\mu_1 = 17$ را بازی می‌کند پس

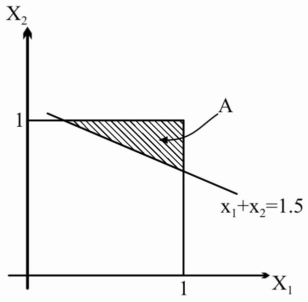
$$n = \frac{\sigma^2\left(Z_{\frac{\alpha}{2}} + Z_\beta\right)^2}{(\mu_1 - \mu_0)^2} = \frac{9\left(Z_{0.025} + Z_{0.01}\right)^2}{(17 - 15)^2} = \frac{9(1.96 + 2.33)^2}{(2)^2} = 41.4 \approx 42$$

۱۶ ـ گزینه ۳ درست است.

طبق تعریف

$$\text{توان} = 1 - \beta = P\ (H_0 \text{ رد} |\ H_0 \text{ نادرستی}) = P\left(X_1 + X_2 \ge 1.5 | \theta = 1\right)$$

$$= \frac{S_A}{S_{\text{مربع}}} = \frac{\frac{1}{8}}{1} = \frac{1}{8}$$



۱۷ ـ گزینه ۳ درست است.

می‌دانیم

$$\operatorname{Cov}(A, B)=\operatorname{Cov}\left(\bar{Y}-B\bar{x}, B\right)=0-\bar{x}\operatorname{Var}(B)=-\frac{\bar{x}\sigma^2}{S_{xx}}$$

که چون $\sum x_i^2=42$ و $\bar{x}=\frac{-2+2+4+3+3}{5}=2$ پس $S_{xx}=42-5(2)^2=22$ لذا جواب می‌شود:

$$-\frac{(2)(4)}{22}=-\frac{4}{11}$$

۱۸ ـ گزینه ۲ درست است.

با توجه به آن‌که $r^2=BB^*=(0.4)(0.9)=0.36$ پس $r=0.6$ (چون $B>0$)، لذا:

$$T=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}=\frac{0.6\sqrt{16}}{\sqrt{0.64}}=3$$

۱۹ ـ گزینه ۳ درست است.

اولاً:

$$\bar{X}_{\bullet\bullet}=\frac{\sum n_i\bar{X}_i}{\sum n_i}=\frac{21+24+5}{7+8+5}=\frac{5}{20}=2.5$$

ثانیاً

$$SST_r=\sum n_i\left(\bar{X}_i-\bar{X}_{\bullet\bullet}\right)^2=7(0.5)^2+8(0.5)^2+5(-1.5)^2=1.75+2+11.25=15$$

۲۰ ـ گزینه ۳ درست است.

می‌دانیم $Y_1=10-X_1\sim b\left(10, \frac{2}{3}\right)$ لذا Y_1 و X_2 مستقلند با $p=\frac{2}{3}$ لذا بنا به رابطه دو جمله‌ای مشروط با فوق هندسی:

$$X_2+Y_1=X_2+10-X_1=17=m$$

$$Y_1\big|(X_2+Y_1=m)\sim HG(N=30, n=m, a=10) \quad\rightarrow\quad E(Y_1)=m\frac{a}{N}=17\frac{10}{30}=\frac{17}{3}$$

پس:

$$E(X_1)=10-E(Y_1)=\frac{13}{3}$$

حل تشریحی سوالات اضافی رشته ریاضی

۲۱ ـ گزینهٔ ۳ درست است.

متغیر تصادفی X مقادیر $1,2,3,4,5$ را اختیار می‌کند و توزیع حاشیه‌ای X برابر است با:

$$f(x)=\sum_{y=0}^{x} f(x,y)=\sum_{y=0}^{x} \binom{x}{y}\left(\frac{1}{2}\right)^{x}\left(\frac{x}{15}\right)=\left(\frac{1}{2}\right)^{x}\left(\frac{x}{15}\right)\sum_{y=0}^{x}\binom{x}{y}=\left(\frac{1}{2}\right)^{x}\left(\frac{x}{15}\right)2^{x}=\frac{x}{15}$$

بنابراین امید ریاضی متغیر تصادفی X برابر است با:

$$E(X)=\sum_{x=1}^{5} xf(x)=\sum_{x=1}^{5} x\left(\frac{x}{15}\right)=\frac{1}{15}\sum_{x=1}^{5}x^{2}=\frac{1}{15}\,\frac{5(6)(11)}{6}=\frac{11}{3}$$

۲۲ ـ گزینهٔ ۴ درست است.

$$E\left(\operatorname{Ln}\prod_{i=1}^{n}X_i\right)=E\left(\sum_{i=1}^{n}\operatorname{Ln}X_i\right)=\sum_{i=1}^{n}E\left(\operatorname{Ln}X_i\right)$$

$$E\left(\operatorname{Ln}X_i\right)=\sum_{x_i}\operatorname{Ln}X_iP\left(X_i=x_i\right)=\operatorname{Ln}\left(e^{-1}\right)P\left(X_i=e^{-1}\right)+\operatorname{Ln}\left(e^{2}\right)P\left(X_i=e^{2}\right)$$

$$=\operatorname{Ln}\left(e^{-1}\right)\frac{1}{2}+\operatorname{Ln}\left(e^{2}\right)\frac{1}{2}=-\frac{1}{2}+1=\frac{1}{2}\quad\Rightarrow\quad E\left(\operatorname{Ln}\prod_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E\left(\operatorname{Ln}X_i\right)=\sum_{i=1}^{n}\frac{1}{2}=\frac{n}{2}$$

۲۳ ـ گزینهٔ ۲ درست است.

X متغیری آمیخته است چون پله‌ای نیست و نقطه گسستگی دارد.

$$P(X=1)=F(1)-F\left(1^{-}\right)=\frac{1}{5}-0=\frac{1}{5}$$

$$P(X=2)=F(2)-F\left(2^{-}\right)=1-\frac{4}{5}=\frac{1}{5}$$

$$f(x)=\frac{2x}{5}\quad,\quad 1<x<2$$

$$E(X)=1\left(\frac{1}{5}\right)+2\left(\frac{1}{5}\right)+\int_{1}^{2}\frac{2x^{2}}{5}dx=\frac{3}{5}+\frac{14}{5}=\frac{23}{15}$$

$$E\left(X^{2}\right)1^{2}\left(\frac{1}{5}\right)+2^{2}\left(\frac{1}{5}\right)\int_{1}^{2}\frac{2x^{3}}{5}dx=1+\frac{15}{10}=\frac{25}{10}$$

$$\operatorname{Var}(X)=\frac{25}{10}-\left(\frac{23}{15}\right)^{2}=\frac{67}{450}$$

۲۴ ـ گزینهٔ ۲ درست است.

اگر X و Y به ترتیب مدت زمان انتظار برای ورود اولین مشتری باشد، با توجه به رابطه نمایی و پواسون داریم :

$$X\sim \operatorname{Exp}(\lambda)\quad,\quad Y\sim \operatorname{Exp}(\mu)$$

در توزیع نمایی می‌توان دید وقتی X و Y مستقل هستند :

$$P(X<Y)=\frac{\lambda}{\lambda+\mu}$$

۲۵ ـ گزینهٔ ۴ درست است.

بنا به فرض $X\sim N(0,\sigma^2)$ پس $\frac{X}{\sigma}\sim N(0,1)$

$$P(|X|<\sigma<10|X|)=P(0.1<|Z|<1)=P(-1<Z<-0.1)+P(0.1<Z<1)=2(\Phi(1)-\Phi(0.1))$$