به نام خداوند بخشنده مهربان

روش‌های آماری و اقتصادسنجی

در تحلیل و مدل‌سازی داده‌های حمل و نقلی

محمد مهدی بشارتی

besharati@iut.ac.ir

احتمالات

فصل سوم

بخش اول: مفاهیم احتمال

بخش دوم: متغیر تصادفی

مفاهیم اولیه احتمال

مدرس: محمدمهدی بشارتی

- ✓ مهندسی حمل‌ونقل: برنامه‌ریزی، طراحی، ساخت، اجرا و نگهداری بخش‌های مختلف سیستم‌های حمل‌ونقلی

- ✓ این کار، مستلزم درک عمیقی از عرضه و تقاضاست؛
  - ○ عرضه سیستم (مثلا طول جاده‌ها، طول خطوط راه‌آهن، تنظیمات چراغ‌های راهنمایی و غیره) و
  - ○ تقاضای سیستم (مثلا تعداد وسایل‌نقلیه‌ای که می‌خواهند از یک بزرگراه استفاده کنند، تعداد قطارهای عبوری از یک مسیر راه‌آهن، تعداد کشتی‌ها در یک بندر و غیره)

- ✓ این امر یک فرآیند بسیار پیچیده و **دربردارنده متغیرهایی است که ماهیت آنها تصادفی و احتمالاتی** است.

- **به عبارت دیگر، مهندسین حمل‌ونقل همچون سایر رشته‌های مهندسی، همواره با مسائل غیرقطعی روبرو هستند.**

سوال: منظور از مسائل غیرقطعی چیست؟

منظور از مسائل غیرقطعی، پدیده‌هایی است که قبل از وقوع، نتیجه آن‌ها قابل پیش‌بینی نبوده و در هر بار رخ‌دادن در شرایط یکسان، می‌توانند نتایج مختلفی را به دنبال داشته باشند. به این‌گونه پدیده‌ها، آزمایش‌های تصادفی گفته می‌شود.

آزمایش تصادفی عبارتست از عملی که در شرایط ثابت، در هر بار تکرار، منجر به یکی از اعضای فضای نمونه (S) می‌شود.

مثال‌هایی از آزمایش‌های تصادفی؟

- تعداد خودروهایی که در ساعت اوج ترافیک صبح، از یک مقطع بزرگراهی عبور می‌کنند.
- نسبت روزانه تعداد مسافرین مذکر به مسافرین مونث در هر مبدا-مقصد سامانه مترو
- تعداد سفرهای روزانه تولید شده در یک ناحیه از یک شهر
- تعداد تصادفات سالانه در یک تقاطع معین
- و ....

مفاهیم اولیه احتمال

باید توجه داشت که:

مهندسین اغلب ترجیح می‌دهند این مسائل غیرقطعی و احتمالاتی را به مدل‌هایی ساده تبدیل‌کرده و مطالعه کنند. بسیاری از فرمول‌های مهندسی که در ظاهر، یک فرمول با فرم بسته به‌نظر می‌رسند؛ درواقع، شامل این مفاهیم احتمالاتی می‌شوند.

- ✓ برای مثال، **آیین‌نامه‌های طراحی**، براساس تحلیل‌های آماری، تئوری احتمال و نمونه‌های آماری توسعه داده می‌شوند. اما در نهایت برای سهولت در استفاده، این روابط را به‌صورت «جبری» و نه «احتمالاتی» برای استفاده مهندسین ارائه می‌کنند.

- ✓ به‌علاوه، **ضریب اطمینانی** که در بسیاری از روابط مهندسی استفاده می‌گردد، از همین مفهوم سرچشمه می‌گیرد.

مفاهیم اولیه احتمال

قبل از اینکه هر آزمایشی صورت گیرد، می‌توان فهرستی از نتایج موردانتظار برای آن را معین کرد و تمامی نتایج آزمایش را در یک مجموعه نشان داد.

فضای نمونه(S)، مجموعه‌ای است متشکل از تمام نتایج ممکنی که با توجه به شرایط آزمایش و برحسب موضوع موردنظر (برای یک مشاهده یا آزمایش تصادفی) قابل تصور است.

توجه شود که هر فضای نمونه، یک مجموعه است. این مجموعه ممکن است قابل شمارش (یا گسسته) و یا غیرقابل شمارش (یا پیوسته) باشد.

مثال:

- (فضای نمونه گسسته): فضای نمونه مدرک تحصیلی افراد متقاضی گواهینامه رانندگی
- (فضای نمونه پیوسته): فضای نمونه سرعت خودروها در یک خیابان، مثلا

$$S = \{X|0 \le x \le 250\}$$

مفاهیم اولیه احتمال

برآمد (Outcome): به هر عضو از فضای نمونه، یک **برآمد یا نتیجه** گویند.

پیشامد (E): به‌هر زیرمجموعه قابل اندازه‌گیری از فضای نمونه، یک پیشامد ($Event$) گفته می‌شود.

سوال: برای دو فضای نمونه زیر، چند پیشامد را مثال بزنید؟

- (فضای نمونه گسسته): فضای نمونه مدرک تحصیلی افراد متقاضی گواهینامه رانندگی
- (فضای نمونه پیوسته): فضای نمونه سرعت خودروها در یک خیابان

- $S = \{X | 0 \le x \le 150\}$

مثلا:

- اگر سرعت مجاز تردد در یک خیابان، ۶۰ کیلومتربرساعت باشد، یک مثال از پیشامد:
- پیشامد اینکه سرعت خودرویی بیشتر از سرعت غیرمجاز باشد $A = \{X | x > 60\}$

مفاهیم اولیه احتمال

پیشامد(E): به‌هر زیرمجموعه قابل اندازه‌گیری از فضای نمونه، یک پیشامد(E) گفته می‌شود.

احتمالِ پیشامد A: یک اندازه عددی که امکان وقوع پیشامد A را نشان می‌دهد.

$$\mathrm{P}(A)=\frac{n(A)}{n(S)}=\frac{\text{تعداد اعضای پیشامد } A}{\text{تعداد اعضای فضای نمونه}}$$

احتمالِ پیشامد A، عددی بین ۰ و ۱ است

برخی از مهم‌ترین روابط حاکم بر پیشامدها عبارتند از:

۱- برای هر پیشامد مانند A، رابطه روبر برقرار است: $0 \leq P(A) \leq 1$

۲- $P(\emptyset) = 0$ ، $P(S) = 1$ ، که در آن، $\emptyset$ پیشامد تهی است.

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

مفاهیم اولیه احتمال

برخی از مهم‌ترین روابط حاکم بر پیشامدها عبارتند از:

٤- اگر A و B دو پیشامد دوبه‌دو ***ناسازگار*** باشند(یعنی: $(A \cap B) = \emptyset$)؛ آنگاه:

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

٥- اگر A' متمم پیشامد A باشد: $P(A') = 1 - P(A)$

٦- درصورتیکه A و B، دو پیشامد ***مستقل*** از هم باشند، آنگاه:

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

سوال: تفاوت دو ویژگی "ناسازگاری" و "استقلال" چیست؟

مفاهیم اولیه احتمال

مدرس: محمدمهدی بشارتی

سوال: تفاوت دو ویژگی "ناسازگاری (Mutual Exclusiveness)" و "استقلال (Independence)" چیست؟

- دو پیشامد A و B را ناسازگار گوییم هرگاه هیچ اشتراکی با هم نداشته باشند، به عبارت دیگر دو پیشامد نتوانند همزمان رخ دهند.

مثال:

در پرتاب یک تاس:

- A: عدد ۲ بیاید
- B: عدد ۵ بیاید

نمی‌توان هم‌زمان هر دو رخ دهند.

- ویژگی استقلال دو پیشامد، یک رابطه عددی است و نه مجموعه‌ای. دو پیشامد A و B مستقل هستند، هرگاه وقوع یکی از آنها، احتمال وقوع دیگری را تغییر ندهد.

مثال:

دو پرتاب سکه مستقل:

- A: در پرتاب اول شیر بیاید
- B: در پرتاب دوم شیر بیاید

وقوع اولی روی دومی اثری ندارد.

احتمال شرطی

- گاهی اوقات، آگاهی از وقوع یک پیشامد، می‌تواند در محاسبه احتمال وقوع پیشامدهای دیگر موثر باشد.
- احتمال وقوع پیشامد B هنگامی که (بدانیم) پیشامد A اتفاق افتاده‌باشد را احتمال شرطی می-نامند.
- احتمال وقوع B به شرط وقوع A از رابطه زیر بدست می‌آید.

$$P(B|A)=\frac{P(A\cap B)}{P(A)} \quad , \qquad P(A)>0$$

تفسیر شهودی:
احتمال شرطی یعنی «احتمال وقوع B در فضای نمونه‌ی جدیدی که فقط شامل حالت‌هایی است که A رخ داده است».

مثال

فرض کنید:

۴۰٪ دانشجویان یک کلاس پسر هستند.

۲۵٪ هم پسر و هم عینکی هستند.

اگر بدانیم یک دانشجو پسر است، احتمال اینکه عینکی باشد چقدر است؟

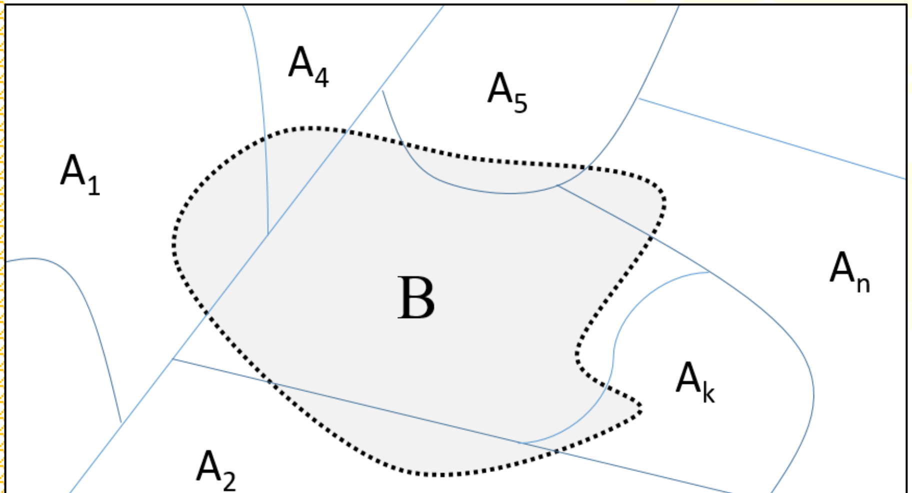
$$P(\text{عینکی}|\text{پسر})=\frac{P(A\cap B)}{P(A)}=\frac{0.25}{0.4}=0.625$$

یعنی در میان پسران کلاس، ۶۲.۵٪ عینکی هستند.

قضیه بیز

باتوجه به شکل روبرو؛

فرض کنید $\{A_i\,, i=1,2,\ldots, n\}$ مجموعه‌ای از پیشامدهای دوبه‌دو ناسازگاری است که فضای نمونه را به‌گونه‌ای تقسیم می‌کنند که به ازای هر i، $P(A_i)$ غیرصفر باشد (و اجتماعشان کل فضای نمونه باشد).



اگر B پیشامدی از فضای نمونه باشد که $P(B)\neq 0$،

آنگاه به‌ازای $k=1,2,\ldots, n$ خواهیم داشت؛

$$P(A_k|B)=\frac{P(A_k\cap B)}{P(A_1\cap B)+P(A_2\cap B)+\cdots+P(A_n\cap B)}=\frac{P(B|A_k)\,.\,P(A_k)}{P(B)}=\frac{P(B|A_k)\,.\,P(A_k)}{\sum_{i=1}^{n}P(B|A_i)\,.\,P(A_i)}$$

قضیه بیز

دو مورد از مهم‌ترین کاربردهای قضیه بیز، عبارتند از:

۱) محاسبه احتمالات شرطی

۲) ارائه برآوردهای دقیق‌تر برای پارامترهای جامعه با ترکیب منابع مختلف اطلاعاتی (**رویکرد مدل‌سازی بیزین**).

$P(A_k)$: توزیع پیشین (Prior)

$P(B|A_k)$: درستنمایی (Likelihood)

$P(A_k|B)$: توزیع پسین (Posterior)

توزیع پیشین　　درستنمایی

$$P(A_k|B) = \frac{P(B|A_k)\,.\,P(A_k)}{\sum_{i=1}^{n} P(B|A_i)\,.\,P(A_i)}$$

توزیع پسین

فرمول بیز را می‌توان به صورت مفهومی زیر نوشت:

Prior × Likelihood ∝ Posterior

مثال:

پلیس راهور در قطعه‌ای از یک راه، راننده‌ای را به جرم تردد با سرعت غیرمجاز متوقف کرده و از او تست الکل می‌گیرد. اگر نتیجه تست مثبت بوده باشد، با توجه به مفروضات زیر می‌خواهیم بدانیم احتمال اینکه او واقعا مست بوده، چقدر است؟ فرض کنید احتمال اینکه دستگاه به اشتباه نتیجه آزمایش را مثبت نشان دهد ۵٪ و احتمال اینکه به اشتباه نتیجه آزمایش را منفی نشان دهد ۱۰٪ می‌باشد. همچنین، فرض کنید از قبل می‌دانیم که به‌طور متوسط ۸٪ از رانندگانی که از این قطعه عبور می‌کنند مست هستند.

پیشامدهای زیر را درنظر می‌گیریم:

A_1: رانندگی درحال مستی

A_2: رانندگی درحال سلامت

B: مثبت بودن نتیجه آزمایش

$$P(A_1) = 0/08 \qquad P(A_2) = 1 - 0/08 = 0/92$$

$$P(B|A_1) = 0/90 \qquad P(B|A_2) = 0/05$$

مثال:

بدین‌ترتیب، براساس قضیه بیز خواهیم داشت؛

$$P(A_1|B) = \frac{p(A_1 \cap B)}{P(A_1 \cap B) + P(A_2 \cap B)} = \frac{P(B|A_1).P(A_1)}{P(B)} = \frac{0/90 * 0/08}{0/90 * 0/08 + 0/05 * 0/92} \simeq 0/61$$

بنابراین، احتمال اینکه راننده‌ای که نتیجه تست الکل او مثبت بوده، واقعا درحال مستی باشد، ۰/۶۱ می‌باشد.

توجه:

احتمال اینکه راننده‌ای با تست مثبت واقعاً مست باشد، علاوه بر **دقت دستگاه**، به **نرخ پایه مستی** در جامعه نیز وابسته است. این وابستگی به نرخ پایه، یکی از مفاهیم بنیادین در استنباط بیزین است.

در آمار بیزین:

- درصد رانندگان مست = **توزیع پیشین (Prior)**
- عملکرد تست = **درستنمایی (Likelihood)**
- احتمال نهایی = **پسین (Posterior)**

احتمالات

فصل سوم

بخش اول: مفاهیم احتمال

بخش دوم: متغیر تصادفی

متغیر تصادفی

- متغیر تصادفی، یکی از مفاهیم مهم در آمار و احتمالات است.
- متغیر تصادفی، تابعی است که به *اعضای فضای نمونه*، اعدادی را اختصاص می‌دهد.
- مرسوم است که: متغیرهای تصادفی را معمولا با حروف بزرگ لاتین(مانند X) نمایش داده و هر مقدار معین از یک متغیر تصادفی را با حروف کوچک (x) نمایش می دهند.
- دو نوع متغیر تصادفی در متون مهندسی رایج هستند:

➢ متغیرهایی که تنها مقادیر **گسسته** می‌گیرند. به عنوان مثال، فرض کنید یک مهندس تعداد تصادفات سالانه یک تقاطع را در نظر دارد. واضح است که مقادیر مورد نظر، اعداد صحیح نامنفی هستند.

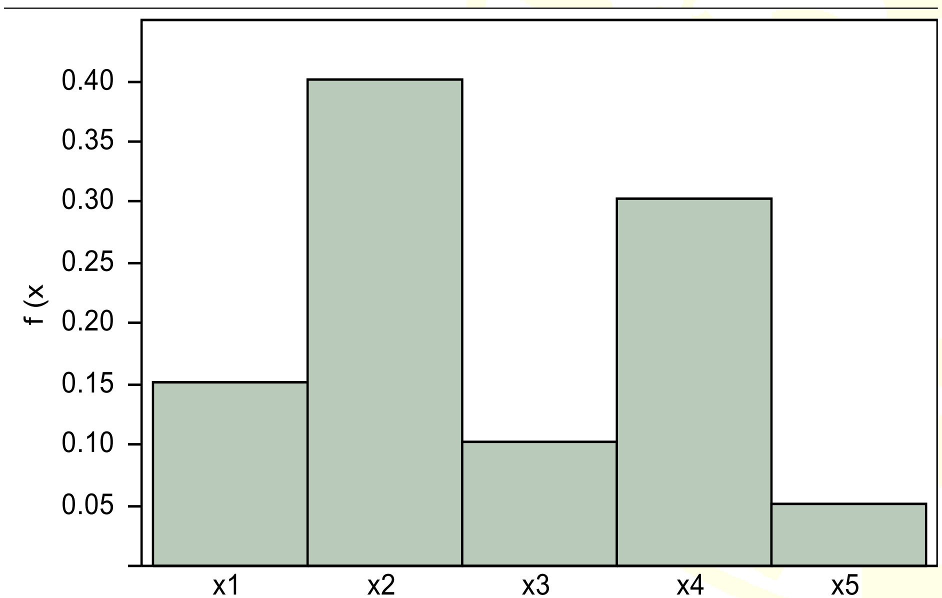
➢ متغیرهایی که می‌توانند تمام مقادیر حقیقی را در بازه مشخصی اختیار نمایند و به **متغیرهای تصادفی پیوسته** موسوم هستند. به عنوان مثال، سرعت وسایل‌نقلیه می‌تواند هر مقدار واقعی (بین صفر و حداکثر سرعت وسایل نقلیه) را اختیار نماید.

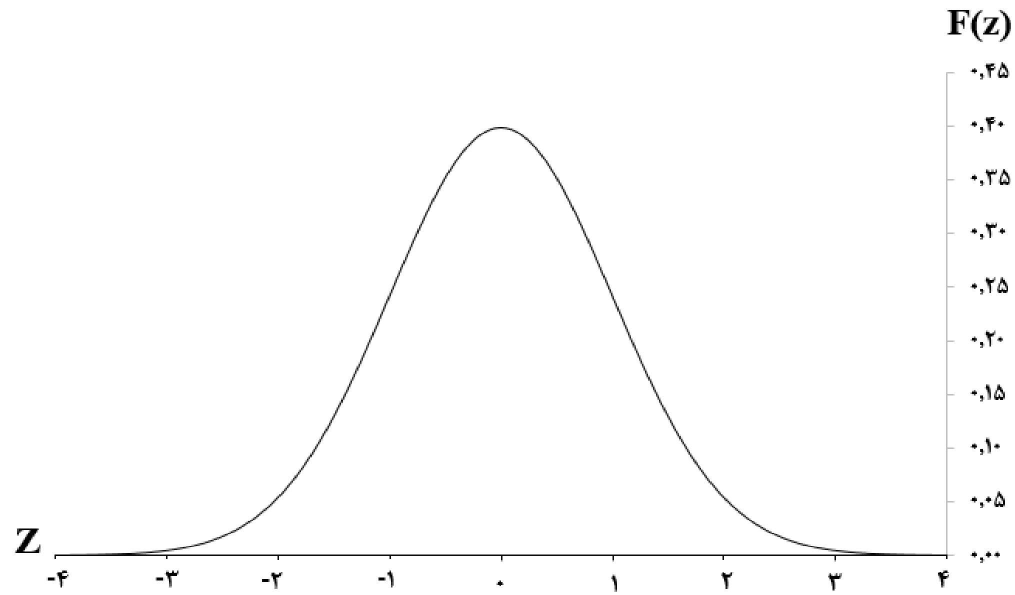
متغیر تصادفی

مدرس: محمدمهدی بشارتی

- ✓ یکی از مهم‌ترین چالش‌هایی که در برخورد با یک متغیرتصادفی با آن مواجه هستیم، برآورد احتمال وقوع (یا مشاهده) هریک از مقادیر ممکن (یا همان پیشامدهای) آن متغیر تصادفی است.

- ✓ برای این منظور از **"توزیع‌های احتمالاتی"** استفاده می‌شود (موضوع فصل بعد).





❖ مشخصه‌های اصلی توزیع های احتمالاتی

۱- امید ریاضی

امید ریاضی یا مقدار مورد انتظار یک متغیر تصادفی برابراست با مجموع حاصلضرب هرمقدار متغیر تصادفی در احتمال متناظر با آن مقدار.

به‌عبارت دیگر، امیدریاضی X میانگین وزنی متغیرX است که در آن، وزن هریک از مقادیر برابراست با احتمال مشاهده مقدار موردنظر.

امیدریاضی، شاخصی از موقعیت مرکزی توزیع احتمال احتمال (**central tendency**) می‌باشد.

❖ مشخصه‌های اصلی توزیع های احتمالاتی

۱- امید ریاضی

اگر X یک متغیرتصادفی **گسسته** و $f(x)$ مقدار احتمال متناظر با مشاهده x باشد ($P(X=x)$)،
آنگاه مقدار موردانتظار X برابر است با

$$E(X) = \sum_{x} x.f(x)$$

درحالتی‌که X متغیرتصادفی **پیوسته** باشد؛ به‌جای عمل جمع، از انتگرال‌گیری استفاده می‌شود

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)dx$$

$f(x)$: تابع چگالی احتمال

❖ مشخصه‌های اصلی توزیع های احتمالاتی

۱- امید ریاضی

اگر X_1، X_2، ...، X_n و نیز Y **متغیرهای تصادفی** و a_i **مقدار ثابت** باشند، آنگاه، مهم‌ترین ویژگی‌های امیدریاضی عبارتند از:

$$E(a_0 + a_1X + a_2X^2 + \cdots)$$

۱) $E(a_0 + a_1X + a_2X^2 + \cdots) = a_0 + a_1E(X) + a_2E(X^2)$

۲) $E(X + Y) = E(X) + E(Y)$

۳) $E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + a_nE(X_n)$

۴) اگر X و Y متغیرهای تصادفی **مستقل** باشند: $E(X.Y) = E(X).E(Y)$

$$E(a_0) = a_0$$

❖ مشخصه‌های اصلی توزیع های احتمالاتی

۲- **واریانس و کوواریانس**

✓ واریانس و کوواریانس، **شاخص هایی از میزان پراکندگی** مقادیر متغیر تصادفی حول امید ریاضی آن است.

اگر X یک متغیرتصادفی **گسسته** و $f(x)$ مقدار احتمال متناظر با مشاهده x باشد، آنگاه مقدار **واریانس** X برابر است با

$$Var(X) = E(X^2) - (E(X))^2$$

درحالتی‌که X متغیرتصادفی **پیوسته** باشد؛ به‌جای عمل جمع، از انتگرال‌گیری استفاده می‌شود.

$$Var(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 . f(x) dx$$

❖ مشخصه‌های اصلی توزیع های احتمالاتی

۲- واریانس و کوواریانس

✓ واریانس و کوواریانس، **شاخص هایی از میزان پراکندگی** مقادیر متغیر تصادفی

کوواریانس متغیرهای تصادفی X و Y از رابطه زیر به دست می‌آید؛

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))] = E(X.Y) - E(X).E(Y)$$

❖ مشخصه‌های اصلی توزیع های احتمالاتی

۲- واریانس و کوواریانس

برخی از مهم‌ترین ویژگی‌های واریانس و کوواریانس عبارتند از:

۱) $Var(a+bX)=b^2Var(X)$

۲) $Var(aX+bY)=a^2Var(X)+b^2Var(Y)+2abCov(X,Y)$

۳) $Cov(aX,bY)= abCov(X,Y)$

۴) اگر X و Y مستقل باشند، $Cov(X,Y)=0$

۵) $Cov(X+a,Y+b)= Cov(X,Y)$

جمع متغیرهای تصادفی

رابطه‌هایی درمورد میانگین و واریانس ترکیبی خطی از n متغیر تصادفی(برای استفاده از مبحث توزیع نمونه‌گیری و استنباط آماری)

اگر X_1، X_2، ...، X_n متغیرهای تصادفی مستقل باشند و $Y = \sum_{i=1}^{n} a_i X_i$

که درآن a_iها مقادیر ثابت هستند؛ آنگاه امیدریاضی و واریانس Y از روابط زیر بدست می‌آید؛

$$var(Y) = \sigma_Y^2 = \sum_{i=1}^{n} a_i^2 . var(X_i)$$

$$E(Y) = \mu_Y = \sum_{i=1}^{n} a_i . E(X_i)$$

مثال

فرض کنید سفر روزانه یک شخص از خانه به محل کار از سه بخش تشکیل شده‌است.

	میانگین (دقیقه)	انحراف استاندارد(دقیقه)
۱. خودروشخصی	۱۰	۳
۲. مترو	۴۰	۷
۳. پیاده‌روی	۷	۲

مشاهده می‌شود که مدت زمان سفر با هر یک از مدها، متغیر بوده و دارای میانگین و انحراف استاندارد معینی می‌باشد. مطلوبست محاسبه «میانگین مورد انتظار» و «واریانس» مدت زمان کل سفر؟

مثال

می‌توان گفت که زمان سفر کل(Y)، یک متغیر تصادفی است؛ که از سه متغیر تصادفی دیگر(X_1، X_2 و X_3) تشکیل شده‌است. بنابراین، طبق روابط ارائه شده، میانگین و واریانس زمان سفر کل بصورت زیر بدست می‌آید؛

$$\mu_Y = \sum_{i=1}^{n} a_i . E(X_i) = 10 + 40 + 7 = 57$$

$$var(Y) = \sigma_Y^2 = \sum_{i=1}^{n} a_i^2 . var(X_i) = (1^2 \times 3^2) + (1^2 \times 7^2) + (1^2 \times 2^2) = 62$$

پیوست– Closed-form expression

در ریاضیات، عبارت فرم بسته (Closed-form expression) عبارتی است که بتوان آن را با استفاده از مجموعه‌ای متناهی از عملیات‌ها محاسبه کرد. عبارت‌های فرم بسته می‌توانند شامل ثابت‌ها، متغیرها، برخی عملیات‌های «آشنا» (مثل، + − × ÷), و توابع ریاضی (مثل، ریشه عدد، توان، لگاریتم، توابع مثلثاتی، و تابع وارون هذلولوی) باشند، ولی معمولاً برای محاسبه آن‌ها نیازی به استفاده از حد نیست.

جواب‌های همه معادله‌های درجه دو که ضریب‌هایشان عدد مختلط باشد را می‌توان با جمع (ریاضی), تفریق، ضرب (ریاضی), تقسیم، و ریشه دوم محاسبه کرد. همه این موارد توابع مقدماتی‌اند. به‌عبارت دیگر، معادله درجه دو

$$ax^2 + bx + c = 0,$$

می‌توان با عبارت فرم بسته (Closed-form solution) زیر محاسبه کرد:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$