فصل ٤ متغیرهای تصادفی پیوسته

٤-١ مقدمه

در این فصل متغیرهای تصادفی پیوسته و ویژگی‌های آنها بررسی می‌شوند. بر خلاف متغیر تصادفی گسسته تکیه‌گاه متغیر تصادفی پیوسته شمارا نیست. متغیر تصادفی x را پیوسته گویند هرگاه تکیه‌گاه آن به صورت یک یا اجتماع چند بازه باشد، به عبارت دیگر تکیه‌گاه آن ناشمارا باشد. بسیاری از متغیرهای فیزیکی قابل اندازه‌گیری مانند طول، وزن، دما، فشار و غیره از نوع متغیرهای پیوسته هستند.

٤-٢ تابع چگالی احتمال متغیر تصادفی پیوسته

رفتار احتمالی یک متغیر تصادفی پیوسته توسط تابعی به نام تابع چگالی تعیین می‌شود. یک تابع چگالی تابعی نامنفی است که انتگرال آن روی تکیه‌گاه متغیر تصادفی برابر با ١ است. به عبارتی تابع $f(x)$ در صورتی یک تابع چگالی است که دو شرط

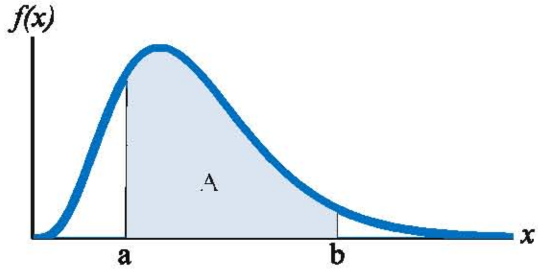
$$f(x) \geq 0 \qquad x \in S_x \qquad (1$$

$$\int_{x \in S_x} (x)dx = 1 \qquad (2$$

صدق کند. در توزیع‌های پیوسته احتمال رخداد پیشامدی مانند A برابر است با انتگرال تابع چگالی در ناحیه A، یا مساحت زیر منحنی تابع چگالی در ناحیه A. به عنوان مثال برای یک تابع چگالی فرضی که شکل آن در نمودار ٤-١ رسم شده است، احتمال اینکه متغیر x بین دو مقدار a و b واقع شود برابر است با سطح زیر منحنی بین این دو نقطه (سطح A) که در ریاضیات به صورت

$$P(a < x < b) = \int_a^b f(x)dx$$

نشان داده می‌شود. با توجه به اینکه برای یک نقطه، هیچ سطحی نمی‌توان متصور شد، در نتیجه در توزیع‌های پیوسته احتمال در نقاط صفر است، یعنی $P(x=a)=۰$. پس در یک توزیع پیوسته رابطه $P(a<x<b)=P(a<x\leq b)=P(a\leq x<b)=P(a\leq x\leq b)=A$ برقرار است.



نمودار ٤-۱. تابع چگالی احتمال $f(x)$ برای متغیر تصادفی پیوسته x.

٤-۳ امید ریاضی و واریانس متغیر تصادفی پیوسته

در فصل پیش بیان شد که امید ریاضی و واریانس متغیر تصادفی گسسته به صورت

$$\mu=E(x)=\sum x.p(x)$$
$$\sigma^{۲}=E(x-\mu)^{۲}=\sum(x-\mu)^{۲}p(x)$$

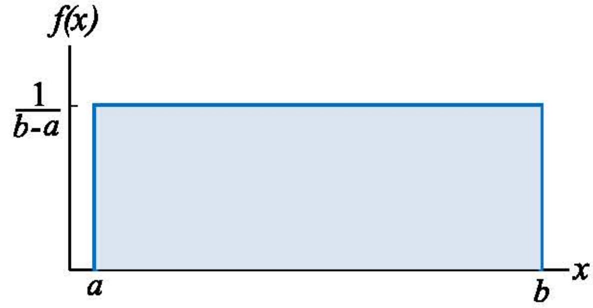
است. چون در تکیه‌گاه متغیرهای تصادفی پیوسته بینهایت مقدار x وجود دارد، میانگین و واریانس توزیع‌های پیوسته از طریق انتگرال‌گیری و به صورت زیر به دست می‌آیند.

$$\mu=E(x)=\int_{-\infty}^{+\infty}x.f(x)dx$$
$$\sigma^{۲}=E(x-\mu)^{۲}=\int_{-\infty}^{+\infty}(x-\mu_x)^{۲}.f(x)dx$$

٤-٤ توزیع یکنواخت یا مستطیلی

هرگاه تابع چگالی احتمال متغیر تصادفی x در محدوده (a,b) برابر عدد ثابت $۱/(b-a)$ و در بقیه نقاط صفر باشد، متغیر تصادفی x دارای توزیع یکنواخت است (نمودار ٤-۲). برای مثال اگر

پزشکی بخواهد در ساعت‌های ۲ تا ۵ بعد از ظهر به طور تصادفی بیماری را معاینه کند، حضور پزشک در هر فاصله یک دقیقه‌ای در طول این ۳ ساعت احتمال یکسانی خواهد داشت.



نمودار ٤-۲. تابع چگالی یکنواخت که مساحت مستطیل مربوطه در آن برابر با $(b-a)(\frac{۱}{b-a})=۱$ است.

سطح مربوط به این توزیع احتمال یعنی مساحت مستطیل حاصل نیز مانند سایر توزیع‌ها برابر با ۱ است. در نتیجه مقادیر مختلف $f(x)$ بر روی محور عمودی باید طوری باشد که این شرط را برقرار کند. توزیع احتمالی یکنواخت ساده‌ترین نوع توزیع است.

تابع چگالی احتمال، میانگین و انحراف معیار در توزیع یکنواخت به صورت زیر است:

$$f(x)=\frac{۱}{b-a} \qquad a\le x\le b$$

$$\mu=\frac{a+b}{۲} \qquad \sigma=\frac{b-a}{\sqrt{۱۲}}$$

مثال ٤-۱. طول یک کانال بتنی آب ۱۲ متر است. فرض کنید متغیر تصادفی x که فاصله یک ترک‌خوردگی از ابتدای کانال را نشان می‌دهد، دارای توزیع یکنواخت باشد، تابع چگالی متغیر تصادفی x و میانگین و انحراف معیار آن به صورت زیر است.

$$f(x)=\frac{۱}{b-a}=\frac{۱}{۱۲-۰}=\frac{۱}{۱۲} \qquad (۰\le x\le ۱۲)$$

$$\mu=\frac{a+b}{۲}=\frac{۰+۱۲}{۲}=۶ \text{ متر}$$

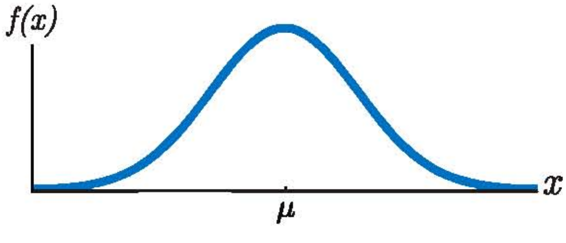
$$\sigma=\frac{b-a}{\sqrt{۱۲}}=\frac{۱۲-۰}{\sqrt{۱۲}}=۳/٤٦٤ \text{ متر}$$

در توزیع یکنواخت احتمال هر پیشامد که در دامنه x قرار بگیرد برابر است با طول آن بازه تقسیم بر دامنه تغییرات x. در مثال بالا احتمال اینکه یک ترک‌خوردگی بیش از دو متر از ابتدای کانال فاصله داشته باشد به صورت زیر است.

$$P(x \geq ۲) = P(۲ \leq x \leq ۱۲) = \frac{۱۲ - ۲}{۱۲ - ۰} = ۰/۸۳۳$$

۴-۵ توزیع نرمال

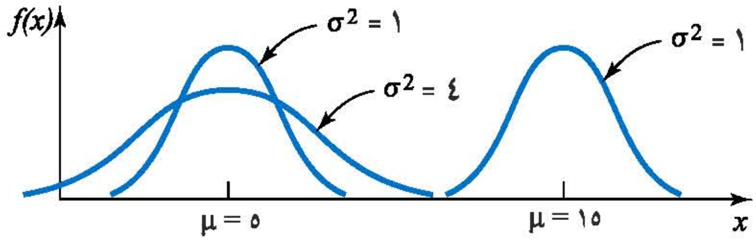
برخی متغیرهای تصادفی پیوسته در طبیعت دارای یک توزیع احتمال زنگوله‌ای شکل و متقارن به نام توزیع نرمال هستند. توزیع نرمال برای اولین بار توسط کارل گوس مطالعه شده و به آن توزیع گاوسی نیز می‌گویند. شکل توزیع نرمال نسبت به میانگین جامعه متقارن است، همچنین میانگین، میانه و مد توزیع با هم برابر هستند. مساحت سطح زیر منحنی نرمال که نمونه‌ای از آن در نمودار ۴-۳ نشان داده شده برابر با یک است.



نمودار ۴-۳. تابع چگالی نرمال $f(x)$ با میانگین μ و انحراف معیار σ.

توزیع نرمال توسط دو پارامتر آن یعنی میانگین (μ) و واریانس $(\sigma^۲)$ مشخص می‌شود و با نماد $x \sim N(\mu, \sigma^۲)$ نشان داده می‌شود.

همانطور که نمودار ۴-۴ نشان می‌دهد با تغییر میانگین توزیع، نمودار بدون آنکه شکل آن تغییر کند، بر روی محور طول جا به جا می‌شود. با تغییر انحراف معیار شکل توزیع نیز تغییر می‌کند. بدین ترتیب که با کاهش انحراف معیار توزیع به سمت بالا کشیده شده و با افزایش آن توزیع پهن‌تر می‌شود.



نمودار ٤-٤. منحنی‌های توزیع نرمال با پارامترهای (میانگین یا واریانس) متفاوت.

فرمول تابع چگالی احتمال منحنی نرمال به صورت زیر می‌باشد.

$$f(x)=\frac{۱}{\sigma\sqrt{۲\pi}}e^{\frac{-(x-\mu)^۲}{۲\sigma^۲}} \qquad -\infty < x < +\infty$$

μ : میانگین متغیر تصادفی x

σ : انحراف معیار

π : عدد پی (تقریباً ۳/۱٤)

e : عدد پندا (تقریباً ۲/۷۱۸)

همانطور که قبلا گفته شد برای محاسبه احتمال یک پیشامد باید در حدود پیشامد از تابع چگالی انتگرال گرفت. برای مثال اگر ارتفاع جامعه بوته‌های کلزا دارای توزیع نرمال با میانگین ٦۸ سانتی‌متر و انحراف معیار ۳ سانتی‌متر باشد یعنی $x \sim N,(٦۸,۹)$، احتمال مشاهده بوته‌هایی با طول ۷۳ تا ۷٦ سانتی‌متر برابر است با مساحت سطح زیر منحنی در فاصله ۷۳ تا ۷٦ که با انتگرال‌گیری به صورت

$$P(۷۳\leq x\leq ۷٦)=\int_{۷۳}^{۷٦}\frac{۱}{\sigma\sqrt{۲\pi}}e^{\frac{-(x-\mu)^۲}{۲\sigma^۲}}dx$$

به دست می‌آید. محاسبه این نوع انتگرال‌ها مشکل است و معمولاً بر اساس روش‌های عددی و به کمک نرم‌افزارها انجام می‌شود. برای راحتی کار می‌توان انتگرال روی محدوده‌های مختلف را در جدولی خلاصه و در ضمیمه کتاب‌های آماری قید کرد، از طرفی چون مقدار این انتگرال‌ها به پارامترهای توزیع یعنی میانگین و انحراف معیار بستگی دارد به تعداد نامتناهی جدول نیاز است.

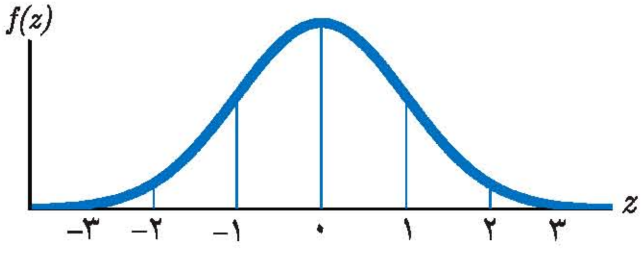
یک راه حل این است که انتگرال‌های مذکور برای یک حالت ساده مثلاً میانگین برابر صفر و انحراف معیار برابر یک محاسبه و احتمال توزیع‌های دیگر از روی آن به دست آید.

٤-٦ توزیع نرمال استاندارد

در فصل اول به استاندارد کردن متغیرها اشاره شد. استاندارد کردن روشی است که متغیر اصلی را با هر میانگین و واریانسی، به متغیری استاندارد با میانگین صفر و واریانس یک تبدیل می‌کند. اگر $x \sim N(\mu,\sigma^{۲})$ آنگاه $z=(x-\mu)/\sigma$ که یک ترکیب خطی روی x است، دارای توزیع نرمال $N(۰,۱)$ می‌باشد (نمودار ٤-٥). این توزیع نرمال خاص را توزیع نرمال استاندارد گویند و فرمول تابع چگالی احتمال آن به صورت

$$f(z)=\frac{۱}{\sqrt{۲\pi}}e^{\frac{-z^{۲}}{۲}} \qquad , \qquad -\infty < z < +\infty$$

است. در جدول ضمیمه ۲ در آخر کتاب، احتمال مشاهده zهای بزرگتر از هر z مثبت آورده شده است. با توجه به تقارن توزیع نرمال و این مطلب که کل مساحت زیر منحنی برابر یک است، احتمال پیشامدهای دیگر نیز به راحتی محاسبه می‌شوند. شکل توزیع نرمال استاندارد به صورت زیر است.



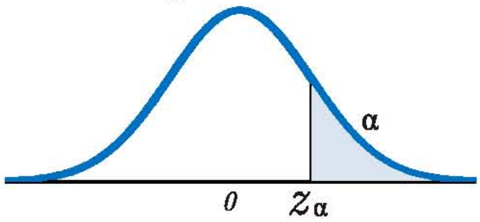
نمودار ٤-٥. توزیع نرمال استاندارد.

٤-٦-٢ محاسبه احتمال در توزیع نرمال استاندارد و نرمال

در جدول نرمال استاندارد (جدول ضمیمه ۲) که قسمتی از آن در جدول ٤-١ آمده است، احتمال تجمعی در دنباله راست به ازای zهای مثبت قید شده است. لذا می‌توان احتمال مشاهده

مقادیر مختلف z را به دست آورد. مقادیر z تا یک رقم اعشار در ستون سمت چپ و صدم آن در ردیف اول آورده شده است. فرض نمایید احتمال مشاهده zهای مساوی یا بزرگتر از ۱/۹۶ را در توزیع نرمال استاندارد یعنی $P(z \geq ۱/۹۶)$ مد نظر است. این احتمال برابر است با مساحت سطح زیر منحنی نرمال استاندارد واقع در دامنه ۱/۹۶+ تا ∞+. با توجه به اینکه این سطح در جدول z برای ۱/۹۶ تا بینهایت برابر با ۰/۰۲۵ می‌باشد، لذا $P(z \geq ۱/۹۶) = ۰/۰۲۵$.

جدول ٤-۱. قسمتی از جدول توزیع نرمال استاندارد.

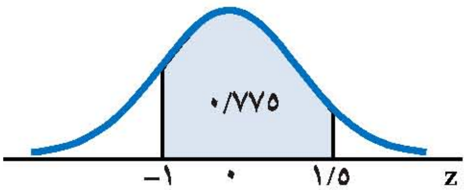


z	۰/۰۰	۰/۰۱	۰/۰۲	۰/۰۳	۰/۰٤	۰/۰۵	۰/۰٦	۰/۰۷	۰/۰۸	۰/۰۹
۰/۰	۰/۵۰۰۰	۰/٤۹٦۰	۰/٤۹۲۰	۰/٤۸۸۰	۰/٤۸٤۰	۰/٤۸۰۱	۰/٤۷٦۱	۰/٤۷۲۱	۰/٤٦۸۱	۰/٤٦٤۱
۰/۱	۰/٤٦۰۲	۰/۵٤٦۲	۰/٤۵۲۲	۰/٤٤۸۳	۰/٤٤٤۳	۰/٤٤۰٤	۰/٤۳٦٤	۰/٤۳۲۵	۰/٤۲۸٦	۰/٤۲٤۷
۰/۲	۰/٤۲۰۷	۰/٤۱٦۸	۰/٤۱۲۹	۰/٤۰۹۰	۰/٤۰۵۲	۰/٤۰۱۳	۰/۳۹۷٤	۰/۳۹۳٦	۰/۳۸۹۷	۰/۳۸۵۹
۰/۳	۰/۳۸۲۱	۰/۳۷۸۳	۰/۳۷٤۵	۰/۳۷۰۷	۰/۳٦٦۹	۰/۳٦۳۲	۰/۳۵۹٤	۰/۳۵۵۷	۰/۳۵۲۰	۰/۳٤۸۳
۰/٤	۰/۳٤٤٦	۰/۳٤۰۹	۰/۳۳۷۲	۰/۳۳۳٦	۰/۳۳۰۰	۰/۳۲٦٤	۰/۳۲۲۸	۰/۳۱۹۲	۰/۳۱۵٦	۰/۳۱۲۱
•	.	.	.	.	.	.	.	.	.	.
•	.	.	.	.	.	.	.	.	.	.
۱/۹	.	.	.	.	.	.	۰/۰۲۵۰	.	.	.
•	.	.	.	.	.	.	.	.	.	.

به طور مشابه این احتمال برای عدد منفی ۱/۹۶- یعنی $P(z \leq -۱/۹۶)$ با توجه به قرینه بودن منحنی احتمال، برابر با ۰/۰۲۵ است. با توجه به اینکه کل سطح زیر منحنی استاندارد برابر با یک است، می‌توان استنباط کرد که در توزیع نرمال استاندارد ٪۹۵ داده‌ها یا (۰/۰۲۵-۰/۰۲۵-۱) سطح زیر منحنی در فاصله ۱/۹۶+ = z تا ۱/۹۶- = z قرار دارند. در حالت کلی رابطه $P(a \leq z \leq b) = P(z \geq a) - P(z \geq b)$ برقرار است. برای هر عدد مثبت مانند a احتمال $P(z \geq a)$ مستقیماً از روی جدول محاسبه و اگر a منفی باشد با توجه به متمم پیشامد و قرینگی توزیع بصورت $P(z \geq a) = ۱ - P(z \leq a) = ۱ - P(z \geq -a)$ محاسبه می‌شود که در آن a- یک عدد مثبت بوده و $P(z \geq -a)$ از روی جدول به دست می‌آید.

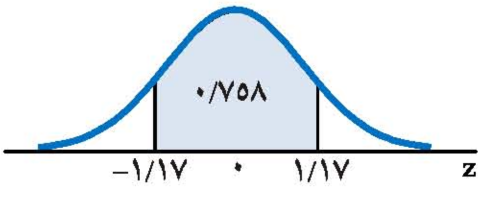
مثال ٤-۲. احتمال مشاهده zهای بزرگتر از ۱- و کوچکتر از ۱/۵+ یعنی $P(-۱ \leq z \leq +۱/۵)$ را به دست آورید.

در ابتدا این احتمال را بر روی منحنی نرمال استاندارد سایه می‌زنیم. از آنجا که سطح زیر منحنی برابر با ۱ است لذا سطح سایه‌دار مورد نظر برابر است با ۱ منهای سطوح سفید در سمت راست و چپ منحنی تابع چگالی. با مراجعه به جدول به راحتی می‌توان احتمال $P(z \geq +۱/۵)$ یعنی فضای سفید سمت راست را به دست آورد. مساحت فضای سفید سمت چپ نیز، $P(z \leq -۱)$، با توجه به تقارن منحنی برابر با $P(z \geq ۱)$ است که با مراجعه به جدول z به دست می‌آید.



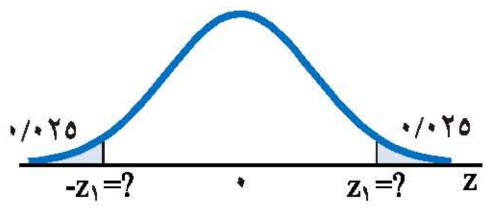
$$
\begin{aligned}
&P(-۱ \leq z \leq ۱/۵) \\
&= P(z \geq -۱) - P(z \geq ۱/۵) \\
&= ۱ - P(z \geq ۱) - P(z \geq ۱/۵) \\
&= ۱ - ۰/۱۵۹ - ۰/۰۶۶۸ = ۰/۷۷۵
\end{aligned}
$$

مثال ٤-۳. احتمال مشاهده zهای بزرگتر از ۱/۱۷- و کوچکتر از ۱/۱۷+ را بدست آورید.



$$
\begin{aligned}
&P(-۱/۱۷ \leq z \leq ۱/۱۷) \\
&= P(z \geq -۱/۱۷) - P(z \geq ۱/۱۷) \\
&= ۱ - P(z \geq ۱/۱۷) - P(z \geq ۱/۱۷) \\
&= ۱ - ۲P(z \geq ۱/۱۷) \\
&= ۱ - ۲(۰/۱۲۱) = ۰/۷۵۸
\end{aligned}
$$

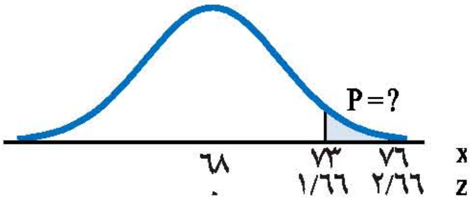
مثال ٤-٤. اگر $P(|z| \geq z_۱) = ۰/۰۵$ باشد مقدار $z_۱$ را بدست آورید. مساحت سطوح سایه خورده در دو طرف برابر با ۰/۰۵ هستند، بنابراین می‌توان نوشت



$$
\begin{aligned}
&P(|z| > z_۱) \\
&= P(z \leq -z_۱) + P(z \geq z_۱) = ۰/۰۵ \\
&P(z \geq z_۱) = ۰/۰۲۵ \\
&z_۱ = ۱/۹۶
\end{aligned}
$$

اگر $x \sim N(\mu,\sigma^{۲})$ آنگاه $z = (x-\mu)\,/\,\sigma$ که یک ترکیب خطی روی x است، دارای توزیع نرمال $N(۰,۱)$ می‌باشد.

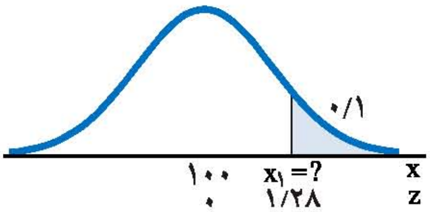
مثال ۴-۵. اگر طول بوته‌های کلزا دارای توزیع نرمال با میانگین ۶۸ و انحراف معیار ۳ سانتی‌متر باشد، احتمال مشاهده بوته‌هایی با طول ۷۳ تا ۷۶ سانتی‌متر را بدست آورید.



$$P(۷۳ \le x \le ۷۶)$$
$$= P(\frac{۷۳-\mu}{۳} \le z \le \frac{۷۶-\mu}{\sigma})$$
$$= P(\frac{۷۳-۶۸}{۳} \le z \le \frac{۷۶-۶۸}{۳})$$
$$= P(۱\,/\,۶۶ \le z \le ۲/۶۶)$$
$$= P(z \ge ۱/۶۶) - P(z \ge ۲/۶۶)$$
$$= ۰/۰۴۸ - ۰/۰۰۴ = ۰/۰۴۴$$

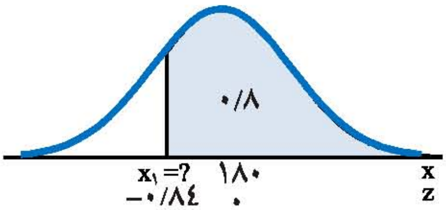
مثال ۴-۶. اگر ارتفاع بوته‌ها در مزرعه گندم متغیری تصادفی با توزیع نرمال با $\mu = ۱۰۰$ و $\sigma = ۱۰$ سانتی‌متر باشد، حداقل طول ۱۰٪ از بلندترین بوته‌ها چند سانتی‌متر است؟

ابتدا ۱۰٪ مساحت سطح زیر منحنی را در سمت راست سایه می‌زنیم. بدین ترتیب ۱۰٪ از بلندترین بوته‌ها را مشخص کرده‌ایم. آنچه در این مثال مجهول است x یا نقطه مربوط به سطح سایه خورده است. با نگاه به جدول نرمال استاندارد، z مربوط به سطح ۰/۱ یعنی ۱/۲۸ به راحتی پیدا می‌شود. با تبدیل این z به x جواب مورد نظر مشخص می‌شود.



$$P(x \ge x_۱) = ۰/۱$$
$$P(\frac{x-\mu}{\sigma} > \frac{x_۱-\mu}{\sigma}) = P(z \ge z_۱) = ۰/۱$$
$$\Rightarrow z_۱ = ۱/۲۸, \quad \frac{x_۱ - ۱۰۰}{۱۰} = ۱/۲۸, \quad x_۱ = ۱۱۲/۸$$

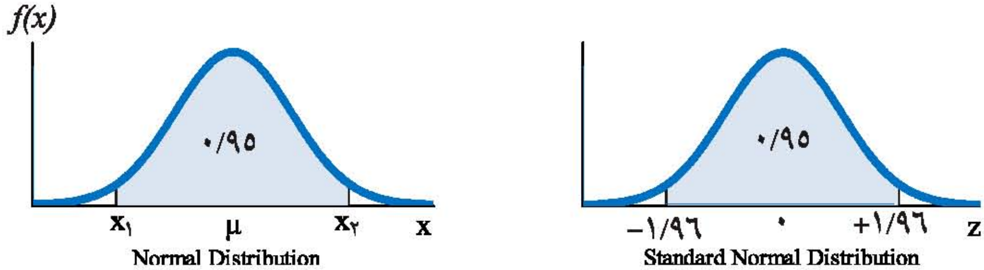
مثال ۴-۷. اگر ارتفاع بوته‌های کلزا دارای توزیع نرمال با میانگین ۱۸۰ و انحراف معیار ۲۰ سانتی‌متر باشد، طولی را که ۸۰٪ گیاهان از آن بیشتر هستند، بدست آورید.



$$P(x \geq x_۱) = P(\frac{x-\mu}{۳} \geq \frac{x_۱-\mu}{\sigma})$$
$$= P(z \geq z_۱) = ۰/۸ \quad \Rightarrow \quad P(z \geq z'_۱) = ۰/۲$$
$$z'_۱ = +۰/۸٤ \quad \Rightarrow z_۱ = -۰/۸٤$$
$$\frac{x_۱ - ۱۸۰}{۲۰} = -۰/۸٤ \quad \Rightarrow x_۱ = ۱٦۳/۲$$

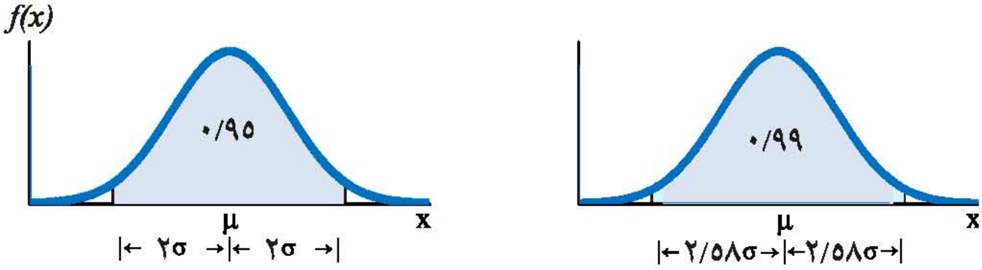
٤-۷ فاصله ۹۵٪ و ۹۹٪ در توزیع نرمال

اغلب دانستن این مطلب که در یک توزیع نرمال با میانگین μ و انحراف معیار σ، ۹۵٪ یا ۹۹٪ از داده‌ها در چه فاصله‌ای از میانگین قرار دارند، مفید و توصیف مناسبی از جامعه است. این فاصله‌ها در نمودار ٤-٦ نشان داده شده‌اند.



نمودار ٤-٦. فاصله ۹۵٪ در توزیع نرمال استاندارد (سمت راست) و توزیع نرمال (سمت چپ).

سطح سایه خورده در داخل توزیع نرمال استاندارد برابر با ۹۵٪ است. اگر سطح سایه‌دار توزیع نرمال (سمت چپ) نیز برابر با ۹۵٪ باشد، $x_۱$ و $x_۲$ در توزیع نرمال به ترتیب معادل ۱/۹٦ و ۱/۹٦- در توزیع نرمال استاندارد خواهند بود یعنی $(x_۱-\mu)/\sigma = ۱/۹٦$ و $(x_۲-\mu)/\sigma = -۱/۹٦$. از اینجا $x_۱$ برابر با $\mu + ۱/۹٦\sigma$ و $x_۲$ برابر با $\mu - ۱/۹٦\sigma$ خواهد بود. به عبارتی در یک توزیع نرمال ۹۵٪ داده‌ها در فاصله $\mu - ۱/۹٦\sigma$ تا $\mu + ۱/۹٦\sigma$ قرار دارند. به همین ترتیب می‌توان دریافت که ۹۹٪ داده‌ها در فاصله $\mu - ۲/۵۸\sigma$ تا $\mu + ۲/۵۸\sigma$ قرار دارند.



نمودار ۴-۷. در یک توزیع نرمال تقریباً ۹۵٪ داده‌ها در فاصله $\mu - ۲\sigma$ تا $\mu + ۲\sigma$ و ۹۹٪ داده‌ها در فاصله $\mu - ۲/۵۸\sigma$ تا $\mu + ۲/۵۸\sigma$ قرار دارند.

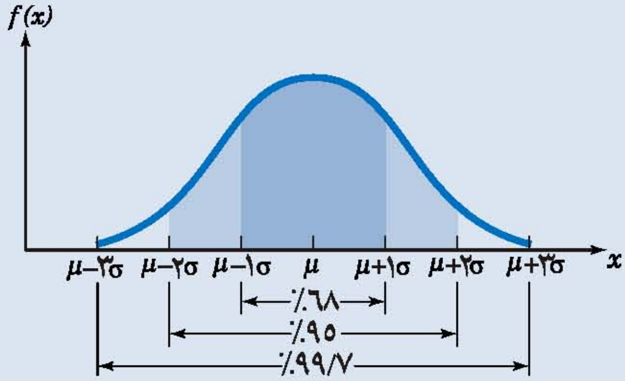
فاصله ۹۵٪ و ۹۹٪ در توزیع نرمال

در یک توزیع نرمال ۹۵٪ داده‌ها در فاصله $\mu - ۱/۹۶\sigma$ تا $\mu + ۱/۹۶\sigma$ و ۹۹٪ داده‌ها در فاصله $\mu - ۲/۵۸\sigma$ تا $\mu + ۲/۵۸\sigma$ قرار دارند، یعنی

$$P(\mu - ۱/۹۶\sigma \leq x \leq \mu + ۱/۹۶\sigma) = ۹۵\%$$

$$P(\mu - ۲/۵۸\sigma \leq x \leq \mu + ۲/۵۸\sigma) = ۹۹\%$$

گاهی فاصله ۹۵٪ در توزیع نرمال را به طور تقریبی به صورت $\mu - ۲\sigma \leq x \leq \mu + ۲\sigma$ یا $\mu \pm ۲\sigma$ در نظر می‌گیرند. همچنین با توجه به اینکه ۹۹/۷٪ داده‌ها در فاصله $\mu - ۳\sigma \leq x \leq \mu + ۳\sigma$ یا $\mu \pm ۳\sigma$ قرار می‌گیرند، می‌توان فواصل مذکور را به صورت زیر نشان داد.

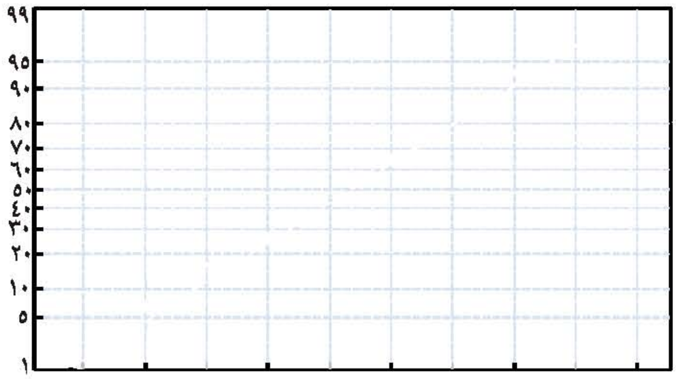


٤-۸ بررسی نرمال بودن داده‌ها

روش‌های مختلفی برای بررسی نرمال بودن داده‌ها وجود دارد. ساده‌ترین روش مشاهده نمودار توزیع است که در صورت نرمال بودن داده‌ها بافت‌نگار یا نمودار ساقه‌برگی آن متقارن و شبیه منحنی نرمال خواهد بود. روش بعدی دامنه میان‌چارکی استاندارد (IQR) از تقسیم تفاوت چارک اول و چارک سوم بر انحراف معیار یعنی $(Q_3 - Q_1) / \sigma$ محاسبه می‌شود. این معیار برای داده‌های نرمال با هر میانگین و واریانسی برابر عدد ۱/۳٤ است. اگر IQR برای مجموعه‌ای از داده‌ها نزدیک به ۱/۳٤ باشد شواهدی مبنی بر نرمال بودن داده‌ها وجود دارد. این مقدار برای توزیع نرمال استاندارد به صورت زیر محاسبه می‌شود.

$$IQR = \frac{Q_3 - Q_1}{\sigma} = \frac{۰/٦۷ - (-۰/٦۷)}{۱} = ۱/۳٤$$

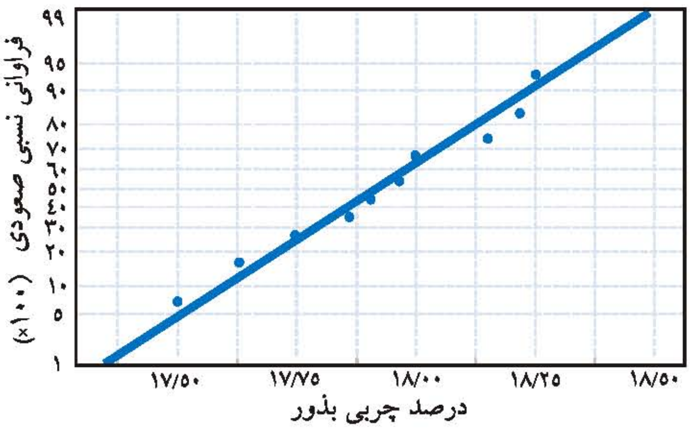
همچنین فرض نرمال بودن داده‌ها را می‌توان با رسم نمودار چندک-چندک، مقادیر مشاهده شده بر روی کاغذ مخصوصی با نام کاغذ احتمال بررسی کرد. برای رسم نمودار چندک-چندک، ابتدا داده‌ها را از کوچک به بزرگ در یک ردیف مرتب کرده و در ردیفی دیگر فراوانی نسبی صعودی را بر اساس فرمول $(i - ۰/۵) / n$ برای آنها محاسبه می‌شود. اگر مشاهدات و مقادیر فراوانی نسبی صعودی به دست آمده را بر روی کاغذ احتمال نرمال (نمودار ٤-۸) به صورت نقاطی رسم شوند، در صورت نرمال بودن داده‌ها، نقاط بر روی یک خط مستقیم قرار خواهند گرفت. در غیر این صورت هر گونه انحراف معنی‌دار از خط مستقیم نشان‌دهنده نرمال نبودن داده‌ها است.



نمودار ٤-۸ کاغذ احتمال نرمال.

مثال ٤-۸ درصد چربی در ۱۰ عدد از نوعی بذر به صورت زیر بوده است. فرض نرمال بودن توزیع مقادیر درصد چربی را در جامعه بذور بررسی کنید.

i	۱	۲	۳	٤	٥	٦	۷	۸	۹	۱۰
x_i	۱۷/٥۰	۱۷/٦۳	۱۷/۷٥	۱۷/۸٦	۱۷/۹۰	۱۷/۹٦	۱۸/۰۰	۱۸/۱٥	۱۸/۲۲	۱۸/۲٥
$(i-۰/٥)/n$	۰/۰٥	۰/۱٥	۰/۲٥	۰/۳٥	۰/٤٥	۰/٥٥	۰/٦٥	۰/۷٥	۰/۸٥	۰/۹٥



نمودار ٤-۹. نمودار چندک-چندک حاصل از داده‌های مثال ٤-۸.

نمودار ٤-۹ که با استفاده از نرم‌افزار Minitab به دست آمده است، نشان می‌دهد که نقاط تقریباً در راستای یک خط مستقیم قرار گرفته و از این رو شواهد مبنی بر نرمال بودن مشاهدات وجود دارد. در کل انحرافات کوچک تا متوسط از خط مستقیم نادیده گرفته شده و توزیع فراوانی داده‌ها، نرمال در نظر گرفته می‌شوند. البته این نمودار در صورتی معتبر است که تعداد داده‌ها حداقل ۲۰ باشد در غیر این صورت حتی اگر داده‌ها از جامعه‌ای نرمال آمده باشند، ممکن است نقاط از خط راست انحراف نشان دهند. این نمودار دو نکته دیگر را هم مشخص می‌کند: ۱. شیب خط با واریانس مشاهدات رابطه عکس دارد، یعنی هر اندازه شیب خط بیشتر باشد، واریانس داده‌ها کمتر است. ۲. اگر نقطه‌ای فاصله زیادی از بقیه نقاط داشته باشد حتی اگر در امتداد آنها (خط راست) هم باشد، به عنوان داده پرت در نظر گرفته می‌شود.

روش‌های بررسی نرمال بودن که در بالا بیان شدند تا حدودی سلیقه‌ای هستند، مثلا ممکن است شخصی عدد ۱/٥ را نزدیک به ۱/۳٤ بداند و شخص دیگری نظر برعکس داشته باشد. در نمودار چندک-چندک نیز ممکن است در مورد نزدیکی نقاط به خط اختلاف نظر وجود داشته باشد. یک روش آماری برای حل این اختلاف نظرها استفاده از آزمون فرض نرمال بودن داده‌ها با

اطمینان خاصی است. با چنین آزمونی که در آن ضریب همبستگی نقـاط (فصـل ۱۱) بـه عنـوان شاخصی از مستقیم بودن خط محاسبه می‌شود، می‌توان استنباط نمود که داده‌ها با اطمینان خاصی (مثلا ٪۹۵) نرمال یا غیر نرمال هستند. روش‌های مختلفی بـرای اجـام ایـن آزمـون در نـرم‌افـزار Minitab در مسیر ... $Stat > Basic\ Statistics > Normality\ Test$ وجود دارد.

تمرین‌ها

۴–۱. فرض کنید x متغیری تصادفی با توزیع نرمال با $\mu = ۳۰$ و $\sigma = ۵$ است. نمره z یا نمره استاندارد را برای هر کدام از xهای مورد نظر را به دست آورید. **الف)** $x = ۳۰$ ب) $x = ۲۰$ د) $x = ۳۵$.

۴–۲. فرض کنید x متغیری دارای توزیع نرمال با $\mu = ۱۰$ و $\sigma = ۲$ است. احتمـالات خواسـته شـده را بـه دست آورید. **الف)** $P(۱۰ \leq x \leq ۱۲)$ ب) $P(۶ \leq x \leq ۹)$ ج) $P(x \leq ۵)$ د) $P(x \geq ۱۳)$.

۴–۳. فرض کنید x متغیری تصادفی با توزیع نرمال با $\mu = ۳۰۰$ و $\sigma = ۳۰$ است. **الف)** با چه احتمالی x در خارج از فاصله دو برابر انحراف معیار از میانگین قرار مـی‌گیـرد؟ (بـه عبـارتی احتمـال اینکـه x کـوچکتر از $\mu - ۲\sigma$ یا بزرگتر از $\mu + ۲\sigma$ باشد چقدر است؟) ب) با چه احتمالی x در خـارج از فاصـله یـک انحـراف معیار از میانگین قرار می‌گیرد؟ ج) دهک هشتم را در توزیع مذکور به دست آورید.

۴–۴. فرض کنید x متغیری دارای توزیع نرمال با میانگین ۱۰۰ و انحراف معیار برابر با ۸ است. **الف)** منحنـی مربوطه به این توزیع را رسم کرده و نقطـه μ را بـر روی آن مشـخص کنیـد. **ب)** فاصـله $\mu \pm \sigma$ را بـر روی منحنی مشخص کنید. ج) $P(\mu - ۲\sigma \leq x \leq \mu + ۲\sigma)$ را به دست آوریـد. د) $P(x \geq \mu + ۲\sigma)$ را بـه دست آورید. ه) $P(۹۲ \leq x \leq ۱۱۶)$ را به دست آورید. و) $P(۷۶ \leq x \leq ۱۲۴)$ را به دست آورید.

۴–۵. تحقیقات نشان داده است که طول ماهی ساردین دوسـاله دارای توزیـع تقریبـاً نرمـال بـا $\mu = ۲۰/۲$ سانتی‌متر و انحراف معیار ۰/۶۵ سانتی‌متر است. **الف)** با چه احتمالی طول یـک مـاهی دو سـاله بـین ۲۰ و ۲۱ سانتی‌متر است؟ **ب)** طول یک ماهی صید شده ۱۹/۸ سانتی‌متر است. آیا این ماهی، دوسـاله اسـت؟ (راهنمـایی: احتمال کمی (٪۵) وجود دارد که طول یک ماهی ساردین دو ساله خارج از محـدوده $\mu - ۲\sigma$ تـا $\mu + ۲\sigma$ قرار گیرد. پس در صورت قرارگیری طول در خارج از این فاصله، ماهی مربوطه دو ساله به حسـاب نمـی‌آیـد)، ج) اگر طول ماهی صید شده ۲۲ سانتی‌متر باشد آیا می‌توان گفت که دوساله است؟

۴-۶. انحراف معیار داده‌های نمونه‌ای برابر با ۵۰، چارک اول برابر با ۱۱ $= Q_1$ و چارک سوم برابر با ۸۰ $= Q_3$ می‌باشد. **الف)** آیا حاصل دامنه میان‌چارکی استاندارد (IQR) تقریباً برابر با ۱/۳۴ است؟ **ب)** این مقدار دلالت بر چه ویژگی از توزیع داده‌ها دارد؟

۴-۷. داده‌های زیر را در نظر بگیرید:

۲۱	۴۰	۶۷	۹۷	۴۳	۸۲	۶۴
۱۶	۱۲	۸۹	۷۶	۳۳	۶۵	۹۴
۵۹	۷۴	۷۳	۴۵	۳۵	۸۴	۱۱
۵۳	۸۶	۸۴	۶۳	۱۱	۱۶	۵۰

الف) مقدار s را در این نمونه محاسبه کنید. **ب)** با استفاده از چارک اول و چارک سوم و انحراف معیار مشخص کنید که آیا داده‌ها تقریباً نرمال هستند یا خیر **ج)** نمودار چندک-چندک را برای داده‌ها با استفاده از نرم‌افزار $Minitab$ به دست آورید و بیان کنید که آیا نرمال بودن داده‌ها تأیید می‌شود یا خیر.

۴-۸. تحقیقی نشان داده است که نوع خاصی از یک سرطان پوست ۶۰٪ مبتلایان را سالانه از بین می‌برد. در یک نمونه ۱۰۰۰۰ نفری، **الف)** امید و واریانس x (افراد مبتلا که در اثر سرطان پوست می‌میرند) را به دست آورید. **ب)** احتمال اینکه در یک نمونه ۱۰۰۰۰ نفری x بیش از ۶۰۴۰ نفر در سال باشد، چقدر است؟

۴-۹. فرض کنید وزن ۶۰۰ نفر دارای توزیع نرمال با میانگین ۶۰ و انحراف معیار ۵ کیلوگرم باشد. مطلوب است تعداد افرادی که وزن آنها بیش از ۵۷ و کمتر از ۷۰ کیلوگرم است.

۴-۱۰. یک محموله بزرگ از کالایی آماده ارسال برای خریدار است. نماینده خریدار به شرطی محموله را قبول می‌کند که در یک نمونه ۱۰ تایی کمتر از ۲ کالای معیوب وجود داشته باشد. اگر محموله ۵٪ کالای معیوب داشته باشد احتمال قبول محموله توسط نماینده خریدار چقدر است؟

۴-۱۱. اگر متغیر وزن بذر ذرت دارای توزیع نرمال با ۵=μ و ۱۶=σ^2 باشد، احتمالات خواسته شده را به دست آورید: **الف)** $P(x \leq 2)$ **ب)** $P(0 \leq x \leq 15)$ **ج)** اگر $P(x \leq x_1) = 0/95$ باشد مقدار x_1 را به دست آورید.