آمار واحتمال مهندسی

رشته برق

جلسه دوم

ساعت کلاس: چهارشنبه ۱۳:۳۰ الی ۱۵:۴۵

استاد: آروین مهدوی

نیمسال دوم ۹۹–۹۸

۳.۱ جدولهای آماری

نمایش داده‌ها را، با نظمی خاص، در چند سطر و ستون یک جدول آماری می‌گویند. یک جدول آماری باید به نحوی تنظیم شود که بتوان به آسانی پاره‌ای از دانسته‌های نهفته در داده‌ها را از روی آن خواند. شماره جدول، نام جدول، عنوانهای سطر و ستون، زیرنویس، و بالاخره مأخذ جدول در صورت لزوم باید روشن و گویا باشند. جدولی که از روی تمام داده‌ها به دست می‌آید، جدول اصلی و جدولی که از روی جدول اصلی، برای بررسی دانسته‌هایی ویژه مشتق می‌شود، جدول فرعی نامیده می‌شود. در کارهای آماری برای این که داده‌ها را خلاصه کنند. آنها را در جدولی به نام جدول فراوانی تنظیم می‌کنند.

۱.۳.۱ فراوانی و فراوانی نسبی

هرگاه n چیز از k نوع $T_۱, T_۲, \cdots, T_k$، با فرض $n \geq k \geq ۲$، به ترتیب با تعدادهای $f_۱, f_۲, \cdots, f_k$ تشکیل شده باشند، این تعدادها را فراوانیها و $\frac{f_۱}{n}, \frac{f_۲}{n}, \cdots, \frac{f_k}{n}$ را فراوانیهای نسبی این چیزها می‌گوئیم. فراوانیهای نسبی را به ترتیب با $r_۱, r_۲, \cdots, r_k$ نشان می‌دهند. واضح است که برای $i = ۱, ۲, \cdots, k$ داریم:

$$\sum_{i=۱}^{k} f_i = n \ , \ ۱ \leq f_i < n \ , \ \sum_{i=۱}^{k} r_i = ۱ \ , \ \frac{۱}{n} \leq r_i < ۱ \ ,$$

۲.۳.۱ فراوانی انباشته و فراوانی نسبی انباشته

با توجه به فراوانی و فراوانی نسبی، برای $i = ۱, ۲, \cdots, k$،

$$s_i = \sum_{j=۱}^{i} r_j \quad , \quad g_i = \sum_{j=۱}^{i} f_j$$

را به ترتیب، فراوانی انباشته و فراوانی نسبی انباشته می‌گویند. واضح است که

$$s_k = ۱ \quad , \quad g_k = n$$

مثال ۱ اگر گروهی ۱۰ نفری از ۳ شیرازی، ۵ تهرانی، و ۲ کرمانی تشکیل شده باشد، فراوانی شیرازیها، تهرانیها و کرمانیها در این گروه به ترتیب ۳، ۵ و ۲ و فراوانی نسبی آنها ۰/۳، ۰/۵،و ۰/۲ می‌باشد. فراوانی انباشته و فراوانی نسبی انباشته شیرازیها و تهرانیها به ترتیب ۸ و ۰/۸ است.

۳.۳.۱ جدول فراوانی

هرگاه داده‌ها را به نحوی تقسیم کرده، آنها را بر حسب فراوانیها در جدولی تنظیم کنیم، این جدول را جدول فراوانی یا جدول توزیع فراوانی می‌نامند. برای اینکه جدول فراوانی را با توجه به داده‌ها از لحاظ داده‌های گسسته و پیوسته، شرح دهیم به مثالهای زیر توجه کنید:

مثال ۲ پزشکی ۲۰ بیمار قلبی دارد که گروه خونی آنها عبارتند از:

$$B \ A \ O \ AB \ O \ A \ A \ A \ O \ O \ A \ A \ B \ B \ AB \ O \ AB \ AB \ O \ O$$

جدول ۱ جدول فراوانی گروه خونی ۲۰ بیمار قلبی با متغیر اسمی x_i

گروه خونی	x_i	f_i	r_i	g_i	s_i
A	1	6	۰/۳۰	۶	۰/۳۰
B	2	3	۰/۱۵	۹	۰/۴۵
AB	3	4	۰/۲۰	۱۳	۰/۶۵
O	4	7	۰/۳۵	۲۰	۱/۰۰
		۲۰	۱/۰۰		

می‌خواهیم از این داده‌ها یک جدول فراوانی بسازیم.

نخست چهار گروه خونی $A,\ B,\ AB,\ O$ را به ترتیب با اعداد ۴، ۳، ۲، ۱ متناظر می‌کنیم تا داده‌های اسمی به دست آیند. این اعداد صرفاً برای گروه‌بندی و نامگذاری است و نمی‌توان روی آنها چهار عمل اصلی را انجام داد. بعد از شمارش لازم، جدول ۱، به نام جدول گروه خونی ۲۰ بیمار قلبی به دست می‌آید. داده‌های این مثال نمونه‌ای از داده‌های گروهی است که در واقع داده‌های گسسته می‌باشند.

مثال ۳ داده‌های زیر تعداد فرزندان ۲۰۰ خانواده در یکی از شهرهای ایران می‌باشند:

۳	۵	۴	۲	۱	۰	۴	۳	۵	۲	۵	۲	۰	۱	۶	۴	۳	۰	۱	۱
۳	۱	۵	۶	۴	۳	۲	۲	۰	۲	۳	۱	۲	۳	۲	۵	۶	۲	۴	۴
۲	۲	۵	۲	۴	۰	۲	۵	۶	۴	۳	۱	۲	۱	۰	۲	۵	۳	۴	۲
۱	۰	۳	۲	۰	۱	۲	۵	۶	۴	۳	۰	۱	۲	۶	۴	۱	۵	۳	۴
۴	۲	۵	۱	۳	۳	۰	۲	۱	۰	۴	۲	۴	۱	۵	۳	۲	۰	۴	۲
۳	۵	۱	۶	۲	۰	۵	۳	۲	۳	۲	۲	۰	۱	۵	۲	۳	۳	۰	۱
۱	۲	۴	۳	۲	۱	۲	۳	۲	۶	۱	۰	۴	۲	۳	۵	۴	۳	۲	۴
۳	۲	۵	۵	۲	۴	۳	۲	۳	۴	۲	۱	۰	۴	۳	۶	۲	۴	۳	۲
۲	۴	۱	۶	۵	۴	۲	۳	۱	۵	۴	۳	۴	۲	۱	۰	۲	۳	۱	۲
۳	۳	۲	۱	۳	۳	۲	۳	۲	۳	۱	۴	۴	۳	۰	۱	۲	۰	۱	۳

در این مثال ویژگی مورد مطالعه، تعداد فرزندان یک خانواده است، که با اعداد ۰، ۱، ۲، ۳، ۴، ۵، و ۶ مشخص شده است. این اعداد برای گروه‌بندی به کار می‌روند و جنبه ترتیبی دارند. داده‌های بالا

جدول ۲ جدول فراوانی تعداد فرزندان ۲۰۰ خانواده با متغیر ترتیبی x_i

x_i	f_i	r_i	g_i	s_i
0	۲۰	۰٫۱۰	۲۰	۰٫۱۰
1	۳۰	۰٫۱۵	۵۰	۰٫۲۵
2	۵۰	۰٫۲۵	۱۰۰	۰٫۵۰
3	۴۰	۰٫۲۰	۱۴۰	۰٫۷۰
4	۳۰	۰٫۱۵	۱۷۰	۰٫۸۵
5	۲۰	۰٫۱۰	۱۹۰	۰٫۹۵
6	۱۰	۰٫۰۵	۲۰۰	۱٫۰۰
	۲۰۰	۱٫۰۰		

نمونه‌ای از داده‌های گسسته هستند که در جدول ۲، به نام جدول فراوانی تعداد فرزندان ۲۰۰ خانواده خلاصه شده‌اند.

مثال ۴ داده‌های زیر اندازه‌های قد ۱۰۰ جوان بیست ساله در یکی از شهرهای ایران می‌باشند، که بر حسب سانتیمتر تا نزدیکترین واحد سر راست شده‌اند.

۱۷۲	۱۸۴	۱۷۰	۱۵۰	۱۶۴	۱۷۱	۱۷۸	۱۷۲	۱۷۷	۱۵۱
۱۵۸	۱۶۵	۱۶۹	۱۷۲	۱۶۶	۱۶۴	۱۶۹	۱۵۹	۱۵۳	۱۷۰
۱۶۲	۱۵۴	۱۵۹	۱۸۲	۱۷۴	۱۶۲	۱۵۱	۱۶۵	۱۷۲	۱۵۶
۱۶۳	۱۷۰	۱۷۷	۱۸۴	۱۷۵	۱۷۱	۱۶۴	۱۵۶	۱۵۰	۱۶۵
۱۷۲	۱۶۹	۱۶۲	۱۷۲	۱۷۶	۱۷۵	۱۷۴	۱۶۹	۱۶۶	۱۵۴
۱۶۲	۱۶۷	۱۸۰	۱۶۹	۱۵۲	۱۵۹	۱۶۱	۱۶۴	۱۷۱	۱۵۶
۱۶۳	۱۷۰	۱۶۵	۱۷۲	۱۸۴	۱۷۰	۱۵۰	۱۶۴	۱۷۱	۱۷۰
۱۸۴	۱۷۷	۱۵۳	۱۷۰	۱۶۲	۱۵۴	۱۷۳	۱۷۷	۱۶۹	۱۷۲
۱۶۶	۱۶۴	۱۷۴	۱۶۹	۱۷۹	۱۷۷	۱۷۹	۱۷۳	۱۶۵	۱۶۵
۱۵۷	۱۶۳	۱۵۳	۱۵۸	۱۵۱	۱۶۲	۱۵۹	۱۶۱	۱۶۰	۱۸۳

این داده‌ها نتیجه اندازه‌گیری با مقیاس نسبتی هستند و آنها را داده‌های پیوسته می‌گویند. برای تشکیل جدول فراوانی به طریق زیر عمل می‌کنیم:

در این داده‌ها عدد ۱۵۰ کوچکترین و ۱۸۴ بزرگترین داده هستند. چون داده‌ها تا نزدیکترین واحد سرراست شده‌اند، می‌توان گفت که اندازه واقعی قدها در فاصله [۱۸۴/۵ ، ۱۴۹/۵] قرار دارند. طول این فاصله یعنی ۳۵ را برد داده‌ها می‌نامیم و به چند فاصله مساوی، مثلاً ۵ فاصله هر یک به طول ۷ سانتیمتر، تقسیم می‌کنیم. هر کدام از فاصله‌های کوچک مثلاً [۱۵۶/۵ ، ۱۴۹/۵] را با ۱۵۶/۵-۱۴۹/۵ نشان داده، آن را یک رده با طول واقعی ۷ می‌نامیم. عدد ۱۴۹/۵ را مرز پائین و عدد ۱۵۶/۵ را مرز بالای این رده می‌خوانیم. تعداد رده‌ها باید طوری انتخاب شود که در هر رده، یک یا چند داده داشته باشیم، و معمولاً تعداد داده‌ها نباید از ۵ کمتر یا از ۲۰ بیشتر باشد. گاهی به جای مثلاً ۱۵۶/۵-۱۴۹/۵ می‌نویسیم ۱۵۶-۱۵۰ و در این حال ۱۵۰ را کرانه پائین و ۱۵۶ را کرانه بالای این رده می‌خوانیم. نقطه وسط هر رده که آن را با x_i نشان می‌دهند، نماینده آن رده نامیده می‌شود. مثلاً نماینده ۱۵۶/۵-۱۴۹/۵ می‌شود ۱۵۳ که x_1 است. در واقع برای فشردگی داده‌ها، به جای تمام داده‌هائی که در رده یاد شده جا دارند، نماینده آنها، یعنی عدد ۱۵۳ را به کار می‌برند. بدین ترتیب داده‌های بالا در جدول ۳ به نام جدول فراوانی اندازه‌های قد ۱۰۰ جوان ۲۰ ساله خلاصه می‌شوند.

جدول ۳ جدول فراوانی قد ۱۰۰ جوان بیست ساله با متغیر نسبتی x_i

رده	x_i	f_i	r_i	g_i	s_i
۱۴۹/۵ – ۱۵۶/۵	۱۵۳	۱۵	۰/۱۵	۱۵	۰/۱۵
۱۵۶/۵ – ۱۶۳/۵	۱۶۰	۲۰	۰/۲۰	۳۵	۰/۳۵
۱۶۳/۵ – ۱۷۰/۵	۱۶۷	۳۰	۰/۳۰	۶۵	۰/۶۵
۱۷۰/۵ – ۱۷۷/۵	۱۷۴	۲۵	۰/۲۵	۹۰	۰/۹۰
۱۷۷/۵ – ۱۸۴/۵	۱۸۱	۱۰	۰/۱۰	۱۰۰	۱/۰۰
		۱۰۰	۱/۰۰		

۴.۱ نمودارهای آماری

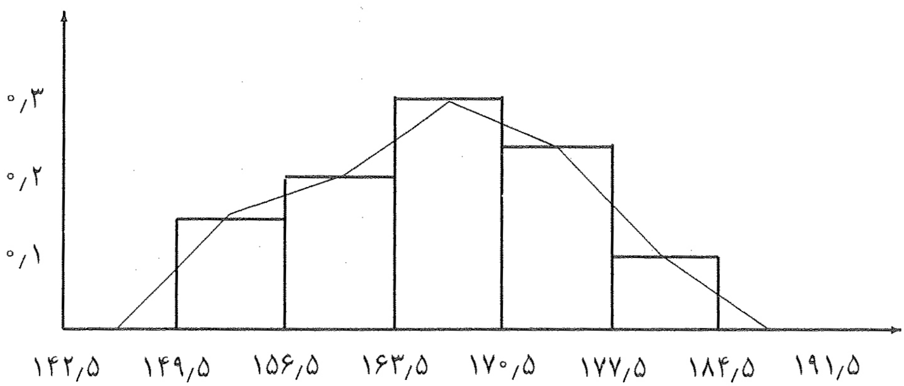
نمایش داده‌ها را، طبق قراردادهای خاص به صورت هندسی، یک نمودار آماری می‌گویند. یک نمودار آماری باید به نحوی ترسیم شود که بتوان به راحتی اطلاعات نهفته در داده‌ها را از روی آن تا حدودی با چشم بدون توضیح و تشریح اضافی دید. هر نمودار آماری، باید دارای شماره، عنوان، و در صورت لزوم زیرنویس و مأخذ باشد. مقیاسهای اندازه‌گیری روی محورهای افقی و عمودی باید مشخص باشند. نمودارهای آماری در امور اقتصادی، صنعتی، بهداشتی، و غیره به کار می‌روند و بر حسب رشته مربوط آنها را به طرق مختلف ترسیم می‌کنند. در اینجا فقط چند نوع نمودار که در آمار و احتمال مورد نیاز می‌باشند شرح می‌دهیم. در پایان این گفتار هم چند نمودار جدید را معرفی می‌کنیم.

۱.۴.۱ هیستوگرام

نموداری است مرکب از چند مستطیل که از روی جدول فراوانی داده‌های پیوسته ساخته می‌شود. تعداد مستطیل‌ها برابر است با تعداد رده‌ها. قاعده هر مستطیل روی محور xها جا دارد و طولش برابر است با طول واقعی رده، که هر چه باشد آن را یک واحد تلقی می‌کنیم، و مرکز آن نماینده رده است. ارتفاع هر مستطیل برابر است با فراوانی نسبی رده مربوط. هیستوگرام را برای داده‌های جدا هم، که بتوان آنها را در یک جدول فراوانی با فواصل مساوی تنظیم کرد، نیز به کار می‌برند. مساحت تمام مستطیل‌های یک هیستوگرام برابر یک واحد مربع است. اگر رده‌ها در جدول فراوانی دارای طولهای مختلف باشند، قاعده‌های تمام مستطیلها برابر نخواهد بود. در این موارد باید ارتفاع مستطیلها را طوری انتخاب کرد که مساحت تمام آنها برابر یک واحد مربع شود(تمرین ۵). در نگاره ۱ هیستوگرام جدول فراوانی ۳ داده شده است. ملاحظه می‌شود که فراوانی‌های نظیر رده‌های ۱۴۲٫۵-۱۴۹٫۵ و ۱۸۴٫۵-۱۹۱٫۵ صفر می‌باشند، ولی برای ترسیم نمودارهای بعدی آنها را منظور می‌داریم.

۲.۴.۱ چندبر فراوانی

اگر نقطه‌های وسط قاعده‌های بالائی مستطیلهای هیستوگرام و نقطه‌های وسط رده‌هائی را که بلافاصله در دو انتهای هیستوگرام بوده، و دارای فراوانی صفر هستند، به هم بپیوندیم، یک خط شکسته به دست می‌آید که آن را چندبر فراوانی می‌نامند. در نگاره ۱ چندبر فراوانی مربوط به جدول ۳ داده شده است.



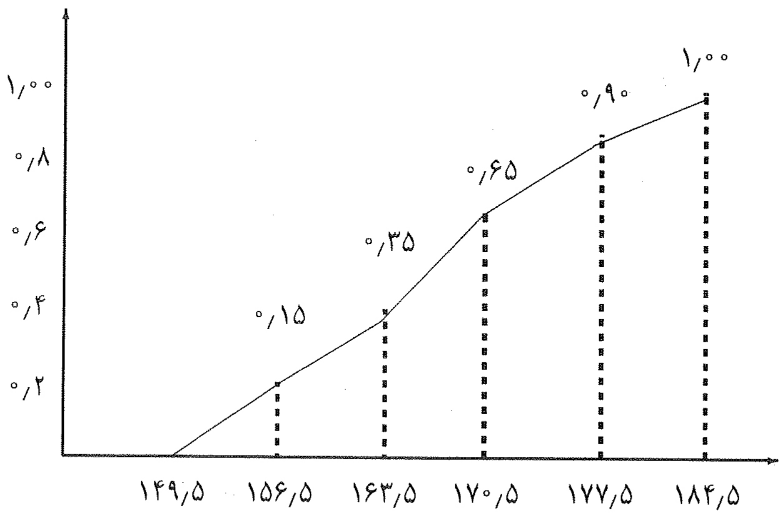
نگاره ۱ هیستوگرام و چندبرفراوانی مربوط به قدها

۳.۴.۱ چندبر فراوانی انباشته

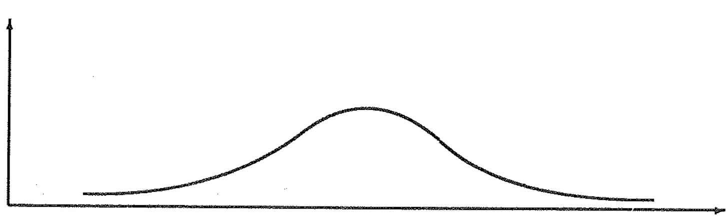
اگر نقاطی را که طول آنها مرز رده‌ها و عرض آنها فراوانی نسبی انباشته تا آن مرز باشد، به هم بپیوندیم یک خط شکسته به دست می‌آید که آن را چندبر فراوانی انباشته می‌نامند. نگاره ۲ چندبر فراوانی انباشته برای جدول ۳ است.

۴.۴.۱ منحنی‌های فراوانی و فراوانی انباشته

اگر تعداد داده‌ها زیاد و طول رده‌ها کوچک و در نتیجه رده‌ها زیاد باشد، چندبر فراوانی و چندبر فراوانی انباشته دارای اضلاع زیاد خواهند شد و می‌توان بر آنها منحنیهائی منطبق کرد که به ترتیب منحنی فراوانی و منحنی فراوانی انباشته نامیده می‌شوند. چون مساحت تمام مستطیلهای هیستوگرام یک واحد مربع می‌باشد، مساحت زیر منحنی فراوانی را هم یک واحد مربع فرض می‌کنند. منحنی فراوانی انباشته، یک



نگاره ۲ چندبرفراوانی انباشته مربوط به قدها



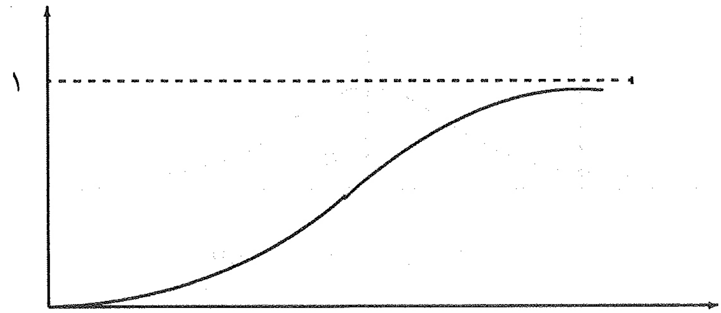
نگاره ۳ منحنی فراوانی

منحنی غیر نزولی است که عرضهای نقاط آن بین صفر و یک می‌باشند. نگاره ۳ و ۴ به ترتیب منحنی فراوانی و منحنی فراوانی انباشته را نشان می‌دهند.

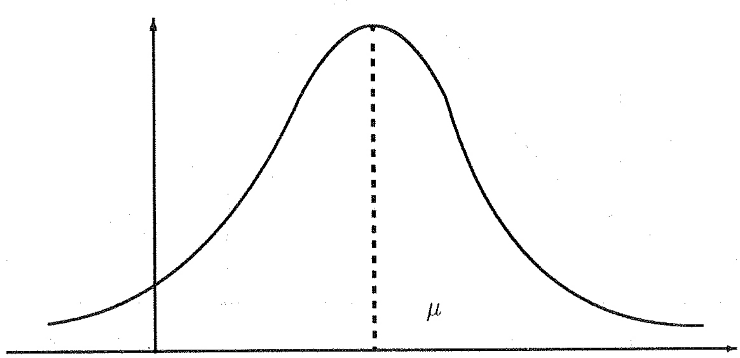
۵.۴.۱ منحنی فراوانی نرمال

اگر منحنی فراوانی دارای معادله مختصاتی

$$y = \frac{۱}{\sqrt{۲\pi\sigma^۲}} e^{-\frac{۱}{۲\sigma^۲}(x-\mu)^۲}$$



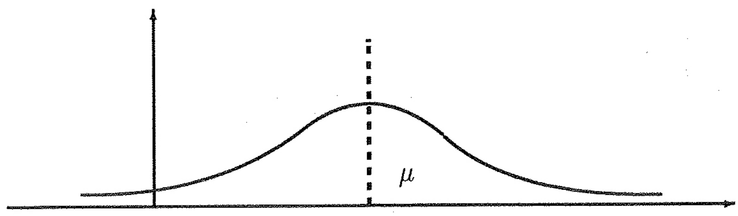
نگاره ۴ منحنی فراوانی انباشته



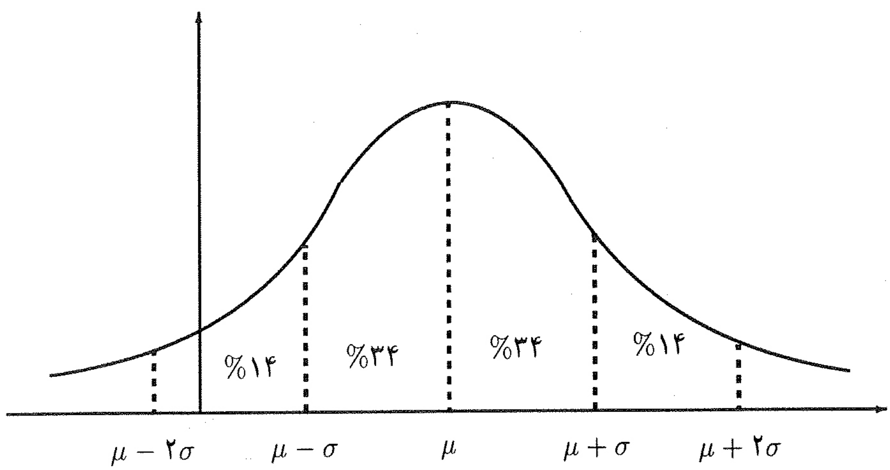
نگاره ۵ منحنی نرمال با $\sigma^{۲}$ کوچک

باشد، آن را منحنی فراوانی نرمال با پارامترهای μ و $\sigma^{۲}$ و متغیر مربوط را متغیر نرمال می‌نامند. این منحنی نسبت به خط $x = \mu$ متقارن، دارای ماکزیمم در نقطه $(\mu, ۱/\sqrt{۲\pi\sigma^{۲}})$، مجانب با محور xها، و زنگ ـ گونه می‌باشد. اگر $\mu = ۰$ و $\sigma^{۲} = ۱$ باشد، منحنی نرمال را منحنی نرمال استاندارد می‌نامند. از عرض نقطه ماکزیمم روشن است که هر چه $\sigma^{۲}$ کوچکتر باشد، منحنی کشیده‌تر بوده، قسمت اعظم مساحت زیر منحنی اطراف $x = \mu$ توزیع می‌شود. نگاره ۵ و ۶ و ۷ نمونه‌هائی از منحنی فراوانی نرمال هستند. نگاره ۸ منهنی نرمال استاندارد را نشان می‌دهد.

کلمه نرمال به معنی طبیعی و عادی است. معمولاً بعضی ویژگیهای طبیعی مانند قد و وزن، تقریباً دارای منحنی فراوانی نرمال هستند، یعنی تعداد افراد معمولی (اطراف μ) زیاد و تعداد افراد غیر طبیعی (دور از μ) کم می‌باشد. نگاره ۸ نحوه توزیع مساحت زیر منحنی نرمال را بر حسب مقیاس σ نسبت به

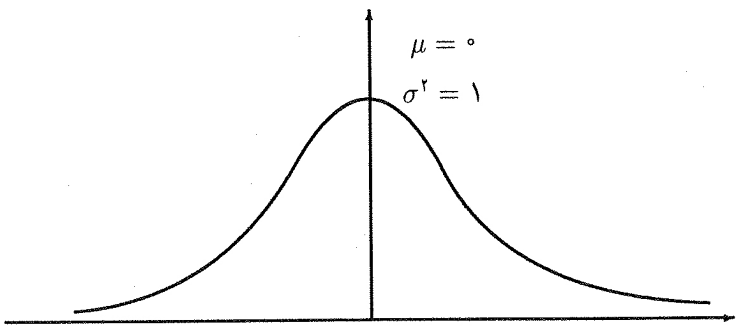


نگاره ۶ منحنی نرمال با $\sigma^۲$ بزرگ



نگاره ۷ نحوه توزیع مساحت زیرمنحنی نرمال

μ نشان می‌دهد.



نگاره ۸ منحنی نرمال استاندارد