

کنترل کیفی داده ها در یک برنامه رجیستری

دکتر سهیل حسنی پور
استادیار اپیدمیولوژی
دانشگاه علوم پزشکی گیلان

کیفیت داده‌ها چیست؟

- ▶ کیفیت داده‌ها معیاری برای اندازه‌گیری درستی وضعیت داده‌ها بر اساس عواملی مانند **صحت**، **کامل بودن**، **سازگاری**، **قابلیت اطمینان** و **به‌روز بودن** است.
- ▶ معیار و سنجش کیفیت داده به سازمان‌ها کمک می‌کند تا **خطاهای داده‌ای** موجود که بایستی برطرف شوند **شناسایی** کنند.
- ▶ سازمان‌ها با استفاده از کیفیت داده در نهایت به این ارزیابی می‌رسند که آیا داده‌های موجود در سامانه‌های نرم‌افزاری **برای اهداف در نظر گرفته‌شده مناسب هستند یا خیر**؟

کیفیت داده‌ها چیست؟

- ▶ به دلیل استفاده فزاینده سازمان‌ها از تجزیه و تحلیل داده‌ها **برای کمک به تصمیم‌گیری‌ها** کیفیت داده‌ها مهم‌تر از گذشته شده است.
- ▶ مدیریت کیفیت داده‌ها **هسته اصلی** فرآیندهای کلی مدیریت داده‌هاست.
- ▶ کارهای انجام‌شده در راستای بهبود کیفیت داده‌ها مرتبط با برنامه‌ها و اصول حاکمیت داده در سازمان می‌باشند.
- ▶ هدف اصلی کیفیت داده‌ها **اطمینان از فرمت صحیح و استفاده مداوم از داده‌ها** در کل سازمان می‌باشد.

تاریخچه کنترل کیفیت

- ▶ می‌توان گفت که تاریخ کنترل کیفیت با **تاریخ صنایع و اختراعات** هم‌زمان بوده، و از روزی که بشر اقدام به ساختن ابزارها و وسایل و ضروریات مورد نیاز خود نموده، همواره در صدد بهبود کیفیت آن‌ها بوده‌است.
- ▶ این موضوع تا زمان انقلاب صنعتی و تولیدات انبوه، خیلی جدی گرفته نشد و هنوز هم جز معدود کشورهای صنعتی پیشرفته مانند ژاپن، آمریکا و انگلیس در بسیاری از کشورها در عمل بهای چندانی به آن داده نمی‌شود، مگر کشورهایی که برای صادر کردن تولیدات خود ناچار به کاربرد آن هستند.

چرا کیفیت داده‌ها مهم است؟

- ▶ داده‌های باکیفیت پایین غالباً باعث می‌شوند تحلیل‌های نادرست از داده ایجاد گردد و استراتژی‌های اشتباه به وجود آید.
- ▶ طی تحقیقاتی در سال ۲۰۱۶ که توسط شرکت IBM انجام شد مشخص گردید هزینه سالیانه مسائل مربوط به کیفیت داده‌ها به ۳,۱ میلیارد دلار رسیده است.
- ▶ توماس ردمن در مقاله‌ای که در سال ۲۰۱۷ توسط دانشگاه MIT منتشر گردید، تخمین زده است که تصحیح خطاهای داده و مقابله با مشکلات تجاری ناشی از داده‌های معیوب و کیفیت داده‌ها به‌طور متوسط سالانه ۱۵ تا ۲۵ درصد درآمد سازمان‌ها را از بین خواهد برد.

کیفیت داده‌های خوب چیست؟

- ▶ **صحت داده‌ها** ویژگی کلیدی و مهم داده‌های باکیفیت بالا است.
- ▶ **برای اجتناب از مشکلات پردازش** تراکنش‌های سیستم‌های عملیاتی و نتایج نادرست ابزارهای تحلیلی، داده‌های استفاده‌شده باید صحیح و درست باشند.
- ▶ **داده‌های نادرست باید شناسایی شده** تصحیح و مستند گردند، تا اطمینان مدیران، تحلیلگران و سایر کاربران جلب گردد که با داده صحیح و اطلاعات خوب کار می‌کنند.

جنبه‌ها و ابعاد مهم کیفیت داده‌ها

کامل بودن داده‌ها : (Data Completeness) وجود مجموعه داده‌هایی که شامل همه عناصر مورد نیاز آنهاست.

سازگاری داده : (Data Consistency) هیچ تضادی بین مقادیر داده‌ای یکسان در سامانه‌های مختلف یا مجموعه داده‌ها وجود نداشته باشد.

داده‌های تکراری : (Data Uniqueness) داده‌های تکراری درون پایگاه داده‌های یکسان وجود نداشته باشد.

رواج داده‌ها : (Data Currency) به معنی اینکه داده‌ها مطابق نیاز به روزرسانی شده تا جریان آنها حفظ شود.

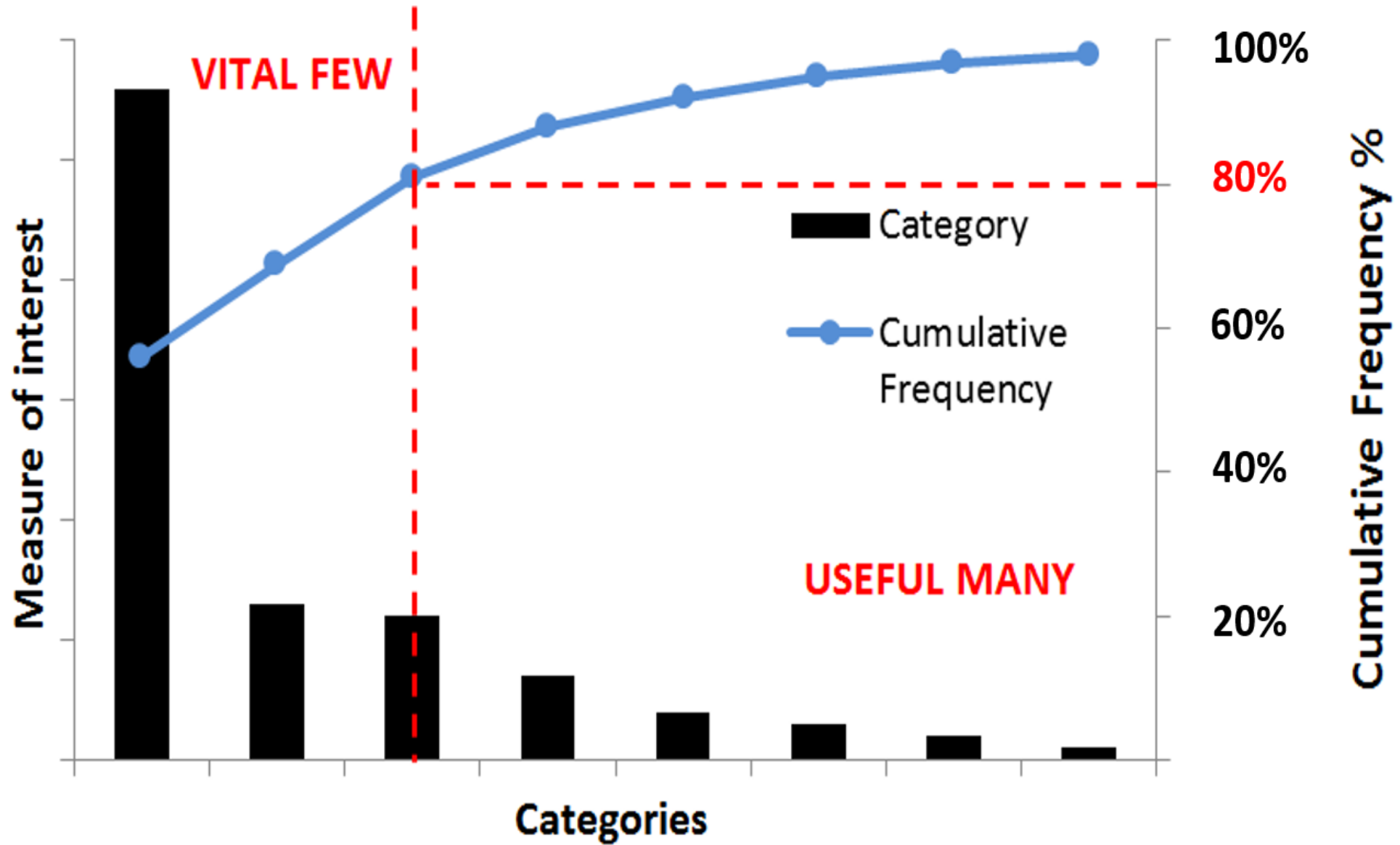
انطباق داده : (Data Conformity) یعنی انطباق داده‌ها با قالب‌های استاندارد که توسط سازمان ایجاد شده است

ابزارهای کنترل کیفیت

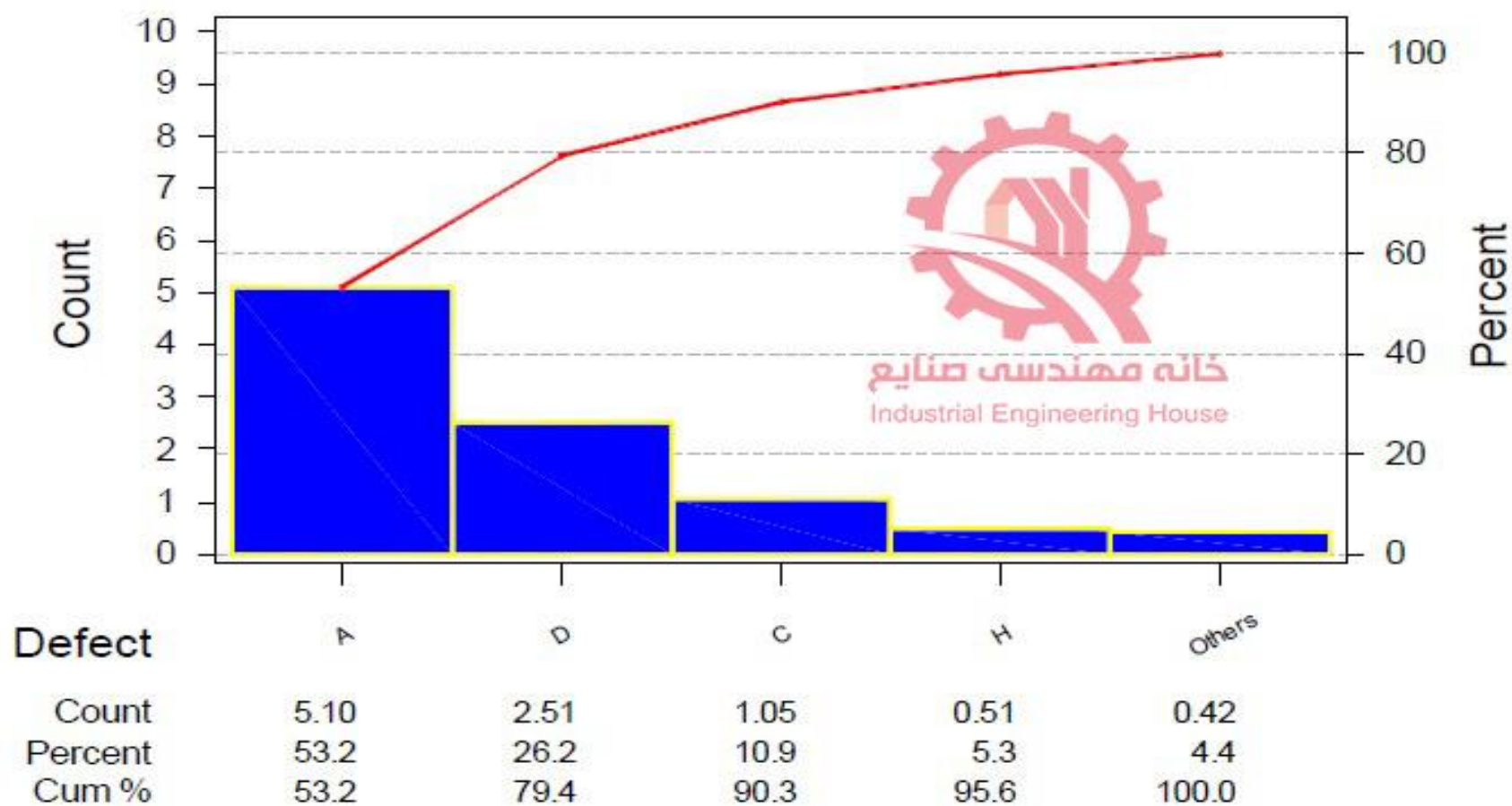
- ▶ نمودار پارتو (Pareto Chart)
- ▶ نمودار پراکندگی (Scatter Plot)
- ▶ نمودار هیستوگرام (Histogram)
- ▶ نمودار کنترل (Control Chart)

نمودار پارتو

- ▶ نمودار پارتو، **نمودار میله ای** است که علل مشکلات به وجود آمده را با فراوانی آنها مقایسه می کند .
 - ▶ نام این نمودار، از یک **دانشمند ایتالیایی** علوم اجتماعی بنام ” ویلفرد پارتو ” گرفته شده است .
 - ▶ براساس اصلی که وی درمورد اقتصاد بیان کرد **۸۰ درصد نتایج و مسائل از ۲۰ درصد علل ناشی می شوند.**
 - ▶ به عبارت دیگر، اگر چه ممکن است برای مسائل موجود، علل بسیار زیادی وجود داشته باشد، ولی اندکی از این علل اهمیت دارند و با رفع آنها میتوان بخش اعظم مسائل را حل کرد .
- به عنوان مثال می توان گفت:
- ▶ **۸۰ درصد خطاها توسط ۲۰ درصد کارکنان انجام می شود.**
 - ▶ **۸۰ درصد ضایعات محصول به علت مشکل در ۲۰ درصد فرایندهای تولیدی آن است.**



PARETO CHART FOR %SCRAB OF MILL BADAMAK

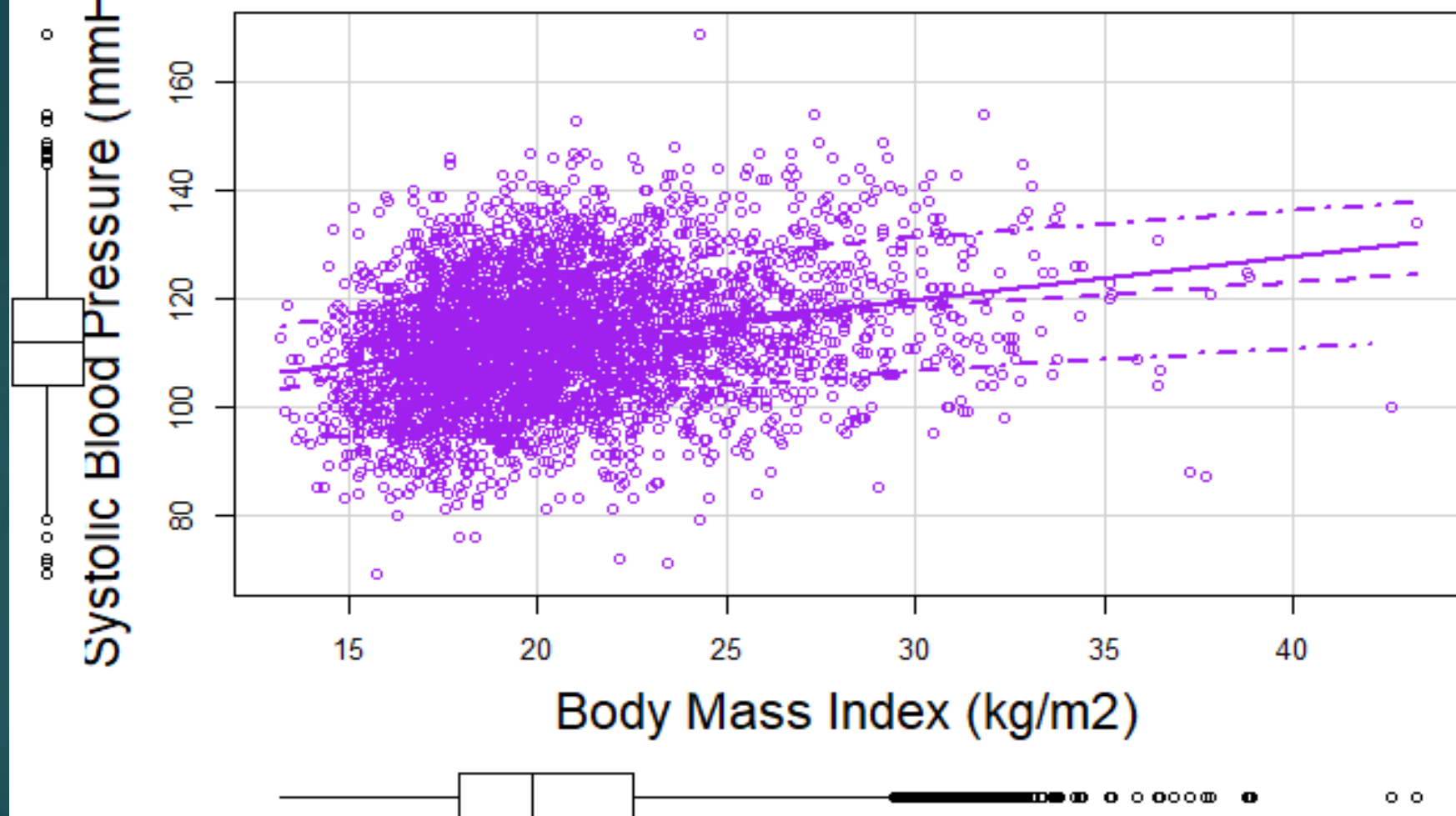


نمودار پارتو درصد ضایعات ریخته گری میل بادامک

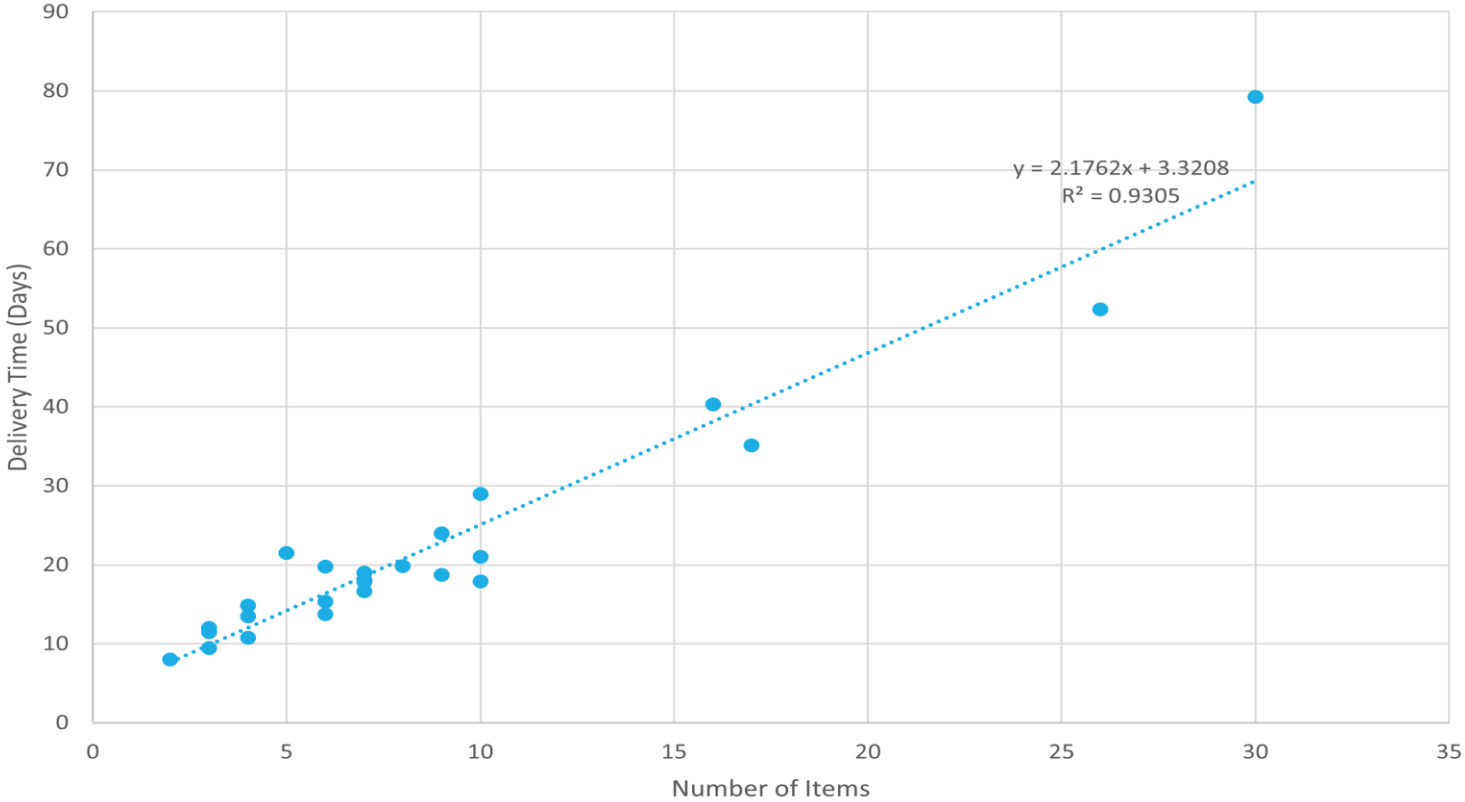
نمودار پراکندگی

- ▶ نمودار پراکندگی یا پراکنش یکی از نمودارهای مفید است که **به منظور پی بردن به رابطه بالقوه بین دو متغیر** استفاده می شود.
- ▶ مقدار Y بر حسب مقدار X بر روی این نمودار رسم می گردد .
- ▶ برای بررسی ارتباط **دو متغیر کمی** مورد استفاده قرار می گیرد.
- ▶ **طریقه رسم نقاط بر روی نمودار پراکندگی بیانگر نوع رابطه موجود بین دو متغیر است.**

ISP Project: Scatter Plot of SBP vs BMI at 13-14 y

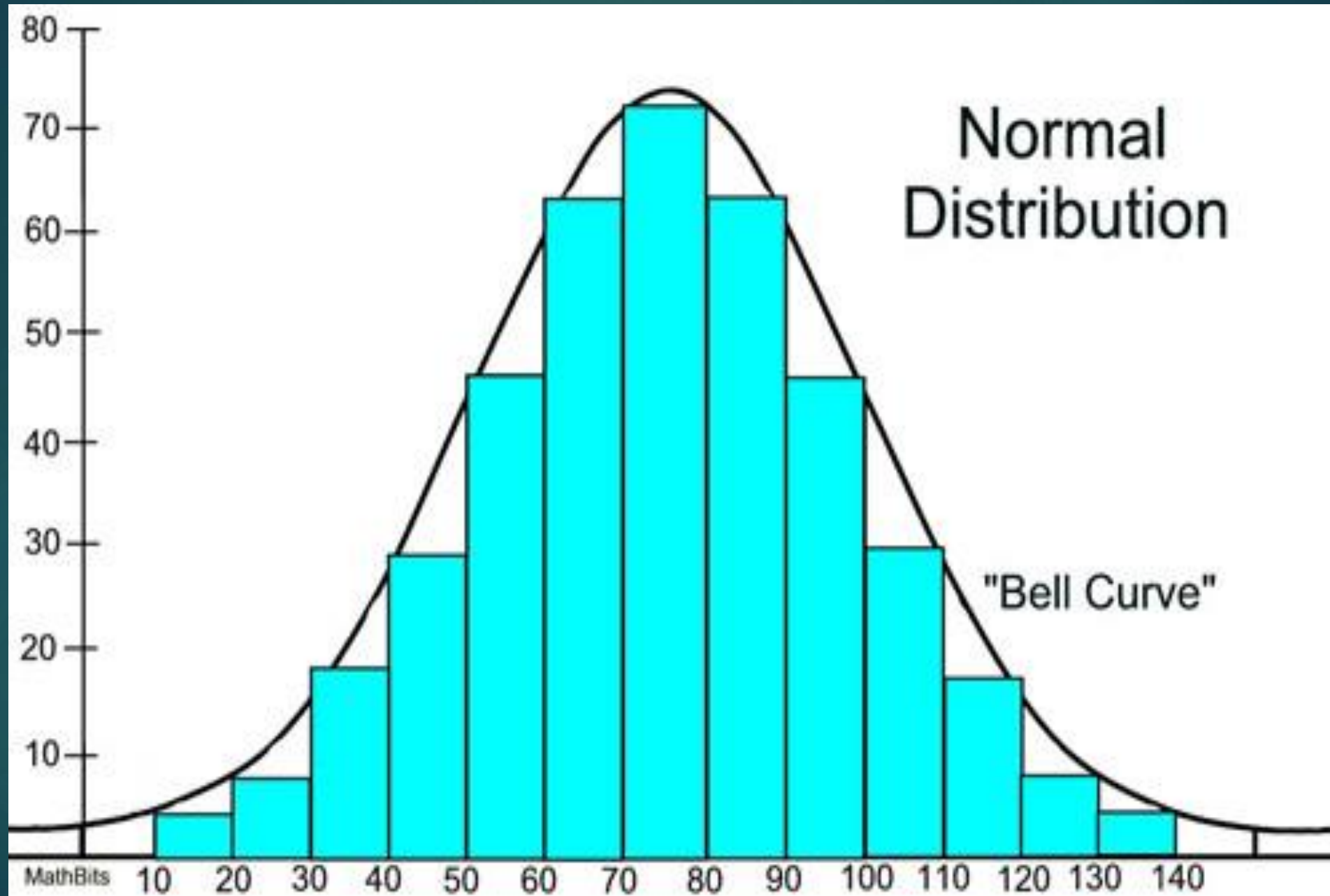


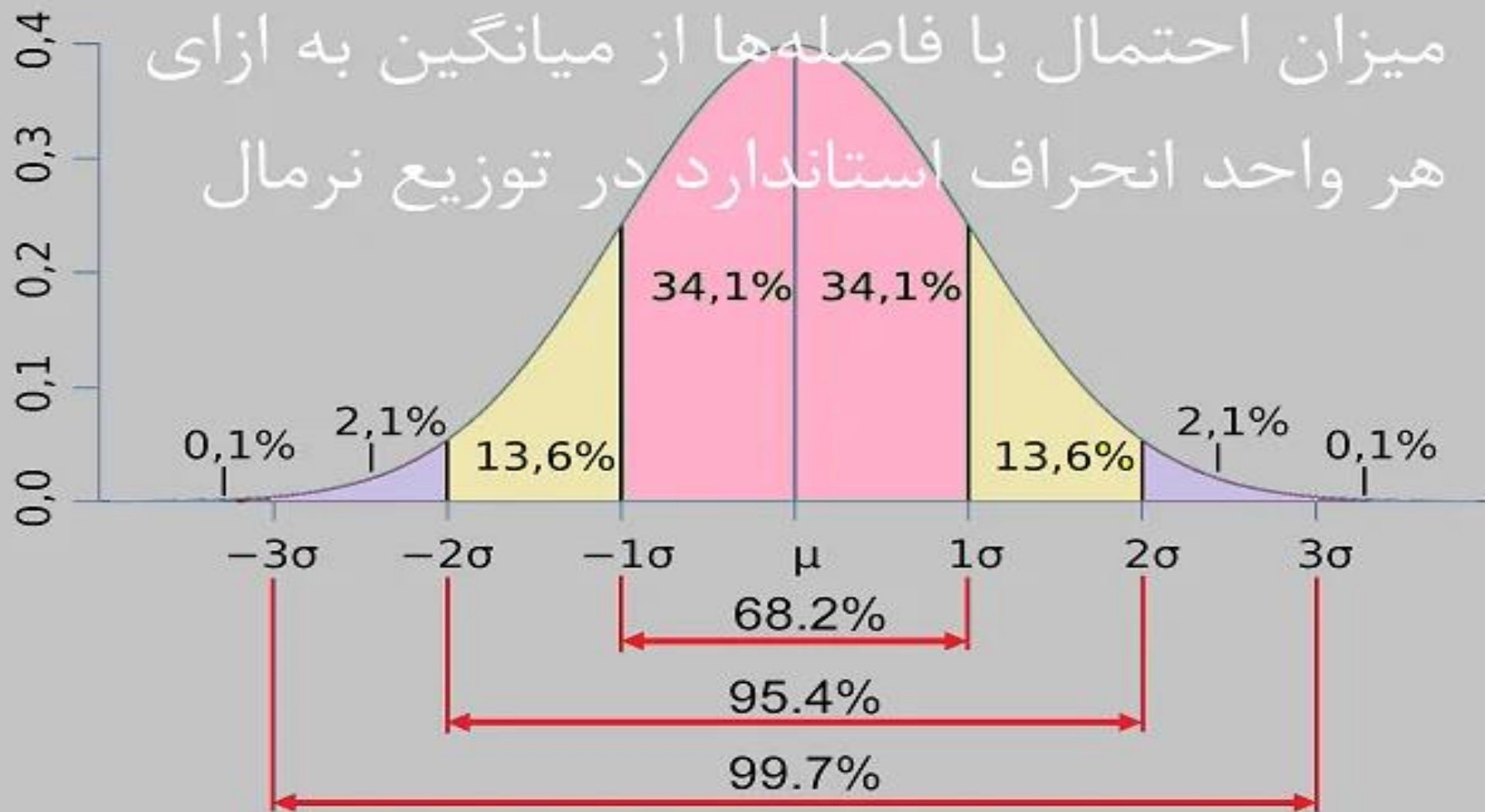
Delivery Time vs. Number of Items



نمودار هیستوگرام

- ▶ هیستوگرام نوعی **نمودار میله ای** است که به کمک آن میتوان داده ها را تشریح کرد.
- ▶ در هیستوگرام، **تعداد زیادی از داده ها** در قالب خاصی طبقه بندی می شوند تا بتوان آنها را ساده تر درک و تحلیل کرد.
- ▶ هیستوگرام، تصویری از داده ها ارائه می کند که توسط آن می توان سه ویژگی زیر را ساده تر مشاهده کرد:
- ▶ ۱- شکل توزیع فراوانی داده ها (Shape)
- ▶ ۲- مکان یا تمایل مرکزی توزیع (Location or Central Tendency)
- ▶ ۳- پراکندگی یا گسترش توزیع (Scatter or Spread)





نمودار کنترل

- ▶ در میان ابزارهای کنترل کیفیت، نمودار کنترل از لحاظ فنی پیچیده ترین آنها است.
- ▶ نمودار شامل یک خط مرکز است که مقدار متوسط مشخصه کیفی را در حالت تحت کنترل نشان می دهد و یا به عبارت دیگر مرحله ای از فرایند را نشان می دهد که فقط خطاهای تصادفی حضور دارند.
- ▶ دو خط افقی دیگر که حد کنترل بالا و حد کنترل پایین نامیده می شوند در این نمودار نشان داده شده اند. این حدود کنترل به گونه ای انتخاب شده اند که اگر فرایند تحت کنترل باشد آنگاه تقریباً کلیه نقاطی که بر اساس اطلاعات نمونه محاسبه شده اند بین این حدود واقع می شوند.
- ▶ تا زمانی که نقاط بین حدود کنترل قرار می گیرند، فرض می شود که فرایند تحت کنترل است و نیازی به فعالیتهای اصلاحی نیست.
- ▶ اگر نقطه ای خارج از حدود کنترل رسم شود نتیجه گیری می شود که فرایند در شرایط خارج از کنترل بسر می برد و اقدامات اصلاحی نیاز است تا منبع ایجاد انحراف یا انحرافات با دلیل تعیین و حذف گردد.

