

### ○ ۳. تحلیل مسئله و گردآوری داده

#### ○ مرحله ۲

#### ○ برنامه بازی: تحلیل عمیق، شناسایی و گردآوری داده و شناخت خطر و عدم اطمینان

مانجو: حالا بیا مرحله ی ۲ رویکرد علم داده را شروع کنیم. قبل از اینکه تحلیل داده ها را شروع کنی، باید اول داده ها را انتخاب کنی. برای این کار، باید اساساً مسئله را بشناسی و انتخاب داده های مربوط را شروع کنی. با دنیای مملو از داده ی در دسترس امروز، اطمینان از این که داده هایی که در پروژه ات استفاده می کنی، واضح، سازمان یافته و مهمتر از همه، باکیفیت هستند، مهم است.

جیم: مانجو، حدس می زنی دو جنبه ی مهم داده را آموزش می دهی، "داده ی درست" و "داده باید واقعی باشد". مانجو: دقیقاً جیم. قطعاً دانستن این که برای مسئله ی معین داده ی درست چیست و آن داده ها چه تناظری با پروژه ات دارند مهم است. حالا بیا درباره ی بخش "تحلیل عمیق" این مرحله حرف بزنیم.

#### ➤ تحلیل عمیق

حالا که هدف را تعیین کردیم، باید اطلاعاتی که معتقدیم برای بررسی های حل مسئله لازمند، انتخاب کنیم. خیلی ساده است، در این مرحله باید بپرسی، مشکل اصلی مربوط به مسئله ی ما چیست. فرض کن یک تولیدکننده ی شبکه های موبایلی هستی و می خواهی بدانی چرا مشتریان به سمت شرکت رقیب رفته اند، مشکل اصلی که شرکت بخواهد به آن رسیدگی کند، قیمت، راحتی، کیفیت شبکه، جایزه ها و غیره است. اینها عوامل، متغیر یا

محرك نامیده می شوند. در مورد شما، تغذیه سالم می خواهد بفهمد چرا پنج موقعیت ضعیف عمل می کنند. محرك هایی که می خواهند بررسی کنند می تواند جامعه، قومیت، قیمت، تمایل به خوردن غذای سالم، کوپن ها و غیره باشد. درک این که چطور و تا چه حد این عوامل بر فروش تأثیر می گذارند، به شروع روند اجرای برنامه ی رفع مشکل کمک می کند.

**جیم: بنابراین آنالیز و تحلیل عمیق، شامل بررسی ریشه ای عوامل احتمالی ایجاد مسئله و درک آن می شود.**

مانجو: فکر می کنم به شکل خوبی بیانش کردی. بیا روی "شناسایی و گردآوری داده" تمرکز کنیم.

## ➤ شناسایی و گردآوری داده

از آنجا که داده به عنوان دارایی مهمی تلقی می شود، بسیاری از سازمان ها مدیری دارند که وظیفه شان مدیریت داده است.

جیم: مانجو، می توانی خلاصه مدیریت داده را با بیان ساده توضیح بدهی؟

مانجو: حتما جیم. از آنجا که داده ها یک دارایی سازمانی مهمی هستند. مدیریت موثر و کارآمد داده مهم است. مدیریت داده باید دو جنبه ای که قبلا گفتیم پوشش دهد- "داده ی درست" و "داده باید واقعی باشد". برای بخش "داده ی درست"، تیمی که در بخش مدیریت داده ها مشغول هست، لازمه که با اشخاص متفاوت / کاربران کسب و کار (تجاری) وارد فعالیت بشود تا بتواند مشخص کند که چه اطلاعاتی لازمه است، که برای حل مشکل در دست جمع آوری قرار بگیرد.

جنبه ی دیگر، "داده باید واقعی باشد"، به کیفیت داده ارتباط دارد که با جزییات درباره اش حرف می زنیم.

کارکردهای مدیریت داده را می توان به چهار نقش اصلی تفکیک کرد: سیاست و حاکمیت داده، استراتژی داده، استانداردهای داده و کیفیت داده. برای مدیریت داده، اطمینان از این که همه ی مؤلفه ها گرد هم می آیند تا هر پروژه علم داده که سازمان انجام می دهد، شانس خوبی برای موفقیت داشته باشد، مهم است.

جیم: مانجو، شرکت ما نسبتاً کوچک است و کارکرد مدیریت داده اختصاصی نداریم. چطور باید داده ها را در این موقعیت کنترل کنیم؟

مانجو: سوال خوبی کردی جیم، اگر شما یک کسب و کار کوچکتر دارید، مهمه که فردی را در داخل سازمان تعیین کنین که مسئولیت او مدیریت داده ها و اطمینان از برآورده شدن الزامات مدیریت داده باشد. فرقی میکنه؟  
جیم: قطعاً.

مانجو: از بین همه ی کارکردهای مدیریت داده، فکر می کنم کیفیت داده، تنها جنبه ی مهم فرایند علم داده است. بدون داده ی مناسب، شاید نتایج تحلیلتان ناقص باشد و تغییراتتان بیهوده باشد. اقدامات کیفیت داده، باید پاکسازی داده و اطمینان از آماده بودن و مناسب بودن داده ها برای تصمیم گیری های مختلف باشد.

داده! داده! با بی حوصلگی گریه کرد: نمی توانم بدون خاک رس آجر بسازم.  
کونان دوئل<sup>۱</sup> - ماجراجویی راش های مسی

جیم: چه جالب. بسیاری از افراد هنگام انجام فعالیت های مربوط به داده درباره ی کیفیت داده صحبت می کنند. با این که وقتی صحبت می کنند سرم را تکان می دهم، اما واقعاً منظورشان را نمی فهمم. می توانی توضیح بدهی؟  
مانجو: قطعاً جیم. اقدامات کیفیت داده را می توان فرآیند حذف و اصلاح داده های نامناسب از مجموعه ی داده دانست. به نظر من حقیر، داده ها باید در سطح منبعشان اصلاح شوند. کسانی که اقدامات کیفیت داده را رهبری می کنند، باید روی اصلاح داده ها در سطح منبع تمرکز کنند. این کار، باعث حذف داده های ضعیفتر در مرحله بعدی می شود.

جیم: چه قدر خوب. می توانی در مورد تغذیه سالم (Eat healthy) (ایت هلثی) چند مثال بیاوری؟  
مانجو: حتماً. بیا در مورد مثال تغذیه سالم بگوییم، داده ی موقعیت مشتری، با استفاده از کدهای پستی ایالات متحده به چند منظور گراوری شدند که باید ۵ رقم داشته باشند. اگر اعداد از ۵ رقم بیشتر یا کمتر باشند یا با

<sup>1</sup> Arthur Conan Doyle, The Adventure of the Copper Beeches - a Sherlock Holmes Short Story

کدهای مشتریان ایالات متحده متناظر نباشند، نادرست خواهند بود و باید تغییر کنند یا حذف شوند. در صورت وجود اعداد گم شده در فیلد کدپستی برخی مشتریان، تحلیل کیفیت داده هم نیاز است.

جیم: مانجو، عالیست. با این بحث، می توانی تعریف داده ی مناسب را توضیح بدهی؟

مانجو: داده های مناسب، داده هایی هستند که کامل، معتبر، مطابق با استانداردها و دقیق هستند. به علاوه، ویژگی بارز داده های مناسب، رفع نیاز مشتری است، یعنی باید برای اطلاعات مطلوب مشتری، در دسترس باشند. منابع ناشناخته، اطلاعات ناقص، خطای گزارش گیری و غیره، همه در ایجاد داده های بد نقش دارند. در سازمان های بزرگتر، اعضای تیم مدیریت داده، کیفیت داده را بررسی می کنند و اطمینان می دهند داده های اصلی با هدف مورد نظر تناسب دارند. در سازمان های کوچکتر، این فرایند شاید کمی سخت تر باشد چون خودتان باید انجامش دهید.

جیم: ممکنه این کار کمی سخت به نظر بیاد اما تجربه یادگیری خوبی می تونه باشه .

مانجو: قطعاً. آمار شناس مطرح جهان، دکتر رائو<sup>۲</sup>، چک لیست کاملی برای بررسی داده ها ارائه کرد، برخی از سوالاتی که او پیشنهاد می کند در زیر لیست شده است.

## ➤ چک لیست رائو

- چطور داده ها گردآوری و ثبت می شوند؟
- آیا سیستم اندازه گیری به خوبی تعریف شده است؟
- داده ها بدون خطای ثبت هستند؟
- داده ها از منابع معتبر و قابل اعتماد به دست آمده اند؟
- آیا هیچ ناهنجاری یا مشکلی در ارتباط با داده ها وجود دارد؟
- اندازه نمونه مناسب است؟ همه ی عوامل گردآوری داده ها در نظر گرفته شده است؟
- نوع درستی از داده ها در اختیار دارید که با هدف مورد نظر تناسب داشته باشد؟

<sup>2</sup> Rao, 1989

جیم: این چک لیست به نظر تمام جنبه های کیفیت داده را پوشش می دهد. خیلی مفید است.

مانجو: بله جیم. تمامی جوانب را تحت پوشش قرار می دهد برای کسب و کارهای کوچکی که داده های ذخیره شده زیادی ندارند، خرید داده یا جمع آوری داده ها ممکن است یک مرحله اضافی در این فرآیند باشد. اگر تصمیم داشتید داده بخرید، اطمینان از قابل اعتماد بودن شرکتی که از آن داده را خریداری می کنید خیلی اهمیت دارد.

خرید داده های استاندارد صنعتی از شرکت گردآوری داده ی معتبر، راه خوبی برای شروع پژوهش است و اطمینان ایجاد می کند که داده ها کیفیت خوبی دارند. اگر کسب و کار کوچکی دارید که به دنبال حل مسئله ای خاص است، شاید گردآوری داده لازم باشد. در این مورد، شخصی که برای مسئولیت مدیریت داده تعیین می کنی باید با استفاده از چک لیست بالا و تحلیل کیفیت داده مطمئن شود کیفیت رضایت بخش است.

جیم: این خوب به نظر می رسد. کیفیت داده ها را چطور دقیق اندازه گیری می کنید؟

مانجو: می خواهم الان درباره ی همین موضوع صحبت (بحث) کنم. [این چیزی است که اکنون در موردش بحث خواهم کرد.] به طور معمول، کیفیت داده ها برای تمام معیارها و متغیرهای وابسته به درصد براساس چهار بعد اندازه گیری می شود: تکمیل (کامل بودن)، انطباق، اعتبار و دقت. برای مثال، مثلاً اگر به دنبال معیاری مثل درآمد باشیم، باید کیفیت داده را با تعیین نمرات درصدی چهار بُعد بالا اندازه گیری کنیم. این درصدها نمرات کیفیت داده (DQ) نامیده می شوند.

جیم: نمرات DQ شاخص عملکرد داده ها هستند؟

مانجو: صد در صد (قطعا یا دقیقا). آنها شاخص های عملکرد داده ها هستند. اجازه بده این چهار بُعد را کوتاه توضیح دهم. بُعد تمامیت اطمینان ایجاد می کند داده های کاملی برای اجرای تحلیل مورد نیاز داریم؛ انطباق اطمینان ایجاد می کند که داده در قالب مورد نیاز موجود است؛ اعتبار اطمینان ایجاد می کند داده ها با مقادیر معتبر ارائه شده اند؛ و دقت اطمینان ایجاد می کند داده ها دقیق هستند و درستی داده ها را نشان می دهد.

ضمناً درصد امتیازات به سادگی تعداد نقاط داده منحصر به فردی هستند، که با الزامات تقسیم بر کل مطابقت می کنند.

جیم: متوجه شدم. می توانید با مثال توضیح دهید؟

مانجو: البته جیم. بنابراین، فرض کنید که شما داده‌های درآمد برای ۱۰۰ نفر در یک منطقه مشخص داشته باشید. اگر ۹۳/۱۰۰ سلول پر شوند، پس امتیاز کامل شما ۰٫۹۳ (یا ۹۳٪) است. اگر ۹۲/۱۰۰ سلول در قالب صحیح باشند (به عنوان مثال: آن را دلار فرض می‌توان کرد). پس امتیاز شما ۰٫۹۲ است (یا ۹۲٪).

روایی نیازمند این است که شما مفهوم معیار اندازه‌گیری را بدانید تا اطمینان حاصل کنید که مقادیر معتبر هستند. بنابراین، برای رسیدن به درآمد، توجه به داده‌های خارج از محدوده (پرت) که احتمالا اشتباه هم هستند، مهم می‌باشد. به عنوان مثال، داشتن یک نقطه داده ۵۰ میلیون در محدوده‌ای می‌تواند باشد، که در آن میانگین درآمد ۵۰ هزار تا است. نهایتاً، دقت به تأیید داده‌ها اشاره دارد. مهم است که اطمینان حاصل شود که هر شخص مقادیر درستی را از میزان درآمد خود ارائه داده باشد. نگاه کردن به فیش حقوقی افراد جهت کسب اطمینان از اینکه مردم در حال افزایش یا کاهش درآمد نیستند مهم است.

جیم: با این مثال‌ها حالا مشخص شد. خیلی جالب است. وقتی نمره‌ی تمام ابعاد را داری، چطور نمره کیفیت کل داده‌ها را محاسبه می‌کنی؟

مانجو: سوال خوبی است. بیا درمورد داده‌های درآمد بگوییم. نمرات ابعاد، به این صورت است: تمامیت ۰٫۹۳؛ انطباق: ۰٫۹۲؛ اعتبار: ۰٫۹۱؛ و دقت: ۰٫۹. پس نمره کل کیفیت داده‌ها برای درآمد، با ضرب تمام این نمره‌ها به دست می‌آید یعنی  $۰٫۹۳ \times ۰٫۹۲ \times ۰٫۹۱ \times ۰٫۹$  که مساوی ۰٫۷۰ یا (۷۰ درصد) است.

جیم، برای کسی که برای بررسی داده‌ها و تأیید این که کیفیت داده‌ها خوب است و با کاربرد مورد نظر تناسب دارد، تحلیل انجام می‌دهد، همیشه ایده‌ی خوبی است.

جیم: بسیار خب، مانجو. من مدیر یکی از مکان‌هایی که ضعیف عمل کرد را می‌شناسم. من با او تماس می‌گیرم تا در مورد داده‌ها به ما کمک کند و بعد می‌توانیم کیفیت داده‌ها را ارزیابی کنیم.

مانجو: این عالی‌ه. درباره‌ی آمار (میزان) افزایش فروش سه ماه قبل و اطلاعات مربوط به عوامل احتمالی مثل قومیت و اطلاعات تخفیف/کوپن سوال کن. اگر این داده‌ها را داشته باشیم، باید بتوانیم کیفیت داده‌های موجود را ارزیابی کنیم.

جیم: چه خبر خوبی. من اخیراً (به تازگی) یک ایمیل از مدیر آنجا گرفتم. فایل داده‌های سه ماهه قبل را می‌توانند بدهند.

مانجو: چه سریع! پس بعد از گرفتن فایل داده‌ها برای ادامه‌ی بحثمان، بیا همدیگر را ببینیم.

جیم: به نظرم خوب است مانجو. همه چیز را برای جلسه ی بعد آماده می کنم.

مانجو: به زودی می بینمت جیم.

(جیم فایل های داده را از مدیر آن موقعیت مکانی می گیرد. بعد از یک هفته با مانجو ملاقات می کند تا بحث را ادامه دهد)

جیم: سلام مانجو. چطوری؟ من فایل های دیتا (اطلاعات یا پرونده داده ها) را دارم.

مانجو: خوبم جیم. بیا اطلاعات داده ها را ببینیم.

جیم: بسیار خب.

مانجو: جیم، با دیدن این فایل داده ها، راحت می توانی بعضی از مشکلات کیفیت داده ها را ببینی. بعضی داده ها گم شده اند و بعضی ها هم معتبر نیستند. جالب است بدانی داده های کامل و دقیقی برای سه ماه دارند که از آنها برای ایجاد نمودار استفاده می کنیم.

جیم: داده های سه ماهه را تیم فروش و بازاریابی گردآوری می کنند.

مانجو: فکر می کنی ممکن است آن موقعیت و تیم بازاریابی به اندازه کافی برای بحث درباره ی چالش ها/ مسائل اطلاعاتی با یکدیگر تعامل نداشته باشند؟

جیم: احتمال دارد.

جدول ۳/۱. تعداد فروش هفتگی سه ماهه آخر

هفته	واحد فروخته شده	% افزایش فروش
هفته قبل	۲۰۰۰	
هفته ۱	۲۰۴۰	۲
هفته ۲	۲۱۰۱	۲/۹۹
هفته ۳	۲۱۸۵	۳/۹۹
هفته ۴		
هفته ۵	۲۴۷۸	
هفته ۶	۲۶۱۴	۵/۴۸
هفته ۷	۲۷۱۸	۳/۹۷
هفته ۸	۲۷۴۵	۰/۹۹۳
هفته ۹	۲۷۶۶	۰/۷۶۵
هفته ۱۰	۲۸۴۸	۲/۹۶
هفته ۱۱	۲۹ Y ۲ (نا معتبر)	
هفته ۱۲	۳۱۱۰	



مانجو: گفتگو درباره این موضوع را برای بعد بگذاریم. در حال حاضر، اجازه بده در مورد موضوعات کیفیت داده ها در فروش هفتگی صحبت کنیم. (جدول ۳,۱).

اگر به داده های مربوط به تعداد واحدهای فروخته شده نگاه کنید، ما داده های گم شده برای هفته ۴ و ورودی نامعتبر برای هفته ۱۱ داریم، جایی که یک حرف به جای یک عدد وجود دارد. این منجر به نمره کامل ۹۲,۳۰٪  $((12/13) \times 100)$  و نمره اعتبار ۸۴,۶۱٪  $(13/11) \times 100$  می شود.

من فکر می کنم برای داده های فروش هفتگی، این دو بعد مرتبط و کافی هستند. با استفاده از این دو بعد، امتیاز کلی کیفیت داده برای واحدهای فروخته شده حاصل ضرب این دو امتیاز است که ۷۸,۱۰ درصد خواهد بود. به دلیل این مسائل، کیفیت داده های افزایش درصد در فروش نیز تحت تأثیر قرار می گیرد، زیرا این مقدار برای چهار هفته قابل محاسبه نیست. بنابراین امتیاز کیفیت داده برای این متغیر (درصد افزایش فروش) ۶۶,۶۷٪  $(12/8) \times 100$  است.

جیم: متشکرم، مانجو برای این آنالیز.

درباره ی فایل (پرونده) بعدی چطور؟ فکر می کنم فایل بعدی اطلاعات مربوط به قومیت (نژاد) مشتریان و اطلاعات مربوط به تخفیف / کوپن باشد.

مانجو: درست است جیم. (تو درست می گویی جیم) فایل بعدی همین اطلاعات را دارد (جدول ۲,۳). این داده ها هم مشکلاتی دارند. اگر به (جدول ۲,۳) نگاه کنی، نمره ی تمامیت و اعتبار برای قومیت ۸۴٪ و ۸۰ درصد و نمره کل کیفیت داده ۶۷/۲٪ است. به همین ترتیب، نمرات تمامیت و اعتبار تخفیف / کوپن ۹۲ و ۸۰ درصد و نمره کل کیفیت داده ها ۷۳/۶٪ است. مثل مورد قبل، ابعاد تمامیت و اعتبار، اینجا هم مرتبط و کافی هستند.

جیم: متشکرم مانجو. این اطلاعات را برای مدیر آن موقعیت می فرستم و از او درخواست بررسی این اطلاعات را خواهم کرد.

مانجو: به نظر خوب می رسد. (ایده خوبی به نظر می رسد).

جدول ۳/۲. داده های نمونه در مورد قومیت مشتری و تخفیف/کوپن

تخفیف/کوپن	قومیت	مشتری
ندارد	سفید پوست	۱
دارد (نا معتبر)	مکزیکي (نا معتبر)	۲
دارد/ ندارد	سفید پوست	۳
ندارد	آسیایی	۴
ندارد	آسیایی	۵
ندارد		۶
دارد	آسیایی	۷
ندارد	آسیایی	۸
	آسیایی	۹
ندارد	آسیایی	۱۰
دارد	آسیایی	۱۱
ندارد	آمریکایی آفریقایی	۱۲
دارد		۱۳
ندارد	اسپانیایی	۱۴
ندارد	سفید پوست	۱۵
	سفید پوست	۱۶
دارد	اسپانیایی	۱۷
دارد	سفید پوست	۱۸
NA (نامعتبر)	آسیایی	۱۹
ندارد	هندي	۲۰
ندارد	آسیایی	۲۱
		۲۲
ندارد		۲۳
دارد	آسیایی	۲۴
تخفیف	آسیایی	۲۵

جیم: من به تازگی از مدیر محل شنیدم. آن‌ها می‌توانند به داده‌های فروش نگاهی بندازند و آن را اصلاح کنند. برای فایل داده‌ی بعدی، وقت بیشتری نیاز دارند. از آنها می‌خواهم فایل داده‌های فروش اصلاح شده را ارسال کنند.

مانجو: عالیست. پس می‌توانیم حداقل تعداد اصلاحات فروش را ببینیم.

**اعداد دروغ نمی‌گویند، اما دروغگوها اعداد را پیدا خواهند کرد.**

ژنرال چارلز گراسونور<sup>۳</sup>

همانطور که احتمالاً تا حالا متوجه شدین، افراد اغلب کیفیت داده‌ها را بدیهی می‌دانند و در نهایت از داده‌های نامناسب برای تحلیل‌ها استفاده می‌کنند. این داده‌های نامناسب می‌تواند افراد را به پاسخ برساند اما ممکن است حقیقت را نشان ندهد.

جیم: شما می‌گویید ما می‌توانیم پاسخ‌هایی را که می‌خواهیم بشنویم، دریافت کنیم و نه لزوماً پاسخ‌هایی که به کسب و کار ما کمک می‌کند.

مانجو: دقیقاً فقط در صورتی می‌توانید مشکل را حل کنید که بدانید مسئله دقیقاً چیست. داده‌های مناسب کمکتان می‌کند مطمئن شوید علت اصلی را پیدا کرده‌اید. از آنجایی که تغذیه سالم یک زنجیره کوچک‌تر با یک تابع مدیریت داده کوچک است، برای آنها به دست آوردن داده‌ها و اطمینان از کیفیت خوب آنها کار دشواری نیست. البته، باید برای تحلیل بیشتر محرک‌هایی انتخاب کنند مثل جامعه، قومیت، متوسط درآمد خانوار، تمایل به غذای سالم و تخفیف/کوپن. به علاوه، هنگام اجرای پروژه، پیش‌بینی کامل مشکلات احتمالی، غیرممکن است، پس همه‌ی اینها، خطر دارد.

جیم: از چه نوع خطری باید مطلع باشیم؟

مانجو: مفاهیم خطر و عدم قطعیت، از گذشته‌های دور وجود داشته است. ما هر از گاهی با عدم اطمینان مواجه می‌شویم و تأثیرات آن را می‌بینیم. اکثر ما توقع داریم، همه چیز همانگونه که تصور می‌کنیم پیش برود. اما در واقعیت همه چیز متفاوت است. خطر و عدم اطمینان همه جا وجود دارد، مثل دنیای علم داده.

<sup>3</sup> General Charles H. Grosvenor

پس حالا توضیح بده چطور می توانیم خطر و عدم اطمینانها را بشناسیم و اقداماتی برای کاهش آنها انجام دهیم.

## ➤ شناسایی خطر و عدم اطمینان

جیم: مانجو، وقتی من خطر و عدم اطمینان را بشناسم و اقدام کنم، مشکل ناشی از خطر حذف می شود؟  
مانجو: نه جیم. برای شناخت خطر و عدم اطمینان و کاهش آنها در پروژه، راه هایی وجود دارد اما هیچ راهی برای حذف کاملشان وجود ندارد. خطر و عدم قطعیت، بخشی از زندگی و کسب و کار هستند.

هدف از این بخش، تجزیه و تحلیل آنچه به نظر من شش منبع اصلی خطر و عدم اطمینان است که ممکن است در پروژه علم داده خود با آن رو به رو شوید.

اگر ما آگاهی نداشته باشیم، نمی توانیم اقدام کنیم.
اگر نتوانیم اقدام کنیم، خطر متضرر شدن بالا می باشد.
اگر آگاهانه عمل کنیم، احتمال ضرر و زیان قابل مدیریت (کنترل) خواهد بود.
اگر ما آگاهانه اقدام نکنیم، سزاوار ضرر و زیان هستیم.
مایکل جی هری <sup>۴</sup> ، موسس و معمار اصلی سیگما شش

## ➤ شش منبع اصلی خطر و عدم اطمینان

۱. خطر و عدم اطمینان در خطای اندازه گیری داده ها
۲. خطر و عدم اطمینان ناشی از وجود تغییر
۳. خطر و عدم اطمینان در پیش بینی، تشخیص و تصمیم گیری
۴. خطر و عدم اطمینان در اجرای فرایند تحلیل
۵. خطر و عدم اطمینان ناشی از اطلاعات ناقص
۶. خطر و عدم اطمینان ناشی از تأخیر

<sup>4</sup> Mikel J Harry

بر اساس اندازه ی کسب و کارت، برخی از این منابع ممکن است کمتر یا بیشتر از بقیه کاربردی باشند. اطلاع از تمام انواع خطر برای جلوگیری از ایجاد مشکل در پروژه ات مهم است.

جیم: مانجو، موافقم که پنج مورد اول به طریقی به علم داده ارتباط دارند. فقط مطمئن نیستم مورد آخر چه ربطی دارد.

مانجو: این چیزی است که به نظر من بسیار مهم است. در زمینه علوم داده ها ، مربوط به تصمیمات یا اقدامات به تاخیر افتاده براساس یافته های تحلیلی می باشد. این بیشتر در مورد اجرا و استقرار یافته های تجزیه و تحلیل است.

مانجو: حالا بیا، به اولین منبع خطر و عدم اطمینان مربوط به خطاهای اندازه گیری داده ها بپردازیم.

## ➤ خطر و عدم اطمینان در خطای اندازه گیری داده

اندازه گیری های دقیق و مناسب برای یک پروژه علمی با داده های خوب ضروری می باشد.

اگر ما داده های صحیح را اندازه نگیریم، در معرض خطر کار کردن بر روی ابتکارات نادرست و نتایج نادرست قرار خواهیم گرفت.

اندازه گیری دو موضوع اصلی دارد: چه چیزی اندازه گیری شود و چطور اندازه گیری شود. چه چیزی اندازه گیری شود، به شناخت مسئله و محرک هایی که برای حل مسئله بررسی خواهد شد، نیاز دارد.

"چگونگی سنجش" در مورد تعیین استانداردهای سازمانی است که تضمین می کند هر کسی که اندازه گیری ها را انجام می دهد یا اندازه گیری ها را می خواند و آنها را به عنوان نقاط داده ثبت می کند این کار را به روشی ثابت (استاندارد) انجام دهد. مهم است که اطمینان حاصل شود که داده ها در جمع آوری و نمایش یکنواخت هستند.

جیم: مانجو، جیم: مانجو، یک لحظه صبر کن. من فکر می کنم در اینجا شما دو جنبه از داده ها را در دست داری، درست است؟ فکر می کنم چیزی که اندازه گیری می شود همان، جنبه ی داده های درست است.

مانجو: قطعاً جیم. درست می‌گویی. چیزی که اندازه‌گیری می‌شود همان جنبه‌ی داده‌های درست است. به هر حال، بخش نحوه‌ی اندازه‌گیری، شرط مهمی برای جنبه‌ی داده‌ای که باید واقعی باشد، است. جنبه‌ی داده باید واقعی باشد، ابعاد دیگر کیفیت داده که قبلاً گفتیم را هم در بر می‌گیرد.

جیم: متوجه شدم (فهمیدم). بخش نحوه‌ی اندازه‌گیری، با روش گردآوری داده‌ها، ابزار مناسب اندازه‌گیری، دقت آن و غیره ارتباط دارد. درست می‌گویم؟

مانجو: این حرفت هم درست است. اگر تغذیه سالم، بخواهد ببیند به طور متوسط هر مشتری برای آمدن به رستوران چه مسافتی باید طی کند و مشخص نباشد که مشتریان قصد دارند هزینه‌ی این مایل و کیلومتر و دقیقه‌ها و غیره را بدهند یا نه، خطای مهمی در داده‌ها ممکن است به وجود بیاید. مثلاً عدد ۶ در فیلد مقدار، می‌تواند هر چیزی باشد یعنی، ۶ دقیقه، ۶ مایل، ۶ ناحیه و غیره. خطا در این موقعیت‌ها اغلب به دلیل تعاریف و مفاهیم ضعیف، سوگیری در اندازه‌گیری و تفاوت ابزار و محققان اندازه‌گیری رخ می‌دهد. وقتی کیفیت داده‌های مشخص شد، لازم است اندازه‌گیری‌ها، درجه‌ی خاصی از اطمینان داشته باشند. به عبارت دیگر، باید مطمئن شویم که چقدر به کیفیت اندازه‌گیری اطمینان داریم.

جیم: می‌توانی درباره‌ی درجه‌ی اطمینان بیشتر صحبت کنی؟

مانجو: حتماً، جیم. درجه‌ی اطمینان، اندازه‌ی احتمال است. روش تعیین مقدار عدم کیفیت است. می‌توانیم برای مقادیر اندازه‌گیری شده با ۹۰٪، ۹۵٪ یا ۹۹٪، حد بالا و پایین به دست بیاوریم. به این ترتیب می‌توانیم اطمینان اطلاعات را برآورد کنیم. به عنوان مثال، می‌توانیم بگوییم با ۹۵٪ اطمینان (احتمال) به طور متوسط یک مشتری بین ۱۰ تا ۱۵ دقیقه برای گرفتن غذا در تغذیه سالم، منتظر می‌ماند. در برآورد مورد کیفیت داده هم می‌توانیم از سطح اطمینان برای تعیین درجه‌ی اطمینان استفاده کنیم. مثلاً، با ۹۵٪ اطمینان، می‌توانیم بگوییم کیفیت داده‌های مسافت طی شده، ۸۵٪ است. سطح اطمینان درجه‌ای است که با آن می‌توانیم دقت برآورد پارامترها را باور کنیم. برآوردهای مبتنی بر سطح اطمینان، بر اساس یک نظریه‌ی برآورد آماری محاسبه میشوند که در آن از داده‌های نمونه با درجه اطمینان خاصی استفاده می‌کنیم (۹۰٪، ۹۵٪ یا ۹۹٪).

جیم: فرض می‌کنم می‌توانم به راحتی از برنامه‌های نرم‌افزاری برای به دست آوردن برآوردها بر اساس سطح اطمینان مختلف، استفاده کنم، درست است؟ این مفهوم خیلی جالب است. با این حال اگر مجبور باشیم از فرمول‌های پیچیده برای محاسبه‌ی برآوردها استفاده کنیم، سخت می‌شود.

مانجو: بسته ها یا برنامه های آماری مختلف ارائه شده، برای مجموعه داده های داده شده به سادگی زمینه اطمینان را فراهم خواهند کرد. پس نیاز نیست نگران فرمول های پیچیده باشید. در واقع، راستش را بخواهی، در بسیاری از مراحل آموزش در دبیرستان به دانش آموزان این مفاهیم و کاربردهایش را آموزش می دهند که همین نشان می دهد چقدر درکشان ساده است..

جیم: وای! بعد از این سوال احساس حماقت کردم.

مانجو: چیزی به اسم یه سوال احمقانه وجود نداره جیم. همه ی سوالات مهم هستند. به موضوع خودمان برگردیم، می توانیم برای اندازه گیری هایی مانند میزان ضعف (نقص) سطح اطمینان تعیین کنیم. ضعف تغذیه سالم می تواند انتظار فرد برای غذا بیشتر از مدت زمان انتظار یا بازه ی انتظار باشد. در آن صورت میزان ضعف، درصد افرادی هستند که بیشتر از مقدار مشخص شده منتظر مانده اند. پس برای تغذیه سالم می توانیم با اطمینان ۹۹٪ بگوییم که تنها ۳٪ تا ۸٪ مشتریان بیشتر از مدت زمان انتظار، منتظر می مانند.

جیم: فهمیدم مانجو.

مانجو: حالا بیا درباره ی نوع دوم خطر در این فهرست حرف بزنیم.

## ➤ خطر و عدم اطمینان ناشی از وجود تغییر

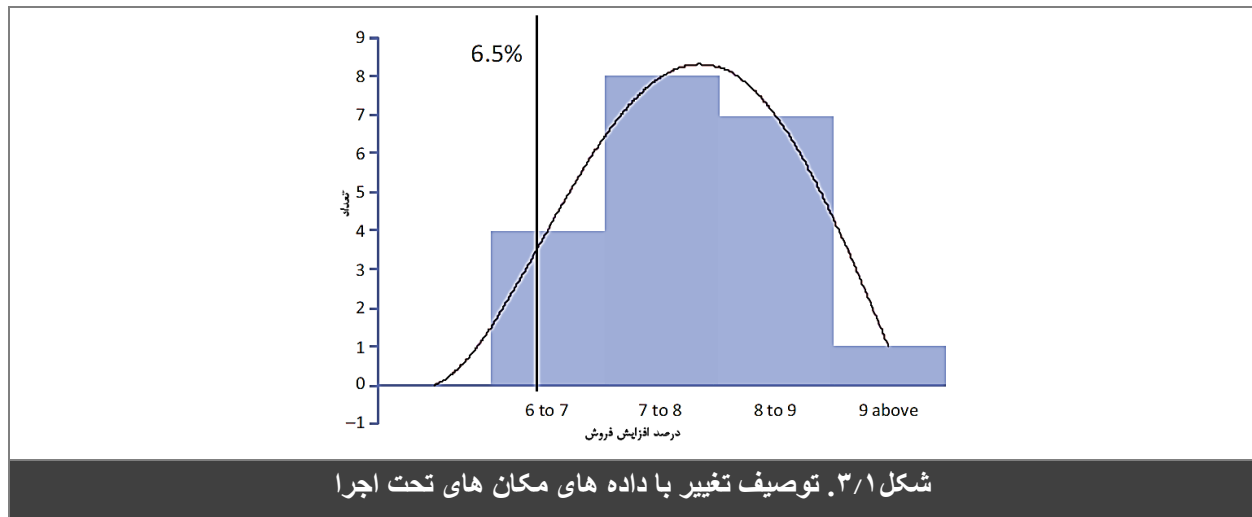
وقتی مسئله را در یکی از جلسات اولمان تدوین می کردیم، به طور خلاصه در مورد تغییر بحث کردیم.

جیم: بله، به یاد دارم. انحراف معیار یا همان مقدار تغییر را توضیح دادی و درباره ی مفهوم کنترل آماری صحبت کردی.

مانجو: بله. عجب حافظه ای! بیا حالا دقیق تر به این موضوع بپردازیم. من از ارقام مشابه استفاده می کنم (برای % فروش در مکان های مورد اجرا) که قبلا از آن استفاده کردیم. برای درک بهتر این موضوع، من از همان اعداد اینجا هم استفاده می کنم (شکل ۳،۱).

جیم: بسیار خب، عالیست.

مانجو: در این اعداد، بیا بگوییم هر مقدار کمتر از ۶/۵٪ برای مکان های تحت اجرا، قابل قبول نیست. به عبارت دیگر، این مقدار، عملکرد هدف را نشان می دهد. هر نوع انحراف از هدف (کمتر از ۶/۵٪ در این مورد)، نامطلوب است و تغییر نامیده می شود. به لحاظ کیفیت، هر انحراف از کیفیت عملکرد (مثلاً عملکرد افراد در روتین روزانه مثل ورزش و عادات غذایی) نامطلوب است و باید اصلاح شود.



جیم: کاملاً منطقی است. معمولاً وقتی شاخص های سلامتی را با تلفن همراه دنبال می کنم، ناهنجاری های ناشی از تغییر را می بینم. موافقم، خیلی مهم است.

مانجو: چه عالی جیم. **دقت کن** که در مورد متغیرهایی مثل قدرت و وزن، انحراف در صورتی اتفاق می افتد که مقادیر از هر دو طرف هدف منحرف شوند. مثلاً اگر وزن هدف ۱۵۰ پوند باشد، هر چیزی کمتر یا بیشتر از این عدد باعث تغییر می شود.

جیم: فهمیدم. در مثال وزن، مشکل اضافه وزن و کمبود وزن است.

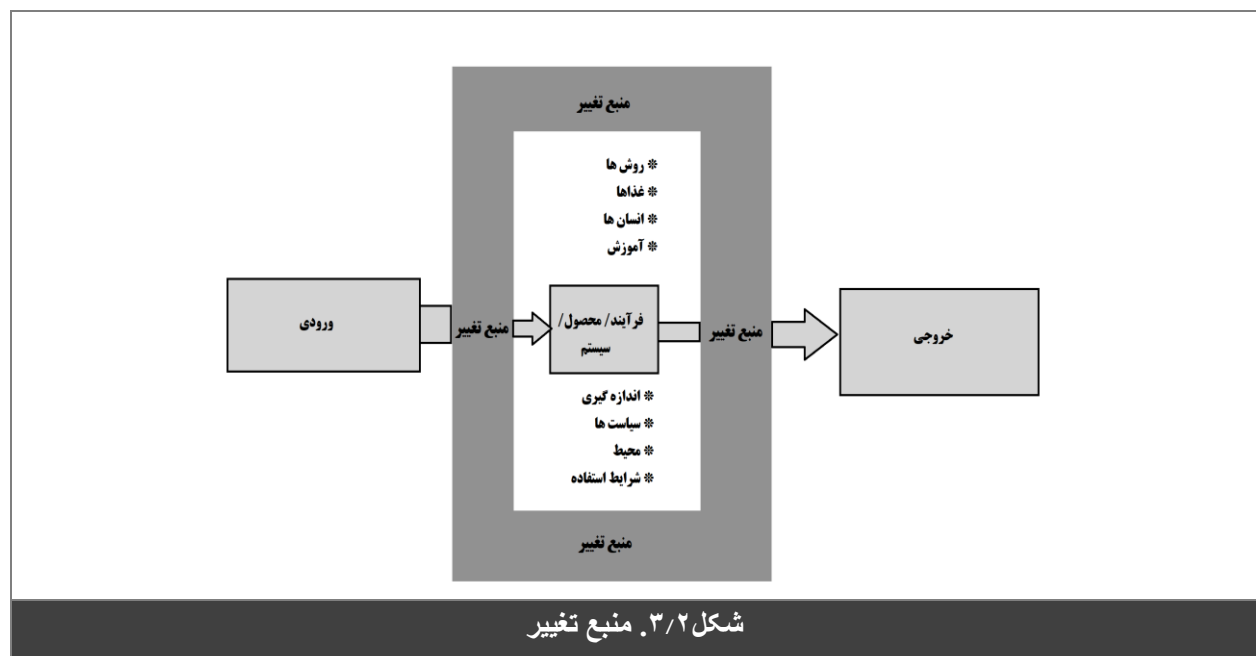
مانجو: درست فهمیدی. درک منبع تغییر طوری که بتوانیم با **اتکا به آنها عمل کنیم** و مشاهدات اندازه گیری هایمان را تا حد امکان سازگار و نزدیک به هدف انجام دهیم، به کاهش خطر ناشی از تغییر کمک زیادی می کند. **منابع تغییرات** همانطور که در شکل ۳،۲ نشان داده شده می تواند عوامل مختلفی مثل روش ها، غذاها، انسان ها، اندازه گیری ها و غیر باشد. شکل ۳،۲، نموداری از فرآیند معمولی است که ورودی، خروجی و منبع تغییر دارد.



در یک فرآیند / سیستم خاص، این فاکتورها تاثیرات متفاوتی خواهند داشت. کاهش خطر تغییر، به درک دقیق فرایند و عوامل مربوطه که تغییر ایجاد می کنند، نیاز دارد.

جیم: مفهوم تنوع جالب است. یک روش خوب برای نظارت بر تغییرات چیست؟

مانجو: سوال خوبی است جیم. کنترل فرآیند آماری (SPC) یک روش موثر برای نظارت بر تغییرات و درک میزان تغییرات است. SPC یک روش داده محور برای اندازه گیری سازگاری و اطمینان از پیش بینی پذیری فرآیندها است. این مفهوم را اولین بار والتر شیوارت<sup>5</sup> در نیمه ی اول قرن بیستم معرفی کرد. کنترل یک فرایند، با کاهش تغییر پذیری فرایند و تشخیص علت تغییر (علت های رایج و علت های خاص)، فرایند را قابل پیش بینی تر می کند. هدف SPC شناسایی تغییرات مربوط به فرایند است.



ابزار اولیه برای SPC ، نمودار کنترل است ؛ یک نمودار سری زمانی که مجموعه ای از اندازه گیری ها را نشان می دهد ، همراه با میانگین ، حد کنترل بالاتر ، و حد کنترل پایین تر ( UCL و lcl ) . این محدوده ها سه انحراف استاندارد (  $\sigma$  ) از میانگین اندازه گیری هستند .

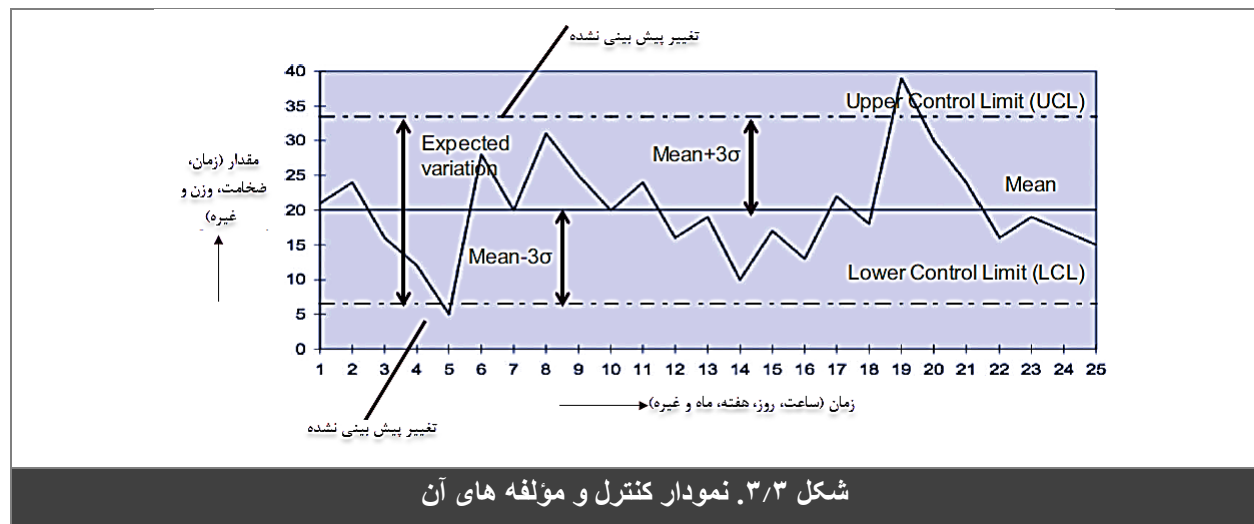
<sup>5</sup> Walter Shewhart

جیم: اوه. بله. یاد بحثم قبلمان درباره ی نمودار و انحراف معیار افتادم که درباره ی تعیین نقاط پرت ( یا داده های خارج از محدوده) با استفاده از این سه حد انحراف معیار حرف زدیم.

مانجو: بله جیم. اینجا هم از همان منطق استفاده کردم. پس نمودار SPC به شناسایی نقاط تغییر غیرعادی با غیرمنتظره (یعنی مقادیر بالای UCL یا پایین LCL) کمک می کند. شکل ۳،۳ مولفه های مختلف نمودار کنترل را نشان می دهد. نمودار کنترل، پایداری، قابلیت پیش بینی و ظرفیت فرایند را با نمایش بصری تغییر نشان می دهد. اگر اندازه گیری ها بین LCL و UCL باشد، می توانیم بگوییم اندازه گیری ها پایدار و قابل پیش بینی هستند، به عبارت دیگر، تحت کنترل هستند و این تغییر پیش بینی شده است. فاصله ی بین LCL و UCL سازگاری اندازه گیری ها را نشان می دهد. هر چه فاصله کمتر باشد به معنی سازگاری بهتر و ظرفیت بالاتر است و برعکس.

جیم: مانجو، حد کنترل را برای بررسی عملکرد سیستم/ فرایند چطور محاسبه می کنی؟

مانجو: معمولاً از داده های گذشته یا **پیشین (قدیمی)** برای ساخت حد کنترل استفاده می کنیم و از این حدها برای بازبینی و کنترل آینده استفاده می کنیم. با داده های **قدیمی** می توان میانگین یا متوسط انحراف معیار ( $\sigma$ ) را حساب کرد. **محدوده های کنترل** در سه برابر انحراف معیار از میانگین قرار داده می شوند. بنابراین، میانگین منهای سه برابر انحراف معیار، LCL و میانگین به علاوه سه برابر انحراف معیار، UCL را می دهد.



جیم: چرا باید سه برابر انحراف معیار را از دو طرف میانگین بگیریم؟

مانجو: سوال بسیار مهمی است. اگر سه انحراف معیار را در دو طرف میانگین در نظر بگیریم، به احتمال ۹۹٫۷۳٪ مشاهدات مشابه در LCL و UCL نهفته است. مشاهداتی که در محدوده LCL و UCL نیستند به دلیل برخی تغییرات علت خاص، داده ها خارج از محدوده (غیر واقعی) هستند. اکثر برنامه های آماری LCL و UCL را برای شما محاسبه و تنظیم می کنند. این فقط توضیح مختصری از معنای آن است.

جیم: فوق العاده است مانجو. وقتی پیشینه (بایگانی) داده ها را نداریم چه کار باید بکنیم؟

مانجو: این هم سوال خوبی است. وقتی داده ی گذشته را نداریم، باید گردآوری داده را در آن سیستم/ فرایند شروع کنی. وقتی داده ی کافی داشتی، حد کنترل را برای کنترل سیستم/ فرایند ایجاد می کنی.

جیم: متوجه شدم. (فهمیدم)

مانجو: در واقع، اگر داده ای داشته باشی، می توانم نشان دهم چطور یک نمودار کنترل بسازیم و چطور تفسیرش کنیم.

جیم: من داده هایی دارم که می توانم با تو به اشتراک بگذارم. این داده های تعداد خطاهای ۱۰۰ واحد در یکی از مکان های تغذیه سالم هنگام سرو غذای گیاهی در سه ماهه گذشته است. همین مکان، داده های فروش و داده های مربوط به سایر متغیرها را در اختیار ما قرار داد. این داده ها در جدول ۳٫۳ قرار دارد.

مانجو: عالیست! بیا برای این داده ها میانگین، انحراف معیار و حد کنترل را حساب کنیم. به این ترتیب، فایل داده ی صحیح را برای داده های فروش این مکان به دست آوردی؟

جیم: اجازه بده پیام هایم را چک کنم. بله. داده های فروش به روز شده را دریافت کردم. جزئیات در جدول ۳٫۴ است.

جدول ۳/۳. خطاهای هفته

خطا	هفته
۴	هفته ۱
۷	هفته ۲
۵	هفته ۳
۲	هفته ۴
۸	هفته ۵
۱۲	هفته ۶
۲	هفته ۷
۴	هفته ۸
۵	هفته ۹
۲	هفته ۱۰
۴	هفته ۱۱
۵	هفته ۱۲

مانجو: عالیست! با این داده ها می توانیم ببینیم بین تعداد خطا و تعداد فروش ارتباط وجود دارد یا نه.

جیم: مانجو من خیلی کنجکاویم که نتیجه را ببینم.

مانجو: قطعاً جیم. بیا مقادیر مورد نیاز را از این داده ها حساب کنیم. برای تعداد خطاها (داده های جدول ۳,۳) میانگین ۵، و انحراف معیار  $2/2$  است. پس  $UCL = 11/6 = 3 + 5 \times 2/2$  و  $LCL = -1/6 = 3 - 5 \times 2/2$  خواهد بود. در مثالی مثل این، نمی توانیم خطای منفی داشته باشیم؛ پس  $LCL$  را صفر در نظر می گیریم. نمودار کنترل را هم برای این داده های می سازم که در شکل ۳,۴ نشان داده شده است.

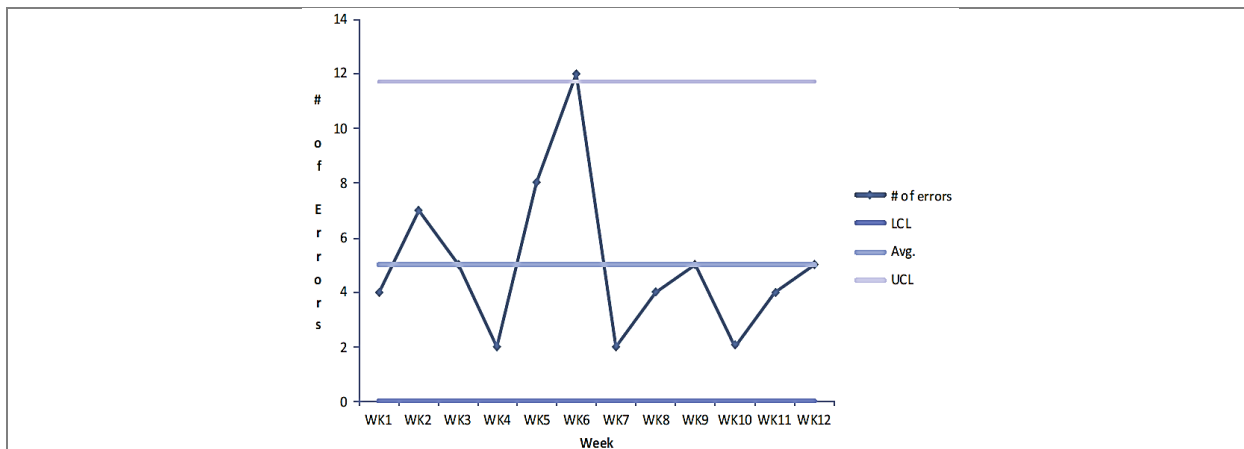
(مانجو چند دقیقه وقت گذاشت تا نمودار کنترل که در شکل ۳,۴ نشان داده شده و نمودار سری زمانی را برای % افزایش داده های فروش که در شکل ۳,۵ نشان داده شده، بسازد.)

مانجو: اگر به نمودار کنترل خطا نگاه کنی، به طور کلی اکثر خطاها کمتر از UCL هستند. هفته ۶ بالاترین و بالاتر از UCL بود. هفته ۵ نیز میزان خطای کمی بالاتر را نشان می دهد. این نشانه ها باید دلیل خاصی داشته باشند.

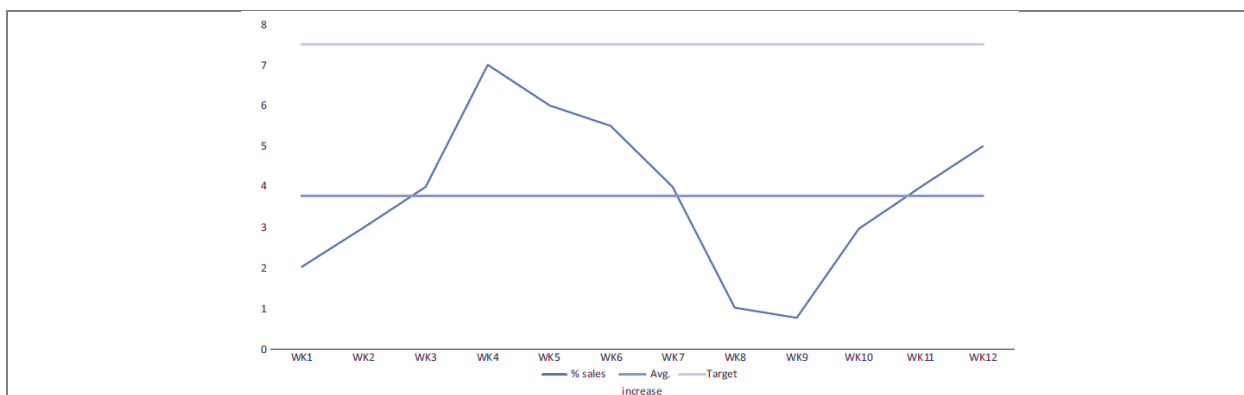
جیم: بله می بینم.

مانجو: اگر به داده های فروش جدول ۳,۴ و شکل ۳,۵ نگاه کنیم، میانگین درصد افزایش فروش ۳/۷۵٪ است. هفته های ۸ و ۹، درصد افزایش فروش بسیار پایینی دارند. گمان می کنم نرخ بالای هفته های ۵ و ۶ بر اعداد فروش هفته های بعد تأثیر گذاشته چون ممکن است مشتری راضی نبوده و برگشته باشد. جیم، در هفته های ۵ و ۶ اتفاق خاصی افتاده است؟

جدول ۳/۴. تعداد فروش هفتگی برای سه ماهه آخر (داده های اصلاح شده)		
هفته	واحد فروخته شده	% افزایش فروش
هفته قبل	۲۰۰۰	
هفته ۱	۲۰۴۰	۲
هفته ۲	۲۱۰۱	۲/۹۹
هفته ۳	۲۱۸۵	۳/۹۹
هفته ۴	۲۳۳۸	۷
هفته ۵	۲۴۷۸	۵/۹۸
هفته ۶	۲۶۱۴	۵/۴۸
هفته ۷	۲۷۱۸	۳/۹۷
هفته ۸	۲۷۴۵	۰/۹۹۳
هفته ۹	۲۷۶۶	۰/۷۶۵
هفته ۱۰	۲۸۴۸	۲/۹۶
هفته ۱۱	۲۹۶۲	۴
هفته ۱۲	۳۱۱۰	۴/۹۹



شکل ۳/۴. نمودار کنترل خطاها



شکل ۳/۵. % افزایش فروش هفتگی (هدف ۷/۵ % است).

جیم: این دو هفته، دو هفته ی اول ماه می است. اوه. بله! در این دو هفته ما افراد جدیدی در آن مکان استخدام کردیم. احتمالاً کامل به سیستم عادت نکرده بودند. می توانم با صحبت با مدیر آنجا، عمیقتر بررسی کنم.

مانجو: شما در حال حاضر می بینید برخی علل خاص ممکن را شناسایی کرده اید. این قدرت نمودار کنترل است. به علاوه، جیم یادت باشد این حد کنترل معمولاً برای کنترل آینده استفاده می شود و باید به طور دوره ای با جدیدترین داده ها به روز شوند.

جیم: چه جالب مانجو. قطعاً می توانم ارزش نمودار کنترل و این که چطور به درک تغییر کمک می کند ببینم. با این رویکرد، می توانیم عملکرد یک معیار را اندازه بگیریم و بر آن نظارت کنیم.

مانجو: کاملاً موافقم. در این زمینه، اجازه بدهید به سرعت سه جنبه از یک معیار را توضیح دهم: حقیقی، قابلیت و هدف برای داده‌های فروش.

درباره این جنبه‌ها قبلاً صحبت کردیم. برای درصد افزایش فروش در داده‌های جدول ۳,۴، عملکرد واقعی، متوسط فروش است که ۳/۷۵٪ است. اگر هفته‌ی ۸ و ۹ را کنار بگذاریم (تغییر با علت خاص)، میانگین ۴/۳۴٪ است که همان ظرفیت است. تا جایی که من می‌دانم هدف ۷/۵٪ است.

جیم: قطعاً. منطقی است.

مانجو: سوالی داری جیم؟

جیم: رویکرد نمودار کنترل برای یادگیری ماشین (ML) و هوش مصنوعی (AI) هم به کار می‌رود؟

مانجو: سوال خوبی است. یادگیری ماشین، شاخه‌ای از هوش مصنوعی (AI) است که با آن توانایی سیستم بر اساس تجربه و داده‌های گذشته افزایش می‌یابد. موضوع AI، ارائه‌ی هوش به سیستم به طوری است که به شکل خودکار تمام عملیات را اجرا کند. من فکر می‌کنم در مورد نمودار کنترل ما همین کار را در سطح خیلی ابتدایی انجام می‌دهیم. داده‌های پیشین را برای محاسبه‌ی حد کنترل به دست می‌آوریم و از آنها برای کنترل و نظارت آینده استفاده می‌کنیم. اگر این رویکرد خودکار شود، سیستم می‌تواند دوره‌ای، جدیدترین داده‌ها را بگیرد و حد کنترل را در صورت نیاز برای کنترل آینده به دست آورد. به عبارت دیگر می‌توانیم بر عملکرد بر اساس داده‌های گذشته با خودکار سازی، نظارت کنیم.

جیم: فکر می‌کنم منطقی است مانجو. بحث خطر و عدم اطمینانناشی از تغییر تمام شد؟

مانجو: بله جیم. حالا بیا درباره‌ی نوع سوم خطر و عدم اطمینانحرف بزنیم که به پیش بینی، تشخیص و تصمیم‌گیری ارتباط دارد.

## ➤ خطر و عدم اطمینان در پیش بینی، تشخیص و تصمیم گیری

خطر و عدم اطمینان در پیش بینی، تشخیص و تصمیم گیری با خطاهای مربوط به متغیرهای مورد استفاده در مدل ارتباط دارد. ما معمولاً متغیرهایی که باید تعیین شود را با  $Y$  (متغیر وابسته) و متغیرهای اثرگذار بر  $Y$  را با  $X_1, X_2, X_3, \dots$  (متغیرهای مستقل) نشان می دهیم و از رابطه ی تابعی زیر برای توصیفش استفاده می کنیم:

$$Y=f(X_1, X_2, X_3, \dots)$$

رابطه ی بالا به صورت  $Y$  تابعی از متغیرهای  $X_1, X_2, X_3$  و ... است بیان می شود.

جیم: گفتم که  $Y$  را می توان با کنترل  $X$  ها، کنترل کرد. می شود گاهی نتوانیم برخی از  $X$  ها را کنترل کنیم با این که بر  $Y$  تأثیر می گذارند. حرفم منطقی است؟

مانجو: قطعاً جیم. هر  $X$  را نمی توان کنترل کرد.  $X$  هایی که می توانیم کنترل کنیم، عوامل کنترل و  $X$  هایی که نمی توانیم کنترل کنیم عوامل غیرقابل کنترل یا اختلال هستند. مثلاً اگر متغیری که باید تعیین شود،  $Y$  زمان انتظار بر حسب دقیقه در رستوران باشد، پس  $X$  می تواند تعداد میزها، مکان، زمان یا روزی از هفته و غیره باشد. درک این عوامل / متغیرها برای تعیین یا داشتن اطلاعاتی در مورد زمان انتظار مهم است. این عوامل را می توان کنترل کرد و به همین دلیل معمولاً به آنها عوامل کنترل گفته می شود.

تغییر  $Y$  می تواند به دلیل این عوامل رخ دهد. تغییر در  $Y$  می تواند بخاطر عوامل دیگری که نمی توان کنترل کرد مثل قطع برق، شرایط آب و هوایی و ترافیک (به دلیل عدم حضور به موقع کارکنان رستوران) رخ دهد که قابل کنترل نیستند.

بنابراین، این عوامل، عوامل غیر قابل کنترل و یا عوامل اختلال نامیده می شوند. تغییرات ناشی از عوامل اختلال باید تا حد امکان با استراتژی های کاهش مناسب کاهش یابد. این همان کاری است که در رویکرد طراحی قوی تاگوچی انجام شده است.

جیم: من درباره ی تاگوچی شنیدم. روشش در کل جهان محبوب است، درست است؟

مانجو: درست است، جیم. دکتر تاگوچی، مهندس بزرگ ژاپنی، کارهای زیادی در زمینه طراحی قوی انجام داد. او به طور موثر آمار را در بسیاری از کاربردهای صنعتی به کار برد و کمک زیادی به حوزه مهندسی کیفیت کرد.



روش های او در سراسر جهان به کار گرفته شده است و شرکت های زیادی میلیون ها دلار به کمک استفاده از روش های او پس انداز کردند.

تاگوچی، ایده ی طرح قدرتمندی را مطرح کرد. هدف این طرح قوی، ارائه ی عملکرد سازگار با محصولات، سیستم ها و حتی مدل ها با به حداقل رساندن تأثیر عوامل غیرقابل کنترل یا عوامل اختلال است. دانش کافی در مورد عوامل غیر قابل کنترل و یا اختلال نیز بسیار مهم است تا طراحی را استحکام بخشد. تاگوچی، آزمایش عملکرد محصول یا سیستم یا مدل تحت شرایط اختلال شدید را هم پیشنهاد می کند تا عملکرد قدرتمندی در همه ی شرایط ایجاد شود.

جیم: فهمیدم. من فکر می کنم «محکم» از دیدگاه تاگوچی به معنای ایجاد سیستم یا محصولی مقاوم در برابر عوامل اختلال می باشد. ، در نتیجه محصول یا سیستم همیشه به همان شیوه عمل خواهد کرد.

مانجو: درست میگی جیم. با مقاوم کردن محصول یا سیستم خود نسبت به عوامل اختلال، به دلیل تغییرات فاکتور اختلال در مقابله با خطر یا در زمینه کاهش ضریب آن عمل می کنید.

به علاوه ما باید برای کنترل تغییرات ناشی از عوامل قابل کنترل هم برنامه داشته باشیم. بنابراین برای این نوع خطر تغییرات، برنامه ی کاهش، باید رابطه ی تابعی بین متغیر وابسته ای که باید تعیین شود ( $Y$ ) و عوامل مستقل ( $X_1, X_2, \dots$ ) شناخته شود و دانش کافی درباره ی این عوامل به دست آید تا اقدام مناسب لحاظ شود. توجه کن که  $X_1$  و  $X_2$  می توانند عوامل کنترل کننده و عوامل اختلال باشند.

جیم: فهمیدم.

مانجو: به نظر من، گاهی اوقات پیش‌بینی‌ها به دلیل عدم درک عوامل اختلال و تأثیرات آنها در شرایط حاد (نا مطلوب) شکست می‌خورد. اگر ما درک خوبی از عوامل اختلال و شرایط حاد مرتبط با آن داشتیم و اقدامات سریعی را علیه آنها انجام می‌دادیم، رویدادهایی مانند حملات ۱۱ سپتامبر و همه‌گیری ویروس کرونا، می‌توانستیم از اثرات نامطلوب آنها جلوگیری کنیم یا به حداقل برسانیم. هیچ کس به طور جدی فکر نمی‌کرد که یک حمله تروریستی ممکن است با استفاده از هواپیماهای تجاری رخ دهد. متعاقباً، دولت‌های تحت رژیم‌های مختلف برای جلوگیری از حملات بزرگ در خاک ایالات متحده، آگاه‌تر و فعال‌تر شدند.

جیم: فکر می‌کنم در مورد همه‌گیری کرونا می‌توان به نتایج مشابهی رسید.

مانجو: قطعاً. اگر پیش‌بینی می‌کردیم و اقدام فوری انجام می‌دادیم، اوضاع خیلی بهتر می‌شد. حالا بیا به سراغ نوع دیگری از خطر و عدم اطمیناندر مورد اجرای فرایند تحلیل برویم.

## ➤ خطر و عدم اطمینان در اجرای فرایند تحلیل

بیا فرض کنیم سعی می‌کنیم طعم برگ‌های گیاهی تغذیه سالم را با چند عامل ارزیابی کنیم و مدیریت می‌خواهد این کار را با تحلیل انجام دهد. برای این کار، مدیریت به ارائه ی یک مدل تحلیلی مناسب نیاز دارد که باید بخشی از فرایند اجرای تجزیه و تحلیل صحیح باشد. به نظرم، اغلب ما بدون داشتن رویکرد اجرایی خوب، تحلیل‌ها را انجام می‌دهیم. به نظرم اشتباه بزرگی است. وقتی شما یک رویکرد منطقی در طی اجرای تجزیه و تحلیل دارید، ما باید قادر باشیم خطر مرتبط با عناصر پردازش مختلف را درک کنیم.

جیم: بسیار خوب. چطور دقیقاً این کار را انجام دهیم؟

مانجو: بیاییم همان مثال ارزیابی طعم همبرگرهای گیاهی را در تصور کنیم. برای انجام این کار، مدیریت تغذیه سالم باید بداند چه نوع خطری را می‌تواند در حین اجرای عملیات تحلیلی پیش‌بینی کند تا مدلی برای ارزیابی ذائقه ارائه دهد. آنها ممکن است عناصر فرآیندی داشته باشند که در شکل ۳،۶ نشان داده شده است. ترفند در اینجا درک و امتیاز دادن به احتمال وقوع مسائل یا مشکلات مربوط به عناصر فرآیند است. امتیازدهی را می‌توان با استفاده از مقیاسی بین ۱ و ۷ انجام داد که ۱ نشان می‌دهد که مشکل ناشی از عنصر فرآیند تقریباً غیر ممکن است و ۷ نشان می‌دهد که احتمال آن بسیار زیاد است.

جیم: حدس می‌زنم که ما نیز باید با دقت بالایی امتیاز بدهیم، درست است؟

مانجو: درست است. ما باید شدت مسائل / مشکلات مربوط به عناصر فرایند را شناخته و امتیازدهی کنیم. مجدداً امتیازدهی را می‌توان با استفاده از مقیاسی بین ۱ تا ۷ انجام داد که ۱ نشان می‌دهد مشکل / مسئله ناچیز است و ۷ نشان می‌دهد مشکل / مسئله مهم است. بعد این دو امتیاز را ضرب می‌کنیم تا عوامل خطر (RF) را به دست آوریم. RF برای اولویت بندی مسائل ناشی از عناصر فرایند استفاده می‌شود. این کار کمک می‌کند تا اقدامات مناسب برای کاهش خطرات مرتبط انجام دهیم. پس چطور امتیاز دهی کنیم؟ امتیازات بر اساس بحث هایی است که در آن تیم متخصصان موضوع گرد هم جمع می‌شوند، همفکری می‌کنند و امتیازدهی می‌کنند.

جیم: چند وقت پیش در مورد آنالیز ( تجزیه و تحلیل) حالت شکست آموزش دیدم. از چیزی شبیه RF برای تعیین کمیت خطر ناشی از عوامل استفاده کردند.

مانجو: این نوع رویکرد اغلب در پروژه های بهبود فرایند برای تعیین و اولویت بندی خطر های پروژه استفاده می‌شود.

حال بیا ببینیم چه چیزی می‌توانیم از ماتریس ارزیابی خطر در شکل ۳,۶ استنباط کنیم. واضح است که «معیارهای تعریف شده» و «کیفیت داده» بالاترین مقادیر RF را دارند و بنابراین به عنوان عناصر فرآیند با عوامل خطر «بسیار بالا» طبقه بندی می‌شوند که ابتدا باید به آنها پرداخته شود و پس از آن اکتساب داده که به عنوان عنصر فرآیند "پرخطر" و غیره طبقه بندی می‌شود.

جیم: راه خیلی خوبی برای فهرست کردن و اولویت بندی خطر های مختلف است پس می‌توانیم اول روی مشکلات مهم کار کنیم. مرحله ی بعد چیست مانجو؟

ارزیابی خطر					ماتریس ارزیابی خطر	
اولویت بندی خطر	عامل خطر (RF)	امتیاز	شدت	امتیاز	احتمال وقوع	عناصر فرایند
بسیار بالا	۴۹	۷	مهم	۷	بسیار محتمل	ماتریس تعریف
بالا	۳۵	۷	مهم	۵	محتمل	دستیابی به داده
بسیار بالا	۴۹	۷	مهم	۷	بسیار محتمل	کیفیت داده
متوسط	۲۱	۷	مهم	۳	غیر محتمل	دسترسی ابزار مناسب
پایین	۳	۳	حاشیه ای	۱	بسیار محتمل	وابستگی فنی
	۱. بسیار محتمل؛ ۳- نامحتمل؛ ۵- محتمل؛ ۷- بسیار محتمل				احتمال وقوع	
	۱. شدت؛ ۳- حاشیه ای؛ ۵- قابل توجه؛ ۷- مهم				شدت	

شکل ۳/۶. ماتریس ارزیابی خطر

مانجو: حالا بیا روی خطر و عدم اطمینانناشی از اطلاعات ناقص تمرکز کنیم.

جیم: عالیست.

## ➤ خطر و عدم اطمینان ناشی از اطلاعات ناقص

مانجو: جیم، من فقط می‌خواهم روشن کنم، که این نوع ریسک مربوط به داده‌های ناقصی نیست که قبلاً به عنوان بخشی از مسائل کیفیت داده در موردش بحث کردیم.

جیم: بسیار خب. این نوع خطر با چه چیزی ارتباط دارد؟

مانجو: این خطر با خطر آماری مربوط به نمونه‌ها و خطر عدم اطلاع از نوع / روش مورد استفاده در مسئله‌ی مورد بررسی ارتباط دارد.

جیم: گرفتم

مانجو: خطر و عدم اطمینان ناشی از اطلاعات ناقص می‌تواند در دو سناریو مورد بحث قرار گیرد. یکی از آن‌ها خطر آماری با نمونه‌های ناکافی یا نامناسب است و دوم خطر ندانستن چیزی است ..

اجازه بده در قسمت اول - خطر آماری را مورد بحث قرار دهیم. ما معمولاً داده‌های نمونه را برای استنباط در مورد کل جمعیت می‌گیریم، کاری که با آمار استنباطی انجام می‌دهیم. وقتی جامعه‌ات با نمونه‌ی خوب نشان داده نمی‌شود، نتایج اشتباه به دست می‌آید و بنابراین، بر تصمیم‌گیری تأثیر می‌گذارد.

جیم: مانجو، نمونه‌ی خوب چیست؟

مانجو: به نظرم، نمونه‌ی خوب، داده‌هایی برای استخراج نتیجه‌های معتبر با درجه اطمینان بالا فراهم می‌کند که می‌توان بر اساس آن عمل کرد. بنابراین، این جنبه‌ی نمونه‌گیری خیلی مهم است و باید هنگام تحلیل داده‌ها به آن توجه کرد. معمولاً تکنیک‌هایی مثل نمونه‌گیری منطقی وجود دارد که برای انتخاب نمونه‌های بی‌طرفانه برای نشان دادن جامعه استفاده می‌شود.

جیم: اگر نمونه‌ی بی‌طرف خوبی داشته باشیم، می‌توانیم مشکلات مربوط به نمونه‌گیری را کاهش دهیم. فهمیدم.

مانجو: خوب. حالا بیا خلاصه درباره‌ی خطر شناختن روش / مدل مورد استفاده حرف بزنیم.

جیم: مشتاقانه منتظرم.

مانجو: بعد از انتخاب نمونه، احتمال خطر دیگری وجود دارد مثل این که چه مدل یا تکنیکی را باید برای مشکل مورد نظر استفاده کرد؟ تعریف واضح مسئله می تواند کمکمان کند سوالات درستی بپرسیم. برای پاسخ دادن به سوالات درست، باید روش / مدل مناسب انتخاب کنیم. در غیر این صورت، خطر ارائه ی پاسخ غلط و نتیجه ی آن تصمیم گیری بد هم وجود دارد. خیلی وقت ها، وقتی داده های خیلی زیادی داریم، استفاده از بسته ی نرم افزاری برای اجرای تحلیل بدون این که بدانیم تحلیل پاسخ صحیحی می دهد، تمایلی طبیعی است. گاهی افراد بدون اینکه بدانند چه کسانی از تحلیل ها استفاده می کنند، تحلیل می کنند. چنین اتفاقاتی بسیار زیاد است. بنابراین انتخاب یک روش / مدل مناسب برای انجام تحلیل های مورد نیاز و ارائه ی پاسخ مناسب بسیار مهم است. علاوه بر این، باید بدانید چه کسی از نتایج استفاده خواهد کرد. خطر ندانستن می تواند صدمات زیادی به بار می آورد. به عنوان مثال، می تواند مانع از دانستن این موضوع شود هر ناهنجاری بالقوه خطرناک است یا نه. اطلاع از این موضوعات، امکان می دهد که تصمیم بگیریم لازم است به آنها رسیدگی کنیم یا نه.

جیم: خیلی وقت ها، ما خیلی چیزها را بدیهی می دانیم و بدون این که کامل درموردشان بدانیم، عمل می کنیم. این اقدامات نتایج بد به دست می آورند و ممکن است خسارت های زیادی به بار بیاروند.

مانجو: چه قدر خوب گفتی. حالا بیا درباره ی نوع آخر خطر که ناشی از تأخیر است صحبت کنیم.

## ➤ خطر و عدم اطمینان ناشی از تأخیر

همانطور که قبلا ذکر کردم، این چیزی است که به نظر من بسیار مهم است. در زمینه علم داده، تصمیمات یا اقدامات تاخیری بر اساس یافته های تحلیلی به دو دلیل صدمات زیادی را به همراه خواهد داشت: (۱) مشکل همچنان وجود خواهد داشت و (۲) شما تمام تلاش خود (زمان، منابع، پول و ... را تلف می کنید). که برای حل مسئله صرف شده است. این روی روحیه افرادی که در جهت رفع این مشکل تلاش کرده اند نیز تأثیر می گذارد.

جیم: موافقم. این اتفاق بارها در زندگی کاریم افتاده با این که ربطی به تصمیمات تحلیلی نداشته است.

مانجو: بله جیم. این چیزها اتفاق می افتد. در صنایع یا دولتها وقتی رهبران نمی توانند برای یک مورد خاص به موقع اقدام کنند، اتفاق می افتد. دانستن کافی نیست. برای اطمینان از نتایج، باید اقدام کنی.

جیم: چه قدر خوب گفتمی مانجو. بحث درباره ی خطر و عدم قطعیت، بحث مهمی است. حالا لطفا بگو فکر می کنی چه خطر هایی در تغذیه سالم قابل استفاده است؟

مانجو: برای تغذیه سالم فکر می کنم کاربردی ترین خطر ها عبارتند از:

۱. خطر و عدم اطمیناندر خطای اندازه گیری داده ها؛
۲. خطر و عدم اطمینانناشی از وجود تغییر؛
۳. خطر و عدم اطمیناندر پیش بینی، تشخیص و تصمیم گیری.

برای کنترل خطر اول، باید مطمئن شوی که سیستم های اندازه گیری دقیق و معتبر هستند. خطای داده ها باید در مراحل ابتدایی اصلاح شوند تا درباره ی داده هایی که تحلیل می شوند، مطمئن شوی. قبلا با فایل داده هایت به این بخش پرداختیم. برای خطر ناشی از وجود تغییر، شاید بخواهی از رویکرد نمودار کنترل استفاده کنی و الگوهای مربوط به فروش را متناظر با مکانی که عملکرد خوبی ندارد، تحلیل کنی. با تحلیل نمودار کنترل به این خطر هم پرداختیم. برای خطر و عدم اطمینان در پیش بینی، تشخیص و تصمیم گیری، شاید بخواهی مطمئن شوی خطاهای متناظر با متغیرهای مربوطه (قابل کنترل و اختلال) در نظر گرفته شده اند.

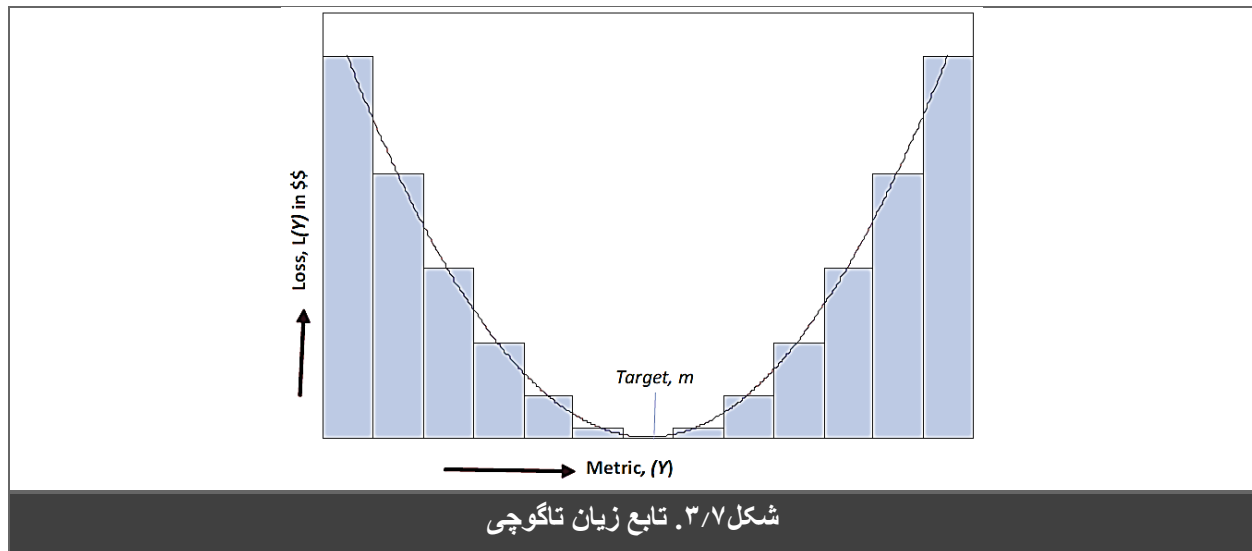
جیم: از توضیحاتت متشکرم مانجو. خیلی واضح بود.

مانجو: خواهش می کنم. پس مدیریت خطر و عدم اطمینانخیلی مهم است. خطر و عدم اطمیناندر محصول و عملکرد و نتایج تحلیل تغییر ایجاد می کند که بر اقدامات تصمیم گیری تأثیر معکوس می گذارد. در اینجا، مایلم بر اهمیت کارکرد از دست رفته تأکید کنم (تابع زیان). جنیچی تاگوچی، با مفهوم تابع زیان، ارتباطی بین تغییر و زیان جامعه برقرار کرد.

جیم: چه جالب. کنجکاو درباره ی مفهوم تابع زیان هم یاد بگیرم.

مانجو: بر اساس تابع زیان، اگر عملکرد محصول و نتایج تحلیل، از هدف منحرف شود، زیان وجود دارد. این زیان می تواند به شکل شکست محصول، نارضایتی مشتری، ورشکستگی شرکت، از دست دادن اعتبار و غیره باشد. این زیان در صورتی که از یک طرف هدف انحراف وجود داشته باشد، مثل آنچه در شکل ۳,۷ نشان داده شده، بیشتر می شود. در این شکل،  $Y$  معیار متناظر با عملکرد محصول یا نتایج تحلیلی را نشان می دهد و  $L(Y)$  زیان جامعه است. در اینجا جامعه، مشتریان، سازمان ها، دولت ها و غیره هستند. وقتی  $Y$  مساوی با  $m$  مقدار هدف باشد،

زیان صفر است. لطفا توجه کن چیزی که در شکل ۳,۷ می بینی، تابع زیان ایده آل است که متقارن است. در عمل، تابع زیان، کاملا متقارن نیست زیرا محاسبات براساس تخمین ها هستند.



جیم: مانجو، یک سوال دارم. برای معیار مثل معیار ما، % افزایش فروش، چطور از تابع زیان استفاده می کنی؟  
 مانجو: سوال فوق العاده ای است جیم. در این موقعیت ها، تنها یک طرف این تابع زیان قابل استفاده است. تنها در صورتی زیان وجود دارد که % افزایش فروش کمتر از مقدار هدف باشد. اگر بالاتر از هدف باشد، موقعیت خوبی است چون تعداد فروش بیشتر را می خواهی.

در مورد تغذیه سالم برای درصد افزایش فروش، اگر مقدار هدف ۷,۵٪ باشد، هر چیزی کمتر از ۷,۵٪ باعث انحراف و در نتیجه ضرر می شود و بنابراین، فقط سمت چپ تابع ضرر (شکل ۳,۵) قابل اعمال است.

به همین شکل، موقعیت هایی وجود دارد که سمت راست تابع زیان برایشان قابل استفاده است. اگر معیارت میزان نقص باشد، هر چیزی کمتر از هدف، مطلوب و هر چیزی بالاتر از هدف نامطلوب است و بنابراین، سمت راست تابع زیان قابل استفاده است.

جیم: سپاسگزارم مانجو. واوو. افق دیدم گشوده تر شد. با تابع زیان همه می توانند اهمیت انحراف (تغییر) از هدف و تأثیرش را بر جامعه درک کنند. در مورد تغذیه سالم، برای مکان هایی که عملکرد ضعیف دارند، من فکر می کنم می توان گفت که ضرر می تواند ناشی از انحراف از عملکرد کلی باشد و می تواند با عواملی مثل عدم شناخت ترجیحات مشتری، طعم بد و غیره ارتباط داشته باشد، درست است؟



مانجو: قطعاً. خوشحالم که سعی می کنی بین موقعیت شرکت و این ایده ها ارتباط برقرار کنی.

جیم: بله. مهم است. این طور بهتر می شود یادگرفت.

مانجو: موافقم. حالا بیا مرحله ۳ رویکرد کلی مان یعنی ایجاد و تحلیل مدل ها را شروع کنیم.