**Phân tích Kinh doanh Thương mại Điện tử bằng Công cụ Mô hình Hóa Thống kê**

**Tiêu đề khóa học:** CS504049 – Business Intelligence (BI)  
**Giảng viên:** Phúc H. Dương  
**Địa điểm:** Cơ sở Tân Phong

**Bối cảnh**

Một công ty thương mại điện tử muốn phân tích mô hình mua hàng của khách hàng để cải thiện chiến lược tiếp thị. Công ty đã thu thập dữ liệu về doanh số bán hàng hàng ngày (tính bằng đô la) trong 6 tháng qua. Bằng cách áp dụng các công cụ mô hình hóa thống kê, họ hy vọng sẽ thu thập được thông tin chi tiết về hành vi của khách hàng và tối ưu hóa các nỗ lực quảng cáo của mình.

**Ứng dụng của các Công cụ Mô hình Hóa Thống kê**

1. **Giá trị trung bình (Doanh số trung bình)**
   * Ứng dụng: Công ty tính toán doanh số bán hàng trung bình hàng ngày để hiểu rõ doanh thu trung bình trong một ngày điển hình.
   * Lợi thế: Cung cấp cái nhìn tổng quan về hiệu suất bán hàng của công ty theo thời gian.
   * Hạn chế: Giá trị trung bình dễ bị ảnh hưởng bởi các giá trị ngoại lệ, chẳng hạn như những ngày có doanh số bán hàng cực kỳ cao hoặc thấp (ví dụ, tăng đột biến trong các dịp lễ), điều này có thể không phản ánh doanh số điển hình.
2. **Trung vị (Giá trị giữa của Doanh số)**
   * Ứng dụng: Để giảm thiểu tác động của các giá trị ngoại lệ, công ty cũng tính toán trung vị của doanh số bán hàng hàng ngày.
   * Lợi thế: Trung vị phản ánh chính xác hơn xu hướng trung tâm khi dữ liệu có giá trị ngoại lệ.
   * Hạn chế: Trung vị không phản ánh sự phân phối toàn bộ của dữ liệu và có thể không nắm bắt được sự lan tỏa của doanh số bán hàng.
3. **Mode (Giá trị doanh số phổ biến nhất)**
   * Ứng dụng: Mode được sử dụng để tìm ra giá trị doanh số phổ biến nhất trong giai đoạn quan sát.
   * Lợi thế: Giúp xác định con số bán hàng điển hình nhất, điều này hữu ích cho mục đích quản lý hàng tồn kho và nhân sự.
   * Hạn chế: Mode có thể không tồn tại hoặc không hữu ích nếu dữ liệu không có giá trị lặp lại hoặc có nhiều mode.
4. **Phương sai (Thước đo về sự lan tỏa của Doanh số)**
   * Ứng dụng: Công ty tính toán phương sai để hiểu rõ mức độ biến động của doanh số bán hàng hàng ngày.
   * Lợi thế: Giúp nắm bắt mức độ phân tán của doanh số bán hàng, giúp công ty hiểu được những biến động trong doanh thu.
   * Hạn chế: Phương sai được tính bằng các đơn vị bình phương, điều này có thể khó giải thích trực tiếp theo đơn vị đô la.
5. **Độ lệch chuẩn (Sự biến thiên của doanh số bán hàng theo đơn vị đô la)**
   * Ứng dụng: Công ty sử dụng độ lệch chuẩn để đánh giá mức độ doanh số lệch so với giá trị trung bình, tính bằng đô la.
   * Lợi thế: Cung cấp một thước đo dễ hiểu về sự biến động của doanh số bán hàng, cho phép công ty chuẩn bị cho những biến động lớn.
   * Hạn chế: Giống như phương sai, độ lệch chuẩn cũng nhạy cảm với các giá trị ngoại lệ và có thể không phản ánh đầy đủ sự phân bố của dữ liệu nếu có các giá trị bán hàng cực đoan.
6. **Mean Absolute Deviation (MAD)**
   * Ứng dụng: Để cung cấp một thước đo mạnh mẽ về sự phân tán, công ty tính toán MAD, ít nhạy cảm với các giá trị ngoại lệ hơn so với phương sai và độ lệch chuẩn.
   * Lợi thế: Cung cấp một thước đo dễ hiểu và ít bị thiên lệch hơn về sự biến động của doanh số bán hàng khi có các giá trị ngoại lệ.
   * Hạn chế: MAD không tính đến hướng của sự lệch (dù trên hay dưới giá trị trung bình).
7. **Biểu đồ Hộp và Râu (Biểu diễn trực quan của phân phối Doanh số)**
   * Ứng dụng: Công ty sử dụng biểu đồ Hộp và Râu để đánh giá trực quan sự phân phối của doanh số, xác định các giá trị ngoại lệ và hiểu rõ sự lan tỏa của dữ liệu bán hàng.
   * Lợi thế: Nó thể hiện hiệu quả xu hướng trung tâm, sự phân tán và các giá trị ngoại lệ trong một biểu đồ duy nhất, giúp dễ dàng truyền đạt phát hiện.
   * Hạn chế: Mặc dù biểu đồ rất trực quan, nhưng nó có thể đơn giản hóa quá mức phân phối dữ liệu và một số bên liên quan có thể gặp khó khăn trong việc giải thích nếu không có kiến thức thống kê.

**Thông tin Chi tiết thu được từ Phân tích**

* Doanh số trung bình khá cao, nhưng trung vị cho thấy hầu hết các ngày có doanh số trung bình, chỉ ra rằng có một số ngày có doanh số cực kỳ cao (có thể là do các đợt khuyến mại).
* Mode cho thấy rằng doanh số bán hàng trong một số ngày nhất định (chủ yếu vào cuối tuần) lặp lại nhiều hơn.
* Phương sai và độ lệch chuẩn cho thấy sự biến động đáng kể trong doanh số bán hàng, điều này đã thúc đẩy công ty tập trung vào việc ổn định doanh thu.
* Khi sử dụng MAD, công ty nhận ra rằng hầu hết các biến động so với giá trị trung bình là vừa phải, nhưng các đợt tăng đột biến đóng góp vào sự biến động chung.
* Biểu đồ Hộp và Râu đã xác định một số giá trị ngoại lệ, chẳng hạn như doanh số vào các ngày Black Friday và Cyber Monday, ảnh hưởng mạnh đến toàn bộ phân tích doanh số.

**Kết luận**

Mỗi công cụ thống kê đều mang lại những thông tin chi tiết độc đáo về mô hình bán hàng của công ty. Tuy nhiên, công ty đã nhận ra rằng không có công cụ đơn lẻ nào có thể giải thích hoàn toàn dữ liệu và sự kết hợp của các công cụ này đã cung cấp cái nhìn toàn diện hơn về hiệu suất của doanh nghiệp.

**Câu hỏi thảo luận**

1. Làm thế nào mà giá trị trung bình khác với trung vị trong việc cung cấp thông tin chi tiết về hiệu suất bán hàng của công ty thương mại điện tử, và tại sao việc sử dụng cả hai là quan trọng?

**+ Trung bình** phản ánh giá trị doanh số tổng thể bằng cách tính tổng doanh số chia cho số lượng ngày, giúp thấy được tổng thể về mức doanh thu hàng ngày. Tuy nhiên, nó dễ bị ảnh hưởng bởi các **giá trị ngoại lệ**, chẳng hạn như những ngày có doanh số đột biến cao hoặc thấp (ví dụ: Black Friday), do đó có thể không đại diện cho các ngày bán hàng điển hình.

**+ Trung vị** đại diện cho giá trị chính giữa của phân phối doanh số, không bị ảnh hưởng bởi các ngoại lệ. Điều này có nghĩa là nó có thể đưa ra bức tranh rõ ràng hơn về **xu hướng trung tâm thực sự** của doanh số, đặc biệt khi có sự biến động lớn.

**+ Tại sao cần sử dụng cả hai?**: Trung bình cho thấy **hiệu suất chung**, trong khi trung vị cho thấy **xu hướng trung tâm thực tế**. Khi dữ liệu có các giá trị ngoại lệ lớn, trung bình có thể bị méo mó và không phản ánh đúng doanh thu bình thường, do đó trung vị sẽ cung cấp cái nhìn chính xác hơn về phần lớn doanh số hàng ngày.

1. Tại sao công ty có thể muốn sử dụng Mean Absolute Deviation (MAD) thay vì phương sai hoặc độ lệch chuẩn khi đánh giá sự biến động của doanh số bán hàng?

+ **MAD (Độ lệch tuyệt đối trung bình)** là một thước đo về sự biến động nhưng **ít bị ảnh hưởng bởi các giá trị ngoại lệ** so với phương sai và độ lệch chuẩn. Nó đơn giản tính độ lệch trung bình của các giá trị so với trung bình, không quan tâm đến hướng (lớn hơn hay nhỏ hơn giá trị trung bình).

**+ Phương sai và độ lệch chuẩn** đều tính toán độ biến động của dữ liệu nhưng có xu hướng **phóng đại sự ảnh hưởng của các giá trị ngoại lệ** (do tính bình phương của các sai số trong công thức tính). Điều này làm cho các giá trị cực đoan có tác động lớn hơn đến kết quả.

+ **Tại sao nên chọn MAD?**: MAD cung cấp một cái nhìn **ít thiên lệch** hơn về sự biến động của dữ liệu, nhất là khi dữ liệu có sự phân tán cao hoặc có các giá trị ngoại lệ lớn (như doanh số tăng đột biến vào các dịp khuyến mại). Nó cung cấp một thước đo ổn định hơn về sự dao động hàng ngày mà không bị ảnh hưởng quá mức bởi các ngoại lệ.

1. Trong những ứng dụng thực tế nào khác, Biểu đồ Hộp và Râu có thể đặc biệt hữu ích, và tại sao?

+ **Biểu đồ Hộp và Râu** rất hữu ích trong nhiều ứng dụng để **trực quan hóa sự phân bố dữ liệu**, bao gồm:

* **Phân tích tài chính**: Giúp so sánh phân phối lợi nhuận hoặc giá cổ phiếu của các công ty, xác định các giá trị ngoại lệ hoặc sự biến động bất thường.
* **Phân tích y tế**: Dùng để so sánh phân phối các chỉ số sức khỏe (như huyết áp, mức đường huyết) giữa các nhóm bệnh nhân, phát hiện sự khác biệt giữa các nhóm dân số.
* **Giáo dục**: Giúp phân tích phân phối điểm số thi của học sinh trong các kỳ thi khác nhau, so sánh giữa các trường hoặc khu vực khác nhau.
* **Phân tích sản xuất**: Được sử dụng để theo dõi sự ổn định của quy trình sản xuất, xác định các lô hàng không đạt tiêu chuẩn hoặc các vấn đề trong quá trình sản xuất.

+ **Lợi ích chính**: Biểu đồ Hộp và Râu có khả năng thể hiện **xu hướng trung tâm**, **sự lan tỏa**, và **các giá trị ngoại lệ** của dữ liệu trong một biểu đồ đơn giản. Điều này làm cho nó trở thành một công cụ hữu ích để **so sánh nhiều tập dữ liệu** cùng một lúc và dễ dàng xác định các điểm bất thường trong dữ liệu.

1. Những hạn chế của việc chỉ sử dụng các thống kê mô tả (chẳng hạn như trung bình, trung vị, mode) trong việc đưa ra quyết định kinh doanh là gì, và những hạn chế này có thể được giải quyết như thế nào?

+ Hạn chế của thống kê mô tả:

* Chỉ cung cấp cái nhìn tổng quan về dữ liệu và không tính đến mối quan hệ giữa các biến số.
* Không giải thích được nguyên nhân của sự biến động hoặc các yếu tố ảnh hưởng đến kết quả.
* Không thể dự đoán được xu hướng trong tương lai hay mô hình hóa các mối quan hệ phức tạp.
* Thiếu thông tin về tương quan và tác động của các yếu tố khác nhau đến hiệu suất kinh doanh.

+ Giải pháp: Để khắc phục các hạn chế này, cần sử dụng các phương pháp phân tích mô hình hóa dự báo và phân tích hồi quy, chẳng hạn như:

* Hồi quy tuyến tính: Để xác định mối quan hệ giữa các yếu tố và kết quả doanh số bán hàng.
* Mô hình chuỗi thời gian: Để dự đoán xu hướng trong tương lai và phát hiện các mẫu chu kỳ theo thời gian.
* Phân tích dữ liệu nâng cao: Sử dụng các kỹ thuật như phân cụm, phân tích nhân tố để hiểu rõ hơn về hành vi của khách hàng và xác định các yếu tố chính tác động đến doanh thu.
* Bằng cách kết hợp thống kê mô tả với các kỹ thuật phân tích tiên tiến, công ty có thể có cái nhìn toàn diện hơn về dữ liệu, từ đó đưa ra các quyết định kinh doanh chính xác hơn.