

BTVN Buổi 4 (BT1): Type I và Type II Error trong Kiểm Định Giả Thuyết

Tác giả: Đinh Thiên Ân - 22520010

Giảng viên hướng dẫn: Cử nhân Trần Quốc Khánh

1. Phân biệt Type I và Type II Error

Type I Error (Lỗi loại I)

- **Định nghĩa:** Lỗi loại I xảy ra khi chúng ta bác bỏ giả thuyết không (H_0) trong khi thực tế giả thuyết không là đúng. Đây còn được gọi là "false positive" (dương tính giả).
- **Hậu quả:** Dẫn đến kết luận sai lầm rằng có một hiệu ứng hoặc mối quan hệ khi thực tế không có.

Type II Error (Lỗi loại II)

- **Định nghĩa:** Lỗi loại II xảy ra khi chúng ta không bác bỏ giả thuyết không (H_0) trong khi thực tế giả thuyết không là sai (giả thuyết đối lập H_1 là đúng). Đây còn được gọi là "false negative" (âm tính giả).
- **Hậu quả:** Dẫn đến việc bỏ lỡ phát hiện một hiệu ứng hoặc mối quan hệ có thật.

2. Ba ví dụ về Type I và Type II Error trong **đề án phân tích dữ liệu việc làm** của nhóm chúng em

Ví dụ 1: Kiểm định sự khác biệt về mức lương giữa các ngành nghề

Trong dự án phân tích dữ liệu việc làm tại Việt Nam, chúng ta muốn kiểm định xem liệu có sự khác biệt đáng kể về mức lương trung bình giữa ngành IT và ngành Tài chính hay không.

- **Giả thuyết không (H_0):** Không có sự khác biệt về mức lương trung bình giữa ngành IT và ngành Tài chính.
- **Giả thuyết đối lập (H_1):** Có sự khác biệt về mức lương trung bình giữa ngành IT và ngành Tài chính.

Type I Error: Kết luận rằng có sự khác biệt về mức lương giữa hai ngành, trong khi thực tế không có. Điều này có thể dẫn đến các quyết định sai lầm về chiến lược tuyển dụng, định hướng nghề nghiệp, hoặc xây dựng cơ cấu lương không phù hợp.

Type II Error: Không phát hiện ra sự khác biệt về mức lương giữa hai ngành, trong khi thực tế có sự khác biệt. Điều này có thể khiến người tìm việc bỏ lỡ cơ hội lựa chọn ngành nghề có mức lương cao hơn, hoặc nhà tuyển dụng đưa ra mức lương không cạnh tranh.

Ví dụ 2: Phân tích mối quan hệ giữa kinh nghiệm và mức lương

Trong nghiên cứu, chúng ta muốn kiểm định mối quan hệ giữa số năm kinh nghiệm và mức lương.

- **Giả thuyết không (H_0):** Không có mối quan hệ tuyến tính giữa số năm kinh nghiệm và mức lương.
- **Giả thuyết đối lập (H_1):** Có mối quan hệ tuyến tính giữa số năm kinh nghiệm và mức lương.

Type I Error: Kết luận rằng kinh nghiệm ảnh hưởng đến mức lương theo quan hệ tuyến tính, trong khi thực tế không có mối quan hệ này. Điều này có thể dẫn đến các chính sách nhân sự không hiệu quả, như chú trọng quá mức vào yếu tố kinh nghiệm trong việc xác định mức lương.

Type II Error: Không phát hiện ra mối quan hệ giữa kinh nghiệm và mức lương, trong khi thực tế có mối quan hệ này. Điều này có thể khiến các nhà tuyển dụng không đánh giá đúng giá trị của kinh nghiệm, dẫn đến việc mất nhân tài có kinh nghiệm cao.

Ví dụ 3: So sánh mức lương giữa Hà Nội và TP.HCM

Trong phân tích dữ liệu việc làm, chúng ta muốn kiểm định xem liệu mức lương trung bình ở Hà Nội có cao hơn đáng kể so với TP.HCM hay không.

- **Giả thuyết không (H_0):** Không có sự khác biệt về mức lương trung bình giữa Hà Nội và TP.HCM.
- **Giả thuyết đối lập (H_1):** Mức lương trung bình ở Hà Nội khác biệt so với TP.HCM.

Type I Error: Kết luận rằng có sự khác biệt về mức lương giữa Hà Nội và TP.HCM, trong khi thực tế không có sự khác biệt đáng kể. Điều này có thể dẫn đến quyết định di cư không cần thiết của người lao động, hoặc các chính sách phân bổ nhân lực không hợp lý của các công ty có nhiều chi nhánh.

Type II Error: Không phát hiện ra sự khác biệt về mức lương giữa hai thành phố, trong khi thực tế có sự chênh lệch đáng kể. Điều này có thể khiến người lao động bỏ lỡ cơ hội cải thiện thu nhập thông qua việc di cư đến khu vực có mức lương cao hơn, hoặc các doanh nghiệp không điều chỉnh chính sách lương phù hợp với thị trường địa phương, dẫn đến khó khăn trong việc thu hút nhân tài.