**ĐỌC SÁCH**

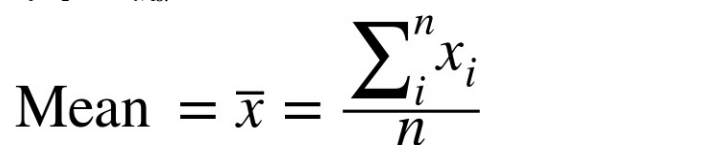
**PRACTICAL STATISTICS FOR DATA SCIENTIST**

**Estimates of Location**

*Các biến số với dữ liệu đo lường hoặc đếm có thể có hàng nghìn giá trị riêng biệt. Một bước cơ bản trong việc khám phá dữ liệu là tìm được giá trị điển hình cho mỗi đặc trưng*

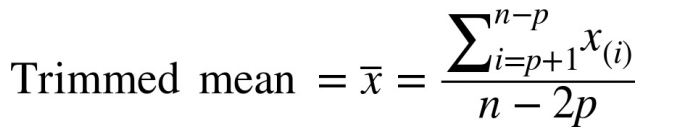
**Mean**

*Đơn giản nhất của estimate of location là mean – giá trị trung bình*.



**Trimmed mean**

*Đối với trimmed mean nó sẽ loại bỏ giá trị nhỏ nhất và lớn nhất. Sau đó tính trung bình với các giá trị còn lại.*

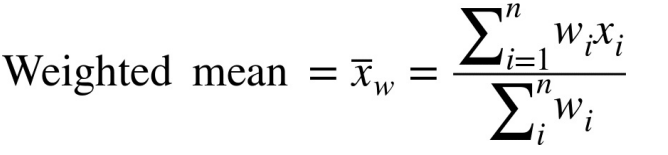


*Việc sử dụng Trimmed mean nhằm để loại bỏ các giá trị cực đoan. Ví dụ trong môn nhảy xa. Để tính điểm của một vận động viên người ta sẽ loại bỏ điểm của 2 giám khảo cho điểm cao nhất và thấp nhất. Sau đó lấy điểm trung bình của các giám khảo còn lại*

*Điều này làm cho việc 1 trọng tài đơn lẻ can thiệp vào điểm số trở nên khó khắn hơn.*

*Weight Mean*

*Tính weight mean bằng cách nhân mỗi giá trị dữ liệu với 1 trọng số và chia tổng của chúng cho tổng trọng số:*



*Ví dụ như trong việc tính GPA của sinh viên, điểm số của sinh viên sẽ nhân với trọng số là số tín chỉ của từng môn học và cuối cùng là chia cho tổng số tín chỉ, chính là tổng trọng số*

**Median**

*Trung vị là số ở giữa một list danh sách đã được sắp xếp. Nếu số lượng dữ liệu là chẵn thì giá trị trung vị có thể không nằm trong tập dữ liệu mà sẽ là một giá trị nào đó chia tập dữ liệu ra thành 2 nửa, nửa trên và nửa dưới . Ví dụ trong sách này là việc xem xét thu nhập của các hộ gia đình ở Medina và Windermere, việc sử dụng trung bình sẽ tạo ra kết quả khác biệt do Bill Gate sống ở Median. Tuy nhiên khi sử dụng trung vị, thì giá trị quan sát là không thay đổi*

**Outliers**

*Trung vị là 1 ước lượng mạnh mẽ vì nó k bị ảnh hưởng bởi giá trị ngoại lai có thể làm sai lệch kết quả. Một giá trị ngoại lai là một giá trị nằm xa các giá trị khác trong tập dữ liệu. Ví dụ như trường hợp Bill Gate ở trên*

**Example: Location Estimates of Population and Murder Rates**

**-------------------------CODING IN MY JUPITER NOTEBOOK-------------------------**