Properties of Maximum Likelihood Estimators

Assumptions

1. $(X_1, X_2, X_3 \ldots)$ is an infinite sequence of I.I.D. continuous random variables
2. $f(X|\theta)$ is a family of probability density functions, with parameter θ in a compact space
3. if $\theta_1 \neq \theta_2$, then the resulting distributions are different ($P\big(f(X|\theta_1)/f(X|\theta_1) \neq 1\big) > 0$)
4. f is continuous in θ.
5. The log-likelihood is integrable ($\mathrm{E}\big[|\ln f(X|\theta)|\big] < \infty$ for all θ)

MLE: Consistent when the Model is True

Theorem: Consistency of MLE

If $(X_1, X_2, X_3 \dots)$ are drawn from $f(X|\theta_t)$ for some parameter θ_t, then under assumptions 1-5,

$$\theta_{MLE} \xrightarrow{p} \theta_t$$

MLE: "Close" when the Model is False

Theorem: MLE and KL divergence

If $(X_1, X_2, X_3 \ldots)$ are drawn from some distribution g, and θ_c minimizes the KL divergence,

$$\text{KL}\left[f(\cdot|\theta) \| g\right] = \int_{-\infty}^{\infty} f(x|\theta) \log \frac{f(x|\theta)}{g(x)} dx$$

Then under assumptions 1-5,

$$\theta_{MLE} \xrightarrow{p} \theta_c$$

Extra Assumptions

6. The parameter θ_0 that maximizes likelihood is in the interior of the parameter space.
7. Likelihood is twice continuously differentiable.
8. Intergration and differentiation operators are interchangeable for likelihood.
9. The Fisher information, $I(\theta_0) = -\mathrm{E}[\frac{\partial^2}{\partial\theta^2} \ln f(X|\theta)|\theta = \theta_0]$ exists and is nonsingular.

MLE: Asymptotically Normal

Theorem: MLE is asymptotically normal

Under assumptions 1-9,

$$\frac{\theta_{MLE} - \theta_0}{\sqrt{n}} \xrightarrow{d} N\big(0, 1/I(\theta_0)\big)$$

MLE: Asymptotically Efficient

Theorem: MLE is asymptotically efficient

Under assumptions 1-9, θ_{MLE} is asymptotically efficient.

Summary of MLE Properties

- Consistent when the model is true
- As close as possible when the model is false
- Asymptotically normal
- Asymptotically efficient

Handling Outliers

Outliers

Outlier definition

An *outlier* is:

- A data point that is unusual (or unlikely)

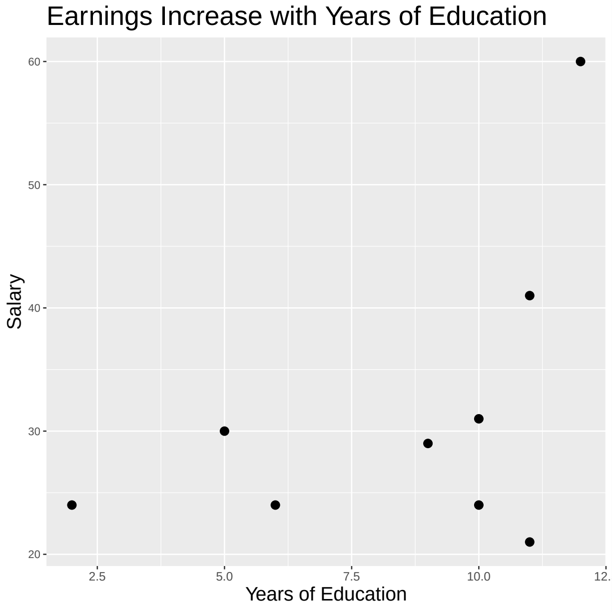
Outliers

Outlier definition

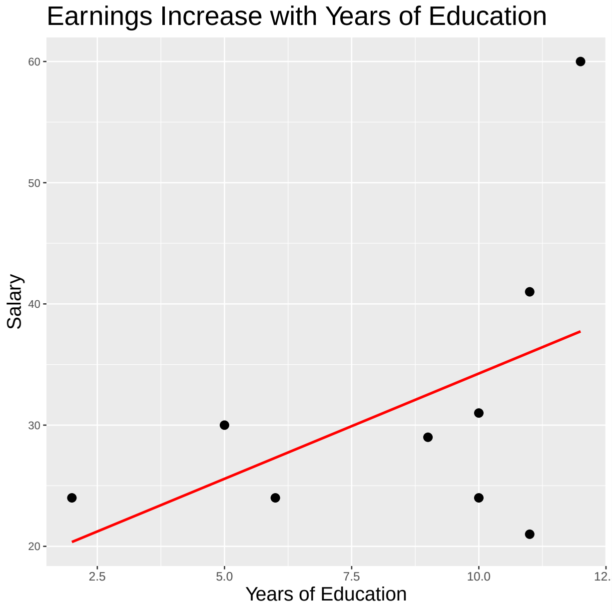
An *outlier* is:

- A data point that is unusual (or unlikely)
- With respect to some statistical model

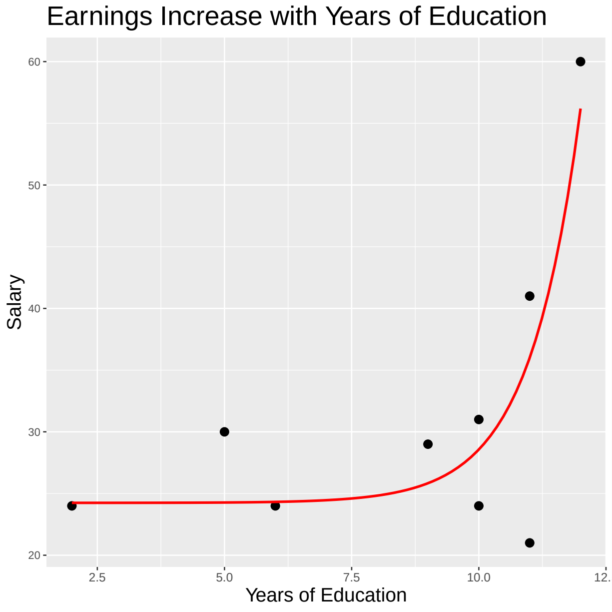
Where is the Outlier? Part I



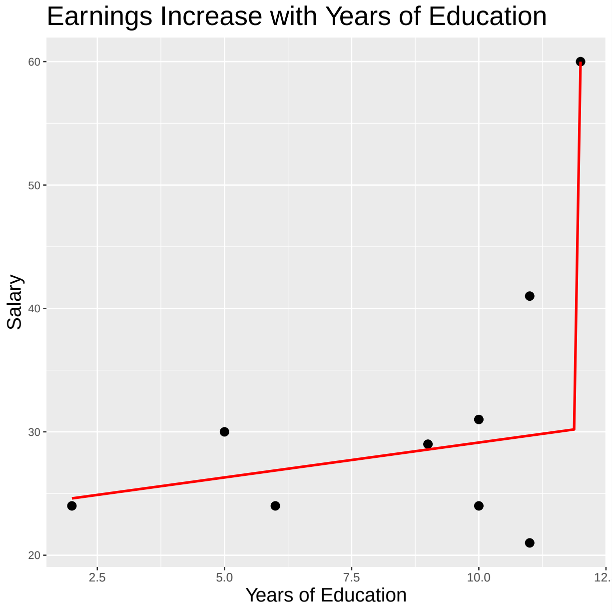
Where is the Outlier? Part II



Where is the Outlier? Part III



Where is the Outlier? Part IV



Lessons

- Outliers are interesting.
- An outlier could be telling you something important.
- **Never** remove a point just because it is an outlier.
    - This is altering the data to fit your idea of what the model is.

How to Handle an Outlier

Is the outlier a coding error, measurement error, or otherwise not a meaningful value from the distribution being studied?

- Then, remove the outlier and document your decision.

Is the outlier a meaningful value?

- Do not remove the outlier.

Alternatives to Removing Outliers

What are our options if we do *not* remove outliers?

1. Apply a log or exponential transform to represent non-linearity.
2. Adopt a more flexible functional form (e.g., polynomial terms).
3. Use indicator variables to treat outlying points separately.
4. Compute outlier robust statistics.

Levels of Measurement

Measuring Soup

Codebook
1 = Shoyu
2 = Shio
3 = Miso

Observation	Broth
1	1
2	3
3	1
4	2
5	1

What is the average broth?

Random Variable

Given probability space (Ω, S, P), a *random variable* is a measurable function from Ω to $\mathbb{R}$.

- $\mathbb{R}$ usually means real numbers with field operations $(+, -, \cdot, ...)$.
- Some operations may not be sensible.

Levels of Measurement

Level	Operations
Nominal	=
Ordinal	$<, =, >$
Interval	$-$, average, weighted average
Ratio	All arithmetic operations

- Stanley Smith Stevens

Nominal Variables

A *nominal variable* defines categories

- Can check if two values are **equal** or **not equal**

Other operations may not be defined

- $<, >$
- $+, -, \cdot, /$

What Can You Do with Nominal Variables?

Valid:

- Value Counts
- Mode

Requires more structure:

- Median
- Mean
- Variance

Ordinal / Rank Variables

An *ordinal variable* defines categories that have an order.

- Can apply $<, =, >$

Intervals may not be meaningful.

- No $+, -, \cdot, /$

What Can You Do with Ordinal Variables?

Valid:

- Median, Quantiles
- Probabilities that A > B, etc.

Requires more structure:

- Mean
- $+, -, \cdot, /$
- Notions of "shape" of distribution

Likert Scales

The interface was easy to navigate

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	2	3	4	5

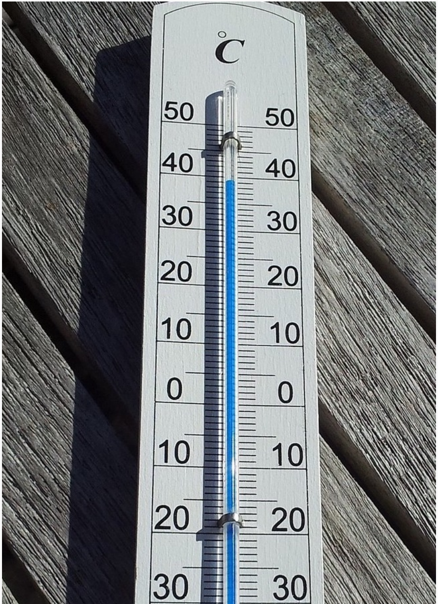
Interval / Metric Variables

An *interval variable* defines categories with consistent intervals

- Can apply $-$, average

May have no notion of zero

- No $+, \cdot, /$



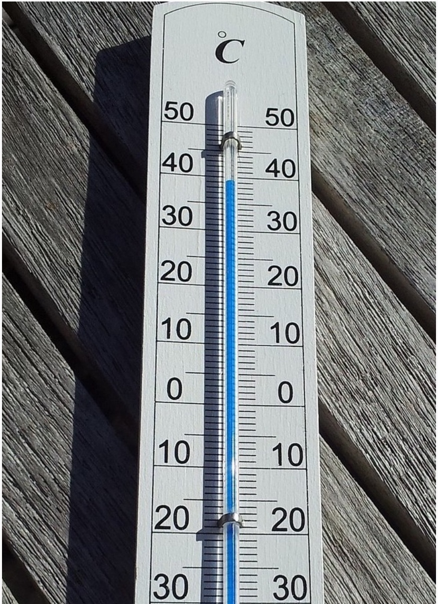
What Can You Do with Interval Variables?

Valid:

- Mean
- Variance
- Most common statistics

Requires more structure:

- Log
- Root-Mean-Square



Ratio Variables

A *ratio variable* defines categories with consistent intervals and a meaningful zero.

- Can apply all arithmetic operations.

Levels of Measurement

Level	Operations
Nominal	=
Ordinal	$<, =, >$
Interval	$-$, average, weighted average
Ratio	All arithmetic operations

- Stanley Smith Stevens