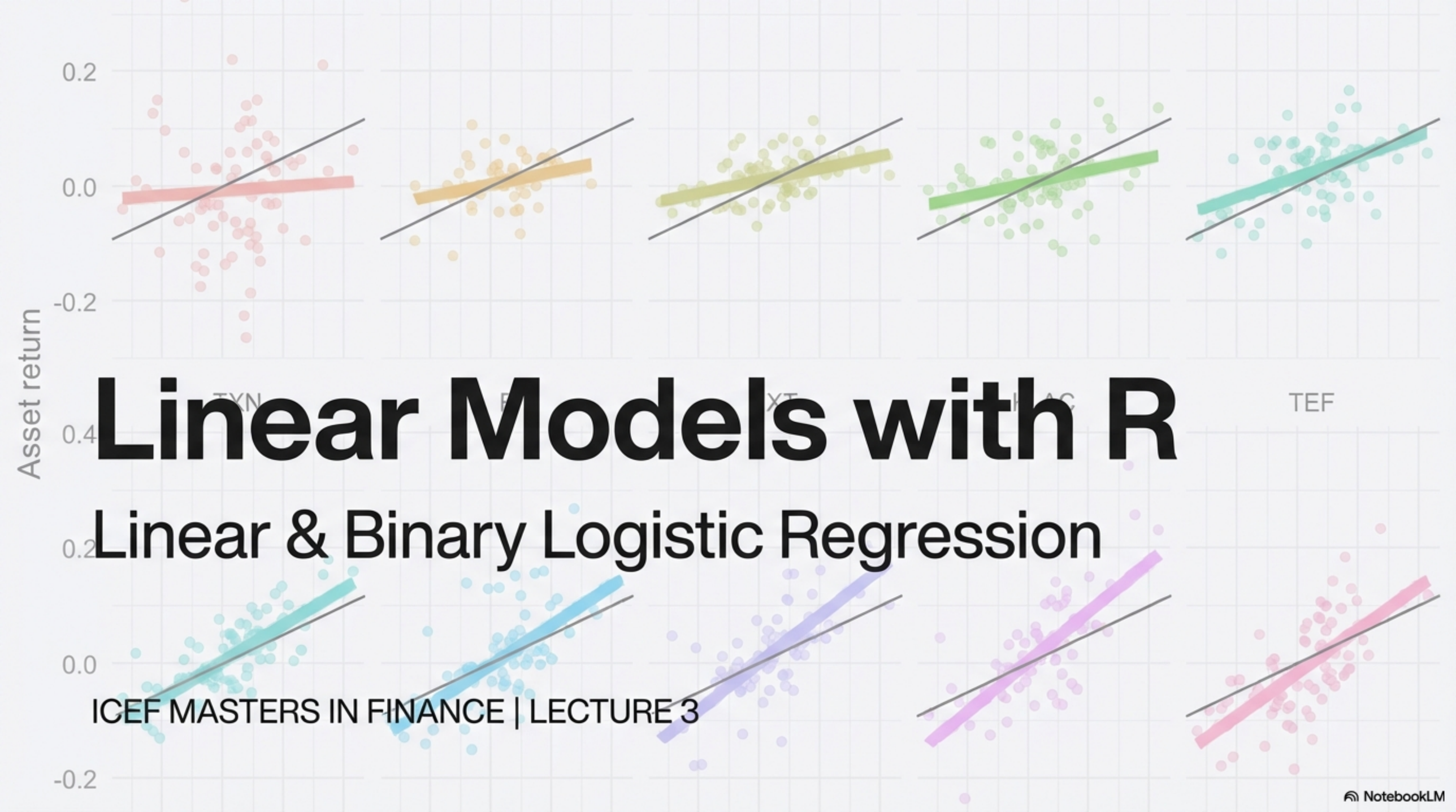
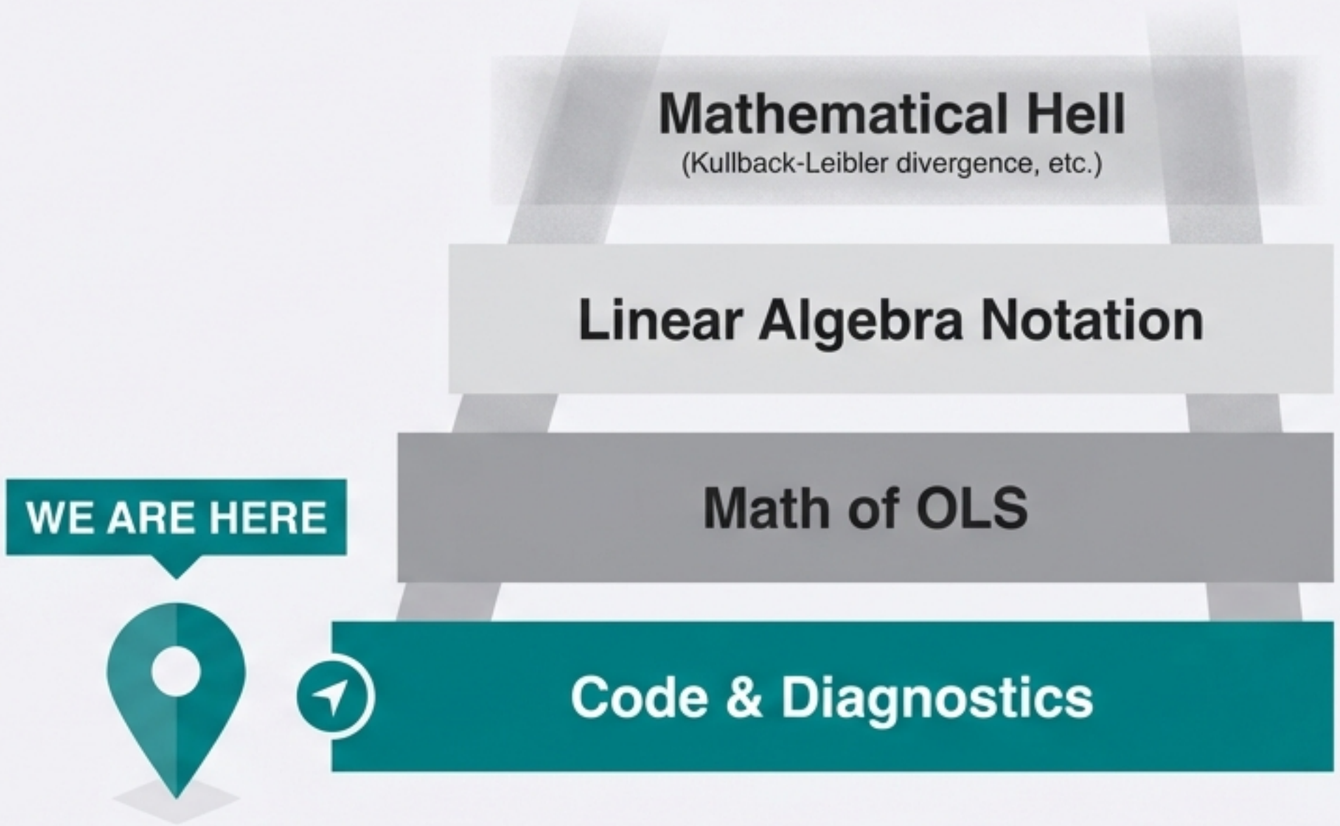


Linear Models with R

Linear & Binary Logistic Regression

ICEF MASTERS IN FINANCE | LECTURE 3

The Complexity Ladder

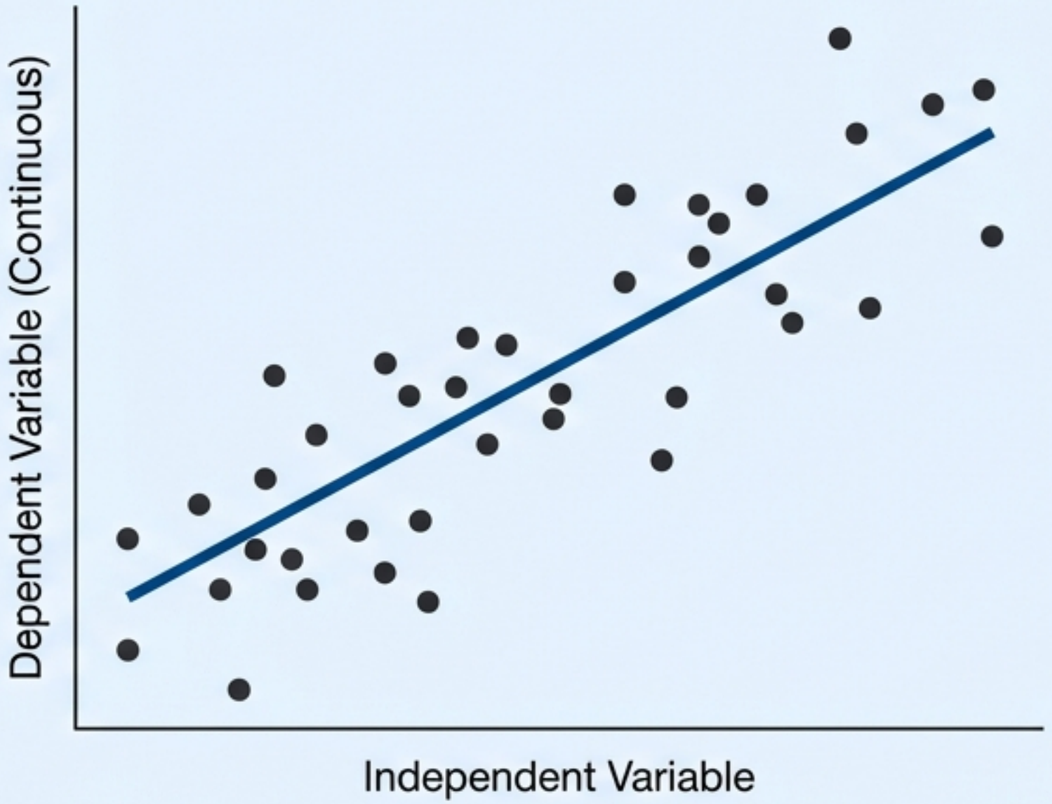


"We acknowledge the depth of the math, but today we master the machine."

Predicting *How Much* vs. Predicting *Which One*

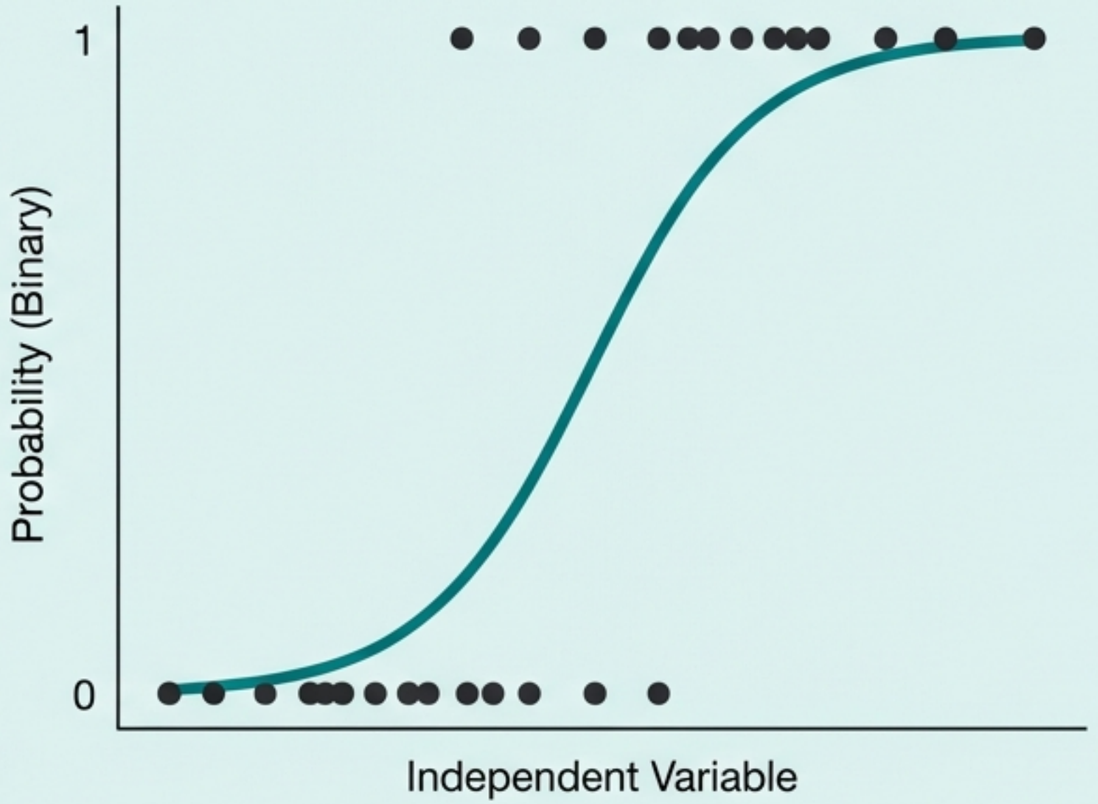
Linear Regression

Target: Continuous (Asset Returns, Price)



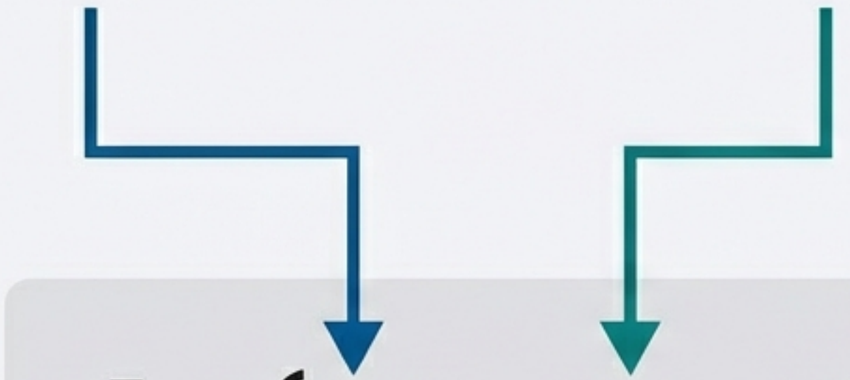
Logistic Regression

Target: Binary (Default/No Default)



The Linear Foundation

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \epsilon_i$$



```
lm(y ~ x, data = df)
```

Key Metrics

- **RSS** (Residual Sum of Squares): Minimize this
- **R^2**: Goodness of Fit
- **MSE**: Mean Squared Error

Coding OLS

```
model <- asset_data |>
  lm(formula = asset_return ~ sp500_return)


# View the results

model |> summary()
```

The Pipe: Data flows into the model.

The Four Pillars of OLS

Linearity

Straight line relationship

Normality

Errors are normally distributed

Homogeneity

Constant variance (Homoscedasticity)

Independence

No pattern in error terms

Alert Coral

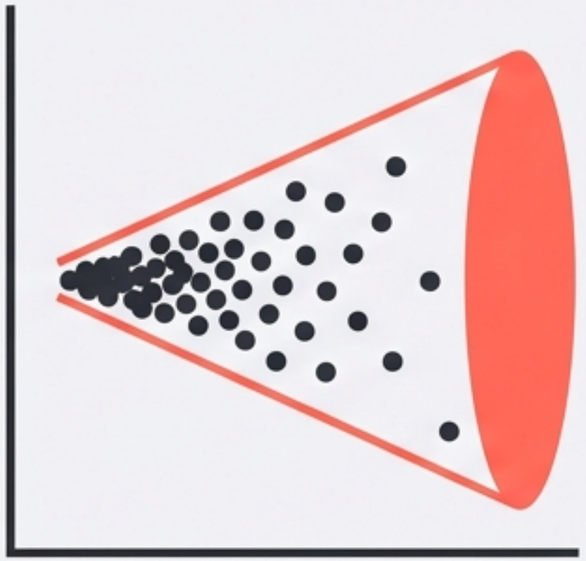
Why Bother? (When Models Fail)

Non-linearity

The model doesn't fit the reality.

Heteroscedasticity

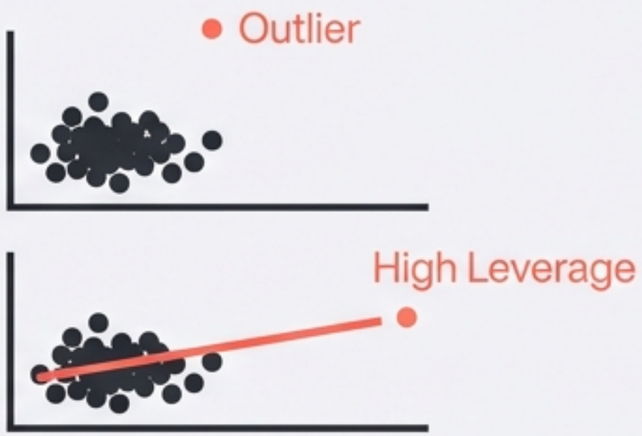
The Variance Explosion.



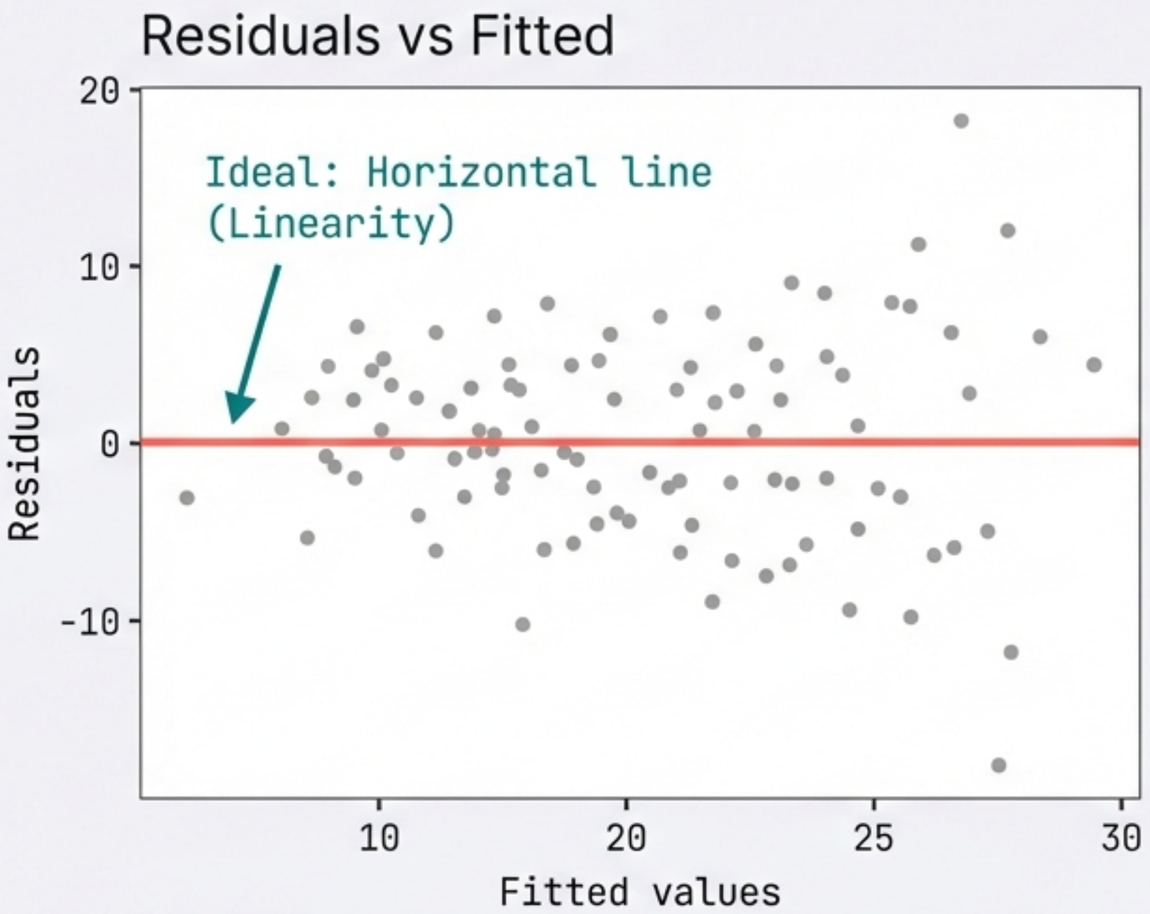
Influential Monsters

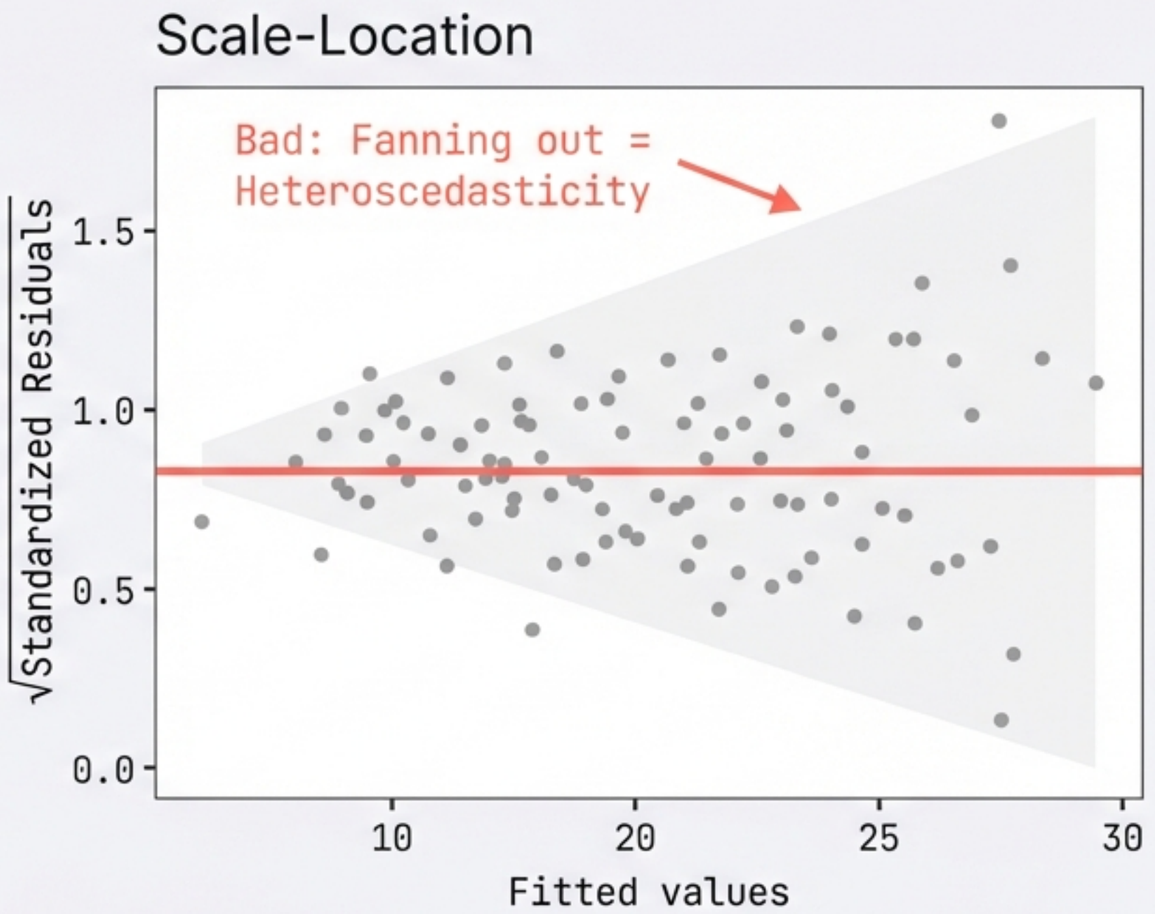
Values that distort the truth.

- **Outliers:** Extreme Y values
- **High Leverage:** Extreme X values

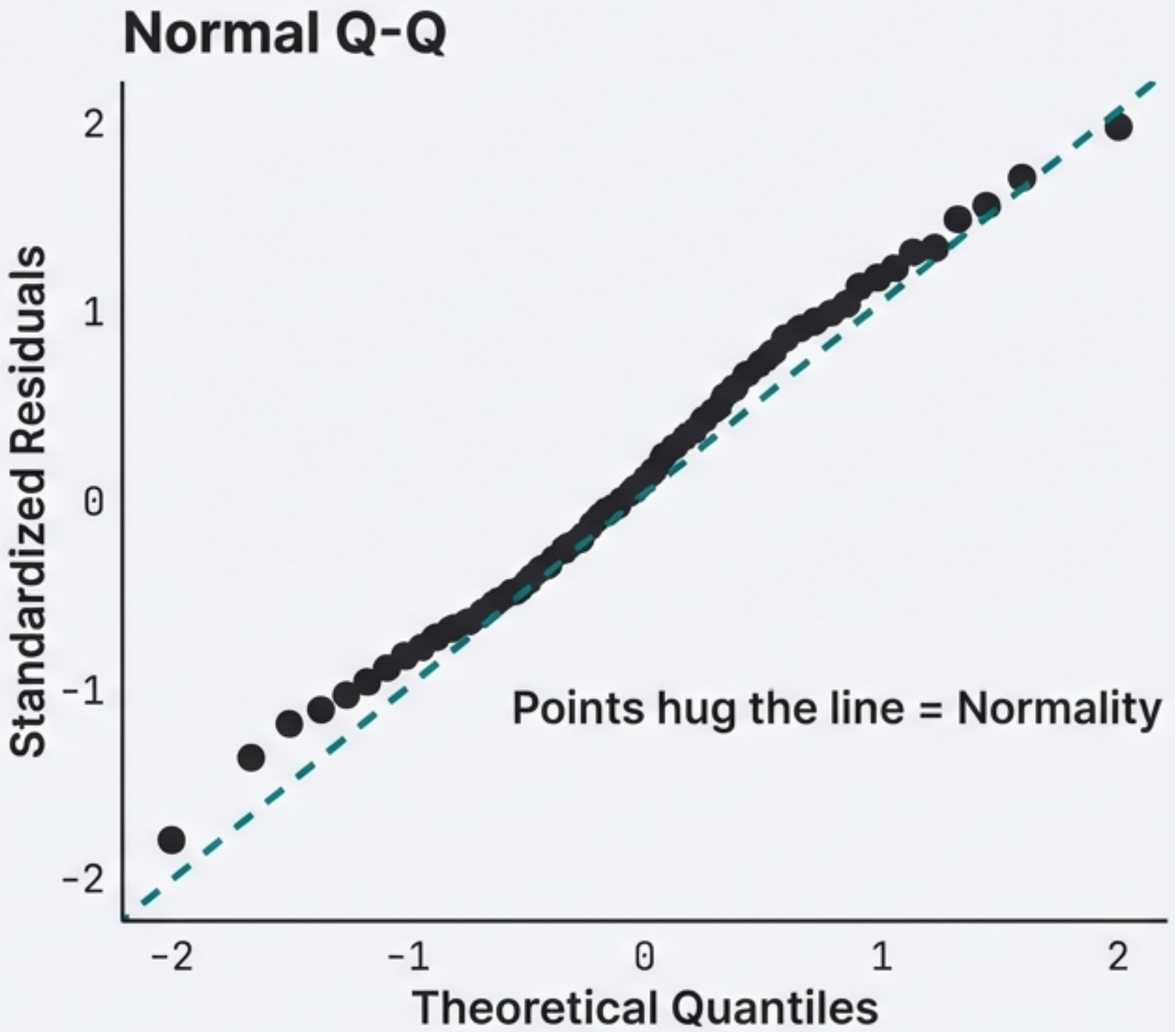


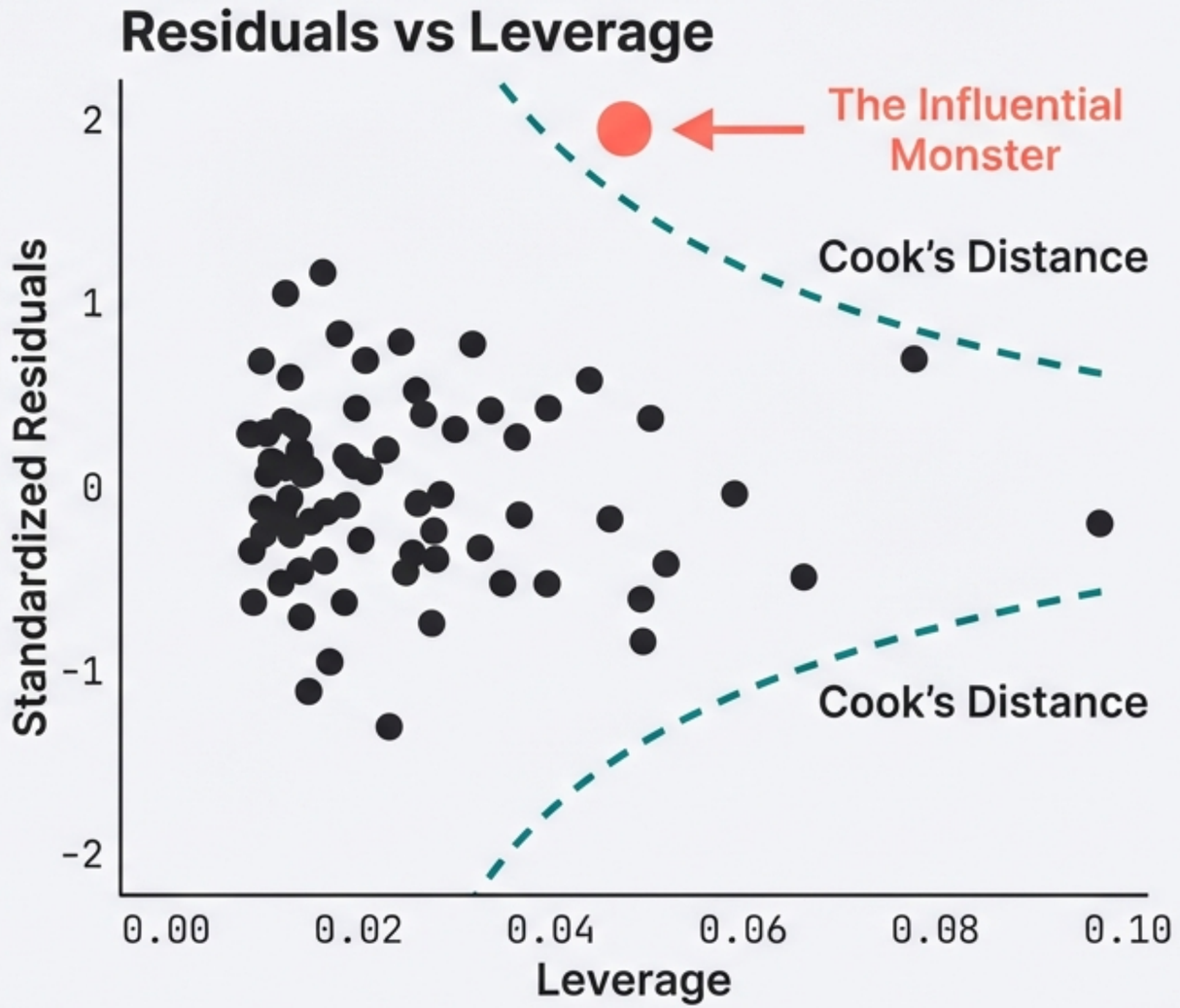
Diagnostics: Linearity & Homogeneity





Diagnostics: Normality & Influence

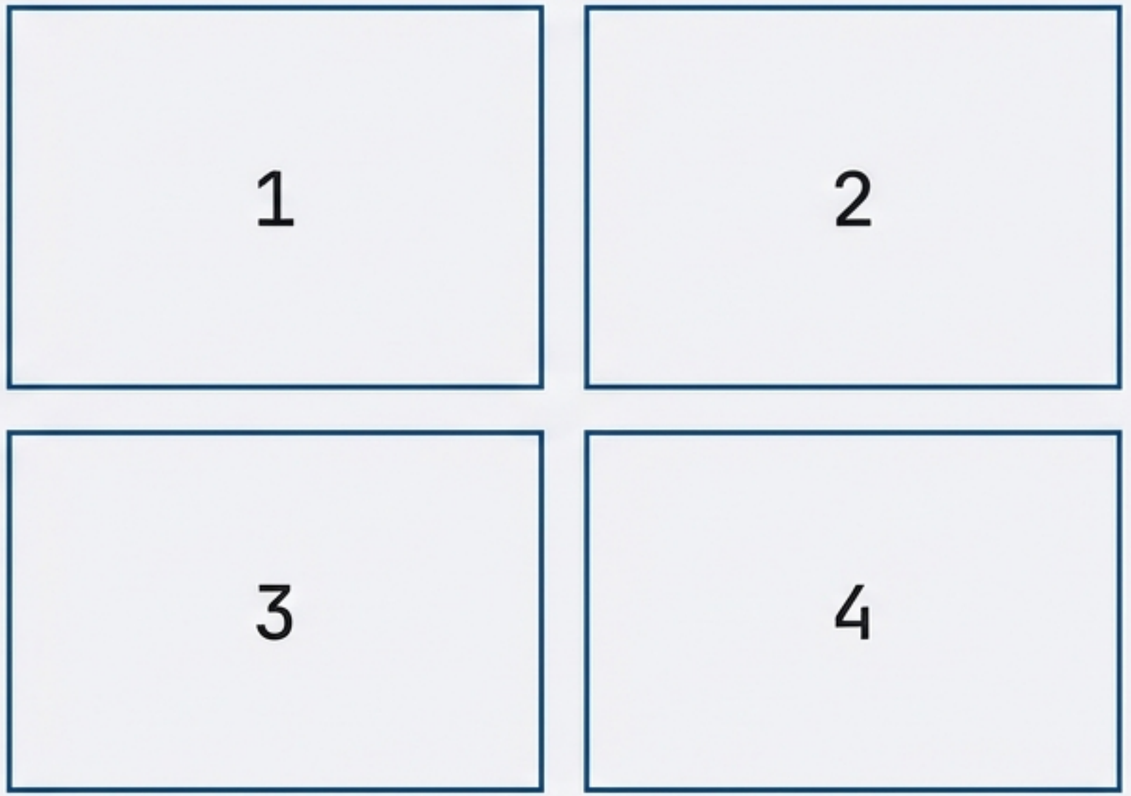




Rapid Diagnostics in R

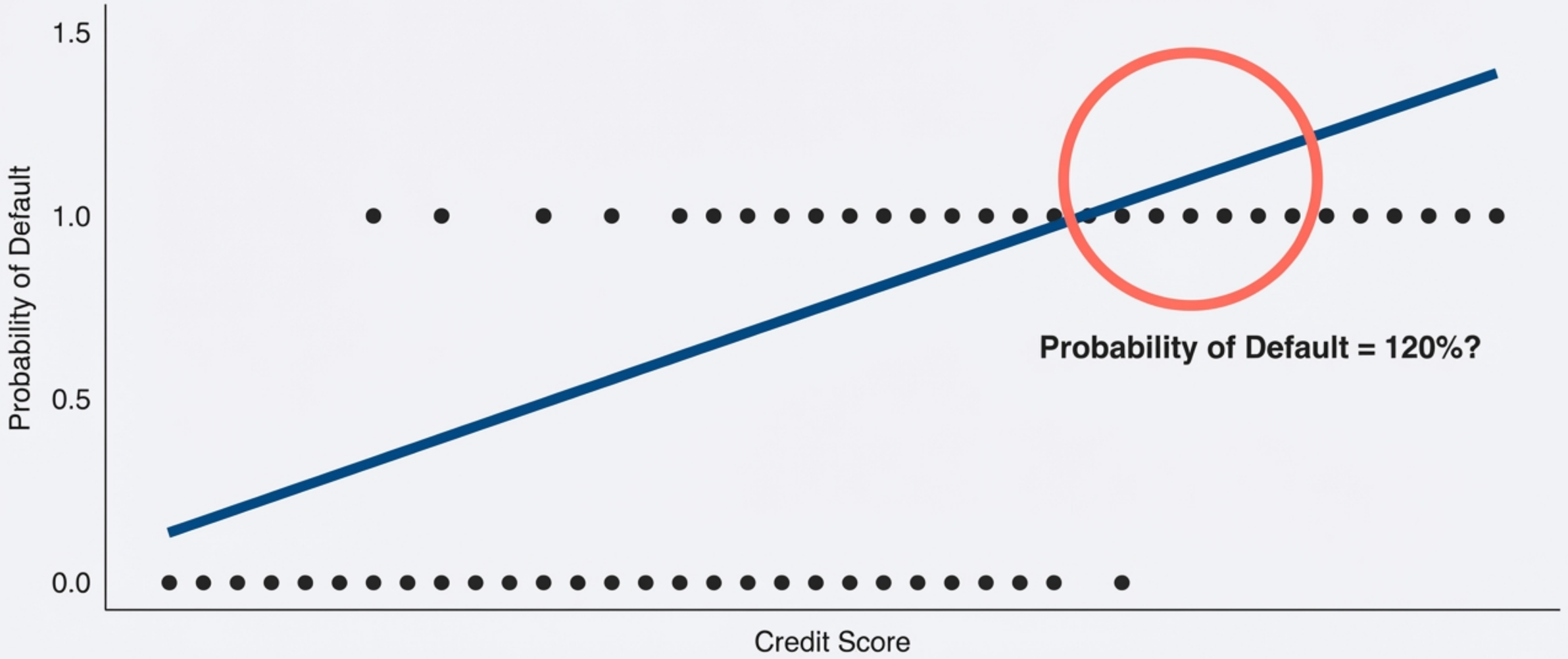
```
# Standard 2x2 diagnostic view
par(mfrow = c(2, 2))
plot(model)
```

Plotting the model object immediately reveals the four diagnostic charts.



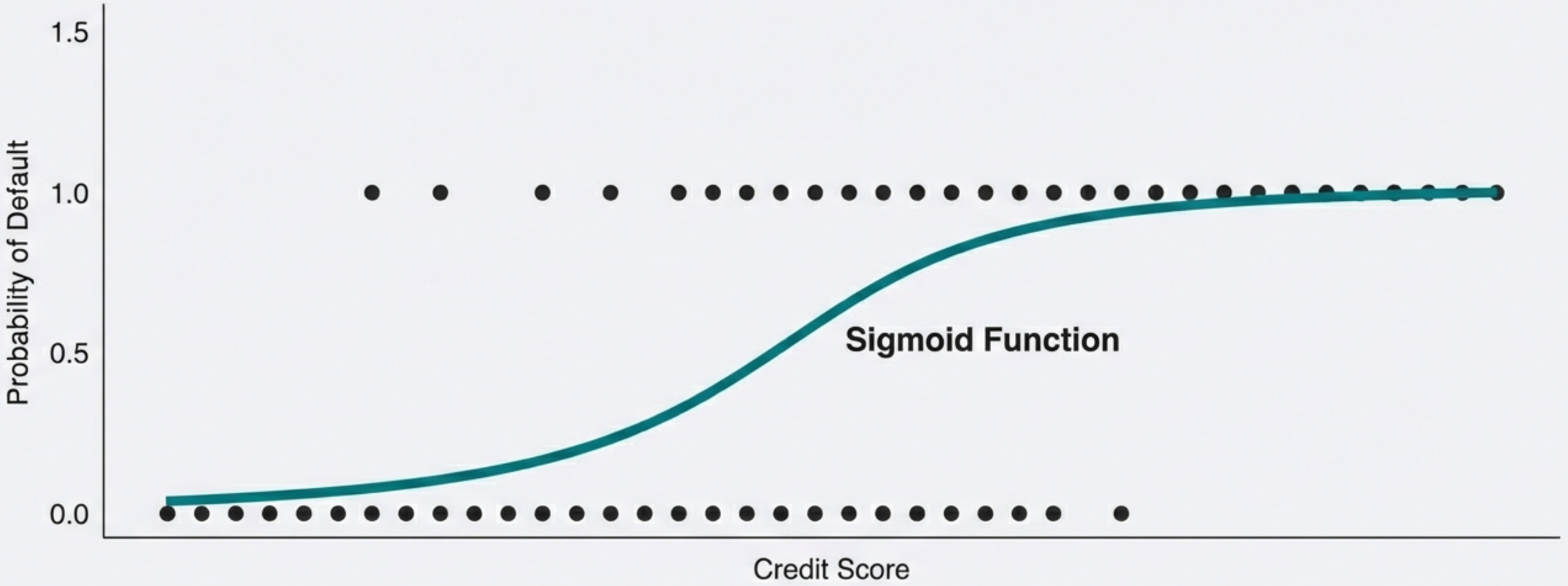
Swiss Financial Modern

The Problem with Linear Lines



Swiss Financial Modern

Enter Logistic Regression



Probability (P) $\longrightarrow$ **Odds** $\left(\frac{P}{1-P}\right)$ $\longrightarrow$ **Log Odds** $(\ln(\text{Odds}))$

Speaking in Odds

Probability	Odds	Log Odds
Range: 0 to 1	**Range:** 0 to ∞	**Range:** −∞ to +∞
Description: Intuitive	**Description:** Gambling / Finance	**Description:** Linear Math Scale
	$P/(1-P)$	$\ln(\text{Odds})$

We model in Log Odds, but we interpret in Probability.

Swiss Financial Modern

Coding Logit

```
logit_model <- credit_data |>
  glm(formula = default ~ credit_score + income,
      family = binomial (link = "logit"))

logit_model |> summary()
```

The Switch: Defines the error distribution & link function.

The Analyst's Summary

Linear (OLS)

- Target: Continuous
- Assumption: Homoscedasticity
- Metric: Minimize RSS

Logistic (Logit)

- Target: Binary
- Transformation: Log-odds
- Metric: Maximum Likelihood

ALWAYS PLOT YOUR RESIDUALS