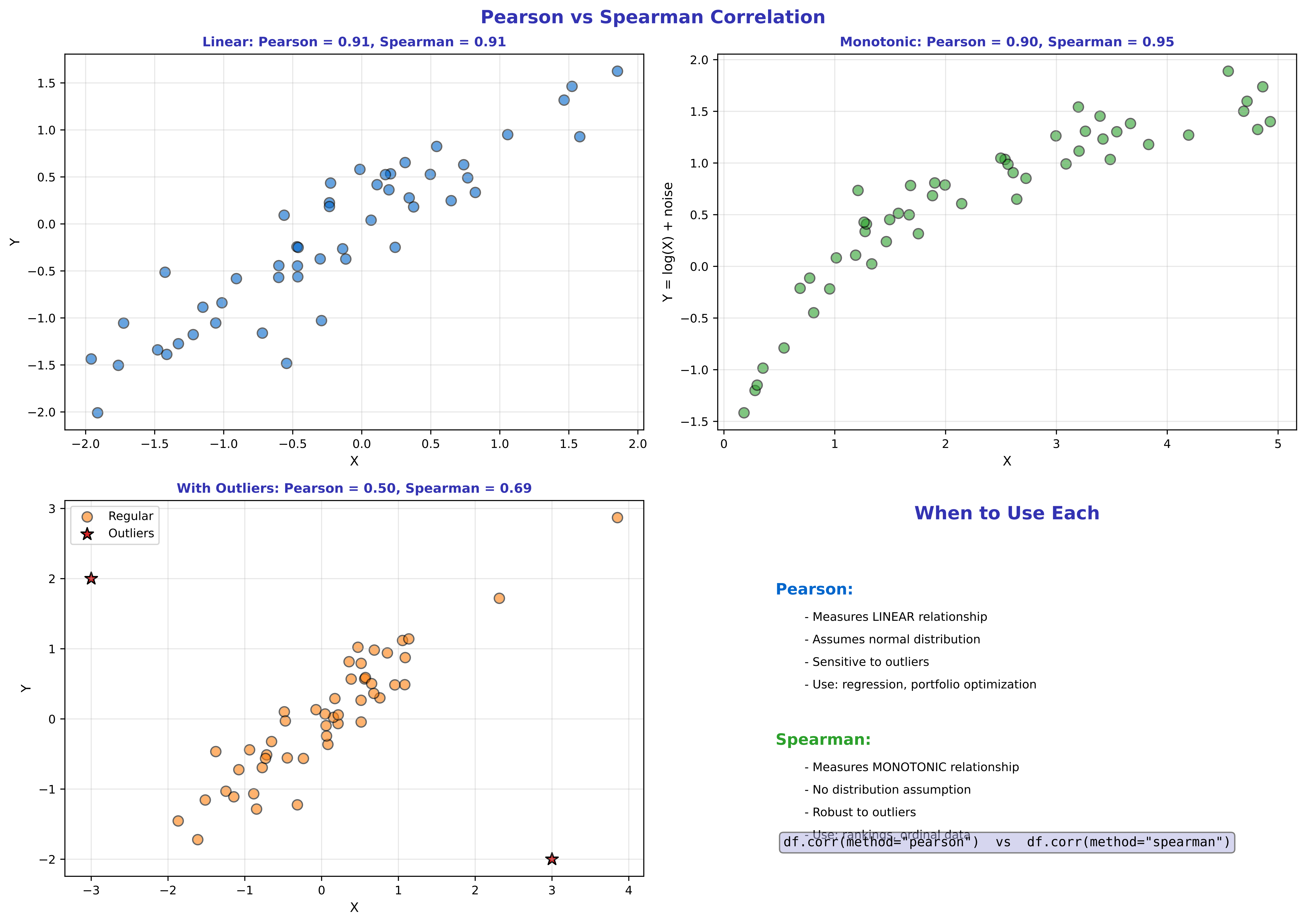
Pearson vs Spearman Correlation
Linear: Pearson = 0.91, Spearman = 0.91
Y
X
Monotonic: Pearson = 0.90, Spearman = 0.95
Y = log(X) + noise
X
With Outliers: Pearson = 0.50, Spearman = 0.69
Regular
Outliers
Y
X
When to Use Each
Pearson:
- Measures LINEAR relationship
- Assumes normal distribution
- Sensitive to outliers
- Use: regression, portfolio optimization
Spearman:
- Measures MONOTONIC relationship
- No distribution assumption
- Robust to outliers
- Use: rankings, ordinal data
df.corr(method="pearson") vs df.corr(method="spearman")