

Para ilustrar o efeito de um *outlier* na estimação dos parâmetros da reta de regressão, serão gerados dois bancos de dados a partir do modelo abaixo, onde a única diferença entre os dados será a observação discrepante.

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

onde $\beta_0 = 3$, $\beta_1 = 10$ e $\epsilon \sim N(0,1)$.

Em seguida serão ajustados dois modelos um de regressão linear simples, utilizando o método de mínimos quadrados ordinários e o outro, de regressão lineares robustas, utilizando os métodos de M-Estimadores e MMEstimadores .

Ajustes para o modelo linear simples via minimos quadrados ordinários

	Valor Verdadeiro	Estimativa sem Outlier	EP sem outlier	Estimativa com Outlier	EP com outlier
(Intercept)	3	3.521714	0.4750086	-9.608952	12.667931
x	10	9.967536	0.0522440	12.429536	1.393287

Tabela 1: Estimativas dos parâmetros utilizando mínimos quadrados

Observando a Tabalea 1, pode-se observar que os parâmetros estimados pelo método dos mínimos quadrados encontram-se próximos dos valores reais, quando não existe o outlier no banco de dados, já quando o outlier é adicionado nota-se que as estimativas obtidas não estão próximas dos valores reais dos parâmetros, mostrando que o método de mínimos quadrados ordinários tende a ter suas estimativas afetadas na presença de outliers, devido a sua sensibilidade a esse tipo de observação. Além disso, é possível observar também um aumento significativo no erro padrão, quando analisado o modelo como a presença do outlier.

O gráfico abaixo mostra as retas estimadas com e sem a presença do *outlier*

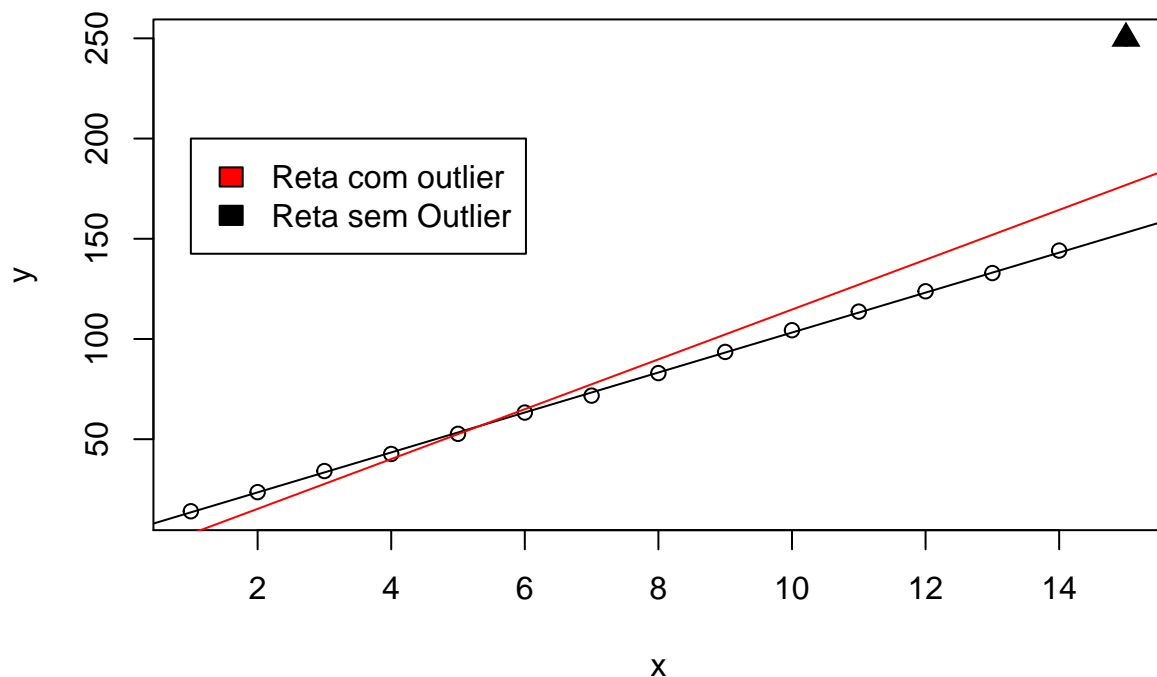


Figura1: Retas ajustadas via mínimos quadrados ordinários

A reta em vermelho apresentada no gráfico (sob influência do outlier) mostra a alteração no coeficiente angular e no intercepto. Além disso, é possível notar um mal ajuste aos dados, visto que a reta não cobre grande parte dos pontos. Enquanto que a reta ajustada para os dados sem o outlier mostra um bom ajuste, cobrindo grande parte dos dados.

Modelo de Regressão Robusta via M-estimadores

Visto a sensibilidade do método de mínimos quadrados ordinários à presença de um outlier, será ajustado agora um modelo de regressão robusta utilizando a classe de M-Estimadores, presentes no pacote MASS do software R. Seguindo o que foi feito para o modelo de Regressão Linear via mínimos quadrados ordinários, serão ajustados modelos para o conjunto de dados com e sem a presença do outlier. A tabela abaixo apresenta as estimativas dos parâmetros, obtidas via M-Estimados.

A tabela abaixo apresneta as etimativas dos parâmetros, obtidas via M-Estimados

	Valor Verdadeiro	Estimativa sem outlier	EP sem Outlier	Estimativa com outlier	EP com Outlier
Intercepto	3	3.497383	0.4806592	3.056253	0.5053115
x	10	9.978729	0.0528655	10.060600	0.0555769

Tabela 2:Estimativas dos pârâmetros utilizando M-Estimadores via mínimos qudrados reponderados

Analisando os coeficientes estimados,pode-se notar que para ambos os conjuntos de dados simulados, com ou sem a presença do outlier, as estimativas estão bem próximas dos valores reais dos parâmetros indicando que

os M-Estimadores possuem menor sensibilidade a observações discrepantes quando comparado com o método de mínimos quadrados ordinário.

O gráfico abaixo reforça o fato de que as retas estimadas sofreram pouca alteração quanto à influência do outlier, visto que não é possível distinguir as retas ajustadas, indicando um bom ajuste aos dados simulados.

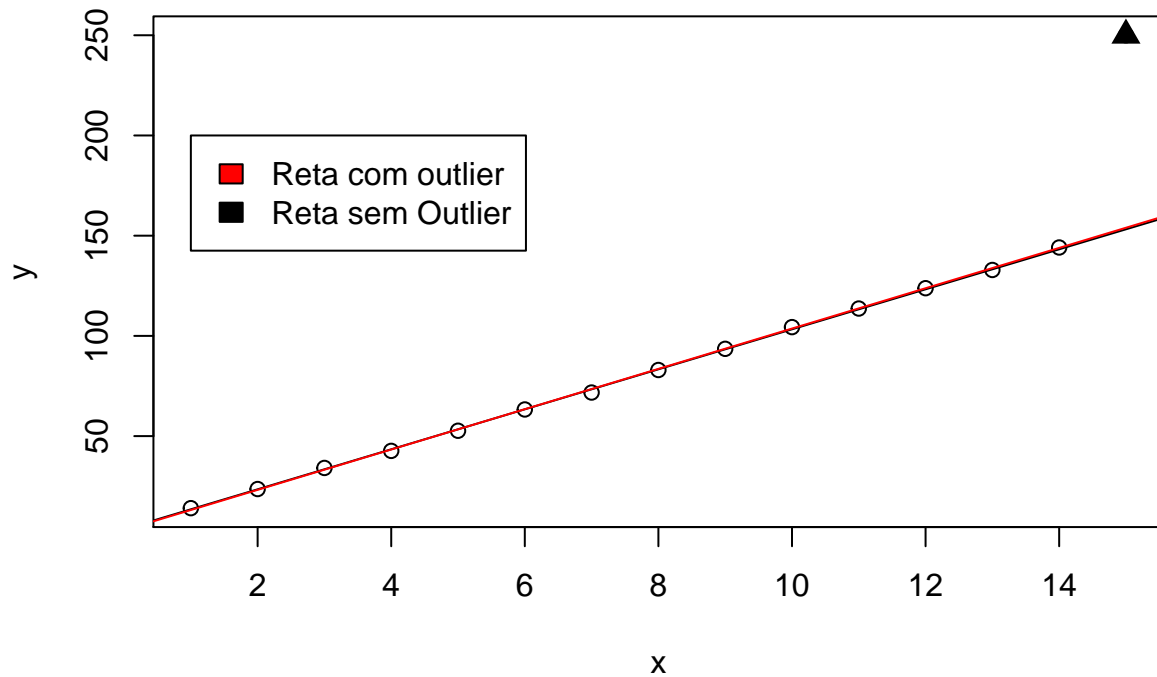


Figura2: Retas ajustadas via M-Estimadores de Huber

Ajustes utilizando MM-Estimadores.

A tabela abaixo apresenta as estimativas dos parâmetros de regressão obtidos via MM-Estimadores:

	Valor Verdadeiro	Estimativa sem Outlier	Estimativa com Outlier
Intercepto	3	3.519254	3.28754
x	10	9.972556	10.01672

Tabela 3: Estimativas obtidas via MM-Estimadores

Analisando a Tabela 3, pode-se observar que as estimativas para ambos os conjuntos de dados apresentam valores próximos dos valores reais dos parâmetros, assim como os M-Estimadores de Huber, indicando um bom ajuste. Além disso, assim como no caso das retas obtidas via M-Estimadores, não é possível distinguir as duas retas estimadas dando ainda mais indícios da boa qualidade do ajuste.

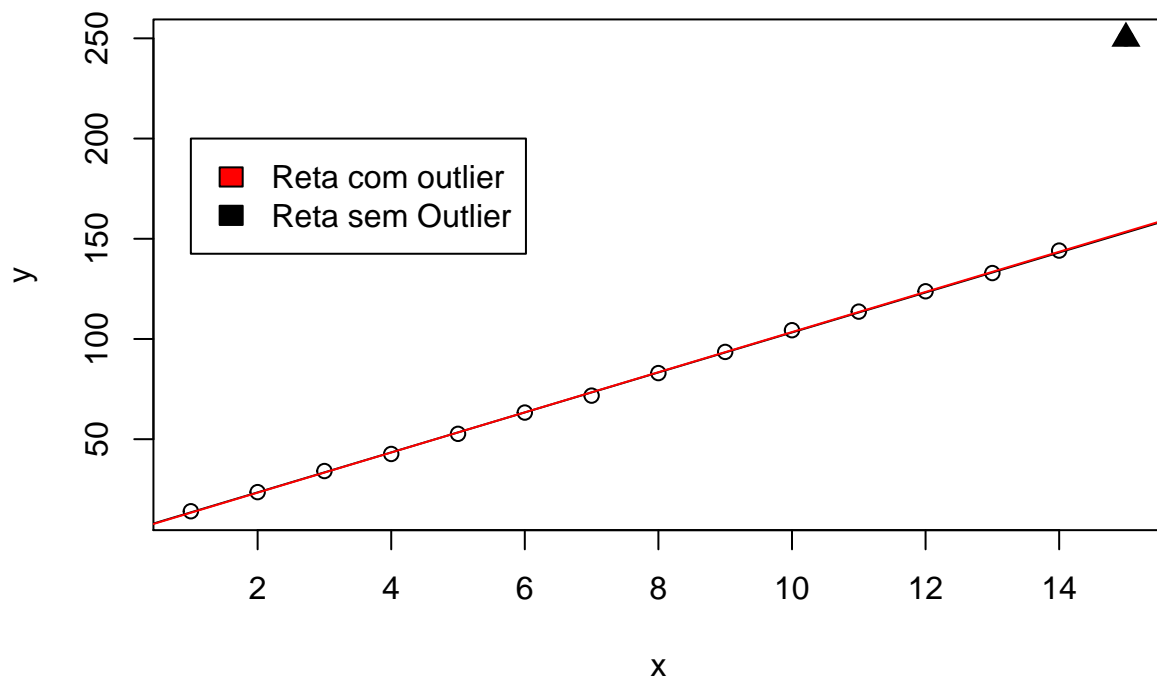


Figura3: Retas ajustadas via MM-Estimadores

Conclusão

As estimativas obtidas para o conjunto de dados simulados sem a presença do outlier foram próximas dos valores reais dos parâmetros para os três métodos de estimação. Já quando inserido o outlier no conjunto de dados simulados é possível notar que os valores estimados dos parâmetros via mínimos quadrados ordinários, sofreu uma alteração significativa e a reta passou a não ser adequada aos dados. Já quando utilizado os métodos de M-Estimadores de Huber e MM-Estimadores foram obtidos valores estimados ainda próximos dos valores reais, indicando que, para este problema, M-Estimadores e MM-Estimadores são boas alternativas para contornar o problema da presença de um outlier