Should We Use Survey Weights When Fitting Models?

Brady T. West

Review: Survey Weights

Recall

Survey weights may be available in data sets collected from **complex probability samples**, accounting for (at least) unequal probabilities of selection into the sample for different cases

Review: Survey Weights

Recall

Survey weights may be available in data sets collected from **complex probability samples**, accounting for (at least) unequal probabilities of selection into the sample for different cases

In theory

These weights are designed to enable unbiased estimation of selected parameters for a **finite target population**

Review: Survey Weights

Recall

Survey weights may be available in data sets collected from **complex probability samples**, accounting for (at least) unequal probabilities of selection into the sample for different cases

In theory

These weights are designed to enable unbiased estimation of selected parameters for a **finite target population**

Therefore

We consider the case where an analyst wants to fit a regression model to a dependent variable (DV) collected from a probability sample

Should We Use Weights in Estimation?

Using survey weights to fit regression models:

- Ensures that the estimated regression parameters will be unbiased with respect to the sample design…
- …but does not protect analysts from poor model specification!

Should We Use Weights in Estimation?

Using survey weights to fit regression models:

- Ensures that the estimated regression parameters will be unbiased with respect to the sample design…
- …but does not protect analysts from poor model specification!

An analyst fits a poorly-specified model,
but still uses the weights in estimation.
What is the result?

Should We Use Weights in Estimation?

Using survey weights to fit regression models:

- Ensures that the estimated regression parameters will be unbiased with respect to the sample design…
- …but does not protect analysts from poor model specification!

Good!

Unbiased estimates of the regression parameters

An analyst fits a poorly-specified model, but still uses the weights in estimation.
What is the result?

Should We Use Weights in Estimation?

Using survey weights to fit regression models:

- Ensures that the estimated regression parameters will be unbiased with respect to the sample design…
- …but does not protect analysts from poor model specification!

An analyst fits a poorly-specified model, but still uses the weights in estimation.
What is the result?

Good!

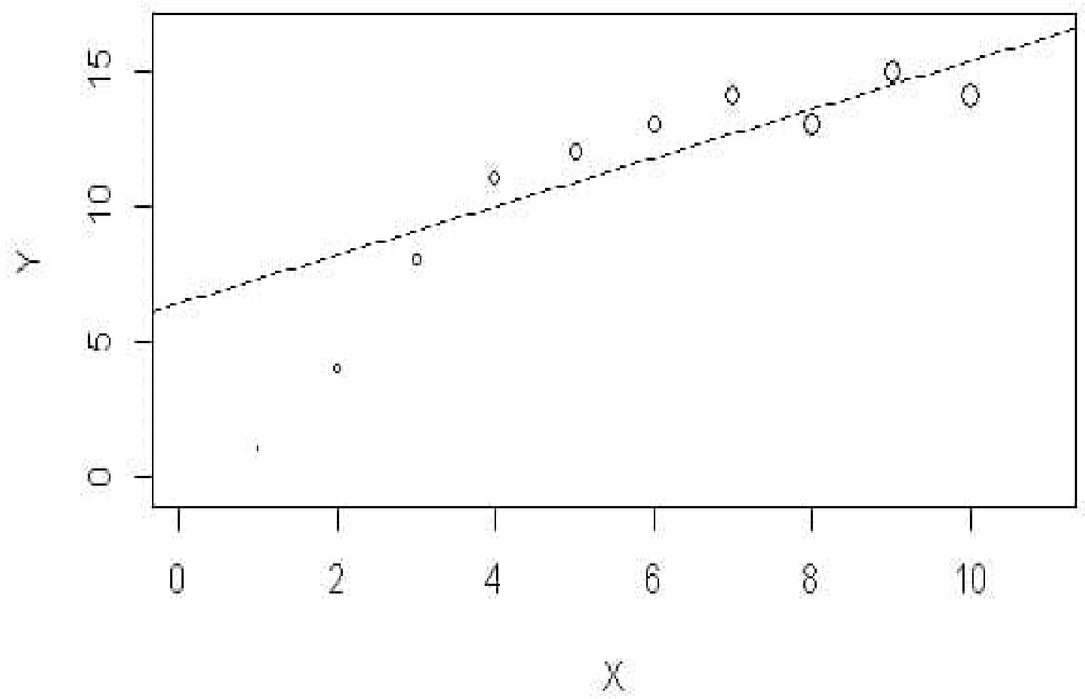
Unbiased estimates of the regression parameters

Incorrect model for the finite population of inference

BAD!

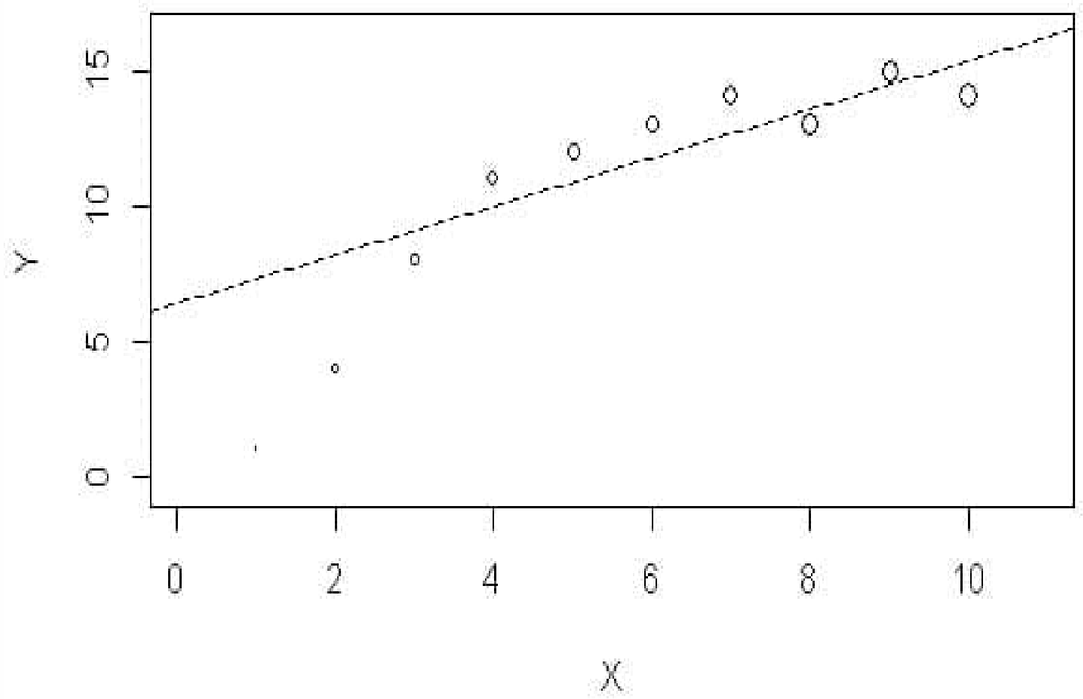
Example: Weights, Poor Model

- In this example data set, the size of the point represents the survey weight for the case



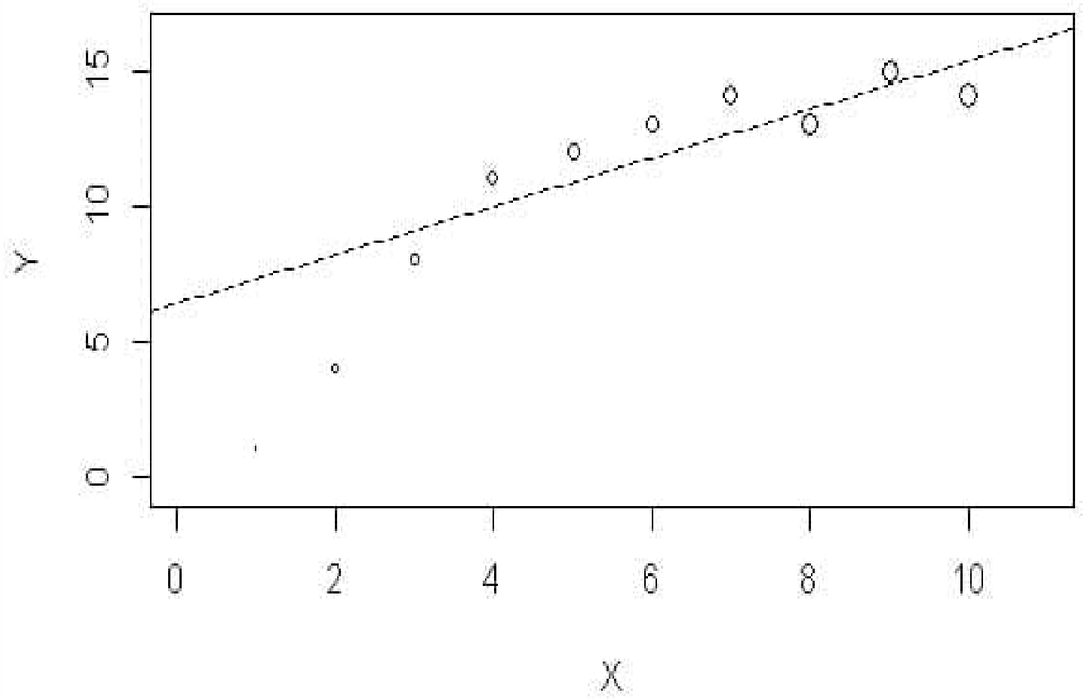
Example: Weights, Poor Model

- In this example data set, the size of the point represents the survey weight for the case
- Non-linear relationship between X and Y, but the analyst misspecifies the model and assumes a linear relationship



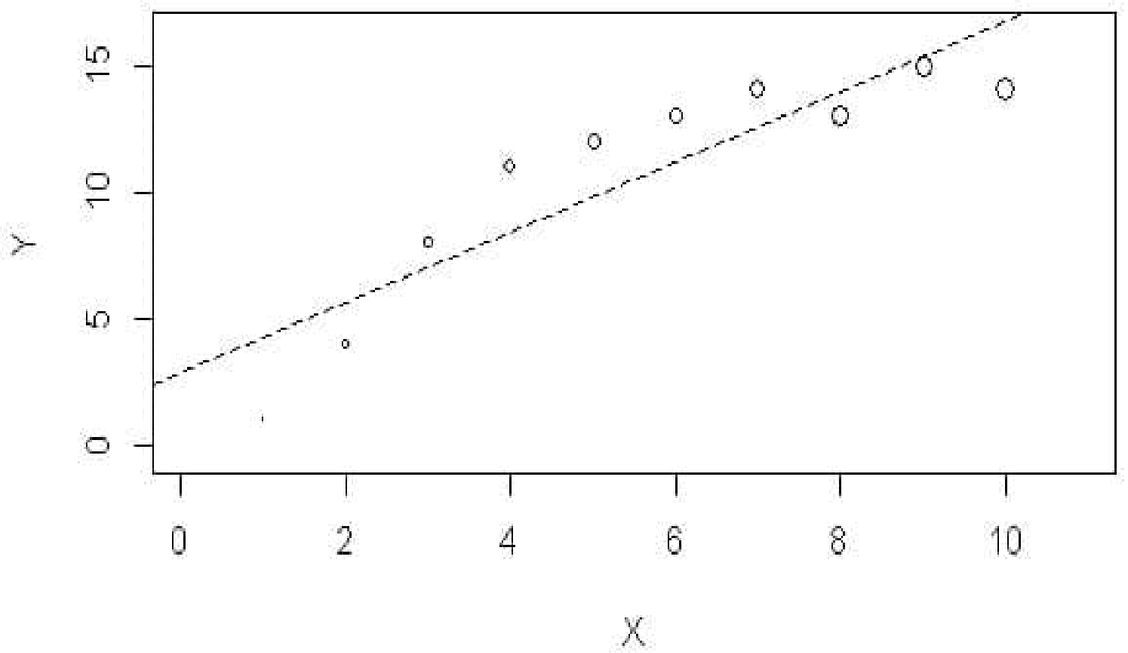
Example: Weights, Poor Model

- In this example data set, the size of the point represents the survey weight for the case
- Non-linear relationship between X and Y, but the analyst misspecifies the model and assumes a linear relationship
- The analyst still uses the weights though! We see an **unbiased** estimate of a **poor** population model.



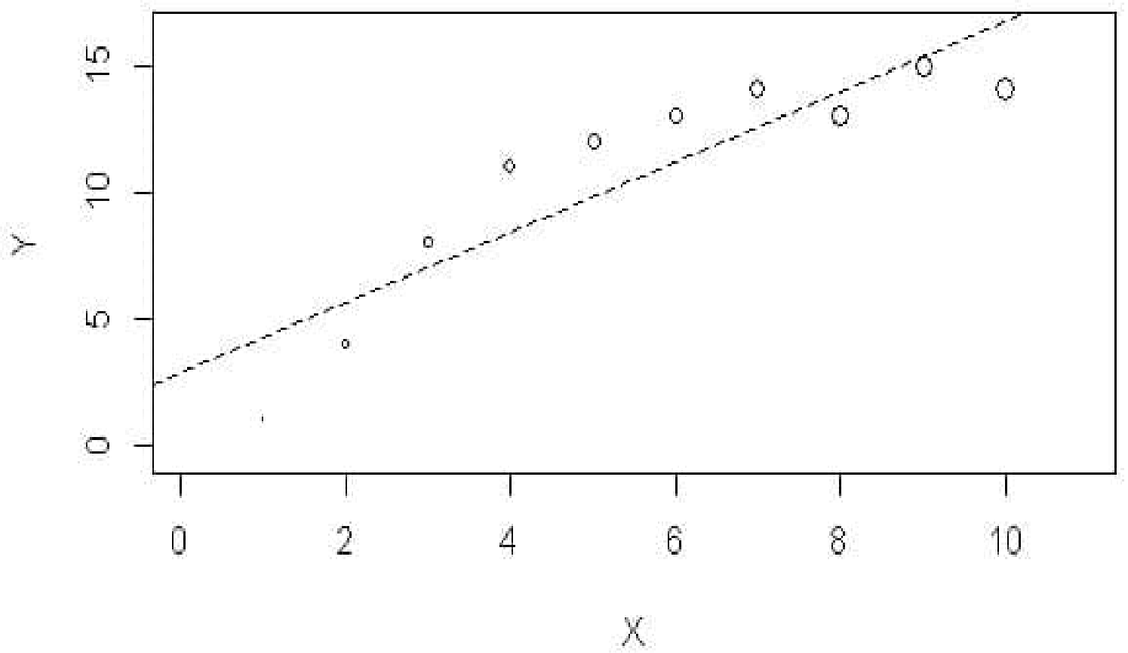
Example: No Weights, Poor Model

- Now, suppose that the analyst ignores the weights AND misspecifies the model…



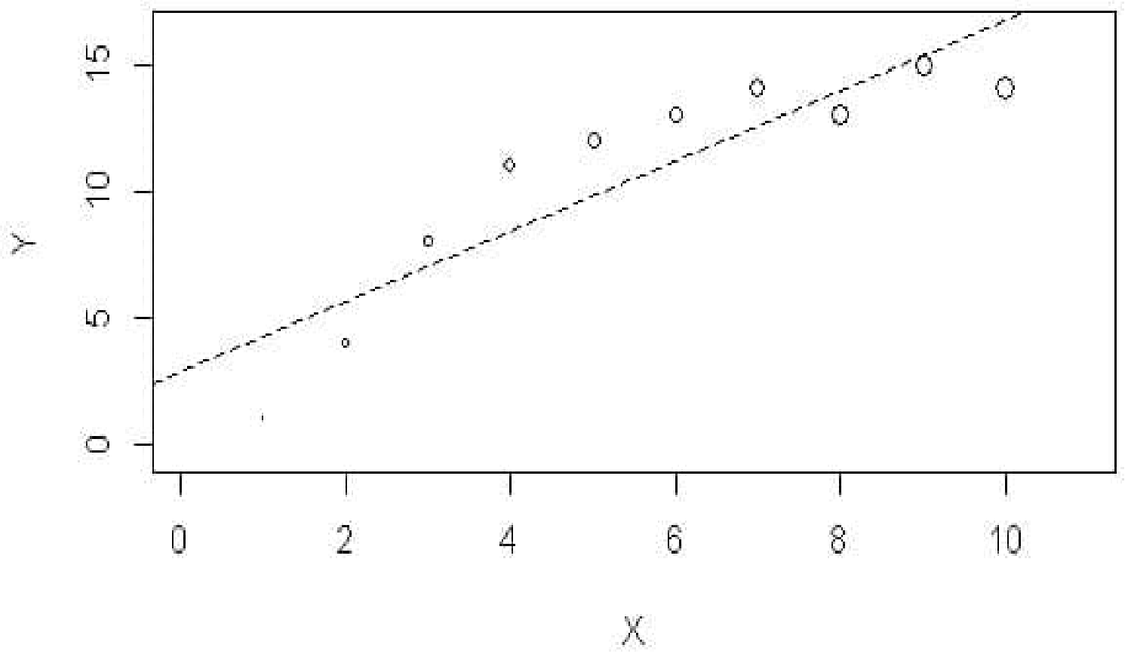
Example: No Weights, Poor Model

- Now, suppose that the analyst ignores the weights AND misspecifies the model…
- We see a **biased** estimate of the relationship based on a **poorly specified model!**



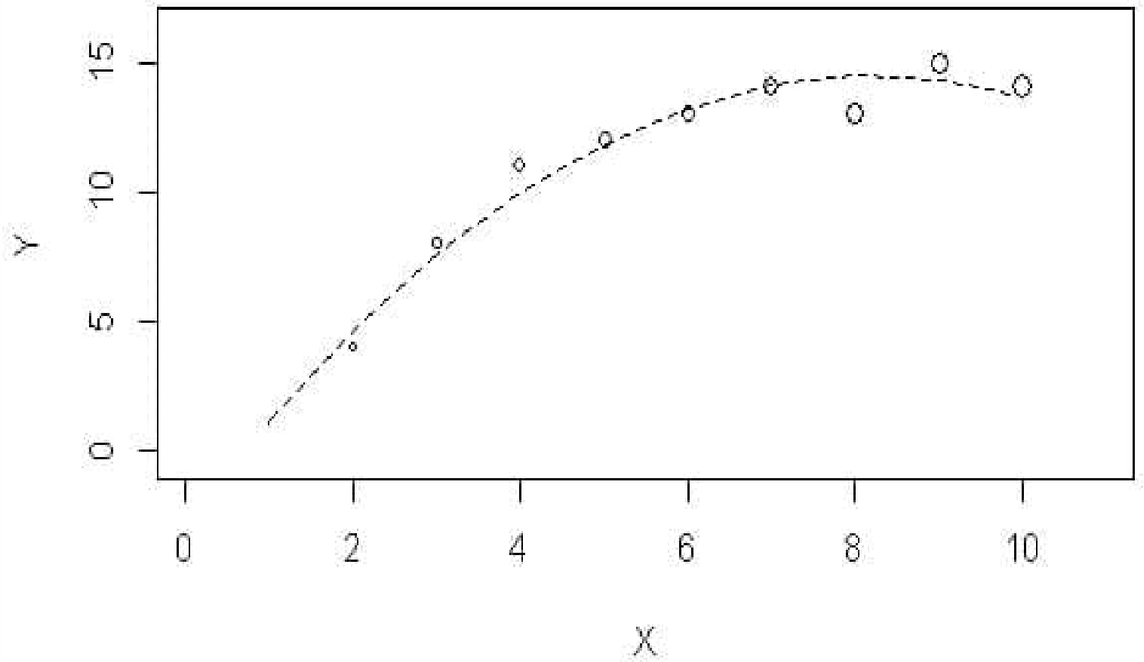
Example: No Weights, Poor Model

- Now, suppose that the analyst ignores the weights AND misspecifies the model…
- We see a **biased** estimate of the relationship based on a **poorly specified model!**
- Note: The fitted regression line is drawn toward the low-weight points (we want to take steps to avoid this situation in practice)



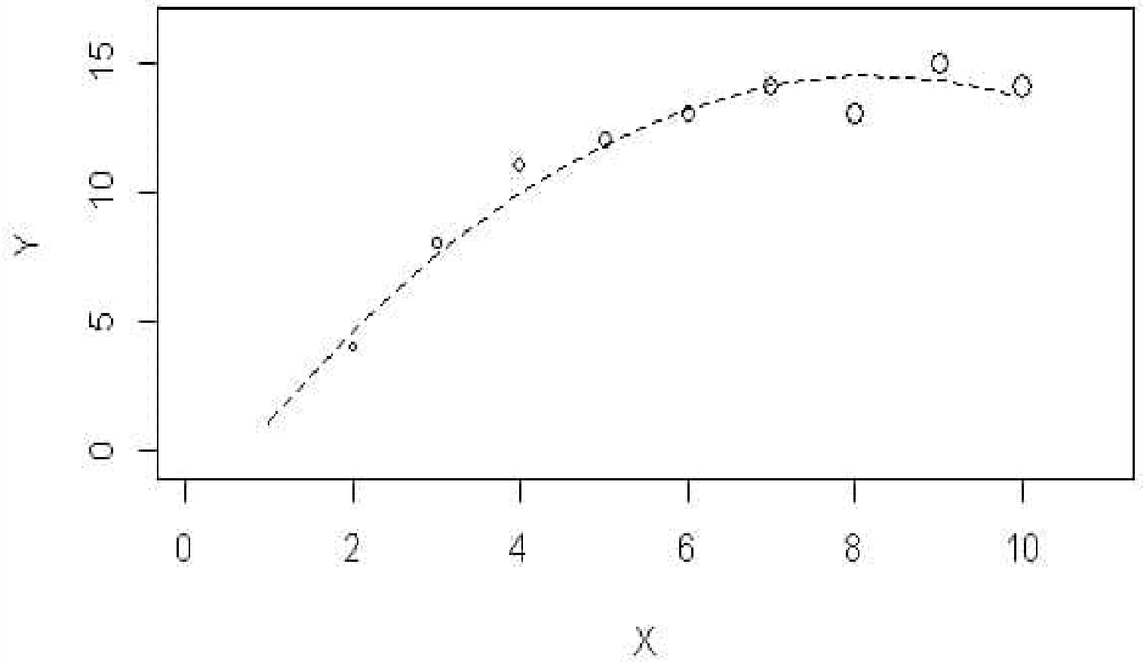
Example: No Weights, Good Model

- Suppose the analyst ignores weights BUT correctly specifies the model



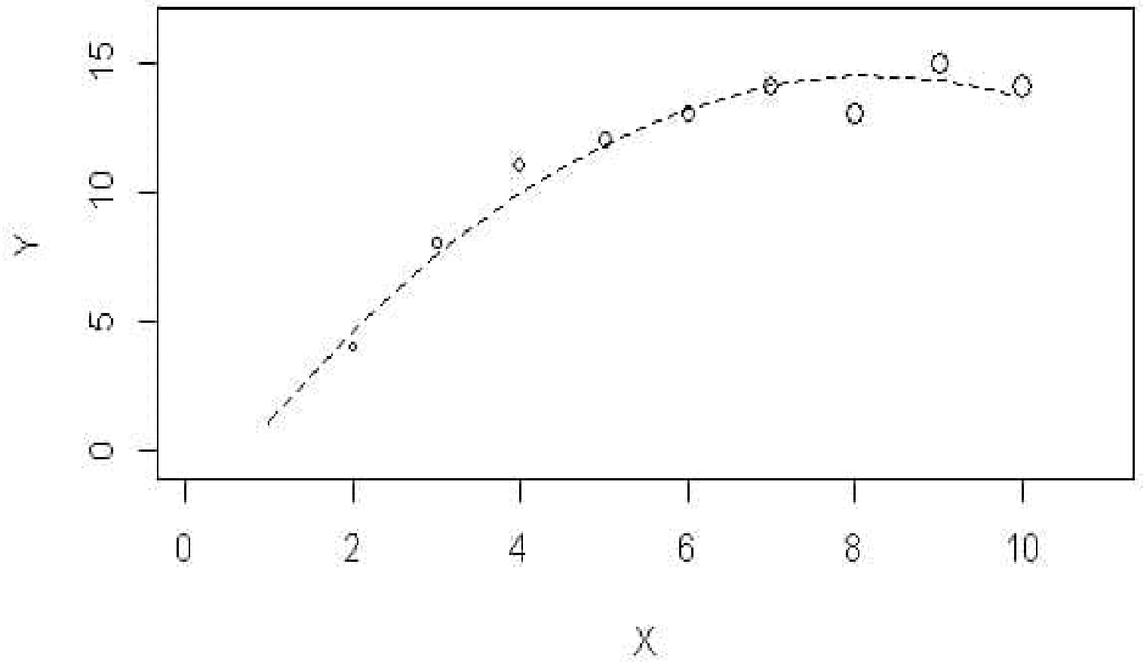
Example: No Weights, Good Model

- Suppose the analyst ignores weights BUT correctly specifies the model
- We see that the **well-specified model** provides a good fit to the observed data!



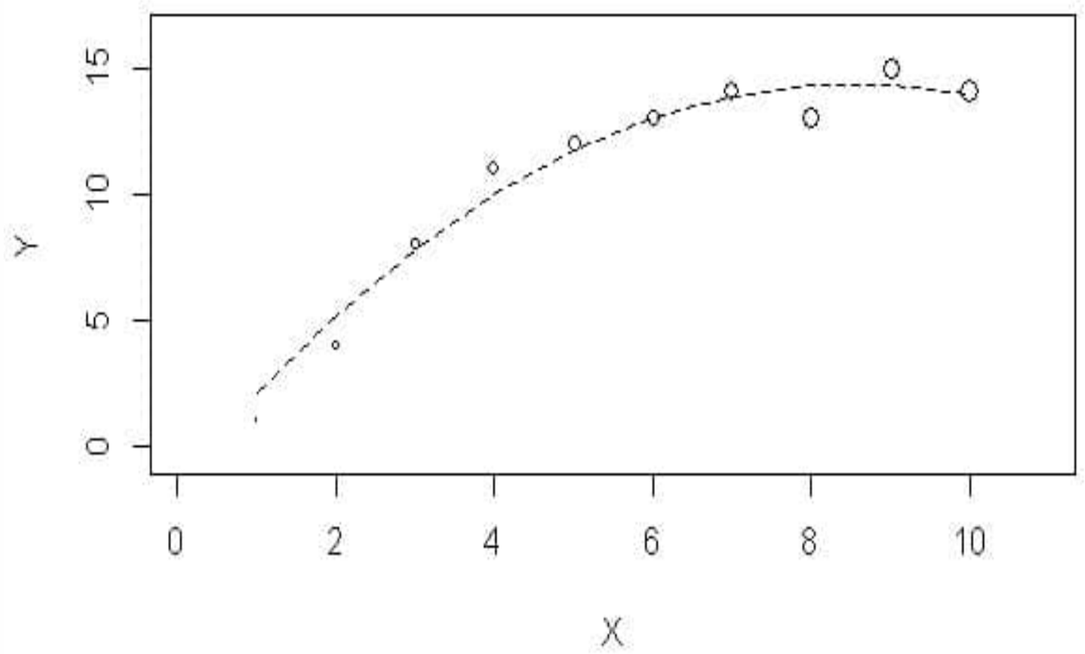
Example: No Weights, Good Model

- Suppose the analyst ignores weights BUT correctly specifies the model
- We see that the **well-specified model** provides a good fit to the observed data!
- In this **model-based approach**, if the model is correctly specified, we may not need the weights to estimate the relationship in the population!



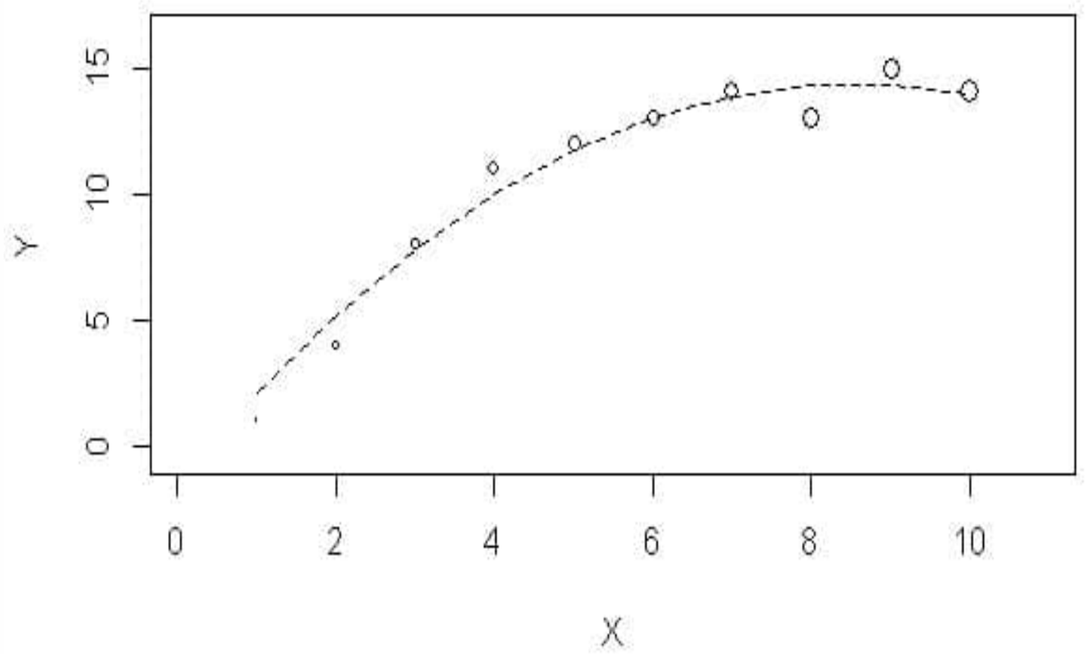
Example: Weights, Good Model

- Suppose the analyst uses the weights to estimate the model AND correctly specifies the model



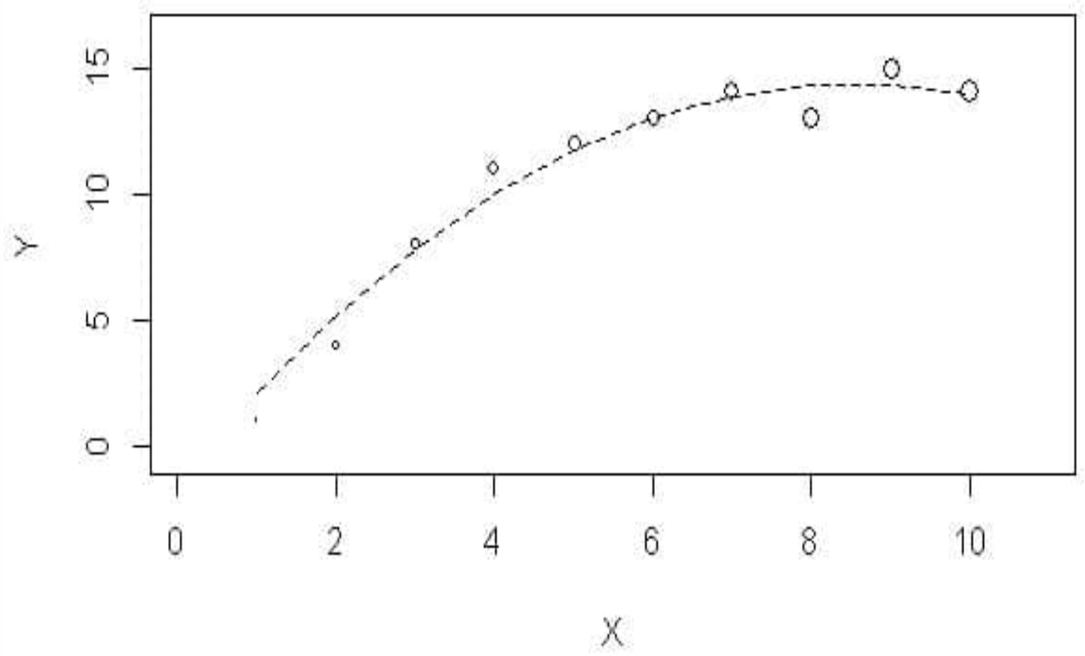
Example: Weights, Good Model

- Suppose the analyst uses the weights to estimate the model AND correctly specifies the model
- We have an **unbiased** estimate of the relationship from a **well-specified model.** But the fit is the same!



Example: Weights, Good Model

- Suppose the analyst uses the weights to estimate the model AND correctly specifies the model
- We have an **unbiased** estimate of the relationship from a **well-specified model.** But the fit is the same!
- In this situation, we would likely find that the estimated coefficients would have **larger standard errors** than in the unweighted case; unnecessary!



Recommendations for Practice

If survey weights are available for a probability sample, and you wish to fit a regression model:

1. Do the best you can to specify the model correctly!

Recommendations for Practice

If survey weights are available for a probability sample, and you wish to fit a regression model:

1. Do the best you can to specify the model correctly!
2. Fit the model with and without using the survey weights.

Recommendations for Practice

If survey weights are available for a probability sample, and you wish to fit a regression model:

1. Do the best you can to specify the model correctly!
2. Fit the model with and without using the survey weights.
3. If estimated coefficients remain similar, but weighted estimates have larger standard errors, model has likely been specified correctly; weights are unnecessary.

And only inflate the sampling variance of your estimates!

Recommendations for Practice

If survey weights are available for a probability sample, and you wish to fit a regression model:

1. Do the best you can to specify the model correctly!
2. Fit the model with and without using the survey weights.
3. If estimated coefficients remain similar, but weighted estimates have larger standard errors, model has likely been specified correctly; weights are unnecessary.
4. If estimated coefficients change substantially, the model may have been misspecified. Weighted estimates should be reported to ensure they are at least unbiased with respect to sample design.

And only inflate the sampling variance of your estimates!

Recommendations for Practice

- It is **difficult** to specify every model correctly: when do we ever know what the true model is, especially with multiple predictors?

"All models are wrong, but some are useful."
George E.P. Box

Recommendations for Practice

- It is **difficult** to specify every model correctly: when do we ever know what the true model is, especially with multiple predictors?
- Modern statistical software (Python) makes it easy to fit unweighted/weighted models, if weights available

"All models are wrong, but some are useful."
George E.P. Box

Recommendations for Practice

- It is **difficult** to specify every model correctly: when do we ever know what the true model is, especially with multiple predictors?
- Modern statistical software (Python) makes it easy to fit unweighted/weighted models, if weights available
- Formal tests also exist for comparing weighted and unweighted estimates
  (See deep-dive reading this week!)

"All models are wrong, but some are useful."
George E.P. Box