

If R is used, present the required output and the (relevant) syntax.

1. The data `Angell.txt`¹, contains information about moral integration of American Cities. Your role is to study the influence of `mobility` and `heterogeneity` on `moralIntegration`.
 - a) Draw and scatterplots, including the least-squares lines, showing the relationship of the response (`moralIntegration`) to each predictor. Are the least-squares lines reasonable summaries of the relationship between the response and each predictor? Explain.
 - b) Compute the simple linear regression of `moralIntegration` on `heterogeneity` and interpret the coefficient estimates and the coefficient of determination.
 - c) Compute the multiple regression using both predictors and interpret the coefficient estimates, corresponding added-variable plots, and the coefficient of determination.
 - d) Explain what information is contained in the R output about the coefficient for `heterogeneity` (using `summary(lm(...))`) in terms of the hypothesis test (and conclusion). In addition, obtain and interpret a 97% confidence interval for this coefficient.
 - e) Call your data frame `Angell` (or change the code below accordingly). Add and run the following syntax to your R script:

```
set.seed(100)
n = dim(Angell)[1]
Angell$social = with(Angell, heterogeneity+mobility+rnorm(n,0,.1))
mod1 = lm(moralIntegration ~ heterogeneity + mobility + social, data= Angell)
```

Perform a hypothesis test for `heterogeneity` and compare it with the one obtained in part d). Assuming you do not know how the predictor `social` was obtained, why do you think you are obtaining these seemingly contradictory results? Explain.

¹The data `Angell.txt`, description `Angell.pdf`, and questions were based on the supplementary material of “Applied Regression Analysis and Generalized Linear Models” 3rd Ed by Fox.