

# A 2008 Applied Statistics Comprehensive Exam Question

From the University of Toronto PhD Comprehensive Exam in Applied Statistics, 2008.

Written by Radford Neal.

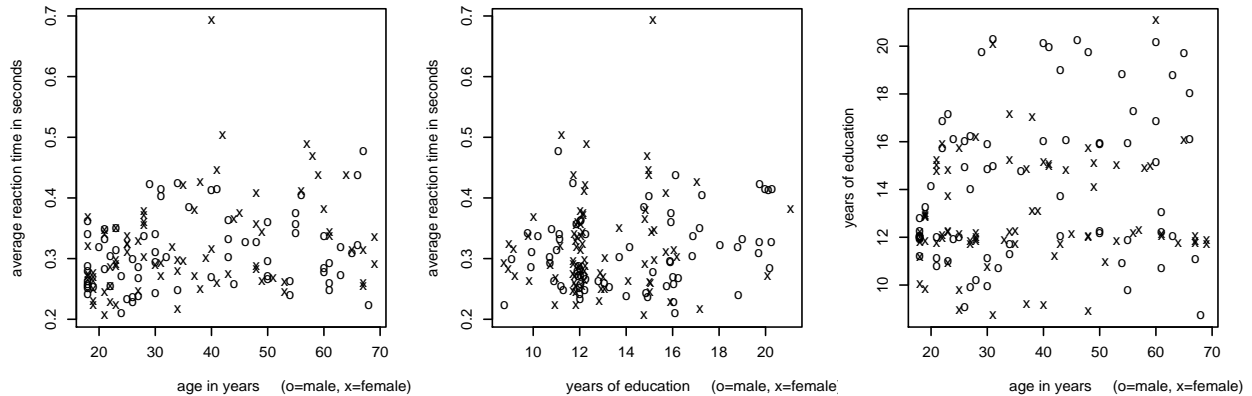
*Note: All the data in this question is fictitious, produced by a simulation program, but you should treat the data as if it were real.*

A researcher (call him Fred) is interested in what factors influence human reaction time to a visual stimulus.

In his first study, Fred recruited 150 subjects from a local shopping mall, all age 18 to 70, asked them their sex, age, and years of education, and had them take a test of reaction time. The following variables were therefore available for each subject:

**sex**            0 = male, 1 = female  
**age**            The subject's age, in years  
**education**    Years of education: eg, 16 = bachelor's degree but no higher  
**time**           Average reaction time in the test, in seconds (lower is better)

Pairwise scatterplots of **age**, **education**, and **time** are shown below, with sex shown by the symbol used, and values for **education** randomly jittered to avoid overlap:



Fred fit a linear regression model for **time** with the other three variables as covariates, using the standard least squares method. The R output from this model fit is shown below:

```
Call:
lm(formula = time ~ sex + age + education)

Residuals:
    Min       1Q   Median       3Q      Max
-0.11421 -0.04709 -0.01406  0.03942  0.36180

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.2243380  0.0306130   7.328 1.47e-11 ***
sex           0.0170760  0.0113831   1.500  0.13574
age           0.0010038  0.0003542   2.834  0.00524 **
education     0.0032729  0.0020417   1.603  0.11109
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.06787 on 146 degrees of freedom  
Multiple R-Squared: 0.08698, Adjusted R-squared: 0.06822  
F-statistic: 4.636 on 3 and 146 DF, p-value: 0.003973

Fred was particularly interested in whether education is related to reaction time, because he thought that more education might indicate greater intelligence, and that greater intelligence and better (ie, lower) reaction times might go together. Fred fit another regression model with only education as a covariate, with the following result:

```
Call:
lm(formula = time ~ education)

Residuals:
    Min       1Q   Median       3Q      Max
-0.11398 -0.05058 -0.01411  0.03568  0.37186

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.265212   0.027756   9.555  <2e-16 ***
education    0.003692   0.002009   1.838  0.0681 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06976 on 148 degrees of freedom
Multiple R-Squared: 0.02231, Adjusted R-squared: 0.0157
F-statistic: 3.377 on 1 and 148 DF, p-value: 0.06814
```

Fred also fit a regression model with only age as a covariate, with the following result:

```
Call:
lm(formula = time ~ age)

Residuals:
    Min       1Q   Median       3Q      Max
-0.12504 -0.04773 -0.01424  0.04140  0.37449

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.2741167   0.0142250  19.270  < 2e-16 ***
age          0.0010964   0.0003498   3.134  0.00208 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06832 on 148 degrees of freedom
Multiple R-Squared: 0.06224, Adjusted R-squared: 0.0559
F-statistic: 9.823 on 1 and 148 DF, p-value: 0.002079
```

Answer the following questions about this study:

- a) Is this an experiment or an observational study? What does your answer to this question say about the ability of this study to answer questions about the causal factors influencing reaction time? Discuss this with reference to the specific variables measured in this study.
- b) Do you see any unusual data points in this data set? What would be an appropriate action to take with regards to any such data points? Does it appear that the results of the regression model fits above will have been strongly influenced by any such unusual point or points?

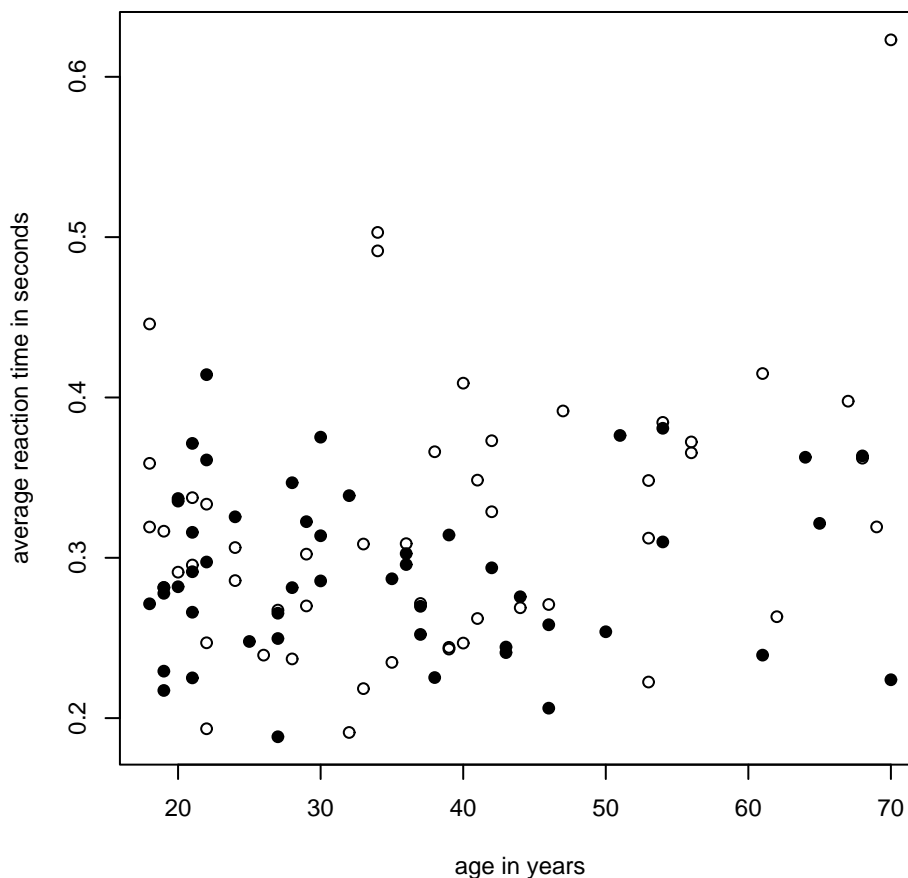
- c) What conclusions do you think Fred should draw from this data? Discuss what can and cannot be concluded with reference to the plots and regression model fits shown above, as well as your answers to (a) and (b) above. Discuss for each of the covariates (**sex**, **age**, and **education**) whether or not there is good reason to believe that that covariate is associated with **time**. In particular, discuss what the data show or do not show regarding Fred's idea that greater years of education might be associated with lower reaction time.

Fred conducted a second study to determine whether or not reaction time is influenced by drinking coffee containing caffeine. He recruited 100 subjects at local coffee shops, restricting these subjects to be between the ages of 18 and 70 years. He randomly split these 100 subjects into two groups of 50. One group of 50 subjects drank a cup of regular coffee (containing caffeine) and 30 minutes later was tested for reaction time. The other group of 50 subjects drank a cup of decaffeinated coffee and 30 minutes later took the same reaction time test. Neither the subjects nor the research assistant (call her Mary), who recorded the data knew whether the coffee drunk by a subject was regular or decaffeinated. (Of course, this information was recorded for later analysis.)

The data obtained consists of the values of three variables for each of the 100 subjects:

**group** Which group the subject was in: 0 = decaffeinated, 1 = regular.  
**age** The subject's age, in years  
**time** Average reaction time in the test, in seconds (lower is better)

The data is plotted below, with black dots for subjects drinking regular coffee, and white dots for subjects drinking decaffeinated coffee:



Fred performed a two-sample  $t$  test of the null hypothesis that the mean value for `time` is the same for `group=0` and `group=1` (using the version in which variances in the two groups are not assumed to be equal) . The result was as follows:

```
data: time[group == 0] and time[group == 1]
t = 1.9952, df = 82.232, p-value = 0.04933
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 8.326383e-05 5.516615e-02
sample estimates:
mean of x mean of y
0.3198746 0.2922499
```

Since the  $p$ -value of 0.04933 is less than the commonly-used standard significance level of 0.05, Fred felt he could conclude that the difference observed in the sample was statistically significant, and that there was therefore good reason to think that drinking coffee containing caffeine improves (ie, reduces) reaction time, on average over the population from which the subjects were drawn, with the magnitude of the improvement estimated to be  $0.319 - 0.292 = 0.027$  seconds.

Fred told Mary (his assistant) of this preliminary conclusion. She wondered if any apparent differences between the groups might just be a result of one group having a greater average age than the other. Mary therefore advocated fitting a linear regression model for `time` with `group` and `age` as covariates, and then seeing what the  $p$ -value was for testing the null hypothesis that the regression coefficient of `group` is zero. The results of a standard least-squares fit of this model in R are as follows:

```
Call:
lm(formula = time ~ group + age)

Residuals:
    Min       1Q   Median       3Q      Max
-0.122739 -0.048999 -0.003313  0.037235  0.271038

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.2816240  0.0199452  14.120 <2e-16 ***
group        -0.0239847  0.0136866  -1.752  0.0829 .
age           0.0010055  0.0004595   2.188  0.0310 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06793 on 97 degrees of freedom
Multiple R-Squared:  0.08425,    Adjusted R-squared:  0.06536
F-statistic: 4.462 on 2 and 97 DF,  p-value: 0.01400
```

Since the  $p$ -value of 0.0829 for the regression coefficient of `group` is greater than the standard significance level of 0.05, Mary thought that the study does not provide good evidence that drinking coffee with caffeine improves reaction time.

Answer the following questions about this second study:

- d) Is this an experiment or an observational study? What does your answer to this question say about the ability of this study to answer questions about the causal factors influencing reaction time?

e) Consider the following three views regarding Fred's analysis using a simple  $t$  test and Mary's analysis using a linear regression model:

- 1) The purpose of randomly assigning subjects to groups is to eliminate any dependence of the group assigned on other variables. The assumption of independence needed to justify a  $t$  test are therefore satisfied for this study. The  $p$ -value that Fred's  $t$  test gives is therefore a valid indication of statistical significance, which can be used to justify Fred's conclusion, regardless of what  $p$ -value might be produced by some other test, such as the one done by Mary.
- 2) Although random assignment usually eliminates problems with groups being associated with other variables, it is possible for randomization to fail to do this, due just to bad luck. We can check whether this has happened for this study by doing a  $t$  test of the null hypothesis that the mean age is the same for the two groups. The result is as follows:

```
data: age[group == 0] and age[group == 1]
t = 1.2121, df = 97.889, p-value = 0.2284
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.30671  9.54671
sample estimates:
mean of x mean of y
 38.04    34.42
```

Since the  $p$ -value of 0.2284 is greater than the standard significance level of 0.05, we do not need to worry about age differences between the two groups. So the conclusion from Fred's  $t$  test is valid. But if the  $p$ -value above had been smaller than 0.05, the regression analysis advocated by Mary would be more appropriate.

- 3) What matters are the actual ages of the subjects in the two groups — the fact that they were assigned randomly does not undo the effect that a difference in average age between groups could have. The regression analysis advocated by Mary is therefore the correct approach, especially since Fred's previous study suggests that age actually is related to reaction time, as is also indicated by the small  $p$ -value for the coefficient of age in Mary's regression model.

Which of these views, if any, is correct? Explain why you think each of the three views above is either correct or incorrect, and discuss any other issues you see regarding whether the analysis done by Fred or the analysis done by Mary is better.

f) Suggest one or more changes to how this study was done that could produce more precise results, without increasing the number of subjects.