

Advanced Quantitative Methods

Lab 2: Linear regression

Johan A. Elkind
jos.elkind@ucd.ie

2 February 2018

1. Open the `films.dta` data set. Create a new variable `highrating`, which is 1 for films rated 3 or higher, 0 otherwise.

```
library(rio)
films <- import("http://www.joselkink.net/wp-content/uploads/2013/01/films.dta")
```

- (a) regress `desclength` on a constant
 - (b) regress `desclength` on `castsize`
 - (c) regress `desclength` on `castsize`, `highrating`, `length`
2. Based on the last regression:
 - (a) Which observation has the largest residual?
Use `residuals(lm(...))`, replacing “...” according to the previous question.
 - (b) Compute mean and median of residuals
 - (c) Compute correlation between residuals and fitted values
Use `fitted(lm(...))`
 - (d) Compute correlation between residuals and `length`
 - (e) All other predictors held constant, what would be the difference in predicted description length between high and low rated movies?
 3. Open the `uswages.dta` data set.

```
us <- import("http://www.joselkink.net/wp-content/uploads/2013/01/uswages.dta")
```

 - (a) Regress `wage` on `educ`, `exper` and `race`.
 - (b) Interpret the results
 - (c) Plot residuals against fitted values and against `educ`
 - (d) Repeat with `log(wage)` as dependent variable
 4. Open the `uswages.dta` data set.
 - (a) Regress `wage` on `educ`, `exper` and `race`.
 - (b) What proportion of the variance in `wage` is explained by these three variables?