Week 9: Linear Regression II

POP88162 Introduction to Quantitative Research Methods

R^2

We will start by revisiting the example from last week. Here we are fitting a bivariate (i.e. with only one independent variable) regression model with country's GDP per capita (`gdp_per_capita`) as an outcome (dependent variable) and the longevity (`democracy_duration`) of its political regime as an explanatory (independent) variable.

```
democracy_gdp_2020 <- read.csv("../data/democracy_gdp_2020.csv")
```

Here we saved the fitted model object under the name `lm_fit_1`. Now we can use `summary()` function to print out detailed model output.

```
lm_fit_1 <- lm(gdp_per_capita ~ democracy_duration, data = democracy_gdp_2020)
summary(lm_fit_1)
```

```
##
## Call:
## lm(formula = gdp_per_capita ~ democracy_duration, data = democracy_gdp_2020)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -44806  -8756  -4944   4820 163717
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)         5051.44    2370.78   2.131   0.0345 *
## democracy_duration   182.22      35.15   5.185 5.99e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20900 on 173 degrees of freedom
##   (20 observations deleted due to missingness)
## Multiple R-squared:  0.1345, Adjusted R-squared:  0.1295
## F-statistic: 26.88 on 1 and 173 DF,  p-value: 5.995e-07
```

What is the coefficient of determination (R^2) for this model? What is its interpretation?

Multiple Linear Regression

Now let's add another explanatory (independent) variable to our model. We will use regime type (`democracy`) to model economic performance measured as GDP per capita.

The syntax for fitting multiple (multivariate) linear regression model is very similar to syntax for fitting a simple linear bivariate model. We can just add another variable to the formula specification that now looks like `Y ~ X_1 + X_2` with the name of the dependent variable being on the left-hand side of the `~` (tilde) and one or more names of independent varable(s) on the right.

```
lm_fit_2 <- lm(gdp_per_capita ~ democracy_duration + democracy, data = democracy_gdp_2020)
lm_fit_2
```

```
##
## Call:
## lm(formula = gdp_per_capita ~ democracy_duration + democracy,
##     data = democracy_gdp_2020)
##
## Coefficients:
##        (Intercept)  democracy_duration           democracy
##            -4971.8               201.8             14649.7
```

What do these numbers tell you? What is a null hypothesis and an alternative hypothesis for this test? What is your decision regarding a null hypothesis given the output above?

How did the calculate R^2 change and why?

Binary Independent Variable

Start by fitting a bivariate linear regression model from the lecture where regime longevity (`democracy_duration`) is the outcome and regime type (`democracy`) is explanatory (independent) variable.

```
lm_fit_3 <- lm(democracy_duration ~ democracy, data = democracy_gdp_2020)
summary(lm_fit_3)
```

```
##
## Call:
## lm(formula = democracy_duration ~ democracy, data = democracy_gdp_2020)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -52.610 -24.610 -10.051   7.949 175.949
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   54.610      4.975  10.976   <2e-16 ***
## democracy     -9.560      6.396  -1.495    0.137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.66 on 193 degrees of freedom
## Multiple R-squared:  0.01144,    Adjusted R-squared:  0.00632
## F-statistic: 2.234 on 1 and 193 DF,  p-value: 0.1366
```

What is your substantive conclusion given this output?

Now recall working with factor variables from our previous workshop. Change the coding of regime type variable in such a way that autocracies now are represented by 1 and democracies by 0. Re-fit the model from above.

What changed? How do the two models compare?