SLR example - Math 150

Jo Hardin

Spring 2017

```
library(readr)  # contains the read_delim package
library(dplyr)
library(ggplot2)
```

The data below represents 10 different variables on health of a country measured on 143 countries. Data taken from Unlocking the Power of Data by Lock5, originally from the Happy Planet Index Project http://www.happyplanetindex.org/. Region of the world is coded as 1 = Latin America, 2 = Western nations, 3 = Middle East, 4 = Sub-Saharan Africa, 5 = South Asia, 6 = East Asia, 7 = former Communist countries. We are going to investigate happiness and life expectancy.

Reading the data into R

```
happy <- read_delim("~/Dropbox/teaching/math150/spring17/happyPlanet.txt", delim="\t")
glimpse(happy)
```

```
## Observations: 143
## Variables: 11
## $ Country        <chr> "Albania", "Algeria", "Angola", "Argentina", "A...
## $ Region         <int> 7, 3, 4, 1, 7, 2, 2, 7, 5, 7, 2, 1, 4, 5, 1, 7,...
## $ Happiness      <dbl> 5.5, 5.6, 4.3, 7.1, 5.0, 7.9, 7.8, 5.3, 5.3, 5....
## $ LifeExpectancy <dbl> 76.2, 71.7, 41.7, 74.8, 71.7, 80.9, 79.4, 67.1,...
## $ Footprint      <dbl> 2.2, 1.7, 0.9, 2.5, 1.4, 7.8, 5.0, 2.2, 0.6, 3....
## $ HLY            <dbl> 41.7, 40.1, 17.8, 53.4, 36.1, 63.7, 61.9, 35.4,...
## $ HPI            <dbl> 47.91, 51.23, 26.78, 58.95, 48.28, 36.64, 47.69...
## $ HPIRank        <int> 54, 40, 130, 15, 48, 102, 57, 85, 31, 104, 64, ...
## $ GDPperCapita   <int> 5316, 7062, 2335, 14280, 4945, 31794, 33700, 50...
## $ HDI            <dbl> 0.801, 0.733, 0.446, 0.869, 0.775, 0.962, 0.948...
## $ Population     <dbl> 3.15, 32.85, 16.10, 38.75, 3.02, 20.40, 8.23, 8...
```

Running the linear model (lm)

```
happy.lm = lm(LifeExpectancy ~ Happiness, data=happy) # summary(happy.lm)  # summary more info
summary(happy.lm)$coef
```

```
##              Estimate Std. Error  t value     Pr(>|t|)
## (Intercept) 28.223143  2.2798520 12.37938 2.760674e-24
## Happiness    6.693042  0.3752837 17.83462 5.782911e-38
```

Ouptut

Some analyses will need the residuals, fitted values, or coefficients individually.

```
head(resid(happy.lm))  # resid(happy.lm)  # for fun, look at all of your residuals
```

```
##           1           2           3           4           5           6
##  11.1651273   5.9958231 -15.3032227  -0.9437395  10.0116481  -0.1981728
```

```
head(fitted(happy.lm)) # fitted(happy.lm) # for fun, look at all of your fitted values.
```

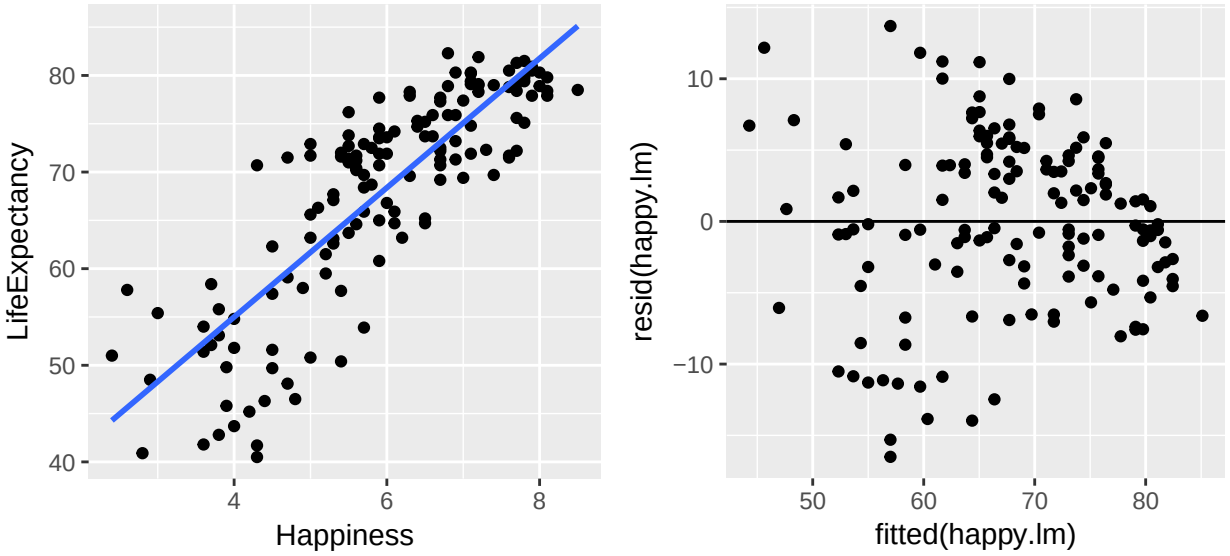
```
##        1        2        3        4        5        6
## 65.03487 65.70418 57.00322 75.74374 61.68835 81.09817
```

```
coef(happy.lm)
```

```
## (Intercept)   Happiness
##   28.223143    6.693042
```

We can plot the main relationship, or we can plot the residuals (to check that technical conditions hold):

```
ggplot(happy, aes(x=Happiness, y=LifeExpectancy)) + geom_point() +
        geom_smooth(method="lm", se=FALSE)
qplot(fitted(happy.lm), resid(happy.lm)) + geom_hline(yintercept=0)
```



Intervals of interest: mean response, individual response, and parameter(s).

```
predict.lm(happy.lm, newdata=list(Happiness=c(4,7)),interval=c("conf"), level=.95)
```

```
##        fit      lwr      upr
## 1 54.99531 53.24675 56.74387
## 2 75.07444 73.78057 76.36830
```

```
predict.lm(happy.lm, newdata=list(Happiness=c(4,7)),interval=c("pred"), level=.95)
```

```
##        fit      lwr      upr
## 1 54.99531 42.72945 67.26117
## 2 75.07444 62.86510 87.28377
```

```
confint(happy.lm, level=.95)
```

```
##                 2.5 %    97.5 %
## (Intercept) 23.716032 32.730255
## Happiness    5.951131  7.434952
```