Assignment 1 DECS-345

Sebastian Salomon

2022-09-24

Assignment 1

1 Learning Checks

Answer the following questions in your own words. None of your answers should be longer than one paragraph.

1. (easy) What is OLS? How is it related to the linear model?

Ordinary Least Squares (OLS) is a least square method for estimating unknown parameters in a linear regression model. It chooses the parameters by minimizing the sum of squares of the error. Error = difference between observed values and predicted values.

2. (medium) What do the coefficients in a linear model tell us?

It tells us whether there is a positive or negative correlation between the independent and the dependent variable. It also tells us the magnitude of such correlation.

3. (hard) Why can the estimated intercept in a linear model sometimes not be given a sensible interpretation?

There are various cases where the x axis can't be 0. In these cases the interpretation of the intercept won't seem realistic in the real world. The model, however, can still be useful to interpret by using reasonable x values.

4. Estimate a linear regression with mortality as outcome and caseweight as independent variable. Based on this regression, what is the expected change in patient mortality if Kletzl Hills hires ten additional FTE nurses?

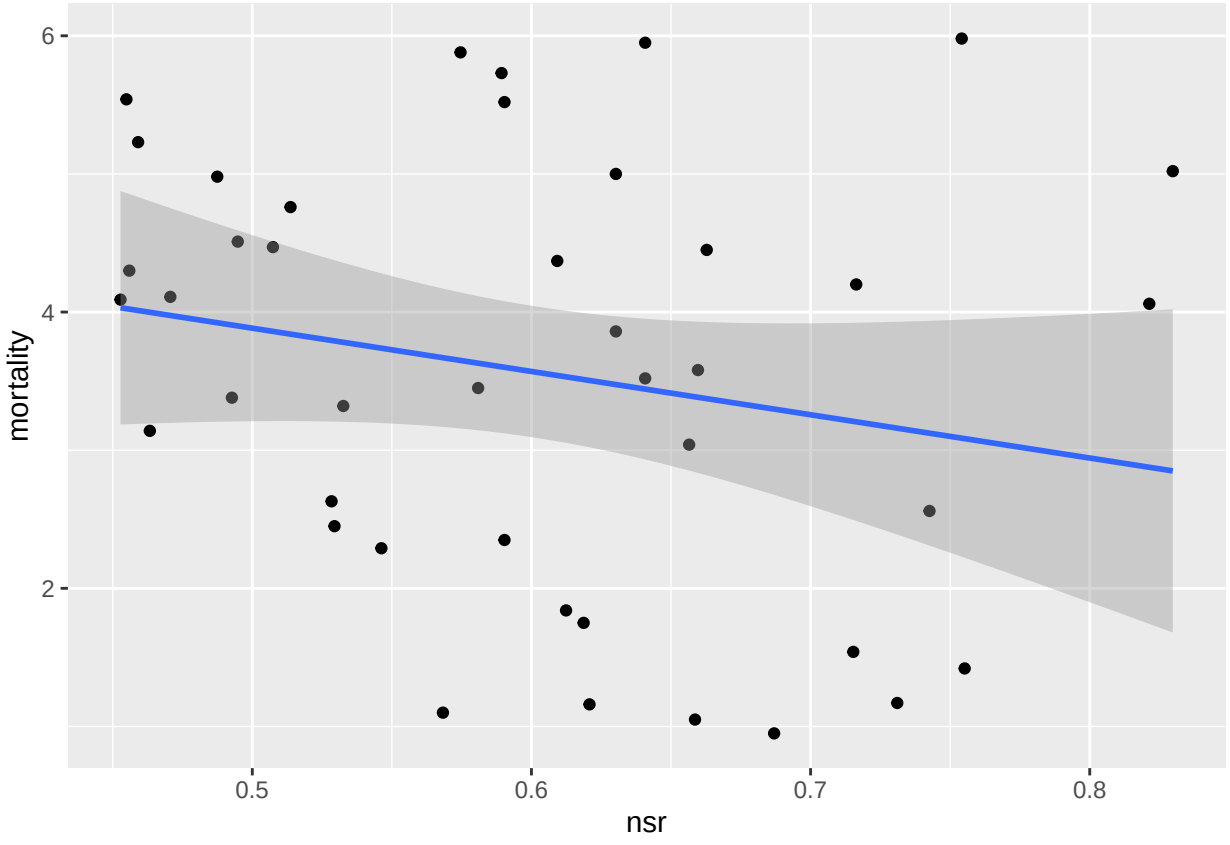
```
library(tidyverse)
library(ggplot2)
library(haven)


hospital_data <- read_dta('C:/Users/ssalo/Documents/2022FA_DECS_435-0_SEC1/hospitalA.dta')

lm(mortality ~ nsr, data = hospital_data) %>% summary()
```

```
##
## Call:
## lm(formula = mortality ~ nsr, data = hospital_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5693 -1.3184  0.1062  1.0719  2.8930
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.450      1.416   3.848 0.000419 ***
## nsr           -3.133      2.320  -1.350 0.184492
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.524 on 40 degrees of freedom
## Multiple R-squared:  0.0436, Adjusted R-squared:  0.01969
## F-statistic: 1.823 on 1 and 40 DF,  p-value: 0.1845
```

```
ggplot(data=hospital_data, aes(x=nsr, y=mortality)) +
  geom_point(stat="identity")+
  geom_smooth(method='lm')
```



Kletzl Hills currently has the following data:

- Caseweight: 1.427

- NSR: 0.6302478 = (Avg full time nurses = x)/(Avg number of staffed hospital beds = 150 beds)
- Mortality: 3.86%

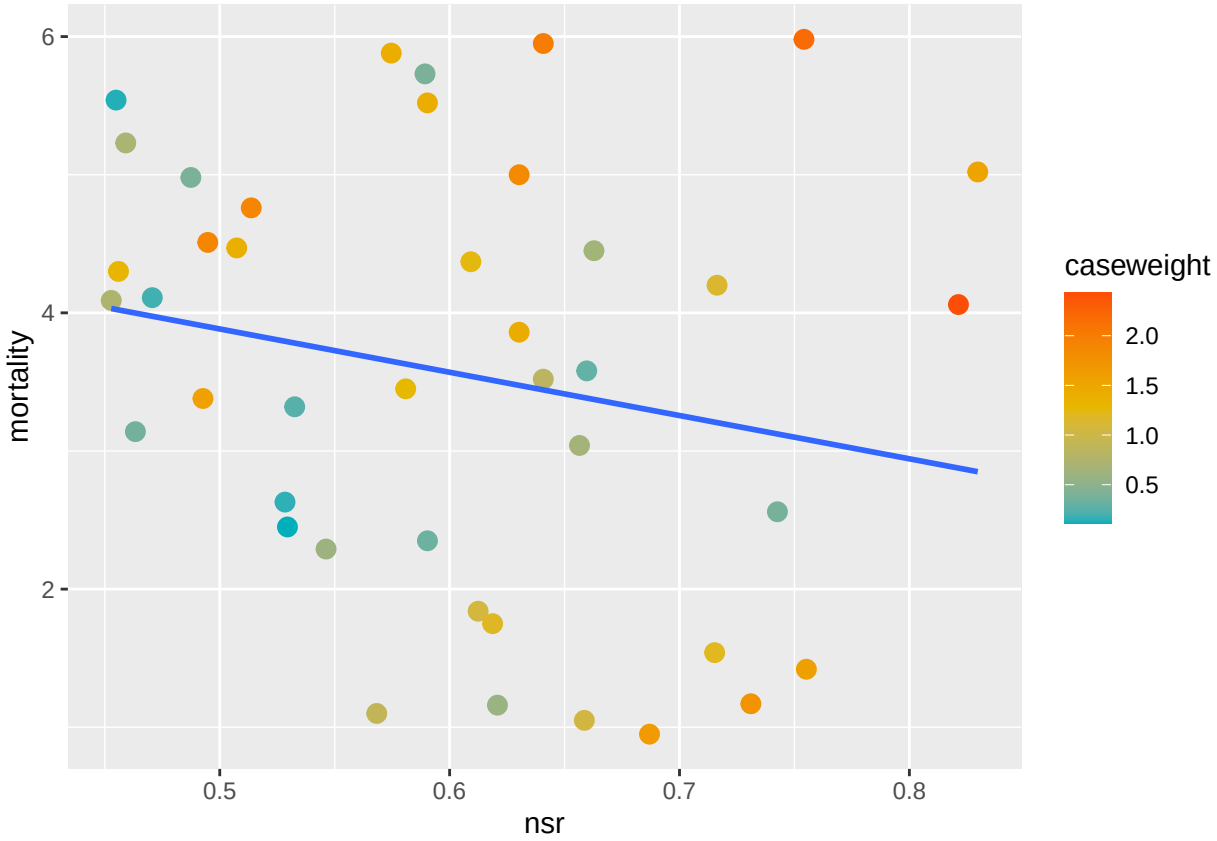
At this moment the average full time nurses is equal to: $x = .63 * 150 = 94.5$ If Kletzl hires 10 extra full time nurses the new NSR would be $104.5/150 = .6966$ That is an increase of $.6966 - .63024 = .06642$ The model tells us that for every 1 point increase in nsr, we get 3.133% decrease in mortality. **So a .06642 increase in nsr would give us a .208% decrease in mortality.**

5. What is the expected change in patient mortality if Kletzl Hills hires ten additional FTE nurses, holding caseweight fixed?

```
fit_caseweight <- lm(mortality ~ nsr + caseweight, data = hospital_data)
summary(fit_caseweight)
```

```
##
## Call:
## lm(formula = mortality ~ nsr + caseweight, data = hospital_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.65797 -1.07184  0.06002  1.11010  2.62685
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.8858     1.3827   4.257 0.000126 ***
## nsr          -5.2861     2.4818  -2.130 0.039539 *
## caseweight    0.8149     0.4064   2.005 0.051939 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.47 on 39 degrees of freedom
## Multiple R-squared:  0.133,  Adjusted R-squared:  0.08851
## F-statistic: 2.991 on 2 and 39 DF,  p-value: 0.0619
```

```
ggplot(data=hospital_data, aes(x=nsr, y=mortality)) +
geom_point(aes(color = caseweight), size = 3) +
  scale_color_gradientn(colors = c("#00AFBB", "#E7B800", "#FC4E07"))+
  stat_smooth(method='lm', se=FALSE)
```



Kletzl Hills currently has the following data:

- Caseweight: 1.427
- NSR: 0.6302478 = (Avg full time nurses = x)/(Avg number of staffed hospital beds = 150 beds)
- Mortality: 3.86%

At this moment the average full time nurses is equal to: $x = .63 * 150 = 94.5$ If Kletzl hires 10 extra full time nurses the new NSR would be $104.5/150 = .6966$ That is an increase of $.6966 - .63024 = .06642$ The model tells us that for every 1 point increase in nsr, we get 5.2861% decrease in mortality. **So a .06642 increase in nsr would give us a .3511% decrease in mortality.**

6. Does controlling for caseweight affect the expected change in patient mortality if Kletzl Hills hires ten additional FTE nurses? Which estimate should you trust more and why?

Yes, an estimated parameter will not change when a new variable is added only if the parameters are completely uncorrelated. In practice this will very rarely be the case, so controlling for caseweight does affect the expected change in patient mortality.

I trust the estimate of the 2nd model, the model where we control for caseweight, more. The adjusted R squared value of the 2nd model is a little higher meaning that more of the data is explained by the model. Also, the p-value for the nsr coefficient is lower in the 2nd model, so it has a greater statistical relationship to the dependent variable than the coefficient of the 1st model does.

7. With the addition of 10 nurses, what is your best estimate for the expected increase in bonus?

The additional 10 nurses would give us a .06642 increase in nsr and the model gives us a .3511% decrease in mortality.

Bonus reduction without increase in staff: $5million Ö 3.86 = 19.3million$

New Bonus reduction with increase in staff: $5million Ö (3.86 - .3511) = 17.544million$

Expected increase in bonus: $19.3 - 17.544 = 1.7555million$

8. Using your best managerial judgement, what course of action would you recommend and why?

The expected increase in bonus calculated using the 2nd model gives the hospital an expected 1.7555 million increase in bonus. But we have to take into account the costs of bringing 10 staff into the hospital. The cost per staff is: $60,000 * 10 = .6million$. Since the expected increase is much higher than the cost of bringing 10 new staff members, I would recommend Dr. Newport to hire the 10 staff members.