Lab 14 - Bivariate Regression & Interpretation

William Bernard

November 28, 2017

Complete the following exercises below and include all code used to find the answers. Knit together the PDF document and commit both the Lab 14 RMD file and the PDF document to Git. Push the changes to GitHub so both documents are visible in your public GitHub repository.

1. Select the main focal relationship you're interested in exploring for your poster project.

    a. Describe the response variable and the explanatory variable and the theoretical relationship you believe exists between these two variables.

I am looking to investigate the theoretical relationship between wheteher or not an individual has experienced the death of a spouse and their overall life satisfaction.

    b. Conduct a simple (bivariate) linear regression on your focal relationship and save the model object. Print out the full results by calling `summary()` on your model object.

```r
sat_life_num <- as.numeric(changing_lives_subset_final$sat_life)

ever_wid_num <- as.numeric(changing_lives_subset_final$ever_widowed)



lm(sat_life_num ~ ever_wid_num) %>%
  summary()
```

```
##
## Call:
## lm(formula = sat_life_num ~ ever_wid_num)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2547 -0.2547 -0.2547  0.7453  2.7932
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.15890    0.06853  31.501   <2e-16 ***
## ever_wid_num  0.04791    0.03756   1.276    0.202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.932 on 3547 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.0004585,  Adjusted R-squared:  0.0001767
## F-statistic: 1.627 on 1 and 3547 DF,  p-value: 0.2022
```

    c. What is the direction, magnitude, and statistical significance of the bivariate association between the explanatory and response variables.

```r
quickplot(sat_life_num, ever_wid_num) +
  geom_jitter()
```

```
## Warning: Removed 68 rows containing missing values (geom_point).
```

```
## Warning: Removed 68 rows containing missing values (geom_point).
```
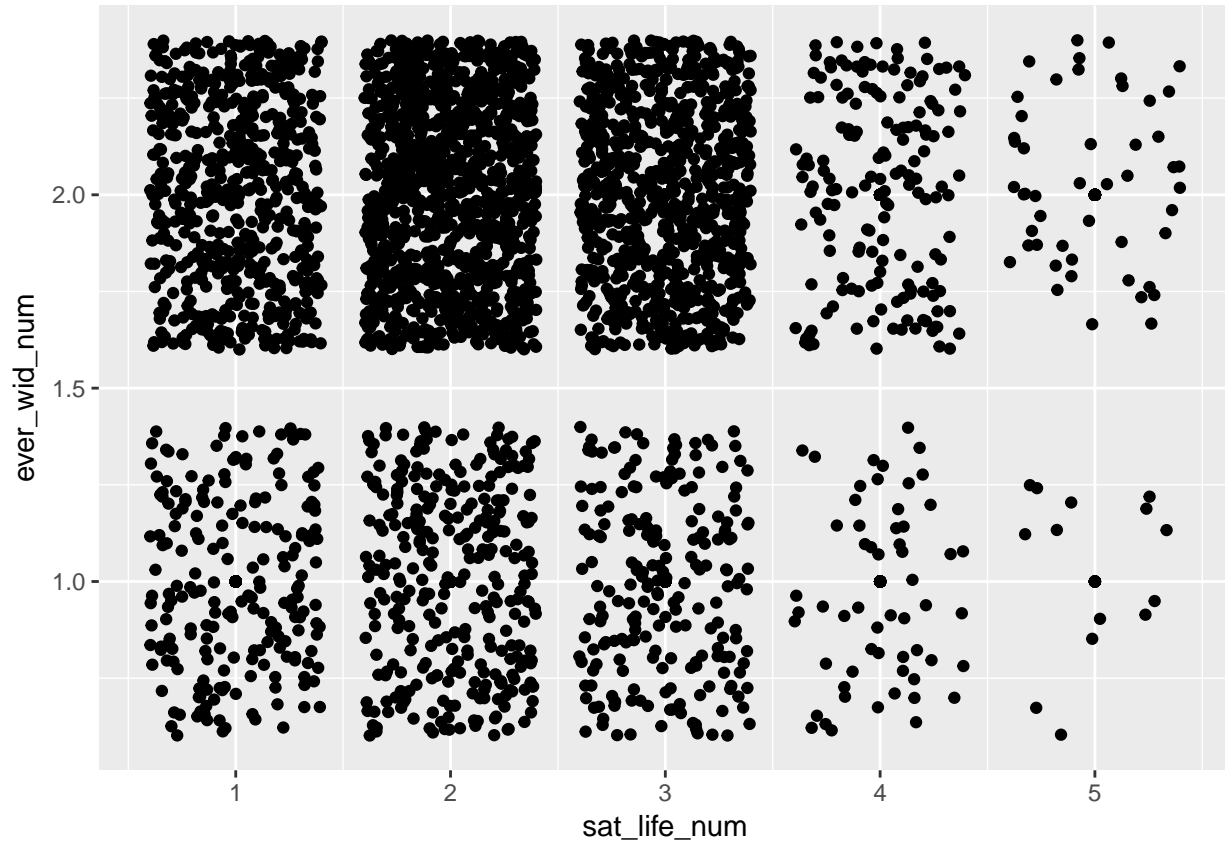

direction of the plot is somewhat meaningless considering the coercion, magnitude is very moderate, and it is not statistically significant at .05

    d. What is the meaning of the model intercept?

at a "ever widowed" value of 0 you would expect to see a satisfaction level of 1.75 which doesnt mean a whole lot.

    e. How well does the bivariate model fit the data? How is this information calculated?

I would tend to think it fits fairly well. i have a very small residual considering the nature of the data and how I had to coerce it. Because of that I think this is appropriate.

    f. Is the observed association between the independent variable and dependent variable consistent with your hypothesis? Why or why not?

the association is .2022, which is far from statistically significant, but not too crazy as to say there is absolutely nothing there. I thought the connection may have been slightly stronger, but its possible that that is also affected by position in the life course.

2. Select a different focal relationship related to your project. This could be:

- **A different response and a different explanatory variable**
- **A different response and the same explanatory variable**
- **The same response and a different explanatory variable**

a. Describe the response variable and the explanatory variable and the theoretical relationship you believe exists between these two variables.

I am looking at the relationship between having a child die and overall life satisfaction.

b. Conduct a simple (bivariate) linear regression on your focal relationship and save the model object. Print out the full results by calling **summary()** on your model object.

```
child_die_num<- as.numeric(changing_lives_subset_final$child_die)

child_sat_lm <- lm(sat_life_num ~ child_die_num)



summary(child_sat_lm)
```

```
##
## Call:
## lm(formula = sat_life_num ~ child_die_num)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.2627 -0.2627 -0.2627  0.7373  2.8163
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.10480    0.07743  27.183   <2e-16 ***
## child_die_num  0.07894    0.04147   1.904    0.057 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9362 on 3590 degrees of freedom
##   (25 observations deleted due to missingness)
## Multiple R-squared:  0.001008,   Adjusted R-squared:  0.0007301
## F-statistic: 3.624 on 1 and 3590 DF,  p-value: 0.05704
```

c. What is the direction, magnitude, and statistical significance of the bivariate association between the explanatory and response variables.

```
quickplot(child_die_num, sat_life_num) +
  geom_jitter()
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```

```
## Warning: Removed 25 rows containing missing values (geom_point).
```
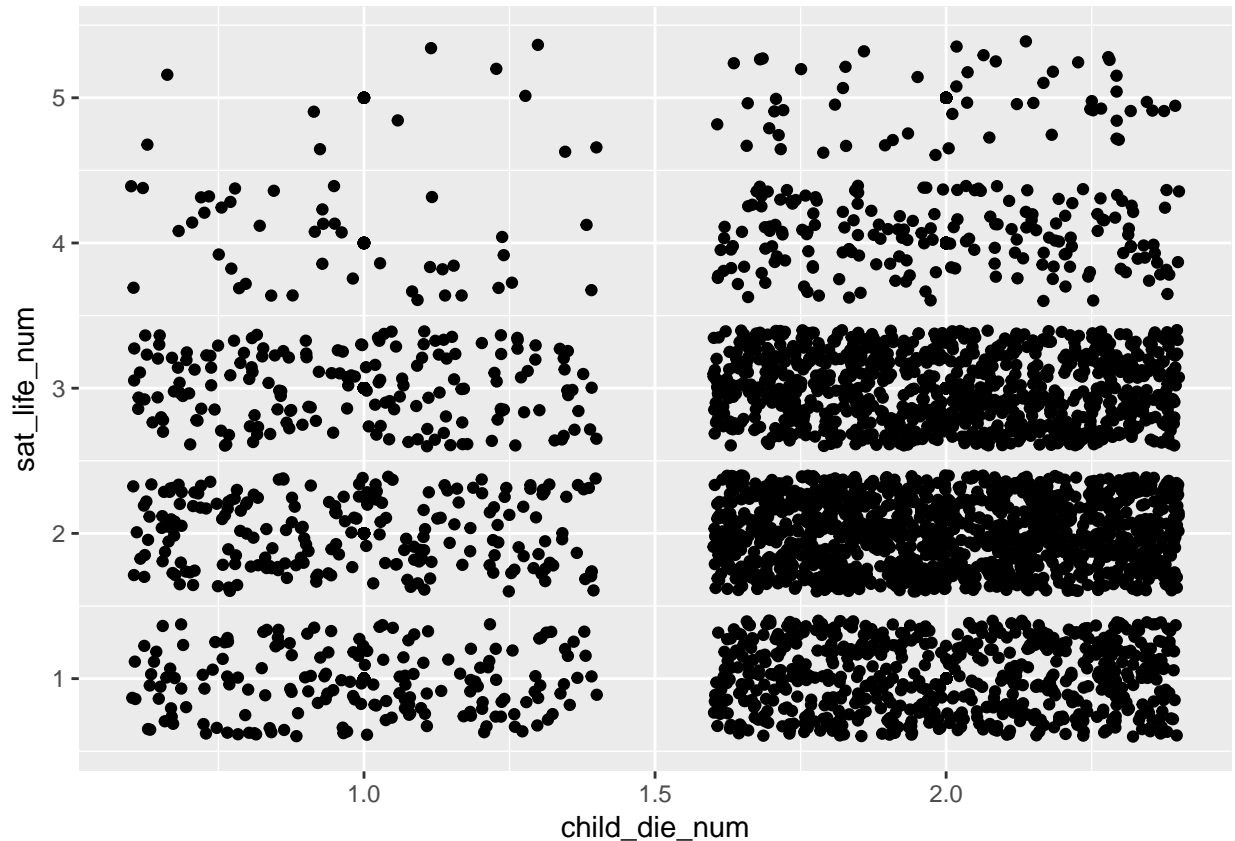
once again using a nominal variable for x leaves us with a linar regression that doesnt have a decernable direction, there is however a strong association between the two variables, and the p value of .057 would tend me towards the claim that it is statistically significant.

    d. What is the meaning of the model intercept?

at no chldren dead you would expect a sat of 2.1

    e. How well does the bivariate model fit the data? How is this information calculated?

it fits fairly well, the r squared value is really small but considering the nature of the data I think this is still okay. r squared is a test of the agrigrate regression, it lets you know how well your model helps predict, and summaraizes discrepencies.

    f. Is the observed association between the independent variable and dependent variable consistent with your hypothesis? Why or why not?

the association is very much in line with my hypothesis, the loss of a child has a strong connection with life satisfaction. which may seem like a no brainer but is a good place to start.