Test 1, Stat 245

Trevor VanVeldhuisen

October 18, 2021

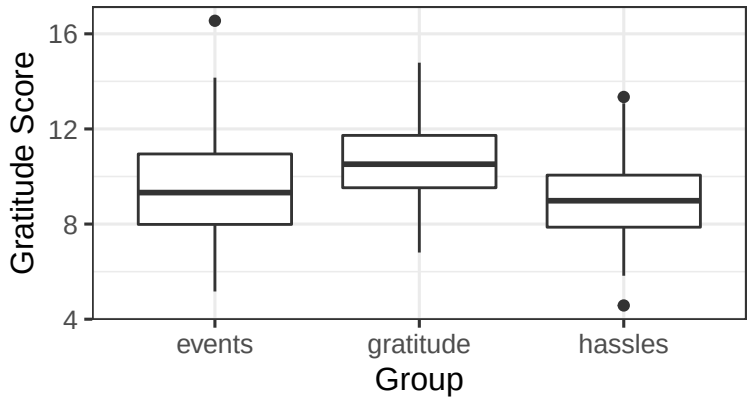
```
gratitude <- read_csv('https://sldr.netlify.app/data/gratitude-experiment.csv')
```

```
glimpse(gratitude)
```

```
## Rows: 196
## Columns: 12
## $ gratitude_score    <dbl> 9.518282, 10.155757, 13.608307, 10.736081, 10.84952~
## $ life_rating        <dbl> 5.500532, 4.111546, 4.591626, 4.160890, 6.699108, 6~
## $ week_rating        <dbl> 6.187401, 4.816418, 5.021152, 6.683120, 4.860684, 6~
## $ illness_score      <dbl> 5.65346133, -2.89429681, -9.61398606, 5.12235596, 6~
## $ exercise_hours     <dbl> 2.333122, 3.442714, 5.569992, 1.353946, 1.463923, 3~
## $ group              <chr> "gratitude", "gratitude", "gratitude", "gratitude",~
## $ jittered_group     <dbl> 2.342593, 1.821113, 1.741265, 1.798877, 1.916003, 1~
## $ overall_mean       <dbl> 9.75556, 9.75556, 9.75556, 9.75556, 9.75556, 9.7555~
## $ group_mean         <dbl> 10.6413, 10.6413, 10.6413, 10.6413, 10.6413, 10.641~
## $ d_from_group       <dbl> -1.12301345, -0.48553801, 2.96701159, 0.09478574, 0~
## $ d_from_overall     <dbl> -0.23727801, 0.40019744, 3.85274704, 0.98052119, 1.~
## $ group_from_overall <dbl> 0.8857354, 0.8857354, 0.8857354, 0.8857354, 0.88573~
```

The predictor variable will be the 3-level group (gratitude, hassles, events as the control group) while the response variable is the gratitude score. Each group in the study was given a different task to do each week, and the goal of that study was to see if there is an association between the group assignment and the overall gratitude score. While the overall goal is to see if there is an association between group and gratitude score, I also added in the predictor variable life rating because I believe life rating would be another predictor in addition to group that would predict gratitude score. The data set was 196 rows of data and there were three categories in my main predictor so I thought including more than adding more predictor variables could have a negative effect the ability to fit the model. There was nothing in the background that stated these variables were important and were not included in the primary research question.

```
gf_boxplot(gratitude_score ~ group , data = gratitude) %>%
  gf_labs(y = 'Gratitude Score',
          x = 'Group')
```



This graph shows the distribution of gratitude score for each group. From this boxplot we can see that the gratitude group has a higher median of gratitude score than the events group as the control group. The hassles group has a pretty similar median gratitude score to the control group.

```
mlr3 <- lm(gratitude_score ~ group + life_rating, data = gratitude)
```

```
summary(mlr3)
```

```
##
## Call:
## lm(formula = gratitude_score ~ group + life_rating, data = gratitude)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7003 -1.3238 -0.1666  1.0438  6.1955
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.5188     0.8501   8.845 5.93e-16 ***
## groupgratitude   0.9397     0.3397   2.767  0.00622 **
## grouphassles    -0.5188     0.3334  -1.556  0.12131
## life_rating      0.4352     0.1766   2.464  0.01461 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.896 on 192 degrees of freedom
## Multiple R-squared:  0.1311, Adjusted R-squared:  0.1175
## F-statistic: 9.654 on 3 and 192 DF,  p-value: 5.764e-06
```

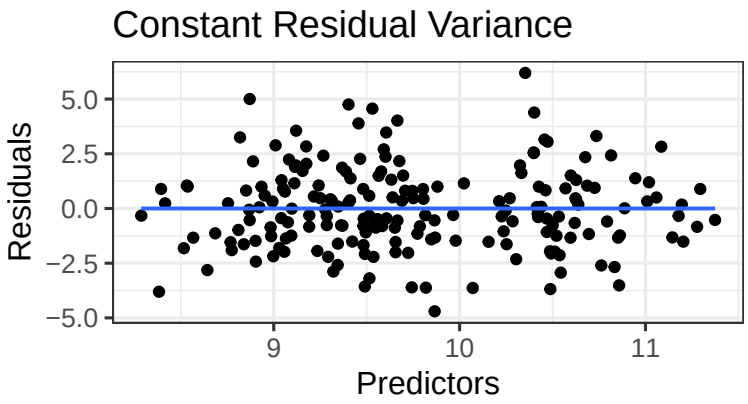
Mathematical Equation: gratitude_score = 7.5188 + 0.9397groupgratitude -0.5188grouphassles + 0.4352life_rating + E groupgratitude{1 if in group_gratitude); 0 otherwise} group_hassles{1 if in group_hassles; 0 otherwise} life_rating is a continuous variable E ~ N(0, 1.896)

The adjusted R-squared = 0.1175. R-squared tells us how well the model fits the data and how much variability the model explains. Since R-squared is low, the data does not fit the data well.

```
gratitude <- gratitude %>%
  mutate(resid = resid(mlr3),
         pred = predict(mlr3))
```

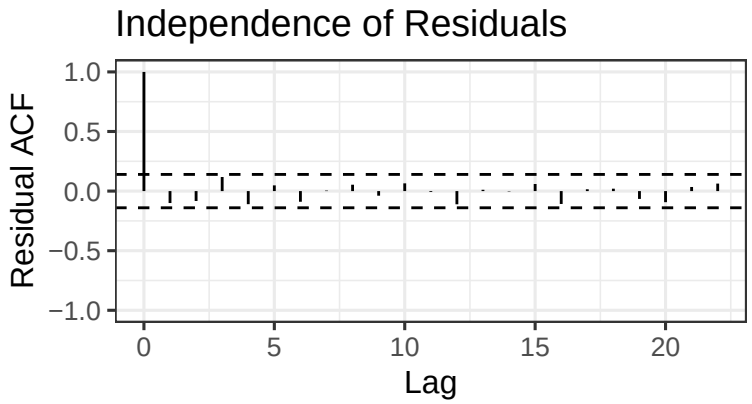
```
gf_point(resid ~ pred, data = gratitude) %>%
  gf_labs(title = 'Constant Residual Variance',
       x = 'Predictors',
          y = 'Residuals') %>%
```

```
  gf_lm()
```



The constant residual variance test does not seem to trumpet out. There is no pattern to this and the points seem evenly scattered around a mean of zero, so there is constant variance.

```
s245::gf_acf(~mlr3) %>%
  gf_lims(y = c(-1,1)) %>%
  gf_labs(title = 'Independence of Residuals')
```



This graph helps show the independence of the residuals, and the values of the residual autocorrelation function stay within the confidence intervals. This means the evidence is consistent with the residuals being independent.

```
g2 <- gratitude %>%
  select(gratitude_score, group, life_rating) %>%
  na.omit() %>%
  mutate(preds = predict(mlr3))
```

```
fake_data <- expand.grid(life_rating = seq(from = 0,
                                           by = 1,
                                           to = 8),
                         group = 'gratitude')
```

```
fake_data <- fake_data %>%
  mutate(pred = predict (mlr3, newdata = fake_data))
```

```
preds <- predict(mlr3, newdata = fake_data, se.fit = TRUE)
```

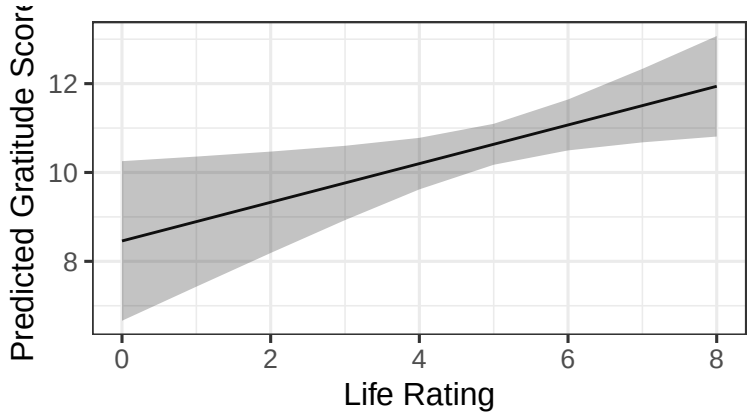
```
fake_data <- fake_data %>%
  mutate(pred = preds$fit,
```

```
         pred.se = preds$se.fit)
```

```
fake_data <- fake_data %>%
  mutate(CI_lower = pred - 1.96*pred.se,
         CI_upper = pred + 1.96*pred.se)
glimpse(fake_data)
```

```
## Rows: 9
## Columns: 6
## $ life_rating <dbl> 0, 1, 2, 3, 4, 5, 6, 7, 8
## $ group       <fct> gratitude, gratitude, gratitude, gratitude, gratitude, gra~
## $ pred        <dbl> 8.458496, 8.893727, 9.328958, 9.764189, 10.199420, 10.6346~
## $ pred.se     <dbl> 0.9164976, 0.7471659, 0.5821827, 0.4266232, 0.2957512, 0.2~
## $ CI_lower    <dbl> 6.662161, 7.429282, 8.187880, 8.928008, 9.619748, 10.17365~
## $ CI_upper    <dbl> 10.25483, 10.35817, 10.47004, 10.60037, 10.77909, 11.09565~
```

```
gf_line(pred ~ life_rating, data = fake_data) %>%
  gf_ribbon(CI_lower + CI_upper ~ life_rating) %>%
  gf_labs( y = 'Predicted Gratitude Score',
           x = 'Life Rating')
```




```
summary(lm(gratitude_score ~ life_rating, data = gratitude))
```

```
##
## Call:
## lm(formula = gratitude_score ~ life_rating, data = gratitude)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.891 -1.466 -0.098  1.219  5.899
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.2008     0.8828   8.157 4.22e-14 ***
## life_rating   0.5295     0.1806   2.932  0.00377 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.98 on 194 degrees of freedom
## Multiple R-squared:  0.04243,    Adjusted R-squared:  0.03749
## F-statistic: 8.596 on 1 and 194 DF,  p-value: 0.003775
```

Interpretation and Conclusion:

First, looking at the box plot, we see that the distribution of the gratitude score is higher in the gratitude group than the hassles and events group. Based on fitting the multiple regression model that included the predictors of group (gratitude, hassles, and events as the control group) and life rating and predicting gratitude score, group gratitude was a statistically significant predictor compared to the control group (p-value of 0.00622). To interpret the estimate, if a participant is in the gratitude group, then the gratitude score is predicted to be 0.9397 higher compared to the events group. Life rating is also a significant predictor of gratitude score, and for every life rating increase by 1 for a participant, we predict an increase in gratitude score by 0.4352 (p-value = 0.01461). Group hassles is not significantly different than control group events (p-value= 0.12131). The adjusted r-squared of the multiple regression model was 0.11, which means that the variables in the model do not fit the data well. Based on examining graphs, there is no evidence that the assumptions of the multiple regression model are violated for the independence test of the error terms and lack of non-linearity for the residuals against the predictors. For the sake of this analysis, I am going to assume it passes the distribution of residuals and the error variance constant tests. The prediction plot shows a good linear relationship with the predicted values of the gratitude score and the life rating predictor which is consistent with the results we saw in the analysis from the multiple regression model. In summary, our analysis answered the research question that there is an association between gratitude score and gratitude group when compared to the control (events) group, but hassles relative to the control group, is not associated with gratitude score.