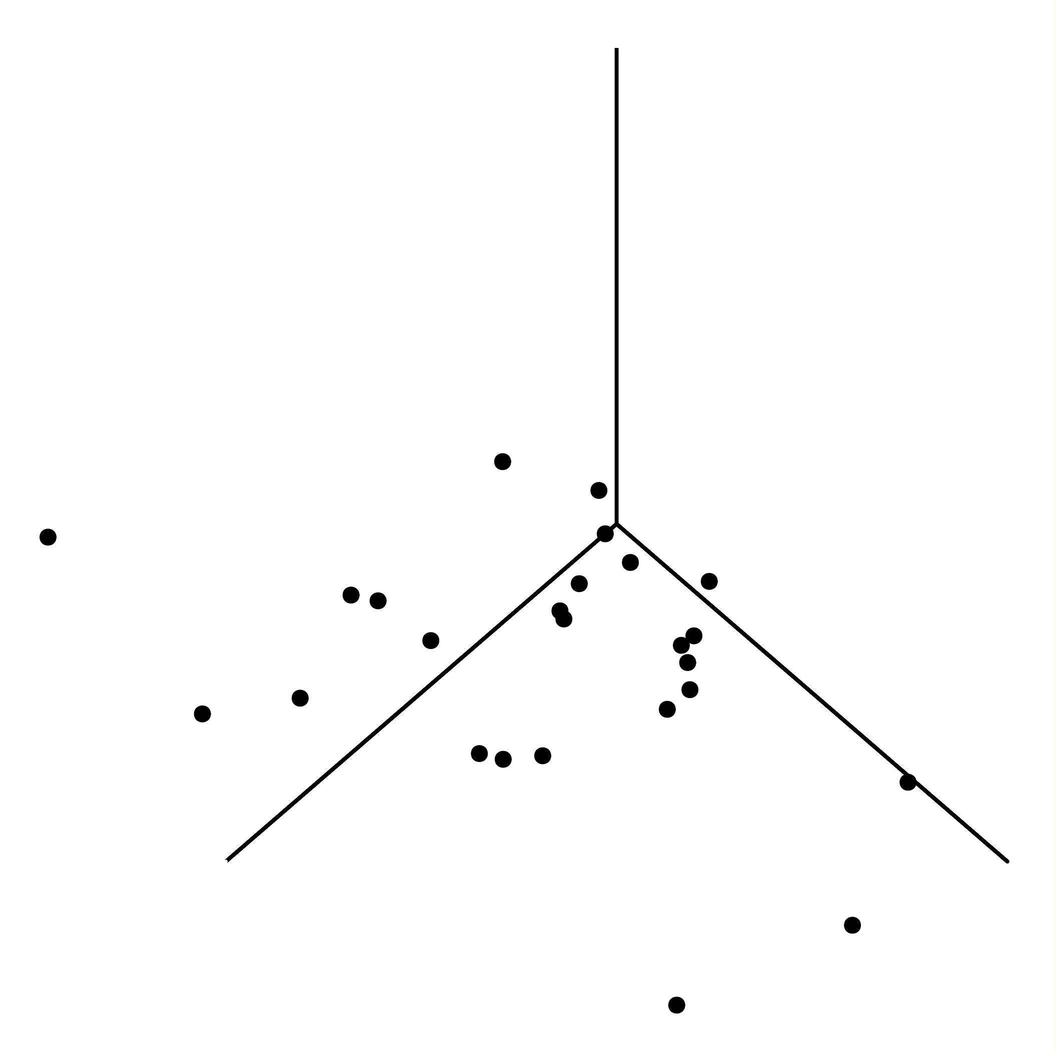
MLR Partial Residual Plots

Stat 230

April 29 2022

Overview



Today:

Partial Residual plots

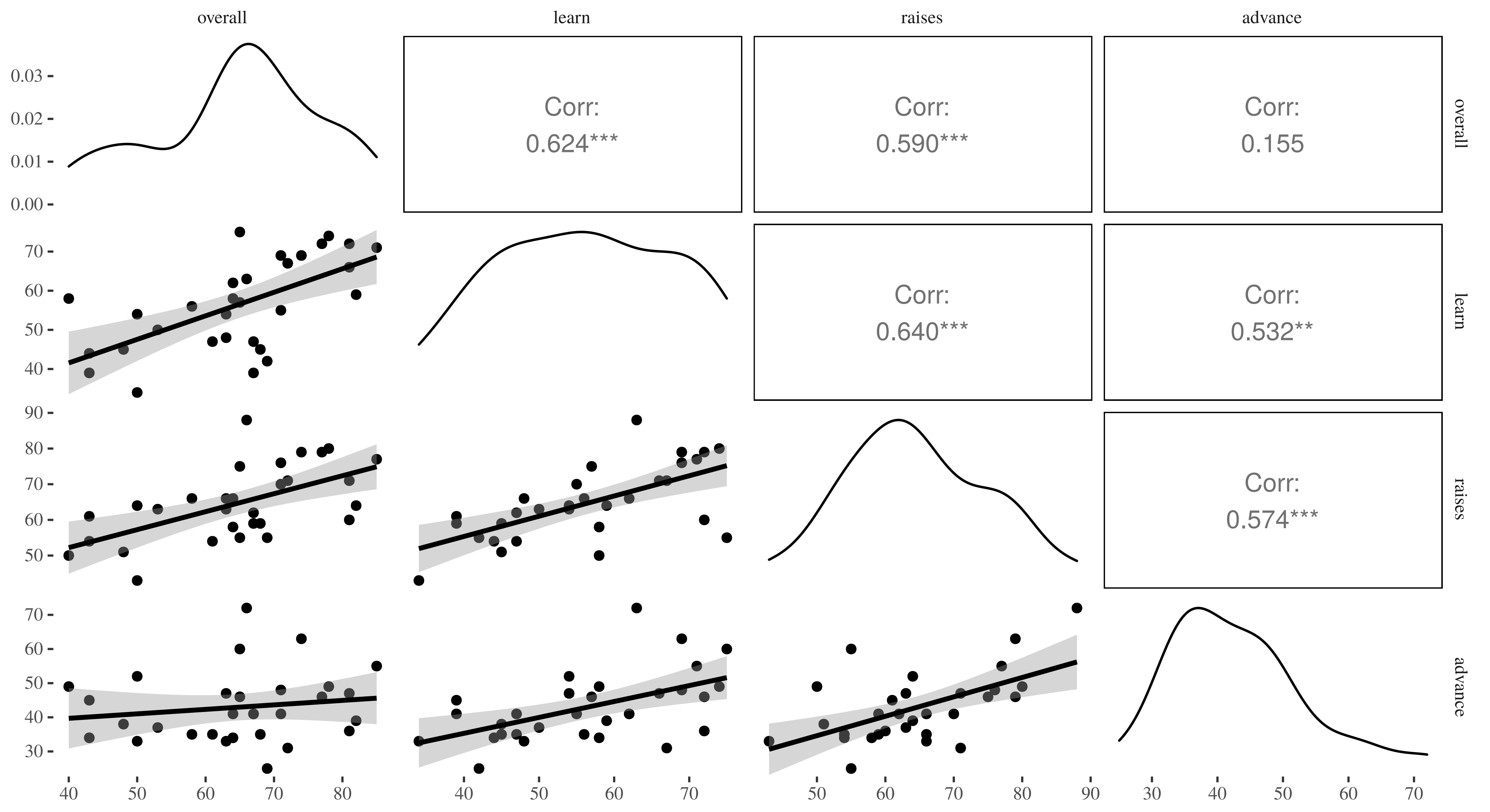
Supervisor dataset

- Employees in a large company were asked to rate their immediate supervisor
- **Response:** overall rating on a scale of 0 (bad) to 100 (good)
- Predictors from survey questions measured on an agreement scale ($0 =$ completely disagree to $100 =$ completely agree)
  - `raises`: "Your supervisor bases raises on performance."
  - `learn`: "Your supervisor provides opportunities to learn new things."
  - `advance`: "I am not satisfied with the rate I am advancing in the company."

Supervisor dataset

```
library(dplyr)
supervisor <- read.csv("https://raw.githubusercontent.com/deepbas/statdatasets/main/supervis
glimpse(supervisor)
```

```
Rows: 30
Columns: 7
$ overall    <int> 43, 63, 71, 61, 81, 43, 58, 71, 72, 67, 64, 67, 69, 68, 77,…
$ complaints <int> 51, 64, 70, 63, 78, 55, 67, 75, 82, 61, 53, 60, 62, 83, 77,…
$ privileges <int> 30, 51, 68, 45, 56, 49, 42, 50, 72, 45, 53, 47, 57, 83, 54,…
$ learn      <int> 39, 54, 69, 47, 66, 44, 56, 55, 67, 47, 58, 39, 42, 45, 72,…
$ raises     <int> 61, 63, 76, 54, 71, 54, 66, 70, 71, 62, 58, 59, 55, 59, 79,…
$ critical   <int> 92, 73, 86, 84, 83, 49, 68, 66, 83, 80, 67, 74, 63, 77, 77,…
$ advance    <int> 45, 47, 48, 35, 47, 34, 35, 41, 31, 41, 34, 41, 25, 35, 46,…
```


overall
learn
raises
advance
Corr:
0.624***
Corr:
0.590***
Corr:
0.155
Corr:
0.640***
Corr:
0.532**
Corr:
0.574***
0.00
0.01
0.02
0.03
40
50
60
70
80
90
30

Supervisor satisfaction

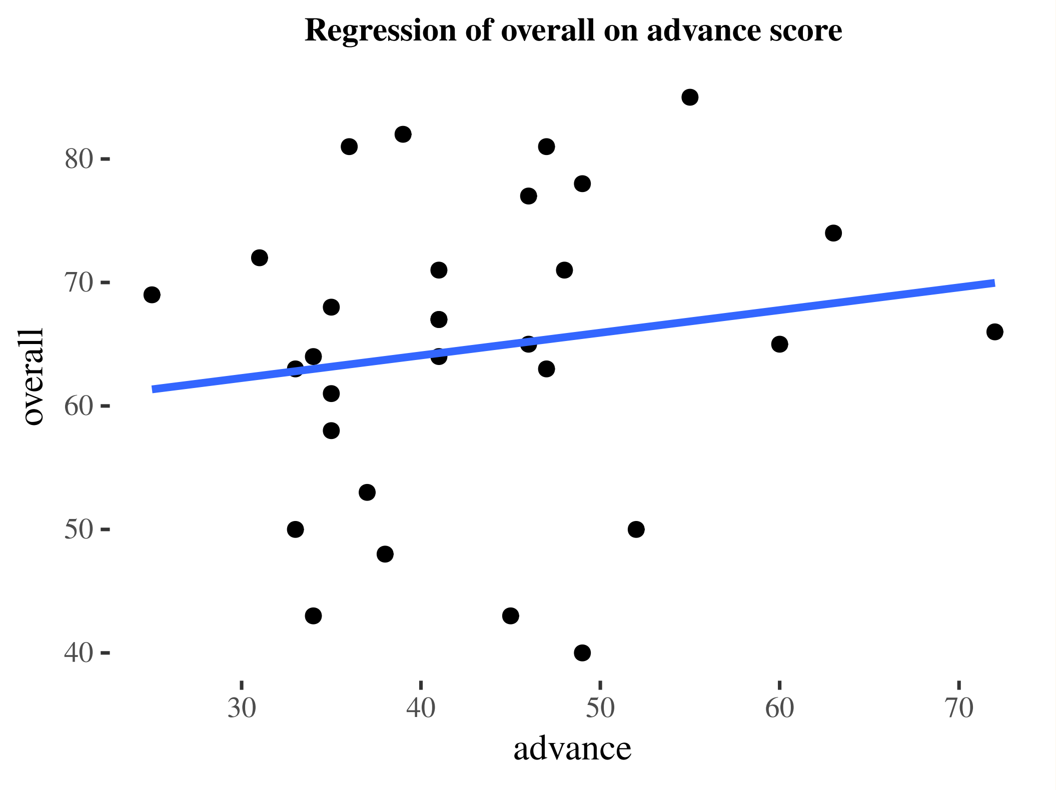
```
supervisor_lm <- lm(overall ~ learn + raises + advance, data = supervisor)
get_regression_table(supervisor_lm)
```

```
# A tibble: 4 × 7
  term      estimate std_error statistic p_value lower_ci upper_ci
  <chr>        <dbl>     <dbl>     <dbl>   <dbl>    <dbl>    <dbl>
1 intercept   17.7      10.2        1.73   0.096   -3.37    38.7
2 learn        0.548     0.184      2.99   0.006    0.171    0.926
3 raises       0.566     0.214      2.64   0.014    0.125    1.01
4 advance     -0.477     0.196     -2.43   0.022   -0.881   -0.074
```

- The estimated effect of `advance` is negative!

$$\hat{\mu}_{\text{overall}\,|x} = 17.7 + 0.548 \text{ learn } + 0.566 \text{ raises } - 0.477 \text{ advance}$$

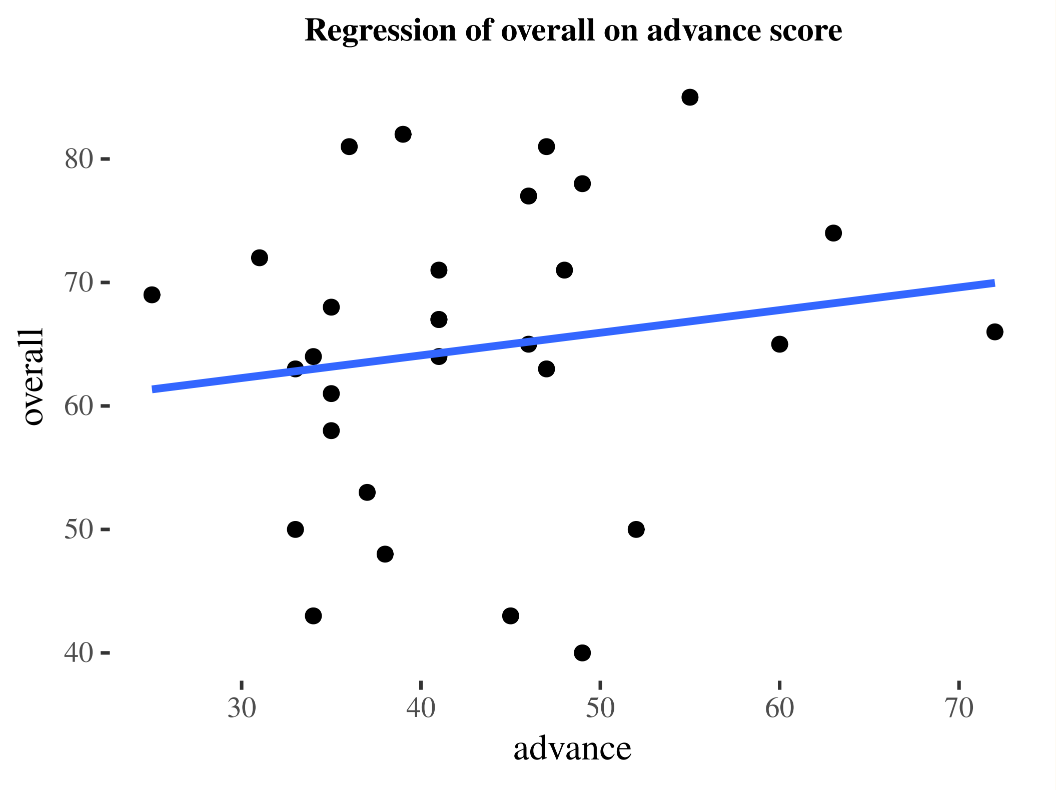
Supervisor satisfaction



Bivariate scatterplot of `overall` against `advance`:

- Across all `learn` and `raises` ratings, increases in the advance rating is associated with an increase in mean overall rating.

Supervisor satisfaction



MLR model:

- Holding `learn` and `raises` ratings fixed, a 10 point increase in the `advance` rating is associated with a decrease in the mean `overall` rating of 0.7 to 8.8 points.

Visualizing SLR

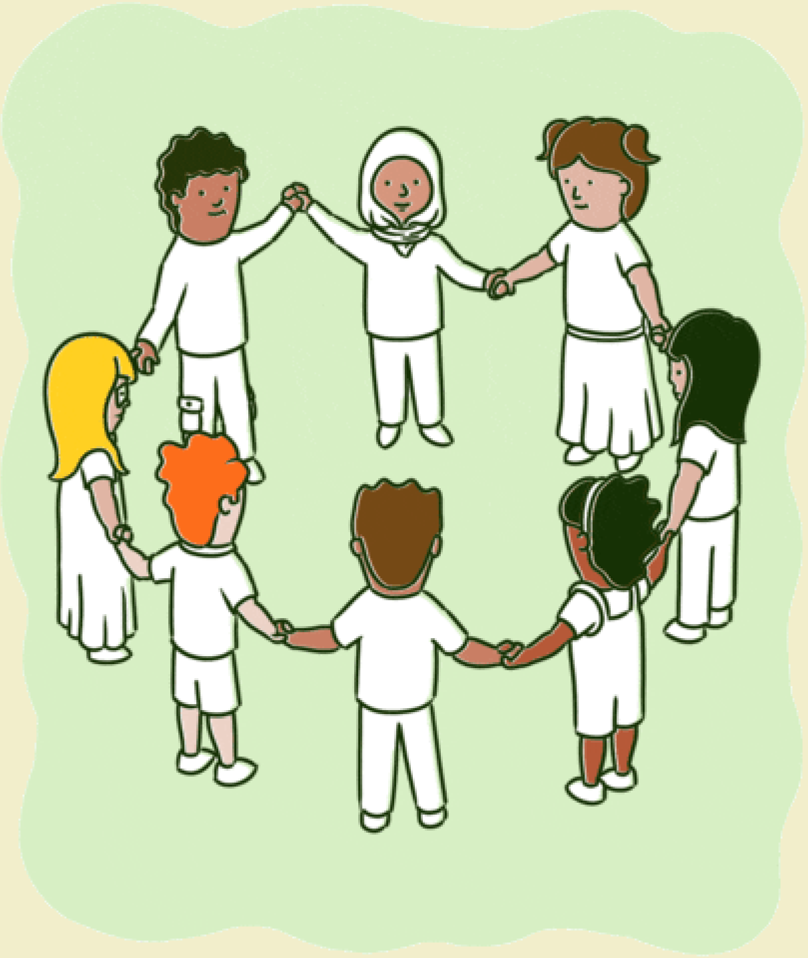
- The main question is how is the mean of Y associated with x ?

$$\mu_{y|x} = f\left(x, \beta' s\right) = ??$$

- A scatterplot of Y vs. x helps us see this relationship in an SLR model (which doesn't account for any other predictors).

Your Turn 1

05:00



- Go over to the in class activity file
- Complete the activity in your group

Visualizing MLR

- The main question is how is the mean of Y associated with, say, x_1 and x_2

$$\mu_{y|x_1,x_2} = f\left(x_1, x_2, \beta' s\right) = ??$$

- We need a 3-d graphic to see this relationship

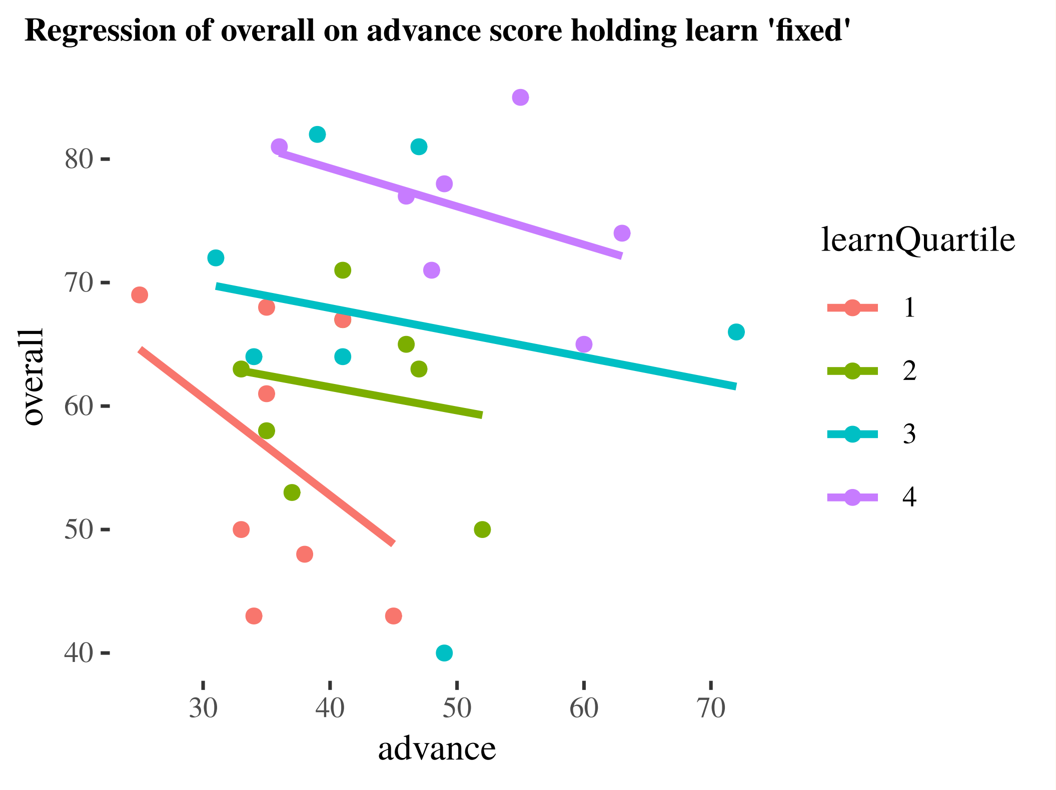
Visualizing MLR

A 2-d scatterplot of Y vs. x_1 doesn't hold x_2 fixed

- Scatterplot of `overall` and `advance` is positive
- because of this, a simple scatterplot doesn't necessary show the effect of x_1 in our MLR model
- The relationship between `overall` and `advance` is negative, holding `learn` and `raises` fixed.

Why do these differ?

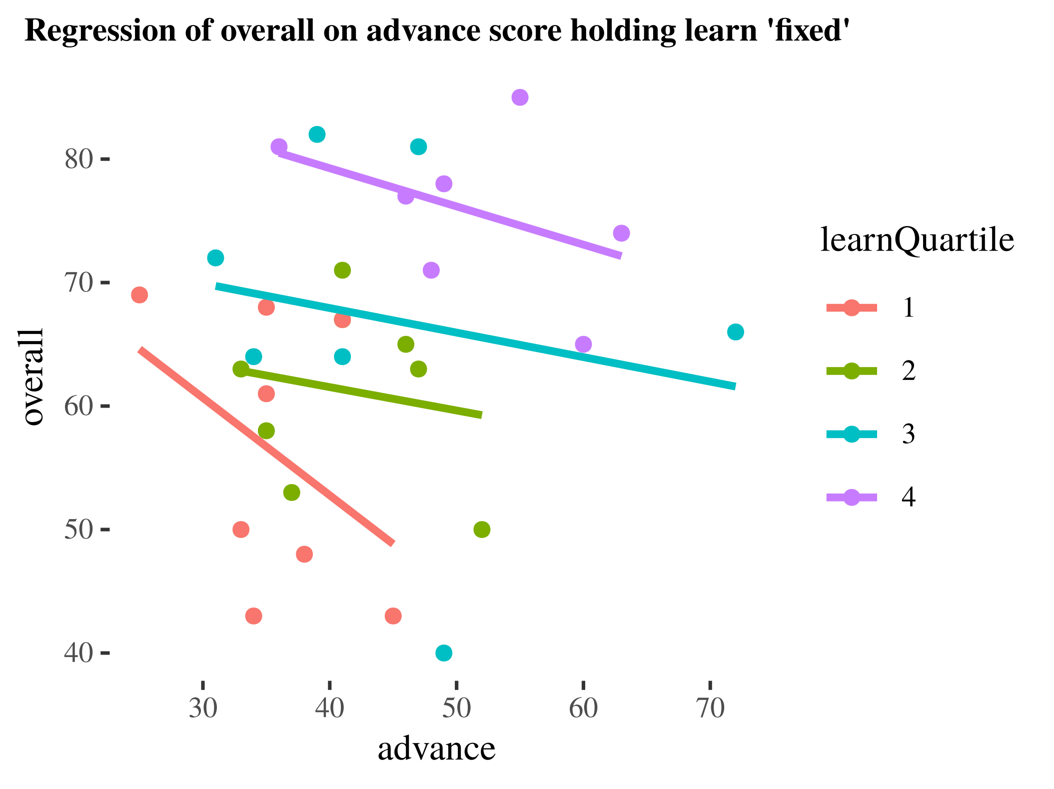
Scatterplot vs. MLR effect



For people with similar `learn` ratings (quartile)

- lower `advance` ratings are associated with higher overall rating
- this is the **negative** effect of `advance` in the MLR!

Scatterplot vs. MLR effect

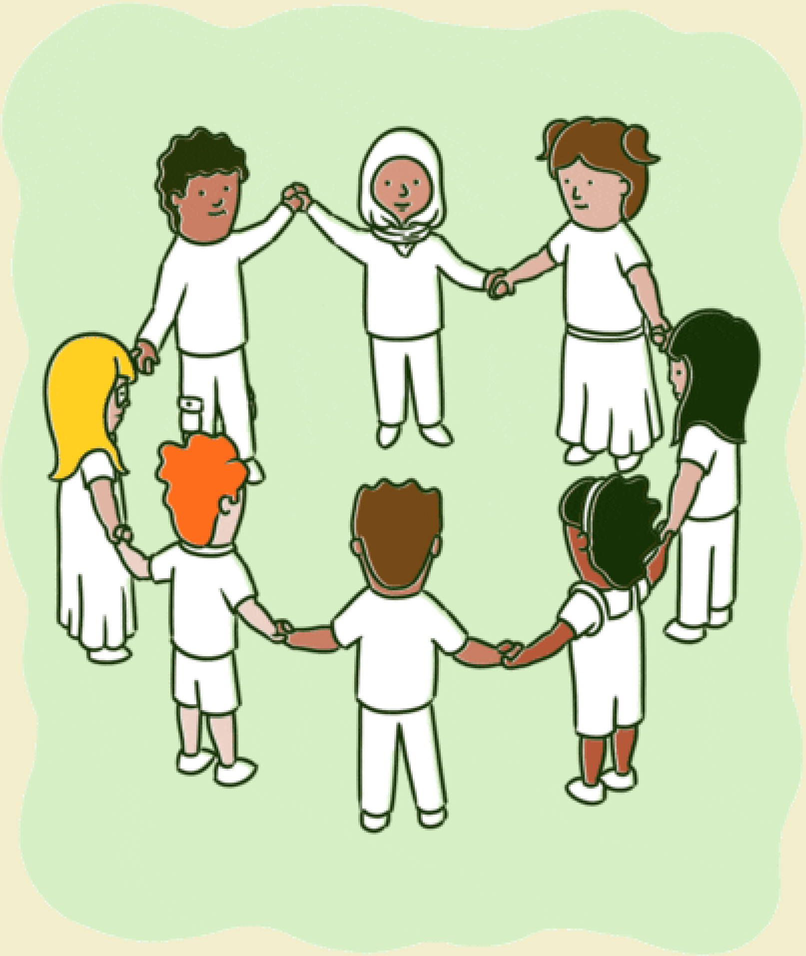


Why does `overall` against `advance` scatterplot show a positive trend?

- As `learn` increases so does `advance` (positively correlated predictors)
- As `learn` increases, it increases `overall` at a faster rate than the decrease in `overall` due to `advance`

Your Turn 2

05:00



- Go over to the in class activity file
- Complete the activity in your group

Visualizing MLR: Partial Residual plot

- The main question is how is the mean of Y associated with, say, x_1 and x_2

$$\mu_{y|x_1,x_2} = f\left(x_1, x_2, \beta' s\right) = ??$$

- One was to visualize the effect of x_1 on y in a MLR model is to plot x_1's partial residuals against x_1 :

$$\text{pres}_{\ 1,i} = y_i - \left(\hat{\beta}_0 + \hat{\beta}_2 x_{2,i}\right)$$

  where $\hat{\beta}'$ s come from the MLR of y on x_1 and x_2.

- Partial residuals "take away" the (linear) effect of x_2 from y to help us see the effect of x_1 after accounting for x_2

Visualizing MLR: Partial Residual plot

- Partial residuals for x_2 look like

$$\text{pres}_{2,i} = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i}\right)$$

- Generalizes to larger MLR models

Partial residual plot in R using `car` package

```
crp(my_lm)
```

- add `layout` to control rows/columns layout
- add `id = TRUE` to `ID` 2 extreme residual and 2 extreme predictor values
- add `smooth = FALSE` to remove the smoother

`crp` plots $pres_i - \overline{pres}$ on y-axis so that these residuals are centered around 0 and have the feel of the "usual" residuals.

`crp` means "component + residual plot" since

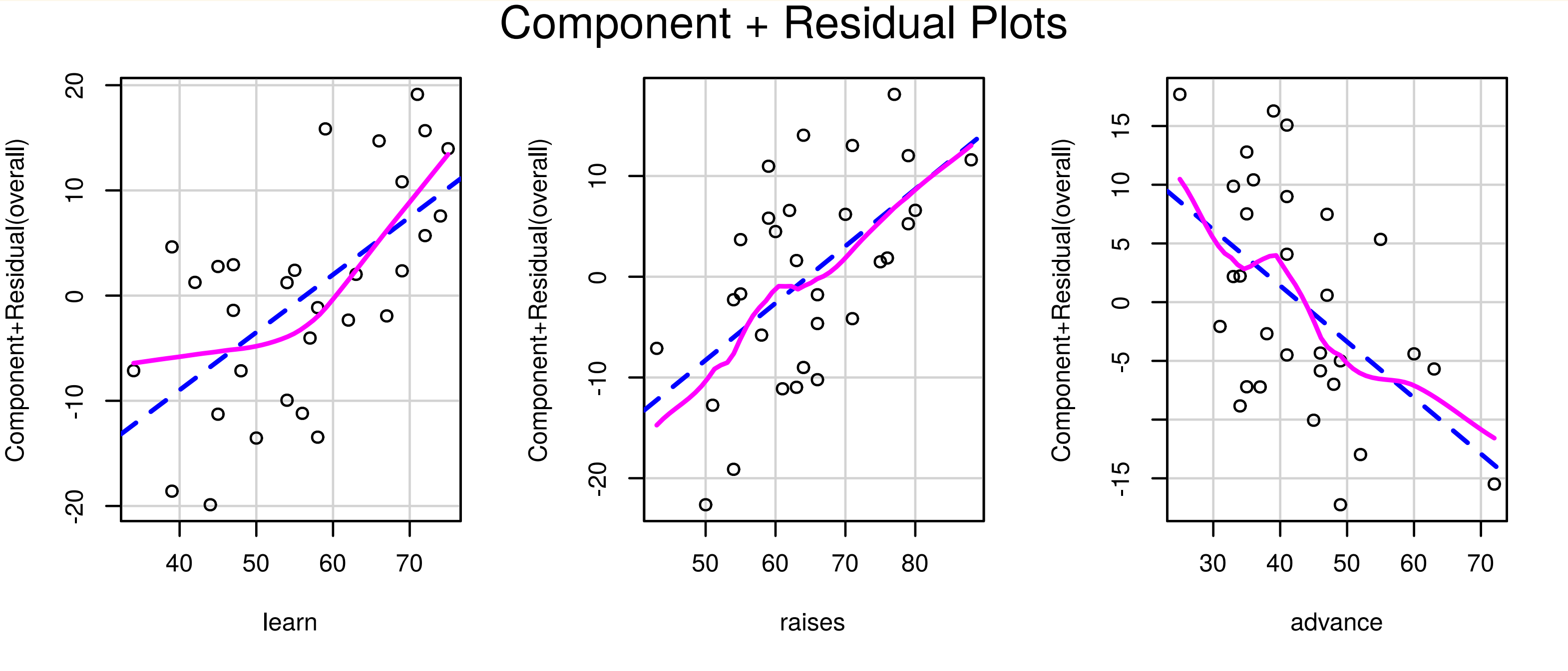
- residual: $r_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i}\right)$
- component (for x_1): $\hat{\beta}_1 x_{1,i}$

$$\text{pres}_{1,i} = y_i - \left(\hat{\beta}_0 + \hat{\beta}_2 x_{2,i}\right) = r_i + \hat{\beta}_1 x_1$$

Supervisor satisfaction

Now we see the effect of `advance` is negative after accounting for `learn` and `raises`

```r
library(car)
crp(supervisor_lm, layout = c(1,3))
```



Visualizing MLR: Partial Residual plot

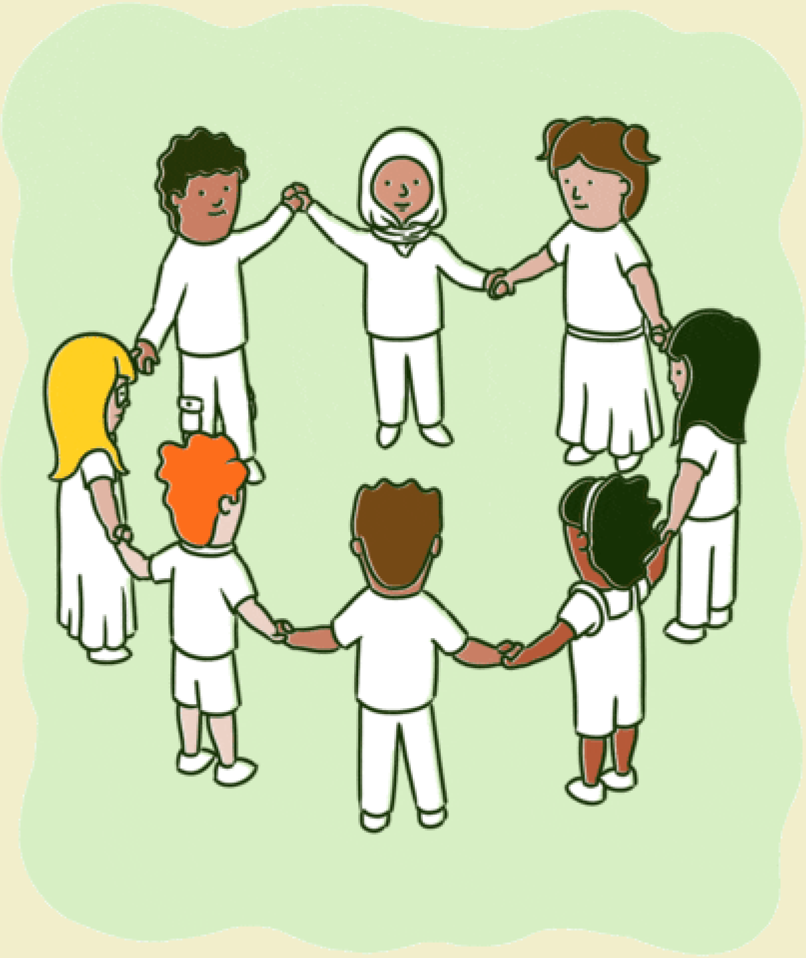
The slope of the least squares line in the partial residuals in this plot for x_j will equal its estimated (linear) effect $\hat{\beta}_j$ in the MLR

Using partial residual plots:

- We can see the "effect" of x_j after adjusting for other model terms.
- We can also see the variation in y that remains after adjusting for other model terms.
- We can look for outliers that could be affecting the estimated effect of x_j
- We can see if the effect of x_j is correctly modeled, signs of non-linearity and/or non-constant variance suggest we need to correct our model form.

Your Turn 3

05:00



- Go over to the rest of in class activity file
- Complete the activity in your group