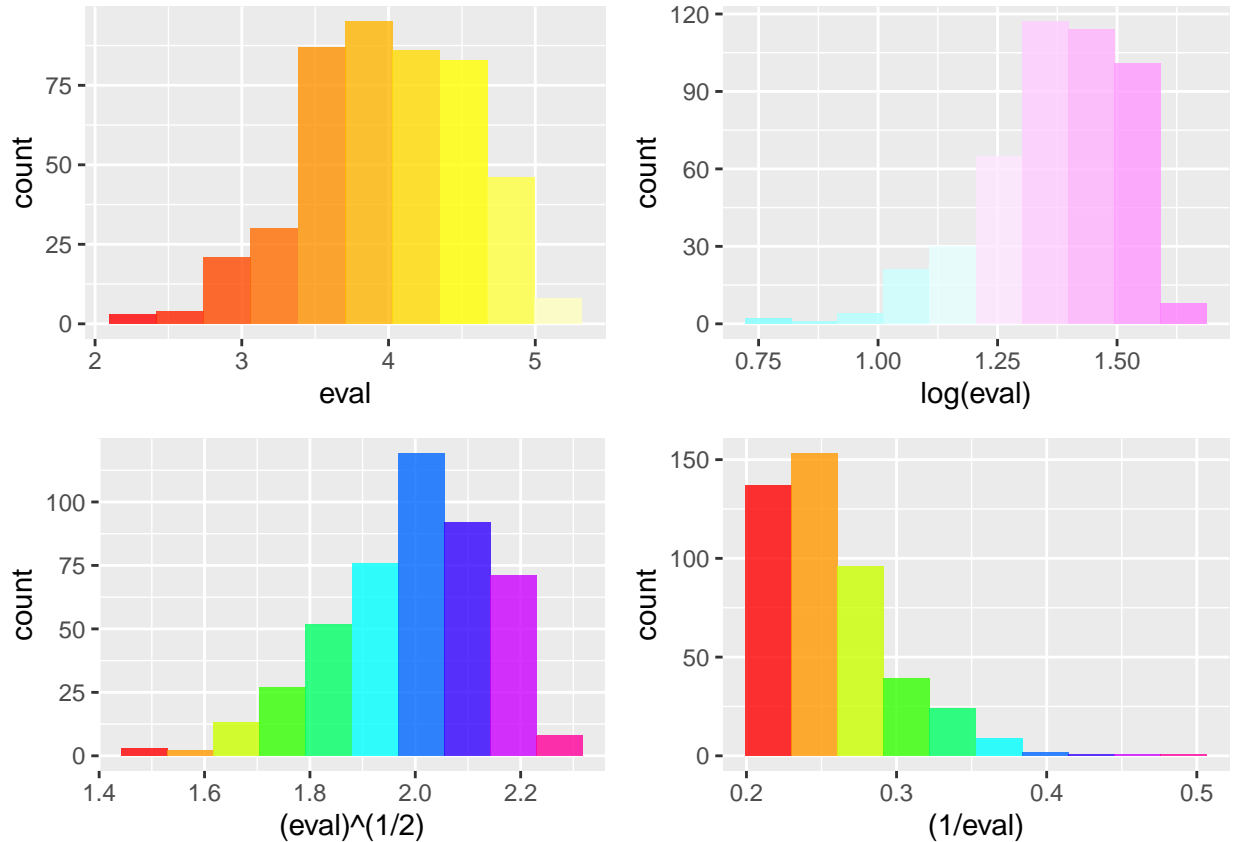


Lab 4: Multilevel Logistic Regression

Ashwini Marathe (asm105) & Srishti Saha (ss1078)

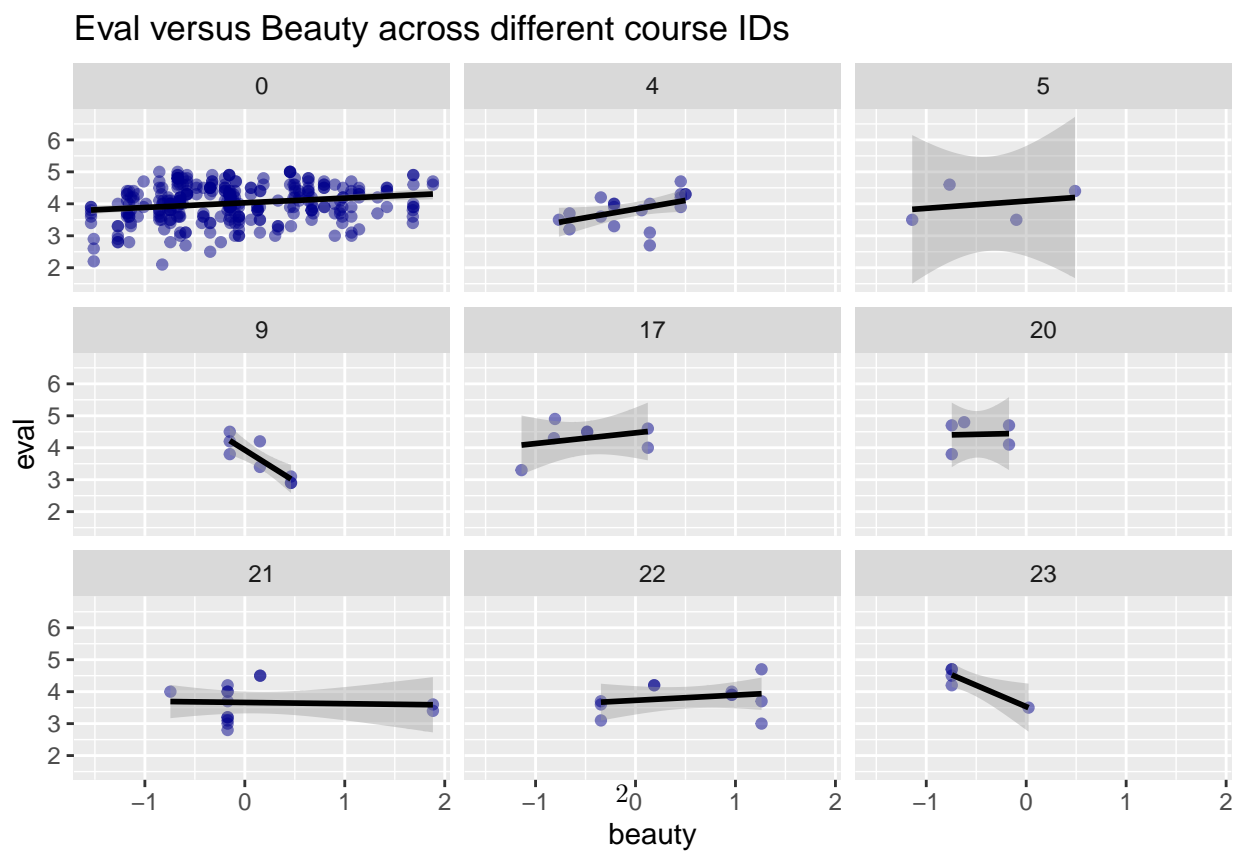
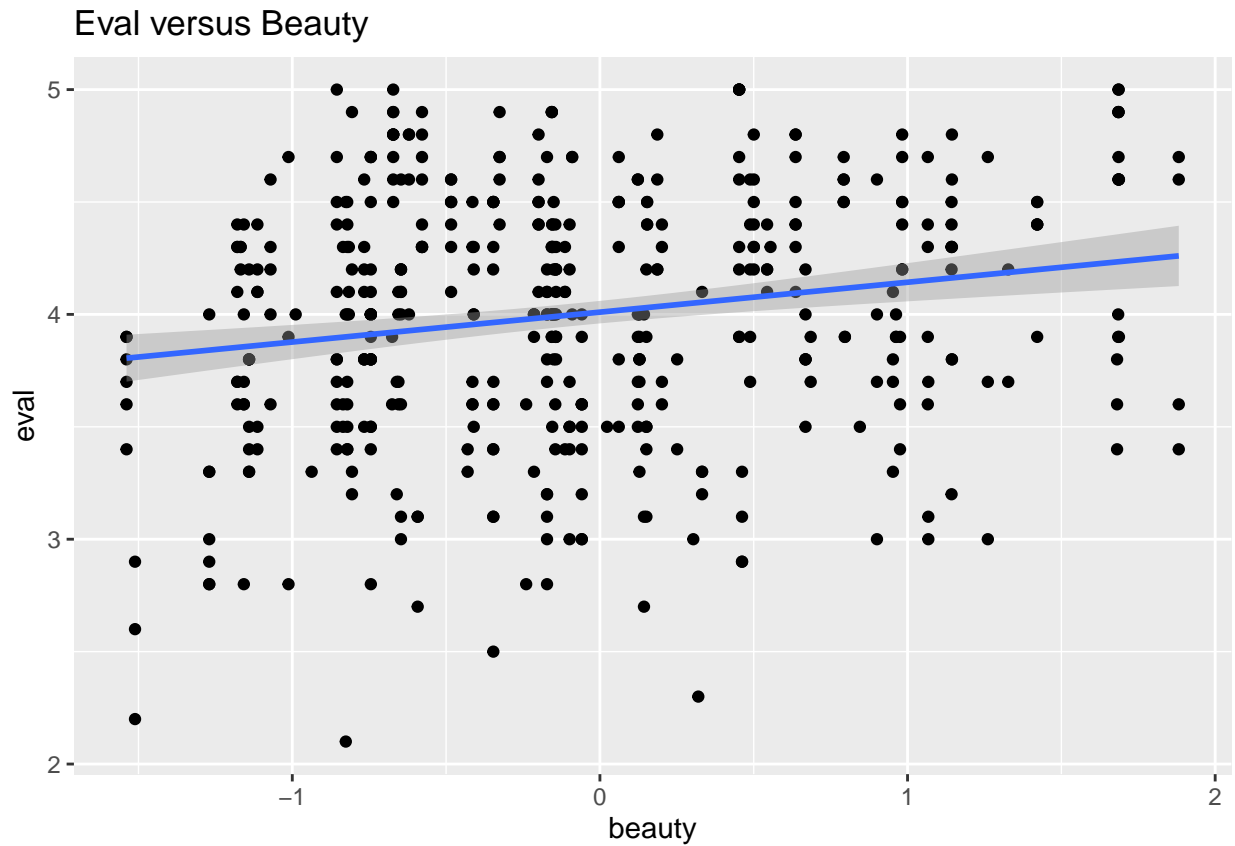
20 October, 2019

Exercise 1: Is the distribution of eval normal? If not, try the log transformation. Does that look more “normal”?



We tried 3 transformations on ‘eval’: log, square-root and inverse. Although ‘eval’ does not look normal, Log-transformation does not look any better. The distribution of ‘eval’ looks the most normal of the four options here. We will go ahead with ‘eval’.

Exercise 2: Describe the overall relationship between eval and beauty. Also examine the same relationship by CourseID. Are there any courses for which the relationship looks potentially different than others?



The relationship between eval and beauty seems positive and linear. Even if we split the data by course IDs and select courses with multiple data points, we see an upward trend between eval and beauty in most course IDs.

The trend of eval versus beauty is downward for courseID 9, 12 and 23. For others, it is either constant or upwards. This discrepancy might also be because of insufficient data in these courses.

The correlation of beauty and eval is 0.19 which means that they have a low-moderate correlation.

Exercise 3: Is it meaningful to fit a model that includes random slopes for beauty by profnumber? Why or why not?

It is not meaningful to fit a model that includes random slopes for beauty by profnumber. This is because beauty is a professor level variable and is constant for a given professor. Thus, it will add no information to the random slopes model at a professor level and will have a constant effect.

Exercise 4: Now, explore the relationship between eval and the other potential predictors, excluding profnumber, profevaluation, and CourseID. Don't include any of the plots, just briefly describe the most interesting relationships. We should not include profevaluation as a predictor for eval. Why?

Numerical variables: students, age, didevaluation, percentevaluating

- The correlation of eval with the following variables is **very low** (absolute value ≤ 0.05): students (linear trend-almost constant); age (linear trend- very slightly downward); didevaluation (linear trend-very slightly upward)
- The correlation of eval with these variables is **moderate to high** (abs value > 0.03): percentevaluating (linear trend-upward)

Factor variables: tenured, female, minority, formal, lower, multipleclass, nonenglish, onecredit, tenuretrack

- Factors that do not have a strong/significant impact: tenured, multiple class.
- If the instructor dressed formally in the picture used for beauty ratings, (formal=1), the median value of eval is slightly higher than non-formal. (however range is similar.)
- Female profs have lower median eval rating as compared to males
- Professor from minority group has less median evaluation as compared to non-minority
- Professors who received an undergraduate education from a non-English speaking country show a lower overall eval rating than the professors who had undergrad training in English speaking countries
- One-credit courses are rated higher overall as compared to the ones who aren't.
- Lower division courses have an overall better eval than the ones taking higher level courses
- Profs who did not work on tenure track show a higher overall eval rating than tenure track professors

We **should not include profevaluation as a predictor for eval** because of the high correlation between them. The correlation coefficient of profevaluation and eval is **0.94**. Furthermore, by definition profevaluation is the instructor rating and eval is course rating. So it is obvious that if the professor was evaluated high, students might have liked the subject as well.

Exercise 5: Fit a varying-intercept model for these data by profnumber with beauty as the only predictor. Interpret the results in the context of the question.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: eval ~ beauty + (1 | profnumber)
## Data: Beauty
##
## REML criterion at convergence: 643.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.6897 -0.6200  0.0688  0.5724  2.4529
##
## Random effects:
## Groups      Name                Variance Std.Dev.
## profnumber (Intercept) 0.1387    0.3724
## Residual              0.1705    0.4129
## Number of obs: 463, groups:  profnumber, 94
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  3.93893    0.04420 84.83439  89.125   <2e-16 ***
## beauty       0.11566    0.05387 86.85890   2.147   0.0346 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## beauty 0.022
```

- Interpreting fixed effects: We have an overall “average” regression line for all courses across all professors which has slope of 0.116 and intercept 3.94. This means just for the fixed effects, the model looks like= $eval_i = 3.94 + 0.12 * beauty_i$. Thus, for any course, the baseline eval (course evaluation) value is 3.93.
 - Then, for any course for profnumber 1, every unit increase in the beauty score increases the eval (course evaluation) value by 0.12 (given all other factors are constant).
- However, we also need to look at the random effects to see how this trend differs for every professor. For example, for profnumber=2, the intercept is -0.21. Thus, estimated regression line for profnumber 2 becomes: $eval_i = (3.94 - 0.21) + 0.12 * beauty_i = 3.73 + 0.12 * beauty_i$
- Interpreting random effects: The estimated standard error $\hat{\sigma} = 0.41$ describes the unexplained variation in average ratings across professors. The estimated $\hat{\tau}_0 = 0.37$ describes the variation in ratings across the different course evaluations for the same instructor.

Exercise 6: Now identify three instructor-level predictors excluding beauty (with profnumber being the instructor-level identifier) that you think we should control for based on your EDA. Fit a varying-intercept model for these data by profnumber and include beauty plus the other three variables you have identified as predictors. Interpret the results in the context of the question.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
```

```

## Formula:
## eval ~ beauty + as.factor(nonenglish) + as.factor(female) + as.factor(tenuretrack) +
##      (1 | profnumber)
##      Data: Beauty
##
## REML criterion at convergence: 636.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.7663 -0.6090  0.0854   0.5502   2.4437
##
## Random effects:
##      Groups      Name      Variance Std.Dev.
##  profnumber (Intercept) 0.1207   0.3474
##      Residual          0.1698   0.4121
## Number of obs: 463, groups:  profnumber, 94
##
## Fixed effects:
##
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)      4.24463    0.10933 71.45689   38.826 < 2e-16 ***
## beauty            0.13681    0.05187 84.91999    2.638 0.00993 **
## as.factor(nonenglish)1 -0.31512    0.16542 91.13033   -1.905 0.05994 .
## as.factor(female)1    -0.22306    0.08646 82.95211   -2.580 0.01165 *
## as.factor(tenuretrack)1 -0.22690    0.11094 72.67664   -2.045 0.04445 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) beauty as.fctr(n)1 as.fctr(f)1
## beauty            0.093
## as.fctr(n)1    0.007  0.037
## as.fctr(f)1   -0.420 -0.168 -0.014
## as.fctr(t)1   -0.855 -0.037 -0.129      0.105

```

We included 3 instructor level predictors: nonenglish, female and tenuretrack as they showed some trend and correlation during the EDA.

- All three predictors were significant in their fixed effects (as they had p-value ≤ 0.05). The most significant predictor was beauty based on its highest absolute t-value.
- Interpreting fixed effects: We have an overall “average” regression line for all courses across all professors which has slope of 0.13 (for beauty), -0.31 (for nonenglish), -0.22 (for female) and -0.23 (for tenuretrack) and intercept 4.24. This means just for the fixed effects, the model looks like $eval_i = 4.24 + 0.13 * beauty_i - 0.31 * nonenglish_i - 0.22 * female_i - 0.23 * tenuretrack_i$. Thus, for any course, the baseline eval (course evaluation) value given it is taught by a male professor trained in english speaking country and working on a non-tenure track (nonenglish=0, female=0, tenuretrack=0), is 4.24.
- Then, for any course for profnumber 1, every unit increase in the beauty score increases the eval (course evaluation) value by 0.13 (given all other factors are constant).
- Then, for any course for profnumber 1, if it is an instructor who studied in a non-english speaking country, the eval score is reduced by -0.31 (given all other factors are constant).
- Then, for any course for profnumber 1, if it is a female instructor, the eval score is reduced by -0.22 (given all other factors are constant).
- Then, for any course for profnumber 1, if it is a non-tenuretrack instructor, the eval score is reduced by -0.23 (given all other factors are constant).

- However, we also need to look at the random effects to see how this trend differs for every professor. For example, for profnumber=2, the intercept is -0.25. Thus, estimated regression line for profnumber 2 becomes: $eval_i = (4.24 - 0.25) + 0.13 * beauty_i - 0.31 * nonenglish_i - 0.22 * female_i - 0.23 * tenuretrack_i$
- Interpreting random effects: The estimated standard error $\hat{\sigma} = 0.41$ describes the unexplained variation in average ratings across professors. The estimated $\hat{\tau}_0 = 0.35$ describes the variation in ratings across the different course evaluations for the same instructor.

Exercise 7: Using the model from question 5, how does the variation in average ratings across professors compare to the variation in ratings across the different course evaluations for the same instructor?

In model in question 5, the two values of estimated standard error and deviation: (0.37, 0.41) are fairly similar. Thus, the variation in average ratings across professors is very similar to the variation in ratings across the different course evaluations for the same instructor. This means that we are considering both the in-group information and the overall information in our model.

Exercise 8: Extend the model from question 6 by also allowing the intercept and coefficient for beauty to vary by CourseID. Fit the model in R and interpret the results: the coefficient estimates and the estimated standard deviations. Did any of the results for the fixed effects change? If yes, why do you think that is?

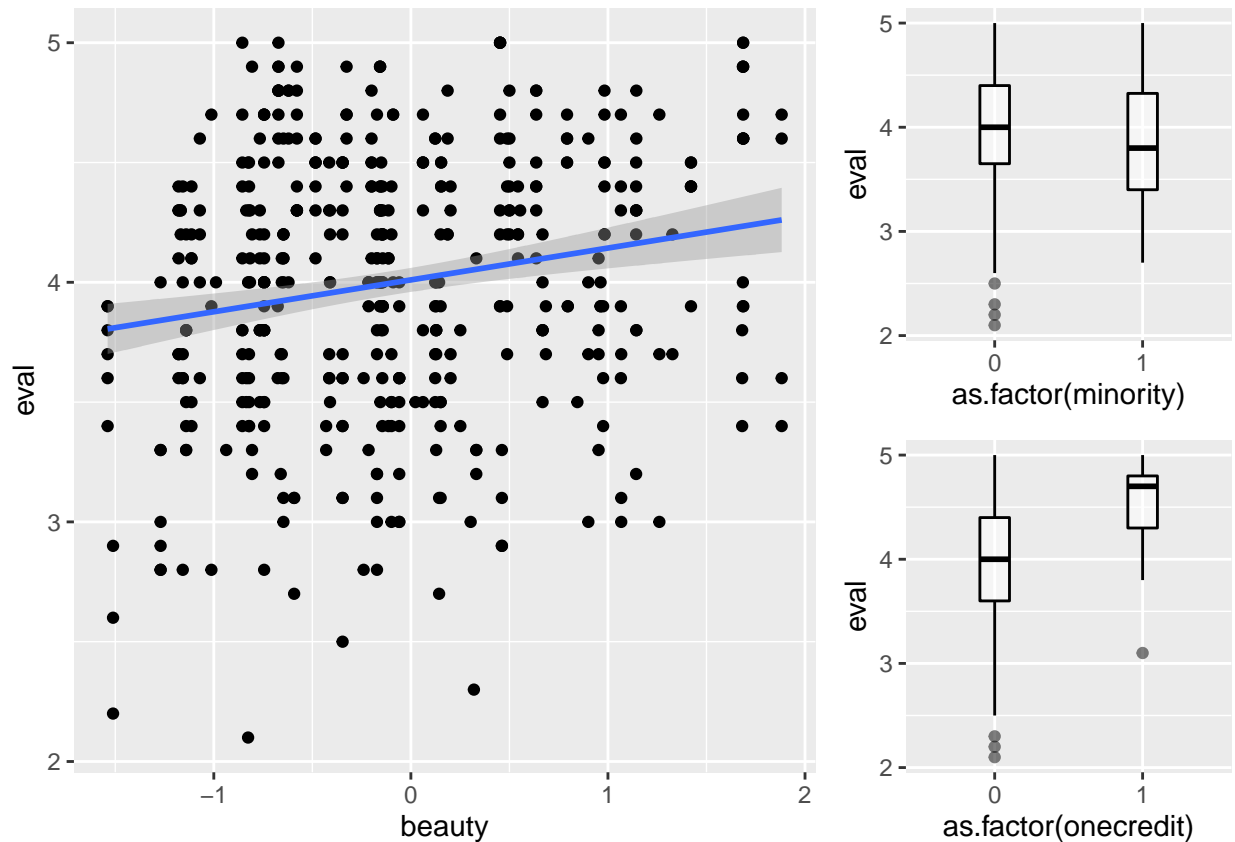
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## eval ~ beauty + as.factor(nonenglish) + as.factor(female) + as.factor(tenuretrack) +
## (1 | profnumber) + (beauty | courseID)
## Data: Beauty
##
## REML criterion at convergence: 622.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.9599 -0.6028  0.0664  0.5628  2.4661
##
## Random effects:
## Groups      Name                Variance Std.Dev. Corr
## profnumber (Intercept) 0.117023 0.34209
## courseID    (Intercept) 0.069497 0.26362
##              beauty      0.003582 0.05985  0.01
## Residual              0.152904 0.39103
## Number of obs: 463, groups:  profnumber, 94; courseID, 31
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)      4.23904    0.12382 87.85081  34.236 < 2e-16 ***
## beauty            0.12048    0.06274  6.46862   1.920  0.09972 .
## as.factor(nonenglish)1 -0.30766    0.16519 93.72754  -1.862  0.06568 .
## as.factor(female)1    -0.23108    0.08564 84.92313  -2.698  0.00841 **
## as.factor(tenuretrack)1 -0.21201    0.10988 74.25688  -1.929  0.05750 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Correlation of Fixed Effects:
##           (Intr) beauty as.fctr(n)1 as.fctr(f)1
## beauty           0.072
## as.fctr(n)1 -0.010  0.026
## as.fctr(f)1 -0.355 -0.127 -0.021
## as.fctr(t)1 -0.752 -0.006 -0.124          0.095
```

Yes, in this model (as compared to model in Q6) the intercept, coefficients and the significance levels have changed. This is because some of the variance that was captured by these predictors is now accounted for after using courseID as a class level identifier. Also, a fraction of the variation in eval is now also explained by the random slope component.

- Interpretation of fixed effects:
 - The intercept estimate dropped marginally from 4.245 to 4.239. The coefficients for beauty, nonenglish, female and tenuretrack have also changed marginally which basically implies that every level (/unit) change in these predictors will have a marginally different effect on the eval score.
 - The significance of beauty has dropped (t-value from 2.64 to 1.92) and now the female predictor is the most significant (abs t-value 2.70)
 - This happened because every course will now contribute to the varying intercept or the baseline eval score which will be also affected by other covariates to get to the actual evaluation score.
- Interpreting random effects:
 - The random effects now include variation by beauty in groups defined by courseID.
 - The estimated standard error 0.3910 describes the unexplained within-instructor and within-course variation.
 - The estimated deviation 0.3421 describes the within-instructor variation attributed to the random intercept.
 - The estimated deviation 0.2636 describes the within-course variation attributed to the random intercept.
 - The estimated deviation 0.0599 describes the within-course variation attributed to the random slope due to the beauty predictor.

Exercise 9: Using two or three (at most!) plots, tell a visual story about your most interesting or surprising findings. If you need to present results by profnumber or CourseID, you should consider either looking at a random sample of them or focusing on the groups that are the most different, since there are so many of them.



- Beauty becomes insignificant in model 8 (high p-value of 0.09). This is contradictory to what we observed in our EDA.
- Minority seemed significant in the EDA (higher eval for minority=0 (non-minor groups)), but was insignificant (based on a high p-value) during model-selection
- There is a huge disparity between the eval scores of one-credit and non-one credit courses. This might be because one-credit courses might have lighter coursework and hence are more likeable by students.

Exercise 10: Identify one class-level predictor (with CourseID being the class-level identifier) that you think we should control for in the model. Include it as a fixed predictor in the model from Question 8. Is the variable significant? If yes, interpret the coefficient.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## eval ~ beauty + as.factor(nonenglish) + as.factor(female) + as.factor(tenuretrack) +
##       as.factor(onecredit) + (1 | profnumber) + (beauty | courseID)
## Data: Beauty
```



```

##
## REML criterion at convergence: 618.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.9561 -0.5982  0.0840  0.5809  2.4703
##
## Random effects:
##   Groups      Name      Variance Std.Dev. Corr
##   profnumber (Intercept) 0.109642 0.3311
##   courseID    (Intercept) 0.055465 0.2355
##             beauty      0.003956 0.0629  0.02
##   Residual              0.153620 0.3919
## Number of obs: 463, groups:  profnumber, 94; courseID, 31
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      4.17187    0.12197  87.64415  34.204 < 2e-16
## beauty           0.12203    0.06212   6.37994   1.964  0.09425
## as.factor(nonenglish)1 -0.31581    0.16121  91.28202  -1.959  0.05317
## as.factor(female)1    -0.22326    0.08357  82.52677  -2.671  0.00910
## as.factor(tenuretrack)1 -0.15642    0.10911  75.99023  -1.434  0.15578
## as.factor(onecredit)1  0.34823    0.13028 355.09882   2.673  0.00787
##
## (Intercept)          ***
## beauty                .
## as.factor(nonenglish)1 .
## as.factor(female)1    **
## as.factor(tenuretrack)1
## as.factor(onecredit)1 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) beauty as.fctr(nn)1 as.fctr(f)1 as.fctr(t)1
## beauty          0.064
## as.fctr(nn)1   -0.004  0.025
## as.fctr(f)1    -0.357 -0.123 -0.022
## as.fctr(t)1    -0.770  0.001 -0.126      0.100
## as.fctr(nc)1   -0.208  0.027 -0.021      0.030      0.195

```

We included onecredit as the class level predictor. The variable is significant according to its low p-value.

- Fixed Effects:
 - The baseline (intercept) changes to 4.17 (from 4.23). This means that the baseline eval score, given that it is taught by male professor trained in english-speaking country who is not a tenure track instructor and the course is a non-one credit course is 4.17
 - The estimate is 0.35. This means that for a one credit course, the eval score increase by 0.35 (given all other factors are constant).
- Random Effects:
 - The estimated standard error 0.3919 describes the unexplained within-instructor and within-course variation.

- The estimated 0.3311 describes the within-instructor variation attributed to the random intercept.
- The estimated 0.2355 describes the within-course variation attributed to the random intercept.