

Logistic Regression Analysis

New Analysis with the change of reference variable (Going from worst to best)

Category: NKNK=0, KBNKW=1, KNK=2, NKK=3, KK=4, Others=5 (New)

Expected Grades: D=0, C=1, B=2, A=3

Actual Grade: D=0, C=1, B=2, A=3

Female=0, Male=1

Spring=0, Fall=1

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.688	-0.391	-0.055	0.322	3.066

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-27.3883	1136.2768	-0.02	0.98077
Grade	0.1103	0.0163	6.78	0.000000000012 ***
Gender1	-0.9607	0.3895	-2.47	0.01363 *
Semester1	1.4995	0.4004	3.75	0.00018 ***
Category1	3.7474	0.7615	4.92	0.000000861036 ***
Category2	3.6672	0.6325	5.80	0.000000006706 ***
Category3	3.7298	0.7926	4.71	0.000002528960 ***
Category4	0.2644	0.5875	0.45	0.65275
Category5	3.9607	0.8916	4.44	0.000008897424 ***
Required1	1.4233	0.8378	1.70	0.08932 .
FirstTime1	0.2729	0.5031	0.54	0.58753
Preparation2	1.0376	2.0585	0.50	0.61423
Preparation3	-0.6124	1.6388	-0.37	0.70865
Preparation4	-0.0139	1.6142	-0.01	0.99311
Preparation5	-0.2144	1.6524	-0.13	0.89679
PerceivedDiffLevel3	15.9400	1136.2740	0.01	0.98881
PerceivedDiffLevel4	16.0792	1136.2738	0.01	0.98871
PerceivedDiffLevel5	15.7939	1136.2739	0.01	0.98891
ExpGrade1	-0.6006	0.9381	-0.64	0.52202
ExpGrade2	-0.5809	1.0002	-0.58	0.56138
TakeUpDiv1	-0.4476	0.4456	-1.00	0.31519
ActualCourseGrade1	-0.0884	0.7651	-0.12	0.90806
ActualCourseGrade2	0.1122	0.8272	0.14	0.89208
ActualCourseGrade3	0.2117	0.7740	0.27	0.78449

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 539.70 on 398 degrees of freedom
Residual deviance: 234.57 on 375 degrees of freedom

AIC: 282.6

Number of Fisher Scoring iterations: 15

Interpretation:

Since the p-values of the coefficients for Grade, Gender, Semester and Category, they have significant association with the grade estimate behavior of the students. The other predictor variables do not have statistically significant association with the students' grade estimation behavior since the p-values of their coefficient are more than 0.05. Our response/dependent variable is behavior, which is a binomial variable. We have set its value to 0 for overestimate and 1 for underestimate of the actual grades. We used the following logistic function to model the Probability($y=1$ given X) or simply $p(X)$. Its value ranges from 0 to 1.

$$P(x) = \frac{e^{\beta_0 + \beta_1 \cdot X}}{1 + e^{\beta_0 + \beta_1 \cdot X}}$$

We can rewrite this equation as

$$\frac{P(X)}{1 - P(X)} = e^{\beta_0 + \beta_1 X}$$

Taking ln both sides, we get

$$\ln\left(\frac{P(X)}{1 - P(X)}\right) = \beta_0 + \beta_1 \cdot X$$

We will call the left side of this equation as log odds.

The explanatory variable X is a vector with values $(Grade, Gender_i, Semester_j, Category_k)$, i, j, k are multi indices with values $i = j = (0,1), k = (0,1,2,3)$. The Grade is a continuous variable, and the rest of the variables are categorical variable. We used the categorical variables with 0 index as the reference variable and the variable with other indices as dummy variables. The coefficient vector β_1 is also a multi index variable.

Being male reduces the log odds by 0.9607. That means that there is a negative effect of Male Gender with under estimate. Likewise, a unit increase in the actual grade increases the log odds by 0.110. This means there is a positive effect of grades on the under estimation. In fall semester the log odds increases by 1.4995. Being a category 1 increases the log odds by 3.7474, being category 2 increases the log odds by 3.6672 and being the category 3 increases the log odds by 3.7298. the coefficient for category 4 is not statistically significant, and being category 5 increased the log odds by 3.9607. Since the p-value of the Grade is the lowest, this suggest that there is the strongest association of the student's actual grade with the students' grade estimation behavior.

Odds Ratios

These represent the change in the probability when the X variable changes from 0 to 1.

	odds ratios
(Intercept)	1.27E-12
Grade	1.116613004
Gender1	0.382624955
Semester1	4.479448786
Category1	42.41067078
Category2	39.14215425
Category3	41.67077318
Category4	1.302649152
Category5	52.49405893
Required1	4.150795489
FirstTime1	1.313768861
Preparation2	2.822435035
Preparation3	0.542048391
Preparation4	0.986196159
Preparation5	0.80702551
PerceivedDiffLevel3	8368623.73
PerceivedDiffLevel4	9618510.721
PerceivedDiffLevel5	7231087.516
ExpGrade1	0.548482448
ExpGrade2	0.559394685
TakeUpDiv1	0.639160297
ActualCourseGrade1	0.915394646
ActualCourseGrade2	1.118736586
ActualCourseGrade3	1.235777096

Interpretation:

One unit increase in the grade increases the odds of being an under estimator (versus being an over estimator) increases by a factor of 1.1166, and so on.

Next we only consider the variable with p value less than 0.05.

Semester, Grade, Gender, Category and Required after removing category 4.

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.921	-0.309	-0.063	0.248	3.010

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-10.8883	1.6901	-6.44	0.00000000012	***
Semester1	2.1851	0.5290	4.13	0.00003610414	***
Grade	0.0993	0.0161	6.18	0.00000000066	***
Gender1	-0.7079	0.4622	-1.53	0.13	
Category1	3.5495	0.7364	4.82	0.00000143652	***
Category2	3.7129	0.6424	5.78	0.00000000748	***
Category3	4.0183	0.8132	4.94	0.00000077505	***
Category5	4.1782	0.9134	4.57	0.00000478161	***
Required1	0.2124	0.9547	0.22	0.82	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 420.45 on 313 degrees of freedom
Residual deviance: 151.45 on 305 degrees of freedom
AIC: 169.4

Number of Fisher Scoring iterations: 7

Odds Ratios:

(Intercept)	Semester1	Grade	Gender1	Category1	Category2	Category3	Category5	Required1
0.000019	8.891966	1.104377	0.492675	34.795740	40.974432	55.605235	65.246913	1.236596

Interpretation:

When the semester changed from spring to fall, the probability of underestimation increased 8.89 times, when the grade increases by one unit, the probability of underestimation increases 1.10 times, when

the gender changed from female to male, the probability became almost half.