Bioimplants Case

Ali Ladha

2024-11-15

Q1

```
library(tidyverse)
library(janitor)


data <- read_csv('/Users/aliladha/Documents/Files/School Work/College/Graduate/Completed Courses/IS 6489-001 Stats & Pred Analytics/bioimplants.csv')

glimpse(data)
```

```
## Rows: 1,470
## Columns: 29
## $ age                         <dbl> 41, 49, 37, 33, 27, 32, 59, 30, 38, 36, 35,…
## $ business_travel             <chr> "Travel_Rarely", "Travel_Frequently", "Trav…
## $ department                  <chr> "Sales", "Research & Development", "Researc…
## $ distance_from_home          <dbl> 1, 8, 2, 3, 2, 2, 3, 24, 23, 27, 16, 15, 26…
## $ education                   <dbl> 2, 1, 2, 4, 1, 2, 3, 1, 3, 3, 3, 2, 1, 2, 3…
## $ education_field             <chr> "Life Sciences", "Life Sciences", "Other", …
## $ environment_satisfaction    <dbl> 2, 3, 4, 4, 1, 4, 3, 4, 4, 3, 1, 4, 1, 2, 3…
## $ gender                      <chr> "Female", "Male", "Male", "Female", "Male",…
## $ job_involvement             <dbl> 3, 2, 2, 3, 3, 3, 4, 3, 2, 3, 4, 2, 3, 3, 2…
## $ job_level                   <dbl> 2, 2, 1, 1, 1, 1, 1, 1, 3, 2, 1, 2, 1, 1, 1…
## $ job_role                    <chr> "Sales Executive", "Research Scientist", "L…
## $ job_satisfaction            <dbl> 4, 2, 3, 3, 2, 4, 1, 3, 3, 3, 2, 3, 3, 4, 3…
## $ marital_status              <chr> "Single", "Married", "Single", "Married", "…
## $ monthly_income              <dbl> 5993, 5130, 2090, 2909, 3468, 3068, 2670, 2…
## $ num_companies_worked        <dbl> 8, 1, 6, 1, 9, 0, 4, 1, 0, 6, 0, 0, 1, 0, 5…
## $ over_time                   <chr> "Yes", "No", "Yes", "Yes", "No", "No", "Yes…
## $ percent_salary_hike         <dbl> 11, 23, 15, 11, 12, 13, 20, 22, 21, 13, 13,…
## $ performance_rating          <dbl> 3, 4, 3, 3, 3, 3, 4, 4, 4, 3, 3, 3, 3, 3, 3…
## $ relationship_satisfaction   <dbl> 1, 4, 2, 3, 4, 3, 1, 2, 2, 2, 3, 4, 4, 3, 2…
## $ stock_option_level          <dbl> 0, 1, 0, 0, 1, 0, 3, 1, 0, 2, 1, 0, 1, 1, 0…
## $ total_working_years         <dbl> 8, 10, 7, 8, 6, 8, 12, 1, 10, 17, 6, 10, 5,…
## $ training_times_last_year    <dbl> 0, 3, 3, 3, 3, 2, 3, 2, 2, 3, 5, 3, 1, 2, 4…
## $ work_life_balance           <dbl> 1, 3, 3, 3, 3, 2, 2, 3, 3, 2, 3, 3, 2, 3, 3…
## $ years_at_company            <dbl> 6, 10, 0, 8, 2, 7, 1, 1, 9, 7, 5, 9, 5, 2, …
## $ years_in_current_role       <dbl> 4, 7, 0, 7, 2, 7, 0, 0, 7, 7, 4, 5, 2, 2, 2…
## $ years_since_last_promotion <dbl> 0, 1, 0, 3, 2, 3, 0, 0, 1, 7, 0, 0, 4, 1, 0…
## $ years_with_curr_manager     <dbl> 5, 7, 0, 0, 2, 6, 0, 0, 8, 7, 3, 8, 3, 2, 3…
## $ attrition                   <chr> "Yes", "No", "Yes", "No", "No", "No", "No",…
## $ employee_number             <dbl> 1, 2, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 1…
```

```
bio <- data %>% clean_names() %>% na.omit() %>%
  mutate(attrition = factor(attrition))



summary(bio)
```

```
##       age        business_travel     department       distance_from_home
##  Min.   :18.00   Length:1470        Length:1470        Min.   : 1.000
##  1st Qu.:30.00   Class :character   Class :character   1st Qu.: 2.000
##  Median :36.00   Mode  :character   Mode  :character   Median : 7.000
##  Mean   :36.92                                         Mean   : 9.193
##  3rd Qu.:43.00                                         3rd Qu.:14.000
##  Max.   :60.00                                         Max.   :29.000
##    education     education_field    environment_satisfaction    gender
##  Min.   :1.000   Length:1470        Min.   :1.000            Length:1470
##  1st Qu.:2.000   Class :character   1st Qu.:2.000            Class :character
##  Median :3.000   Mode  :character   Median :3.000            Mode  :character
##  Mean   :2.913                      Mean   :2.722
##  3rd Qu.:4.000                      3rd Qu.:4.000
##  Max.   :5.000                      Max.   :4.000
##  job_involvement   job_level       job_role         job_satisfaction
##  Min.   :1.00    Min.   :1.000   Length:1470        Min.   :1.000
##  1st Qu.:2.00    1st Qu.:1.000   Class :character   1st Qu.:2.000
##  Median :3.00    Median :2.000   Mode  :character   Median :3.000
##  Mean   :2.73    Mean   :2.064                      Mean   :2.729
##  3rd Qu.:3.00    3rd Qu.:3.000                      3rd Qu.:4.000
##  Max.   :4.00    Max.   :5.000                      Max.   :4.000
##  marital_status     monthly_income  num_companies_worked  over_time
##  Length:1470        Min.   : 1009   Min.   :0.000        Length:1470
##  Class :character   1st Qu.: 2911   1st Qu.:1.000        Class :character
##  Mode  :character   Median : 4919   Median :2.000        Mode  :character
##                     Mean   : 6503   Mean   :2.693
##                     3rd Qu.: 8379   3rd Qu.:4.000
##                     Max.   :19999   Max.   :9.000
##  percent_salary_hike performance_rating relationship_satisfaction
##  Min.   :11.00       Min.   :3.000      Min.   :1.000
##  1st Qu.:12.00       1st Qu.:3.000      1st Qu.:2.000
##  Median :14.00       Median :3.000      Median :3.000
##  Mean   :15.21       Mean   :3.154      Mean   :2.712
##  3rd Qu.:18.00       3rd Qu.:3.000      3rd Qu.:4.000
##  Max.   :25.00       Max.   :4.000      Max.   :4.000
##  stock_option_level total_working_years training_times_last_year
##  Min.   :0.0000     Min.   : 0.00       Min.   :0.000
##  1st Qu.:0.0000     1st Qu.: 6.00       1st Qu.:2.000
##  Median :1.0000     Median :10.00       Median :3.000
##  Mean   :0.7939     Mean   :11.28       Mean   :2.799
##  3rd Qu.:1.0000     3rd Qu.:15.00       3rd Qu.:3.000
##  Max.   :3.0000     Max.   :40.00       Max.   :6.000
##  work_life_balance years_at_company years_in_current_role
##  Min.   :1.000     Min.   : 0.000   Min.   : 0.000
##  1st Qu.:2.000     1st Qu.: 3.000   1st Qu.: 2.000
##  Median :3.000     Median : 5.000   Median : 3.000
##  Mean   :2.761     Mean   : 7.008   Mean   : 4.229
##  3rd Qu.:3.000     3rd Qu.: 9.000   3rd Qu.: 7.000
##  Max.   :4.000     Max.   :40.000   Max.   :18.000
##  years_since_last_promotion years_with_curr_manager attrition  employee_number
##  Min.   : 0.000             Min.   : 0.000          No :1233   Min.   :   1.0
##  1st Qu.: 0.000             1st Qu.: 2.000          Yes: 237   1st Qu.: 491.2
```

```
##  Median : 1.000            Median : 3.000                    Median :1020.5
##  Mean   : 2.188            Mean   : 4.123                    Mean   :1024.9
##  3rd Qu.: 3.000            3rd Qu.: 7.000                    3rd Qu.:1555.8
##  Max.   :15.000            Max.   :17.000                    Max.   :2068.0
```

```
#What is the attrition rate for employees at BI? (A rate, remember, is expressed as a pr
oportion.)
# Calculate overall attrition rate.


#For Attrition (Employees leaving)
bio %>%
  summarize(attrition = mean(attrition =="Yes"))
```

```
## # A tibble: 1 × 1
##   attrition
##       <dbl>
## 1     0.161
```

```
#For Attrition (Employees Staying)
bio %>%
  summarize(attrition = mean(attrition == 'No'))
```

```
## # A tibble: 1 × 1
##   attrition
##       <dbl>
## 1     0.839
```

```
#237/(1233 + 237)

#1233/(1233 + 237)
```

The attrition rate for employees leaving Bioimplants is 0.16 The majority class for attrition is No as that is 0.84 which means that more employees stay than leave.

```
#Create a summary table of conditional attrition rates by department and job role. (The
table should have 3 columns: department, job role, and the calculated conditional attrit
ion rate.) Sort this table by attrition rate in descending order.


bio %>% group_by(department, job_role) %>%
  summarize(attrition = mean(attrition == 'Yes')) %>%
  arrange(desc(attrition))
```

```
## `summarise()` has grouped output by 'department'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 11 × 3
## # Groups:   department [3]
##    department             job_role                  attrition
##    <chr>                  <chr>                         <dbl>
##  1 Sales                  Sales Representative        0.398
##  2 Research & Development Laboratory Technician       0.239
##  3 Human Resources        Human Resources             0.231
##  4 Sales                  Sales Executive             0.175
##  5 Research & Development Research Scientist          0.161
##  6 Research & Development Manufacturing Director      0.0690
##  7 Research & Development Healthcare Representative   0.0687
##  8 Research & Development Manager                     0.0556
##  9 Sales                  Manager                     0.0541
## 10 Research & Development Research Director           0.025
## 11 Human Resources        Manager                     0
```

```
# Note: The simplest possible classification model would be to use the attrition majorit
y class—“Yes” or “No”—as the prediction. This is called “majority class” prediction. The
in-sample accuracy of the majority class model is simply the proportion of the majority
class. This is an important performance benchmark.
```

The sales department with the job role as a representative has the highest attrition rate of 0.40.

The research & development department with the job role as a Laboratory Technician has the second highest attrition rate of 0.24.

The human resources department with the job role being involved in HR has the third highest attrition rate of 0.23.

Q2

```
# Fit a logistic regression model of attrition using all the predictors. (Note: employee
_number is NOT a predictor!)

logistic_bio <- glm(ifelse(attrition=='Yes',1,0) ~. -employee_number,
                    data = bio,
                    family = binomial)

summary(logistic_bio)
```

```
##
## Call:
## glm(formula = ifelse(attrition == "Yes", 1, 0) ~ . - employee_number,
##     family = binomial, data = bio)
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      -1.060e+01  3.836e+02  -0.028 0.977965
## age                              -3.133e-02  1.352e-02  -2.317 0.020478 *
## business_travelTravel_Frequently  1.925e+00  4.102e-01   4.692 2.70e-06 ***
## business_travelTravel_Rarely      1.041e+00  3.781e-01   2.752 0.005919 **
## departmentResearch & Development  1.279e+01  3.836e+02   0.033 0.973411
## departmentSales                   1.261e+01  3.836e+02   0.033 0.973777
## distance_from_home                4.584e-02  1.073e-02   4.271 1.95e-05 ***
## education                         3.184e-03  8.745e-02   0.036 0.970961
## education_fieldLife Sciences     -7.910e-01  8.030e-01  -0.985 0.324577
## education_fieldMarketing         -3.666e-01  8.522e-01  -0.430 0.667048
## education_fieldMedical           -8.939e-01  8.025e-01  -1.114 0.265365
## education_fieldOther             -8.716e-01  8.614e-01  -1.012 0.311602
## education_fieldTechnical Degree   1.182e-01  8.207e-01   0.144 0.885452
## environment_satisfaction         -4.334e-01  8.268e-02  -5.242 1.59e-07 ***
## genderMale                        3.879e-01  1.839e-01   2.109 0.034939 *
## job_involvement                  -5.312e-01  1.222e-01  -4.348 1.37e-05 ***
## job_level                        -7.592e-02  3.147e-01  -0.241 0.809386
## job_roleHuman Resources           1.402e+01  3.836e+02   0.037 0.970841
## job_roleLaboratory Technician     1.492e+00  4.833e-01   3.087 0.002024 **
## job_roleManager                   3.909e-01  8.867e-01   0.441 0.659319
## job_roleManufacturing Director    2.584e-01  5.301e-01   0.487 0.625962
## job_roleResearch Director        -1.051e+00  1.002e+00  -1.049 0.294010
## job_roleResearch Scientist        5.535e-01  4.943e-01   1.120 0.262804
## job_roleSales Executive           1.202e+00  1.126e+00   1.068 0.285482
## job_roleSales Representative      2.144e+00  1.180e+00   1.817 0.069265 .
## job_satisfaction                 -4.184e-01  8.118e-02  -5.154 2.55e-07 ***
## marital_statusMarried             3.213e-01  2.657e-01   1.209 0.226602
## marital_statusSingle              1.160e+00  3.439e-01   3.372 0.000746 ***
## monthly_income                    8.216e-06  8.116e-05   0.101 0.919373
## num_companies_worked              1.935e-01  3.868e-02   5.002 5.68e-07 ***
## over_timeYes                      1.970e+00  1.929e-01  10.211  < 2e-16 ***
## percent_salary_hike              -2.192e-02  3.907e-02  -0.561 0.574786
## performance_rating                1.068e-01  3.966e-01   0.269 0.787631
## relationship_satisfaction        -2.571e-01  8.240e-02  -3.120 0.001808 **
## stock_option_level               -2.087e-01  1.568e-01  -1.331 0.183054
## total_working_years              -6.131e-02  2.940e-02  -2.085 0.037031 *
## training_times_last_year         -1.918e-01  7.304e-02  -2.626 0.008633 **
## work_life_balance                -3.633e-01  1.234e-01  -2.943 0.003249 **
## years_at_company                  9.443e-02  3.892e-02   2.426 0.015257 *
## years_in_current_role            -1.518e-01  4.521e-02  -3.357 0.000789 ***
## years_since_last_promotion        1.780e-01  4.205e-02   4.234 2.30e-05 ***
## years_with_curr_manager          -1.346e-01  4.707e-02  -2.859 0.004253 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1298.58  on 1469  degrees of freedom
## Residual deviance:  860.85  on 1428  degrees of freedom
## AIC: 944.85
##
## Number of Fisher Scoring iterations: 14
```

```
# Report accuracy for this model with a decision threshold of .5. (Accuracy is defined a
s the proportion of correct predictions.)
predict(logistic_bio, type = 'response') %>% head()
```

```
##          1          2          3          4          5          6
## 0.71708957 0.01524084 0.67736633 0.14476907 0.34904441 0.06889458
```

```
ifelse(predict(logistic_bio, type = 'response') > .5, 'Yes', 'No') %>% head()
```

```
##     1     2     3     4     5     6
## "Yes"  "No" "Yes"  "No"  "No"  "No"
```

```
(ifelse(predict(logistic_bio, type = 'response') > .5, 'Yes', 'No') == bio$attrition) %
>% mean()
```

```
## [1] 0.892517
```

```
# Comment on whether the model offers an improvement over predicting with the majority c
lass.
```

The coefficients of the model is expressed in log odds due to logistical regression.

The model offers an improvement over predicting with just the majority class since it analyzes all the predictors in relation to attrition. The majority class prediction was 0.84 which was to not leave the company. In comparison the logistic model with an accuracy of 0.89 is doing better than the majority class prediction.

Q3

```
# The upside of standardizing inputs by centering and scaling is that it allows you to c
ompare coefficient effect sizes easily—they are all on the same scale. (The downside is
that they are no longer scaled in the original units, and interpretation changes.) Even
though the coefficients are expressed in log odds in this case, after standardization th
ey can still be compared for effect sizes on a relative basis.
#
# There are a lot of coefficients to type into the model formula. A shortcut to automati
cally include all the predictors in the dataset is ., as in: glm(target ~ ., family = bi
nomial, data = ...). However, this shortcut doesn’t allow you to standardize also. The e
asiest solution to create a new data set in which all the continuous variables are cente
red. For this a version of mutate() is useful: mutate_if(). The code would go like this:
#
# data %>%    mutate_if(is.numeric, scale)
#
# In English: if the variable is numeric, then scale it.
#
# Notice that some of the standard errors and coefficients in the model above have explo
ded. (You can see this more easily if you adjust the number of digits printed in the out
put with options(scipen = 3).) The SEs for some of the department and job_role coefficie
nts are over 380. Why has this happened? Multicolinearity! Some of the levels of the dep
artment variable are correlated with levels in job_role. For example, since most of the
people in the Human Resources department also have a job title of Human Resources, the i
nformation from department is redundant: by definition, if we know job_role we also know
department and vice versa. This is a textbook example of how multicollinearity makes inf
erence difficult—we can’t compare the coefficients because some of them are wacky. The s
olution? Remove the redundant variable. Refit the model without department

bio_scaled <- bio %>%
  mutate_if(is.numeric, function(x) scale(x) %>% as.vector())

glimpse(bio_scaled)
```

```
## Rows: 1,470
## Columns: 29
## $ age                        <dbl> 0.44619856, 1.32191535, 0.00834016, -0.4295…
## $ business_travel            <chr> "Travel_Rarely", "Travel_Frequently", "Trav…
## $ department                 <chr> "Sales", "Research & Development", "Researc…
## $ distance_from_home         <dbl> -1.01056544, -0.14709966, -0.88721318, -0.7…
## $ education                  <dbl> -0.89138490, -1.86779013, -0.89138490, 1.06…
## $ education_field            <chr> "Life Sciences", "Life Sciences", "Other", …
## $ environment_satisfaction   <dbl> -0.6603060, 0.2545383, 1.1693826, 1.1693826…
## $ gender                     <chr> "Female", "Male", "Male", "Female", "Male",…
## $ job_involvement            <dbl> 0.379543, -1.025818, -1.025818, 0.379543, 0…
## $ job_level                  <dbl> -0.05776789, -0.05776789, -0.96115930, -0.9…
## $ job_role                   <chr> "Sales Executive", "Research Scientist", "L…
## $ job_satisfaction           <dbl> 1.1528613, -0.6606284, 0.2461164, 0.2461164…
## $ marital_status             <chr> "Single", "Married", "Single", "Married", "…
## $ monthly_income             <dbl> -0.1083127, -0.2916193, -0.9373347, -0.7633…
## $ num_companies_worked       <dbl> 2.1244130, -0.6778187, 1.3237753, -0.677818…
## $ over_time                  <chr> "Yes", "No", "Yes", "Yes", "No", "No", "Yes…
## $ percent_salary_hike        <dbl> -1.15016269, 2.12858163, -0.05724792, -1.15…
## $ performance_rating         <dbl> -0.426085, 2.345353, -0.426085, -0.426085, …
## $ relationship_satisfaction  <dbl> -1.5836393, 1.1910327, -0.6587487, 0.266142…
## $ stock_option_level         <dbl> -0.9316973, 0.2419060, -0.9316973, -0.93169…
## $ total_working_years        <dbl> -0.42149902, -0.16445544, -0.55002081, -0.4…
## $ training_times_last_year   <dbl> -2.1712429, 0.1556541, 0.1556541, 0.1556541…
## $ work_life_balance          <dbl> -2.4929720, 0.3379811, 0.3379811, 0.3379811…
## $ years_at_company           <dbl> -0.164557109, 0.488341541, -1.143905083, 0.…
## $ years_in_current_role      <dbl> -0.06327437, 0.76473737, -1.16729002, 0.764…
## $ years_since_last_promotion <dbl> -0.67891464, -0.36858985, -0.67891464, 0.25…
## $ years_with_curr_manager    <dbl> 0.2457504, 0.8062671, -1.1555415, -1.155541…
## $ attrition                  <fct> Yes, No, Yes, No, No, No, No, No, No, No, N…
## $ employee_number            <dbl> -1.700704, -1.699043, -1.695721, -1.694060,…
```

```
# Which of the centered and scaled predictors has the largest effect size?
# Interpret the coefficient with the largest effect size. Since you are working with sta
ndardized coefficients, the interpretation for continuous predictors will be: a 1 unit
(that is, after scaling, a 1 standard deviation) increase in x is associated with a coef
ficient-sized change in the log odds of y, on average, while holding the other predictor
s constant. The coefficient represents the change in the log odds of the outcome associa
ted with an increase from the reference level in the categorical variable.

bio_scaled_mod <- glm(attrition ~. -department -employee_number, data = bio_scaled, fami
ly = binomial)

summary(bio_scaled_mod)
```

```
##
## Call:
## glm(formula = attrition ~ . - department - employee_number, family = binomial,
##     data = bio_scaled)
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)    
## (Intercept)                      -5.255992   0.965177  -5.446 5.16e-08 ***
## age                              -0.286685   0.123481  -2.322 0.020250 *  
## business_travelTravel_Frequently  1.916549   0.409938   4.675 2.94e-06 ***
## business_travelTravel_Rarely      1.029756   0.377867   2.725 0.006427 ** 
## distance_from_home                0.373074   0.086912   4.293 1.77e-05 ***
## education                         0.001838   0.089633   0.021 0.983637    
## education_fieldLife Sciences     -0.636669   0.756292  -0.842 0.399883    
## education_fieldMarketing         -0.210456   0.805873  -0.261 0.793975    
## education_fieldMedical           -0.745371   0.756263  -0.986 0.324331    
## education_fieldOther             -0.728186   0.822241  -0.886 0.375827    
## education_fieldTechnical Degree   0.263370   0.777694   0.339 0.734870    
## environment_satisfaction         -0.472694   0.090303  -5.235 1.65e-07 ***
## genderMale                        0.384540   0.183942   2.091 0.036569 *  
## job_involvement                  -0.380629   0.086838  -4.383 1.17e-05 ***
## job_level                        -0.091126   0.348695  -0.261 0.793834    
## job_roleHuman Resources           1.300441   0.674242   1.929 0.053762 .  
## job_roleLaboratory Technician     1.482158   0.483232   3.067 0.002161 ** 
## job_roleManager                   0.184607   0.786224   0.235 0.814362    
## job_roleManufacturing Director    0.253002   0.529989   0.477 0.633097    
## job_roleResearch Director        -1.030455   0.998159  -1.032 0.301905    
## job_roleResearch Scientist        0.543235   0.494240   1.099 0.271710    
## job_roleSales Executive           1.018424   0.446029   2.283 0.022412 *  
## job_roleSales Representative      1.956923   0.551146   3.551 0.000384 ***
## job_satisfaction                 -0.463738   0.089445  -5.185 2.16e-07 ***
## marital_statusMarried             0.318912   0.265685   1.200 0.230008    
## marital_statusSingle              1.144697   0.343560   3.332 0.000863 ***
## monthly_income                    0.042897   0.381930   0.112 0.910572    
## num_companies_worked              0.486656   0.096599   5.038 4.71e-07 ***
## over_timeYes                      1.973530   0.193041  10.223  < 2e-16 ***
## percent_salary_hike              -0.079948   0.142665  -0.560 0.575212    
## performance_rating                0.037690   0.142885   0.264 0.791949    
## relationship_satisfaction        -0.276487   0.088957  -3.108 0.001883 ** 
## stock_option_level               -0.185089   0.133146  -1.390 0.164491    
## total_working_years              -0.481526   0.228213  -2.110 0.034860 *  
## training_times_last_year         -0.243712   0.094064  -2.591 0.009572 ** 
## work_life_balance                -0.257597   0.087273  -2.952 0.003161 ** 
## years_at_company                  0.553308   0.235679   2.348 0.018889 *  
## years_in_current_role            -0.538649   0.163006  -3.304 0.000952 ***
## years_since_last_promotion        0.579743   0.134794   4.301 1.70e-05 ***
## years_with_curr_manager          -0.469430   0.168172  -2.791 0.005249 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##     Null deviance: 1298.58  on 1469  degrees of freedom
## Residual deviance:  862.21  on 1430  degrees of freedom
## AIC: 942.21
##
## Number of Fisher Scoring iterations: 7
```

The coefficient with the largest positive effect size is OvertimeYes. OvertimeYes is when employees work more than 40 hours a week For every 1 unit increase in overtime which in this case is a 1 standard deviation increase (as the numeric variables were scaled) is associated with a change in attrition rate of: 1.97 log odds while holding all the other predictors constant.

The coefficient with the largest negative effect is for jobs that are for research directors. However, not every employees can be a research director. Therefore, the largest negative predictor which is reasonable is: Education field medical. It appears that those who studied in the medical field are less likely to leave by -0.75 log odds while holding all other predictors constant.

Q4

```
# Based on the above logistic regression model (and, specifically, on the coefficient wi
th the largest effect size that you identified above), how might company policy be chang
ed to reduce employee attrition?
#
# Describe your proposed policy change.
# Estimate and explain the change in churn probability associated with that policy chang
e.
```

Based on the logistic model, I recommend to remove the implementation of overtime.

```
# Current/Baseline attrition probability via Delta Method
base <- predict(logistic_bio, type = 'response') %>% mean() #0.16
base
```

```
## [1] 0.1612245
```

```
#removal of overtime rate
remove_ot <- predict(logistic_bio,
        newdata = mutate(bio, over_time = 'No'),
        type = 'response') %>% mean() # 0.10

remove_ot
```

```
## [1] 0.1026007
```

The reduction in attrition by removing Overtime is approximately 6 points: .16 - .10 = .06