Nick Videtti
IST-772 Final Exam
Spring 2022

1. Create the following new variables:
   - public_dummy = Indicator variable that is 1 for public school 0 for private.
   - relig_exempt = Indicator that is 1 when religious > 0, and 0 otherwise.

See R code for how I did this.

2. Calculate means for each of the numerical variables in the data and the dummy variables that you created in Question 1 and report those in your solutions.

```
> for (i in 1:ncol(data))
+   {if (class(data[,i]) == 'numeric') {print(c(colnames(data[i]),mean(data[,i])))}}
[1] "allvaccs"          "89.0391067672636"
[1] "conditional"       "6.72976528626361"
[1] "medical"           "0.193064825224928"
[1] "religious"         "4.03806312143553"
[1] "dptMiss"          "8.7121545432765"
[1] "polMiss"           "8.43657180362321"
[1] "mmrMiss"           "8.88925712568195"
[1] "hepMiss"           "6.44134113139398"
[1] "varMiss"           "5.89255375116905"
[1] "public_dummy"      "0.83810888252149"
[1] "relig_exempt"      "0.567335243553009"
```

3. Develop a linear regression model that predicts medical exemptions ("medical") from "pubpriv" and "enrollment". Paste the outputs in your solutions.

```
Call:
lm(formula = medical ~ pubpriv + enrollment, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-0.3657 -0.2363 -0.1751 -0.1206 15.1251

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.2128964  0.0853441   2.495   0.0128 *
pubprivPUBLIC -0.1841863  0.1019180  -1.807   0.0712 .
enrollment     0.0017019  0.0009059   1.879   0.0607 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8555 on 695 degrees of freedom
Multiple R-squared:  0.006456,  Adjusted R-squared:  0.003597
F-statistic: 2.258 on 2 and 695 DF,  p-value: 0.1053
```

4. Provide a brief written interpretation of the results making sure to incorporate a discussion of hypothesis tests and p-values.

This is a very weak model, as we see the Multiple R-squared and the Adjusted R-squared are 0.006456 and 0.003597, respectively. This means that less than 1% of the change in "medical" is accounted for by this model. In fact, the p-value of the F-statistic shows that we do not even have statistically significant evidence (at the 0.05 level of significance) to reject the null hypothesis that R-squared is different from 0, which makes this model completely useless altogether.

Also, the coefficients go through the hypothesis test with the null hypothesis that their actual value is 0, and (at the 0.05 level of significance) only the intercept has statistically significant evidence to reject that null hypothesis.

5. Run a logistic regression model using public_dummy and enrollment to predict relig_exempt. Report the results in your solutions.

```
Call:
glm(formula = relig_exempt ~ public_dummy + enrollment, family = "binomial",
    data = data)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.444  -1.274   1.009   1.073   1.153

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.103742   0.200710   0.517    0.605
public_dummy -0.075529   0.239898  -0.315    0.753
enrollment    0.002927   0.002155   1.358    0.174

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 954.94  on 697  degrees of freedom
Residual deviance: 952.89  on 695  degrees of freedom
AIC: 958.89

Number of Fisher Scoring iterations: 4
```

6. Provide a write up of your results from question 5, with an interpretation of the question: “how do the odds of students claiming a religious exemption to vaccination increase or decrease with public school status?”. Be sure to incorporate a discussion of hypothesis tests , p-values and odds ratios.

We see that at the 0.05 level of significance, that none of the coefficients have statistically significant evidence to reject the null hypothesis that they are different from 0. Therefore, we do not have statistically significant evidence to make any conclusions about the odds of students claiming a religious exemption to vaccination based on public school status.

If we were to incorrectly assume that we did have statistically significant evidence to reject the null hypothesis that our coefficients are equal to 0, we would first need to transform our coefficients by exponentiating. This is because logistic regression gives us coefficients that are in terms of log odds.

```
> exp(coef(Question5))
 (Intercept) public_dummy   enrollment
   1.1093143    0.9272532    1.0029318
```

Under this incorrect assumption of significance, we would answer our question about the odds of students claiming a religious exemption to vaccination based on public school status by looking at the value above corresponding to “public_dummy”. As we recall, this is 0 for private schools and 1 for public schools. This result can be interpreted as students at public schools have approximately 0.927:1 odds of claiming a religious exemption compared to that of private schools. That is, the odds of a student claiming a religious exemption to vaccination decrease from private school students to public school students.