Modeling Binomial Counts

Logistic regression – Stat 230

In this class, you will fit logistic regression models to binomial count data rather to a binary response variable. To learn the basic idea, we'll consider the cancer data displayed in Table 7.3 (p. 231) of the textbook.

You can load the data using the command

```
table73 <- read.csv("https://aloy.rbind.io/kuiper_data/Table7.3.csv")
```

Fitting a binomial logistic regression model

To fit a binomial logistic regression model in R, we again use the `glm()` function; however, the response variable must now be a proportion and we must also specify the binomial denominators. Below is the basic recipe to fit this class of model:

```
binom_glm <- glm(prop ~ x, data = df, family = "binomial", weights = denoms)
```

where

- `prop` is the name of the column containing the sample proportions (of successes)
- `x` is the name of the predictor variable, and if there are multiple we add them to our formula as we always have
- `df` is the name of the data set
- `denoms` is the name of the column that contains the binomial denominators (i.e., group sizes)

> 💡 Tip
>
> Sometimes there is not a proportion column already in the data set. In this case, you can use the recipe
>
> ```
> binom_glm <- glm(y/denoms ~ x, data = df, family = "binomial", weights = denoms)
> ```
>
> where `y` denotes the observed counts (i.e., number of successes).

Task 1. Adapt the above recipe to complete chapter 7 activity 26. **Note:** The `Proportion.Malignant` column has been rounded, and R requires the proportions to be exact so you should adapt the recipe in the "Tip" box above.

Evaluating residuals

As you learned in the daily prep, there are two types of residuals that we can use to investigate the linearity of the log odds and to help identify potential outliers: Pearson residuals and deviance residuals. While these are different types of residuals than we have seen before, the good news is that we can still use the `residualPlot()` and `residualPlots()` commands from the {car} package to create them.

To create Pearson residual plots, run

```
residualPlots(binom_glm, type = "pearson", smooth = FALSE)
```

To create deviance residual plots, run

```
residualPlots(binom_glm, type = "deviance", smooth = FALSE)
```

Task 2. Complete chapter 7 activity 27.

Task 3. Complete chapter 7 activity 28.

Assessing the fit of a binomial model

At the beginning of class we discussed the the deviance goodness-of-fit test that can be run on binomial logistic regression models. Let's use this test to assess our fitted logistic regression model. (Note: In this class we won't use the Pearson goodness-of-fit test or the Hosmer-Lemeshow tests.)

Task 4. Complete chapter 7 activity 30. Be sure to state the hypotheses, test statistic, degrees of freedom, p-value, and a brief conclusion about the adequacy of your model. To help you organize your calculations, look at Table 7.4 and equation 7.26.

Task 5. Verify your calculations using the output from `summary()` to run this goodness-of-fit test.