**Reviewing Odds**

Imagine there are 100 zip codes in a city; 28 of them have a park, and the remaining 72 don’t. In this case, we’re dealing with a binary variable *Y*, where *Y = 1* indicates that there’s a park in the zip code and *Y = 0* indicates that there isn’t a park in the zip code.

**Q1.** What is the *probability* that a randomly selected zip code has a park?

**A1.**

**Q2.** What is the *probability* that a randomly selected zip code does *not* have a park?

**A2.**

**Q3.** What are the *odds* that a randomly selected zip code that has a park?

**A3.**

Or, we can write

**Q4.** What are the log odds that a randomly selected zip code has a park?

**A4.**

**Review of Logistic Regression**

Imagine we want to see whether having a park in a zip code is predicted by 4 predictors:

* MEDHVAL Median house value in the zip code, in $1000 *Continuous*
* %POVERTY % individuals living in poverty *Continuous*
* URBAN =1 if zip code urban; =0 otherwise *Binary*
* COMMERCIAL =1 if zip code commercial; =0 otherwise *Binary*

Below are the regression results.

**Coefficients:**

**Estimate Std. Error z value P-value OR**

INTERCEPT -8.320 2.320 -3.586 0.000 0.0002

MEDHVAL 0.923 0.328 2.814 0.005 2.5168

%POVERTY -0.354 0.113 -3.133 0.002 0.7019

URBAN 0.581 0.274 2.118 0.034 1.7878

COMMERCIAL -0.438 0.128 -3.422 0.001 0.6453

***Interpretation of the output***

As we can see, all 4 predictors are significant. We can interpret each one of them as follows:

INTERCEPT:

* The estimate is -8.320. In this case, the estimated coefficient for the intercept (-8.320) is the log odds of there being a park in a zip code where all predictors are 0.
* Said differently, -8.320 is the log odds of there being a park in a zip code where
  + MEDHVAL = 0 (i.e., median house value is 0), AND
  + %POVERTY = 0 (i.e., poverty rate is 0), AND
  + URBAN = 0 (i.e., zip code is not urban), AND
  + COMMERCIAL = 0 (i.e., zip code is not commercial)
* Said differently, the odds that a zip code where
  + MEDHVAL = 0 (i.e., median house value is 0), AND
  + %POVERTY = 0 (i.e., poverty rate is 0), AND
  + URBAN = 0 (i.e., zip code is not urban), AND
  + COMMERCIAL = 0 (i.e., zip code is not commercial)

has a park are e-8.320 = 2.72-8.320 = 0.0002.

MEDHVAL:

* The estimate is .923. This means that for a 1 unit ($1,000) increase in MEDHVAL, the log odds of there being a park in the zip code go up by .923, holding COMMERCIAL, %POVERTY and URBAN constant.
* Said differently, for a 1 unit ($1,000) increase in MEDHVAL, the odds of there being a park in the zip code go up by a factor of e.923 = 2.5168, holding COMMERCIAL, %POVERTY and URBAN constant.
  + E.g.: Imagine we have 2 zip codes with identical values of COMMERCIAL, %POVERTY and URBAN. Imagine further that in zip code 1, MEDHVAL = *a*, and in zip code 2, MEDHVAL = *a* + 1. Then, if we were to divide the odds of there being a park in zip code 2 by the odds of there being a park in zip code 1, the ratio of those odds (i.e., *odds ratio*) would be 2.5168.
* Said differently, for a 1 unit ($1,000) increase in MEDHVAL, the odds of there being a park in the zip code go up by 151.68%, holding COMMERCIAL, %POVERTY and URBAN constant.

%POVERTY:

* The estimate is -.354. This means that for a 1 unit (1%) increase in %POVERTY, the log odds of there being a park in the zip code decrease by .354, holding MEDHVAL, COMMERCIAL and URBAN constant.
* Said differently, for a 1 unit (1%) increase in %POVERTY, the odds of there being a park in the zip code go down by a factor of e-.354 = .7019, holding MEDHVAL, COMMERCIAL and URBAN constant.
  + E.g.: Imagine we have 2 zip codes with identical values of MEDHVAL, COMMERCIAL and URBAN. Imagine further that in zip code 1, %POVERTY = *c* and in zip code 2, %POVERTY = *c* + 1. Then, if we were to divide the odds of there being a park in zip code 2 by the odds of there being a park in zip code 1, the ratio of these odds (i.e., the *odds ratio*) would be .7019.
* Said differently, for a 1 unit (1%) increase in %POVERTY, the odds of there being a park in the zip code change by -29.81% -- that is, they go down by 29.81%, holding MEDHVAL, COMMERCIAL and URBAN constant.

URBAN:

* The estimate is 0.581. This means that for a 1 unit increase in URBAN (i.e., as we go from a non-urban to urban zip code), the log odds of there being a park in the zip code go up by 0.581, holding COMMERCIAL, %POVERTY and MEDHVAL constant.
* Said differently, for a 1 unit increase in URBAN (i.e., as we go from a non-urban to urban zip code), the odds of there being a park in the zip code go up by a factor of e0.581 = 1.7878, holding COMMERCIAL, %POVERTY and MEDHVAL constant.
  + The ratio between the odds of there being a park in an urban zip code and the odds of there being a park in a non-urban zip code is 1.7878.
* Said differently, for a 1 unit increase in URBAN (i.e., as we go from a non-urban to urban zip code), the odds of there being a park in the zip code go up by , holding COMMERCIAL, %POVERTY and MEDHVAL constant.

COMMERCIAL:

* The estimate is -0.438. This means that for a 1 unit increase in COMMERCIAL (i.e., as we go from a non-commercial zip code to a commercial zip code), the log odds of there being a park in the zip code go down by 0.438, holding URBAN, %POVERTY and MEDHVAL constant.
* Said differently, for a 1 unit increase in COMMERCIAL (i.e., as we go from a non-commercial zip code to a commercial one), the odds of there being a park in the zip code decrease by a factor of e-0.438 = 0.6453, holding URBAN, %POVERTY and MEDHVAL constant.
* Said differently, the odds of there being a park in a commercial zip code are 0.6453 of the odds of there being a park in a non-commercial zip code, holding URBAN, %POVERTY and MEDHVAL constant.
  + The ratio between the odds of there a being a park in a commercial zip code and the odds of there being a park in a non-commercial zip code is 0.6453.
* Said differently, the odds of there being a park in a commercial zip code are lower than the odds of there being a park in a non-commercial zip code.
* Said differently, the odds of there being a park in a non-commercial zip code are 1/0.6453 = 1.5497 higher than the odds of there being a park in a commercial zip code, holding URBAN, %POVERTY and MEDHVAL constant.
  + The ratio of the odds of there being a park in a non-commercial zip code and the odds of there being a park in a commercial zip code is 1.5497.
* Said differently, the odds of there being a park in a non-commercial zip code are higher than the odds of there being a park in a commercial zip code, holding URBAN, %POVERTY and MEDHVAL constant.

**A few notes on log odds and odds:**

* The log odds (i.e., the ) can range between negative infinity and infinity.
  + Negative values indicate that there’s a negative association between predictor *i* and the dependent variable
  + A value of 0 indicates that there is no relationship between predictor *i* and the dependent variable
  + Positive values indicate that there’s a positive association between predictor *i* and the dependent variable
* The odds ratios (i.e., the exponentiated ) can range between 0 and infinity – that is, they cannot be negative.
  + (Non-negative) odds ratios < 1 indicate that there’s a negative association between predictor *i* and the dependent variable (this corresponds to < 0)
  + Odds ratio = 1 indicates that there’s no relationship between predictor *i* and the dependent variable (this corresponds to = 0)
  + Odds ratios > 1 indicate that there’s a positive association between predictor *i* and the dependent variable (this corresponds to > 0)