2019R1 Discrete Data Analysis (STAT5107) Assignment 4

Yiu Chung WONG 1155017920

```r
set.seed(5107);
```

1.

$$\frac{\pi(x+1)}{\pi(x)}$$

is the relative risk between x+1 and x.

In a logit model

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}$$

assuming $0 \neq \Pr[Y = 1] \neq 1$, the log-odd is defined as

$$logit(\pi(x)) = \log \frac{\Pr[Y = 1]}{\Pr[Y = 0]} = \alpha + \beta x.$$

if nudge the variable by a small value, we have

$$logit(\pi(x + \delta))$$

and the difference between the nudged log-odd and the original log-odd is defined as

$$\begin{aligned} logit(\pi(x+\delta)) - logit(\pi(x)) &= (\alpha + \beta(x+\delta)) - (\alpha + \beta x) \\ &= \beta\delta \end{aligned}$$

- β is the change in log-odd when x changes by δ unit.

let δ be 1 and we have

$$\begin{aligned} \exp(\beta) &= \exp(\beta \times 1) \\ &= \exp(logit(\pi(x+1)) - logit(\pi(x))) \\ &= \frac{\frac{\pi(x+1)}{1-\pi(x+1)}}{\frac{\pi(x)}{1-\pi(x)}} \end{aligned}$$

$exp(\beta)$ is then the odds ratio for a one-unit increase in x.

- For a small success rate, the relative risk and odds ratio are close.

2a.

For

$$logit(\pi(x)) = \alpha + \beta \log(d)$$

Then it can be rewritten as:

$$\begin{aligned} log\frac{\pi(x)}{1-\pi(x)} &= \alpha + \beta \log(d) \\ \frac{\pi(x)}{1-\pi(x)} &= \exp(\alpha) \times \exp(\log(d^\beta)) \\ &= \exp(\alpha) \times d^\beta \end{aligned}$$

Then d will be 1 for the first draft pick

$$\exp(\alpha) \times 1^\beta = \exp(\alpha)$$

2b.

On average, first drafts have probability of success

$$\exp(\alpha)$$

- For basketball, this is `exp(2.3)` = 9.9741825.
- For baseball, this is `exp(0.7)` = 2.0137527.

On average, for each unit increase in d, the probability of success is multiplied by d^β.

- For basketball, this is `d^{-1.1}`. This mean the probability of success has an inverse relationship with d.
- For baseball, this is `d^{-0.6}`. Probability of success also falls as d increase but at a slower rate.

3.

```
alpha <- -3.7771;
beta <- 0.1449;
prob <- function(li)
{
  1/(1 + exp(-(alpha + li*beta)));
}
```

3a.

```
li <- 8;
prob(li);
```

```
## [1] 0.06799525
```

3b.

```r
-alpha/beta;
```

```
## [1] 26.06694
```

3c.

```r
li_0.009 <- 8;
pi_0.009 <- prob(li_0.009);
rate_0.009 <- beta*(pi_0.009 * (1 - pi_0.009));

li_0.036 <- 26;
pi_0.036 <- prob(li_0.036);
rate_0.036 <- beta*(pi_0.036 * (1 - pi_0.036));

kable(data.frame(rate_0.009, rate_0.036))
```

rate_0.009	rate_0.036
0.0091826	0.0362241

3d.

```r
kable(data.frame(prob(14), prob(28)));
```

prob.14.	prob.28.
0.1482365	0.5695707

3e.

```r
exp(beta);
```

```
## [1] 1.155924
```

3f.

```r
beta <- 0.1449;
se <- 0.0593;
exp( qnorm(c(.025, .975), beta, se))
```

```
## [1] 1.029089 1.298391
```

- Odds ratios are also not normally distributed. As a result, we must take the natural log of the odds ratio and first compute the confidence limits on a logarithmic scale.

3g.

H_0: β is not sigificantly different to zero. H_1: β is sigificantly different to zero.

```
H_0 <- 0;
z_wald <- (beta - H_0) / se;
cutoff <- qnorm(.975, 0, 1);
```

- For a large sample, the Wald statastic $z = \hat{\beta}/se(\hat{\beta})$ has a standard normal distribution when $\beta = 0$.
- The wald statistic is at 2.4435076 which is greater than the cutoff at 1.959964.
- There is enough evidence to reject H_0.

3h.

- H_0: All βs are zero.
- H_1: At least a βs is not zero.

```
intercept_only <- 34.372;
intercept_with_covariates <- 26.073;
deviance_diff <- intercept_only - intercept_with_covariates;
df <- 1;
deviance_diff > qchisq(.95, df);
```

```
## [1] TRUE
```

- The test statistic is greater than the cutoff. There is enough evidence to reject H_0; we conclude at least a βs is not zero.

4a.

```
intercept <- -3.5961;
def_beta <- -.8678;
vic_beta <- 2.4044;

death_prob <- function(def_race = 0, vic_race = 0)
{
  prob <- 1 / (1 + exp(-(intercept + def_beta * def_race + vic_beta * vic_race)));
  return(prob);
}
```

- Holding all other predictors constant, on average, the odds of death penalty for white over black defendant is 0.4198743; white defendants are 41.9874257% as likely to face death penalty.
- Holding all other predictors constant, on average, the odds of death penalty for white victim over black victim s 11.0717852, cases of white victim are 1107.1785218% as likely to face death penalty.

```
outer(X = 0:1, Y = 0:1, FUN = death_prob);
```

```
##             [,1]     [,2]
## [1,] 0.02669815 0.232955
## [2,] 0.01138622 0.113096
```

- Combination with highest probability of death penalty is: defendant = 0, victim = 1.

4b.

```
vic_lower <- 1.3068;
vic_upper <- 3.7175;
```

- Conditioned on defendant's race, the confidence interval for the effect on the odds from black to white victim equals $(e^{vic\ lower}, e^{vic\ upper}) = (3.6943329, 41.161362)$. We infer that cases with white victims have at least has at least 269.4332912% increase and at most 41 fold in the odds that a given defendant faces death penalty.

4ci.

- H_0: β for defendant is zero.
- H_1: β for defendant is not zero.

```
def_se <- 0.3671;
def_z <- def_beta/def_se;
normal_cutoff <- qnorm(c(.025, .975));
def_z < normal_cutoff[1];
```

```
## [1] TRUE
```

- At 0.5 α level, the Wald test statistic is lower than the left side cutoff; there is enough evidence to reject H_0 and we conclude β for defendant is not zero.

4cii.

- H_0: β for defendant is zero.
- H_1: β for defendant is not zero.

```
chi_cutoff <- qnorm(.95, 1);
def_chi <- 5.59;
def_chi > chi_cutoff;
```

```
## [1] TRUE
```

- With degrees of freedom of 1, the Chi-Square cutoff is at 2.6448536, which is lower than the Chi-Square statistic of the β of the defendant variable.
- There is enough evidence to reject H_0 and we conclude β for defendant is not zero.