

Vaccination Rate v.s. Covariates

Log Odds v.s. covariates

```
f <- '~/github-repo/w210_capstone/mids_capstone/vaxxfacts/cleaning_etl/city_correlation.csv'
d <- read.csv(f)
d$city <- as.factor(d$city)
d$school_year <- as.factor(d$school_year)
#d$year <- as.factor(d$school_year)
d2012 <- d[d$year == 2012,]
d2013 <- d[d$year == 2013,]
d2014 <- d[d$year == 2014,]
d2015 <- d[d$year == 2015,]
d2016 <- d[d$year == 2016,]
d2017 <- d[d$year == 2017,]

#head(d)
library(Hmisc)

## Warning: package 'Hmisc' was built under R version 3.3.2
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 3.3.2
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:base':
##
##      format.pval, round.POSIXt, trunc.POSIXt, units

#describe(d)

library(betareg)

y.transf.betareg <- function(y){
  n.obs <- sum(!is.na(y))
  (y * (n.obs - 1) + 0.5) / n.obs
}

## 2012
d2012$vac_rate <- y.transf.betareg(d2012$vac_pct / 100)

mod2012 <- betareg(vac_rate ~ male_pct + median_age +
  under_5_pct + X5_9_pct +
  X10_14_pct + X15_19_pct +
  hispanic_latino_pct +
  white_pct + black_pct +
  aian_pct + asian_pct +
  nhopi_pct + other_pct +
```

```

        median_income + no_insurance_pct,
        data=d2012)

summary(mod2012, type = "deviance")

##
## Call:
## betareg(formula = vac_rate ~ male_pct + median_age + under_5_pct +
##       X5_9_pct + X10_14_pct + X15_19_pct + hispanic_latino_pct + white_pct +
##       black_pct + aian_pct + asian_pct + nhopi_pct + other_pct + median_income +
##       no_insurance_pct, data = d2012)
##
## Deviance residuals:
##      Min      1Q  Median      3Q      Max
## -13.1765  0.2534  0.5605  0.8404  2.4586
##
## Coefficients (mean model with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -9.637e+00  1.299e+01  -0.742  0.45800
## male_pct       1.893e-02  6.191e-03   3.058  0.00223 **
## median_age     5.837e-03  6.795e-03   0.859  0.39028
## under_5_pct   -3.280e-03  1.131e-02  -0.290  0.77186
## X5_9_pct      -7.835e-03  1.164e-02  -0.673  0.50089
## X10_14_pct     1.405e-02  1.102e-02   1.275  0.20214
## X15_19_pct    -1.077e-06  1.149e-02   0.000  0.99993
## hispanic_latino_pct 1.197e-01  1.294e-01   0.925  0.35499
## white_pct      1.036e-01  1.294e-01   0.801  0.42339
## black_pct      1.121e-01  1.306e-01   0.858  0.39073
## aian_pct       1.043e-01  1.295e-01   0.805  0.42071
## asian_pct      1.194e-01  1.297e-01   0.920  0.35747
## nhopi_pct      1.671e-01  1.350e-01   1.238  0.21585
## other_pct      4.996e-02  6.522e-02   0.766  0.44367
## median_income  1.806e-06  1.211e-06   1.491  0.13586
## no_insurance_pct -1.863e-03  4.157e-03  -0.448  0.65397
##
## Phi coefficients (precision model with identity link):
##      Estimate Std. Error z value Pr(>|z|)
## (phi) 13.4098    0.7088  18.92  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 1590 on 17 Df
## Pseudo R-squared: 0.27
## Number of iterations: 43 (BFGS) + 8 (Fisher scoring)

## 2013
d2013$vac_rate <- y.transf.betareg(d2013$vac_pct / 100)

mod2013 <- betareg(vac_rate ~ male_pct + median_age +
                    under_5_pct + X5_9_pct +
                    X10_14_pct + X15_19_pct +
                    hispanic_latino_pct +
                    white_pct + black_pct +

```

```

        aian_pct + asian_pct +
        nhopi_pct + other_pct +
        median_income + no_insurance_pct,
data=d2013)

summary(mod2013, type = "deviance")

##
## Call:
## betareg(formula = vac_rate ~ male_pct + median_age + under_5_pct +
##       X5_9_pct + X10_14_pct + X15_19_pct + hispanic_latino_pct + white_pct +
##       black_pct + aian_pct + asian_pct + nhopi_pct + other_pct + median_income +
##       no_insurance_pct, data = d2013)
##
## Deviance residuals:
##      Min      1Q  Median      3Q      Max
## -12.7020 -0.2215  0.4387  0.7098  2.7649
##
## Coefficients (mean model with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.492e+01  9.845e+00  -1.515 0.129659
## male_pct       2.130e-03  6.186e-03   0.344 0.730537
## median_age     1.853e-02  6.334e-03   2.926 0.003432 **
## under_5_pct    3.106e-02  1.134e-02   2.738 0.006184 **
## X5_9_pct      -3.744e-04  1.216e-02  -0.031 0.975432
## X10_14_pct     4.236e-02  1.230e-02   3.443 0.000575 ***
## X15_19_pct     2.546e-03  1.079e-02   0.236 0.813399
## hispanic_latino_pct 1.707e-01  9.794e-02   1.743 0.081389 .
## white_pct      1.510e-01  9.790e-02   1.542 0.123003
## black_pct      1.484e-01  9.870e-02   1.503 0.132825
## aian_pct       1.408e-01  9.817e-02   1.434 0.151549
## asian_pct      1.652e-01  9.809e-02   1.685 0.092055 .
## nhopi_pct      2.219e-01  1.040e-01   2.134 0.032851 *
## other_pct      8.949e-02  5.021e-02   1.782 0.074702 .
## median_income  1.457e-06  1.206e-06   1.208 0.226996
## no_insurance_pct 1.560e-03  4.434e-03   0.352 0.724967
##
## Phi coefficients (precision model with identity link):
##      Estimate Std. Error z value Pr(>|z|)
## (phi) 13.1361    0.6835  19.22  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 1406 on 17 Df
## Pseudo R-squared: 0.3177
## Number of iterations: 30 (BFGS) + 7 (Fisher scoring)
## 2014
d2014$vac_rate <- y.transf.betareg(d2014$vac_pct / 100)

mod2014 <- betareg(vac_rate ~ male_pct + median_age +
                    under_5_pct + X5_9_pct +
                    X10_14_pct + X15_19_pct +

```

```

        hispanic_latino_pct +
        white_pct + black_pct +
        aian_pct + asian_pct +
        nhopi_pct + other_pct +
        median_income + no_insurance_pct,
data=d2014)

summary(mod2014, type = "deviance")

##
## Call:
## betareg(formula = vac_rate ~ male_pct + median_age + under_5_pct +
##       X5_9_pct + X10_14_pct + X15_19_pct + hispanic_latino_pct + white_pct +
##       black_pct + aian_pct + asian_pct + nhopi_pct + other_pct + median_income +
##       no_insurance_pct, data = d2014)
##
## Deviance residuals:
##      Min      1Q   Median      3Q      Max
## -11.2458  0.0945  0.4900  0.7186  2.8532
##
## Coefficients (mean model with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -2.483e+01  9.725e+00  -2.553 0.010683 *
## male_pct         5.578e-03  6.513e-03   0.856 0.391735
## median_age       2.051e-02  6.599e-03   3.108 0.001884 **
## under_5_pct      4.439e-02  1.241e-02   3.577 0.000348 ***
## X5_9_pct        -6.799e-03  1.250e-02  -0.544 0.586612
## X10_14_pct       2.308e-02  1.255e-02   1.839 0.065898 .
## X15_19_pct       2.682e-02  1.167e-02   2.298 0.021579 *
## hispanic_latino_pct 2.675e-01  9.681e-02   2.763 0.005727 **
## white_pct        2.481e-01  9.675e-02   2.565 0.010325 *
## black_pct        2.464e-01  9.777e-02   2.520 0.011745 *
## aian_pct         2.280e-01  9.690e-02   2.353 0.018643 *
## asian_pct        2.637e-01  9.700e-02   2.718 0.006565 **
## nhopi_pct        1.745e-01  1.043e-01   1.673 0.094425 .
## other_pct        1.387e-01  4.933e-02   2.812 0.004920 **
## median_income     6.290e-07  1.283e-06   0.490 0.623819
## no_insurance_pct  -1.092e-02  4.914e-03  -2.223 0.026249 *
##
## Phi coefficients (precision model with identity link):
##      Estimate Std. Error z value Pr(>|z|)
## (phi) 10.6950    0.5556  19.25  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 1255 on 17 Df
## Pseudo R-squared: 0.2979
## Number of iterations: 33 (BFGS) + 5 (Fisher scoring)
## 2015
d2015$vac_rate <- y.transf.betareg(d2015$vac_pct / 100)

mod2015 <- betareg(vac_rate ~ male_pct + median_age +

```

```

        under_5_pct + X5_9_pct +
        X10_14_pct + X15_19_pct +
        hispanic_latino_pct +
        white_pct + black_pct +
        aian_pct + asian_pct +
        nhopi_pct + other_pct +
        median_income + no_insurance_pct,
data=d2015)

summary(mod2015, type = "deviance")

##
## Call:
## betareg(formula = vac_rate ~ male_pct + median_age + under_5_pct +
##       X5_9_pct + X10_14_pct + X15_19_pct + hispanic_latino_pct + white_pct +
##       black_pct + aian_pct + asian_pct + nhopi_pct + other_pct + median_income +
##       no_insurance_pct, data = d2015)
##
## Deviance residuals:
##      Min      1Q  Median      3Q      Max
## -6.4740 -0.3324  0.4456  0.6550  2.8230
##
## Coefficients (mean model with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.702e+01  9.051e+00  -1.880 0.060107 .
## male_pct       6.362e-03  6.316e-03   1.007 0.313760
## median_age     2.217e-02  6.660e-03   3.329 0.000871 ***
## under_5_pct    4.755e-02  1.407e-02   3.379 0.000727 ***
## X5_9_pct       3.279e-02  1.357e-02   2.416 0.015697 *
## X10_14_pct    -7.510e-03  1.337e-02  -0.562 0.574333
## X15_19_pct     3.528e-02  1.274e-02   2.769 0.005620 **
## hispanic_latino_pct 1.887e-01  8.980e-02   2.101 0.035639 *
## white_pct      1.705e-01  8.972e-02   1.900 0.057369 .
## black_pct      1.589e-01  9.057e-02   1.754 0.079428 .
## aian_pct       1.816e-01  8.984e-02   2.022 0.043197 *
## asian_pct      1.829e-01  8.991e-02   2.034 0.041924 *
## nhopi_pct      1.029e-01  9.991e-02   1.030 0.303218
## other_pct      1.008e-01  4.635e-02   2.175 0.029663 *
## median_income  1.376e-06  1.234e-06   1.115 0.264991
## no_insurance_pct -7.348e-03  5.445e-03  -1.349 0.177231
##
## Phi coefficients (precision model with identity link):
##      Estimate Std. Error z value Pr(>|z|)
## (phi) 16.2594    0.8633  18.83  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 1490 on 17 Df
## Pseudo R-squared: 0.2696
## Number of iterations: 32 (BFGS) + 8 (Fisher scoring)
## 2016
d2016$vac_rate <- y.transf.betareg(d2016$vac_pct / 100)

```

```

mod2016 <- betareg(vac_rate ~ male_pct + median_age +
                    under_5_pct + X5_9_pct +
                    X10_14_pct + X15_19_pct +
                    hispanic_latino_pct + white_pct +
                    black_pct + aian_pct +
                    asian_pct + nhopi_pct +
                    other_pct + median_income +
                    no_insurance_pct,
                    data=d2016)

summary(mod2016, type = "deviance")

##
## Call:
## betareg(formula = vac_rate ~ male_pct + median_age + under_5_pct +
##         X5_9_pct + X10_14_pct + X15_19_pct + hispanic_latino_pct + white_pct +
##         black_pct + aian_pct + asian_pct + nhopi_pct + other_pct + median_income +
##         no_insurance_pct, data = d2016)
##
## Deviance residuals:
##      Min      1Q  Median      3Q      Max
## -7.4922 -0.3816  0.4242  0.6193  3.1258
##
## Coefficients (mean model with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.879e+01  9.408e+00  -1.997  0.04578 *
## male_pct       2.610e-03  6.008e-03   0.434  0.66396
## median_age     2.149e-02  6.059e-03   3.547  0.00039 ***
## under_5_pct    2.662e-02  1.265e-02   2.104  0.03540 *
## X5_9_pct       2.334e-02  1.234e-02   1.892  0.05855 .
## X10_14_pct    -1.725e-02  1.244e-02  -1.387  0.16548
## X15_19_pct     2.425e-02  1.241e-02   1.955  0.05060 .
## hispanic_latino_pct 2.127e-01  9.418e-02   2.258  0.02395 *
## white_pct      1.940e-01  9.409e-02   2.061  0.03928 *
## black_pct      1.902e-01  9.487e-02   2.005  0.04496 *
## aian_pct       1.992e-01  9.409e-02   2.117  0.03427 *
## asian_pct      2.041e-01  9.415e-02   2.167  0.03020 *
## nhopi_pct      2.792e-01  1.018e-01   2.742  0.00610 **
## other_pct      1.063e-01  4.794e-02   2.218  0.02658 *
## median_income  2.476e-06  1.119e-06   2.212  0.02699 *
## no_insurance_pct -8.686e-03  5.999e-03  -1.448  0.14767
##
## Phi coefficients (precision model with identity link):
##              Estimate Std. Error z value Pr(>|z|)
## (phi)    19.216      1.014    18.94  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 1587 on 17 Df
## Pseudo R-squared: 0.2186
## Number of iterations: 32 (BFGS) + 6 (Fisher scoring)

```

```
## 2017
d2017$vac_rate <- y.transf.betareg(d2017$vac_pct / 100)

mod2017 <- betareg(vac_rate ~ male_pct + median_age +
                    under_5_pct + X5_9_pct +
                    X10_14_pct + X15_19_pct +
                    hispanic_latino_pct +
                    white_pct + black_pct +
                    aian_pct + asian_pct +
                    nhopi_pct + other_pct +
                    median_income + no_insurance_pct,
                    data=d2017)

summary(mod2017, type = "deviance")

##
## Call:
## betareg(formula = vac_rate ~ male_pct + median_age + under_5_pct +
##         X5_9_pct + X10_14_pct + X15_19_pct + hispanic_latino_pct + white_pct +
##         black_pct + aian_pct + asian_pct + nhopi_pct + other_pct + median_income +
##         no_insurance_pct, data = d2017)
##
## Deviance residuals:
##      Min      1Q   Median      3Q      Max
## -13.6295  0.0909  0.3153  0.4278  1.6984
##
## Coefficients (mean model with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.121e+01  5.622e+00  -1.995  0.04606 *
## male_pct       1.038e-03  3.962e-03   0.262  0.79328
## median_age    -1.960e-03  3.954e-03  -0.496  0.62005
## under_5_pct    1.689e-02  8.337e-03   2.026  0.04279 *
## X5_9_pct      -1.926e-02  8.255e-03  -2.334  0.01962 *
## X10_14_pct     1.641e-02  8.319e-03   1.973  0.04849 *
## X15_19_pct     6.875e-03  8.106e-03   0.848  0.39639
## hispanic_latino_pct 1.429e-01  5.613e-02   2.547  0.01087 *
## white_pct      1.334e-01  5.609e-02   2.378  0.01740 *
## black_pct      1.448e-01  5.655e-02   2.560  0.01046 *
## aian_pct       1.366e-01  5.608e-02   2.437  0.01482 *
## asian_pct      1.437e-01  5.614e-02   2.559  0.01050 *
## nhopi_pct      1.157e-01  6.228e-02   1.857  0.06326 .
## other_pct      7.064e-02  2.856e-02   2.473  0.01339 *
## median_income  3.454e-06  7.005e-07   4.931 8.17e-07 ***
## no_insurance_pct -9.923e-03  3.828e-03  -2.592  0.00954 **
##
## Phi coefficients (precision model with identity link):
##              Estimate Std. Error z value Pr(>|z|)
## (phi)    34.530      1.279      27    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 3128 on 17 Df
## Pseudo R-squared: 0.2562
```

Number of iterations: 34 (BFGS) + 4 (Fisher scoring)