# Logistic + Poisson Models

Shawn Fries

# Logistic model for Alzheimer's diagnosis

- Can't just blindly use every variable; certains variables are redundant (perfectly multicollinear) or nearly so (e.g. redundant columns like WHODIDDX and DXMETHOD, and the different cognitive function variables are worth a close look):

```
> print(vif(alzheimers_model))
                     GVIF Df GVIF^(1/(2*Df))
factor(apoe)     1.435405  5        1.036806
cdr_memory       1.892249  1        1.375590
cdr_orientation  1.640946  1        1.280994
cdr_judgment     1.626847  1        1.275479
cdr_commun       1.966447  1        1.402301
cdr_homehobb     2.229619  1        1.493191
cdr_perscare     1.566590  1        1.251635
cdr_global       2.184471  1        1.477996
gds_15           1.059261  1        1.029204
MMSELAN          1.093752  1        1.045826
mmse_time        1.765170  1        1.328597
mmse_place       1.422324  1        1.192612
mmse             2.237386  1        1.495789
animal           1.749781  1        1.322793
vegetable        1.748620  1        1.322354
factor(DXMETHOD) 1.139365  3        1.021983
factor(ethnicity) 1.291906  6        1.021573
age_visit        1.185519  1        1.088815
sex              1.223339  1        1.106047
educ             1.106424  1        1.051867
```

# Logistic (not using too many variables)

```
> print(vif(alzheimers_model_simplified))
          apoe      cdr_memory cdr_orientation    cdr_judgment      cdr_commun      cdr_global            mmse           anima
      1.034390        1.677119        1.458887        1.474171        1.547833        1.711366        1.238774         1.17964
      DXMETHOD       ethnicity
      1.077291        1.044188
```

- Alternatively, generally not worth using variables that aren't significant on their own (z-test e.g. from base R summary of model)
- Results from model with too many variables:

```
AIC: 493.83
```

```
> PseudoR2(alzheimers_model)
 McFadden
0.8044248
```

- Results from slightly simplified model (no multicollinearity) (low AIC is better)

```
AIC: 479.1
```

```
> PseudoR2(alzheimers_model_simplified)
 McFadden
0.7920167
```

# Poisson regression for count of vegetables named

- Estimates log of expected value of mean and variance of the number of vegetables a patient named in 60 seconds

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.8764494  0.0232874 123.519  < 2e-16 ***
DEMENTED    -0.5744257  0.0235202 -24.423  < 2e-16 ***
ethnicity   -0.0035131  0.0005161  -6.807 9.95e-12 ***
age_visit   -0.0074711  0.0002565 -29.130  < 2e-16 ***
sex          0.1796270  0.0065301  27.508  < 2e-16 ***
educ         0.0029733  0.0004008   7.419 1.18e-13 ***
---
```

```
AIC: 43700

Number of Fisher Scoring iterations: 4


> PseudoR2(vegetable_model_simpler)
  McFadden
0.05906536
```

- Not the best model because the data turn out to be closer to normally distributed than Poisson