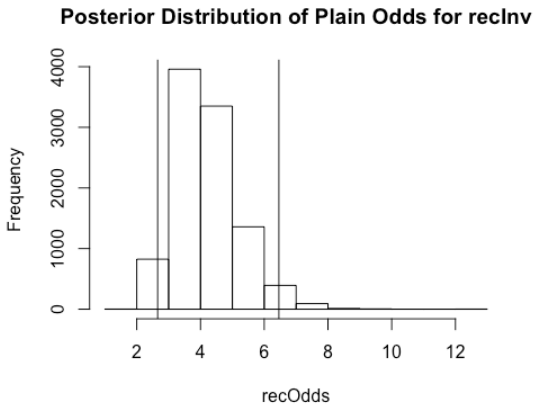
Homework

- The homework for week 9 is based on exercises 1, 5, 6, and 7 on page 234 but with changes as noted in the notebook.
- Second practice exam for the final (fourth one for the course) is today

7

Bayesian results predicting hiredata$hired

```
> summary(exp(bayesLogitOut))

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each
variable, plus standard error of the mean:

               Mean       SD  Naive SE Time-series SE
(Intercept) 0.01748 0.009976 9.976e-05      0.0002938
recInv      4.18235 0.984962 9.850e-03      0.0303707

2. Quantiles for each variable:

              2.5%    25%     50%     75%  97.5%
(Intercept) 0.0046 0.0104 0.01551 0.02188 0.0423
recInv      2.6555 3.4897 4.04834 4.73614 6.4598
```



5

Paper of the Week – Hsieh, Block, & Larsen 1998

- Power Analysis: How much data do you need to collect in order to have a good chance of affirming your hypothesis?
- Alpha is the level of Type 1 error, beta is the level of Type II error and 1 - beta is power; Convention: Go for 0.80

STATISTICS IN MEDICINE
Statist. Med. **17**, 1623–1634 (1998)

A SIMPLE METHOD OF SAMPLE SIZE CALCULATION FOR LINEAR AND LOGISTIC REGRESSION

F. Y. HSIEH[1]*, DANIEL A. BLOCH[2] AND MICHAEL D. LARSEN[3]

6

Breakout – Run Logistic Regression Models

- Open `2. week_9_logistic_clinic.Rmd`
- Read in the hiring data
- Create an initial model using the hiring recommendation to predict the actual hiring decision
- Add a Bayesian model using MCMCpack
- Find a second predictor that improves the prediction quality
- Share your code on https://codeshare.io/aJDyRX

All handouts for this class: **https://tinyurl.com/IST772crowston**

3

Results from predicting hiredata$hired

```
> PseudoR2(glmOut)

Nagelkerke
0.21358554
```

```
> anova(glmOut, test="Chisq")
Analysis of Deviance Table
Model: binomial, link: logit

Response: hired

Terms added sequentially (first to last)


          Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
NULL                        294     332.33
recommend  1   45.989       293     286.34 1.189e-11 ***
```

```
> exp(coef(glmOut))
(Intercept)   recommend
  3.9962895   0.2513514
> exp(confint(glmOut))
Waiting for profiling to be
done...
                 2.5 %     97.5 %
(Intercept) 1.8465086 8.8680474
recommend   0.1582932 0.3856539
```

Why is the whole span of the CI smaller than 1?

4

All handouts for this class: **https://tinyurl.com/IST772crowston**

IST772 Logistic Regression (Week 9)

Pre-class activity:

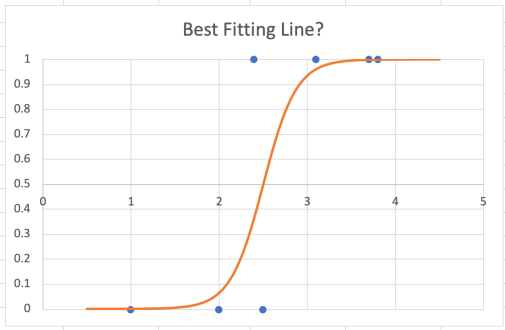
- Grab a copy of 1. Week9logistic.xlsx from the handouts area
- Adjust the coefficient with the buttons to see the effects on the deviance residuals
- Examine the prediction equation

1

Logistic Demonstration

Log odds when X = 0.

	A	B	C	D	E	F	G	H	I	J	K
1	X	Y	a + bX	logit'(a+bX)=P(X)	Log Likelihood						
2	1	0	-8.019	0.000	0.000		intercept (a)=	-13.359	Pre-calculated		
3	2	0	-2.679	0.064	-0.066		coefficient (b)=	5.340	Controlled by up/down		
4	2.4	1	-0.543	0.367	-1.001						
5	2.5	0	-0.009	0.498	-0.688						
6	3.1	1	3.195	0.961	-0.040						
7	3.7	1	6.399	0.998	-0.002						
8	3.8	1	6.933	0.999	-0.001						
9											
10				Total Log Likelihood	-1.799						
11					Maximize						
12											
13	Change										
14	Coefficient										
15											
16											
17											



2