Agenda

1. Presentations and final reports
2. A little more logistic regression
3. Putting it all together

Project Some important things to keep in mind:

- For the presentation, you will only have 8 minutes. Going over time is a surefire way to lose points. Going under time is fine.
- Please focus your presentation only on one model (your current favorite). You should convince us the problem is interesting, interpret the coefficients in the context of the problem, and (briefly!) show us the conditions for regression.
- We'll spend some time next week talking about ways to make your final project report look pretty, but remembering `message=FALSE` is a big one.

Coefficient interpretation Because interpreting coefficients is so important for both your project and your test, lets practice writing out the "script" for a variety of cases.

1. Multiple regression: categorical variable

2. Multiple regression: numeric variable

3. Multiple regression: interaction term

4. ANOVA α_i

5. ANOVA μ_i

6. Logistic regression: categorical variable

7. Logistic regression: numeric variable

8. Logistic regression: interaction term

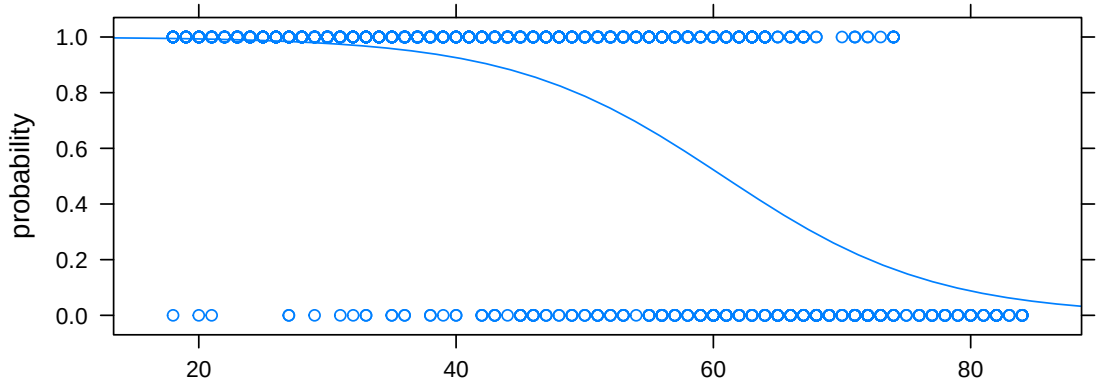
Logistic regression Let's pull the example of med school acceptance through the three 'spaces' for logistic regression to see how it applies.

```
##            Sex
## Acceptance  F  M
##          0 10 15
##          1 18 12
```

Probability

$$\pi = \frac{e^{\beta_0+\beta_1 \cdot X}}{1+e^{\beta_0+\beta_1 \cdot X}}$$

The probability form is how the model gets fit, but it does not have an easy interpretation for what happens with a change in x.

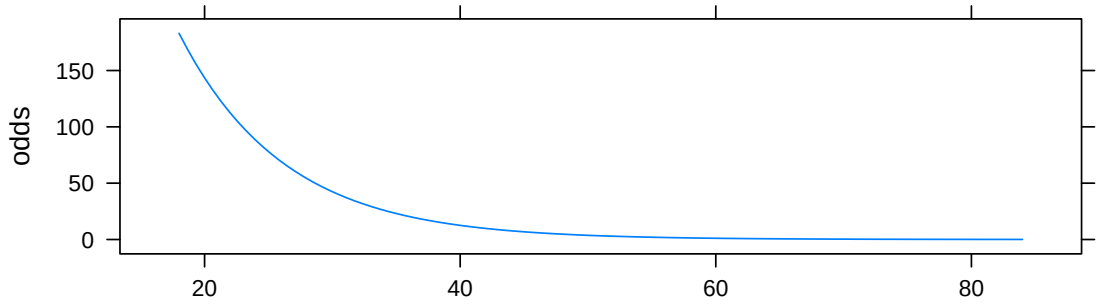


What is the probability of getting into med school for a woman? For a man?

Odds

$$\frac{\pi}{1-\pi} = e^{\beta_0+\beta_1 \cdot X}$$

Odds are useful when you want to interpret the slope coefficient. We can use the interpretation, "A one unit increase in x is associated with changing the odds by a factor of e^{β_1}.

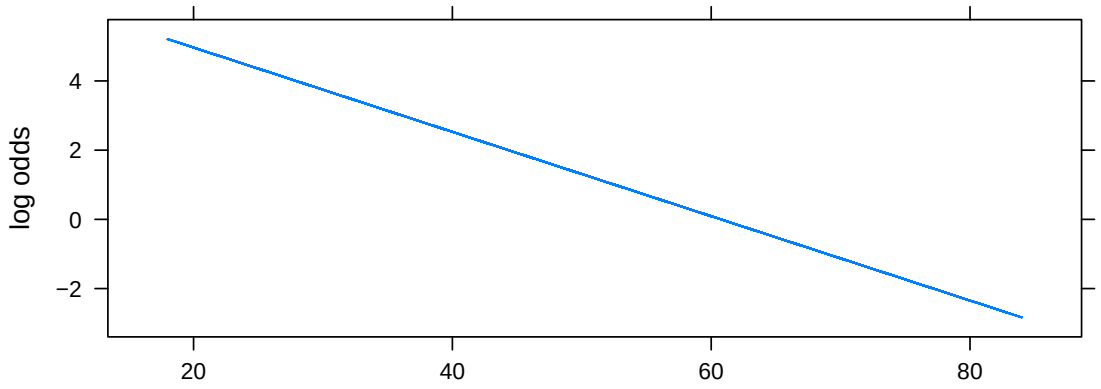


What are the odds of getting into med school for a woman? For a man? What is the odds ratio?

Log odds

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 \cdot X$$

Thinking about log odds is useful when you want a linear form of a regression line. You can interpret the coefficients in the standard way we have been doing for linear regression, "A one unit increase in x is associated with a β_1 increase in the log odds of y"



What is the empirical logit for getting into med school for a woman? For a man?

```
##
## Call:
## glm(formula = Acceptance ~ Sex, family = binomial, data = MedGPA)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.435  -1.084   0.940   0.940   1.274
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.5878     0.3944   1.490    0.136
## SexM         -0.8109     0.5528  -1.467    0.142
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 75.791  on 54  degrees of freedom
## Residual deviance: 73.594  on 53  degrees of freedom
## AIC: 77.594
##
## Number of Fisher Scoring iterations: 4
```

Write out the equation for the model in each of the three formats.

How do you interpret the coefficient on `SexM`?

Putting it all together Lets consider the material we have covered throughout this course. We have tried to focus mostly on applied ideas, some of which are computational (more useful in practice, hard to test) and some which are slightly more theoretical (needed so you understand what you can apply, easier to test).

Lets try to categorize some of the tasks we have learned how to perform. Where in the CFAU process do these fit?
Interpreting multiple regression coefficients, nested F-tests, added variable plots, stepwise regression, calculating Cook's distance, calculating Fisher's LSD, calculating Tukey's HSD, the Bonferroni correction, randomization tests, bootstrap methods, fitting ANOVA, rewriting ANOVA models in different ways, checking ANOVA p-values, fitting a logistic regression model, translating from log-odds to odds, translating from odds to probability, calculating odds ratios, looking at p-values on logistic regression, interpreting logistic regression coefficients, considering forms of models.
Choose

Fit

Assess

Use