Scribe Report – April 28, 2014

1. BaseballBets.CSV
    - First we plot the Home Team Victory Win (1 or 0) and the point spread
        - Noted that the better odds with the greater point spread
    - Second we will place the data into buckets
        - Put the victory into buckets based on the point spread
    - Review: Fitted values
        - $\widehat{y_i} = B_0 + B_1 x_i$
        - Where $\widehat{y_i}$ fitted value is the conditional expected value
    - Expected value for binary outcome is the weighted average of the expected probabilities
        - For a binary outcome

x_i	w_i
0: Loss	Nothing
1: Win	$B_0 + B_1 x_i$

        - E(Y|X) = $B_0 + B_1 x_i$ + 0(-)
            - = $B_0 + B_1 x_i$
        - Conditional expected value of a binary outcome is exactly the probability of getting a 1
    - Fit the model with the with the binary outcome for victory – we see that the home team wins 52% of the time with a 2% increase with every point spread increase
    - Fit the simple linear model
    - Issues:
        - Probabilities are against the rules of mathematics
        - Problem – model on the right side can take any real number but it should only be between 0 and 1 – probabilities are between 0 and 1
            - Pr(y_i= 1 | x_i) should be on (0,1)
            - But $B_0 + B_1 x_i$ - can only be a real number
        - We need our regression function to have a "speed limit" so it does not go past 0 and 1
        - Make an S shaped curve so the probabilities are not outside 0 and 1
    - Logistic Regression

- $\Pr(y_i = 1 \mid x_i) = \frac{e^{(B_0 + B_1 x_i)}}{1 + e^{(B_0 + B_1 x_i)}}$
- $= g(B_0 + B_1 x_i)$ ← speed limit function or "link function"
- where $g(s) = \frac{e^{(s)}}{1 + e^{(s)}}$
- Use the link function so that way you cannot go above 1 or below 0

- In R:
  - glm ← generalized linear function
  - family ← binomial
  - When imposed using the link function you get an S shaped curve

```
# A simple fix: fit a logistic regression model instead
glm1 = glm(homewin~spread, data=bballbets, family=binomial)
```