STATS 331

Welcome back!

Introduction to Bayesian Statistics
Semester 2, 2016

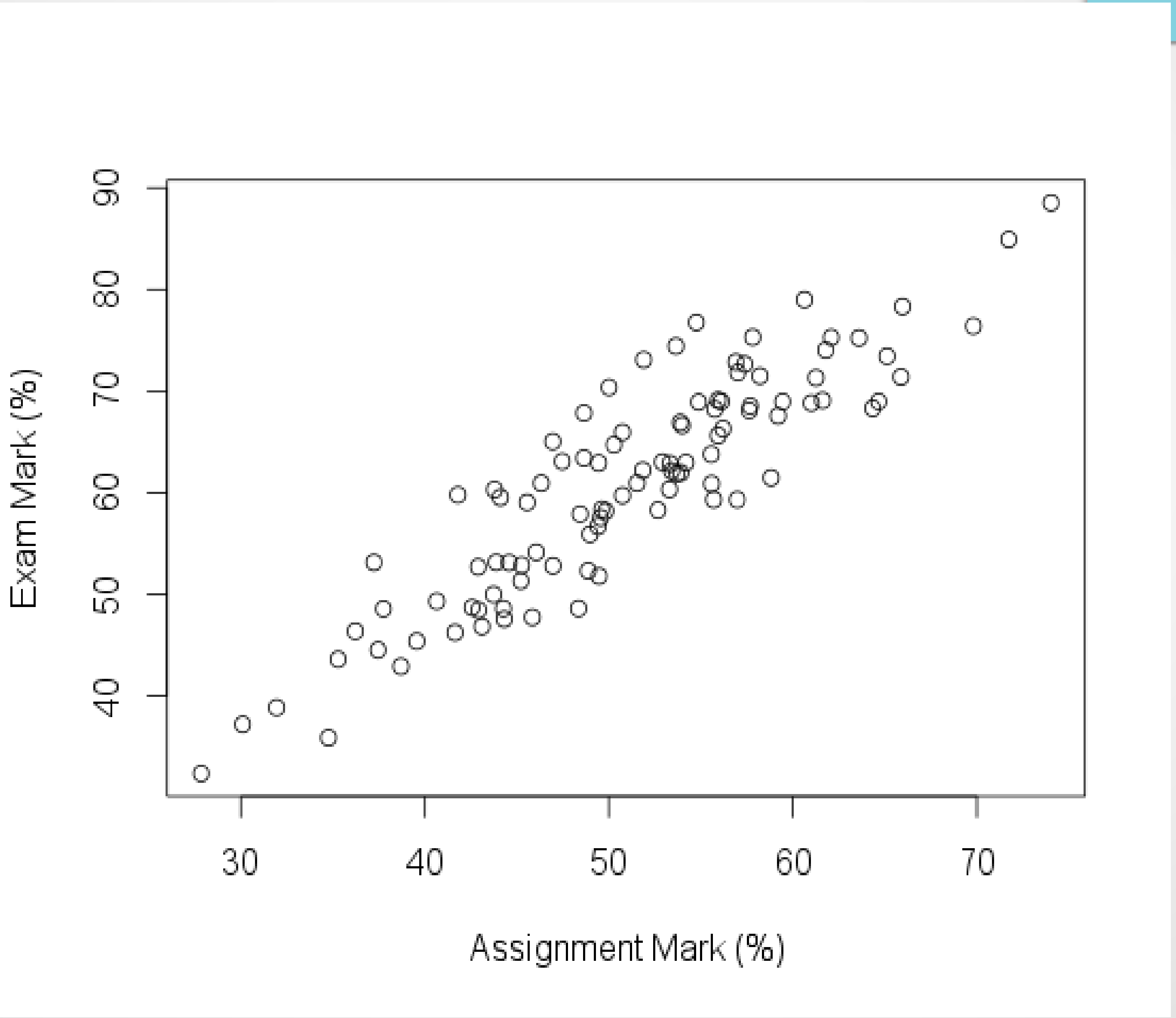
Today's Lecture

Bayesian Linear Regression

It's All Models Now

- We have seen most of the basic principles
- Now it's application time!

Linear Regression, aka “Fitting a straight line”



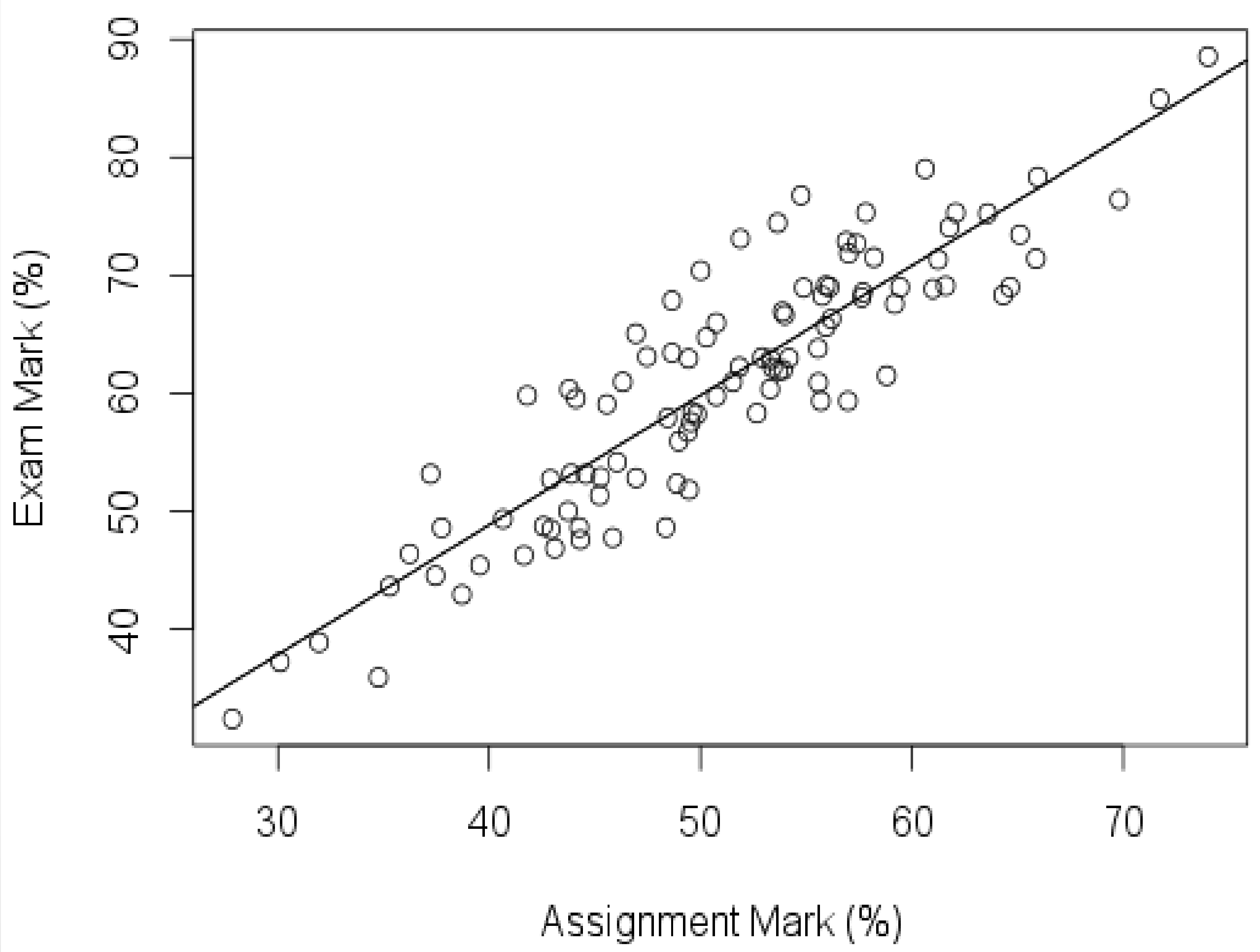
Least Squares

- Conventional estimator for slope and intercept

```
reg = lm(y ~ x)

summary(reg)

Call:

lm(formula = y ~ x)

Residuals:

    Min      1Q  Median      3Q     Max

-9.3842 -3.0688 -0.6975  2.6970 11.7309

Coefficients:

            Estimate Std. Error t value Pr(>|t|)    (Intercept)
   4.83805    2.79361   1.732   0.0865

  x            1.09947    0.05386  20.412   <2e-16
```

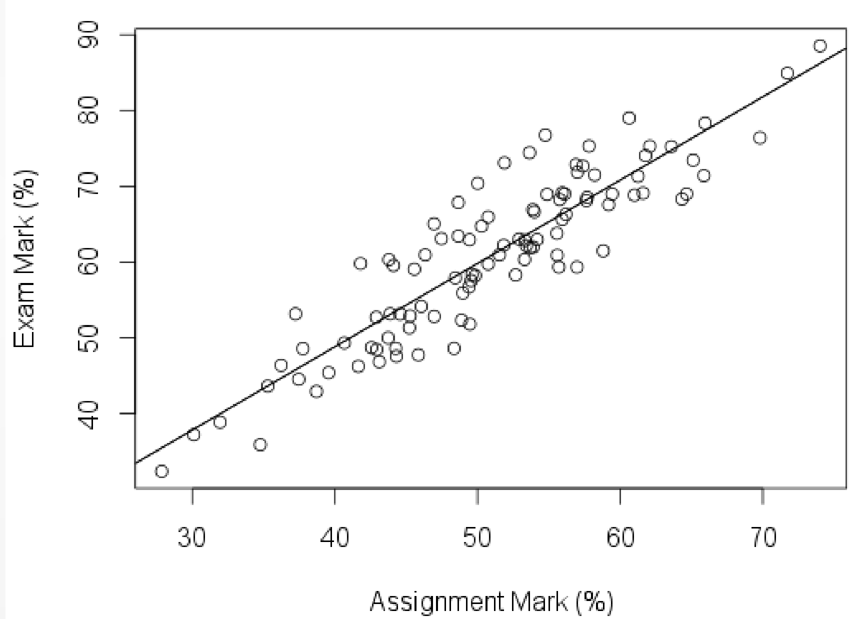
Line of Best Fit

```
abline(reg)
```



Prediction

- We have a nice point estimate
- Can predict new data (aka extrapolate)
- Put standard deviation around best fit line prediction

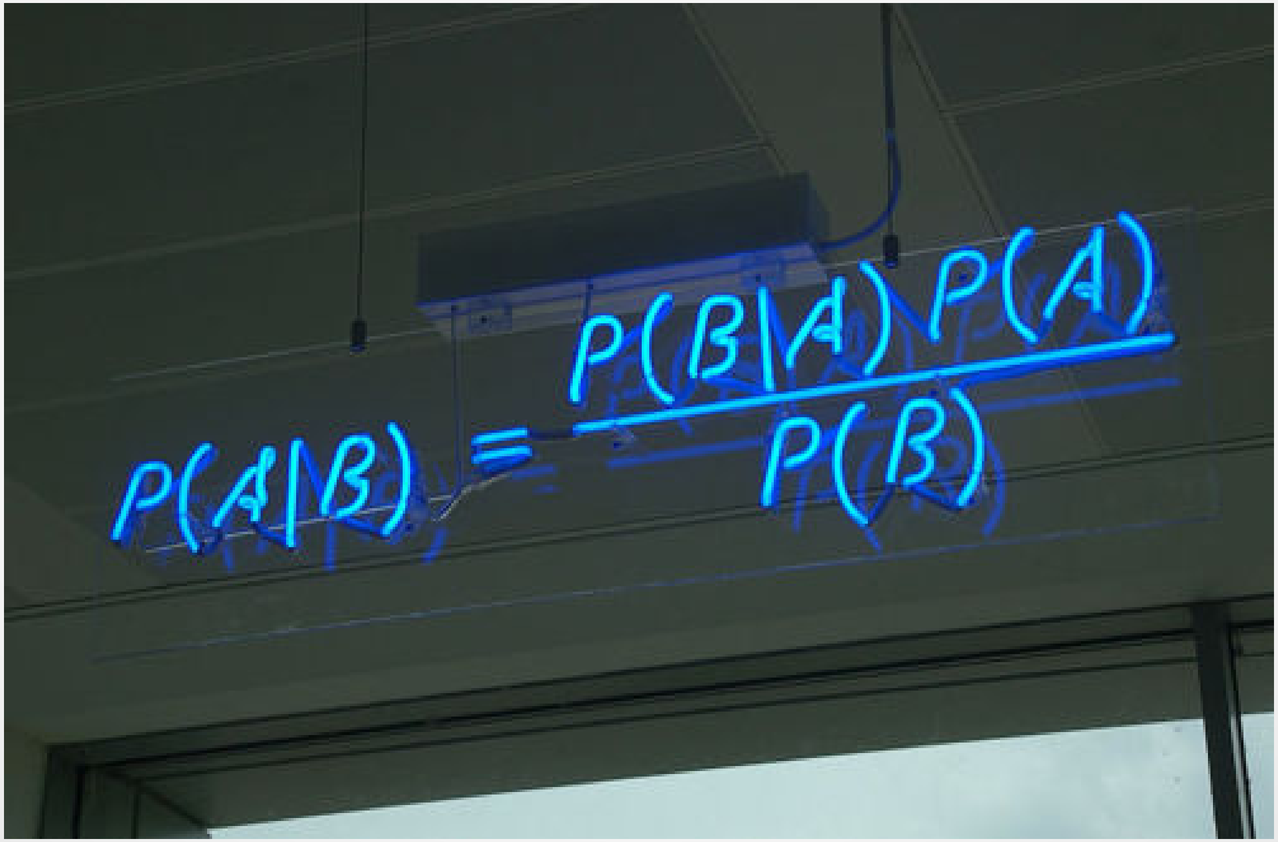


However...

This doesn't account for the uncertainty about the parameters*!

*There are classical ways to do this but we won’t be discussing them

Bayesian Approach



What is the Question?

- Want to infer the intercept β_0 and the slope β_1
- Have data $\{y_1, y_2, \ldots, y_N\}$

  [and prior information N and $\{x_1, x_2, \ldots, x_N\}$]

- Use Bayesian parameter estimation

Need a Prior

- Want to infer the intercept β_0 and the slope β_1

$$p(\beta_0, \beta_1)$$

Sampling Distribution/Likelihood

- If we knew the parameters, how would we predict data?

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Or...

$$y_i|\beta_0, \beta_1 \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2)$$

If we knew the slope and intercept of our straight line, then our probability distribution for the data would be a normal distribution around the straight line.

Analytical

- Bayes' rule

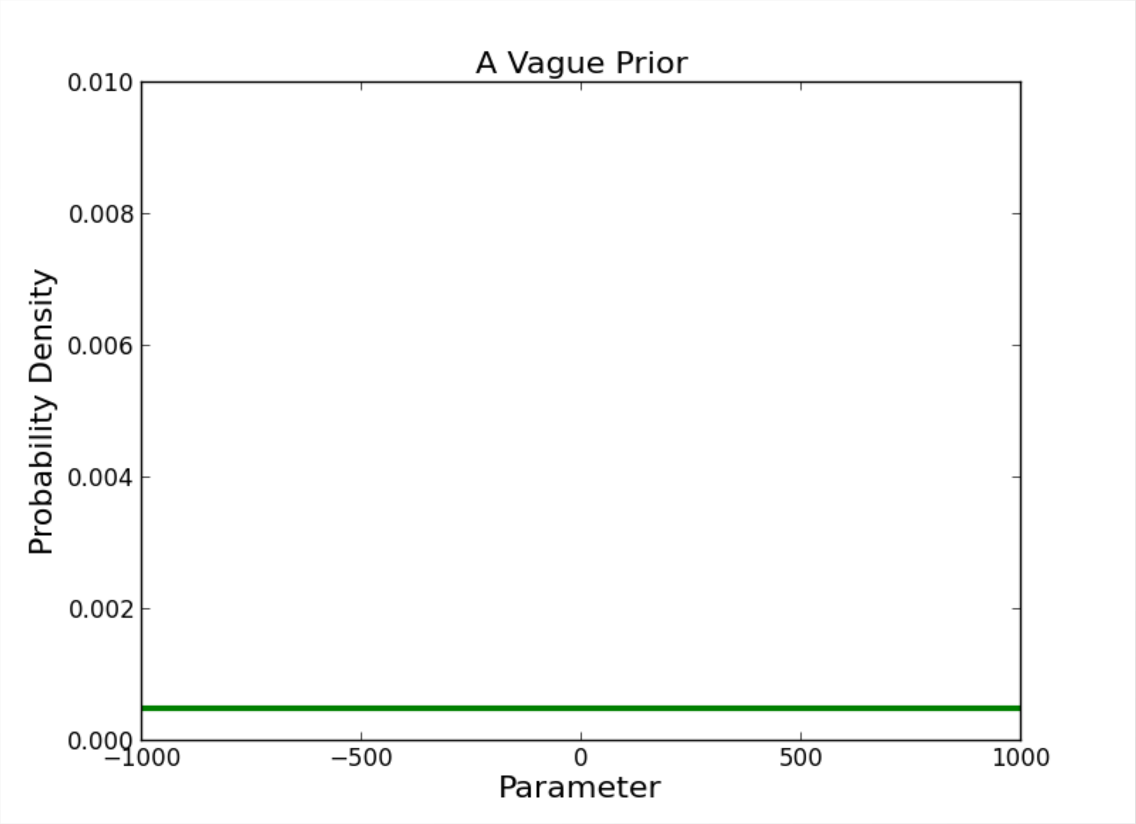
posterior α prior x likelihood

$$p(\beta_0, \beta_1 | y_1, y_2, ..., y_N) \propto p(\beta_0, \beta_1) \times p(y_1, y_2, ..., y_N | \beta_0, \beta_1)$$

Choice of Prior

- Let's be naïve

$$p(\beta_0, \beta_1) \propto 1$$



If we don't specify the endpoints, this is called an “improper” prior.

Likelihood

$$p(y_1, y_2, \ldots, y_N | \beta_0, \beta_1) \quad = \quad \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right]$$

$$\propto \quad \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right]$$

Prior x Likelihood

- Just proportional to likelihood in this case, due to uniform prior

- Parameter values that result in *small residuals* are **more probable**

$$\propto \quad \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right]$$

Least Squares

- The posterior mode is the least squares estimate!

Can interpret classical method as implicitly making certain assumptions i.e. flat prior, posterior mode, known standard deviation

Implementation in JAGS

- We can implement Bayesian linear regression in JAGS, including making the standard deviation unknown

Simple Linear Regression – JAGS Model

```
model
{
    beta0 ~ dnorm(0, 1/1000^2)
    beta1 ~ dnorm(0, 1/1000^2)

    log_sigma ~ dunif(-10, 10)
    sigma <- exp(log_sigma)
    for(i in 1:N)
    {
      mu[i] <-  beta0 + beta1*x[i]
      y[i] ~ dnorm(mu[i], 1/sigma^2)
    }
}
```

Normal Distributions in JAGS

- The normal distribution is available with dnorm

- The first argument is the mean, the second argument is 1/(standard deviation)^2   [sometimes called the “precision”]

Over to RStudio

- Let's use the simple linear regression model on the 20X 'road' dataset

SHOUTING

- TEST ON WEDNESDAY!
- BRING CALCULATOR AND PENS!
- ARRIVE ON TIME AT THE CORRECT ROOM!
- GOOD LUCK!