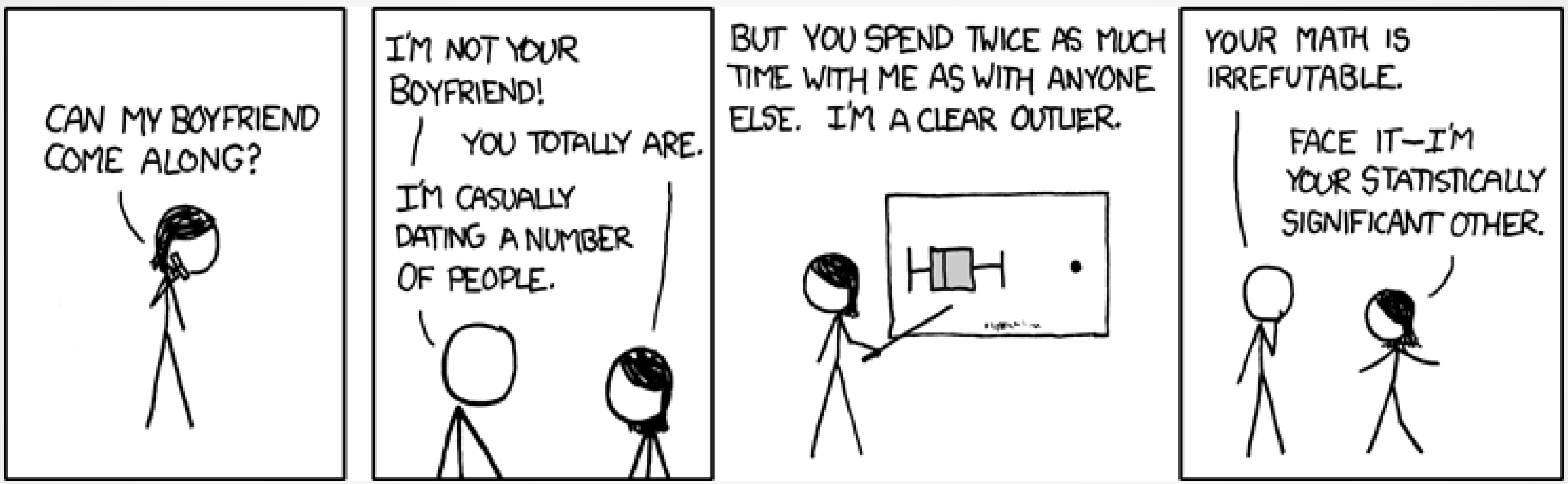
STATS 331



Introduction to Bayesian Statistics
Semester 2, 2016

Today's Lecture

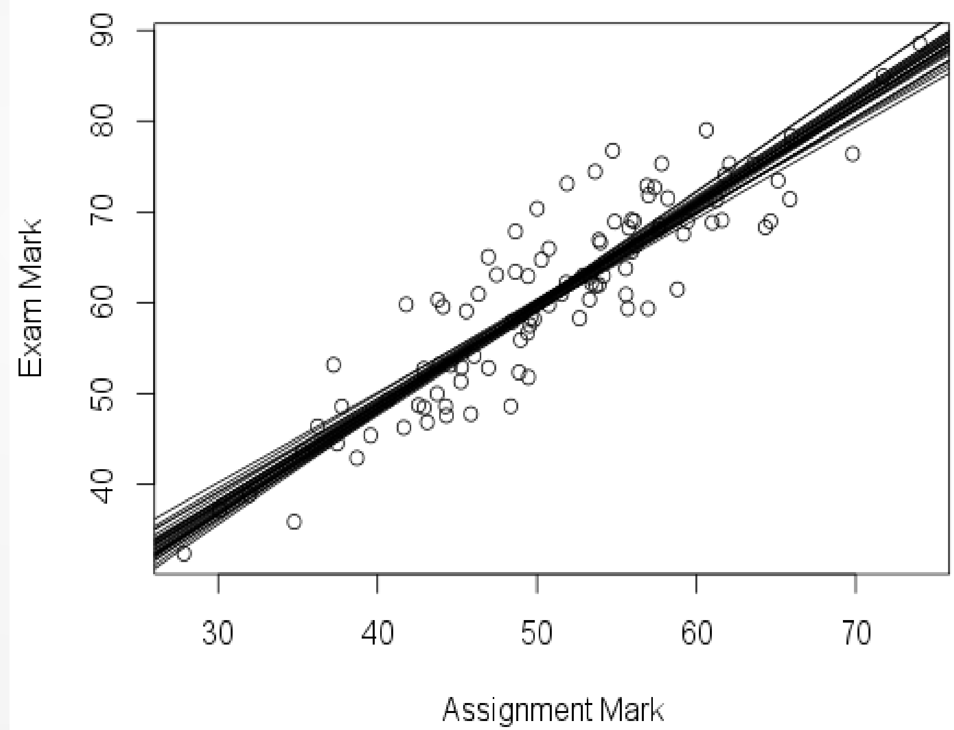
- Midterm test comments
- Multiple linear regression
- Nonlinear regression

Simple Linear Regression

"Fitting a straight line"

Least squares == posterior mode (with specific assumptions)

We developed a JAGS model which can give us the full posterior distribution



Multiple Linear Regression!

Just like simple linear regression, but with more than one "input variable"

Output = linear combination of inputs + scatter

Fitting a straight plane (-:

Regression Examples

(height, weight) → (age at death)

(team height, distance from home city) → (basketball score)

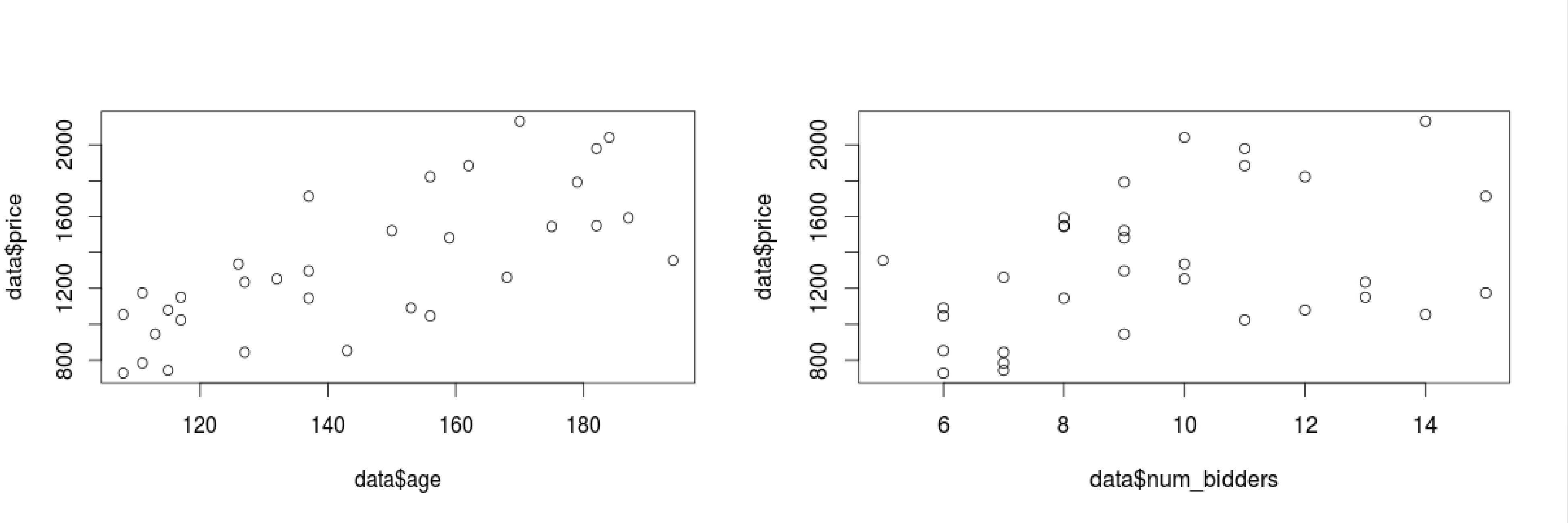
(number of mentions of Nigeria, number of mentions of Viagra) → (spam or not-spam)

…

Models like this are very important! There are whole courses about them (e.g. STATS 330)

Example: Really Important Science

How does the age of a grandfather clock, and the number of bidders, affect the sale price?



Modelling the relationship

y_i = output/response (sale price of clock)

$x_{i,1}$ = input 1 (age of clock)

$x_{i,2}$ = input 2 (number of bidders)

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i$$

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

What is the Question?

Want to infer β_0, β_1 and β_2

Will need prior for these!

From data $y_1, y_2, \dots, y_N$.

Prior info $x_{1,1}, \dots, x_{N,1}$ (ages), $x_{2,2}, \dots, x_{N,2}$, (number of bidders), N

Code Up Likelihood in JAGS

```
for(i in 1:N)
{
    mu[i] <- beta0 + beta1*age[i] +
      beta2*num_bidders[i]
    y[i] ~ dnorm(mu[i], 1/sigma^2)
}
```

Vague Priors

```
beta0 ~ dnorm(0, 1/1000^2)
beta1 ~ dnorm(0, 1/1000^2)
beta2 ~ dnorm(0, 1/1000^2)

log_sigma ~ dunif(-10, 10)
sigma <- exp(log_sigma)
```

Over to Rstudio

Warnings about priors

- If the number of input variables is large, you need a prior for a large number of coefficients
- Vague priors can become more dodgy as the number of parameters increases!
- "Hierarchical priors" often needed – first example tomorrow.

Interaction Term

Something might happen only if both predictors are "switched on"

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,1} x_{i,2} + \epsilon_i$$

Interaction Term

Not fitting a plane anymore...the surface can curve

Interaction Term: Changes

```
beta3 ~ dnorm(0, 1/1000^2)
for(i in 1:N)
{
    mu[i] <- beta0 + beta1*age[i] +
  beta2*num_bidders[i] +
  beta3*age[i]*num_bidders[i]
}
```

Results with interaction term...

- Check out the trace plots. Things are starting to get unhappy due to a very strong correlation in the posterior
- For now, brute force (more steps) will work...
- `beta1` is now practically zero and `beta2` negative! Everything has gone into the interaction term.

Trick to improve MCMC efficiency

In regression models, *centering* the variables can help.

Note: meaning of `beta0` changes.

```
mu[i] <- 1327 + beta0 + beta1*(age[i] - 145) +
  beta2*(num_bidders[i] – 9.5) + beta3*(age[i] -
  145)*(num_bidders[i] – 9.5)
```

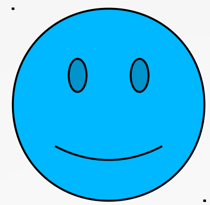
145 is the mean age

9.5 is the mean number of bidders

1327 is the mean sale price

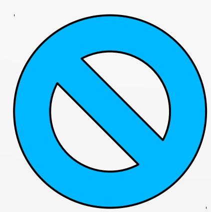
Can we do model selection?

$$H_0 : \beta_3 > 0$$
$$H_1 : \beta_3 < 0$$



Why? Prior prob of exactly zero is either
1 (in the first model) or
0 (in the second model)

$$H_0 : \beta_3 = 0$$
$$H_1 : \beta_3 \neq 0$$



Could use marginal likelihoods,
but JAGS can't calculate them!

Ideas

Look at marginalised posterior - is there significant probability *density* at 0?

If yes, the model without the term would probably be fine.

Another idea: testing prior in JAGS

Can we make a prior with which allows beta3 to be either zero (with, say, 50% prior probability) or non-zero?

Yes!

We'll do this next week during the “t-test” section.

See You Tomorrow!

See you tomorrow!