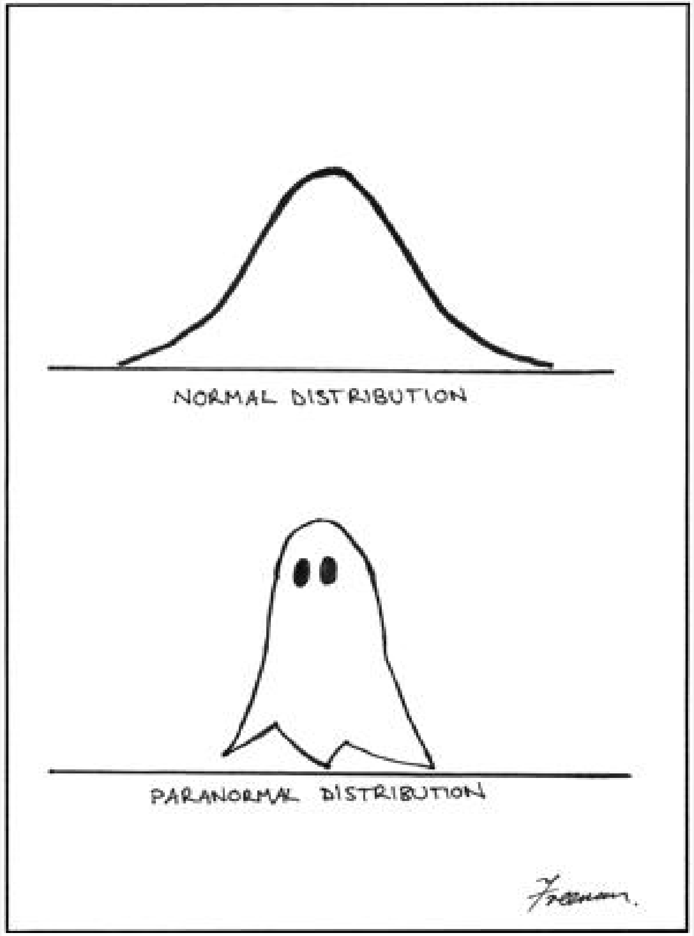
STATS 331

By Matthew Freeman
(Flinders University)

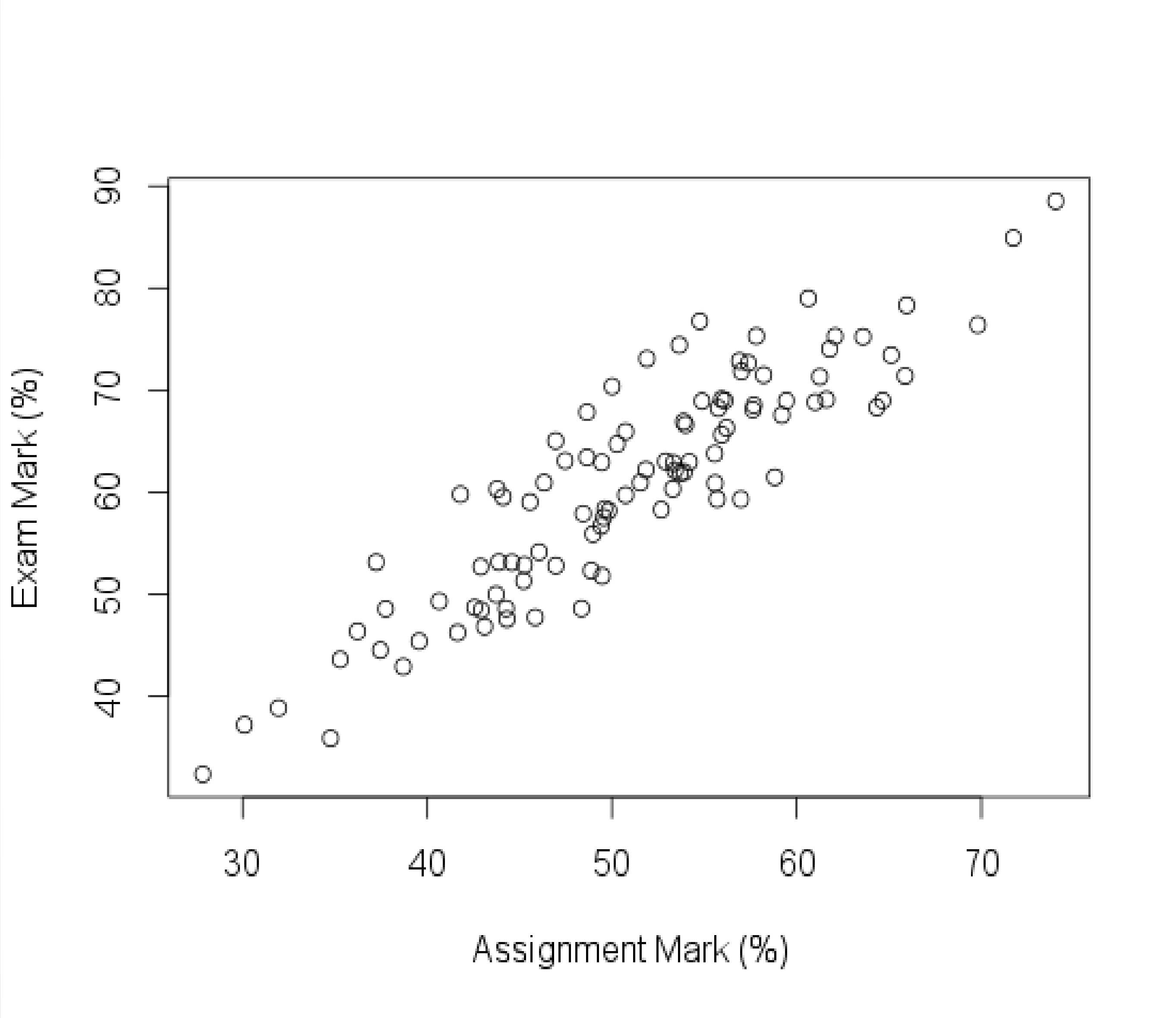


Introduction to Bayesian Statistics
Semester 2, 2016

Today's Lecture

- Prediction in JAGS
- Bayesian ‘robustness’ and ‘outliers’

aka “Fitting a straight line”



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

Prediction

If we knew the parameters...

we could predict y_{new} using the sampling distribution:

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

Predictive Distribution

$$
\begin{aligned}
p(y_{\text{new}}|y_1,\ldots,y_N) &= \int p(y_{\text{new}},\beta_0,\beta_1|y_1,\ldots,y_N)\,d\beta_0\,d\beta_1 \\
&= \int p(\beta_0,\beta_1|y_1,\ldots,y_N)p(y_{\text{new}}|\beta_0,\beta_1,y_1,\ldots,y_N)\,d\beta_0\,d\beta_1 \\
&= \int p(\beta_0,\beta_1|y_1,\ldots,y_N)p(y_{\text{new}}|\beta_0,\beta_1)\,d\beta_0\,d\beta_1
\end{aligned}
$$

- Do prediction conditional on parameters
- Average over all possible parameter values using the posterior (end up with a *mixture* distribution)

Prediction in JAGS

- Suppose we wanted to predict `y_new` at `x=90` (aka extrapolation)

A New Node

```
model
{
      # Other stuff...

      y_new ~ dnorm(beta0*90 + beta1,
         1/sigma^2)
}
```

Prediction in JAGS

- Doing prediction in JAGS is like adding *one extra data point, with a different name* to the model

- How it actually works: Each iteration of the MCMC, based on the current values of the parameters `beta0`, `beta1`, and sigma, JAGS will simulate a value for `y_new`.

- Don't forget to “monitor” the new variable!

Over to RStudio

- Let's do the prediction

- The uncertainty in the prediction (e.g. measured by the posterior standard deviation) is generally *larger* than a classical point estimate of `sigma` – why?

Outliers

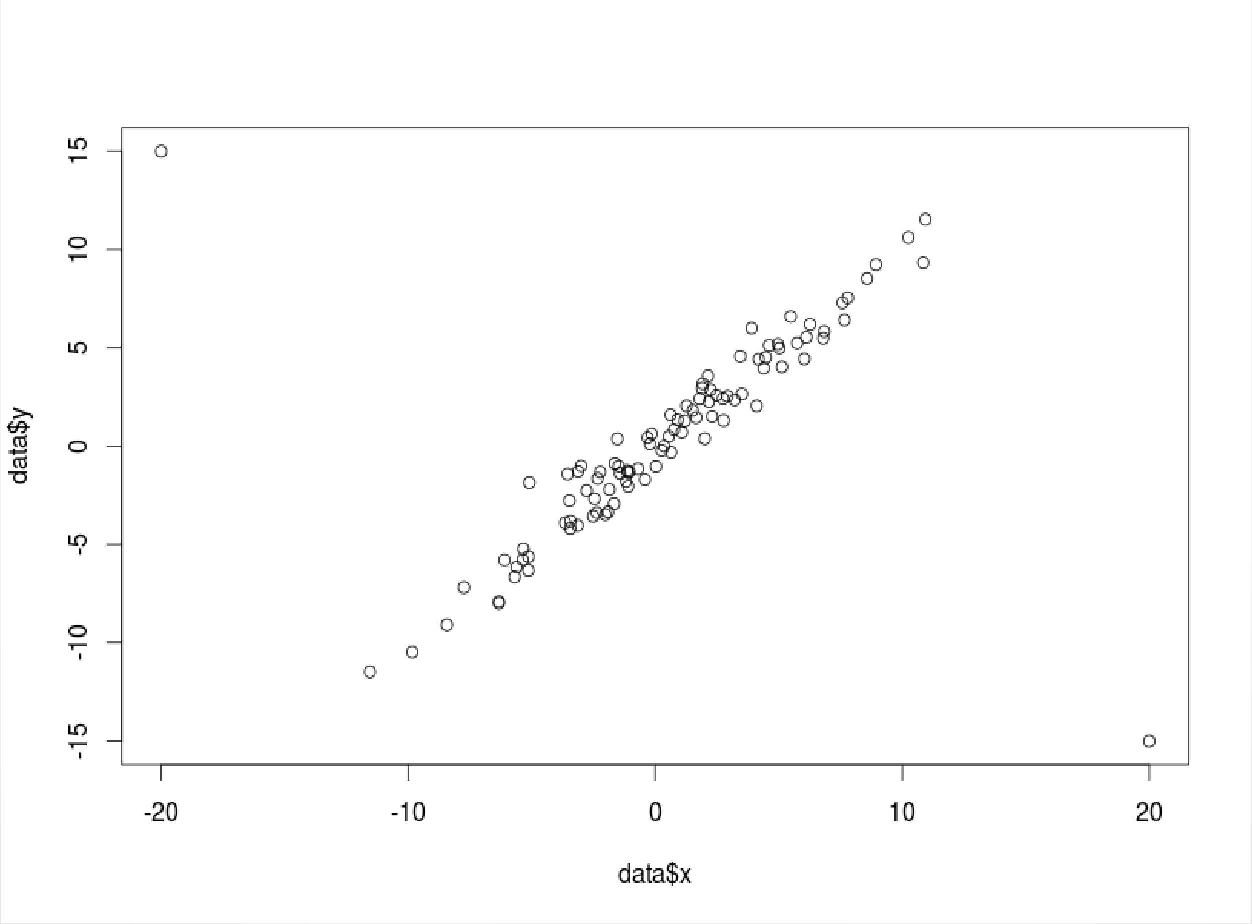
Outliers

- Outliers are an important aspect of real data sets
- There are many methods for “detecting” and “removing” outliers
- We will look at *one* way of handling outliers in Bayesian data analysis

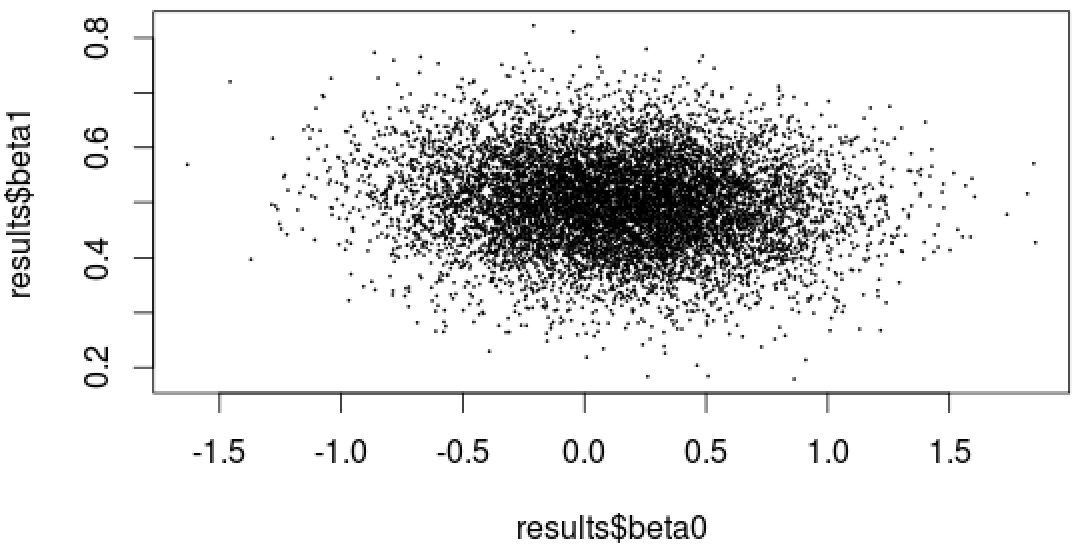
Data with outlier

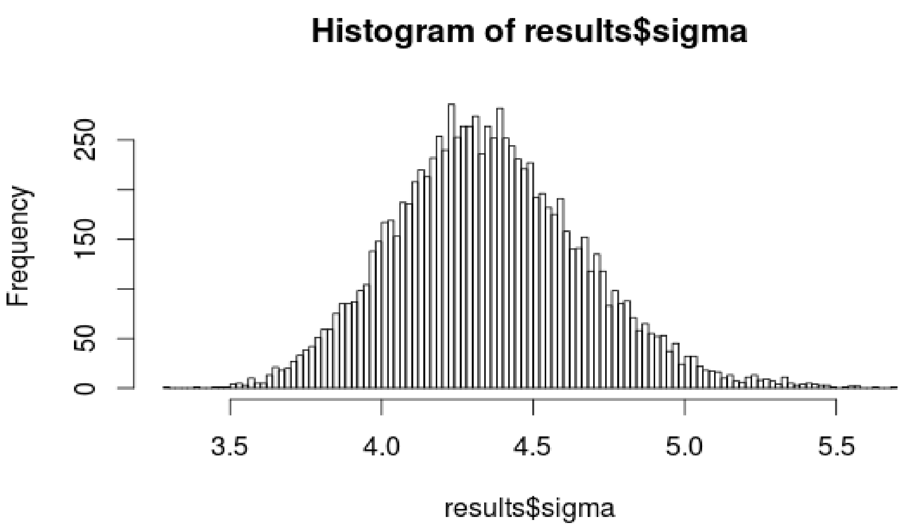
I simulated 100 data points from y = x

and then replaced two data values



Results With Outlier



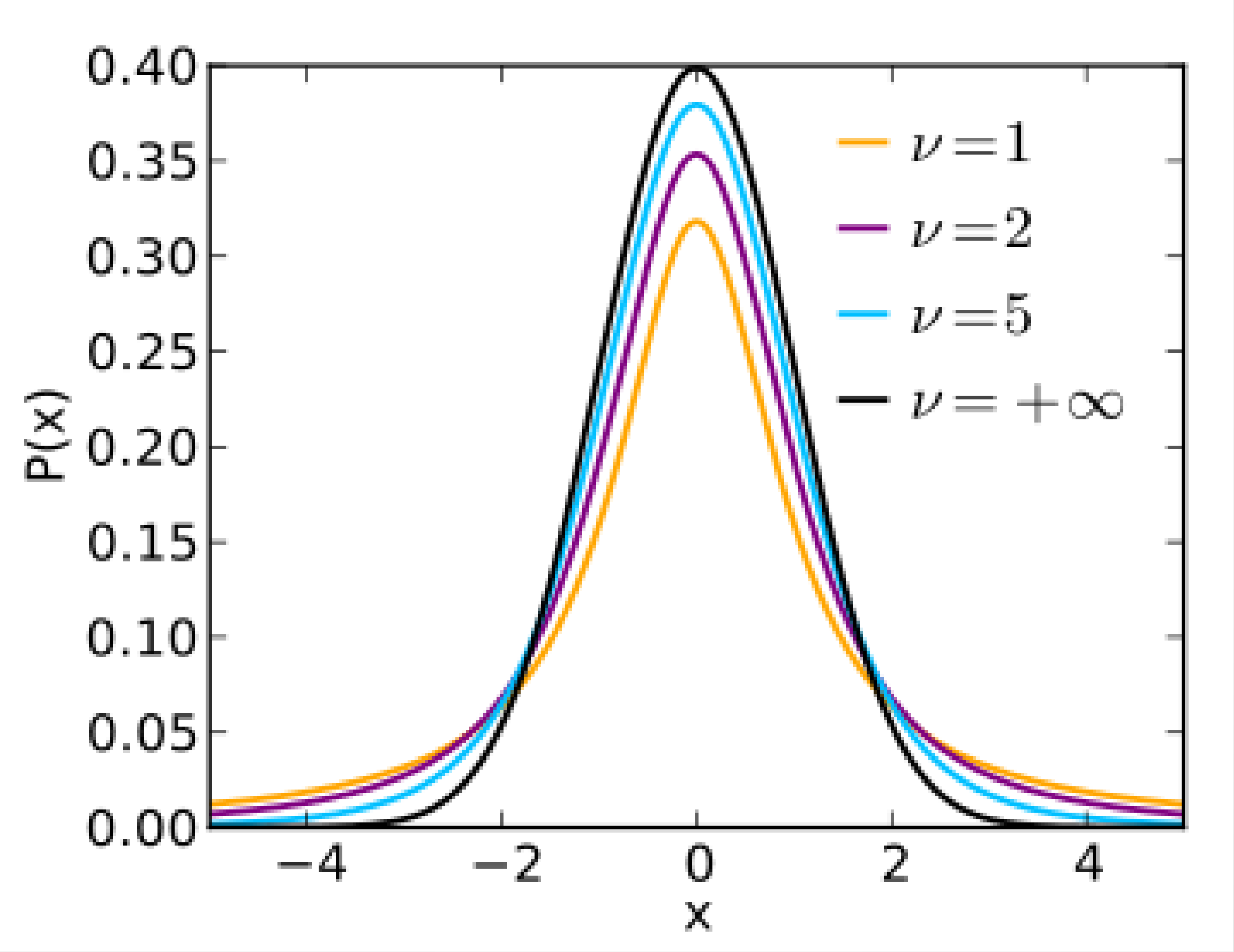


The Outlier Has a Large Effect

- The true known solution (slope = 1, intercept = 0, sigma = 1), isn't within the high probability region of the posterior!

- However, this posterior distribution is correct given the assumptions that went into it. **If the assumptions are really believed, this is the answer, and that's that.**

t-distributions from Wikipedia



t-distributions in jags

```
dt(mu, 1/sigma^2, nu)
```

instead of

```
dnorm(mu, 1/sigma^2)
```

- Note extra parameter (which will need a prior)

t Model Results

- Much better!

- Why? Model "expects" discrepant points, and can explain them by lowering nu, rather than my changing the other parameters.

Wisdom

"Sometimes outliers are bad data, and should be excluded, such as typos. Sometimes they are Wayne Gretzky or Michael Jordan, and should be kept."

- Neil McGuigan, on stackexchange.com

My (less catchy) advice

- “Robustness” is a term in statistics that means parameter estimates aren't largely influenced by outliers
- Whether your method should be robust or not **depends entirely on your prior information**, as expressed in your choice of prior and likelihood.
- One size does not fit all!