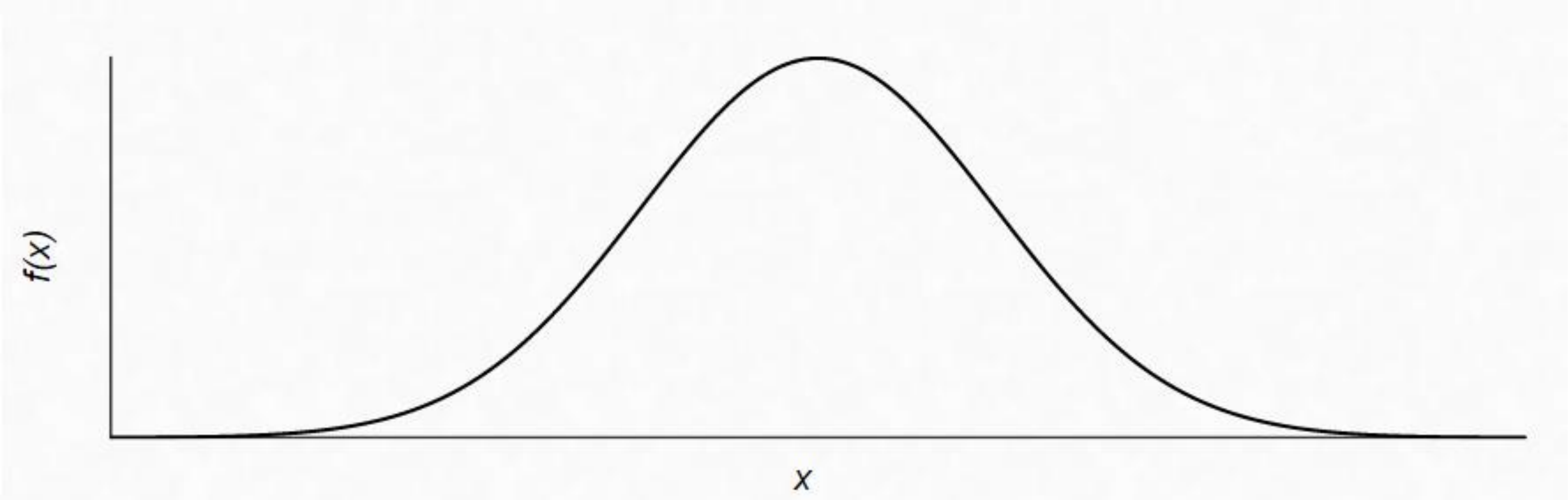
Categorical covariates: ‘random’ and ‘fixed’ effects



Outline

1. **Basic introduction**
   - *Key concepts*
     - *Defining ‘random’ and ‘fixed’ effects*
     - *Normal distributions*
     - *Hierachical models and hyper-parameters (hey-O!)*
     - *Shrinkage*

2. **Case study: adult size variation among populations**

3. **Perils and pitfalls, review take home ideas**

Definitions and distinctions

The terms ‘fixed’ and ‘random’ effect are used fairly widely (also, wildly) across a number of ‘fields’...

Gelman **(2005)** *The Annals of Statistics*

Definitions and distinctions

1. Fixed effects are constant across individuals, random effects vary (Kreft and de Leeuw 1998)

Gelman **(2005)** *The Annals of Statistics*

Definitions and distinctions

1. Fixed effects are constant across individuals, random effects vary (Kreft and de Leeuw 1998)
2. Effects are fixed if they are interesting in themselves and random if there is interest in the underlying population (Searle, Casella, and McCulloch 1992)

Gelman **(2005)** *The Annals of Statistics*

Definitions and distinctions

1. Fixed effects are constant across individuals, random effects vary (Kreft and de Leeuw 1998)
2. Effects are fixed if they are interesting in themselves and random if there is interest in the underlying population (Searle, Casella, and McCulloch 1992)
3. When a sample exhausts the population, use fixed effects. When the sample is small (i.e., negligible), the variable is random (Green and Tukey 1960)

Gelman **(2005)** *The Annals of Statistics*

Definitions and distinctions

1. Fixed effects are constant across individuals, random effects vary (Kreft and de Leeuw 1998)
2. Effects are fixed if they are interesting in themselves and random if there is interest in the underlying population (Searle, Casella, and McCulloch 1992)
3. When a sample exhausts the population, fixed. When the sample is small (i.e., negligible), the variable is random (Green and Tukey 1960)
4. If an effect is assumed to be a realized value of a random variable, it is called a random effect (LaMotte 1983)

Gelman **(2005)** *The Annals of Statistics*

Definitions and distinctions

1. Fixed effects are constant across individuals, random effects vary (Kreft and de Leeuw 1998)
2. Effects are fixed if they are interesting in themselves and random if there is interest in the underlying population (Searle, Casella, and McCulloch 1992)
3. When a sample exhausts the population, fixed. When the sample is small (i.e., negligible), the variable is random (Green and Tukey 1960)
4. If an effect is assumed to be a realized value of a random variable, it is called a random effect (LaMotte 1983)
5. Fixed effects are estimated using least squares, and random effects are estimated with shrinkage (Snijders and Bosker 1999)

Gelman **(2005)** *The Annals of Statistics*

What is a random effect?!

```
lm(y ~ as.factor(x))
```

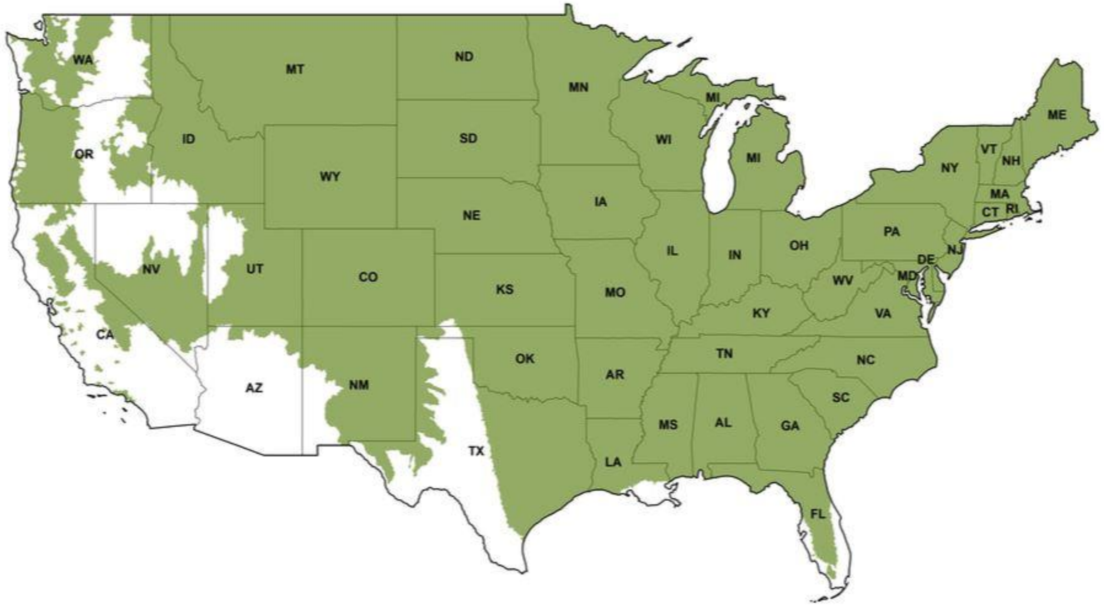
Fixed effect of x

```
lmer(y ~ (1 | as.factor(x)))
```

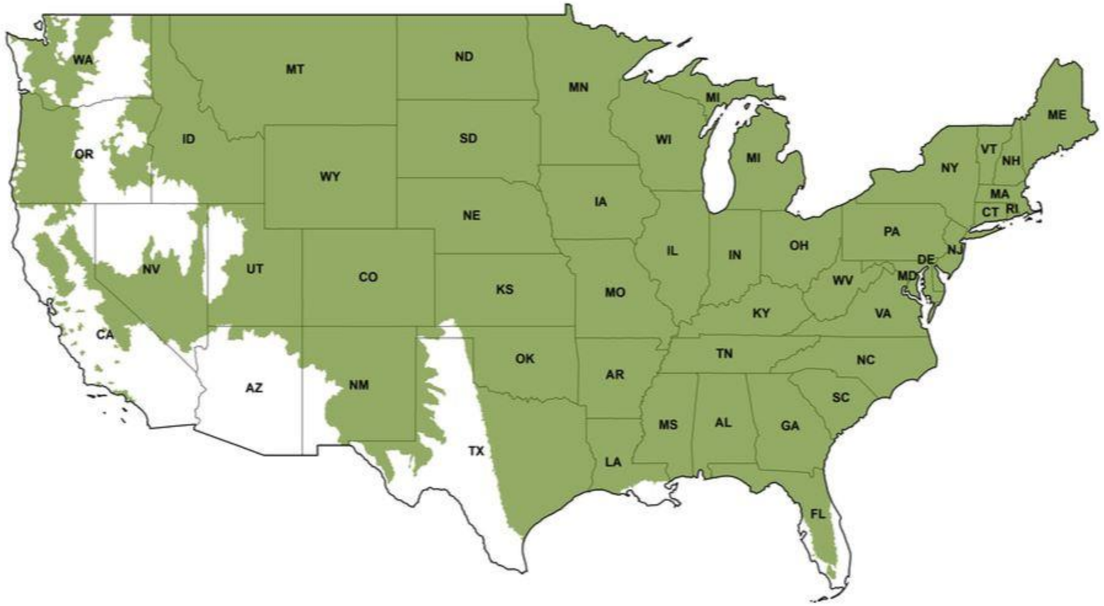
Random effect of x

Using them in a Bayesian context can help clarify what they are!

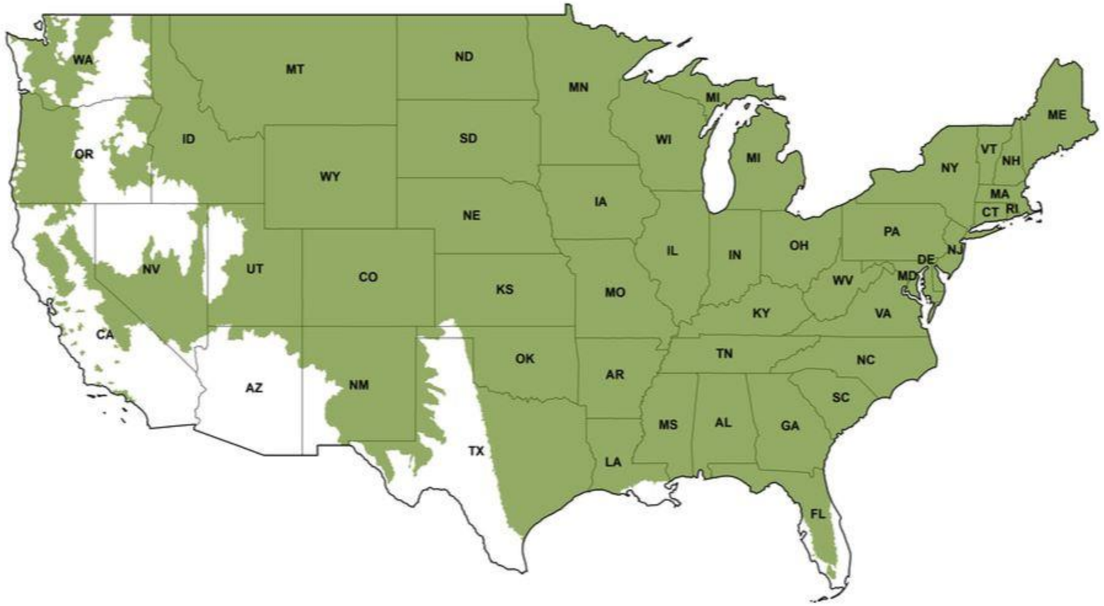
First, let’s describe our data



We'll sample 30 populations of foxes, and weigh a certain number of adult females from each population



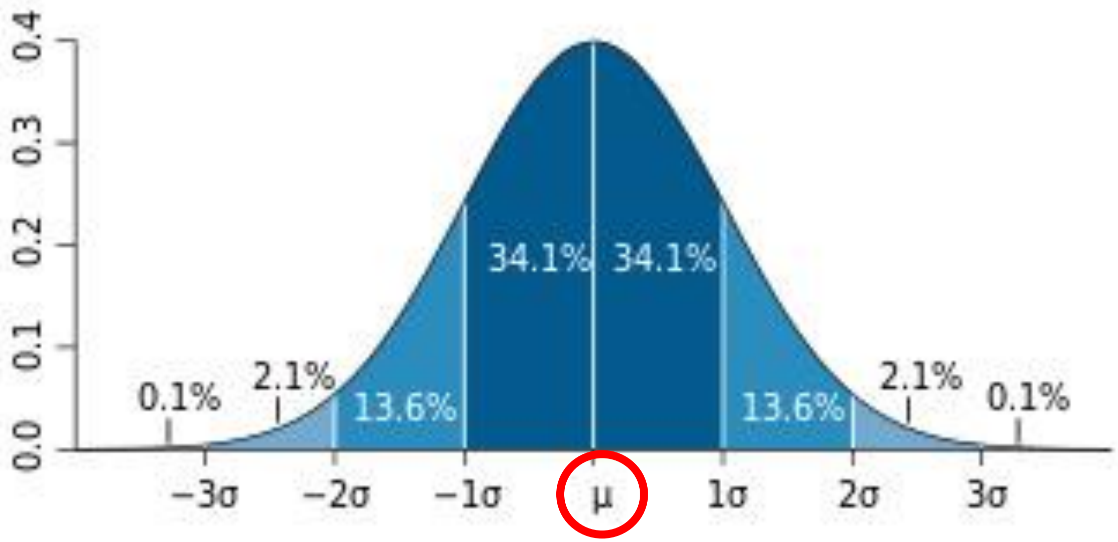
The data (y) will be female mass. We'll have two associated covariates; latitude of the population (continuous, x) and a population identifier (categorical, p).



So what is a random effect?

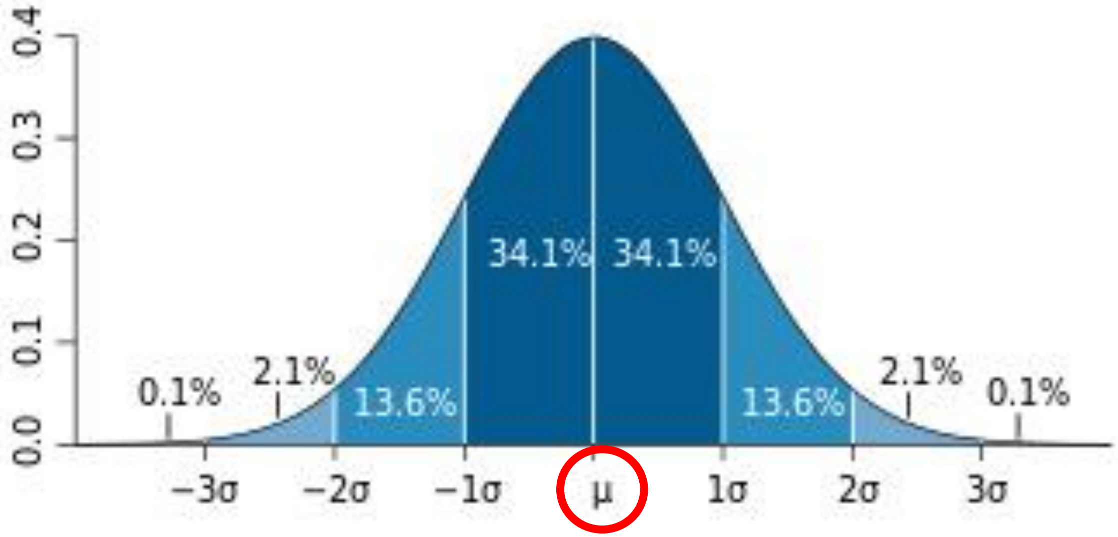
Let's remember a beautiful and simple idea…

$$y_i \sim \text{normal}(\mu, \sigma^2)$$



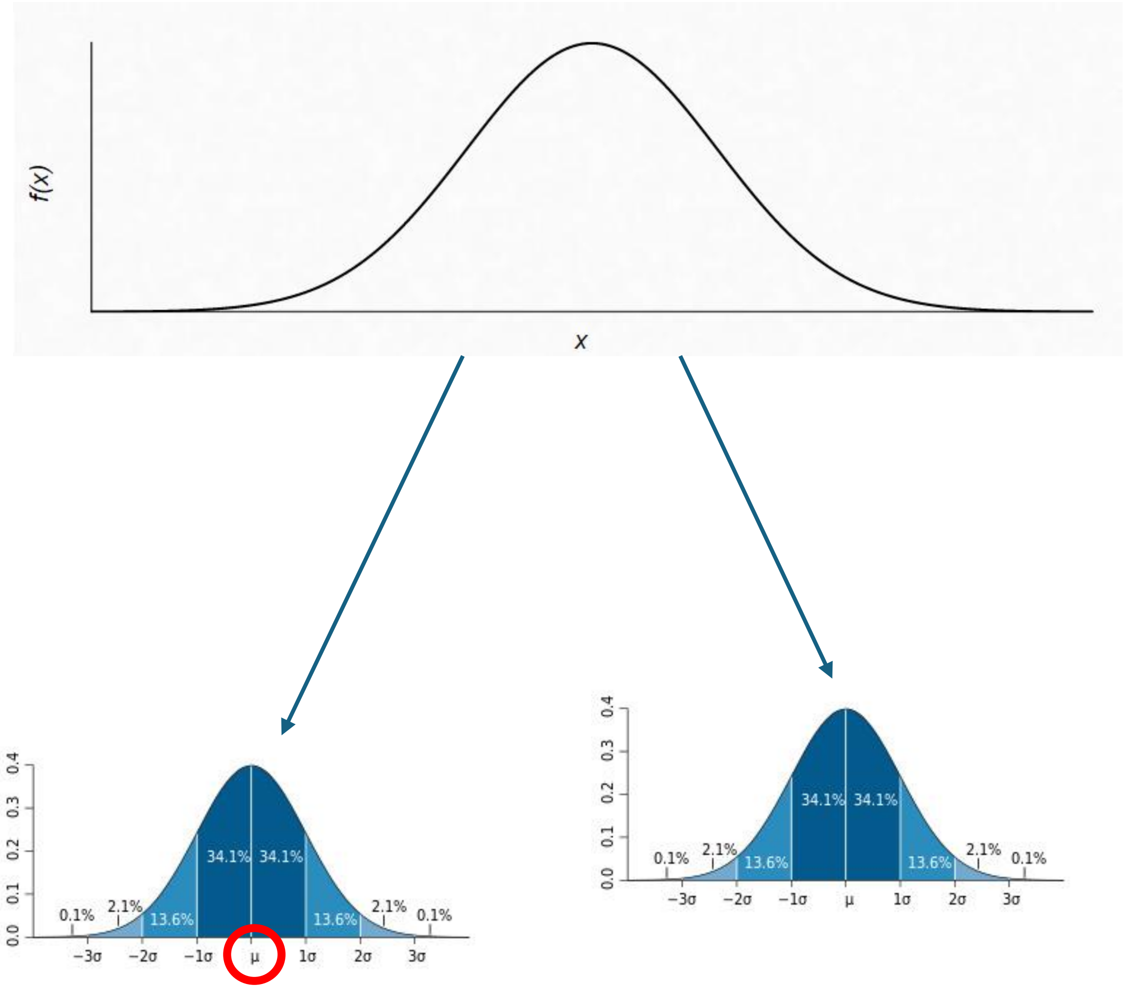
Now let's imagine we have 30 populations, each with a mean…

$$y_i \sim \mathrm{normal}(\mu, \sigma^2)$$



Each of those populations is a sample from the total population!

$$\mu_j \sim \text{normal}(\mu^*, \sigma^2)$$



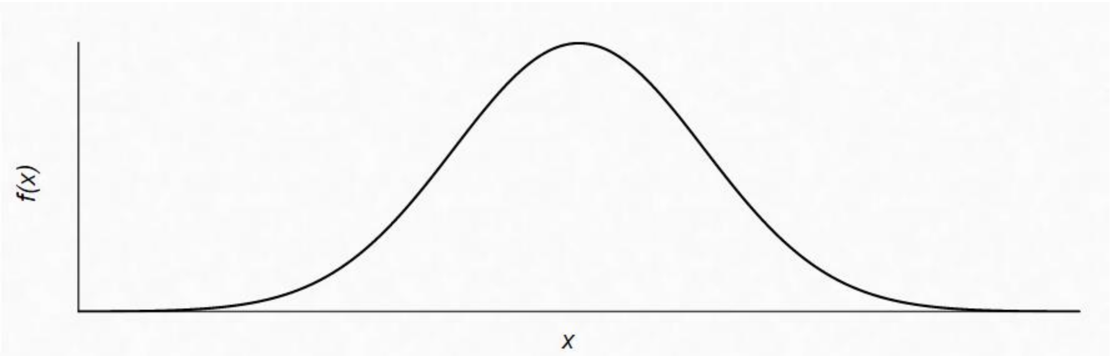
μ_j: each populations mean

μ^*: an average female fox

σ^2: variance among pops

Now let's imagine we have 30 populations, each with a mean…

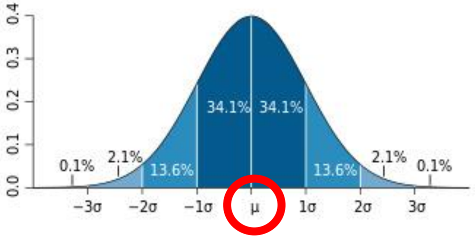
$$\mu_j \sim \text{normal}(\mu^*, \sigma^2)$$

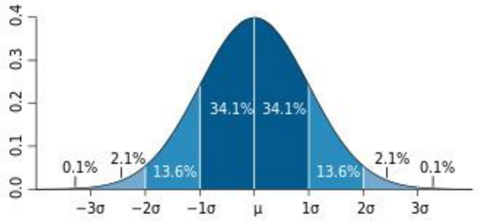


y_i: **the weight of each fox**

ς^2: **variance within pops**

$$y_i \sim \text{normal}(\mu_j, \varsigma^2)$$

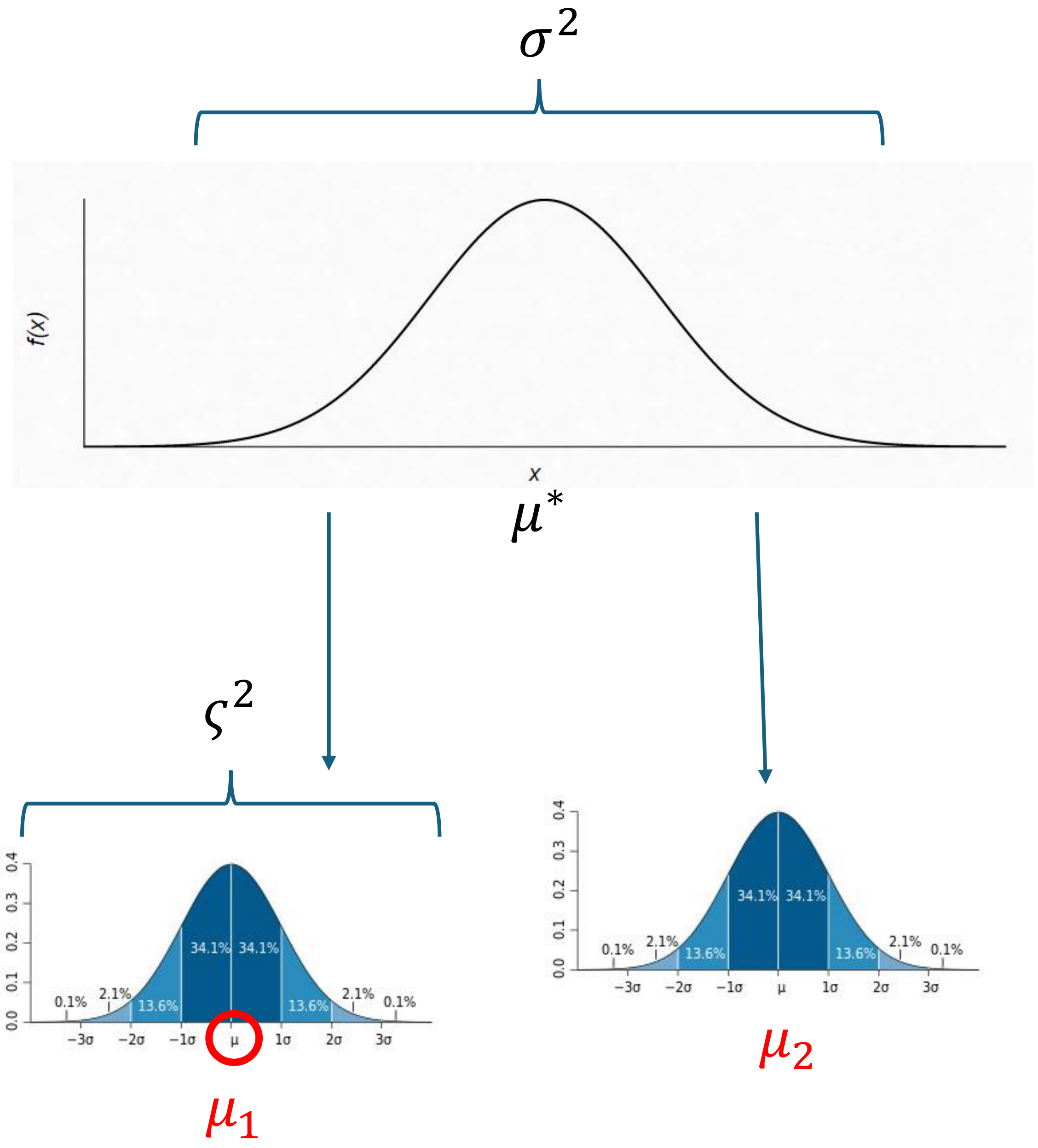




Now let's imagine we have 30 populations, each with a mean…

$$\mu_j \sim \text{normal}(\mu^*, \sigma^2)$$

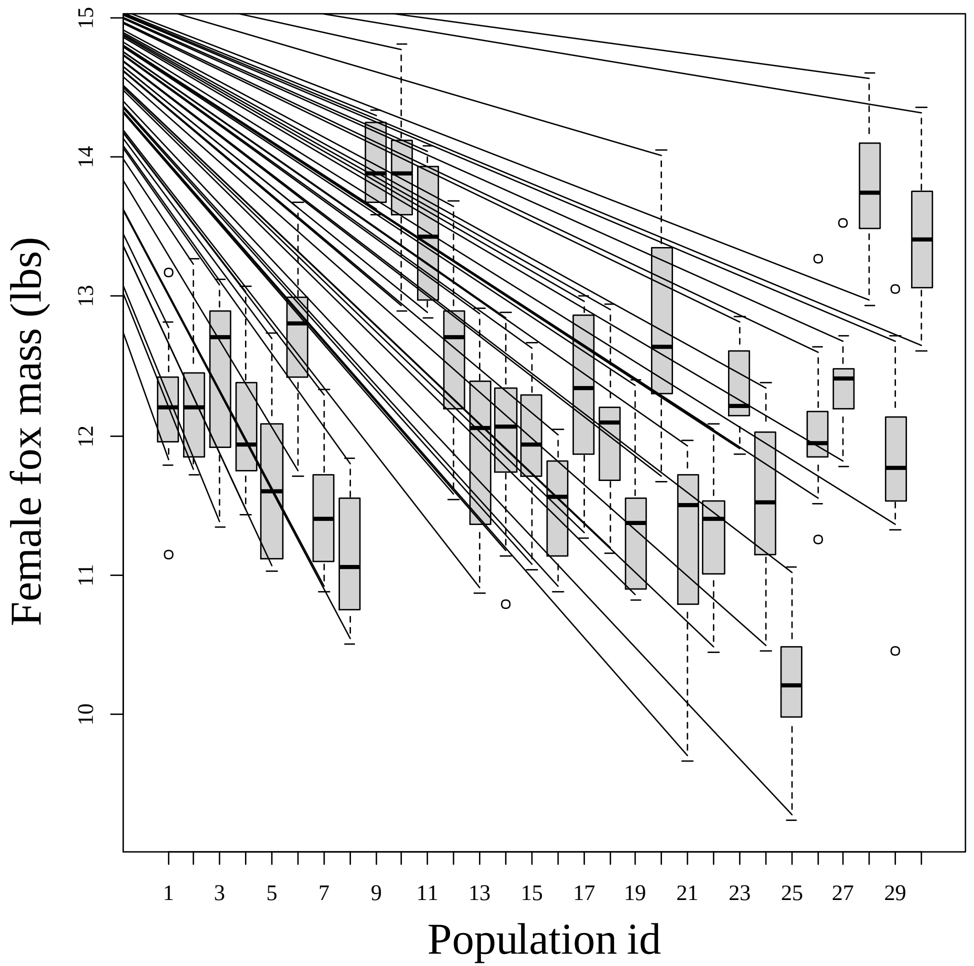
$$y_i \sim \text{normal}(\mu_j, \varsigma^2)$$



The foxes in each population are normally distributed

$$\mu_j \sim \text{normal}(\mu^*, \sigma^2)$$

$$y_i \sim \text{normal}(\mu_j, \varsigma^2)$$

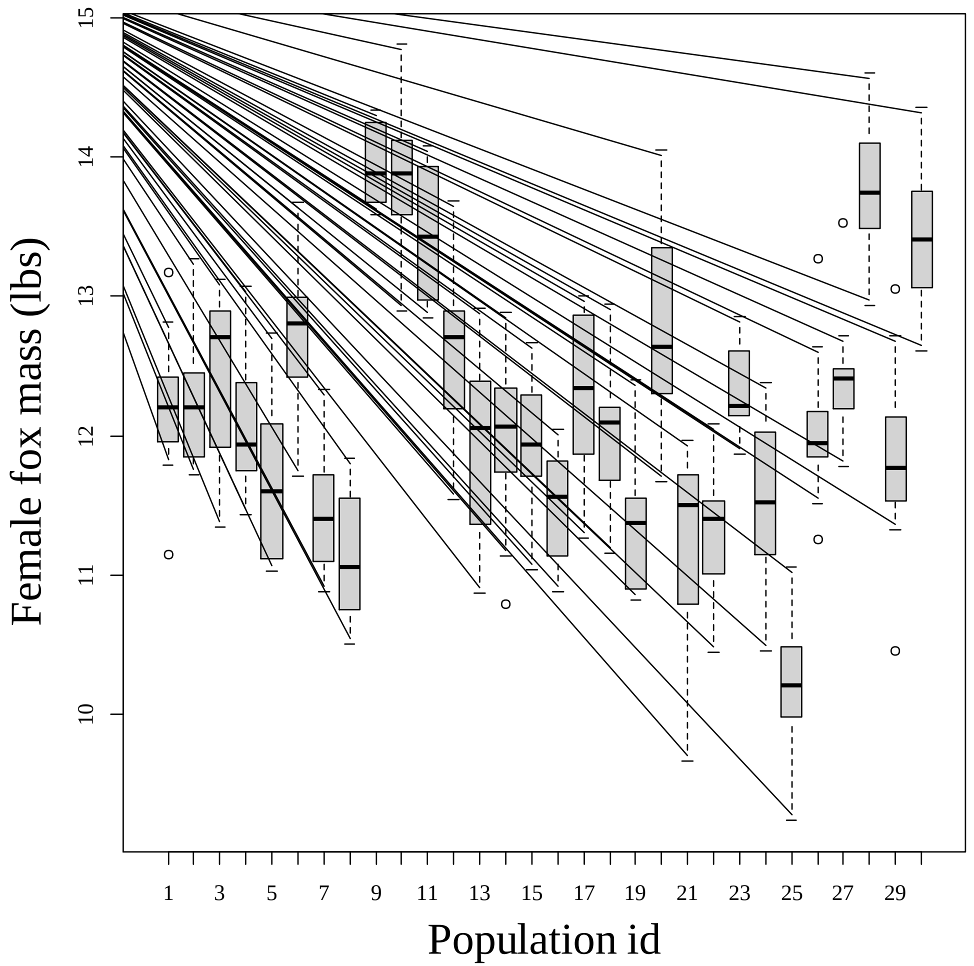


The populations are normally distributed (random effects)

Why do we need a certain number (7?, 10?, more?) of groups…

$$\color{red}{\mu_j} \sim \mathrm{normal}(\mu^*, \sigma^2)$$

$$y_i \sim \mathrm{normal}(\color{red}{\mu_j}, \varsigma^2)$$

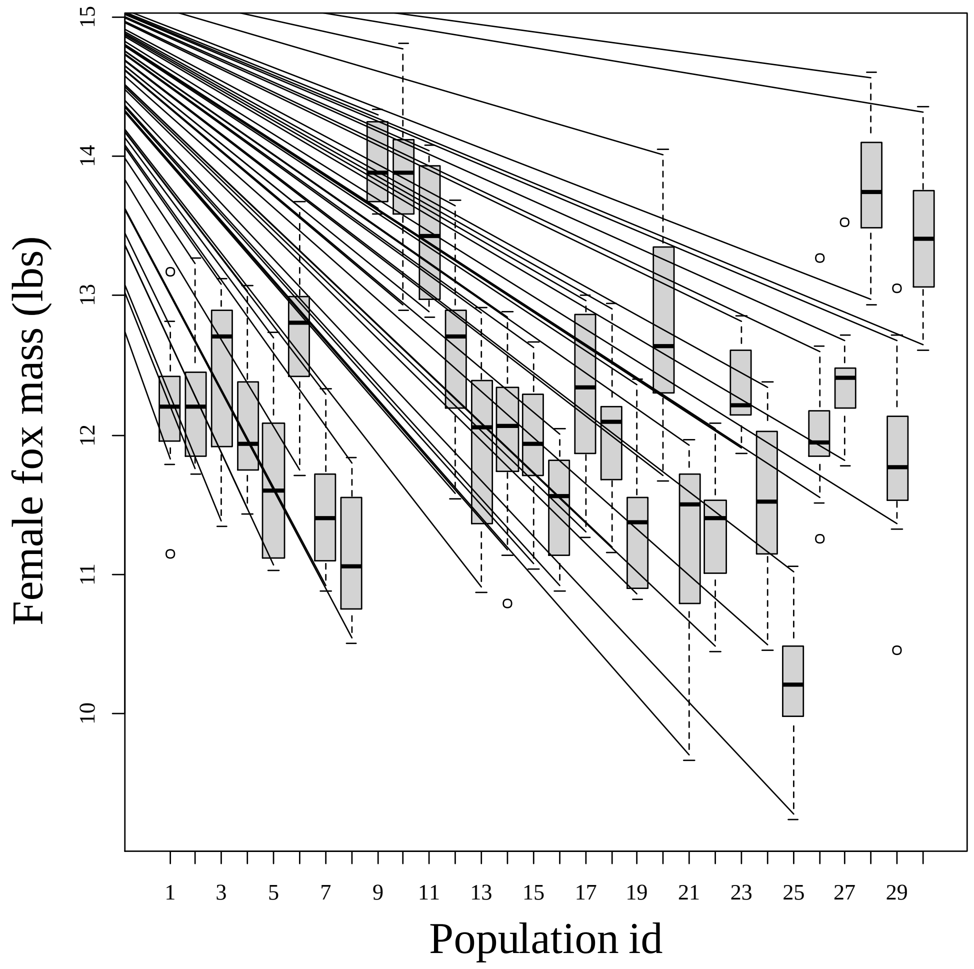


to estimate random effects?

So, what is a fixed effect?!

$$\mu_j \sim \text{normal}(12, 10)$$

$$y_i \sim \text{normal}(\mu_j, \varsigma^2)$$

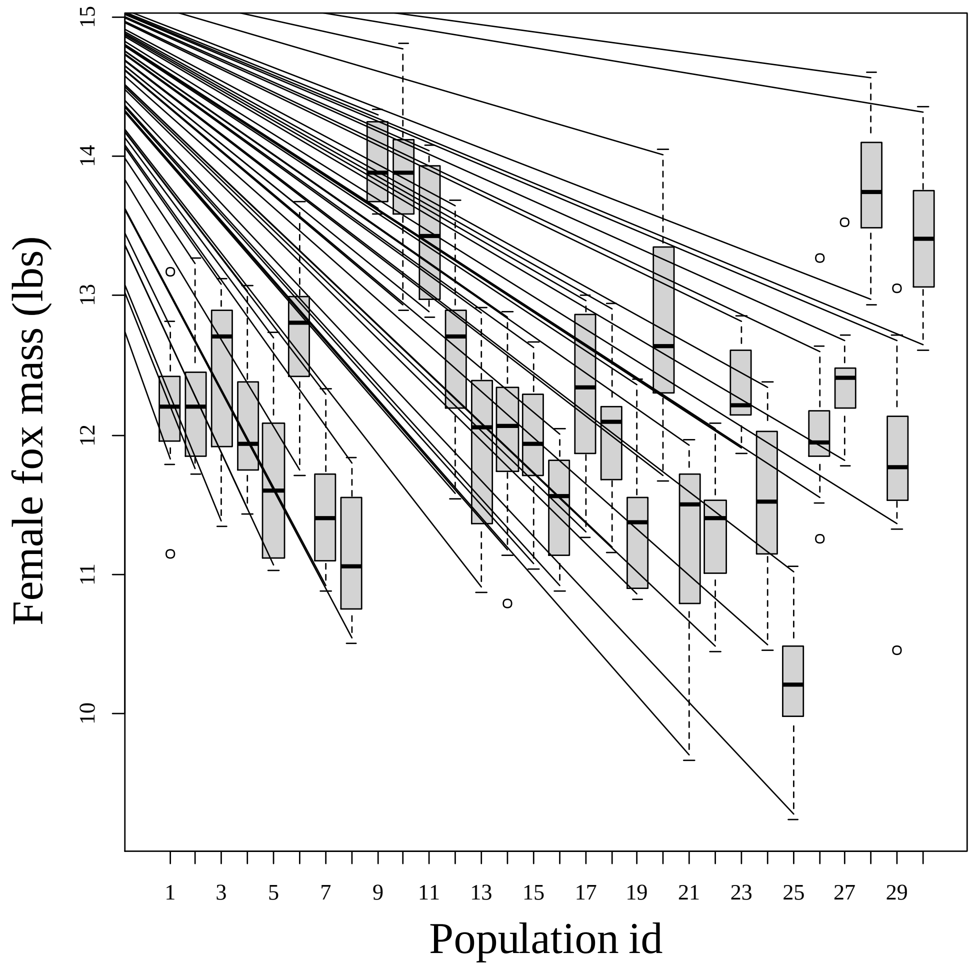


Fixed effects assume that each group is independent.

Random effects are nearly ~equivalent to fixed effects if…

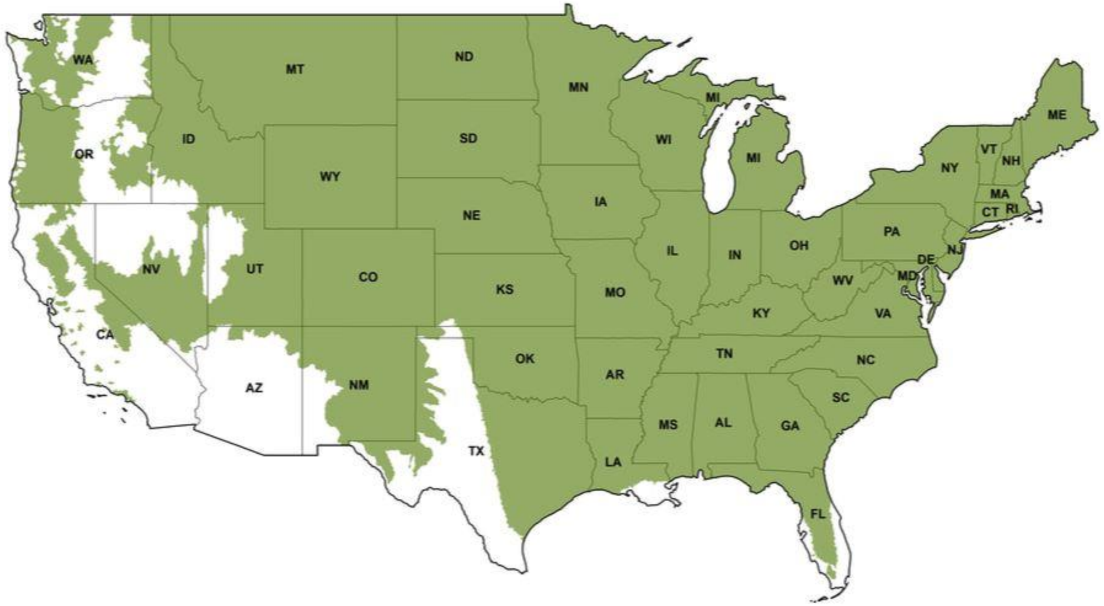
$$\mu_j \sim \text{normal}(\mu^*, \infty)$$

$$y_i \sim \text{normal}(\mu_j, \varsigma^2)$$



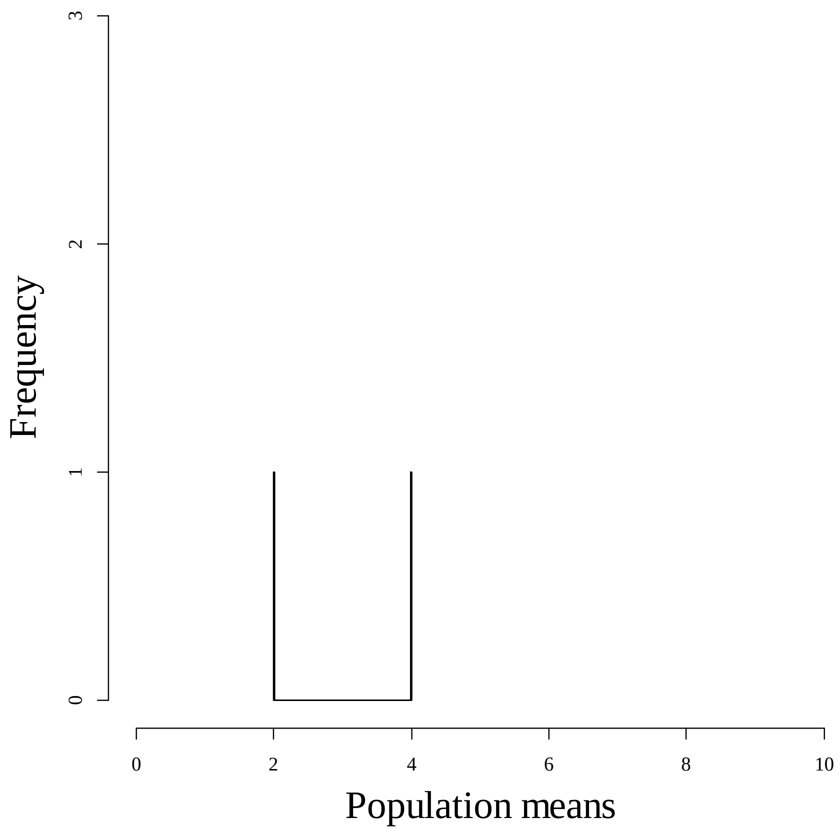
Or as the number of samples increases…

Case study I: adult size variation among populations



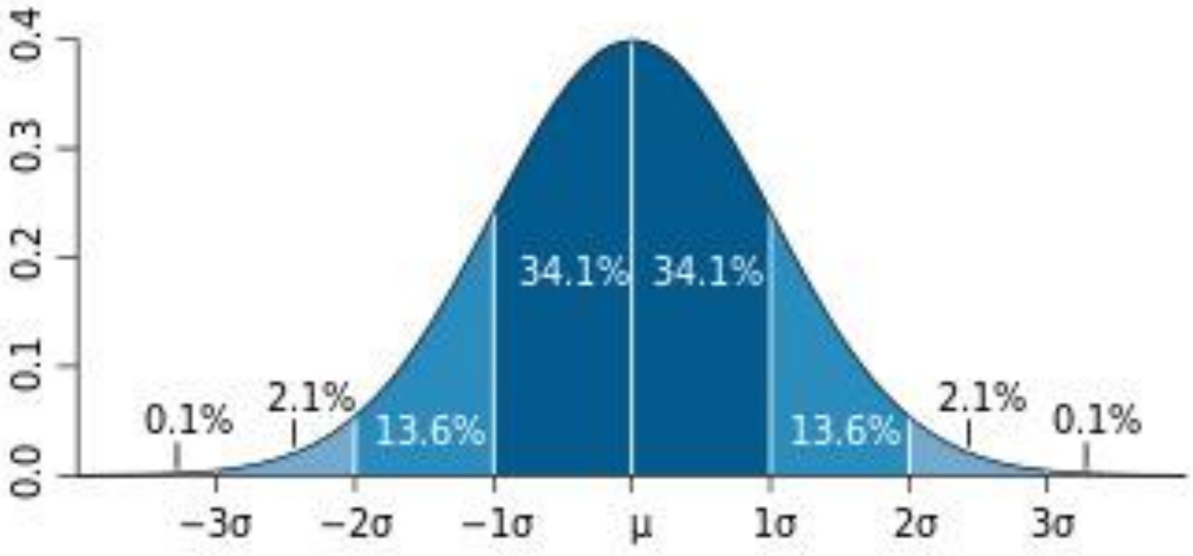
To JAGS!

Take-home idea: <u>there's a reason you need *n* groups!</u>



Imagine trying to estimate the mean (3?) and variance (?) of two groups?

Take-home idea: <u>REs are just another normal distribution ☺!</u>



Take-home idea: <u>estimates 'shrink' more when *n* is small</u>

Take-home idea: as *n* increases, random fx ~= fixed fx

Take-home idea: <u>there are LOTS of bad explanations of random fx</u>

1. Fixed effects are constant across individuals, random effects vary (Kreft and de Leeuw 1998)
2. Effects are fixed if they are interesting in themselves and random if there is interest in the underlying population (Searle, Casella, and McCulloch 1992)
3. When a sample exhausts the population, fixed. When the sample is small (i.e., negligible), the variable is random (Green and Tukey 1960)
4. If an effect is assumed to be a realized value of a random variable, it is called a random effect (LaMotte 1983)
5. Fixed effects are estimated using least squares, and random effects are estimated with shrinkage (Snijders and Bosker 1999)

Seeing the actual model helps us understand what they are.