

# Load your modules in Julia (packages in R)

---

```
• begin
•   using CSV      # module for reading and writing CSV files
•   using Plots    # module for plotting
•   using StatsPlots
•   using DataFrames # ... working with data frames
•   using Random   # ... random numbers
•   using Turing   # Module for Bayesian statistics in Julia
•   using RCall    # Module to run R code in Julia.
• end
```

## 2. Get your data.

---

In this example, we will follow Richard McElreath's example on the partial census of the Dobe are !Kung San. These data was complaed b Nancy Howell in the late 1960s. The !Kung San are the most famus foragin hyman population of the 20<sup>th</sup> century.

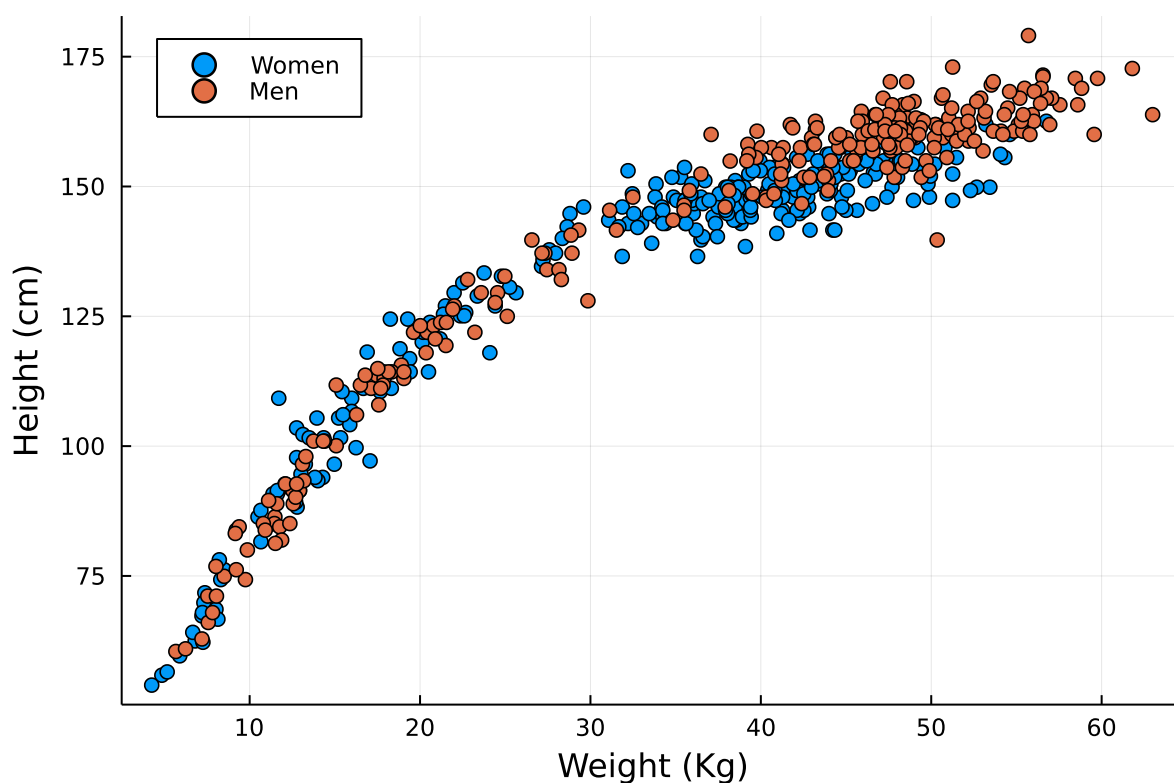


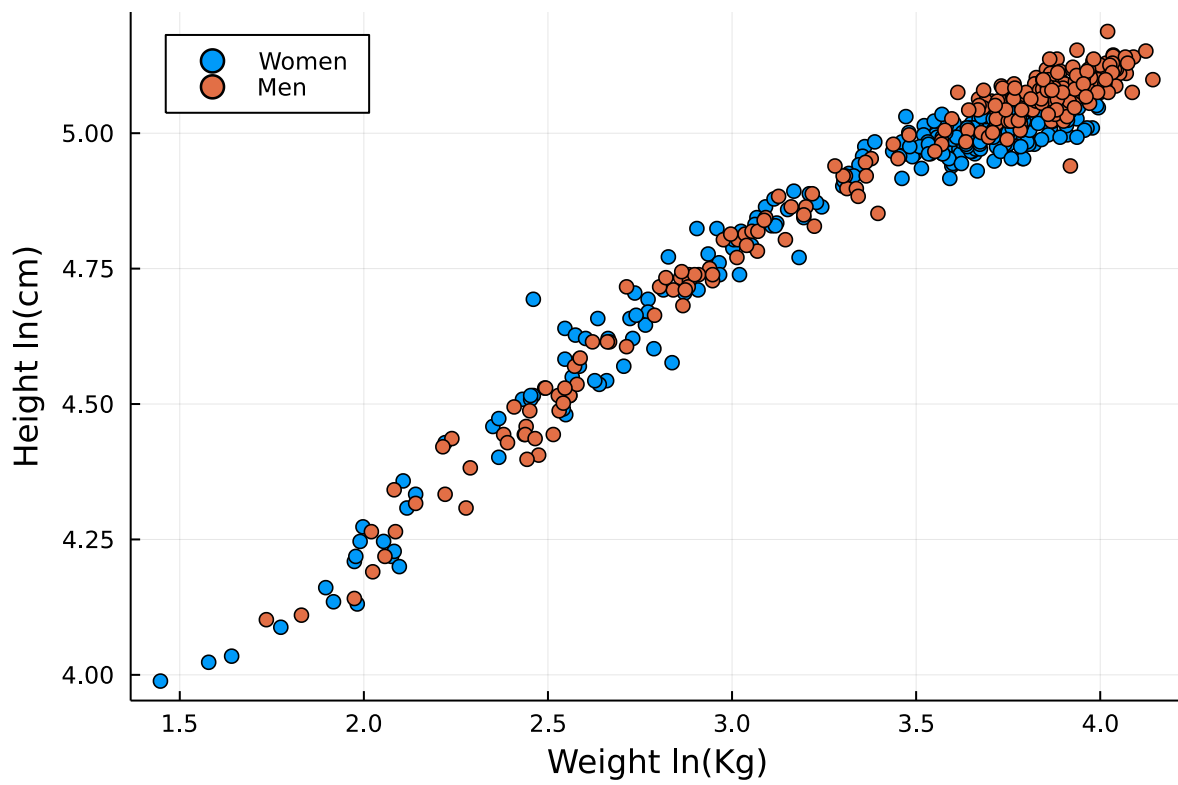
	Column1	height	weight	age	male
1	1	151.765	47.8256	63.0	1
2	2	139.7	36.4858	63.0	0
3	3	136.525	31.8648	65.0	0
4	4	156.845	53.0419	41.0	1
5	5	145.415	41.2769	51.0	0
6	6	163.83	62.9926	35.0	1
7	7	149.225	38.2435	32.0	0
8	8	168.91	55.48	27.0	1
9	9	147.955	34.8699	19.0	0
10	10	165.1	54.4877	54.0	1
more					
544	544	158.75	52.5316	68.0	1

```
• begin
• data = CSV.read("Howell1.csv", DataFrame; delim = ",");
• #data = filter!(x -> x.age >= 18, data);
• end
```

```
• Men = filter(x -> x.male == 1, data);
```

```
• Women = filter(x -> x.male == 0, data);
```





```
• begin
•   scatter(log.(Women.weight), log.(Women.height), lab= "Women", legend = :topleft)
•   scatter!(log.(Men.weight), log.(Men.height), lab= "Men")
•   xlabel!("Weight ln(Kg)")
•   ylabel!("Height ln(cm)")
• end
```

# The model

We want to model the relationship between weight and height and test whether this relationship is different between men and women.

## Our model:

$$x_i = \text{weight}_i - \overline{\text{weight}}$$

,

$$Y \approx \text{Normal}(\mu, \sigma)$$

,

$$\mu_i = \alpha + \beta_w x_i + \beta_m \text{Male}_i + \beta_{mxw} \text{Male}_i x_i$$

$$\alpha \approx \text{Normal}(0, 10)$$

,

$$\beta_w \approx \text{Normal}(0, 10)$$

,

$$\beta_m \approx \text{Normal}(0, 10)$$

,

$$\beta_{mxw} \approx \text{Normal}(0, 10)$$

,

## Define our model

In Julia, we basically make a function that looks almost exactly as in our writing model definition

Kung\_San (generic function with 1 method)

```
• @model function Kung_San(height, weight, male)
•   x = weight .- log(45) # define our continuous predictor
•   σ ~ truncated(Cauchy(0, 10), 0, Inf) # priors for error
•   α ~ Normal(4.93, 10) # priors for intercept
•   β_w ~ Normal(0, 10) # prior for weight effect (slope)
•   β_m ~ Normal(0, 10) # prior for sex effect (slope)
•   β_wm ~ Normal(0, 10) # prior for the interaction effects
•   μ = α .+ β_w .* x .+ β_m .* male .+ β_wm .* (x.*male) # linear model
•   height .~ Normal.(μ, σ) # likelihood
• end
```

## Create a model, with your data.

.....

```
• My_model = Kung_San(log.(data.height), log.(data.weight), data.male);
```

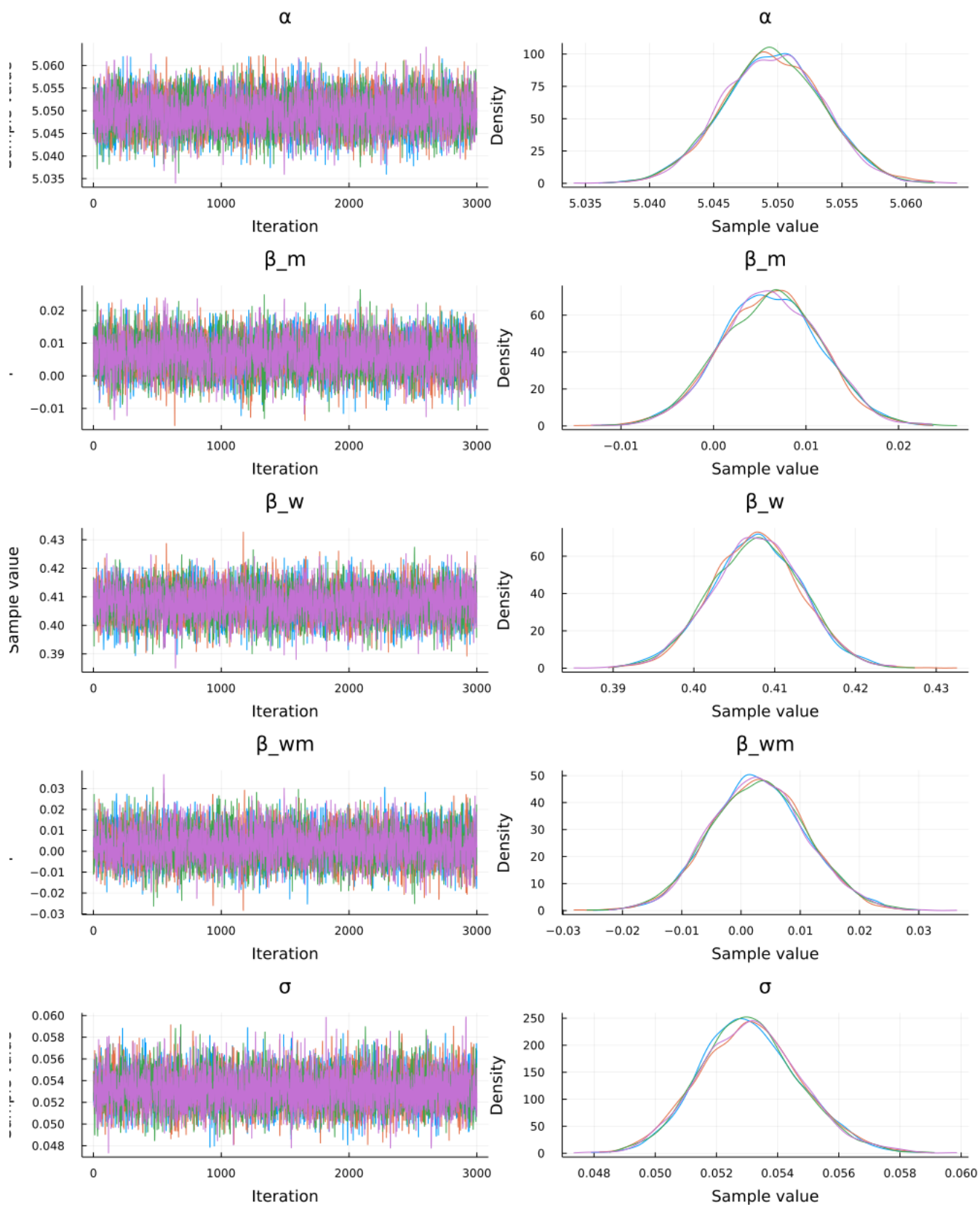
## Now, we run the model, and collect our posterior samples.

.....

```
• # Run the model  
• # Multiple threads 4  
• # 3000 samples  
• chns = sample(My_model, NUTS(), MCMCThreads(), 3000, 4);
```

## Now, we check if the MCC chains converged

.....



• `plot(chns)`

# Check your summary statistics

	parameters	mean	std	naive_se	mcse	ess	rhat
1	$:\alpha$	5.04962	0.00387235	3.53495e-5	5.01242e-5	7361.59	0.999982
2	$:\beta_m$	0.00618889	0.0053693	4.90148e-5	7.28096e-5	7091.96	1.00007
3	$:\beta_w$	0.407733	0.00559922	5.11137e-5	7.0735e-5	6773.15	0.99996
4	$:\beta_{wm}$	0.00294581	0.00802802	7.32855e-5	0.000106298	6862.81	1.00002
5	$:\sigma$	0.0531063	0.00164275	1.49962e-5	1.68475e-5	9234.18	0.999735

```
• describe(chns)
```

## Let's test our hypothesis

Research hypothesis: the relationship between height and weight is stronger in males.

```
• poste = DataFrame(chns);
```

p\_link (generic function with 1 method)

```
• ## model from the posteriors
• function p_link(weight, male,  $\alpha$ ,  $\beta_w$ ,  $\beta_m$ ,  $\beta_{wm}$ ) # weight in Kg, Male==1 female==0,
  ...
•   x = log.(weight) .- log.(45)
•   p =  $\alpha$  .+  $\beta_w$  .* x .+  $\beta_m$  .* male .+  $\beta_{wm}$  .* (x.*male)
•   return exp.(p)
• end
•
```

Get summary statistics such as complementary intervals (CI)

HDI (generic function with 1 method)

```
• function HDI(samples; credible_mass=0.95)
•   # Computes highest density interval from a sample of representative values,
•   # estimated as the shortest credible interval
•   # Takes Arguments posterior_samples (samples from posterior) and credible mass
•   (normally .95)
•   # Originally from https://stackoverflow.com/questions/22284502/highest-
•   posterior-density-region-and-central-credible-region
•   # Adapted to JuliaLang
•   sorted_points = sort(samples)
•   ciIdxInc = Int(ceil(credible_mass * length(sorted_points)))
•   nCIs = length(sorted_points) - ciIdxInc
•   ciWidth = repeat([0.0],nCIs)
•   for i in range(1, stop=nCIs)
•       ciWidth[i] = sorted_points[i + ciIdxInc] - sorted_points[i]
•   end
•   HDImin = sorted_points[findfirst(isequal(minimum(ciWidth)),ciWidth)]
•   HDImax = sorted_points[findfirst(isequal(minimum(ciWidth)),ciWidth)+ciIdxInc]
•   return([HDImin, HDImax])
• end
•
```

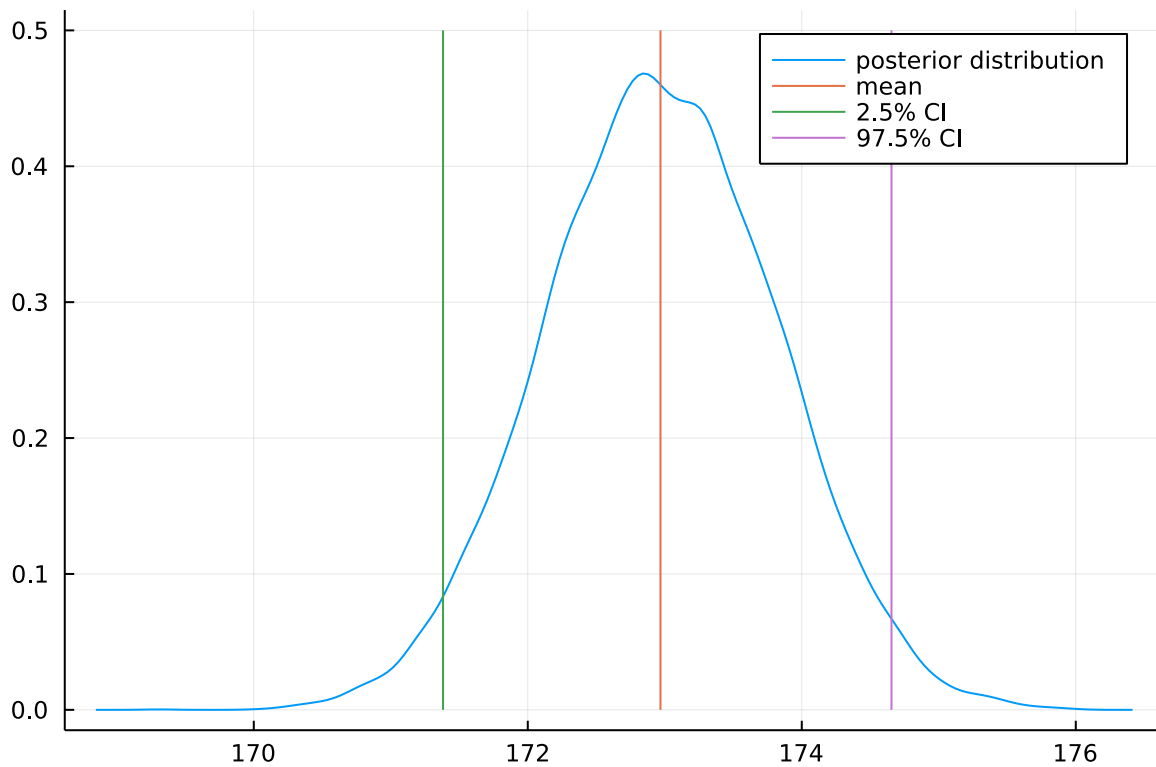


# What is the predicted height of a 58.0 Kg woman of the !Kung San population



We use our new predicted function to estimate the most likely size range of a 58.0 Kg women.

```
prediction_W_58kg =  
[172.067, 172.103, 174.075, 173.892, 173.696, 174.298, 172.823, 173.291, 173.391, 172.44:  
• prediction_W_58kg = p_link(58.0, 0, poste.α, poste.β_w, poste.β_m, poste.β_wm)
```



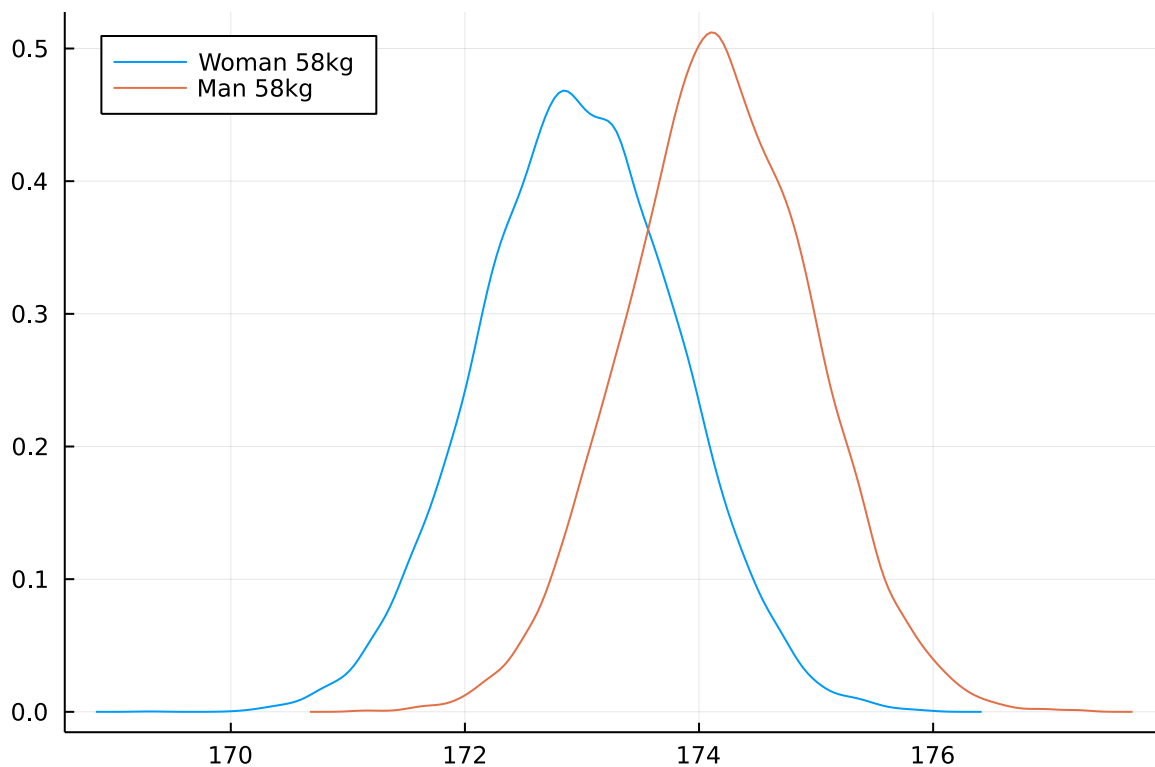
## Time to test a research hypothesis (not a statistical hypothesis):

A man that weights 58 Kg is smaller than a woman of the same size

`prediction_M_58kg =`

```
[173.651, 173.66, 174.197, 174.043, 176.632, 173.758, 174.808, 173.213, 174.092, 172.988
```

```
• prediction_M_58kg = p_link(58.0, 1, poste.α, poste.β_w, poste.β_m, poste.β_wm)
```



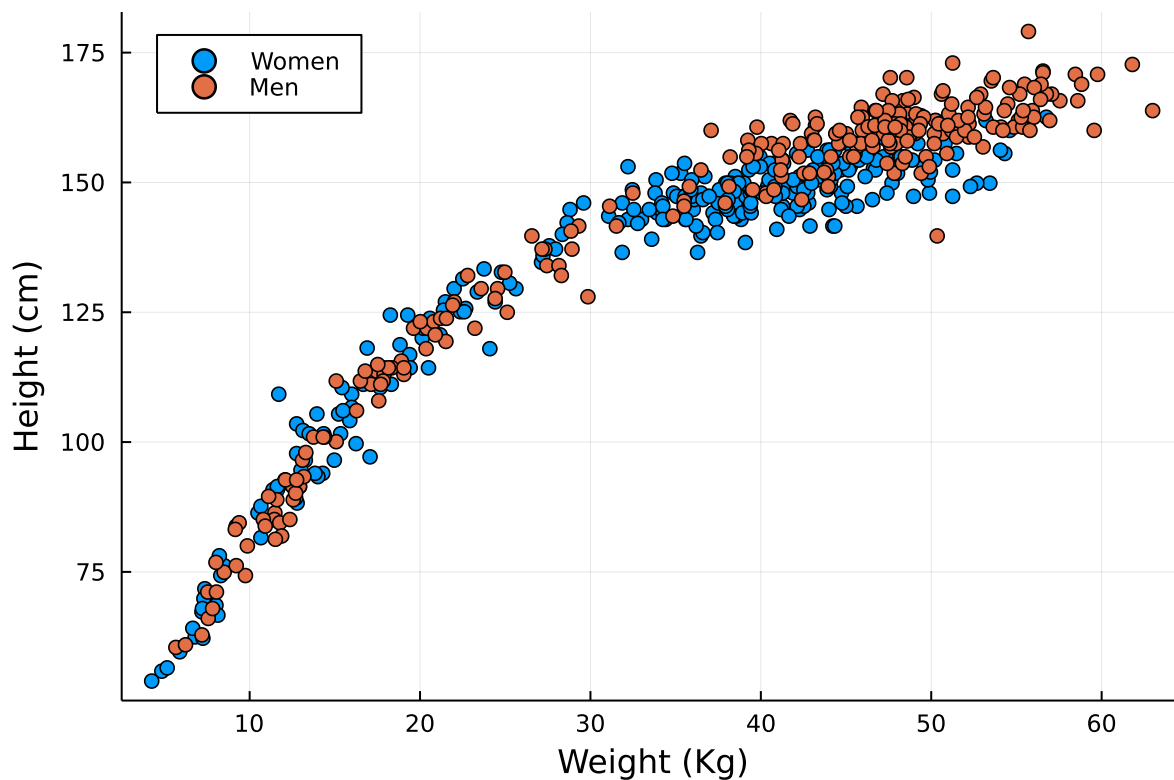
85.175

```
• # formal test  
• 100 * sum(prediction_M_58kg .- prediction_W_58kg .> 0) / length(prediction_M_58kg)  
•
```

**Our model predictions show that a 58 Kg man is much taller than a 58 Kg woman with a 100%**

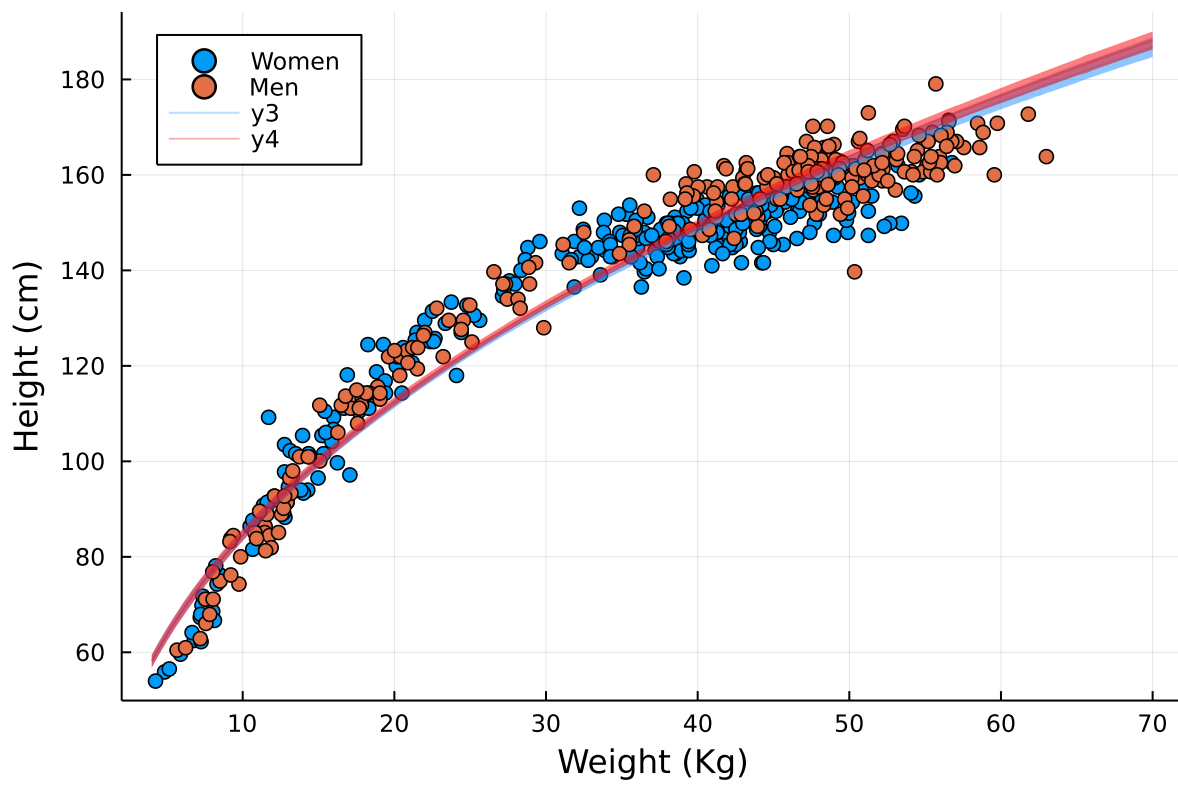
**Are men always taller than women?**

.....



```
• begin
•   scatter(Women.weight, Women.height, lab= "Women", legend = :topleft)
•   scatter!(Men.weight, Men.height, lab= "Men")
•   xlabel!("Weight (Kg)")
•   ylabel!("Height (cm)")
• end
```

This plot shows that men are generally taller, but some women are taller than some men. **So, what is the probability that a woman is taller than a man when we control for their weight?**



79.2114465408805

```
• 100*sum(Matrix(raw_m .- raw_w)[:] .> 0) / length(Matrix(raw_m .- raw_w)[:])
```