**Bài toán 2 (Bắt buộc). Viết chương trình sử dụng thuật toán Euler tìm nghiệm của hệ SIR hoặc mở rộng của nó với tham số đầu vào gồm biến thời gian t, các hệ số tiếp xúc β, hệ số phục hồi γ và điều kiện đầu là số ca mắc bệnh I(t0) và số ca phục hồi R(t0) của mô hình tính tại thời điểm đầu tiên ghi nhận được các ca nhiễm bệnh. Giá trị trả về là mảng chứa số người nhiễm bệnh I(t) và số người đã hồi phục R(t) tính tại thời điểm t ≥ t0. Cho một số ví dụ về điều kiện đầu và các hệ số trong mô hình và dùng chương trình đã viết để tìm nghiệm xấp xỉ. Biểu diễn nghiệm xấp xỉ bằng cách vẽ đồ thị. Trường hợp là hệ SIR mở rộng như SIRD thì cần trả về I(t), R(t) và D(t) là số ca tử vong tại thời điểm t ≥ t0.Trình bày chi tiết kết quả trong báo cáo.**

*Input:*

*+ Thời gian ban đầu: t0*

*+ Thời gian: t*

*+ Hệ số tiếp xúc: β*

*+ Hệ số hồi phục: γ*

*+ Số người có nguy cơ bị lây nhiễm tại thời điểm t0: S0*

*+ Số ca mắc bệnh tại thời điểm t0: I0*

*+ Số ca hồi phục tại thời điểm t0: R0*

*Output:*

*+ Số ca mắc bệnh tại thời điểm t: I*

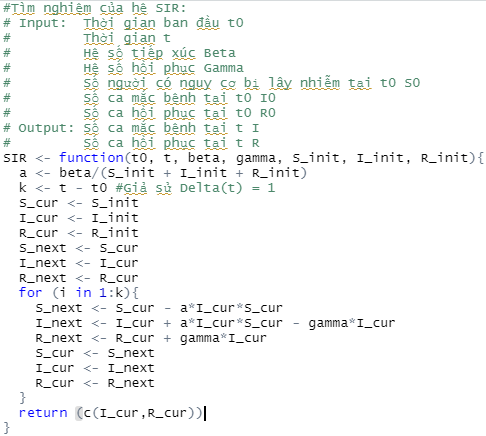
*+ Số ca hồi phục tại thời điểm t: R*

**\*Hiện thực hàm:**

Ta sẽ sử dụng công thức Euler để tính toán như sau:

Giả sử rằng Δt = 1 ta có:

, với

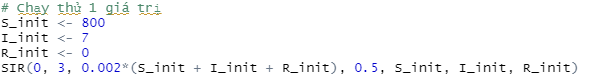


**\*Chạy thử chương trình:**

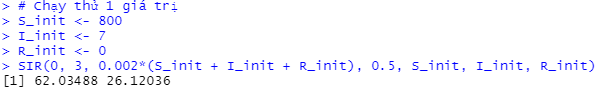
- Đầu tiên ta sẽ chạy thử chương trình với các tham số S0, I0, R0 trong đề bài để kiểm tra xem chương trình chạy đúng hay chưa.

Ta thiết lập như sau: S0 = 800, I0 = 7, R0 = 0, chọn thời điểm ban đầu là t0 = 0, a = 0.002 và γ = 0.5.

Ta cần tìm I(3) và R(3)

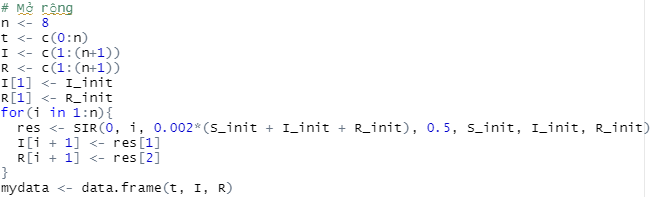


Kết quả thu được:

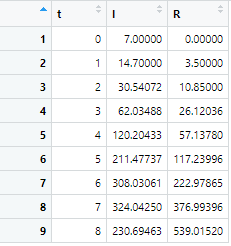


Kết quả này đúng với kết quả thu được trong đề bài.

- Ta có thể mở rộng hơn bằng việc lập bảng các I, R tại nhiều thời điểm t. Với các đầu vào tương tự như trên, ta lập được bảng sau:

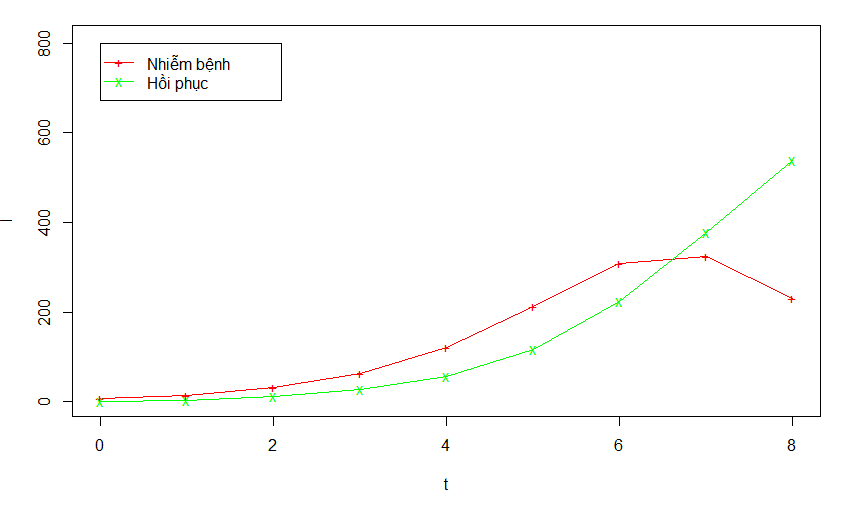


Bảng các I, R tại các thời điểm t:

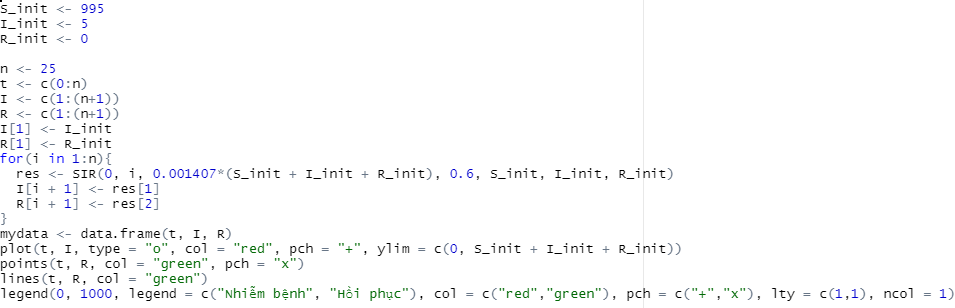


- Ta cũng có thể vẽ đồ thị cho bảng trên như sau:

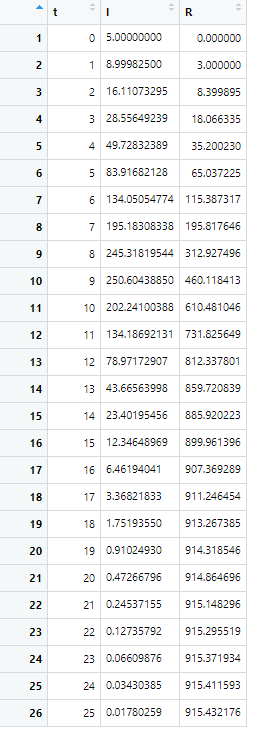


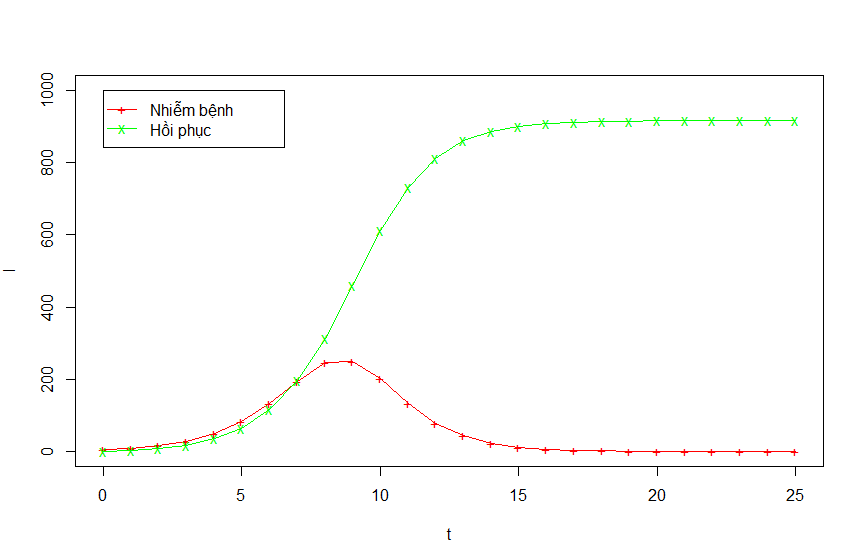


- Ta sẽ lấy thêm một ví dụ khác. Với các thông tin đầu vào: S0 = 995, I0 = 5, R0 = 0, chọn thời điểm ban đầu là t0 = 0, a = 0.001407 và γ = 0.6, ta cần tìm I(t), R(t) với t0 < t ≤ 25.



Kết quả thu được:





**Bài toán 3 (Bắt buộc). Viết chương trình theo ngôn ngữ tự chọn để lấy mẫu sử dụng thuật toán Metropolis–Hastings với tham số đầu vào là phân bố xác suất tiên nghiệm π(β, γ) cho trước. Giá trị trả về là một mẫu gồm các cặp β và γ có phân bố xác suất π(β, γ). Vẽ biểu đồ thể hiện quá trình chọn mẫu. Trình bày chi tiết kết quả trong báo cáo.**

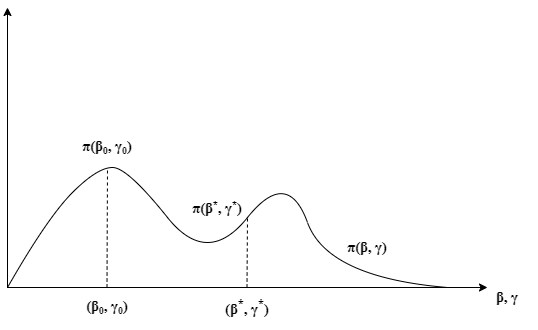
*Input: hàm phân bố xác suất tiên nghiệm π(β, γ).*

*Output: mẫu bao gồm các cặp β và γ.*

**\*Cơ sở lý thuyết:**

**\*Ý tưởng:**

Giả sử ta xác định được β và γ phân bố xác suất theo 1 hàm π(β, γ). Ta có đồ thị của hàm π(β, γ) như sau:



Ta chọn ngẫu nhiên 1 điểm (β, γ) trên trục hoành của đồ thị làm điểm ban đầu (β0, γ0), ta có β = β0 và γ = γ0, từ đó xác định π(β0, γ0) trên đồ thị.

Ta lấy phân phối xác suất bất kỳ p(β, γ), ta chọn xác suất bất kỳ p(β, γ) có dạng phân phối Chuẩn để việc tính hệ số r dễ dàng hơn. Lấy ngẫu nhiên 1 điểm (β\*, γ\*) mang xác suất bất kỳ p(β, γ), từ đó ta xác định π(β\*, γ\*).

Vì mang phân phối chuẩn nên p(β\*, γ\*| β, γ) = p(β, γ | β\*, γ\*).

→ Xác suất chấp nhận (β\*, γ\*) r được tính theo công thức:

**r = min{1, } = min{1, }**

+ Nếu π(β\*, γ\*) ≥ π(β, γ) thì r = 1, ta luôn chấp nhận di chuyển đến (β\*, γ\*)

→ βi+1 = β\* và γi+1 = γ\*

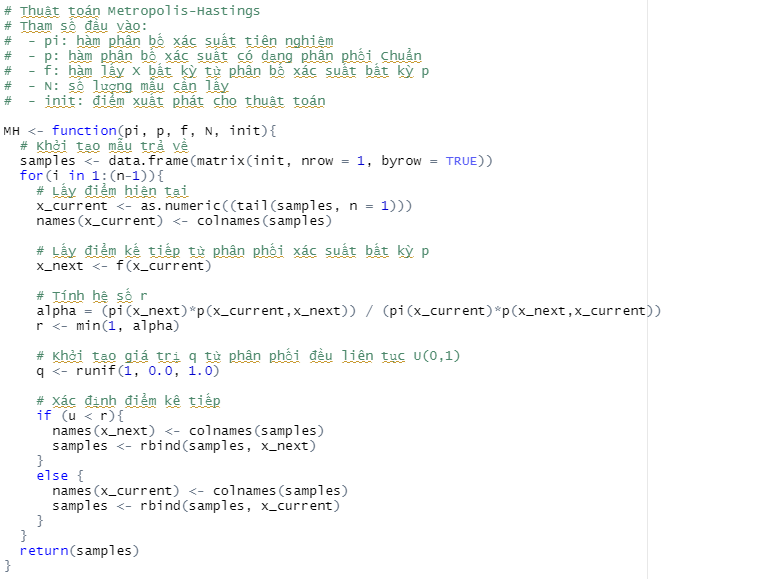
+ Nếu π(β\*, γ\*) < π(β, γ) thì r = < 1, ta sẽ di chuyển với mức lưỡng lự có xác suất r hoặc giữ nguyên vị trí cũ với xác suất (1-r). Để quyết định di chuyển hay đứng yên, ta khởi tạo một giá trị q ngẫu nhiên từ phân phối đều liên tục U(0,1), hay nói đơn giản là lấy một giá trị q ngẫu nhiên nằm trong khoảng [0,1].

* Nếu q < r, ta chấp nhận di chuyển đến (β\*, γ\*) → βi+1 = β\* và γi+1 = γ\*.
* Ngược lại, ta giữ nguyên vị trí hiện tại (βi, γi) → βi+1 = βi và γi+1 = γi.

Ta sẽ lặp lại quá trình trên cho đến khi i = n (n là kích thước mẫu ta muốn khảo sát), ta sẽ thu được một mẫu kết quả bao gồm các cặp β, γ.

Lưu ý mỗi lần di chuyển ta thêm giá trị β\*, γ\* vào mẫu kết quả.

**\*Hiện thực hàm:**



**\*Kiểm thử:**

Ta sẽ chạy thử đoạn Code trên với các bước như sau:

*1. Thiết lập hàm phân bố xác suất tiên nghiệm:*

- Giả sử ta có xác suất tiên nghiệm π(β, γ) có dạng phân phối Chuẩn N(a,).

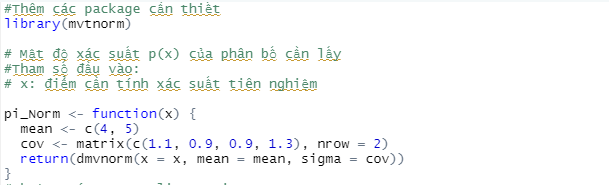
- Vì hàm xác suất tiên nghiệm phụ thuộc vào 2 tham số β, γ, để biểu diễn hàm phân phối dạng Chuẩn N(a,), ta sử dụng mà trận tương quan tuyến tính (**Square variance-covariance matrix**) có kích thước 2x2 để biểu diễn phương sai và sử dụng mà trận có kích thước 2x1 bao gồm 2 giá trị trung bình của β, γ để đưa vào giá trị trung bình a. Để tính phân bố xác suất tiên nghiệm tại 1 điểm βi, γi, ta sử dụng hàm **dmvnorm** trong thư viện **mvtnorm** (thư viện nhằm hỗ trợ xử lý các phân phối Chuẩn). Hàm **dmvnorm** bao gồm các tham số:

* x: điểm cần tính phân bố xác suất
* mean: giá trị trung bình của x
* sigma: phương sai của x

- Ta sẽ chạy thử chương trình với ví dụ cụ thể như sau:

+ Giá trị trung bình của β, γ lần lượt là 4.0 và 5.0

+ Ma trận tương quan tuyến tính: (1.1 0.9; 0.9 1.3)



→ Hàm **pi\_Norm** sẽ tính xác suất tiên nghiệm π(βi, γi) tại điểm (βi, γi), với tham số đầu vào **x** là một ma trận kích thước 2x1.

*2. Thiết lập hàm phân bố xác suất bất kỳ:*

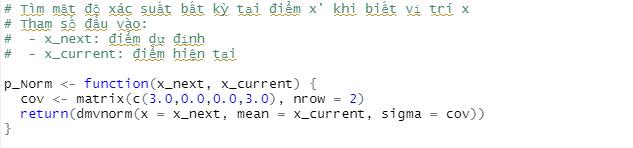
- Giả sử ta có xác suất bất kỳ của p(β, γ) có dạng phân phối Chuẩn N(a,).

- Tương tự như tìm xác suất tiên nghiệm π(β, γ), ta cũng sử dụng hàm **dmvnorm** trong thư viện **mvtnorm**.

- Ta sẽ chạy thử chương trình với ví dụ cụ thể như sau: Phân phối xác suất ngẫu nhiên p(β, γ) là phân phối Chuẩn **N(x\_current, )**

+ **x\_current** là giá trị trung bình của phân phối

+ Ma trận tương quan tuyến tính: (3.0 0.0; 0.0 3.0)



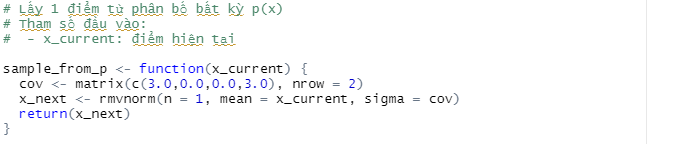
→ Hàm **p\_Norm** sẽ tính xác suất dự định p(β\*, γ\*| βi, γi), với tham số đầu vào bao gồm **x\_next** là một ma trận kích thước 2x1, là điểm dự định di chuyển đến (β\*, γ\*) và **x\_current** là một ma trận kích thước 2x1, là điểm hiện tại (βi, γi).

*3. Tìm một điểm ngẫu nhiên từ phân phối xác suất* *p(β, γ):*

Khi đã thiết lập được phân phối ngẫu nhiên p(β, γ), ta lấy một điểm ngẫu nhiên (β\*, γ\*) từ phân phối đó. Để làm được điều này, ta sẽ sử dụng hàm **rmvnorm** trong thư viện **mvtnorm.** Hàm **rmvnorm** bao gồm các tham số sau:

* n: số lượng điểm ngẫu nhiên cần tạo
* mean: giá trị trung bình x
* sigma: phương sai của x

Ta sẽ chạy thử chương trình với ví dụ cụ thể như sau: Phân phối xác suất ngẫu nhiên p(β, γ) là phân phối Chuẩn **N(x\_current, )** đã thiêt lập ở hàm **p\_Norm**.



*4. Khởi tạo các giá trị ban đầu:*

Ta thiết lập các tham số còn lại của thuật toán Metropolis-Hastings: số lượng mẫu và điểm khởi tạo. 

Để vẽ đồ thị biểu diễn mẫu chọn được ta thêm thư viện **ggplot2**

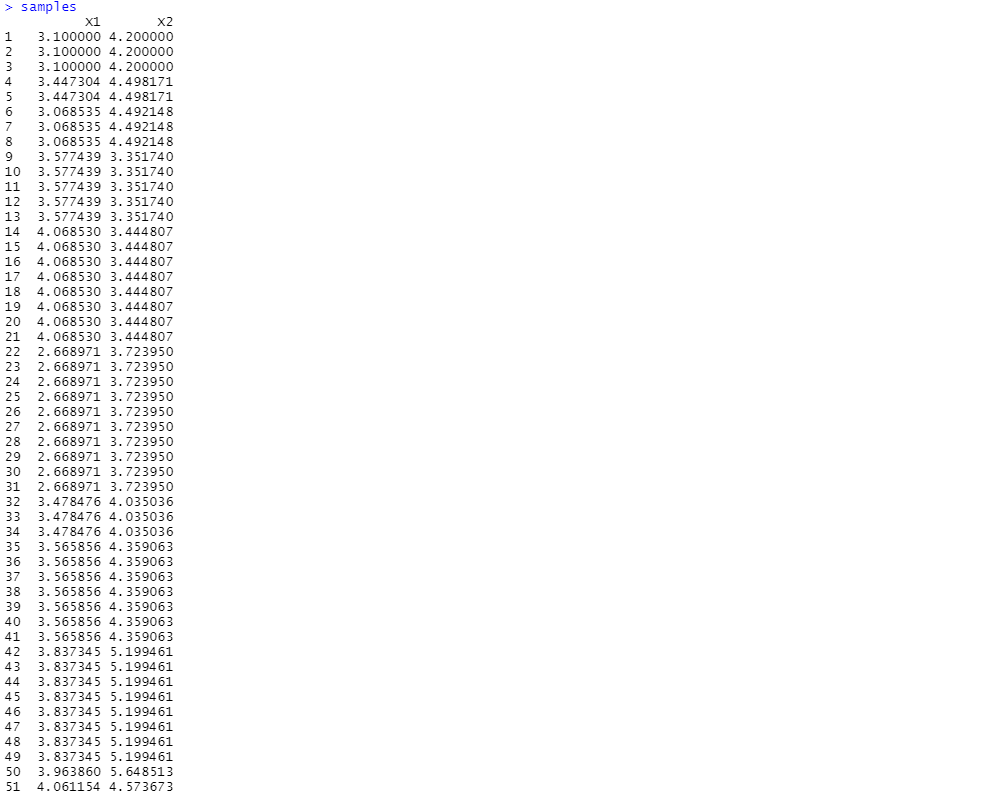


*5. Kết quả thu được:*

Sau khi đã hiện thực các tham số, ta tiến hành chạy chương trình.

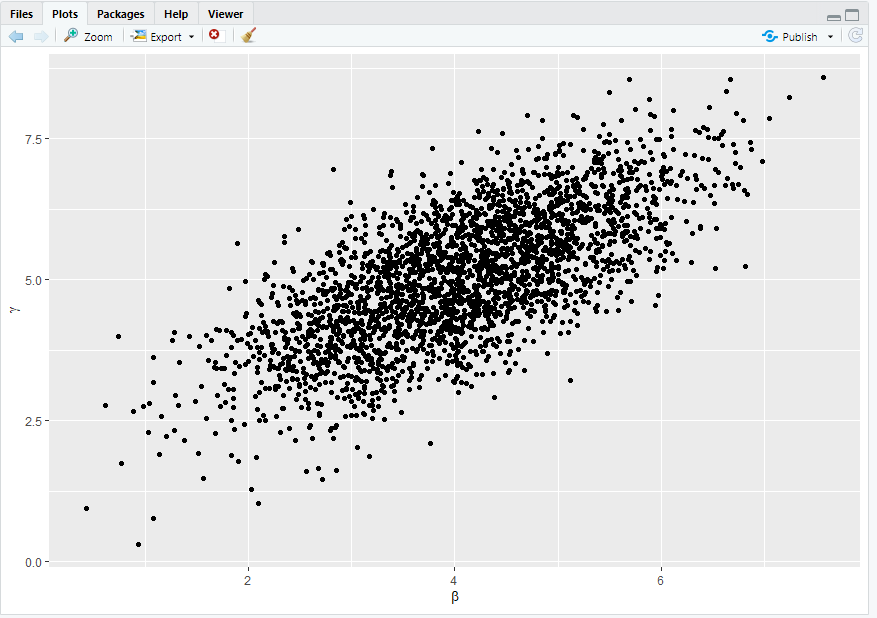
Kết quả thu được như sau:

- Để xem mẫu các β, γ ta nhập lệnh **samples** ở Console:



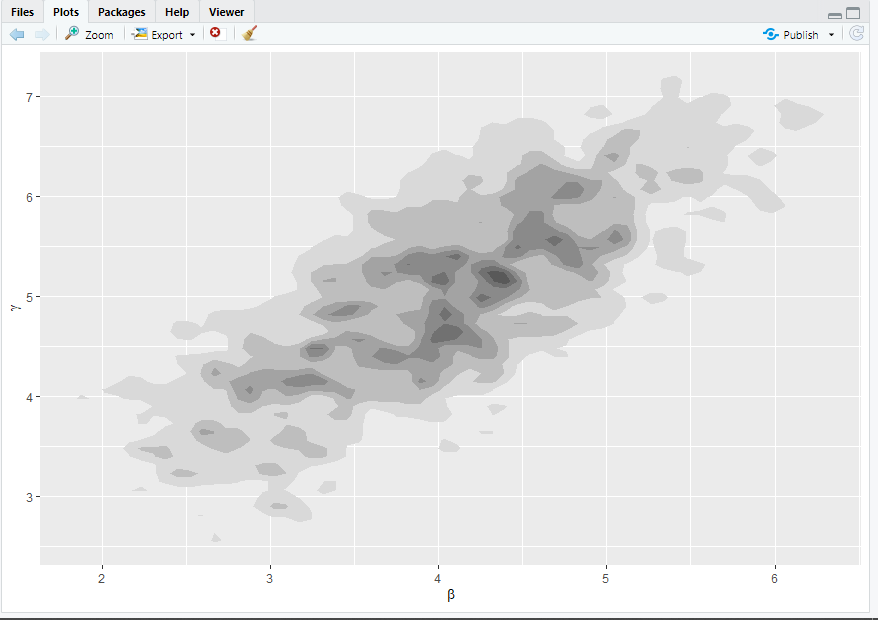
- Để xem đồ thị dạng các điểm mẫu ta chạy lệnh sau: **ggplot(data = samples,aes(x = X1, y = X2)) + geom\_point() + xlab(expression(beta)) + ylab(expression(gamma))**

→ Kết quả thu được:



- Để xem đồ thị dạng mật độ ta chạy lệnh sau: **ggplot(data = samples,aes(x = X1, y = X2)) + stat\_density2d(aes(fill = ..level..), geom = "polygon", h = 0.26) + scale\_fill\_gradient(low = "grey85", high = "grey35", guide = FALSE) + xlab(expression(beta)) + ylab(expression(gamma))**

→ Kết quả thu được:



**Bài toán 4 (Bắt buộc). Mỗi nhóm hãy tự chọn lấy một khu vực gồm một số quốc gia, dùng các chương trình ở Bài tập 3 để ước lượng giá trị trung bình hệ số R0 trong (20) ở khu vực này. Phân tích rõ chính sách hạn chế đi lại và cách ly đã ảnh hưởng đến hệ số R0 ở khu vực này như thế nào. Nêu rõ dẫn chứng. Xem tham khả trong [JKG20]. Chương trình mẫu viết bằng ngôn ngữ R có thể tham khảo trong [LM05]. Trình bày chi tiết kết quả trong báo cáo.**

**\*Ý tưởng:**

Ta có công thức tính giá trị trung bình của R0:

**E(R0) = R0(β, γ)d(β, γ)**

Tuy nhiên, tích phân (20) không thể được tính toán một cách trực tiếp. Thay vào đó chúng ta sẽ sử dụng công thức xấp xỉ (21):

**E(R0) ≈ (X | βi, γi)**

Với **(βi , γi)** được lấy ra dựa trên phân bố xác suất tiên nghiệm **π(β, γ)** và **m** là kích thước mẫu, mẫu sử dụng là kết quả thu được từ Bài toán 3.

*Tuy nhiên từ mẫu SIR ban đầu, để xác định* ***(X | βi, γi)*** *ta cần xác định phân bố xác suất của X mang phân phối gì. Ta có thể giả sử nó mang phân phối loại gì và tiến hành kiểm định xem liệu giả sử như vậy có hợp lý hay không với một mức ý nghĩa nào đó. Cụ thể các bước như sau như sau:*

*- Từ dữ liệu thống kê dịch bệnh, ta lập bảng các giá trị β, γ. Lập các khoảng giá trị sao cho tần số xuất hiện lớn hơn hoặc bằng 5.*

*- Ta gọi giả thuyết kiểm định H0: Mẫu có phân phối F(X), H1: Mẫu không có phân phối F(X).*

*- Xác định miền bác bỏ:*  ***= ((k – r – 1), +∞)****. Với (k - r - 1) là giá trị hàm Chi bình phương nằm ở cột , hàng (k – r – 1) nằm trong bảng Chi bình phương*

*(tham khảo tại* [*https://people.richland.edu/james/lecture/m170/tbl-chi.html*](https://people.richland.edu/james/lecture/m170/tbl-chi.html)*),*

*k là số các khoảng đã chia và r là số tham số của phân phối F(X).*

*- Tính tiêu chuẩn kiểm định* ***X2qs = ,*** *với Oi là tần số từ thực nghiệm, Ei là tần số lý thuyết nếu H0 đúng. Nếu X2qs  thì ta bác bỏ giả thuyết H0 chấp nhận giả thuyết H1, ngược lại ta chấp nhận giả thuyết H0.*

Tuy nhiên, ta sẽ giả sử rằng X mang phân bố xác suất Γ(λ, ν) sau khi tiến hành kiểm định. Nếu giả thuyết trên được chấp nhận, ta có thể tính được (X | βi, γi) theo công thức (15)

**π(X | βi, γi) = X(ti)^(βi−1) exp{−γX(ti)}**

, từ đó dễ dàng tính được giá trị trung bình E(R0) theo công thức (21)

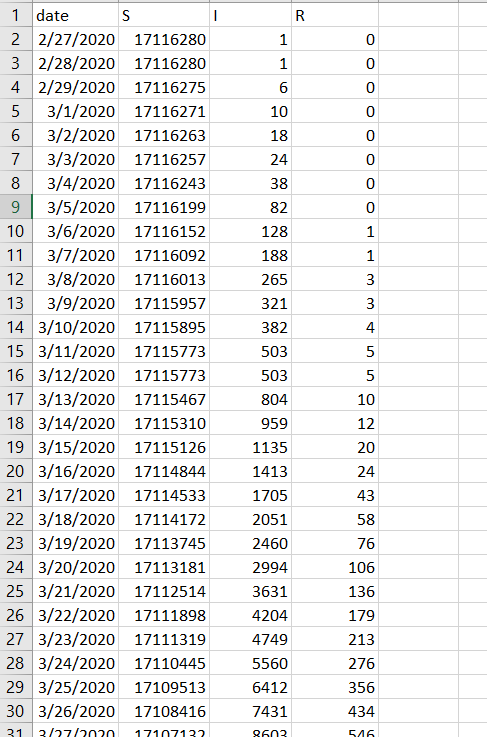
**\*Hiện thực:**

- Ở Bài tập lớn này, ta sẽ lựa chọn quốc gia để lấy thông tin khảo sát là Netherlands (Hà Lan).

*1. Đọc dữ liệu đầu vào:*

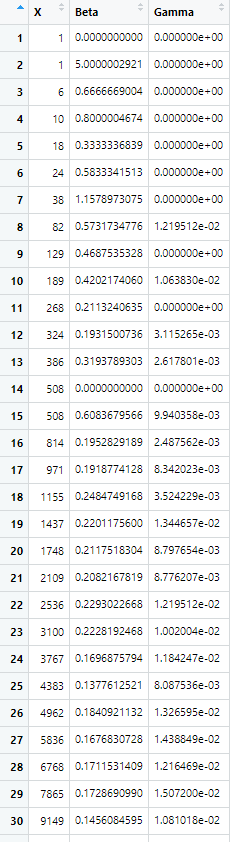
- Trước hết, ta tiến hành đọc dữ liệu đầu vào từ file Excel đã được hiệu chỉnh format để dễ dàng nhập dữ liệu, bằng cách sử dụng thư viện **xlsx.** File Excel này bao gồm 4 cột:

* **Date**: ngày khảo sát. Ngày đầu tiên được khảo sát là ngày 27/2/2020, là ngày đầu tiên xuất hiện nhiễm bệnh đầu tiên.
* **S**: ta giả định toàn bộ công dân Hà Lan đều nằm trong diện có nguy cơ bị nhiễm bệnh. Dân số Hà Lan được sử dụng là dân số vào năm 2020 với số dân là 17.116.281.
* **I**: số người mắc bệnh.
* **R**: số người đã không còn bị mắc bệnh nữa. May mắn, Hà Lan chưa xuất hiện ca tử vong nào cả nên R chính là số người hồi phục và còn sống.



- Sau đó, ta tạo một mẫu thống kê gồm 3 cột chứa thông tin như sau:

* X: chứa tổng số người nhiễm bệnh và số người hồi phục sau khi nhiễm bệnh
* Beta: chứa giá trị Beta của 2 ngày kế tiếp được tính gần đúng bằng công thức Euler:
* Gamma: chứa giá trị Beta của 2 ngày kế tiếp được tính gần đúng bằng công thức Euler:



→ Đoạn code hiện thực như sau:



*2. Hiện thực các hàm phân phối:*

Ta giả sử rằng β, γ có phân phối Gamma Γ(λ, ν). Ta thiết lập hàm phân phối Gamma như sau:

- Các đặc trưng của phân phối Gamma:

Giá trị trung bình E(X) = λ / ν, phương sai D(X) = λ / ν2

**→ ν = E(X) / D(X) và λ = E(X) \* ν**

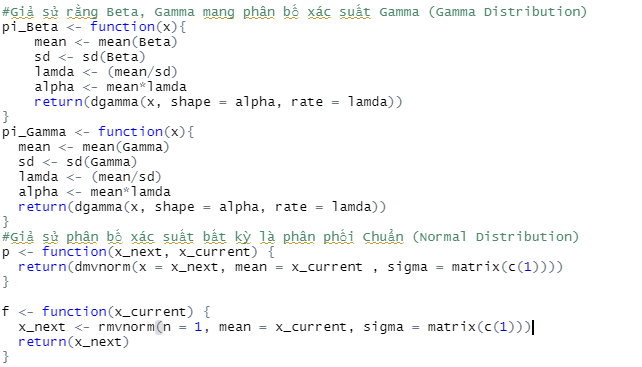
- Để tính xác suất tại một điểm theo phân phối Gamma, ta sử dụng hàm **dgamma**.

Ta cũng chọn phân phối bất kỳ có dạng phân phối Chuẩn. Để tính xác suất bất kỳ p(β\*, γ\*| βi, γi) ta sẽ biến đổi thành p(β\*, γ\*) ~ N((βi, γi), ), có thể dễ dàng tính bằng hàm **dmvnorm**.

Hàm **p** này giúp tìm ra vị trí ngẫu nhiên (β\*, γ\*) để xét cho thuật toán Metropolis-Hastings, nhưng sẽ không xuất hiện trong công thức tính hệ số xác suất di chuyển r vì phân phối Chuẩn là phân phối đối xứng.

Hàm **f** sẽ lấy ra vị trí ngẫu nhiên (β\*, γ\*) từ phân phối xác suất **p**.

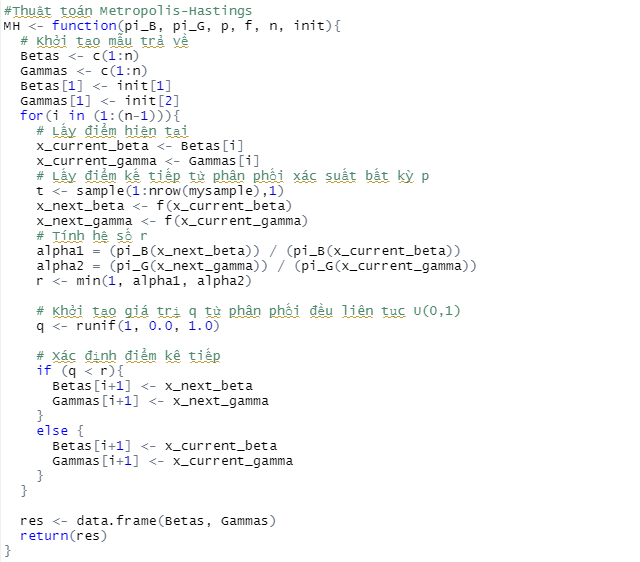
Hiện thực code của các hàm như sau:



*3. Hiện thực hàm giải thuật Metropolis-Hastings:*

Ta sẽ áp dụng giải thuật Metropolis-Hastings đã được trình bày ở Bài tập 3 để thiết lập mẫu các β, γ để tính giá trị R0.

Đoạn code hiện thực như sau:



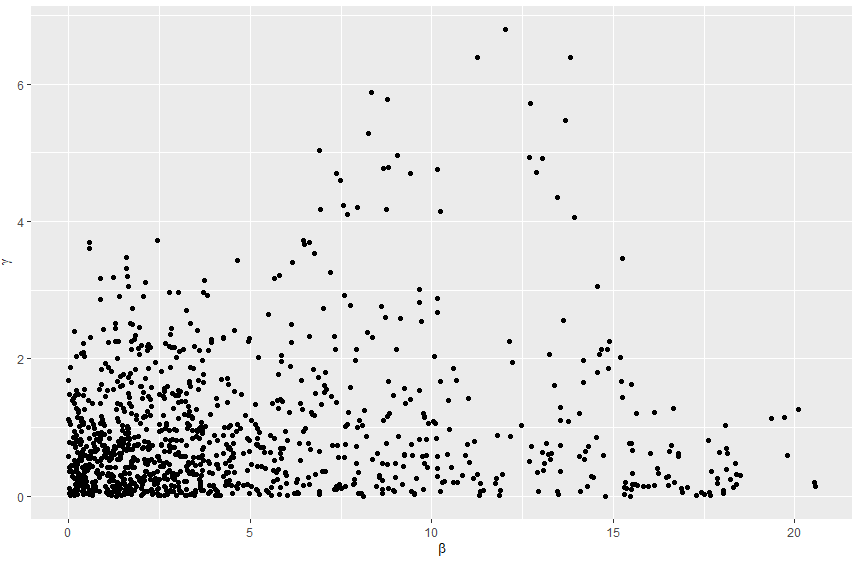
*4. Thiết lập các giá trị ban đầu:*

Ta cũng thiết lập các giá trị ban đầu cho hàm MH. Ta chọn ngẫu nhiên một điểm (β, γ) ngẫu nhiên và thiết lập kích thước mẫu n.



*5. Kết quả thu được:*

- Mẫu thu được từ việc tạo mẫu bằng thuật toán Metropolis-Hastings dưới dạng đồ thị:



- Ước lượng R0 theo công thức cho một lần chạy thử:



- Kết quả này chỉ là ước lượng, và mỗi lần chạy khác đều cho mỗi kết quả khác nhau. Sau đây là một vài kết quả khác thu được:







- Giá trị trung bình của các hệ số R0 này sẽ tính bằng công thức (21). Mặc dù các R0 thu được các trường hợp là khác nhau, nhưng đều có điểm chung là các R0 đều lớn hơn 1. Ta có thể rút ra các nhận xét đối với chính sách hạn chế đi lại và cách ly COVID-19 của nước Hà Lan như sau:

…