Twitter and Formal Demography

IUSSP Workshop on Web, Social Media Data and Demographic Methods

Emilio Zagheni

August 31, 2016

European Population Conference, Mainz, Germany

- We saw how Tweets are stored and how we can access them via the API

- We saw how Tweets are stored and how we can access them via the API
- These data can be useful to understand population processes

- We saw how Tweets are stored and how we can access them via the API
- These data can be useful to understand population processes
- Demographic methods can help us understand these data

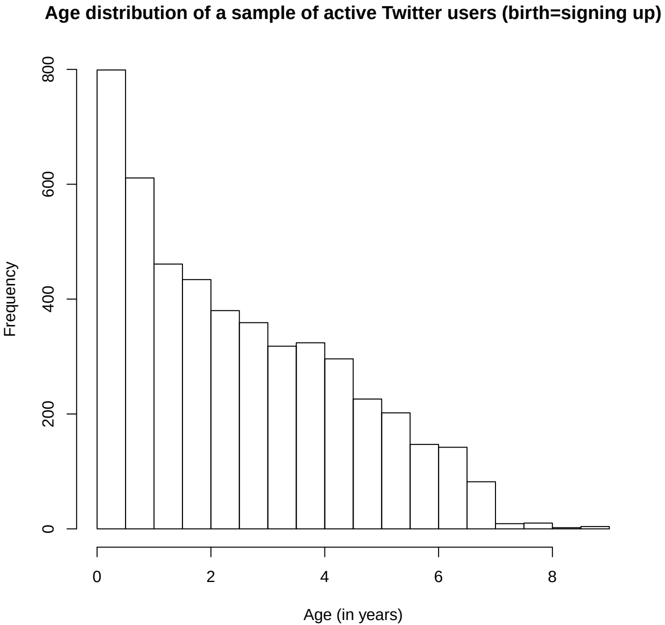
- We saw how Tweets are stored and how we can access them via the API
- These data can be useful to understand population processes
- Demographic methods can help us understand these data
- An example: estimating Twitter growth rate from a cross section of Tweets

The U2 band has 'lived" in Twitter for 7 years and 3 months

Robert Moffitt was “born” in Twitter 2 years and 3 months ago

Barack Obama has been on Twitter for 9 years

```
{ "created_at": "Wed Nov 07 04:16:18 +0000 2012",
  "id": 266031293945503744,
  "text": "Four more years. http://t.co/bAJE6Vom",
  "source": "web",
  "user": {
    "id": 813286,
    "name": "Barack Obama",
    "screen_name": "BarackObama",
    "location": "Washington, DC",
    "description": "This account is run by Organizing for Action staff.
        Tweets from the President are signed -bo.",
    "url": "http://t.co/8aJ56Jcemr",
    "protected": false,
    "followers_count": 40873124,
    "friends_count": 654580,
    "listed_count": 202495,
    "created_at": "Mon Mar 05 22:08:25 +0000 2007",
    "time_zone": "Eastern Time (US & Canada)",
    "statuses_count": 10687,
    "lang": "en" },
  "coordinates": null,
  "retweet_count": 783488,
  "favorite_count": 295026,
  "lang": "en"
}
```


Age distribution of a sample of active Twitter users (birth=signing up)
Frequency
0
200
400
600
800
0
2
4
6
8
Age (in years)

Estimating population growth rate from one census

- Problem: Given the number of individuals P_x at age x and P_y at age y, at time t, find the rate at which the births were increasing between years $t - x$ and $t - y$;
- Consider the situation where y is greater than x.

Estimating population growth rate from one census

- Problem: Given the number of individuals P_x at age x and P_y at age y, at time t, find the rate at which the births were increasing between years $t-x$ and $t-y$;
- Consider the situation where y is greater than x.
- We have

$$\underbrace{B(t-x)}_{\substack{\text{births at} \\ \text{time t-x}}} \quad \underbrace{L_x}_{\substack{\text{fraction surviving} \\ \text{x years to time t}}} \quad = \quad \underbrace{P_x}_{\substack{\text{Population size} \\ \text{of age x at time t}}}$$

$$B(t-y)L_y = P_y$$

We have

$$B(t-x)L_x = P_x$$

$$B(t-y)L_y = P_y$$

We have

$$B(t-x)L_x = P_x$$

$$B(t-y)L_y = P_y$$

which can be expressed as:

$$\frac{B(t-x)}{B(t-y)}\frac{L_x}{L_y} = \frac{P_x}{P_y}$$

We have

$$B(t-x)L_x = P_x$$

$$B(t-y)L_y = P_y$$

which can be expressed as:

$$\frac{B(t-x)}{B(t-y)}\frac{L_x}{L_y} = \frac{P_x}{P_y}$$

or

$$\frac{B(t-x)}{B(t-y)} = \frac{P_x}{P_y}\frac{L_y}{L_x}$$

Let's then assume that the birth function is exponential:

$$B(t) = B(0)e^{rt}$$

Let's then assume that the birth function is exponential:

$$B(t) = B(0)e^{rt}$$

Then

$$\underbrace{\frac{B(t-x)}{B(t-y)}}_{e^{(y-x)r}} = \frac{P_x}{P_y}\frac{L_y}{L_x}$$

Let's then assume that the birth function is exponential:

$$B(t) = B(0)e^{rt}$$

Then

$$\underbrace{\frac{B(t-x)}{B(t-y)}}_{e^{(y-x)r}} = \frac{P_x}{P_y}\frac{L_y}{L_x}$$

Thus

$$r = \frac{1}{y-x} log(\frac{P_x}{P_y}\frac{L_y}{L_x})$$

Let's then assume that the birth function is exponential:

$$B(t) = B(0)e^{rt}$$

Then

$$\underbrace{\frac{B(t-x)}{B(t-y)}}_{e^{(y-x)r}} = \frac{P_x}{P_y}\frac{L_y}{L_x}$$

Thus

$$r = \frac{1}{y-x} log(\frac{P_x}{P_y}\frac{L_y}{L_x})$$

For the specific Twitter sample we get $r \approx 0.3$

Toy example, but the message is that the demographer's toolbox can be relevant outside of standard applications