

データから未来を予測する

【できるようになること】

データは集計や計算することで、利用できるということを学びます。

新型コロナウイルスのデータを使うことで、未来を予測することができます。

※ もちろん、未来は変えることができます。

【致死率(致命率)とは?】

報道などで、致死率という言葉、数値を目にしたことはありますか?

Wikipedia では、「特定の疾病に罹患した母集団のうち、その感染が死因となって死亡する割合」と記載されています。

国ごとに集計したデータを見ます。

C	D	E	F	G	H	I	J	K	L
Province_	Country_	FLast_Upd:	Lat	Long_	Confirmed	Deaths	Recovered	Active	Combined
	US 集計				2422299	124410	663562	1558389	
	Brazil 集計				1228114	54971	679524	493619	
	Russia 集計				613148	8594	374557	229997	
	India 集計				490401	15301	285637	189463	
	United Kingdom 集計				309455	43314	1361	264780	

Wikipedia では、致死率の計算式は以下とされています。

$$\text{致命率} = \text{死亡数} / \text{罹患数}$$

上記の表から、US(アメリカ合衆国)の数値を計算してみます。

$124,410 / 2,422,299 = 0.05136...$ で、約 5.1% となります。

しかし、これは「今」の数値です。


Wikipedia ページの「注釈」にも以下のように記載があります。

ある期間における致命率、とすべき。時間を考慮しないなら rate ではなく proportion だから「致命割合」とすべき、という意見もある(「新型」インフルエンザ対策の公衆衛生学的視点)。

データから未来を予測する

【時間を考慮した致死率】

以下の表を考えます。

日付	5月1日	5月5日	5月10日	5月15日
感染者数	10人 	10人	10人	10人
死亡者数	1人	2	3	5
回復者数	1人	2	3	5
致死率(その時点)	10%	20%	30%	50%
致死率(予測)	50%	50%	50%	50%

計算している項目は以下。

$$\begin{aligned} \text{致死率(その時点)} &= \text{死亡者数} / \text{感染者数} \\ \text{致死率(予測)} &= \text{死亡者数} / (\text{死亡者数} + \text{回復者数}) \end{aligned}$$

- 感染者数が増えない状況でのシミュレーションです。
- 致死率(その時点)は、まだ死亡も回復もしていない、結果のでていない人数も母数に含まれているため、この感染症の本当の致死率は、全員が回復もしくは死亡の結果が出ないとわからない。(途中の数値は低く出る可能性がある)
- 致死率(予測)の計算では、結果が出た人を母数とするため、この感染症自体の危険性を判断するには適しているのではないかと考えられる。(この資料の筆者の永原はそう考えている)
- ただし、以下のように致死率(予測)は増減(変動)することを理解して使う必要がある。(未来は変えられる部分ともいえる)

日付	5月1日	5月5日	5月10日	5月15日
感染者数	10人	10人	10人	10人
死亡者数	1人	2	4	5
回復者数	2人	2	3	5
致死率(その時点)	10%	20%	40%	50%
致死率(予測)	33%	50%	57%	50%

データから未来を予測する

【致死率の計算】

CSV データから、致死率(その時点)と致死率(予測)を計算してみます。

さきほどの CSV に「L 列」と「M 列」を追加しました。

(ファイルを名前を付けて保存で、Excel として保存しておくのをお勧めします)

$$\begin{aligned} \text{L 列 (致死率(その時点))} &= \text{I 列} / \text{H 列} \\ \text{M 列 (致死率(予測))} &= \text{I 列} / (\text{I 列} + \text{H 列}) \end{aligned}$$

D	H	I	J	K	L	M
Country_Region	Confirmed	Deaths	Recovered	Active	致死率(その時点)	致死率(予測)
US 集計	2,422,299	124,410	663,562	1,558,389	0.051360299	0.157886321
Brazil 集計	1,228,114	54,971	679,524	493,619	0.044760503	0.074841898
Russia 集計	613,148	8,594	374,557	229,997	0.014016192	0.022429799
India 集計	490,401	15,301	285,637	189,463	0.031200997	0.05084436
United Kingdom 集計	309,455	43,314	1,361	264,780	0.139968655	0.969535534
Peru 集計	268,602	8,761	151,225	108,616	0.032617032	0.054761042

- US(アメリカ合衆国)の致死率(その時点)=5.1%
- US(アメリカ合衆国)の致死率(予測)=15.7%

大きな差があります。

【致死率の注意点】

国によっては、データが正しくないのでは?と思えるケースもあります。

例えば、United Kingdom(イギリス)の致死率(予測)は96.9%です。

さすがにそれはおかしい。

データを良く見ると、「Recovered」(回復者数)がほとんどありません。

おそらく、イギリスでは回復者数は正確に集計していないのだと思われます。

データから未来を予測する

【致死率(予測)からの計算】

致死率(予測)が計算できたということは、未来の死亡者数を予測することもできるはずですが。

ただし、ここでもいくつかの注意点があります。

- データ時点での感染者数のため、感染者数が増えれば、数値は変わっていく。
(良くない要素)
- 人々の行動によって、致死率(予測)は減らすことができ、死亡者数(予測)も減る可能性がある。
(良い要素)

以下の式で死亡者数予測を計算して「R列」に追加します。

$$\begin{aligned} \text{致死率(予測)} &= \text{死亡者数} / (\text{死亡者数} + \text{回復者数}) \\ \text{死亡者数(予測)} &= \text{感染者数} \times \text{致死率(予測)} \end{aligned}$$

D	H	I	J	K	L	M	R
Country_Region	Confirmed	Deaths	Recovered	Active	致死率(その時点)	致死率(予測)	死亡者数(予測)
US 集計	2,422,299	124,410	663,562	1,558,389	0.051360299	0.157886321	382,448
Brazil 集計	1,228,114	54,971	679,524	493,619	0.044760503	0.074841898	91,914
Russia 集計	613,148	8,594	374,557	229,997	0.014016192	0.022429799	13,753
India 集計	490,401	15,301	285,637	189,463	0.031200997	0.05084436	24,934
United Kingdom 集計	309,455	43,314	1,361	264,780	0.139968655	0.969535534	300,028
Peru 集計	268,602	8,761	151,225	108,616	0.032617032	0.054761042	14,709
Chile 集計	259,064	4,903	219,327	34,834	0.018925825	0.021865941	5,665

US(アメリカ合衆国)は382,448人の予測になります。

このデータの時点での死亡者数が124,410人なので、まだ、これから大変な状況が続くことが想定されます。

ただし、ワシントン大学保健指標評価研究所(IHME)の新型コロナウイルス死亡者数予測を見ると、公共の場所で95%の人がマスクをすれば、かなり死亡者予測が減るシミュレーションをしていることなどがわかります。

<https://covid19.healthdata.org/united-states-of-america>

数値を鵜呑みにせず、うまく扱えるようになりたいと思います。

データから未来を予測する

【日本の数値】

ここで、日本のデータを見たいと思います。(2020-05-25 のデータ)

D	H	I	J	K	L	M	R
Country_Region	Confirmed	Deaths	Recovered	Active	致死率 (その時点)	致死率 (予測)	死亡者数 (予測)
Japan 集計	18,055	971	16,142	942	0.053780116	0.05674049	1,024

死亡者数(予測)も実際の死亡者数に近くなっています。

致死率(その時点)と致死率(予測)の数値も、かなり近いものになっています。

これをどう読むか、考えてみるのも良いかと思います。