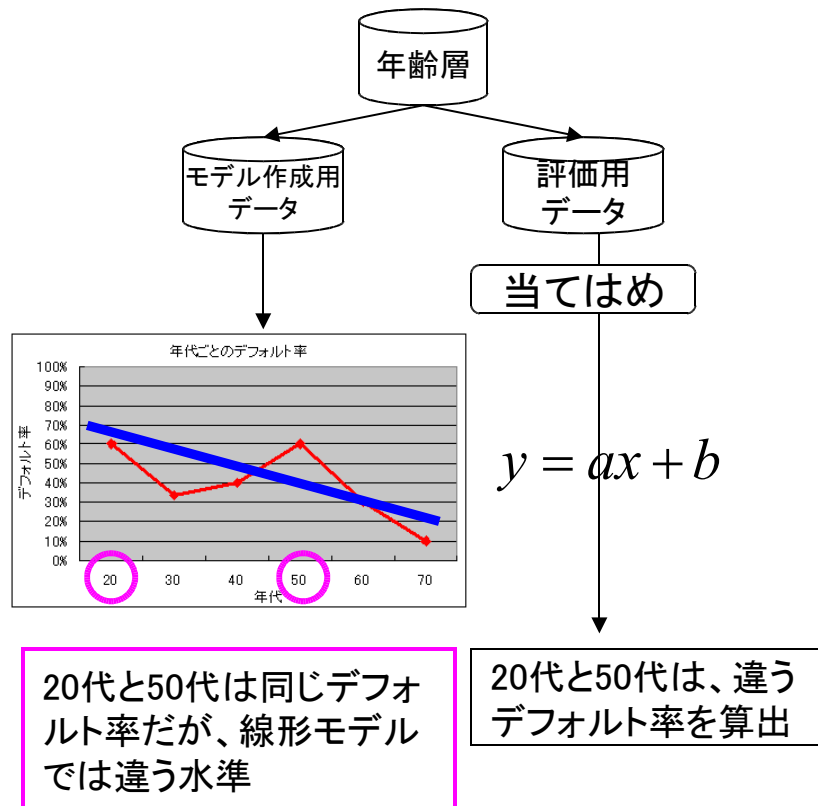
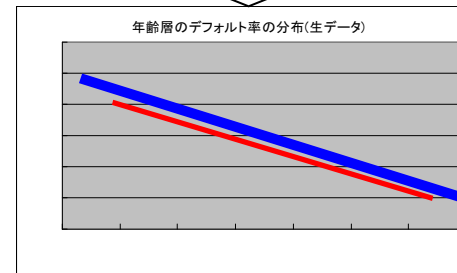
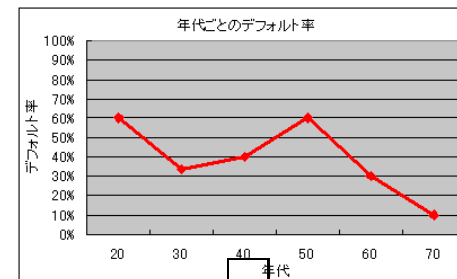


## 2 発想の原点

1変数で考えた場合



変数の値とデフォルト率の関係を線型にできれば...



**WOE変換**

生データの場合より  
精度up

### 3 WOEとは

#### 由来

1950年、ジョン・グッド(統計学者)が提唱、地質学の分野で応用。  
その後、医療の分野でも利用。  
近年、信用リスクの分野でも外資系の金融機関で利用され始めている。

WOE(Weights Of Evidence) ターゲットとする変数が、デフォルトか正常かの2値の場合

$$\text{Weights Of Evidence}_{\text{attribute}} = \log\left(\frac{\text{P\_Default}_{\text{attribute}}}{\text{P\_Normal}_{\text{attribute}}}\right)$$

$$\text{P\_Normal}_{\text{attribute}} = \frac{\text{Number Of Normal}_{\text{attribute}}}{\text{Number Of Normal}_{\text{total}}}$$

$$\text{P\_Default}_{\text{attribute}} = \frac{\text{Number Of Default}_{\text{attribute}}}{\text{Number Of Default}_{\text{total}}}$$

年齢層	カテゴリ値1	count	Normals	P_Normal attribute	Defaults	P_Default attribute	Default rate	WOE値	カテゴリ値2
missing	missing	50	42	2.3%	8	4.1%	16.0%	0.579	0.579
20-22	20	200	152	8.4%	48	24.9%	24.0%	1.084	1.084
23-26	23	300	246	13.6%	54	28.0%	18.0%	0.720	0.720
27-29	27	450	405	22.4%	5	2.6%	1.1%	-2.158	-2.158
30-34	30	500	475	26.3%	25	13.0%	5.0%	-0.708	-0.708
35-44	35	350	339	18.8%	11	5.7%	3.1%	-1.191	-1.191
44+	44	150	147	8.1%	3	1.6%	2.0%	-1.655	-1.655
合計		2000	1807		193		9.7%		