

## 欠測の対処法

- ①削除法 : データがない組み合わせは削除
- ②補完法 : 回帰式に入れて得られた推測値を穴埋めする
- ③モデル化法: 欠測の理由を盛り込む

# 欠測処理

## 削除法

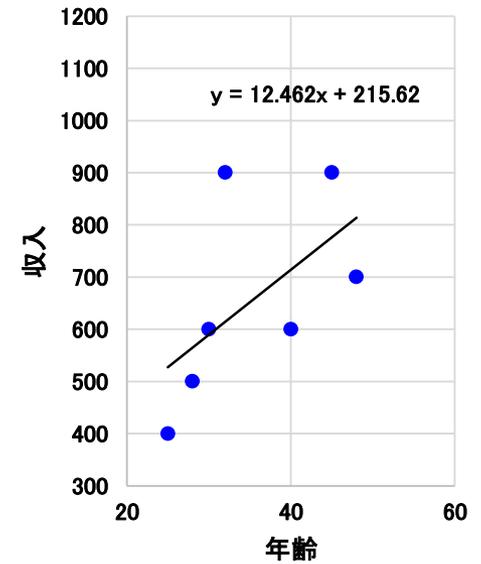
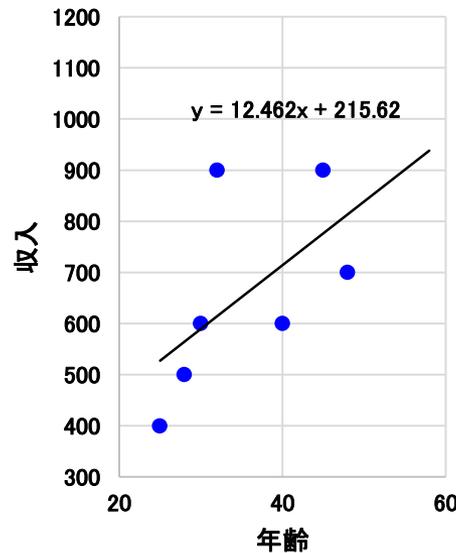
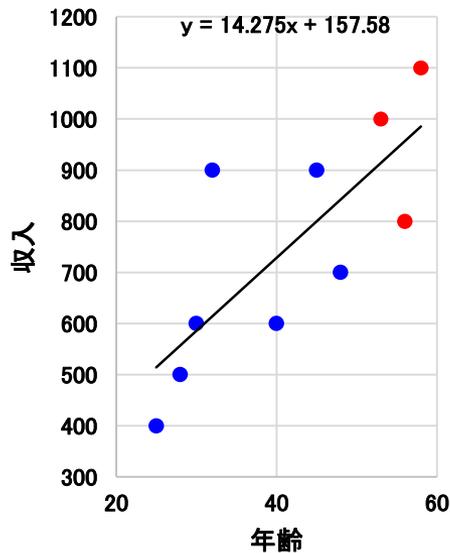
ID	年齢	収入
1	25	400
2	28	500
3	30	600
4	32	900
5	40	600
6	45	900
7	48	700
8	53	1000
9	56	800
10	58	1100
平均	41.5	750.0
標準偏差	12.2	227.3
相関係数	0.769	

ID	年齢	収入
1	25	400
2	28	500
3	30	600
4	32	900
5	40	600
6	45	900
7	48	700
8	53	
9	56	
10	58	
平均	41.5	657.1
標準偏差	12.2	190.2
相関係数	0.583	



ID	年齢	収入
1	25	400
2	28	500
3	30	600
4	32	900
5	40	600
6	45	900
7	48	700
平均	35.4	657.1
標準偏差	8.9	190.2
相関係数	0.583	

欠測データ



# 補完法

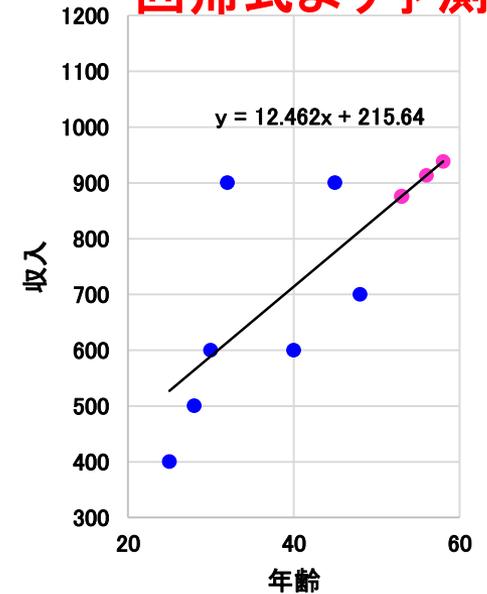
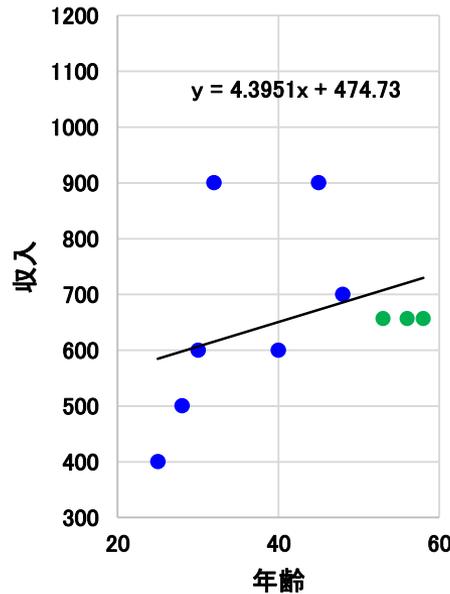
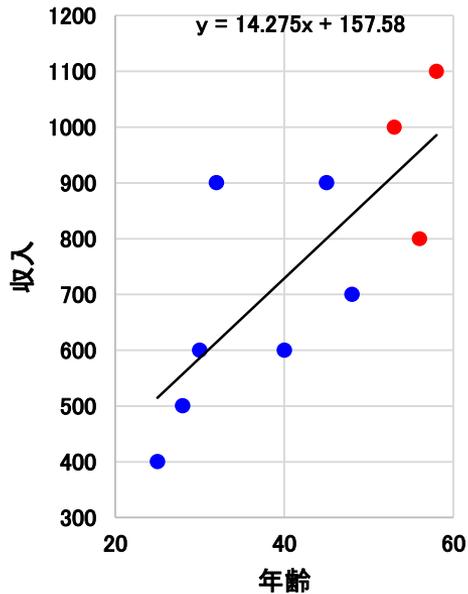
ID	年齢	収入
1	25	400
2	28	500
3	30	600
4	32	900
5	40	600
6	45	900
7	48	700
8	53	1000
9	56	800
10	58	1100
平均	41.5	750.0
標準偏差	12.2	227.3
相関係数	0.769	

ID	年齢	収入
1	25	400
2	28	500
3	30	600
4	32	900
5	40	600
6	45	900
7	48	700
8	53	657.1
9	56	657.1
10	58	657.1
平均	41.5	657.1
標準偏差	12.2	155.3
相関係数	0.346	

ID	年齢	収入
1	25	400
2	28	500
3	30	600
4	32	900
5	40	600
6	45	900
7	48	700
8	53	876.1
9	56	913.5
10	58	938.4
平均	41.5	732.8
標準偏差	12.2	198.0
相関係数	0.771	

平均値を入れる

$y = 12.462x + 215.62$   
回帰式より予測



n=10個の部品をd=31日間、寿命試験し、31日以前に寿命のm=8個は日数を記録、31日目で故障していないものは2個ありました。寿命は？

No.	1	2	3	4	5	6	7	8	9	10
Days	1	2	3	7	8	12	24	30	31+	31+

平均: 10.9 (m=8)

平均: 14.9 (n=10)

平均: 18.6 (m=8)

寿命は平均値 $\mu$ の指数分布に従うとすると、

$$\begin{aligned}
 \hat{\mu} &= \frac{1}{m} \{ (x_1 + x_2 + \cdots + x_m) + (n - m)d \} \\
 &= \frac{1}{8} \{ (1 + 2 + \cdots + 30) + (10 - 8) \times 31 \} \\
 &= 18.6
 \end{aligned}$$