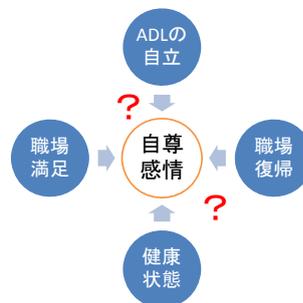


重回帰分析

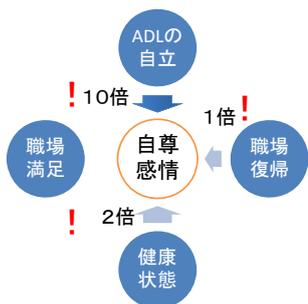
15

自尊感情は何に影響を受けてる？



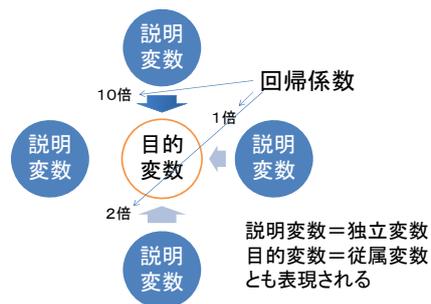
16

自尊感情は何に影響を受けてる？



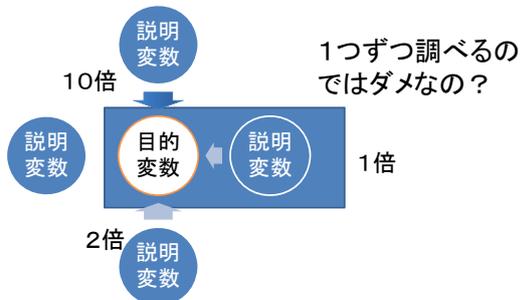
17

重回帰分析



決定係数: 説明変数で、目的変数がどれくらい説明できたか？¹⁸

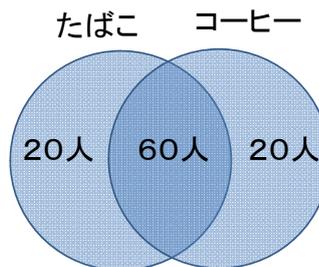
重回帰分析の必要性



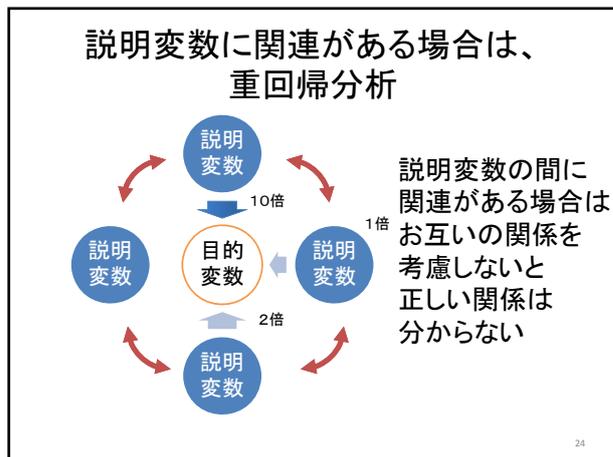
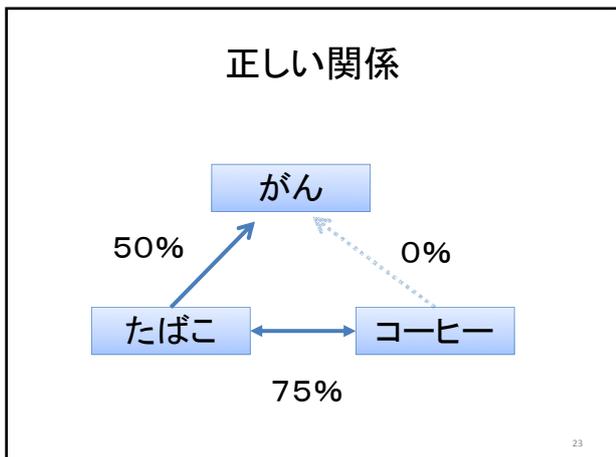
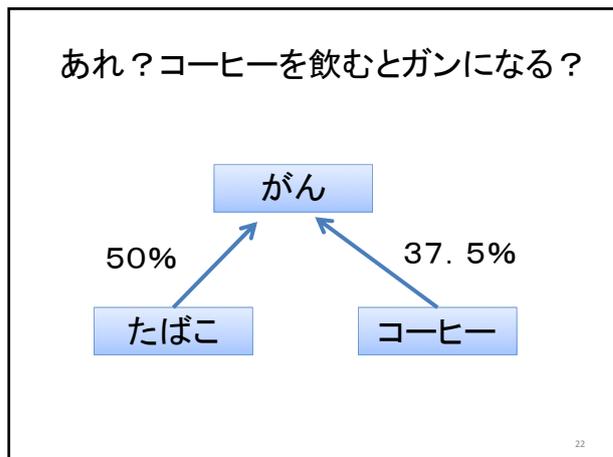
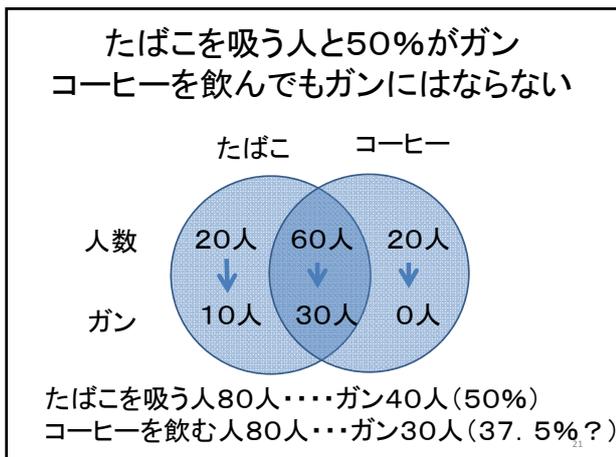
19

事例) たばことコーヒーとがん

100人中、たばこを吸う人が80人、コーヒーを飲む人は80人、両方の人が60人いたとします。



20



どの説明変数を入れるのが難しい

- 説明変数同士に関連があるので、説明変数を1つ増やしたり減らしたりすると、他の回帰係数が変わってしまう
- そこで、説明変数を増やしたり、減らしたりして、どの組み合わせがよいか調べる方法の1つがステップワイズ法

どうやってよいモデルと判定する？

- 変数の組み合わせが2つある場合、どちらがよい組み合わせなのか判定できないと意味がありません。
- 判定指標
 - 決定係数
 - 自由度調整済み決定係数
 - AIC(赤池情報量規準)

1) 決定係数

- 重回帰分析では、説明変数によって目的変数の予測値が計算できません。
- 目的変数の予測値と実際の値の相関係数を「重相関係数」といいます。
- 重相関係数の二乗が決定係数です。
- “目的変数の分散のうち、何%の分散を説明できているか”という意味になります。

27

2) 自由度調整済み決定係数

- 決定係数の欠点
 - 説明変数が多いほど決定係数は高くなるので、説明変数を2つ入れるよりも3つ入れた方が決定係数は高くなります。そうすると、ほとんど予測に役立っていない説明変数であってもたくさん説明変数を入れた方が決定係数が高くなります。
- そこで、説明変数の数によって決定係数を調整する(多くの決定係数を入れた場合は低くなるように調整する)。これが自由度調整済み決定係数です。

28

3) AIC(赤池情報量基準)

- 情報量という観点からモデルの当てはまりの良さを計算する方法
- 予測値と実値が等しければ0になる。小さいほどよい指標
- いくつであればよいという絶対的な基準ではなく、複数のモデルのうちより小さい方がよいという判定に使う。

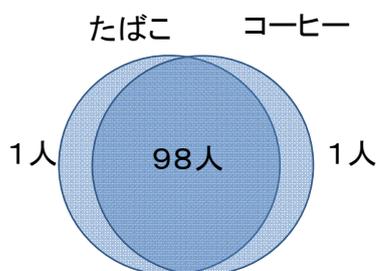
29

多重共線性とは

- 2つの説明変数の間に非常に高い相関があるとどうなるか？
- ↓
- どちらの変数の効果が判定できない。

30

多重共線性のイメージ



この場合、どんなに詳しく調査してもたばこの影響なのかコーヒーの影響なのか分からない。 31

多重共線性が起こっているか？

- 説明変数間で相関が高いかどうかを調べる
 - 相関行列を作る
- 複数の説明変数間で関連があるか調べる
 - VIFを調べる

32

多重共線性があったらどうする？

- 非常に高い相関がある2変数の場合、どちらか1変数のみを使う
- いくつかの変数を合成した合成変数を作る
 - 主成分分析
 - 因子分析でグループ別に因子得点を求める

33

例) 看護師の臨床能力は何で説明できる

- 臨床能力 ← 経験年数、年齢……
 - 目的変数: 臨床能力
 - 説明変数: 経験年数、年齢、……
- 年齢 = 経験年数は非常に高い相関があるので、重回帰分析を行うと、
 - ・ 経験年数が回帰係数が高く、年齢は0に近い
 - ・ 経験年数の回帰係数がほぼ0で、年齢が高いのどちらかになる。

34