

第5回 回帰分析

2007年10月09日

2007/10/09

前回の復習

- 散布図を作成した
 - 散布図と相関の関係について
- 相関係数を求めた
 - 散布図と相関係数の関係について
- 単回帰分析を行った
 - 一方のデータから他方のデータを予測する

2007/10/09

今日は

- 予定を変更して回帰分析について勉強する
- データは
 - http://www.ae.keio.ac.jp/~satoru_y/sho uei/05.xls
 - からダウンロード
- データの出典
 - マンガでわかる統計学
 - [回帰分析編]より
 - (高橋信 著, オーム社)



2007/10/09

回帰分析とは

- いくつかの変数から別の変数を予測する
 - 予測に用いる変数...説明変数, 独立変数
 - 予測される変数...目的変数, 従属変数
- 回帰式と呼ばれる式を求めて予測に用いる
- 単回帰分析...説明変数が1つ
- 重回帰分析...説明変数が複数

2007/10/09

単回帰分析

- 手順
 1. 説明変数と目的変数にがあるかを散布図を描いて調べる
 - 関係がないなら回帰分析をする意味がない
 2. 回帰式を求める
 3. 回帰式の精度を確認する
 4. 回帰係数の検定を行う
 5. 母回帰を推定する
 6. 予測する

2007/10/09

回帰式を求める

- 回帰式
 - 売上 = $a \times \text{最高気温} + b$
- 計算によって求まるが, 面倒なので分析ツールを使う
 - 分析ツールの回帰分析
 - Yの範囲...売上
 - Xの範囲...最高気温
 - 得られた結果の3つ目の表の「係数」が回帰式
 - 売上 = $3.7 \times \text{最高気温} - 36.3$

2007/10/09

精度の確認

- 得られた結果の一番上の表
 - 重相関R・・・相関係数(の絶対値)
 - 重決定R²・・・「重相関R」の2乗(決定係数, 寄与率という)
- これらの値が1に近い
 - 説明変数によって目的変数がよく表わされている
 - あてはまりが良い
- 一応の目安・・・決定係数が0.5以上

2007/10/09

回帰係数の検定

- 仮設: (最高気温の)回帰係数が0ではないか?
 - もし0だったら, 売上が温度に関係ないことになる
- 回帰係数が0かどうかを調べる
 - 回帰係数の検定
- 得られた結果の「分散分析表」に注目
 - 「回帰」の行の「有意 F」の項目
 - この値が0.05以下なら回帰係数が0でないと言える
 - 0.05より大きい場合, 回帰係数が0でないとは言いきれない

2007/10/09

母回帰の推定, 予測

- 難しいのでごく簡単に
 - 母回帰の推定・・・例えば最高気温が31°Cだった時に $A \times 31 + B$ の値がだいたいどの区間におさまっていることはわかる. その区間(信頼区間)を求めること.
 - 予測・・・例えば明日の予想最高気温が27°Cだったとする. その時に, 売上がどのくらいになるかを予測することができる. また, その予測区間を求めることができる.
 - ちなみに, 信頼区間より予測区間のほうがちょっとだけ広い.

2007/10/09

重回帰分析

- 基本的な流れは単回帰分析とおなじ.
 1. 各説明変数と目的変数に関係があるかを散布図を描いて調べる
 - できれば説明変数の数だけ散布図を描く
 2. 回帰式を求める
 3. 回帰式の精度を確認する
 4. 回帰係数の検定を行う
 5. 母回帰を推定する
 6. 予測する

2007/10/09

回帰式を求める

- 回帰式
 - 売上 = $a_1 \times \text{面積} + a_2 \times \text{距離} + a_3 \times \text{年齢} + b$
- 分析ツールを使う
 - 分析ツールの回帰分析
 - Yの範囲・・・売上
 - Xの範囲・・・面積, 距離, 年齢
 - 得られた結果の3つ目の表の「係数」が回帰式
 - 売上 = $42.2 \times \text{面積} + 0.33 \times \text{距離} + 1.09 \times \text{年齢} - 17.7$

2007/10/09

精度の確認

- 得られた結果の一番上の表
 - 重相関R・・・重相関係数
 - 重決定R²・・・「重相関R」の2乗(決定係数, 寄与率という)
 - 補正 R²・・・自由度調整済み重相関係数
 - 補正 R²という言葉はない
 - 説明変数を増やせば増やすほど重相関係数の値は大きくなる性質がある
 - 目的変数に無関係な説明変数でも!
 - それを補正した値

2007/10/09

精度の確認

- 自由度調整済み重相関係数の値が1に近い
 - 説明変数によって目的変数がよく表わされている
 - あてはまりが良い
- 一応の目安・・・0.5以上

2007/10/09

回帰係数の検定(包括的)

- 仮設: 回帰係数が全て0ではないか?
 - もし0だったら, 各説明変数が売上に関係ないことになる
- 得られた結果の「分散分析表」に注目
 - 「回帰」の行の「有意 F」の項目
 - この値が0.05以下なら回帰係数が0でないと言える
 - 0.05より大きい場合, 回帰係数が0でないとは言い切れない

2007/10/09

回帰係数の検定(個別)

- 仮設: 各回帰係数が0ではないか?
 - もし0だったら, 各説明変数が売上に関係ないことになる
- 得られた結果の一番下の表に注目
 - 各説明変数の「P-値」
 - この値が0.05以下ならその回帰係数が0でないと言える
 - 0.05より大きい場合, その回帰係数が0でないとは言い切れない
 - 年齢は0.05より大きいので, この分析では不要な説明変数の可能性が非常に高い

2007/10/09

母回帰の推定, 予測

- 単回帰と同じで省略

2007/10/09