

第4回 相関係数と単回帰分析

2007年10月02日

2007/10/02

前回の続き

- 前回は離散型分布の2項分布, ポアソン分布について形と性質の勉強をした
- 2項分布
 - 平均=試行回数×当たる確率
 - 分散=試行回数×当たる確率×(1-当たる確率)
- ポアソン分布
 - 平均=分散=事象が起こる平均回数

2007/10/02

連続型

➤ 正規分布

$$N(\mu, \sigma^2), f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- 例:身長・体重の誤差
- 分析ツール→乱数発生→正規
 - 変数の数, 乱数の数
 - 平均
 - 標準偏差

2007/10/02

他にも

➤ 連続型でよく出てくる分布

- t分布
- カイ2乗分布
- F分布
- etc.

2007/10/02

何で分布が必要なのか

- 検定をするときに使う
 - 検定をしたい値(たとえば平均)がある分布に従っていることが分かっているから
 - 有意か有意でないか(差があるかないか)の判断の時に分布の値に基づいた判断が行われる
- 分布についてよくわからない場合・・・
 - そういうものだと思ってください
 - むしろ「検定」の方法・使い方・結果の読み取り方を覚えておいたほうが役に立つ・・・はず

2007/10/02

ここまでが前回の話

2007/10/02

散布図(相関図)

- 2種類の変量について調べたデータを平面に図示したもの
 - (例)身長と体重, 試験の数学の点と英語の点
 - 散布図は
 - 正の相関
 - 負の相関
 - 無相関(相関がない)
- の3種類に分けられる

といってもピンとこないので, 実際に散布図を作成してみよう

2007/10/02

散布図の作り方

- 準備として
 - http://www.ae.keio.ac.jp/~satoru_y/sho/uei/04.xls
- をダウンロード
- このデータは永田・棟近著, 多変量解析入門, サイエンス社のP132の表9.1のデータを参考に値を変更

2007/10/02

散布図の作り方

- まず, 国語と英語の散布図を作る
 1. A2からB11まで選択
 2. 「挿入」→「グラフ」
 3. グラフ ウィザードで「散布図」を選び「次へ」
 4. データ範囲や名前を確認し「次へ」
 5. グラフのタイトルや各軸の名前等を設定し「次へ」
 6. 最後にグラフの作成場所を選び「完了」
 7. できたグラフを適当に調整して完成
- 同様に, 英語と数学, 英語と理科の散布図を作る

2007/10/02

散布図の形状を比べてみよう

- 国語と英語
 - 右上がり
 - 正の相関
- 英語と数学
 - 右下がり
 - 負の相関
- 英語と理科
 - ばらばら
 - 無相関

2007/10/02

相関係数の前に・・・

- 平均 (Excelの関数だとAVERAGE)
 - 各データの和をデータの個数で割ったもの
- 分散 (Excelの関数だとVAR)
 - 各データから平均を引いて2乗したものとの和を **データの個数-1**で割ったもの
- 共分散 (Excelの関数だとCOVAR)
 - (1つ目の各データから平均を引いたもの) × (2つ目の各データから平均を引いたもの)の和を **データの個数**で割ったもの

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

2007/10/02

共分散と相関係数

- 相関係数 (Excelの関数だとCORREL)

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \approx \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}}$$

- 注意: 分散と共分散から相関係数は計算できるが, 分散でVARを使っている場合, **データの個数-1**で割っているため上の式の右辺の値は異なる

2007/10/02

相関係数とは

- 実際に相関係数を計算してみよう
 - 国語と英語(正の相関)
 - 0.923
 - 英語と数学(負の相関)
 - -0.833
 - 英語と理科(無相関)
 - -0.056

2007/10/02

相関係数とは

- 正の相関がある時
 - 1に近い
- 負の相関がある時
 - -1に近い
- 無相関の時
 - 0に近い

2007/10/02

相関係数から単回帰分析

- 正または負の相関がある時、散布図の点を直線で表す
 - 単回帰分析という
 - 一方の値からもう一方を予測したい時に使う
 - (例)気温からビールの売上
- 細かい計算方法は省略
 - ちょっと難しい話になるので興味があればご自分で勉強しましょう

2007/10/02

回帰直線の求め方

- 一番簡単な方法
 1. Excelの散布図上で散布図の点を右クリック
 2. 開いたメニューから「近似曲線の追加」
 3. 「線形近似」を選択
 4. オプションで「グラフに数式を表示する」にチェック
- その他の方法
 - 分析ツールの「回帰分析」
 - LINESTという関数

2007/10/02