QC検定2級　受験前講習会

4.相関分析・単回帰分析

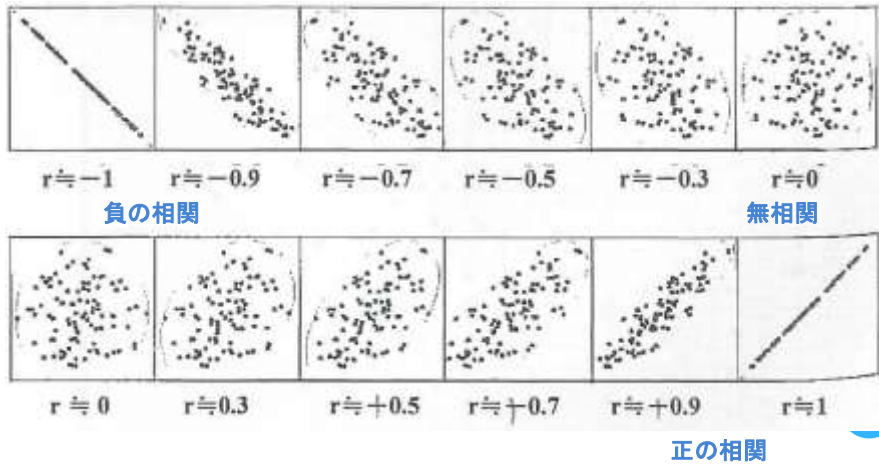
会津オリンパス株式会社
古川　英嗣

このテキストには著作権があり、許可無く資料のコピー・掲載はできません。
また、セミナーの録音もこれに該当し、録音できません。

相関分析

テキストP.154～156

相関分析：対応する2つの変数(x , y)間の直線的な関係の強さを調べる手法で，**散布図**を作成し，**相関係数(r)**を使って，その関係の強さを表現する



相関係数の計算式

テキストP.157

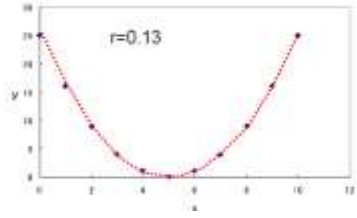
相関係数　$r = \dfrac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$　　(-1≦r≦1)

$S_{xx} = \sum x^2 - \dfrac{(\sum x)^2}{n}$　　xの平方和

$S_{yy} = \sum y^2 - \dfrac{(\sum y)^2}{n}$　　yの平方和

$S_{xy} = \sum xy - \dfrac{\sum x \cdot \sum y}{n}$　　xとyの積和

※注意
相関係数が低い＝2変数に関係がないではない。
直線的な関係の強さを示す指標。

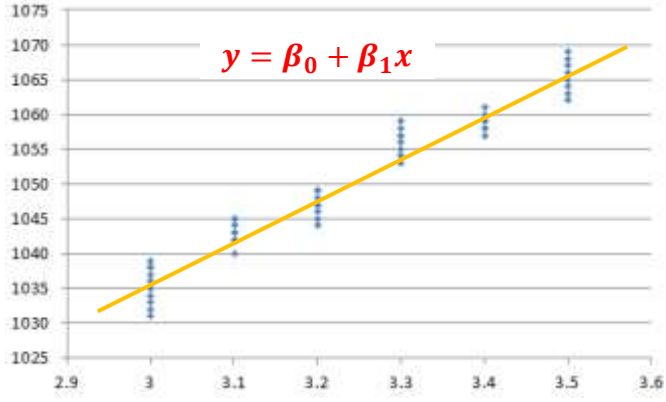


3

単回帰分析

テキストP.171～177

単回帰分析：説明変数(x)が目的変数(y)に及ぼす**影響度の強さ**を調べる手法。散布図を作成し，それらの関係を**回帰式**で表現する。
説明変数が一つ⇒単回帰分析



4

回帰式の推定式

$$y_i = \beta_0 + \beta_1 x + \varepsilon_i$$

各組のデータから傾き(β_1)と切片(β_0)を算出する(**最小二乗法**)。

※次ページで補足説明。

β_0とβ_1の推定値を、b_0とb_1とする。

$$\hat{y} = b_0 + b_1 \cdot x$$

$$b_1 = \frac{S_{xy}}{S_{xx}}$$

$$b_0 = \bar{y} - b_1 \cdot x = \bar{y} - \frac{S_{xy}}{S_{xx}} \cdot \bar{x}$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} \quad \text{xの平方和}$$

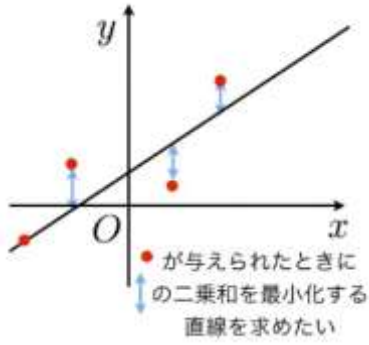
$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n} \quad \text{yの平方和}$$

$$S_{xy} = \sum xy - \frac{\sum x \cdot \sum y}{n} \quad \text{xとyの積和}$$

寄与率 $R^2 = \frac{S_R}{S_{yy}} = \left(\frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}\right)^2$

5

補足説明;最小二乗法について



2つセットのデータの組 (x_i , y_i) が n 個与えられた状況を考えます。

そして x_i と y_i **に直線的な関係があると推察できるときに，ある意味で一番もっともらしい直線を引く**のが最小二乗法です。

直線とy方向の誤差(ズレ)の二乗和が最小になるのが最もらしい直線であると考える。

6

回帰分析における検定

テキストP.178～180

2変数間に直線関係があるかどうか？

回帰式の検定⇒**分散分析表**を用いた**F検定**により検定される

[仮説]

帰無仮説　$H_0: \beta_1 = 0$　　直線関係が無いという仮説

対立仮説　$H_1: \beta_1 \neq 0$　　直線関係が有るという仮説

分散分析表

複数の母集団に対してそれぞれの有意性を検討する手法。各要因によって変化するバラつきを意味のない変動(**誤差**)と意味のある変動(**要因効果**)の分散を分け，その分散比(**F検定**)を求めることで，**要因効果が誤差に比べて大きければ有意であると判定する**方法。

7

分散分析表

回帰分析の問題でかなりの頻度で登場。
この後の実験計画法でも必須。

要因	平方和	自由度	平均平方(分散)	分散比F_0
回帰 R	$S_R = \frac{S_{xy}^2}{S_{xx}}$	$\varphi_R = 1$	$V_R = {}^{S_R}/_{\varphi_R}$	${}^{V_R}/_{V_e}$
残差 e	$S_e = S_{yy} - S_R$	$\varphi_e = \varphi_T - \varphi_R$	$V_e = {}^{S_e}/_{\varphi_e}$	
計 T	$S_T = S_{yy}$	$\varphi_T = n - 1$		

$F_0 = {}^{V_R}/_{V_e}$ の棄却域は，

$F_0 \geqq F(\varphi_R, \varphi_e; \alpha)$

であれば，帰無仮説を棄却し，直線関係に意味があると判定する。

8

例題

製品HKの製造工程において、水分含有量と不純物量は重要な管理項目となっている。最近の生産品より、ランダムにn=32のサンプルを抜き取り、それぞれのサンプルから水分含有量(x)と不純物量(y)を測定したところ、以下の統計量が得られた。

xとyの間に直線関係が存在するかどうかを検討する。有意水準は5%とする。

(統計量)

$$\sum x = 164.0 \quad \sum x^2 = 950.0 \quad \bar{x} = 5.125 \quad S_{xx} = 109.50$$

$$\sum y = 1280.0 \quad \sum y^2 = 147.0 \quad \bar{y} = 40.0 \quad S_{yy} = 97.00$$

$$\sum xy = 746.0 \quad S_{xy} = 90.00$$

9

例題

回帰母数の予測値

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{90.00}{109.50} = 0.82$$

$$b_0 = \bar{y} - b_1\bar{x} = 40.0 - 0.82 \times 5.125 = 35.8$$

$$b_0 + b_1\bar{x} = 35.8 + 0.82x$$

分散分析表作成のための計算

$$S_R = \frac{S_{xy}^2}{S_{xx}} = \frac{90^2}{109.50} = 73.97 \qquad \varphi_R = 1$$

$$S_T = S_{yy} = 97.00 \qquad \varphi_T = n - 1 = 31$$

$$S_e = S_{yy} - S_R = 97.00 - 73.97 = 23.03 \qquad \varphi_e = \varphi_T - \varphi_R = 30$$

10

例題

要因	平方和	自由度	平均平方(分散)	分散比F_0
回帰 R	$S_R = 73.97$	$\varphi_R = 1$	$V_R = 73.97$	$V_R / V_e = 96.3$
残差 e	$S_e = 23.03$	$\varphi_e = 30$	$V_e = 0.768$	
計 T	$S_T = 97.00$	$\varphi_T = 31$		

$F(\varphi_R, \varphi_e; \alpha)$=F(1,30:0.05)=4.17

$F_0 = 96.3 > 4.17$ より回帰による変動Rは有意である。

⇒xとyの間に直線関係があると判定する。

11

演習問題
[問題13,14]

12

まとめ

・相関分析とは対応する2つの変数(x , y)間の直線的な関係の強さを調べる手法。**相関係数(r)**を使って，その関係の強さを表現する

⇒相関係数の違いによる散布図の変化

⇒相関係数の算出

・単回帰分析は説明変数(x)が目的変数(y)に及ぼす**影響度の強さ**を調べる手法。それらの関係を**回帰式**で表現する。

⇒回帰式の算出

・回帰式における検定は，2変数間に直線関係があるかどうか？を検定する。

⇒検定の考え方

・検定は分散分析表によるF検定が用いられる。

⇒分散分析表に必要な計算

⇒分散分析表の読み方

13