第 13 回(2009/7/2)

・期末試験の予定を知らせた.

期末試験に関する重要な知らせ：
１）日時：７月２３日(木)、８：４０〜１０：１０
２）持込：電卓、筆記用具
３）資格：欠席した小テストと欠席のレポートの提出が必要
４）成績：小テストと宿題は４０%、期末テスト６０%
５）試験場所：B209
６）試験範囲：本講義の内容のすべて

・宿題の説明があった.

４）ワイブル分布

ある物体の疲労強度はワイブル分布（Weibull distribution）に従う．ワイブル分布の確率密度関数は，(「工学統計解析」，吉本著より)
ワイブル分布の確率分布関数 *F(x)*は

$$F(x) = 1 - \exp\{1 - (\frac{x-\gamma}{\alpha})^m\} \qquad (x > \gamma)$$

この場合，確率密度関す f(x)は F(x)の微分である.

$$f(x) = \frac{m}{\alpha}(\frac{x-\gamma}{\alpha})^{m-1} \exp\{-(\frac{x-\gamma)}{\alpha})^m\} \qquad (x > \gamma)$$

α＝１，γ＝０の場合における f(x)のグラフは以下の図に示す．*m*は形状パラメータであり，*m=1* のときは指数分布になる.

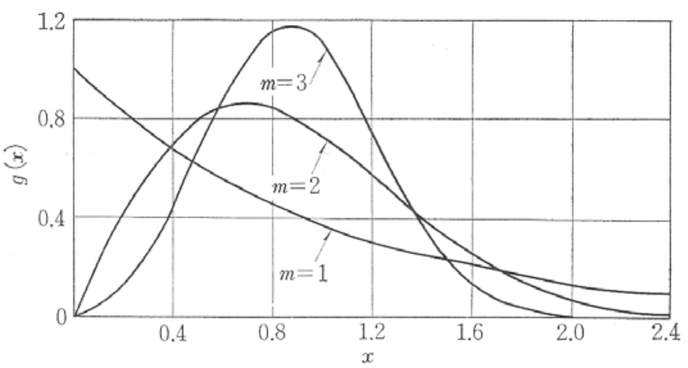


図 5.4 $\alpha=1, \gamma=0$ の場合のワイブル分布

ワイブル分布の平均と分散化は以下の式で計算できる．

$$E(x) = \gamma + \alpha\Gamma(\frac{1}{m} + 1)$$

$$V(x) = \alpha^2\{\Gamma(\frac{2}{m} + 1) - \Gamma^2(\frac{1}{m} + 1)\}$$

$$\Gamma(k) = \int_0^\infty e^{-y} y^{k-1} dy$$

$\Gamma(k)$はガンマ関数である．

今日まで多くの寿命データがワイブル分布に従うことが実証されていて，たとえば，機械が故障するまでの時間の分布，すなわち寿命のモデル分布としてよく知られています．

５）確率紙

確率紙は事象がどの分布に従うかを判断し，その分布の母数（平均や分散など）を容易に推定するために工夫されたグラフ用紙である．確率用紙には正規確率紙，対数正規確率紙，ワイブル確率紙などがある．それぞれの分布に当てはまる場合に確率と変数との関係が直線になるように縦・横軸の目盛りが調整されている．

正規確率紙：

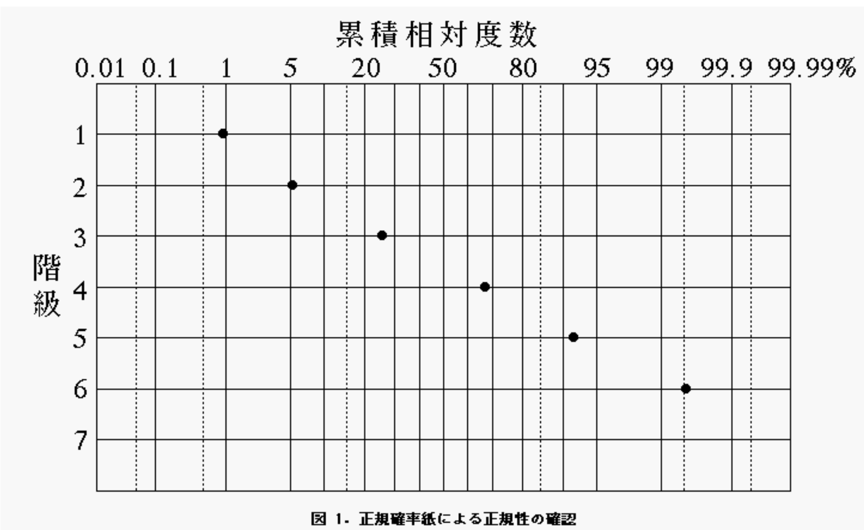
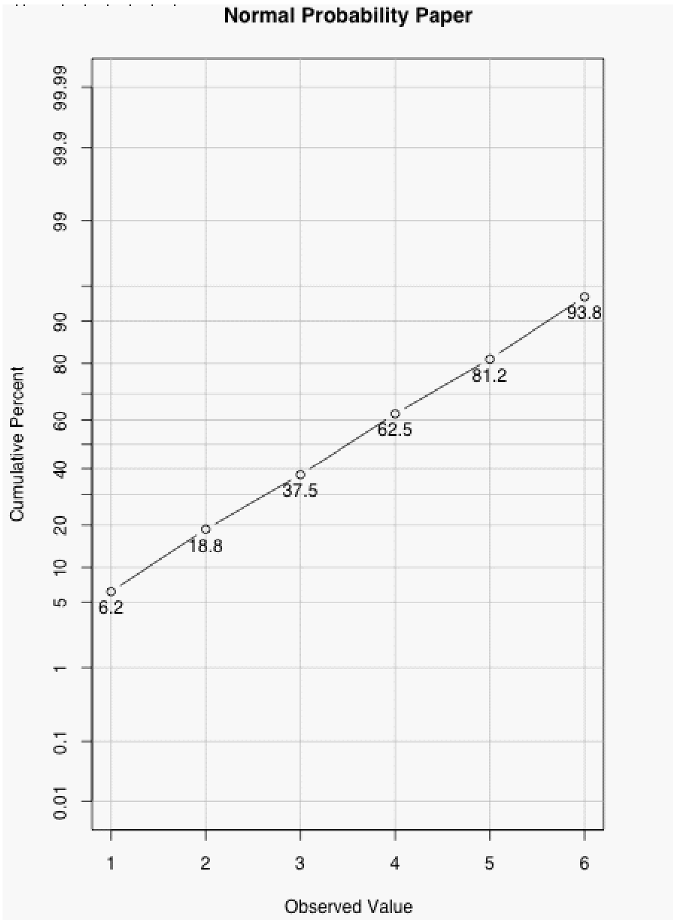
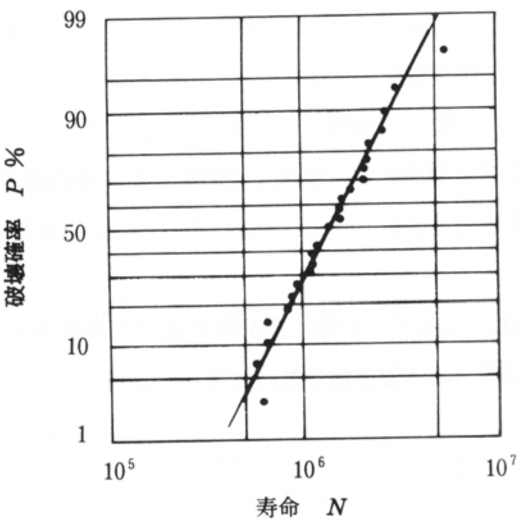


図 1. 正規確率紙による正規性の確認



・確率紙による分布型に対する適合性の検討



(a) 対数正規確率紙にプロットした寿命データ

・小テストIVを行った.

第 14 回(2009/7/9)

・小テスト 4 についての説明があった.

・レポートや宿題，小テストの未提出を該当者に知らせた.

４．回帰分析

４．１線形回帰

一つの変数は他の変数にどのように関係しているかを推測することは研究においてしばしば必要である．このような関数間の関係を推測するために回帰分析という方法が用いられる．１次関数を想定した線形回帰(linear regression)においては，Xの値が決まったからと言って Y の値が決まるのではなく，Yは確率変数である．すなわち，

$$Y = \alpha + \beta x + \varepsilon$$

であり，誤差(error) ε をともなう.

誤差 ε は平均と分散が $E(\varepsilon) = 0, \qquad V(\varepsilon) = \sigma^2$ となる確率変数である.

従って, Y の平均 $E(Y)$については, $E(Y) = \alpha + \beta x$ となり， α， β を回帰係数という.

直線 $y = \alpha + \beta x$ を母回帰直線(population regression line)という.

ε はしばしば正規分布 $N(0, \sigma^2)$に従うと仮定されるが, そのとき Yは $N(\alpha+\beta x, \sigma^2)$に従うことになる.

例題 7.1　ある会社の 10 個所の販売店の従業員数 X (人) と年間売上高 Y (億円) の関係は下の表のようになっている．散布図を描け．

標本番号 i	従業員数 X	年間売上高 Y
1	1	3
2	2	4
3	4	5
4	5	6
5	6	7
6	7	7
7	8	8
8	8	9
9	9	9
10	10	12

[**解**]　xy 座標面上に点 (x, y) をプロットすると，散布図 (図 7・1) が得られる．

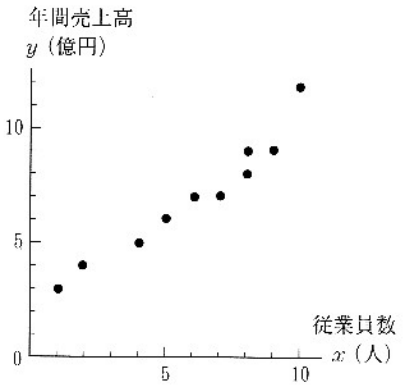


図 7・1　散布図

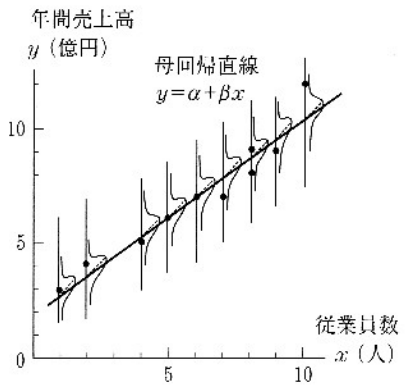


図 7・2　母回帰直線

４．２最小自乗法による線形回帰

観測値(x_i, y_i)を回帰直線 Y=a+bX で近似することを考えることにする．

Y を確率変数，X を変数として，X=x_i のときの Y の推定値は

$$Y=a+bx_i$$

であり，観測値 y_i と推定値 Y との差 e_i を偏差を言う．すなわち，

$$e_i = y_i - (a + bx_i), \qquad (i = 1,2,3,\cdots,n)$$

最小自乗法とは偏差 e_i の二乗和を最小にするように直線または曲線を決める方法である．すなわち

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \{y_i - (a + bx_i)\}^2 \ \Rightarrow\ \min$$

とする a,b を求める．偏差（上の式）を a,b で偏微分して

$$na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \ ,$$

$$a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

という 2 つの方程式を得る．のこ連立 1 次方程式を正規方程式(normal equation)という．正規方程式を解くと

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} = r\frac{s_y}{s_x}$$

$$a = \frac{1}{n}\sum_{i=1}^{n} y_i - b\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{y} - b\bar{x}$$

$$r = \frac{s_{xy}}{s_x s_y}$$

ここで，$\bar{x},\bar{y}$とs_x^2，s_y^2はそれぞれ変数 x,y の標本平均と標本分散であり，s_{xy}と r は変数 x,y の標本共分散と標本相関係数である．($r = \frac{s_{xy}}{s_x s_y}$)

求めた標本回帰直線は観測値（生データ）にどのぐらい当てはまっているか，すなわち，「フィットの良さ」，「フィットの妥当性」を検証するために決定係数(coefficient of determination)r^2を用いる．

$$r^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$\bar{y}$は標本の平均値，$\hat{y}_i$は回帰直線による推定値，y_iは標本の観測値である．上の式を次のように変形できる．

$$r^2 = \frac{\{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})\}^2}{\{\sum_{i=1}^{n}(x_i - \bar{x})^2\}\{\sum_{i=1}^{n}(y_i - \bar{y})^2\}^2} = \frac{S_{xy}^2}{(S_x^2 \cdot S_y^2)}$$

従って, 決定係数は相関係数の自乗であることがわかる.

・最終試験のための説明があった.
・最終試験場は B209.

・ 期末試験の注意事項についての説明があった.
・ 授業評価を実施した.