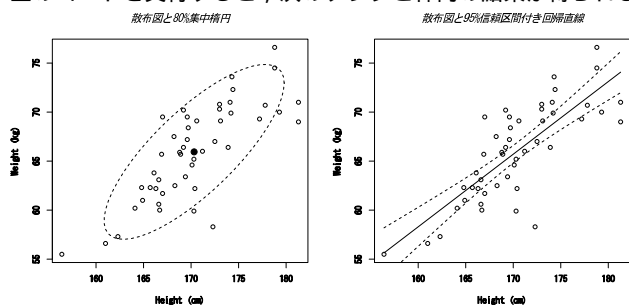


医学情報処理演習第9回「計数データと比率の解析」2007年12月3日 前回の課題の回答例

<http://phi.med.gunma-u.ac.jp/medstat/kadai8-2007.R>

```
dat <- read.delim("http://phi.med.gunma-u.ac.jp/medstat/p02.txt")
dat <- subset(dat,complete.cases(dat))
males <- subset(dat,sex=="M")
attach(males)
layout(t(1:2))
cor.test(wt,ht)
summary(res <- lm(wt~ht))
predict(res,list(ht=165),interval="confidence")
htt <- seq(min(ht),max(ht),length=50)
library(car)
data.ellipse(ht,wt,levels=0.8,col="black",lty=2,lwd=1,xlab="Height (cm)",ylab="Weight (kg)",main="散布図と 80% 集中楕円")
plot(wt~ht,xlab="Height (cm)",ylab="Weight (kg)",main="散布図と 95% 信頼区間付き回帰直線")
matlines(htt,predict(res,data.frame(ht=htt),interval="confidence"),col="black",lty=c(1,2,2))
detach(males)
```

上のコードを実行すると、次のグラフと枠内の結果が得られる（情報量の多い行は削除済み）。



```
Pearson's product-moment correlation
data: wt and ht
t = 8.636, df = 46, p-value = 3.476e-11
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6468709 0.8750522
sample estimates:
      cor
0.7864562
```

(1) 相関については、左側のグラフと上枠内の結果から、ピアソンの積率相関係数が 0.79 (95% 信頼区間 [0.64,0.88]) と、有意な ($p = 3.5 \times 10^{-11}$) 正の相関があることがわかる。

(2) 回帰分析については、右側のグラフと下枠内の結果から、体重の分散の 61% は身長によって説明され ($p = 3.5 \times 10^{-11}$ なので 5% 水準で有意)、回帰式は体重=0.74・身長-60.2 となる。この説明力では予測に十分ではないが、敢えて実行すると、この集団の身長 165cm の男性の体重は 62 kg (期待値の 95% 信頼区間は [60.7,63.3]) と推定される。

```
Call:
lm(formula = wt ~ ht)
Residuals:
    Min     1Q  Median     3Q     Max
-9.1082 -1.8814 -0.1752  2.2808  6.0176
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -60.21562   14.61515   -4.120 0.000156 ***
ht           0.74071    0.08577    8.636 3.48e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3.052 on 46 degrees of freedom
Multiple R-Squared:  0.6185, Adjusted R-squared:  0.6102
F-statistic: 74.58 on 1 and 46 DF, p-value: 3.476e-11

      fit      lwr      upr
[1,] 62.00103 60.72401 63.27804
```