

RjpWiki アーカイブス

【グラフィックス参考実例集：箱型図 boxplot】

1 イントロ

ご存知(とはいえ世間での認知度はほとんど零)箱型図(箱ひげ図)。関数 `boxplot` は箱ひげ図と呼ばれるグラフを作図します。このグラフは、いくつかのデータの分布の違いを比較するのに有効です。`boxplot` は任意個の引数をとって、それぞれに対する箱ひげ図を描きます。上下のひげは、それぞれ上(下)側四分位数の位置から、極値までの間に引かれます。極値とは、上(下)側四分位数から四分位範囲の 1.5 倍以内にあるデータのうちの最大(小)値です。極値よりも大きい、または小さい値は外れ値としてひげの先にプロットされます。なお、`boxplot` に `old=T` と指定することによって、S version 3.2 以前と同じスタイルの箱ひげ図を描くことができます。

2 サンプルプログラム

2.1 刻み目 (notches) 付きの箱型図

```
boxplot1 <- function() {
  #モデル式による箱型図
  oldpar <- par(no.readonly = TRUE) # 現在のグラフィックスパラメータ退避
  on.exit(par(oldpar)) # (関数がエラー中断しても) 終了時にパラメータ復帰
  data(InsectSprays) # データ読み込み
  png("boxplot1.png") # png デバイスを開く
  boxplot(count ~ spray, data = InsectSprays, col = "lightgray")
  boxplot(count ~ spray, data = InsectSprays,
          notch = TRUE, add = TRUE, col = "blue")
  dev.off() # デバイスを閉じる
}
```

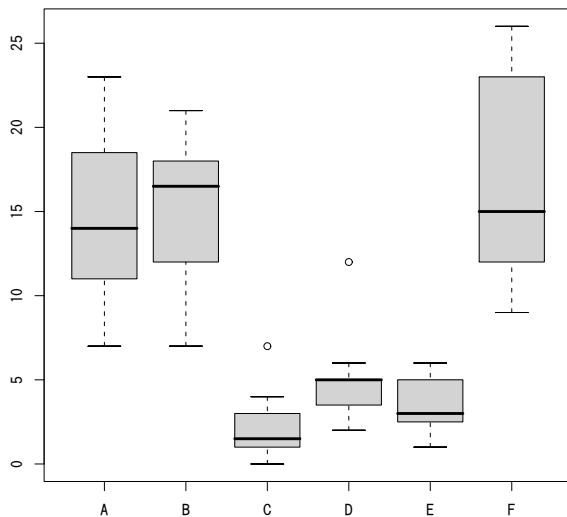


図 1: 刻み目 (notches) のない箱型図 (1)

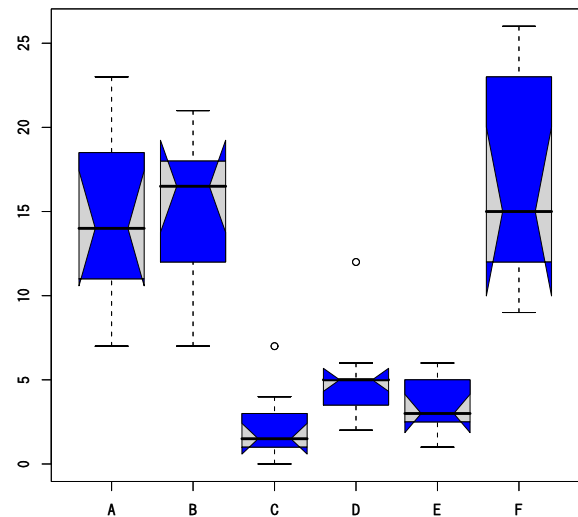


図 2: 刻み目 (notches) 付きの箱型図 (2)

2.2 箱型図に平均+標準偏差範囲を矢印で加える

```

boxplot2 <- function() {
  oldpar <- par(no.readonly = TRUE) # 現在のグラフィックパラメータ退避
  on.exit(par(oldpar)) # (関数がエラー中断しても) 終了時にパラメータ復帰
  data(OrchardSprays) # データ読み込み
  png("boxplot2.png") # png デバイスを開く
  # 箱型図 (縦軸は対数目盛)
  # 箱型図の基本情報取り出しのため変数 rb に付値
  rb <- boxplot(decrease ~ treatment, log="y", data = OrchardSprays,
                col="bisque")
  title("Comparing boxplot()s and non-robust mean +/- SD")
  # 各ブロックの平均値と標準偏差のベクトル計算
  mn.t <- tapply(OrchardSprays$decrease, OrchardSprays$treatment, mean)
  sd.t <- tapply(OrchardSprays$decrease, OrchardSprays$treatment, sd)
  # 箱の中心の x 座標
  xi <- 0.3 + seq(rb$n)
  # 平均値を点プロット
  points(xi, mn.t, col = "orange", pch = 18)
  # 平均値±標準偏差の範囲を矢印で図示
  arrows(xi, mn.t - sd.t, xi, mn.t + sd.t,
        code = 3, col = "pink", angle = 75, length = .1)
  dev.off() # デバイスを閉じる
}
    
```

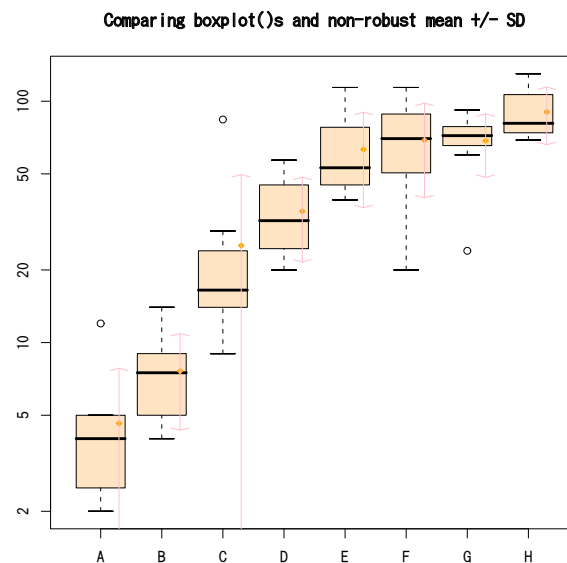


図 3: 平均+標準偏差範囲付きの箱型図

2.3 行列を元に箱型図を描く

```
boxplot3 <- function() {  
  oldpar <- par(no.readonly = TRUE) # 現在のグラフィックスパラメータ退避  
  on.exit(par(oldpar)) # (関数がエラー中断しても) 終了時にパラメータ復帰  
  ## 各種分布によるそれぞれ 100 個の疑似乱数を並べた行列  
  mat <- cbind(Uni05 = (1:100)/21, Norm = rnorm(100),  
               T5 = rt(100, df = 5), Gam2 = rgamma(100, shape = 2))  
  png("boxplot3.png") # png デバイスを開く  
  boxplot(data.frame(mat), main = "boxplot(data.frame(mat), main = ...)")  
  par(las=1) # 軸ラベルをすべて水平に  
  boxplot(data.frame(mat), main = "boxplot(*, horizontal = TRUE)",  
          horizontal = TRUE)  
  dev.off() # デバイスを閉じる  
}
```

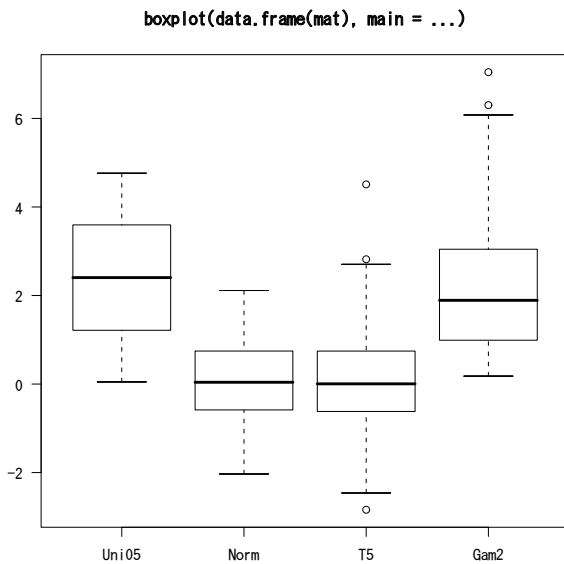


図 4: 行列を元に箱型図を描く (1)

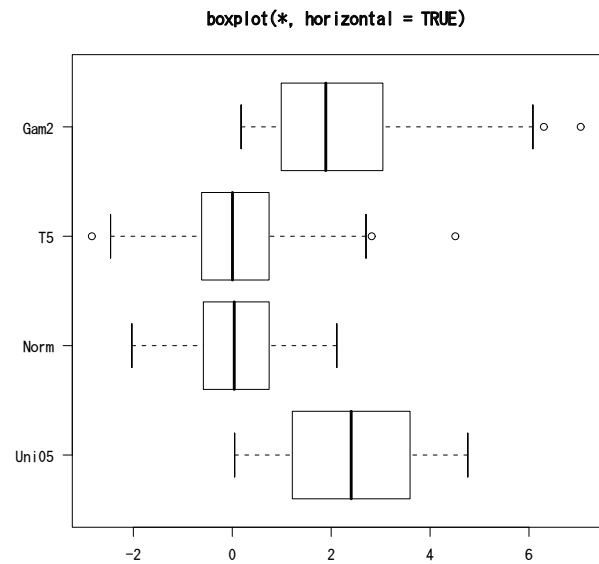


図 5: 行列を元に箱型図を描く (2)

2.4 オプション 'at =' の使用法

```
boxplot4 <- function() {  
  data(ToothGrowth) # データ読み込み  
  ## 箱の中心位置を at で指定  
  ## VC を与えた結果だけをまず描く (オプション subset= supp == "VC")  
  png("boxplot4.png") # png デバイスを開く  
  boxplot(len ~ dose, data = ToothGrowth,  
          boxwex = 0.25, at = 1:3 - 0.2,  
          subset= supp == "VC", col="yellow",  
          main="Guinea Pigs' Tooth Growth",  
          xlab="Vitamin C dose mg",  
          ylab="tooth length", ylim=c(0,35))  
  ## オプション add=TRUE で重ね描き指定  
  ## オレンジジュースを与えた結果を次に描く (オプション subset = supp == "OJ")  
  boxplot(len ~ dose, data = ToothGrowth, add = TRUE,  
          boxwex = 0.25, at = 1:3 + 0.2,  
          subset= supp == "OJ", col="orange")  
  ## 凡例を加える  
  legend(2, 9, c("Ascorbic acid", "Orange juice"),  
        fill = c("yellow", "orange"))  
  dev.off() # デバイスを閉じる  
}
```

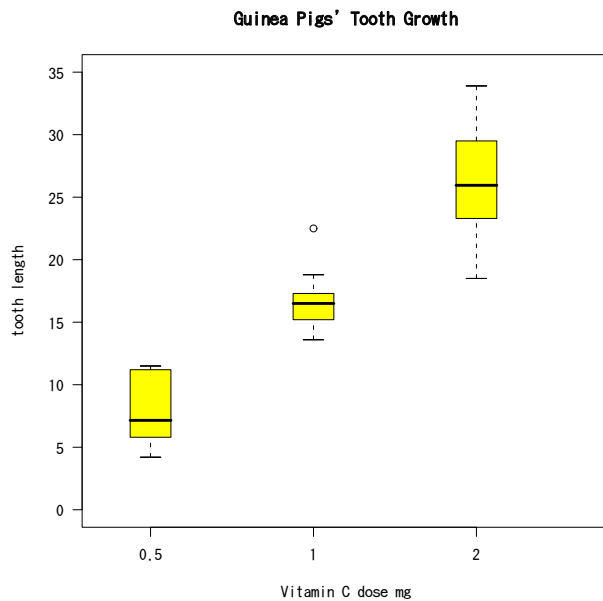


図 6: オプション 'at =' の使用法 (1)

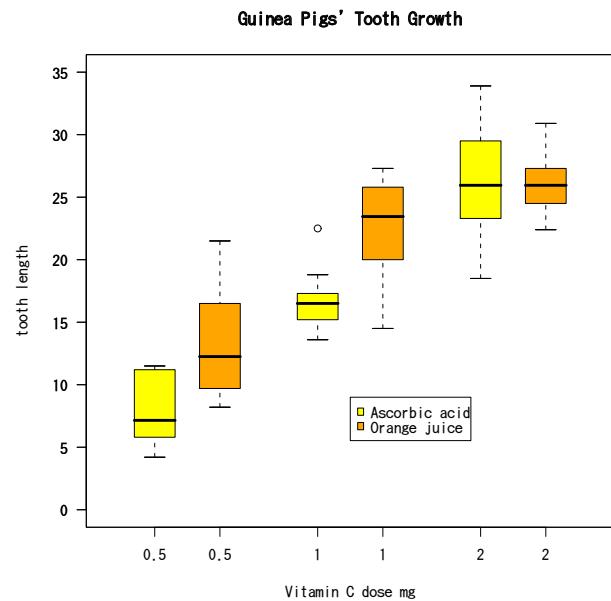


図 7: オプション 'at =' の使用法 (2)

2.5 箱型図を軸の修飾に使う (Statistiques avec R より、2004.3.4)

おまけにラグプロットまで。どうするとこういう使い方を思いつけるのか、ただ脱帽。注意すると最初に layout 関数で作図領域を 4 分割し、そのうちの 3 つに順に本図、y データの箱型図、x データの箱型図を描いている!

```
data(cars)
op <- par()
layout( matrix( c(2,1,0,3), 2, 2, byrow=T ), c(1,6), c(4,1),)
par(mar=c(1,1,5,2))
plot(cars$dist ~ cars$speed, xlab='', ylab='', las = 1)
rug(side=1, jitter(cars$speed, 5) )
rug(side=2, jitter(cars$dist, 20) )
title(main = "cars data")
par(mar=c(1,2,5,1)) # この使い方がすごい boxplot を置く領域を指定している
boxplot(cars$dist, axes=F)
title(ylab='Stopping distance (ft)', line=0) # その領域にタイトルを置く工夫
par(mar=c(5,1,1,2)) # 同じく x 軸の下に boxplot を置く領域を指定している
boxplot(cars$speed, horizontal=T, axes=F)
title(xlab='Speed (mph)', line=1)
par(op)
```

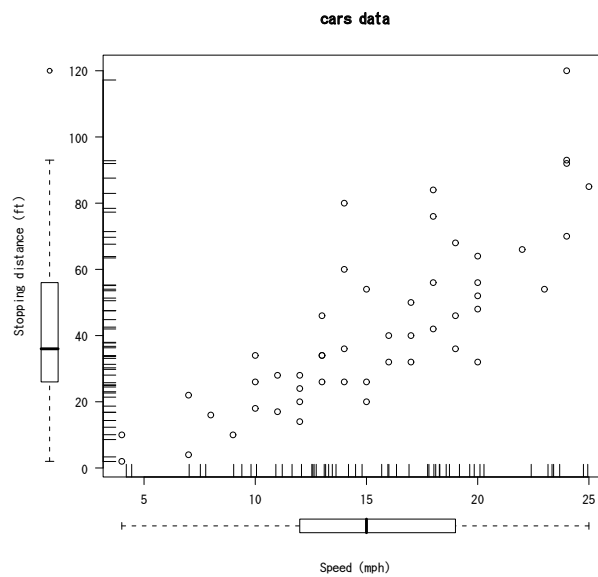


図 8: 箱型図を軸の修飾に使う

2.6 平均と標準偏差を使った箱ひげ図

箱ひげ図には平均値と標準偏差を用いたものもありますが、これを R で作図することは可能でしょうか？

```
bp <- function(x, px, wx=0.25)
{
  m <- mean(x)
  s <- sd(x)
  mx <- max(x)
  mn <- min(x)
  x1 <- px-wx
  y1 <- m-s
  x2 <- px+wx
  y2 <- m+s
  rect(x1, y1, x2, y2)
  lines(c(px-wx, px+wx), c(m, m))
  arrows(px, m+s, px, mx, angle=90, length=2*wx)
  arrows(px, m-s, px, mn, angle=90, length=2*wx)
}
x <- rnorm(100)
plot(c(0,1), c(min(x), max(x)), type="n", xaxt="n", xlab="",
      bty="n", ylim=c(min(x), max(x)))
bp(x, 0.5, 0.4)
```

```
z <- matrix(rnorm(300), nr=100, nc=3)
plot(c(0,3), c(min(z), max(z)), , type="n", xaxt="n", xlab="",
      bty="n", xlim=c(0,4), ylim=c(min(z),
      max(z)), ylab="foo bar baz")
for (i in 1:3) {
  bp(z[,i], i, wx=0.12)
}
```

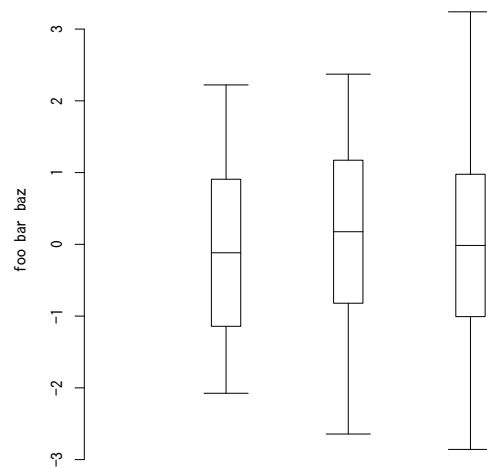
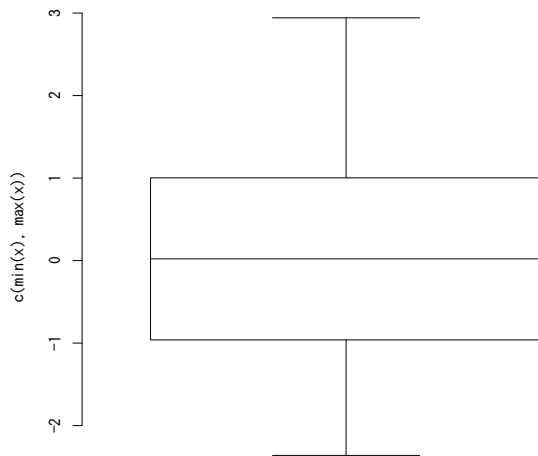


図 9: 平均と標準偏差を使った箱ひげ図 (1)

図 10: 平均と標準偏差を使った箱ひげ図 (2)

2.7 箱ひげ図と散布図の合成

```
a1=rnorm(30)
a2=rnorm(30)
a3=rnorm(30)
x=c(rep(1,30),rep(2,30),rep(3,30))
boxplot(a1,a2,a3)
#par(new=T)
points(x,c(a1,a2,a3))
```

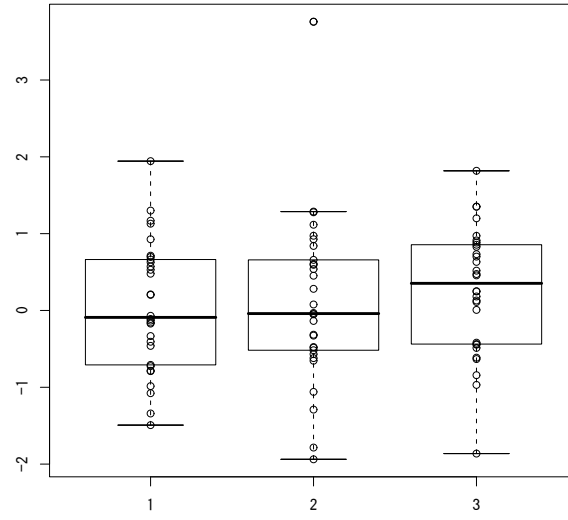


図 11: 箱ひげ図と散布図の合成