

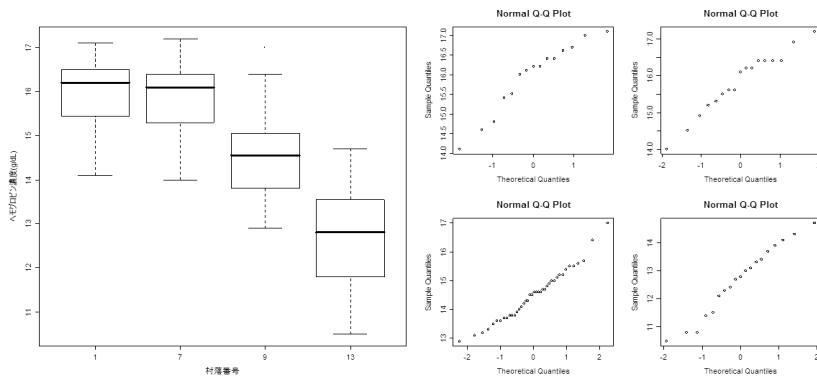
## 医学情報処理演習第7回課題回答例

以下のようにして、まずデータを読み込み(1行目)、attachしてから(2行目)変数 VIL を要因型に変換する(3行目)ことで、分析の準備が整う。

```
dat <- read.delim("http://phi.med.gunma-u.ac.jp/medstat/p07.txt")
attach(dat)
VIL <- as.factor(VIL)
```

次に下枠内のように画面を5分割して、左半分に村落ごとの層別箱ひげ図を描き、右半分の4区分に村落ごとの正規確率プロットを描く。層別箱ひげ図をみると、村落間でヘモグロビン濃度には差がありそうにみえる。正規確率プロットはどれも概ね直線状にみえる。

```
layout(matrix(c(1,1,1,1,2,4,3,5),nr=2))
boxplot(HB ~ VIL, ylab="ヘモグロビン濃度 (g/dL)", xlab="村落番号")
tapply(HB, VIL, qqnorm)
```



そこで以下のように、村落ごとに Shapiro-Wilk の検定をする。

```
tapply(HB, VIL, shapiro.test)
```

下枠内のように、どの村でもヘモグロビン濃度が正規分布に従っているという帰無仮説は棄却されない結果になる(関係部分のみ表示した)

```
$'1' W = 0.9274, p-value = 0.2492
$'7' W = 0.9597, p-value = 0.6255
$'9' W = 0.975, p-value = 0.5104
$'13' W = 0.9637, p-value = 0.6463
```

そこで(以下を各自続けよ),

bartlett.test(HB, VIL) として、Bartlett の検定により、「村落間でヘモグロビン濃度の分散に差がない」という帰無仮説を検定すると、Bartlett's K-squared = 3.7251, df = 3, p-value = 0.2927 が得られるので、有意水準 5% で帰無仮説は棄却されない。summary(aov(HB~VIL)) により一元配置分散分析を行うと、下枠内の通り、村のヘモグロビン濃度への効果は有意である。

```
          Df Sum Sq Mean Sq F value    Pr(>F)
VIL         3  122.820   40.940  43.478 < 2.2e-16 ***
Residuals  87   81.920    0.942
```

どの村とどの村の間で差があるかを検討するため、TukeyHSD(aov(HB~VIL)) として Tukey の HSD 法により多重比較する。

```
      diff      lwr      upr    p adj
7-1 -0.1282353 -1.028646  0.7721758 0.9821751
9-1 -1.4425000 -2.212059 -0.6729415 0.0000248
13-1 -3.2663158 -4.144232 -2.3883998 0.0000000
9-7 -1.3142647 -2.050165 -0.5783645 0.0000616
13-7 -3.1380805 -3.986647 -2.2895139 0.0000000
13-9 -1.8238158 -2.532015 -1.1156170 0.0000000
```

より\*1, 1村と7村の間ではヘモグロビン濃度に有意差がないが、それ以外のすべての組み合わせで有意差があることがわかる。

\*1 演習室の R-2.1.1 では多重性調整済みの有意確率 (p adj) は表示されないが、R-2.3.1 では上の結果が得られる。なお、レポートなどに記載する値はこんなに長く書く必要はなく、有効数字を考えて適切な小数以下数桁を記載すれば十分である。