

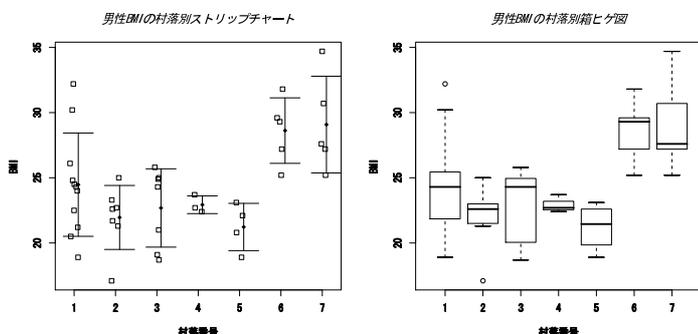
前回の課題の回答例

以下のコードを実行する。

```
http://phi.med.gunma-u.ac.jp/medstat/kadai7-2007.R

dat <- read.delim("http://phi.med.gunma-u.ac.jp/medstat/bmi.txt")
males <- subset(dat,SEX=="M")
males$VG <- as.factor(males$VG)
attach(males)
ix <- 1:7+.1
mx <- tapply(BMI, VG, mean)
sx <- tapply(BMI, VG, sd)
pdf("kadai7g2.pdf", width=12, height=6)
par(family="Japan1GothicBBB")
layout(t(1:2))
stripchart(BMI~VG, vert=T, method="jitter", xlab="村落番号", ylab="BMI",
  main="男性 BMI の村落別ストリップチャート")
points(ix, mx, pch=18)
arrows(ix, mx-sx, ix, mx+sx, code=3, angle=90)
boxplot(BMI~VG, main="男性 BMI の村落別箱ヒゲ図", xlab="村落番号", ylab="BMI")
dev.off()
tapply(BMI, VG, shapiro.test)
print(vareq <- bartlett.test(BMI~VG))
ifelse(vareq$p.value<0.05, oneway.test(BMI~VG), summary(aov(BMI~VG)))
pairwise.t.test(BMI, VG)
TukeyHSD(aov(BMI~VG))
detach(males)
```

結果は次のグラフと枠内の通り（情報量の多い行は削除済み）。



Shapiro-Wilk の検定を村別に行なった結果、すべての村で p 値が 0.05 よりずっと大きく、分布の正規性は棄却されなかった（結果は省略）。次に Bartlett の検定で p 値が 0.2939 と 0.05 よりずっと大きいので（下枠内）、村によって分散に差がないという仮説も棄却されない。

Bartlett's K-squared = 7.3007, df = 6, p-value = 0.2939

上の R のコードでは Bartlett の検定結果から自動判定されて、通常の ANOVA が行なわれ、 p 値が 0.0005578 と得られるので、村落が BMI に与える効果が有意水準 5% で統計学的に有意であることがわかる（下枠内）。

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
VG	6	307.71	51.29	5.3066	0.0005578 ***
Residuals	35	338.26	9.66		

Holm の方法で検定の多重性を調整した多重比較の結果、すべての 2 村落ずつの組み合わせについて、平均値の差の検定の有意確率として下表が得られる。これより、有意水準 5% で判定すると、2-6、3-6、5-6、2-7、3-7、5-7 の組み合わせで有意差があるといえる。

1	2	3	4	5	6
2	1.0000	-	-	-	-
3	1.0000	1.0000	-	-	-
4	1.0000	1.0000	1.0000	-	-
5	0.9046	1.0000	1.0000	1.0000	-
6	0.2218	0.0156	0.0397	0.2218	0.0204
7	0.1413	0.0084	0.0211	0.1458	0.0122
					1.0000

Tukey の HSD で多重比較をした結果を下枠内に示すが、有意水準 5% での判定では、Holm の場合と同じく、2-6、3-6、5-6、2-7、3-7、5-7 の組み合わせで有意差があるといえる。

	diff	lwr	upr	p adj
2-1	-2.5155844	-7.2141006	2.182932	0.6372249
3-1	-1.7870130	-6.4855292	2.911503	0.8936407
4-1	-1.5393939	-7.8690030	4.790215	0.9872586
5-1	-3.2477273	-8.9217246	2.426270	0.5636876
6-1	4.1472727	-1.0941415	9.388687	0.2000918
7-1	4.6072727	-0.6341415	9.848687	0.1161921
3-2	0.7285714	-4.4658337	5.922977	0.9993812
4-2	0.9761905	-5.7297577	7.682139	0.9992350
5-2	-0.7321429	-6.8231228	5.358837	0.9997454
6-2	6.6628571	0.9726714	12.353043	0.0131507
7-2	7.1228571	1.4326714	12.813043	0.0066876
4-3	0.2476190	-6.4583291	6.953567	0.9999998
5-3	-1.4607143	-7.5516942	4.630266	0.9881548
6-3	5.9342857	0.2441000	11.624471	0.0362672
7-3	6.3942857	0.7041000	12.084471	0.0192848
5-4	-1.7083333	-9.1304579	5.713791	0.9904469
6-4	5.6866667	-1.4102418	12.783575	0.1886493
7-4	6.1466667	-0.9502418	13.243575	0.1263544
6-5	7.3950000	0.8760733	13.913927	0.0177038
7-5	7.8550000	1.3360733	14.373927	0.0099235
7-6	0.4600000	-5.6861030	6.606103	0.9999842

本日の課題

<http://phi.med.gunma-u.ac.jp/medstat/p02.txt> は、男女 50 人ずつ、100 人についての身長 (ht) と体重 (wt) の計測値である。性別を表す変数は sex となっている。欠損値を含んでいる。

```
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)
```

上の枠内の通り R で実行すれば、欠損値を含まない男性だけのデータを attach することができ、wt で男性の体重ベクトル、ht で男性の身長ベクトルを得ることができる。このデータについて次のいずれか（または両方）を実行せよ。結果は例によって図を Powerpoint または Word に貼り付け、統計解析の結果とその解釈を文章にして記入し、学籍番号も打って印刷し、署名して提出すること。

1. 身長と体重に相関関係があるかどうか検討せよ。
2. 体重が身長によって説明されるかどうか検討せよ。また、もし十分に説明されるとしたら、この集団に属する身長 165 cm の男性の体重は何 kg と予測されるか、予測値の 95% 信頼区間とともに推定せよ。