

2群の分散の比較 — 素データを使う場合

`var.test`(変数1, 変数2) または `var.test`(変数 ~ 群分け変数, データフレーム名)

対応のないt検定においては、等分散性が仮定できるか否かが問題となる。等分散性の検定を行って分散が等しいという帰無仮説が保持されれば（棄却されなければ）、t検定において `var.equal=TRUE` という指定をすることにより、検出力を高めることができる。

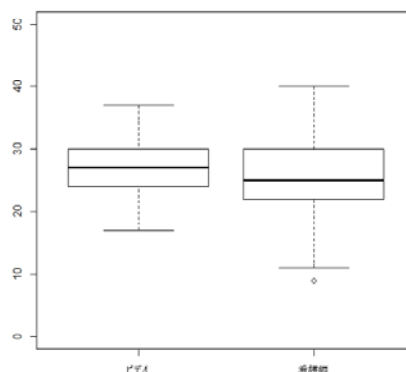
よくわからないときは、等分散性を仮定しない（何も指定しない）。その場合、自由度は小数点以下を切り捨てて整数にしておいたほうが無難。

```
> setwd("i:\\documents\\scripts\\")
> d1 <- read.table("対応のない2群の平均値データ.csv", header=TRUE, sep=",")
> head(d1)
  番号 説明 不安
1    1 ビデオ 26
2    2 看護師 24
3    3 ビデオ 29
4    4 看護師 23
5    5 ビデオ 25
6    6 看護師 23
>
>
> # 記述統計量
> library(psych)
> describeBy(d1[,c("不安")], d1$説明)
group: ビデオ
  vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
1     1 103 27.04 4.32    27   27.12 4.45  17  37    20 -0.1   -0.44 0.43
-----
group: 看護師
  vars   n mean   sd median trimmed  mad min max range skew kurtosis   se
1     1  94 25.44 6.22    25   25.43 5.93   9  40    31 -0.03  -0.15 0.64
>
> # 分散
> tapply(d1[,c("不安")], d1$説明, var)
ビデオ 看護師
18.62593 38.70018
>
> # 等分散性の検定
> var.test(不安 ~ 説明, d1)

      F test to compare two variances

data: 不安 by 説明
F = 0.4813, num df = 102, denom df = 93, p-value = 0.0003419
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3220954 0.7164806
sample estimates:
ratio of variances
 0.4812879
>
>
> #箱ひげ図
> boxplot(不安 ~ 説明, d1, ylim=c(0, 50))
>
```

	A	B	C
1	番号	説明	不安
2	1	ビデオ	26
3	2	看護師	24
4	3	ビデオ	29
5	4	看護師	23
6	5	ビデオ	25
7	6	看護師	23
8	7	ビデオ	37
9	8	看護師	26
10	9	看護師	20
11	10	ビデオ	36
12	11	ビデオ	28
13	12	ビデオ	27
14	13	ビデオ	27
15	14	ビデオ	33
16	15	看護師	38
17	16	ビデオ	30
	17	看護師	23
	18	看護師	28
	19	ビデオ	28
	20	ビデオ	23



2群の分散の比較 — 既知の値を使う場合

自作関数を使う

known.var.test(人数ベクトル, 分散の値ベクトル)

```
#-----
known.var.test <- function(n, u) {
  if (u[1] < u[2])
    {temp <- u[1]; u[1]<- u[2]; u[2] <- temp; df1 <- n[2]-1; df2 <- n[1]-1}
  else { df1 <- n[1]-1; df2 <- n[2]-1}
  v <- u*u
  F <- u[1]^2/u[2]^2
  p.value <- round(2*(1 - pf(F, df1, df2)),4)
  title <- "Equality Test : equality of two variances"
  statistic.1 <- data.frame(n, u, v)
  statistic.2 <- data.frame(F, df1, df2, p.value)

  out <- list(title, statistic.1, statistic.2)
  return(out)
}
# -----
>

> #各群の人数
> n <- c(103, 94)
>
> #各群の標準偏差（不偏分散の平方根）
> u <- c(4.32, 6.22)
>

> known.var.test(n, u)
[[1]]
[1] "Equality Test : equality of two variances"

[[2]]
      n      u      v
1 103 6.22 38.6884
2  94 4.32 18.6624

[[3]]
      F df1 df2 p.value
1 2.073067  93 102  4e-04
```

多群の分散の比較 — 素データを使う場合

`bartlett.test(変数1, 変数2)` または `bartlett.test(変数 ~ 群分け変数, データフレーム名)`

分散分析においては、等分散性が仮定できるか否かが問題となることがある。等分散性の検定を行って分散が等しいという帰無仮説が保持されれば（棄却されなければ）、分散分析において `var.equal=TRUE` という指定をすることにより、検出力を高めることができる。

よくわからないときは、等分散性を仮定しない（何も指定しない）。その場合、自由度は小数点以下を切り捨てて整数にしておいたほうが無難。

```
> setwd("i:¥¥Rdocuments¥¥scripts¥¥")
> d1 <- read.table("1B平均値データ.csv", header=TRUE, sep=",")
> head(d1)
  番号 group utsu
1    1    60   23
2    2    50   24
3    3    50   28
4    4    40   29
5    5    60   36
6    6    50   22
>
```

```
> # 記述統計量
> library(psych)
> describeBy(d1[,c("utsu")], d1$group)
group: 40
  vars  n mean  sd median trimmed mad min max range skew kurtosis  se
1    1 73 26.56 5.71    26   26.59 5.93  14  41   27 0.08   -0.38 0.67
-----
group: 50
  vars  n mean  sd median trimmed mad min max range skew kurtosis  se
1    1 81 29.26 7.09    29   29.25 7.41  13  47   34 0.03   -0.3 0.79
-----
group: 60
  vars  n mean  sd median trimmed mad min max range skew kurtosis  se
1    1 84 27.51 6.69    27   27.43 7.41  11  44   33 0.1   -0.53 0.73
>
> # 分散
> tapply(d1[,c("utsu")], d1$group, var)
      40      50      60
32.61073 50.31944 44.80709
>

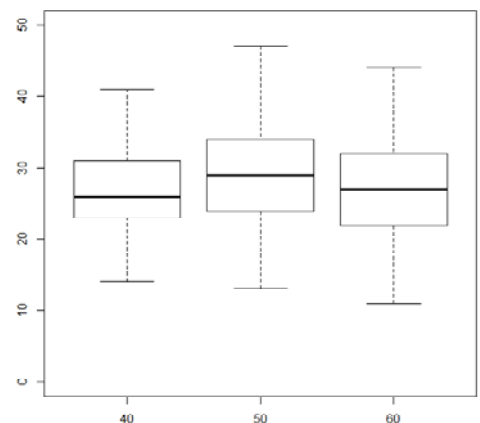
> #等分散性の検定
> bartlett.test(utsu ~ group, d1)
```

Bartlett test of homogeneity of variances

```
data: utsu by group
Bartlett's K-squared = 3.6091, df = 2, p-value = 0.1646
```

```
>
> #箱ひげ図
> boxplot(utsu ~ group, d1, ylim=c(0, 50))
>
>
>
```

	A	B	C
1	番号	group	utsu
2	1	60	23
3	2	50	24
4	3	50	28
5	4	40	29
6	5	60	36
7	6	50	22
8	7	60	36
9	8	50	23
10	9	60	28
11	10	60	27
12	11	60	27
13	12	50	36
14	13	50	33
15	14	40	32
16	15	40	22
17	16	60	44
18	17	50	25
19	18	60	32
20	19	60	20
21	20	40	28



多群の分散の比較 — 既知の値を使う場合

自作関数を使う

`known.bartlett.test`(人数ベクトル, 分散の値ベクトル)

```

#-----
known.bartlett.test <- function(n, u) {
  v <- u*u
  n1 <- n-1
  sum.Q<- sum(n1 * v)
  q <- length(n)
  df <- q-1
  f <- sum(n)-q
  X.2 <- f*log(sum.Q/f) - sum(n1 * log(v))
  B <- X.2 / (1+1/3/df*(sum(1/n1)-1/f))
  p.value <- round(1 - pchisq(B, df), 4)
  title <- "Equality Test : equality of variances"
  statistic.1 <- data.frame(n, u, v)
  statistic.2 <- data.frame(B, df, p.value)
  out <- list(title, statistic.1, statistic.2)
  return(out)
}
#-----

>
> #各群の人数
> n <- c(73, 81, 84)
>
> #各群の標準偏差（不偏分散の平方根）
> u <- c(5.71, 7.09, 6.69)
>

> known.bartlett.test(n, u)
[[1]]
[1] "Equality Test : equality of variances"

[[2]]
      n      u      v
1 73 5.71 32.6041
2 81 7.09 50.2681
3 84 6.69 44.7561

[[3]]
      B df p.value
1 3.594569 2 0.1657

>
>

```