

第9章：t検定

9.1 箱ひげ図の作成

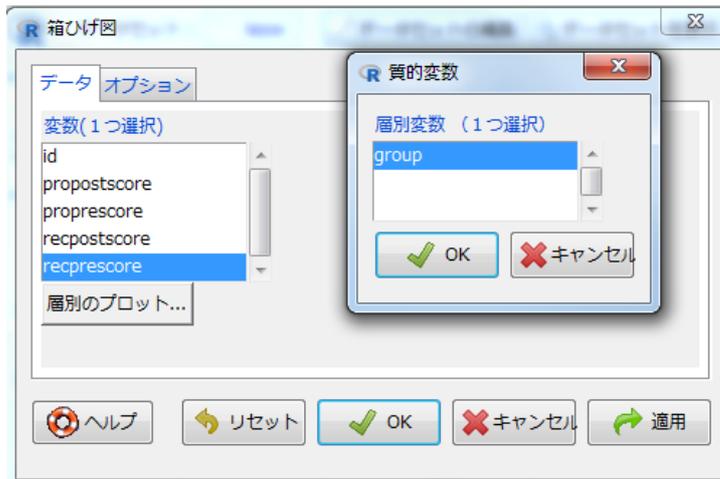
Leow and Morgan-Short 2004 SPP データを読み込む

leow として保存する

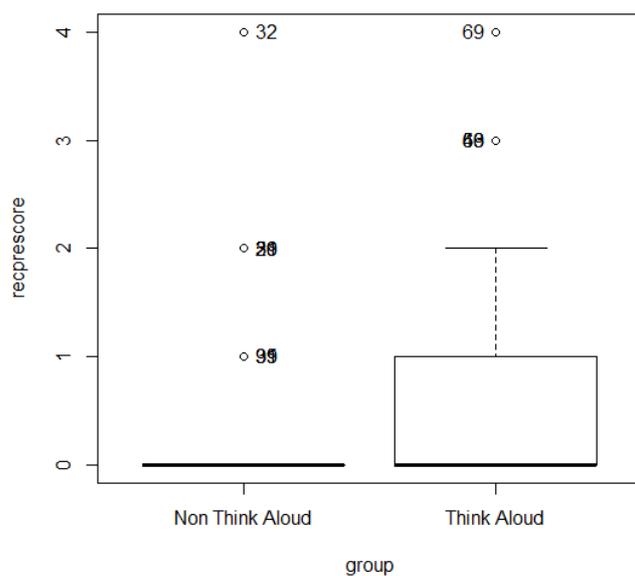
コマンドでグラフ→箱ひげ図

「データ」タブで `recpresscore` を選択

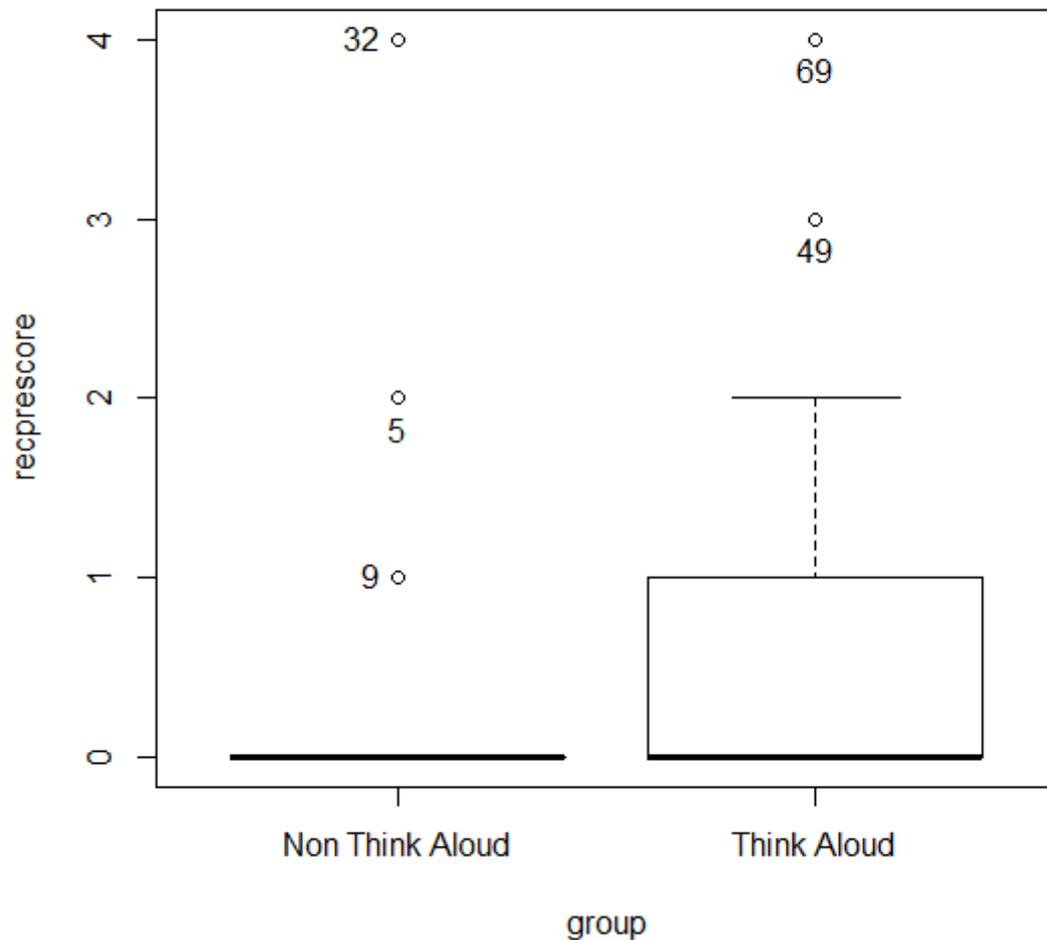
「層別」で `group` を選択



グループ別に1つの従属変数を以下のように表記する：



読みにくいので「オプション」タブで「外れ値の特定」を「マウスで」選択：



箱ひげ図の解釈

- 太線は中央値を示す。不連続点テストであるため、y軸の点に合わせてしか線が現れない
- Non Think Aloud グループ：値がほぼ全て（36中39）0に集中している（被験者がテストの項目を知らなかったことを反映している）。上の3つの点を外れ値とする。
- Think Aloud グループ：38人のほとんどが0点だった
- 箱の部分はデータの75%を占めているので少なくとも28人が1点か0点を取ったことがわかる
- ひげの部分は2点まで伸びる。これが散布の最高値とし、3点と4点の値は外れ値とする

- 両グループの箱ひげ図は対称性に欠けている。つまり、値の過半数が低い値に集中している。これを「正の歪」(positively skewed)と言う
- 箱とデータの中央値の非対称性、ひげの非対称性、外れ値の存在からすれば、正規分布をなしているとは言えない

コンソールでのコマンド：

```
boxplot(recprescore~group, ylab="recprescore", xlab="group", data=leow)
identify(leow$group, leow$recprescore)
```

boxplot(x~y)	y の変数 x の箱ひげ図を作成
recprescore~group,	recprescore が group の変数
ylab="recprescore"	y 軸のラベル
xlab="group"	x 軸のラベル
data=leow	データ名
identify(x)	マウスで外れ値などを選択可
leow\$group, leow\$recprescore	変数名

9.1.2 従属変数の連鎖の箱ひげ図 (対応ある t 検定)

French and O'Brien 2008 SPP data を読み込む

french として保存する

変数を加えるのでコンソールで作業する

注意点：当該のデータセットがコマンダーで選択されていても、コンソールで attach コマンドを入力必要がある

```
>attach(french)
```

```
>boxplot(anwr_1,anwr_2,enwr_1,enwr_2,ylab="Score on test out of 40",
names=c("Arabic Time 1", "Arabic Time 2", "English Time 1", "English Time 2"),
las=1,notch=TRUE,col="grey", boxwex=.5, ylim=range(c(1,40)),medcol="white")
```

boxplot(anwr_1,anwr_2,enwr_1,enwr_2,	箱ひげ図 (変数)
ylab="Score on test out of 40",	y 軸のラベル
names=c("Arabic Time 1", ...),	x 軸の変数にラベルを付ける
las=1,	ラベルを軸と並行に表示する
notch=TRUE,	「刻み目」で 95%信頼区間*を表示する

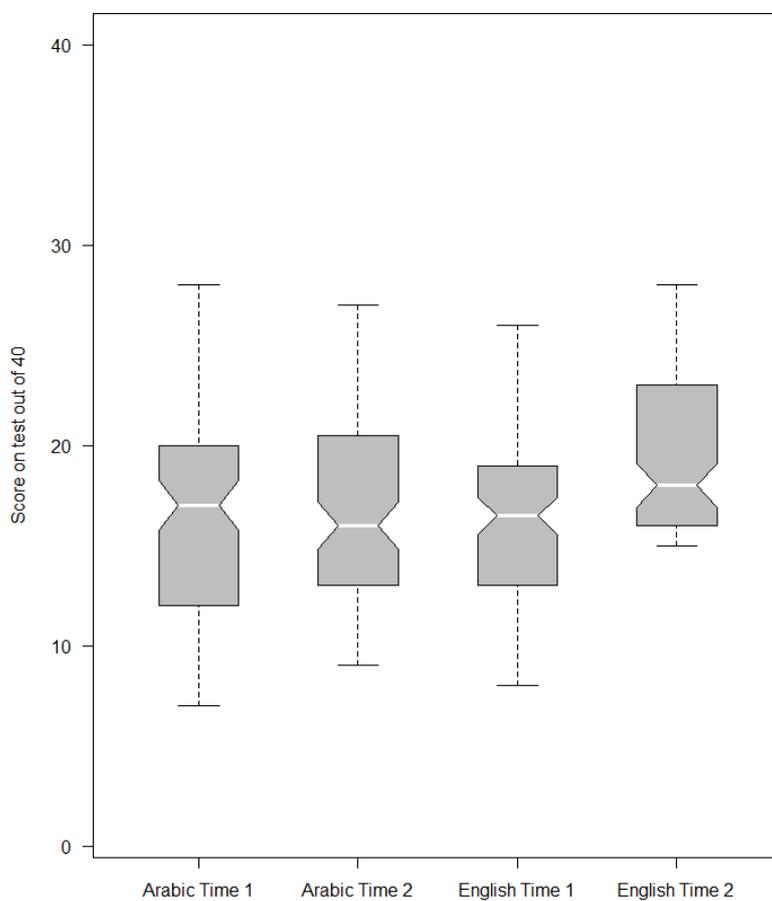
col="grey",	箱ひげ図の色設定
boxwex=.5,	箱ひげ図を細くする
ylim=range(c(1,40)),	y 軸を 1 から 40 に設定する
medcol="white")	中央値を表す線の色設定

* 中央値の 95%信頼区間のことである

「実際の中央値が刻み目で表示される区間にある確率が 95%

2 つの刻み目が重ならない場合、それらの中央値に相違があると言える (.05 有意水準)」

<http://www.cs.utsa.edu/~cs1173/lessons/BoxPlotQuestions/BoxPlotQuestions.html>



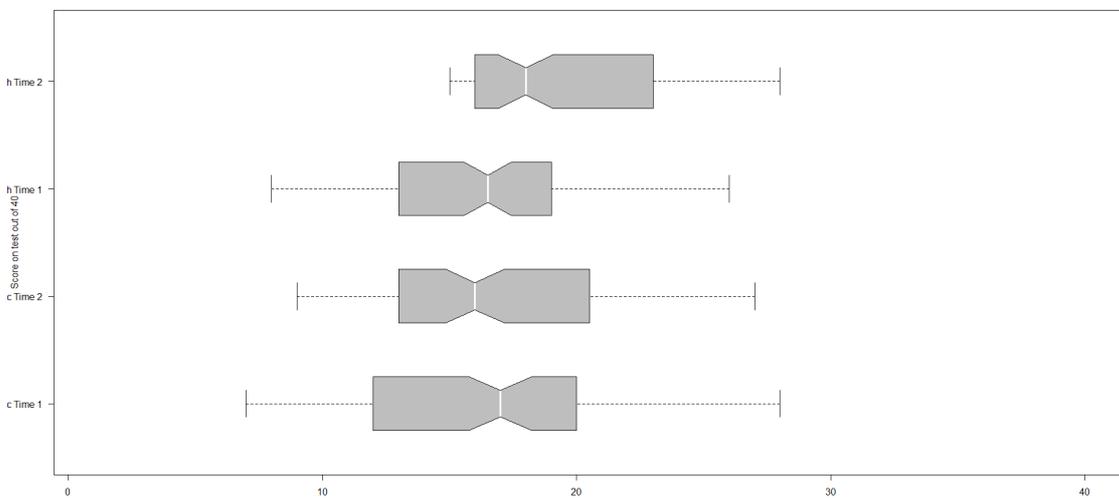
箱ひげ図の解釈

- English Time 2 を除いて音韻論的記憶を計る変数の分布が正規のようである（外れ値なし）
- Arabic 1 と Arabic 2 の中央値が近い一方、後者の分布の範囲が前者のそれより小さい
- English 2 の中央値は確実に English 1 の中央値より高く、分布の範囲も小さい

その他のコマンド

<code>varwidth=TRUE</code>	箱ひげ図の幅をデータ数に合わせる
<code>outline=FALSE</code>	外れ値を未表示にする
<code>names=FALSE</code>	ラベルを未表示にする
<code>horizontal=FALSE</code>	箱ひげ図を横にする
<code>par(mfrow=c(1,2))</code>	画面に図が2つ表示できるようにする c(行数,列数)

以下、`horizontal=TRUE` と `varwidth=true` を追加：

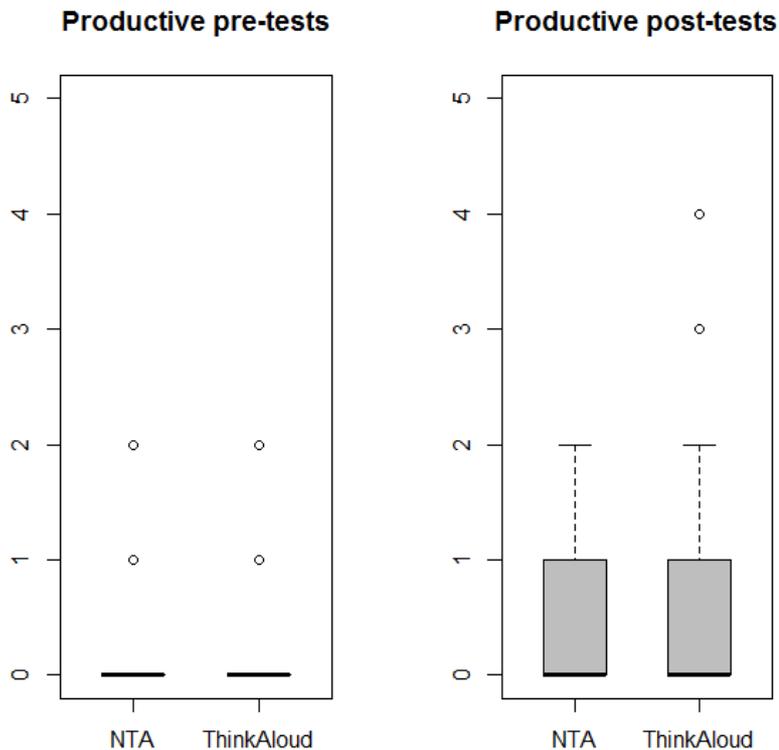


9.1.3 グループに分かれた従属変数の箱ひげ図

従属変数が2つ以上あり、スコアがグループ別に分かれている場合に用いる

`leow` を選択する

`par(mfrow)` コマンドを用いて2つの変数を表示させる



解釈

- Pre-test では、ほとんど全ての被験者は 0 点だった。0 点以外の被験者は外れ値とされている
- Post-test では両グループは同じ分布をなしているが、Think Aloud のグループのほうが外れ値が多い
- Post-test の中央値が 0 点にあることから、被験者の少なくとも 50% が 0 点だったことがわかる

コード：

```
par(mfrow=c(1,2))  
levels(leow$group)=c("NTA", "ThinkAloud") #Make labels shorter so they'll print  
boxplot(proprescore~group,data=leow,ylim=range(c(0,5)),col="gray", main="Productive  
pre-tests",boxwex=.5)  
boxplot(propostscore~group,data=leow,ylim=range(c(0,5)),col="gray", main="Productive  
post-tests",boxwex=.5)
```

注意点：

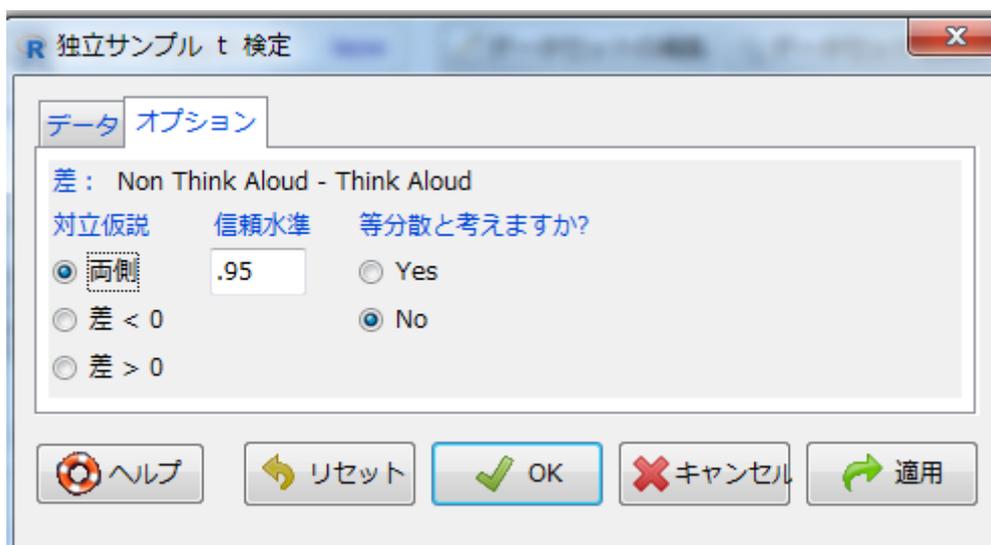
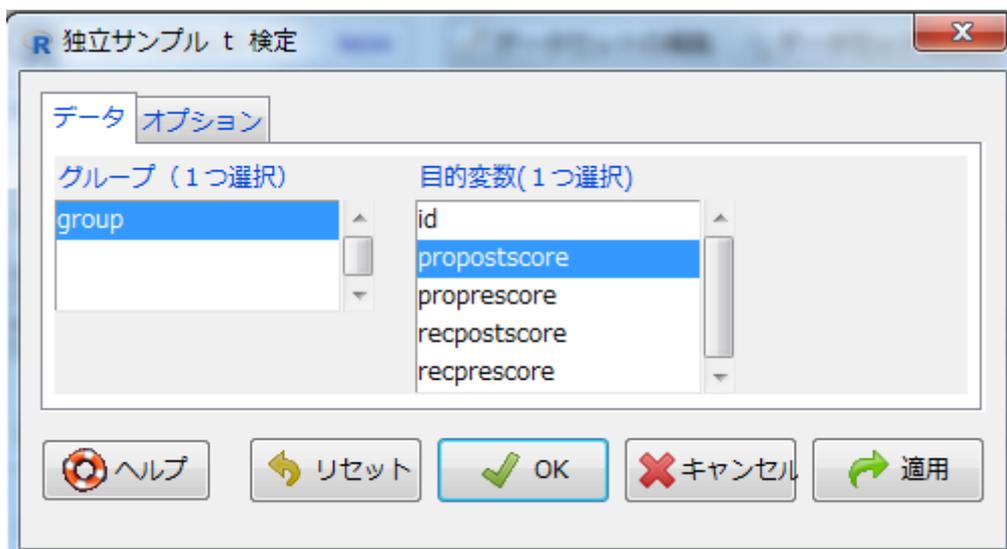
- `par(mfrow=c(1,2))`というのは、箱ひげ図を「1行2列」で並べること
- グループを分けるために、`(proprescore~group)`のように「従属変数~独立変数」を用いる
- `par(mfrow)`コマンドを用いる場合、両方の箱ひげ図が同じ目盛りで表示されるように `ylim` を設定する必要がある

(9.2 箱ひげ図の実践)

9.3 独立サンプル t 検定の実行

(leow)

統計量→平均→独立サンプル t 検定



「等分散と考えますか」を選択すると、伝統的な t 検定を行うことになる。今回は No を選択してより信頼性の高いウェルチ t 検定を行うことにする。

特定のグループが勝る前提はないので「対立仮説」を両側にする。

出力：

```
Welch Two Sample t-test

data:  propostscore by group
t = -0.25134, df = 73.747, p-value = 0.8022
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.680795  1.304277
sample estimates:
mean in group Non Think Aloud      mean in group Think Aloud
                1.153846                1.342105
```

- 2 グループにおける平均値に差があるが、この差が有意かどうかを知るために信頼水準を確認する
- Non Think Aloud と Think Aloud のグループ間の実際の差は 0.19
- 95%信頼水準では約-1.68 から 1.30 という、実際の差よりも広い範囲の平均値が期待される
- この信頼区間が 0 をまたがるため帰無仮説は棄却できない
- t 検定の結果を報告するときに t 値と自由度も表記する（ウェルチ t 検定を実行する場合は、自由度が整数にならないことがある）
- 信頼区間を報告すれば p 値を報告必要はない

コード：

```
t.test(propostscore~group, alternative='two.sided', conf.level=.95,
var.equal=FALSE, data=leow)
```

t.test(x,...)	t 検定を実行する
propostscore~group	従属変数～独立変数(group)
alternative='two.sided'	両側対立仮説 その他'less' (差<0), 'greater' (差>0)

conf.level=.95	信頼水準
var.equal=FALSE	等分散を想定しないウェルチ t 検定を実行する
data=leow	データ選択

注意点:

- 「グループ」の一覧表には、2次元だけを持つ変数のみが表示される（この場合、groupにある Non Think Aloud と Think Aloud）
- 以下のように、LarsonHall.Forgetting.sav のデータにある status をグループとして用いたい場合に、status を一つずつ除外する必要がある

	id	sex	age	status	erguse	usyears	returnage	triptime	accentr	accentl	sentenceaccent	rlwtest
1	DN312	Male	20	Non	6	NA	NA	0.86	8.00	10.60	3.000	67
2	EH307	Female	19	Non	11	NA	NA	0.00	3.60	9.20	3.125	51
3	HM311	Male	19	Non	12	NA	NA	0.00	4.00	8.20	1.000	50
4	KK306	Male	20	Non	8	NA	NA	0.00	2.80	8.40	1.500	47
5	KM303	Female	20	Non	6	NA	NA	2.00	7.40	11.40	3.875	82
6	MM309	Female	20	Non	8	NA	NA	0.00	7.40	11.60	3.375	80
7	MN314	Male	20	Non	11	NA	NA	0.00	4.80	9.40	2.125	86
8	MS308	Female	20	Non	8	NA	NA	0.00	3.40	8.60	1.375	56
9	NI301	Male	20	Non	13	NA	NA	4.00	8.00	10.60	2.250	63
10	NN313	Male	19	Non	11	NA	NA	0.00	5.00	7.80	2.625	70
11	RU305	Female	21	Non	12	NA	NA	3.29	8.80	10.80	4.375	91
12	TH310	Female	19	Non	11	NA	NA	0.00	6.00	12.00	3.375	78
13	TN316	Female	21	Non	10	NA	NA	0.00	9.80	10.60	2.625	67
14	YK302	Male	20	Non	14	NA	NA	3.00	3.60	7.20	1.500	57
15	YT315	Male	18	Non	8	NA	NA	2.00	5.40	8.60	1.500	71
16	CK704	Female	20	Late	20	2.0	NA	0.00	9.00	9.40	3.500	77
17	EA601	Female	22	Late	16	3.0	NA	4.00	8.60	10.80	4.125	89
18	HI500	Female	21	Late	21	3.5	NA	2.00	10.20	11.60	4.875	89
19	UN602	Male	22	Late	13	4.0	NA	0.00	6.40	9.60	2.250	64
20	MC705	Female	23	Late	13	4.5	NA	1.00	7.80	9.20	3.375	84
21	mh502	Female	21	Late	8	3.0	NA	3.00	4.75	5.25	2.875	73
22	MT707	Female	23	Late	19	3.5	NA	4.00	9.20	13.00	4.250	75
23	SH702	Female	20	Late	8	1.5	NA	0.00	6.60	12.60	2.750	58
24	TI501	Male	25	Late	16	2.5	NA	12.00	7.20	10.60	2.125	63
25	TK603	Female	20	Late	13	1.5	NA	4.00	7.00	9.40	2.375	90
26	YK600	Female	21	Late	12	3.0	NA	1.00	8.40	9.20	2.625	65
27	YM701	Female	29	Late	7	4.0	NA	0.00	6.80	10.20	1.625	66
28	YM706	Female	20	Late	16	1.0	NA	54.00	7.00	12.00	4.000	57
29	YN703	Female	22	Late	15	3.0	NA	0.00	8.00	9.20	2.250	36
30	YT604	Female	20	Late	16	1.0	NA	2.00	6.20	10.80	3.500	52
31	AS206	Female	20	Early	13	1.0	6	NA	9.00	12.20	4.375	95
32	CS207	Female	21	Early	6	3.0	6	NA	9.80	12.20	5.250	93
33	EH201	Female	21	Early	10	4.0	4	NA	7.60	10.80	3.875	89
34	EI203	Female	21	Early	6	3.5	7	NA	9.80	10.20	4.875	89
35	EI213	Female	23	Early	10	1.0	5	NA	8.40	12.00	4.000	70
36	HI205	Male	19	Early	11	3.0	6	NA	9.40	11.60	4.750	78
37	KI202	Female	18	Early	5	3.5	7	NA	4.60	10.80	4.750	47
38	MH214	Female	18	Early	10	4.0	6	NA	9.00	11.80	6.750	90
39	MN210	Female	18	Early	4	5.0	7	NA	5.40	10.40	3.625	87
40	MN212	Female	22	Early	3	2.5	7	NA	9.80	12.40	2.875	64
41	RS208	Female	19	Early	10	4.0	7	NA	10.20	12.60	6.250	95
42	SI211	Female	21	Early	15	1.0	6	NA	8.20	11.00	2.875	68
43	YO209	Female	18	Early	4	4.0	7	NA	6.80	10.00	4.750	76
44	YW204	Female	24	Early	9	1.0	4	NA	7.60	10.60	5.250	91

```
>forgetNoNon <- subset(forget, subset=status!="Non")
>forgottenNoNon <- subset(forgotten, subset=status!="Non")
```

「!=」とは≠という意味

	id	sex	age	status	anguse	usyears	returnage	triptime	accentr	accentl	sentenceaccent	rlwtest
16	CK704	Female	20	Late	20	2.0	NA	0	9.00	9.40	3.500	77
17	EA601	Female	22	Late	16	3.0	NA	4	8.60	10.80	4.125	89
18	HI500	Female	21	Late	21	3.5	NA	2	10.20	11.60	4.875	89
19	JN602	Male	22	Late	13	4.0	NA	0	6.40	9.60	2.250	64
20	MC705	Female	23	Late	13	4.5	NA	1	7.80	9.20	3.375	84
21	mh502	Female	21	Late	8	3.0	NA	3	4.75	5.25	2.875	73
22	MT707	Female	23	Late	19	3.5	NA	4	9.20	13.00	4.250	75
23	SH702	Female	20	Late	8	1.5	NA	0	6.60	12.60	2.750	58
24	TI501	Male	25	Late	16	2.5	NA	12	7.20	10.60	2.125	63
25	TK603	Female	20	Late	13	1.5	NA	4	7.00	9.40	2.375	90
26	YK600	Female	21	Late	12	3.0	NA	1	8.40	9.20	2.625	65
27	YM701	Female	29	Late	7	4.0	NA	0	6.80	10.20	1.625	66
28	YM706	Female	20	Late	16	1.0	NA	54	7.00	12.00	4.000	57
29	YN703	Female	22	Late	15	3.0	NA	0	8.00	9.20	2.250	36
30	YT604	Female	20	Late	16	1.0	NA	2	6.20	10.80	3.500	52
31	AS206	Female	20	Early	13	1.0	6	NA	9.00	12.20	4.375	95
32	CS207	Female	21	Early	6	3.0	6	NA	9.80	12.20	5.250	93
33	EH201	Female	21	Early	10	4.0	4	NA	7.60	10.80	3.875	89
34	EI203	Female	21	Early	6	3.5	7	NA	9.80	10.20	4.875	89
35	EI213	Female	23	Early	10	1.0	5	NA	8.40	12.00	4.000	70
36	HI205	Male	19	Early	11	3.0	6	NA	9.40	11.60	4.750	78
37	KI202	Female	18	Early	5	3.5	7	NA	4.60	10.80	4.750	47
38	MH214	Female	18	Early	10	4.0	6	NA	9.00	11.80	6.750	90
39	MN210	Female	18	Early	4	5.0	7	NA	5.40	10.40	3.625	87
40	MN212	Female	22	Early	3	2.5	7	NA	9.80	12.40	2.875	64
41	RS208	Female	19	Early	10	4.0	7	NA	10.20	12.60	6.250	95
42	SI211	Female	21	Early	15	1.0	6	NA	8.20	11.00	2.875	68
43	YO209	Female	18	Early	4	4.0	7	NA	6.80	10.00	4.750	76
44	YW204	Female	24	Early	9	1.0	4	NA	7.60	10.60	5.250	91

(forgetNoNon)

統計量→平均→独立サンプル t 検定
 (変数を選択)

9.4 Robust 独立サンプル t 検定の実行

実際のデータは正規分布に従うことは少ない。

正規分布に従わず、外れ値も含めるデータでも利用できる robust t 検定を実行する

Wilcox(2003)の方法に従う：20% 刈り込み平均値(trimmed mean*)とノンパラメトリック bootstrap との組み合わせが最適している

* 「運動競技などの採点で、極端な点数を付ける審査員の影響を少なくするために、点数を大きさの順に並べて、両側から同数ずつ削除してから平均を求めることがあります。このような平均を、トリム平均 (トリムド平均)、調整平均、刈り込み平均 (trimmed mean) などと呼びます」

<https://oku.edu.mie-u.ac.jp/~okumura/stat/basics.html>

以下のコードを直接にコンソールに入力する

```
>install.packages("WRS",repos="http://R-Forge.R-project.org")
>library(WRS)
```

警告: package 'WRS' is not available (for R version 3.2.2)

警告: package 'WRS' is not available (for R version 2.15.3)

以下のサイトで見つかったインストールの方法でもインストールできない

<http://www.r-bloggers.com/installation-of-wrs-package-wilcox-robust-statistics/>

コード:

```
>Library(WRS)
```

```
>propostscoreNTA <- subset(leow, subset=group=="NTA", select=c(propostscore))
```

```
#n=30
```

```
propostscoreTA <- subset(leow, subset=group=="ThinkAloud", select=c(propostscore))
```

```
#n=37
```

```
trimpb2 (propostscoreNTA, propostscoreTA, tr=.2, alpha=.05, nboot=2000, win=F)
```

このコマンドを用いて、データを分けてベクターが 1 グループのデータのみを含ませる

出力:

```
$p.value  
[1] 0.743  
  
$ci  
[1] -0.59 0.36  
  
$est.dif  
[1] -0.05
```

信頼区間(ci)は広く、0 も通ることから、グループ間には差がないと解釈される

\$est.dif というのは、20%の刈り込み平均

P 値: 「グループ間に差がない」という帰無仮説に対して計算される

trimpb2(x,y,...)	=“Trimmed percentile bootstrap”
x,y	変数 x,y は1つのグループのデータだけを含むベクターでなければならない
tr=.2	20%水準(外れ値が多い場合にはより高い水準を選ぶ)

alpha=.05	有意水準
nboot=2000	ブートストラップを行う回数 Wilcox によれば 500 回が足りる
WIN=F	TRUE にするとデータを Windsorize する (極端な値を除外するのではなくその値を 和らげる方法)
win=.1	Windsorize する場合、その水準を設定する (10%)

Robust 検定の結果を報告するとき、方法、データ数、各グループの平均値とグループ間にある平均の差の信頼水準を記す (教科書 p.176 参照)

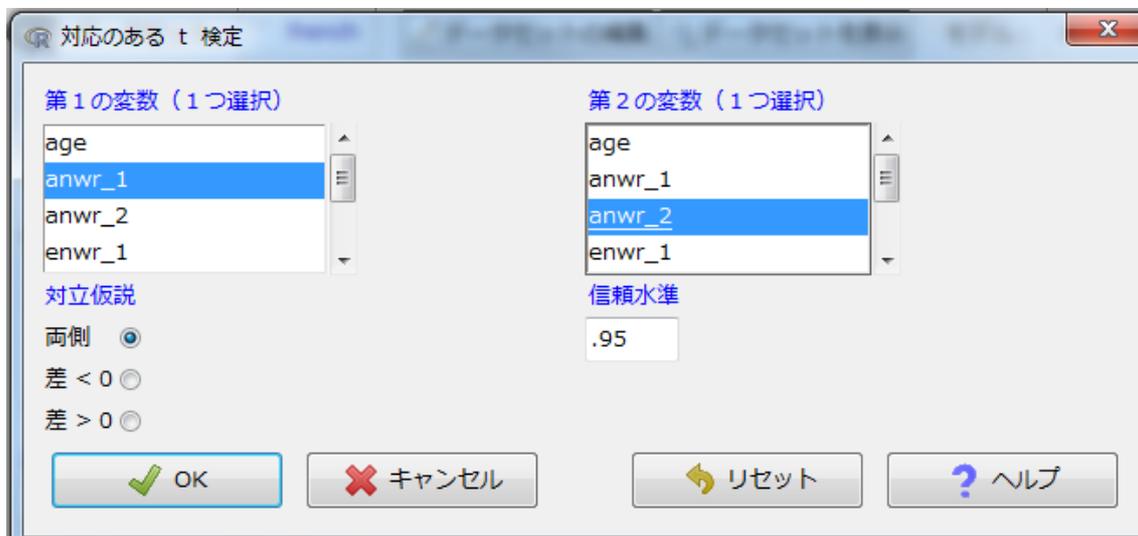
(9.5 独立サンプル t 検定の実践)

9.6 対応のある t 検定の実行

統計量→平均→対応のある t 検定

第 1 の変数 : anwr_1

第 2 の変数 : anwr_2



コード :

```
>t.test(french$anwr_1,french$anwr_2,alternative='two.sided',
>conf.level=.95, paired=TRUE)
```

出力(アラビア語) :

```
Paired t-test

data:  anwr_1 and anwr_2
t = -1.8319, df = 103, p-value = 0.06985
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.52065542  0.02065542
sample estimates:
mean of the differences
                -0.25
```

解釈

- 95%信頼区間 : [-.52, .02] → 平均の差は.52 から (逆方向の) .02 になり得る
- 40点 t 検定にしては広くないが、ゼロに近い
- p 値を参考にすると帰無仮説が棄却される

コード :

```
>t.test(french$enwr_1,french$enwr_2,alternative='two.sided',
>conf.level=.95, paired=TRUE)
```

出力(英語) :

```
Paired t-test

data:  enwr_1 and enwr_2
t = -14.292, df = 103, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-3.536743 -2.674796
sample estimates:
mean of the differences
                -3.105769
```

解釈

- 以上のアラビア語スコアの t 検定よりかなり広い 95%信頼区間
- enwr スコアが Time 1 と Time 2 の間に有意に上達したと言える

注意点：

片側テストを用いると検定力が高まる

その場合、信頼区間の上限が示されない（差の程度を問題としない）

p 値は既に片側検定に合わせた値になる

9.7 Robust 対応のある t 検定

9.4 同様、Wilcox の WRS パッケージを用いる

Wilcox によれば対応のある t 検定の場合は 2 つの仮説を検定できる

- ① グループ間の差が 0 ($H_0 : \mu_D = 0$)
- ② グループの平均値が相当する (equivalent) ($H_0 : \mu_1 = \mu_2$)

① の場合は各ペアの差を、再サンプルしてブートストラップを行う

② の場合は観測値のペアを再サンプルする

① あるいは②の選択は帰無仮説の棄却・採択に関わることがある

R や SPSS では帰無仮説①の検定を行う（以上の出力では alternative hypothesis: true difference in means is not equal to 0 と書いてある）

French and O'Brien (2008)データを読み込む

french として保存する

コード：

```
# 両方の仮説の検定を行う
```

```
rmmcpgb(french$ANWR_1, french$ANWR_2, alpha=.05, est=mean, tr=.2, dif=T,  
nboot=2000, BA=T, hoch=F)
```

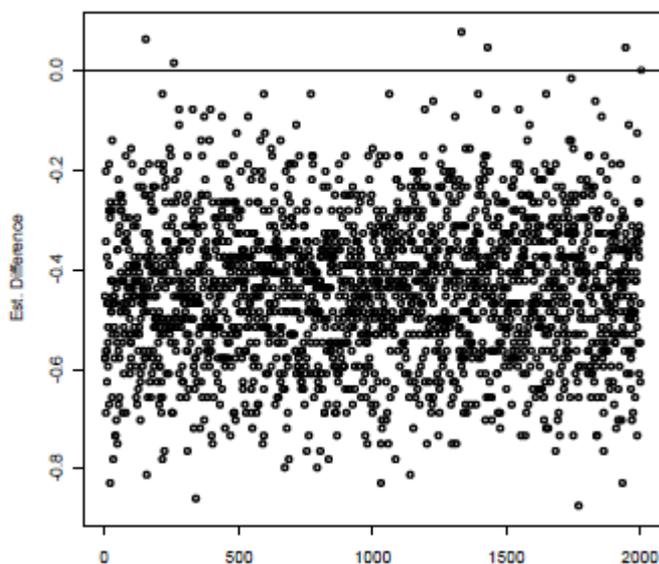
rmmcpgb(x, y=NA, ...)	対応のあるデータを比較するための%ブートストラップを計算する
french\$ANWR_1 french\$ANWR_2	ベクター

alpha=.05	有意水準
est=mean	推定
tr=.2	Mean を用いる場合その刈り込みの程度 (20%)
plotit=T	95%信頼区間を示す図でブートストラップデータを表記する(以下の図)
dif=T	dif=T : ($H_0 : \mu_D = 0$)を検定する dif=F : ($H_0 : \mu_1 = \mu_2$)を検定する
nboot=2000	ブートストラップの回数
BA=F	dif=F の場合、その調整を行う
hoch=F	dif=F の場合、TRUE にするといい

その他のコード：

seed=F	TRUE にすると seed を獲得 (ブートストラップを再出力できる)
seed(2)	seed(番号)の形式で seed を用いる

plotit=T の出力：



Robust 検定の結果を報告するとき、方法、データ数、各グループの平均値とグループ間にある平均の差の信頼水準を記す

(9.8 対応のある t 検定の実践)

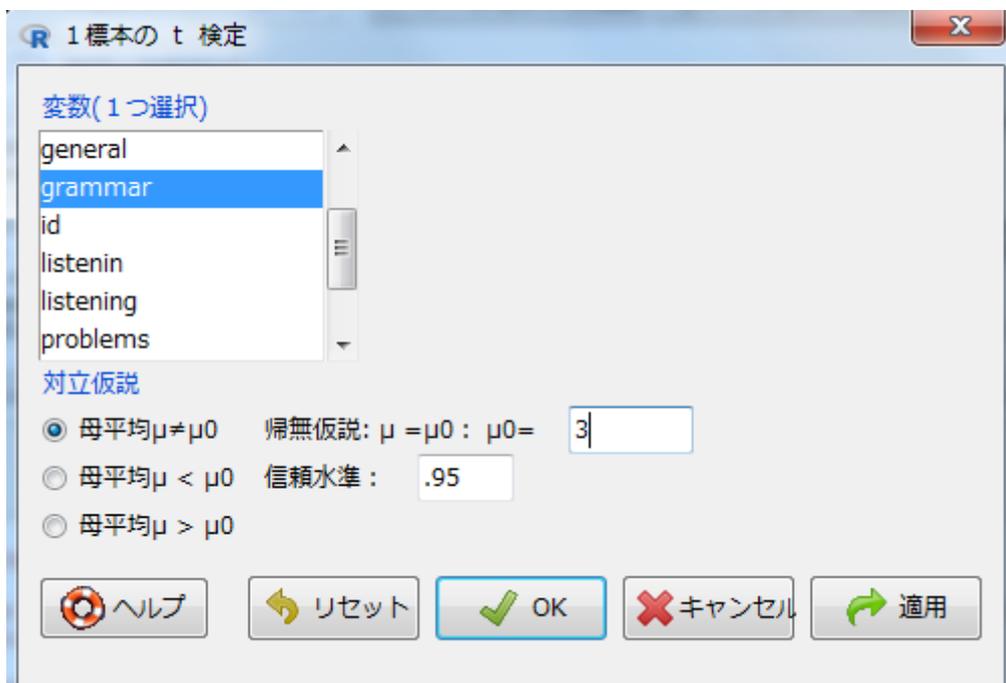
9.9 1 標本 t 検定を実行する

ESL 学習者が母語話者の教師を好んだか非母語話者の教師を好んだか調べる

Torres のデータを読み込む

Torres として保存する

統計量→平均→1 標本 t 検定



「 μ_0 」を「3」にする（このデータでは、「3」が“no preference”の名義変数になっている）

出力:

```
One Sample t-test

data: grammar
t = 1.7724, df = 101, p-value = 0.07934
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 2.976625 3.415532
sample estimates:
mean of x
 3.196078
```

解釈

- Grammar (x)の名義変数の平均は約 3.20→母語話者の教師を好む若干な傾向がみられる
- Df 足す 1=被験者の人数
- p 値が.05 以上→帰無仮説を棄却できない
- 信頼区間：0 ではなくて選択された 3 の値を軸にしている
- 平均値の差が 3 を通るので帰無仮説は棄却できない

コード：

```
t.test(torres$grammar, alternative='two.sided', mu=3, conf.level=.95)
```

t.test(x,...)	全種類の t 検定のコード
torres\$grammar	Grammar という変数を選択
alternative='two.sided'	両側検定（その他 “less”, “greater”（片側検定）
mu=3	サンプルサイズを 1 にする Compares your measured mean score to an externally measured mean ???
conf.level=.95	平均の差の信頼水準

効果量(effect size)の計算

出力の x の平均値引く $\mu \rightarrow 3.19-3=0.19$

割る集団の標準偏差 $\rightarrow 0.19/1.12=.17$

Grammar の効果量は.17 というとても小さい値

9.10 Robust1 標本 t 検定の実行

```
>trimpb(torres$grammar, tr=.2, alpha=.05, nboot=2000,WIN=F,null.value=3)
```

出力：

```
[1] "The p-value returned by the this function is based on the"  
[1] "null value specified by the argument null.value, which defaults to 0"  
[1] "Taking bootstrap samples. Please wait."  
$ci  
[1] 2.967742 3.483871  
  
$p.value  
[1] 0.102
```

標本平均：「2.97, 3.48」（95%信頼区間）

p 値：実際の平均値が 3 であるという帰無仮説を検定する

ブートストラップが行われる前の刈り込み平均を計算する：

```
> mean(torres$PRON,tr=.2)  
[1] 4.596774
```

注意点：以上の平均値は普段出力される平均値とはまた違う

```
> mean(torres$pron)  
[1] 4.313725
```

(9.11 標本 t 検定の実践)