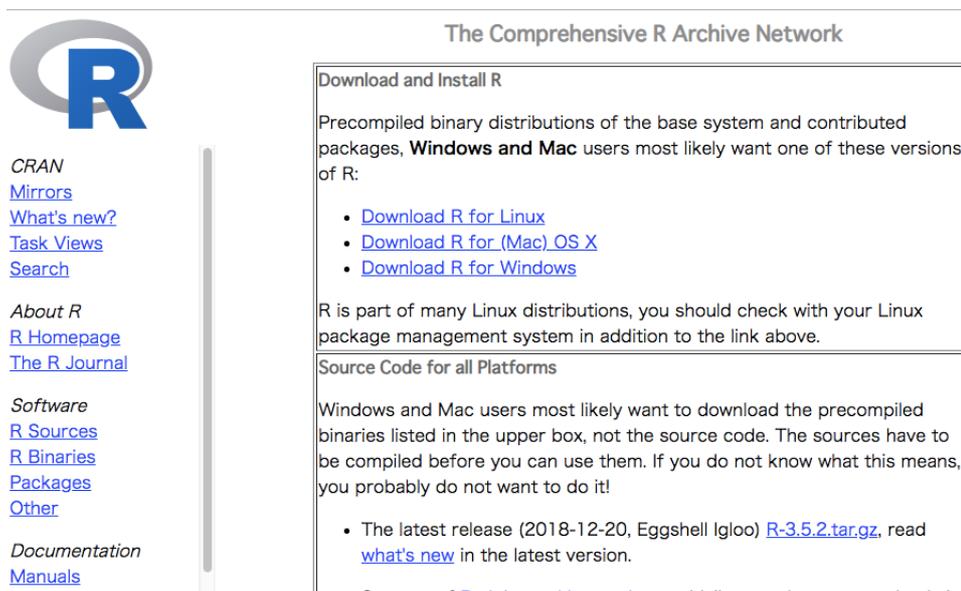


教員支援ツールとしての統計

・統計解析フリーソフト「R」 <https://cran.r-project.org/>

無料で持続可能な統計環境を提供するためのフリーソフト。作図もきれい。簡単なプログラミングを必要とするが、特別なことをしないのであれば、パッケージ化されたいくつかのプログラムをカット&ペーストすることで使用可能。教員にとっては、例えば、試験結果をまとめて学生にわかりやすい形で返したり、複数の試験の相関を調べたり、教育効果について測定したデータの分析などに役立つ。Rのインストール後、いくつかのパッケージもインストールすると便利。ただし、本格的な研究に使う場合は、ちゃんと統計の勉強もしましょう。



パッケージリストのサイト <https://cran.r-project.org/>

最初にインストールするパッケージ

```
install.packages("psych")
```

```
install.packages("beeswarm")
```

それぞれ統計解析用あるいは蜂群図作成用パッケージ

`install.packages()`関数を使ってインストール

インストールしたら読み込ませておく

```
library(psych)
```

```
library(beeswarm)
```

`library()`関数でパッケージに入っている全関数を使えるようにする

ベクトルを作る

複数データを `c()`内にコンマで並べ、`<-` を使って1つのベクトル変数に格納

```
weight<-c(54,91,38,67,61,72)
```

データフレームの作り方

1) `data.frame` を使いデータフレームを作る

```
weight<-c(54, 91, 38, 67, 61, 72)
```

```
height<-c(164, 183, 148, 173, 165, 175)
```

```
age<-c(22, 18, 14, 23, 59, 42)
```

```
data01<-data.frame(weight, height, age)
```

2) できたデータフレームを確認する

```
data01
  weight height age
1     54    164  22
2     91    183  18
3     38    148  14
4     67    173  23
5     61    165  59
```

データフレーム内のデータの記述統計量を求める

1) あらかじめ **psych** を読み込み、その中の **describe** を使い記述統計量を計算する

```
library(psych)
describe(data01)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
weight	1	6	63.83	17.81	64	63.83	13.34	38	91	53	0.07	-1.34	7.27
height	2	6	168	12.03	169	168	8.15	148	183	35	-0.41	-1.3	4.91
age	3	6	29.67	17.31	22.5	29.67	9.64	14	59	45	0.67	-1.42	7.06

mad: median absolute deviation (中央値絶対偏差)

se: standard error (標準誤差)

skew: skewness (歪度) 分布の偏り具合。正規分布 (それが 0) より左に偏っていたらプラス、その反対はマイナス。

kurtosis (尖度) 分布のとがり具合。正規分布 (それが 0) よりとがっていた場合はプラス、その反対はマイナス

2) R で出力されたデータを選択し、コピー&ペーストで Excel の表に張る。右下に出るアイコンをクリックし、「テキストフィルウィザードを使用する」を選び、「次へ」ボタンを押して最後に「完了」を押す。

エクセルのクリップボードから表データをデータフレームとして読み込み

Excel 等でファイルをクリップボードに読み込み、以下のコマンドを実行する。

1) Windows の場合

```
data02<-read.table("clipboard",header=T)
```

2) Mac の場合

```
data02<-read.table(pipe("pbpaste"),header=T)
```

図を描く (ここでは右の表 1 をデータフレームとして data02 に読み込んだものを使っている)

1) 箱ひげ図の描き方

```
boxplot(data02) 4つのテスト結果をすべて1つの図に並べて描く(図1)
```

```
boxplot(data02[,2]) テスト2のみ描き出す
```

```
boxplot(data02[,-3]) テスト3以外は描かない
```

```
boxplot(data02[,2:4]) テスト2~4の範囲のみ描き出す
```

2) ヒストグラムの描き方

```
hist(data02[,1]) テスト1のヒストグラムを描く(図2)
```

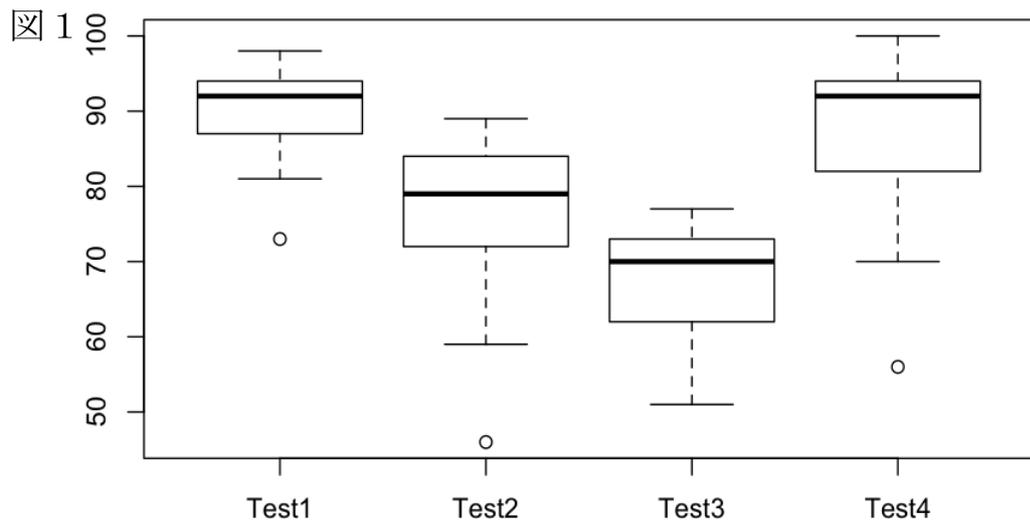
3) 蜂群図の描き方

```
library(beeswarm)
```

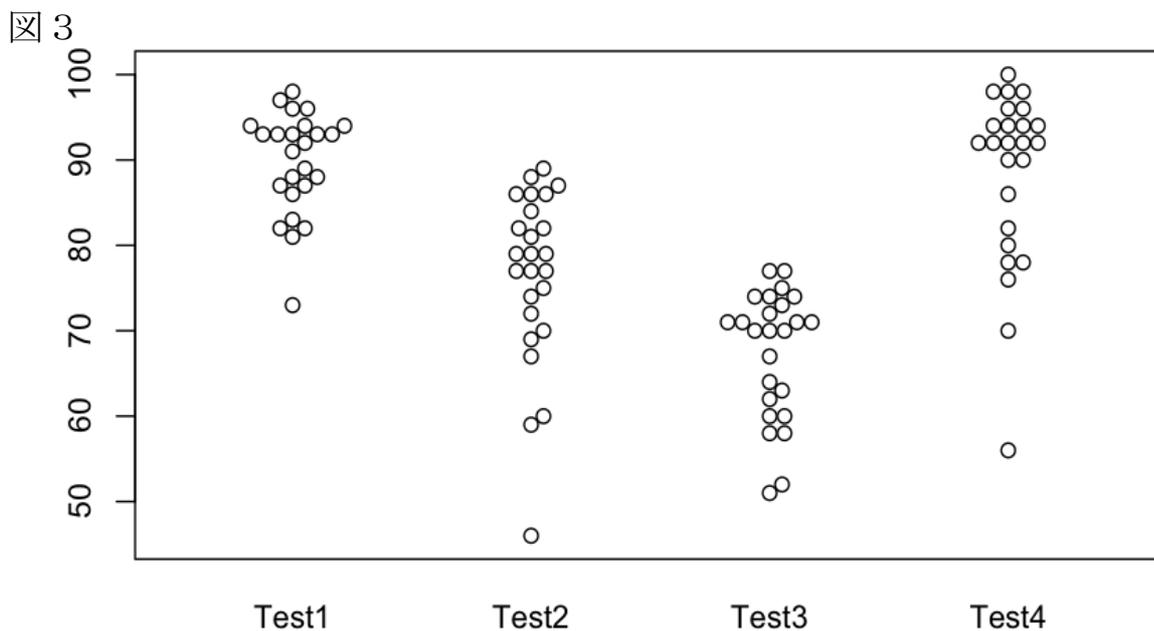
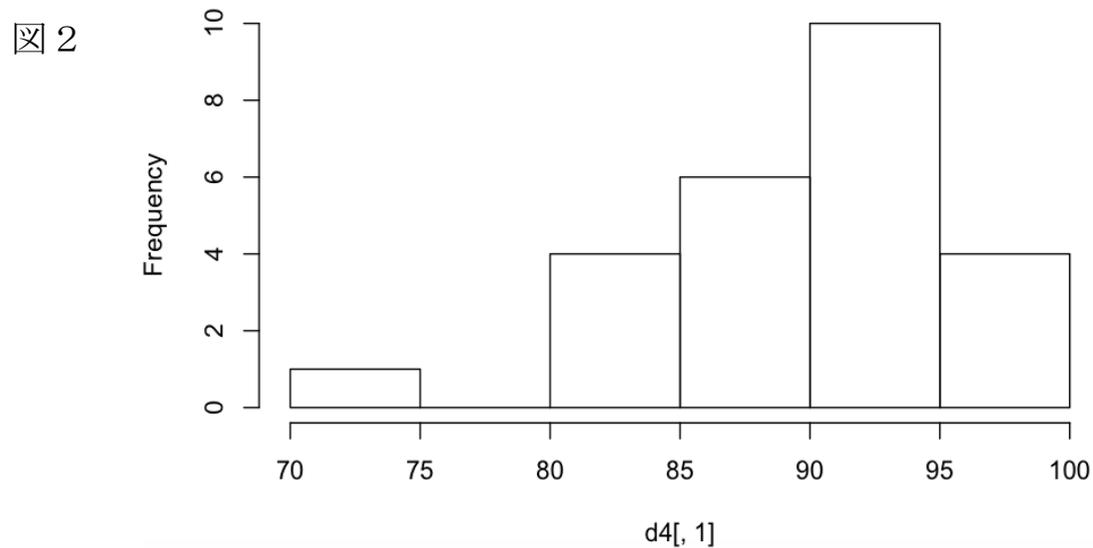
```
beeswarm(data02) 4つのテスト結果を1つの図に並べて描く(図3)
```

表 1

Test1	Test2	Test3	Test4
97	89	77	100
94	86	77	94
87	84	75	90
93	86	74	94
86	75	74	96
96	86	74	78
92	81	73	98
81	77	72	94
96	87	71	98
94	82	71	92
93	74	71	92
93	79	71	96
89	82	70	92
93	88	70	98
98	79	70	92
94	77	67	80
82	60	64	56
91	70	63	86
88	77	62	92
87	67	60	78
93	72	60	82
88	79	58	76
83	46	58	90
73	69	52	70
82	59	51	94



Histogram of d4[, 1]



4) 折れ線グラフの描き方 (図4) (なお、ここでは下の表2をデータフレームとして data03 に読み込んだものを使っている)

```
data03<-read.table(pipe("pbpaste"),header=T)
st1 <- data03 $st1
st2 <- data03 $st2
st3 <- data03 $st3
st4 <- data03 $st4
st5 <- data03 $st5
st6 <- data03 $st6
st7 <- data03 $st7
st8 <- data03 $st8
st9 <- data03 $st9
st10 <- data03 $st10
```

表 2

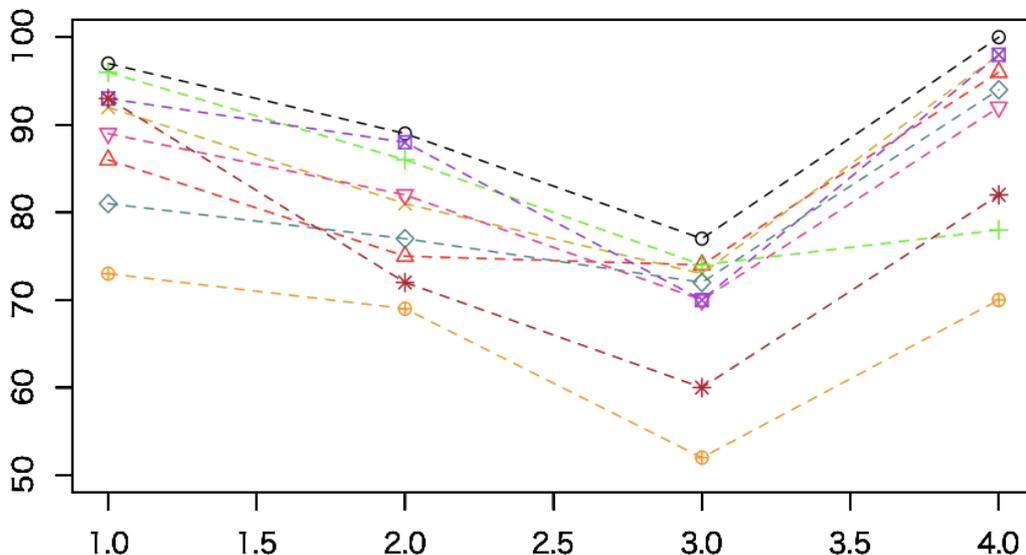
st1	st2	st3	st4	st5	st6	st7	st8	st9	st10
97	86	96	92	81	89	93	93	88	73
89	75	86	81	77	82	88	72	79	69
77	74	74	73	72	70	70	60	58	52
100	96	78	98	94	92	98	82	76	70

```
par(family="HiraKakuPro-W3") # Mac で日本語表記する

plot(st1,type="o", xlim=c(1, 4), ylim=c(50, 100),col=rgb(0,0,0) , pch=1, lty=2)
par(new=T)
plot(st2,type="o", ann=F, xlim=c(1,4), ylim=c(50, 100), col="red", pch=2, lty=2)
par(new=T)
plot(st3,type="o", ann=F, xlim=c(1, 4), ylim=c(50, 100), col="green", pch=3, lty=2)
par(new=T)
plot(st4,type="o", ann=F, xlim=c(1, 4), ylim=c(50, 100), col="gold3", pch=4, lty=2)
par(new=T)
plot(st5,type="o", xlim=c(1, 4), ylim=c(50, 100),col="cadetblue4", pch=5, lty=2)
par(new=T)
plot(st6,type="o", xlim=c(1, 4), ylim=c(50, 100),col="deeppink2", pch=6, lty=2)
par(new=T)
plot(st7,type="o", xlim=c(1, 4), ylim=c(50, 100),col="blueviolet", pch=7, lty=2)
par(new=T)
plot(st8,type="o", xlim=c(1, 4), ylim=c(50, 100),col="brown", pch=8, lty=2)
par(new=T)
plot(st9,type="o", xlim=c(1, 4), ylim=c(50, 100),col="seafreen4" , pch=9, lty=2)
par(new=T)
plot(st10,type="o", xlim=c(1, 4), ylim=c(50, 100),col="orange2", pch=10, lty=2)
par(new=T)
title("4 つのテストの相関")
```

図 4

4つのテストの相関



5) 相関の求め方

cor を使って2つのテスト (テスト1 とテスト2) の相関係数を出す

```
cor(Test1, Test2)
0.6372167
```

すべてのテストの相関係数の行列を出す (round(__,2)の部分は2桁まで表示させるためのもの)

```
round(cor(data02),2)
  Test1 Test2 Test3 Test4
Test1  1.00  0.64  0.63  0.44
Test2  0.64  1.00  0.74  0.43
Test3  0.63  0.74  1.00  0.50
Test4  0.44  0.43  0.50  1.00
```

plot を使って2つのテスト (テスト1 とテスト2) の散布図を描く (図5)

```
plot(Test1, Test2)
```

pairs を使ってすべてのテストの散布図行列を描く (図6)

```
pairs(data02)
```

図5

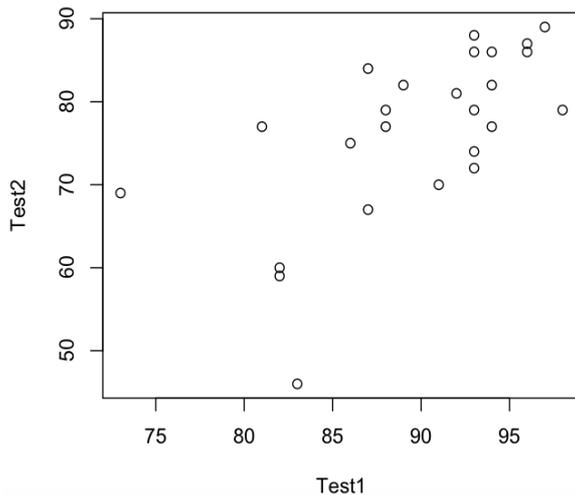
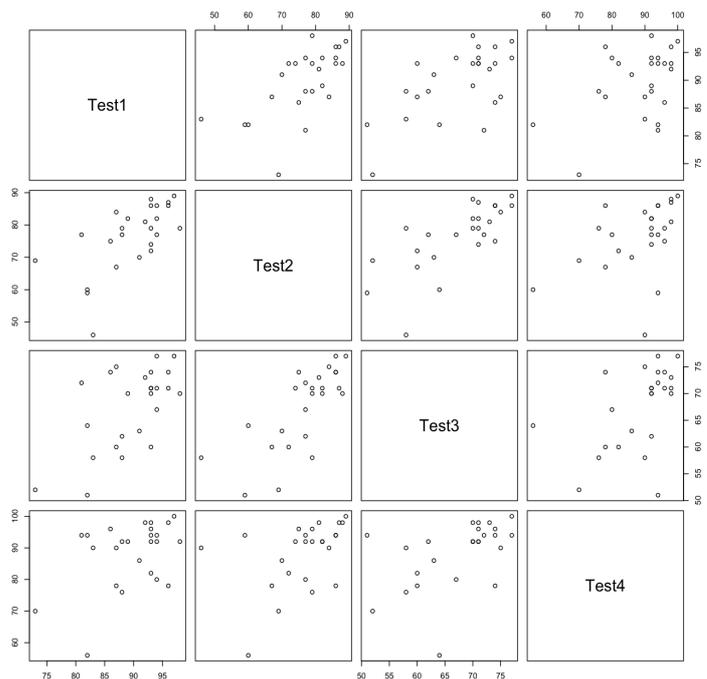
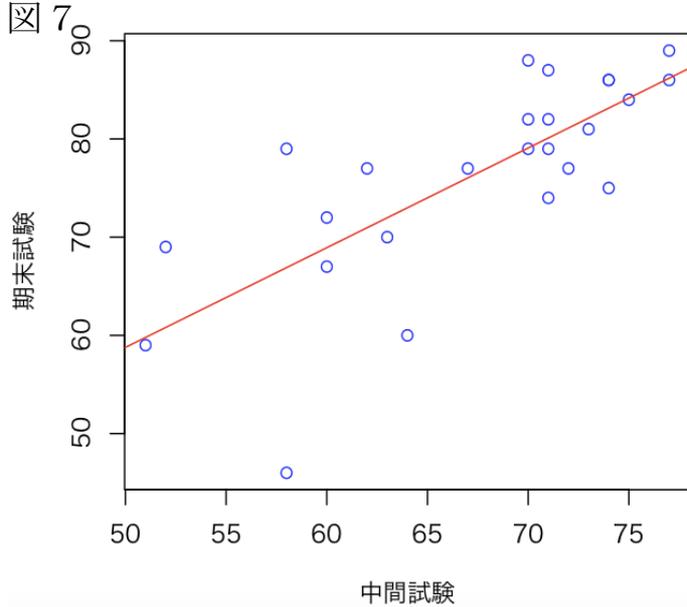


図6



散布図にラベルを付けたり点の色や形を変えたり回帰直線を加えたりして見栄えを良くする (図7)

```
par(family="HiraKakuPro-W3") # Mac で日本語表記する
plot(Test2~Test3,xlab="中間試験",ylab="期末試験", xlim=c(50,90), ylim=c(50,90), col="blue")
abline(lm(Test2~Test3), col="red")
```



相関係数と散布図とヒストグラムをいっぺんに描く その1

```
pairs.panels(d4) # 全てのペアで描く場合 (図8)
pairs.panels(d4[, 1:2]) # ペアを指定して描く場合 (図9)
```

図8

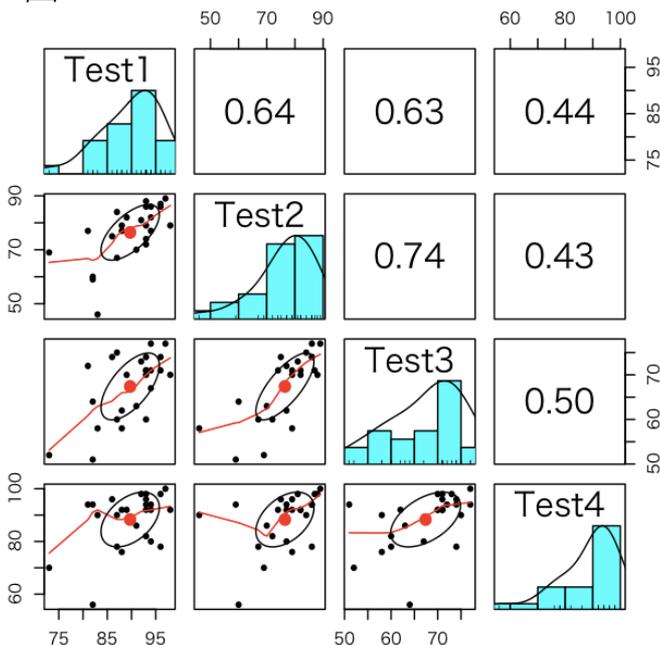
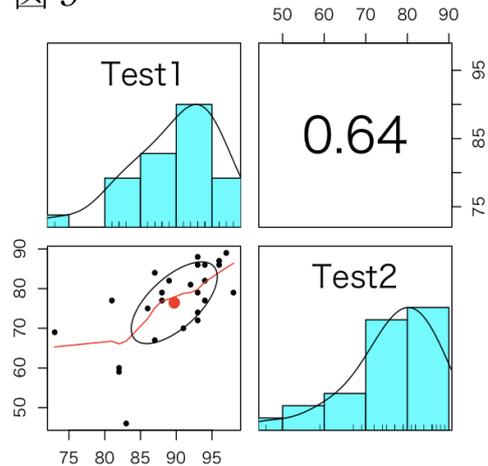


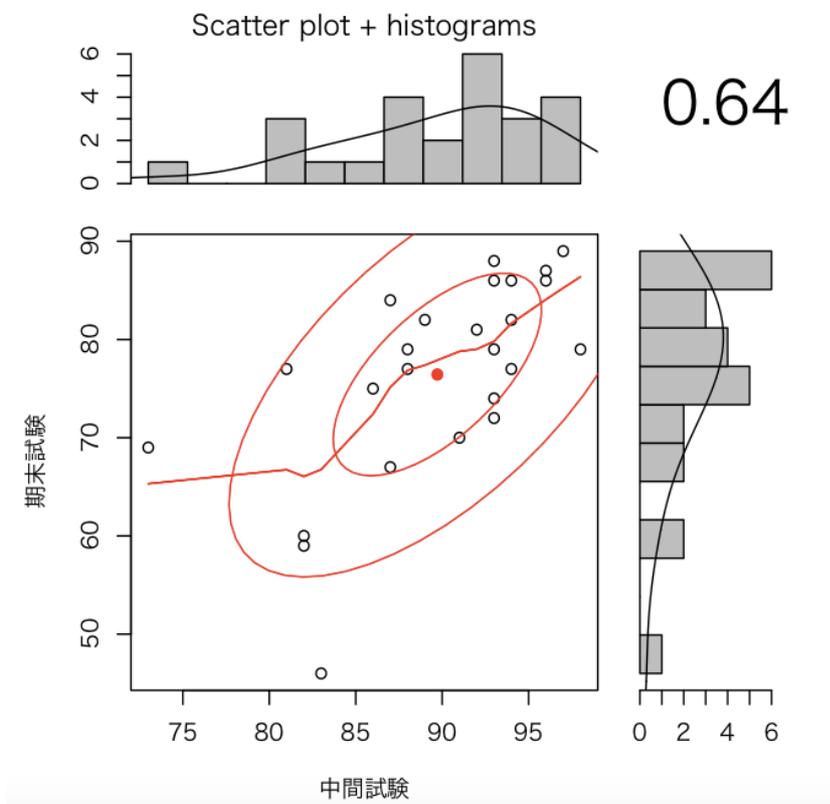
図9



相関係数と散布図とヒストグラムをいっぺんに描く その2
`scatter.hist(d4[,1:2],xlab="中間試験",ylab="期末試験")`

ペアを指定して描く (図10)

図 10



統計についてきちんと学びたい場合は、広島大学で行われている「言語教育データ分析勉強会」等に参加して勉強しましょう。

言語教育データ分析勉強会 2018

日時 : 毎週木曜日 18:00~19:30

場所 : 広島大学総合科学部 J307

講師 : 草薙邦広先生・ 鬼田崇作先生・ 阪上辰也先生

(広島大学 外国語教育研究センター)

<https://sites.google.com/site/hiroshima2018dataanalysis/>