

Exercise 2

Code ▼

問題

講座内容に関するアンケート直近50名分dataを基に

評価が変化したかを検定する

↓

平均値とかこの平均値を比べる

↓

それぞれの数値に有意差があるかを検定で確かめる

- 1. dataを読み込む
 - 。csv 自体でも内容を確認
- 2. histogrameで確認
 - 。色々階級を変えて確認
- 3. 平均と標準偏差を求め検定に必要な統計量を算出
 - 。それぞれの統計量が棄却域に入るかを確認
- 4. Rの関数 t.test を使って確認

1. csv file 読み込み

- working dirctory 確認

Hide

```
getwd()
```

```
[1] "/home/rstudio/r_rstudio/statistics/02_inferential"
```

Hide

```
data <- read.csv("data02_mac.csv")
data
```

参加者 <int>	年齢 <int>	性別 <chr>	設問1 <int>	設問2 <int>	設問3 <int>	設問4 <int>	設問5 <int>
1	22	男性	5	2	4	1	5
2	40	女性	4	2	4	1	4
3	28	男性	4	2	4	4	3
4	34	女性	3	1	4	4	4
5	33	女性	4	2	5	3	4
6	38	男性	4	1	4	5	5
7	38	男性	5	3	4	3	4
8	42	男性	5	2	5	2	3
9	50	女性	5	3	4	4	4
10	20	女性	4	3	5	3	5

1-10 of 50 rows

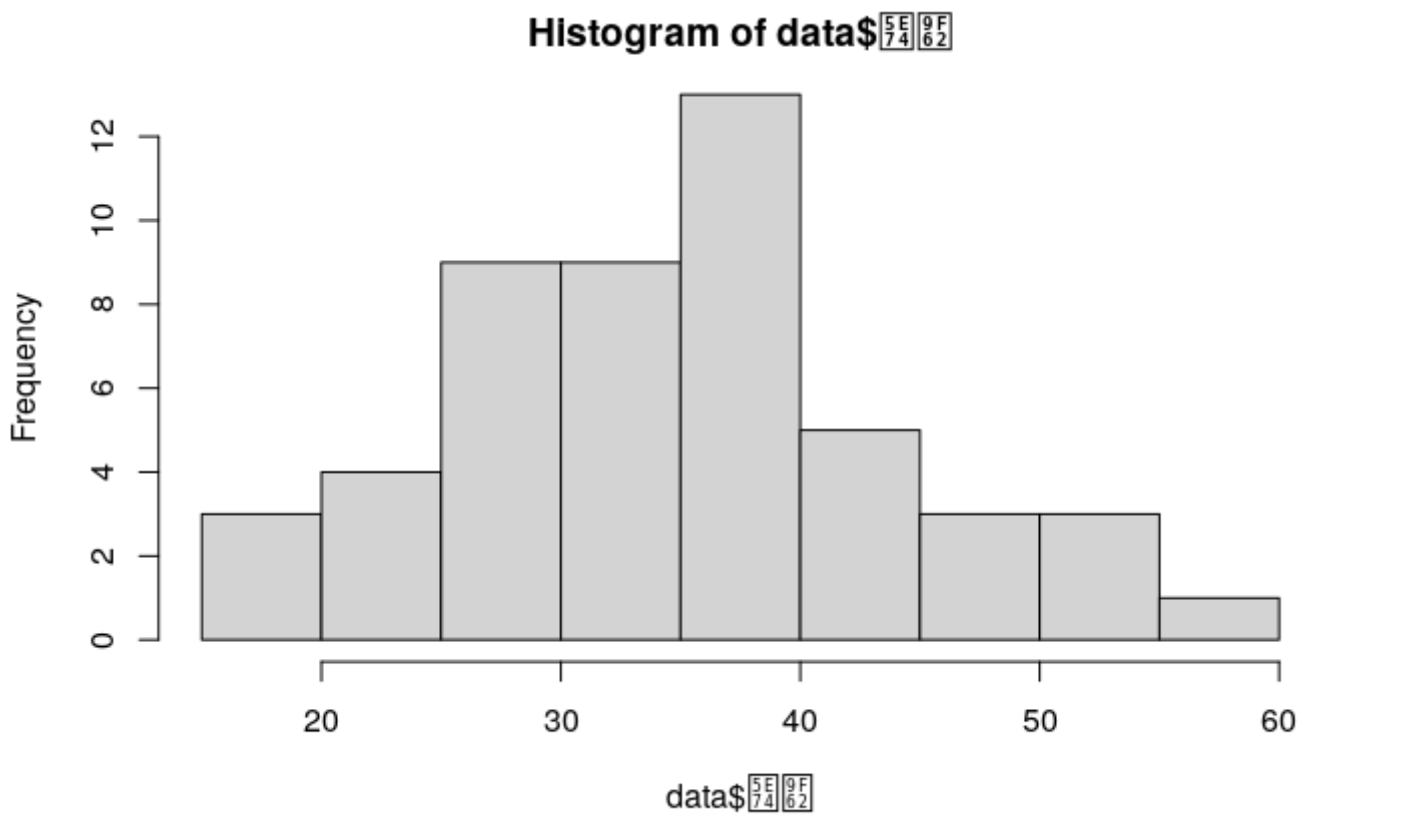
Previous12345Next

2. histograme で確認

- まずは histograme で確認
- 年齢

Hide

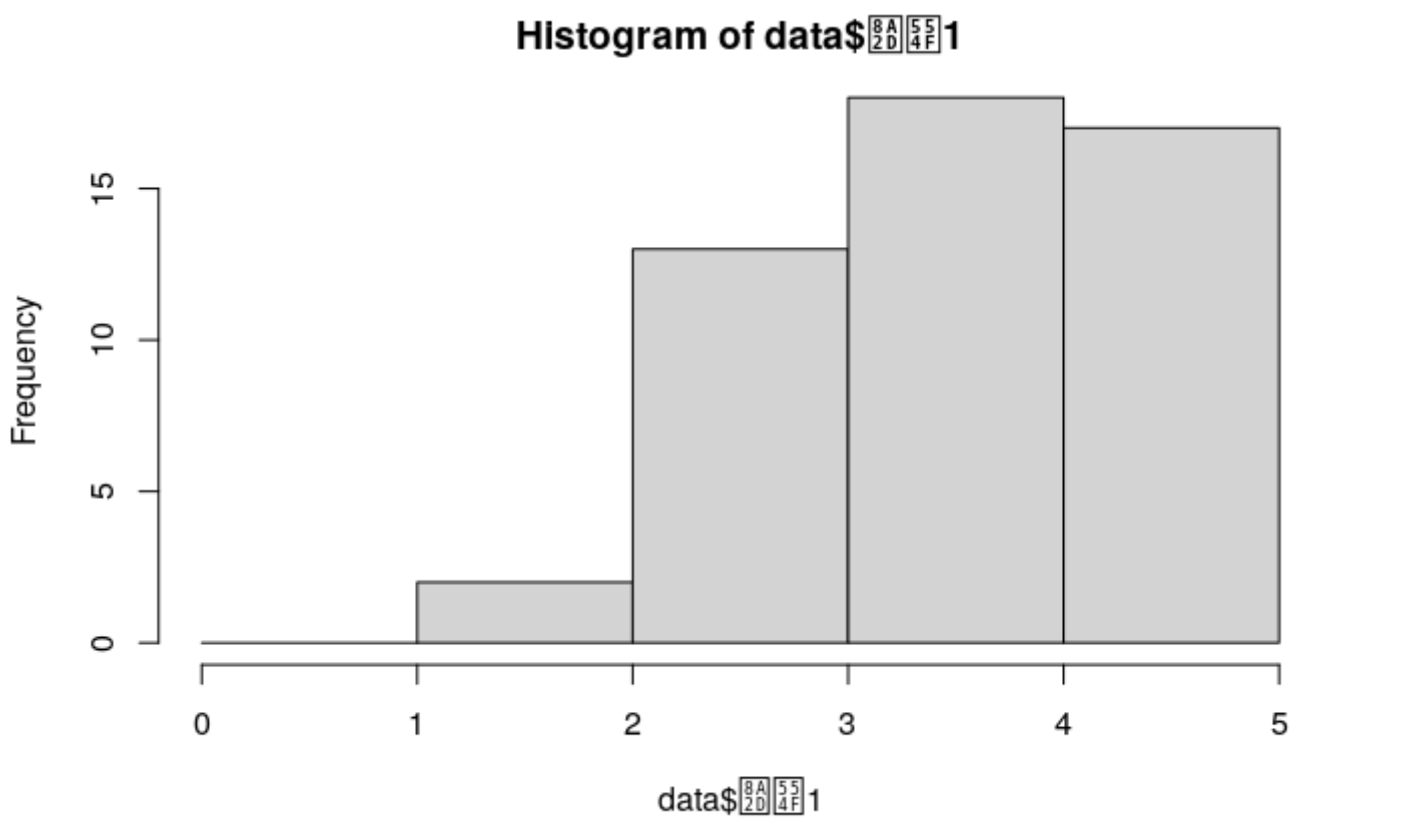
```
hist(data$年齢)
```



- 設問1
 - 3, 4, 5 の高評価が多い

Hide

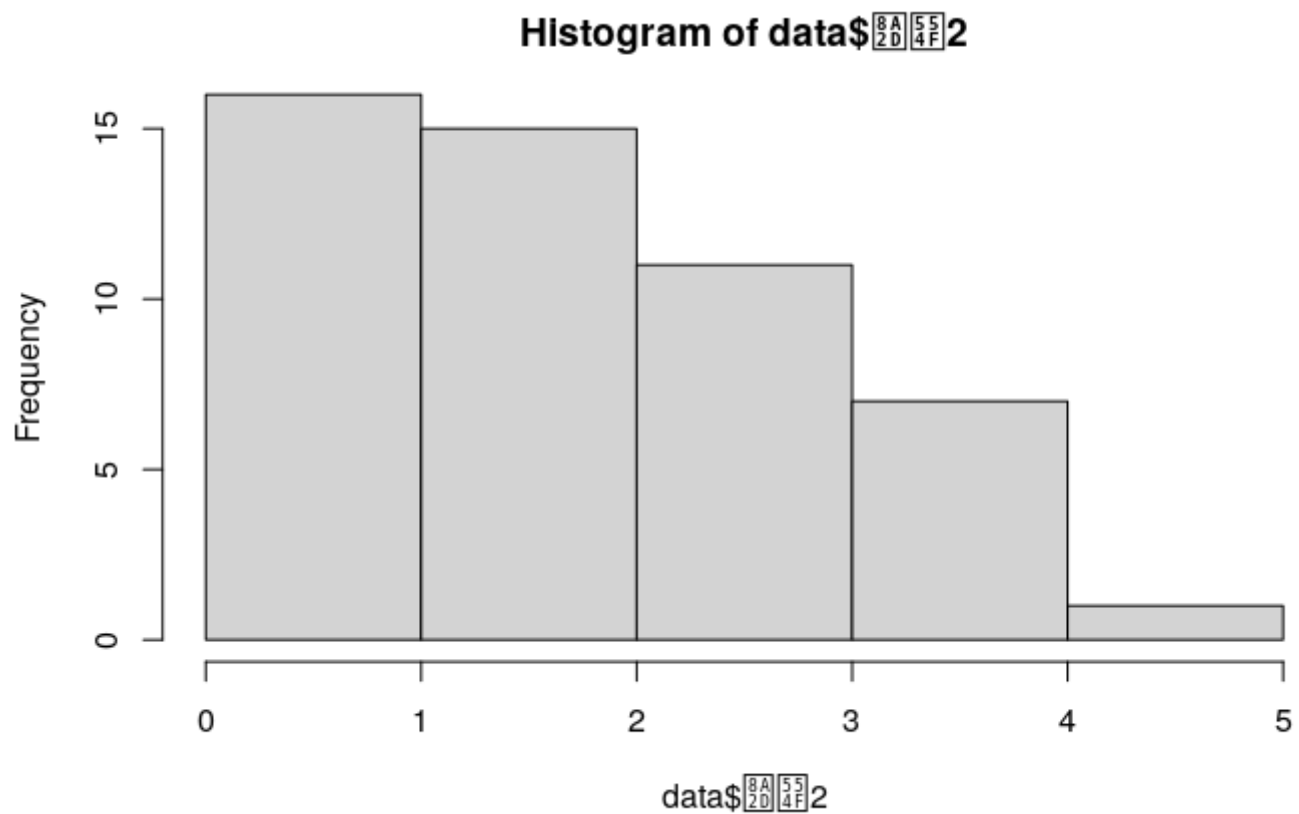
```
hist(data$設問1, breaks = seq(0, 5, 1))
```



- 設問2
 - 1, 2, 3 の低評価が多い

Hide

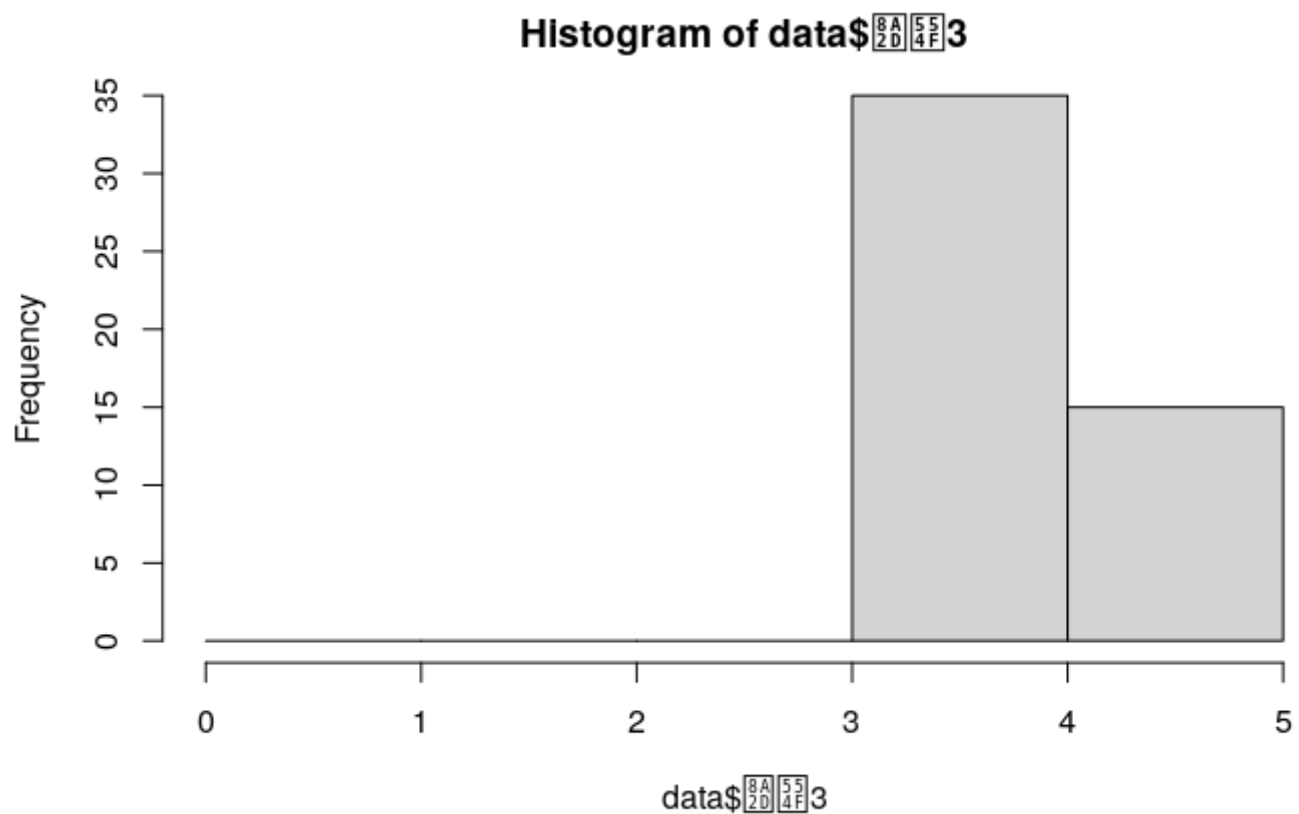
```
hist(data$設問2, breaks = seq(0, 5, 1))
```



- 設問3
 - 4, 5 のしかない

Hide

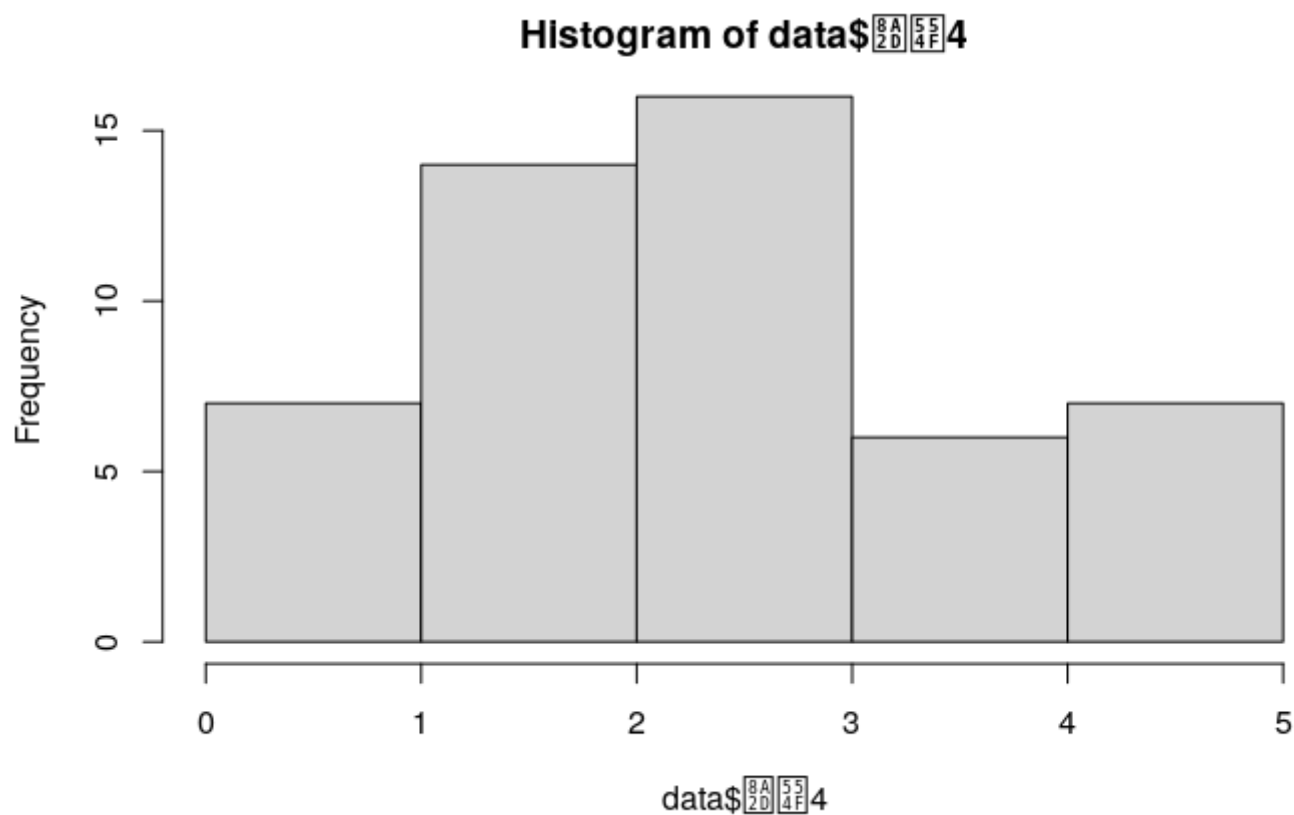
```
hist(data$設問3, breaks = seq(0, 5, 1))
```



- 設問4
 - 満遍なく広がっている

Hide

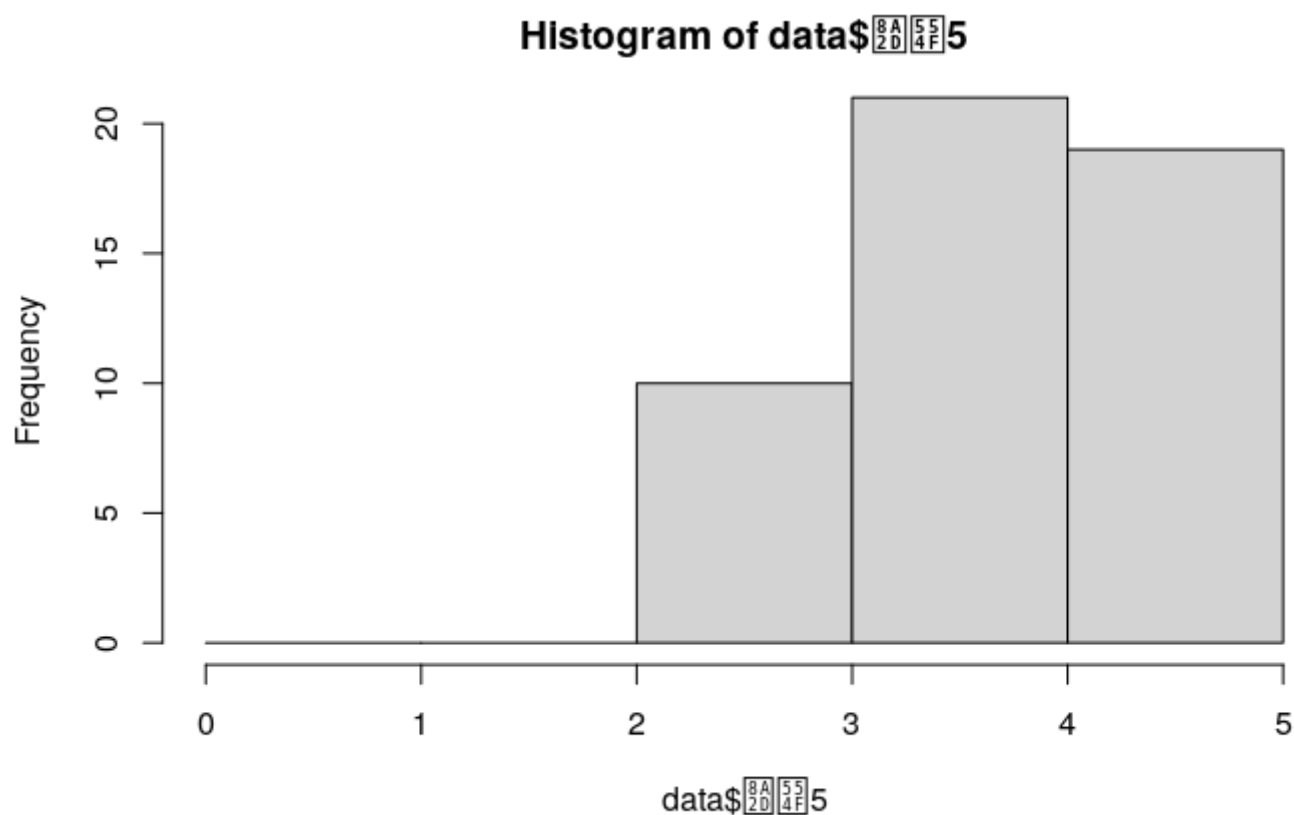
```
hist(data$設問4, breaks = seq(0, 5, 1))
```



- 設問5
 - 3, 4, 5 が多い傾向がある

Hide

```
hist(data$設問5, breaks = seq(0, 5, 1))
```



3. 平均と標準偏差を求め検定に必要な統計量を算出

- それぞれの統計量が棄却域に入るかを確認

設問	これまでの平均
1	3.5
2	2.5
3	4.2
4	3.2

設問	これまでの平均
5	3.7

検定

設問 1	値
現在の平均値	4
標準偏差	0.8806306
これまでの平均値	3.5
標本数	50
統計量	4.0147753

Hide

```
xm <- mean(data$設問1)
xm
```

[1] 4

Hide

```
s1 <- sd(data$設問1)
m1 <- 3.5
n1 <- nrow(data)
```

標準化

=

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Hide

```
z1 <- (xm - m1)/(s1 / sqrt(n1))
```

統計量の結果から

- 統計量 : 4.0147753
 - 95%で検定をした場合, **珍しいことが起きている区間にあたる** ので...
 - 平均値が 3.5 -> 4 に上がっていたということは, **有意差** があり意味がある数である

上記を for文で回す

- for 文で 設問 1 ~ 5までの統計量を求める

Hide

```
mm = c(3.5, 2.5, 4.2, 3.2, 3.7)
mm
```

[1] 3.5 2.5 4.2 3.2 3.7

Hide

```
xx = rep(0, 5)
xx
```

[1] 0 0 0 0 0

Hide

```
zz = rep(0, 5)
zz
```

```
[1] 0 0 0 0 0
```

- **mm** : 3.5, 2.5, 4.2, 3.2, 3.7
 - 基の平均値を配列に入れる
- **xx**
 - 0の配列を作成 & 平均を入れる入れ物を作成
- **zz**
 - 0の配列を作成 & 統計量を入れる入れ物を作成

Hide

```
for (i in 1:5) {
  xx[i] = mean(data[,3+i])
  zz[i] = (mean(data[,3+i]) - mm[i]) / (sd(data[,3+i]) / sqrt(n1))
}
xx
```

```
[1] 4.00 2.24 4.30 2.84 4.18
```

Hide

```
zz
```

```
[1] 4.014775 -1.646132 1.527525 -2.061423 4.540541
```

for文後の検定結果

統計量	数値	基の μ	現在 μ	有意差
設問1 統計量	4.0147753	3.5	4	有(+なので良い)
設問2 統計量	-1.6461322	2.5	2.24	無し(よくある事)
設問3 統計量	1.5275252	4.2	4.3	無し(よくある事)
設問4 統計量	-2.061423	3.2	2.84	有 (-なので注意)
設問5 統計量	4.5405405	3.7	4.18	有(+なので良い)