
演習問題 解答

演習問題：一変数の記述統計 1.

1

1. サンプルデータbusinessは、国別の産業データ（1993年）です。

- ① 国籍、産業の度数分布表とそのグラフを作成してみましょう。
- ② 売上高を以下のようにカテゴリ化して、度数分布表を作成してみましょう。また、ヒストグラムを作成してみましょう。

1000ドル未満

1000ドル以上-5000ドル未満

5000ドル以上-10000ドル未満

10000ドル以上

演習問題：一変数の記述統計 1. 解答 (1/2)

2

>#1. サンプルデータbusinessの読み込み

>ビジネス=read.csv("business.csv")

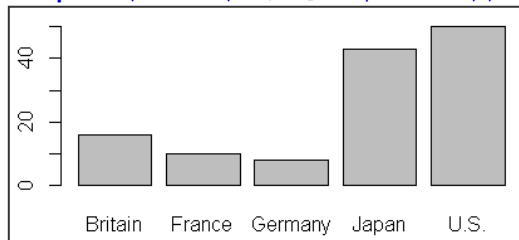
>#1. ①国籍毎の度数分布

>table(ビジネス\$NATION)

Britain	France	Germany	Japan	U. S.
16	10	8	43	50

>#1. ①国籍毎のグラフ

>barplot(table(ビジネス\$NATION))



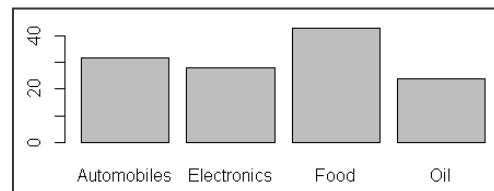
>#1. ①産業毎の度数分布

>table(ビジネス\$INDUSTRY)

Automobiles	Electronics	Food	Oil
32	28	43	24

>#1. ①産業毎のグラフ

>barplot(table(ビジネス\$INDUSTRY))



演習問題：一変数の記述統計 1. 解答 (2/2)

3

>#1. ②カテゴリ化して度数分布表を作成

>カテゴリ化=cut (ビジネス\$SALES, breaks=c (0, 1000, 5000, 10000, Inf), right=FALSE)

>levels (カテゴリ化)=c ("1000ドル未満", "1000ドル以上-5000ドル未満", "5000ドル以上-10000ドル未満", "10000ドル以上")

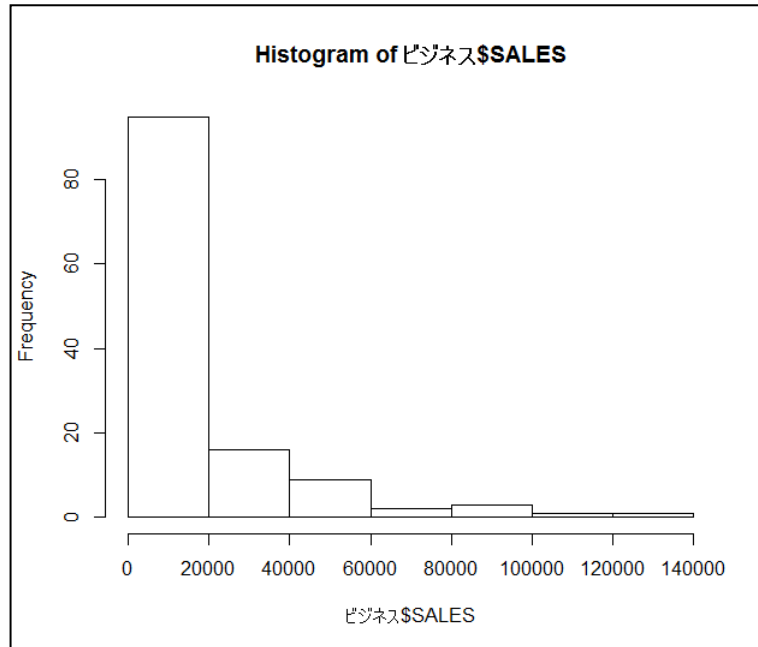
>table (カテゴリ化)

カテゴリ化

1000ドル未満	1000ドル以上-5000ドル未満	5000ドル以上-10000ドル未満	10000ドル以上
0	19	47	61

>#1. ②カテゴリ化してヒストグラムを作成

>hist (ビジネス\$SALES)



演習問題：一変数の記述統計 2.

4

2. サンプルデータcolorは、ヨーロッパの2つの地域（Region）の子供の髪（Hair）と目の色（Eyes）を調査し、その度数（Count）を集計したデータです。子供の髪の色、目の色の度数分布表とそのグラフを地域別に作成してみましょう。

ヒント①：subset関数を使用し、地域別のデータを作成してください。

例：subset(身長, Region==1)

ヒント②：グループ毎の集計は、グループ化されたデータ処理が可能なtapply関数を使用してください。

指定方法：tapply(【集計したい項目】 , 【グループ指定】 , 【関数】)

例：tapply(Count, 地域毎のHair, sum)

演習問題：一変数の記述統計 2. 解答

5

>#サンプルデータcolorの読み込み

>カラー=read.csv("color.csv")

>#読み込んだデータを、地域別に作成

>カラー地域1=subset(カラー, Region==1)

>カラー地域2=subset(カラー, Region==2)

>#地域別の子供の髪の色を度数分布表

>地域1_髪色=apply(カラー地域1\$Count, カラー地域1\$Hair, sum)

>地域1_髪色

black	dark	fair	medium	red
3	65	76	83	19

>地域2_髪色=apply(カラー地域2\$Count, カラー地域2\$Hair, sum)

>地域2_髪色

black	dark	fair	medium	red
19	117	152	134	94

>#地域別の子供の目の色を度数分布表

>地域1_目色=apply(カラー地域1\$Count, カラー地域1\$Eyes, sum)

>地域1_目色

blue	brown	green
65	123	58

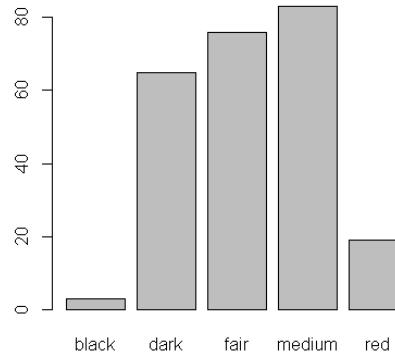
>地域2_目色=apply(カラー地域2\$Count, カラー地域2\$Eyes, sum)

>地域2_目色

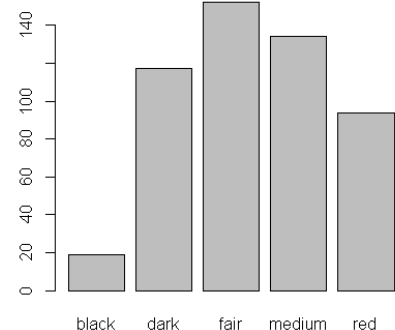
blue	brown	green
157	218	141

>#地域別の子供の髪の色 グラフ

>barplot(地域1_髪色)

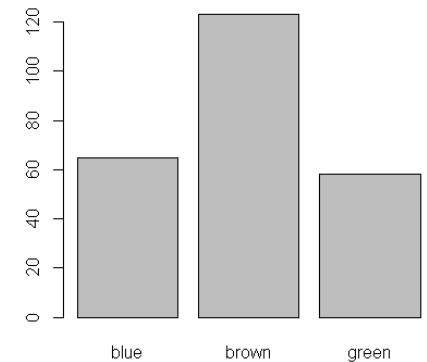


>barplot(地域2_髪色)

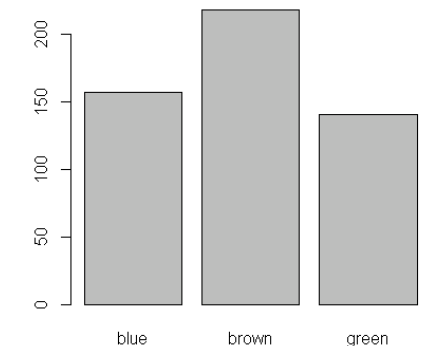


>#地域別の子供の目の色 グラフ

>barplot(地域1_目色)



>barplot(地域2_目色)



演習問題：一変数の記述統計 3.

6

3. サンプルデータ business に対して、

- ① 売上高 (SALES) と利益 (PROFITS) の要約統計量を計算し、変数の分布を確認してみましょう。
- ② 産業 (INDUSTRY) 別に売上高のヒストグラムと箱ひげ図を描いてみましょう。
- ③ 国籍 (NATION) および産業別に売上高の平均、標準偏差、変動係数を計算しましょう。日本の電子産業と食品産業ではどちらの売上高の方がばらつきが大きいといえるでしょうか。また、アメリカの電子産業と石油産業とではどうでしょうか。

演習問題：一変数の記述統計 3. 解答 (1/5)

7

```
>#サンプルデータbusinessの読み込み
```

```
>ビジネス=read.csv("business.csv")
```

```
>#①売上高の要約統計量
```

```
>summary(ビジネス$SALES)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2978	6148	9554	18488	19957	133622

```
>#①利益の要約統計量
```

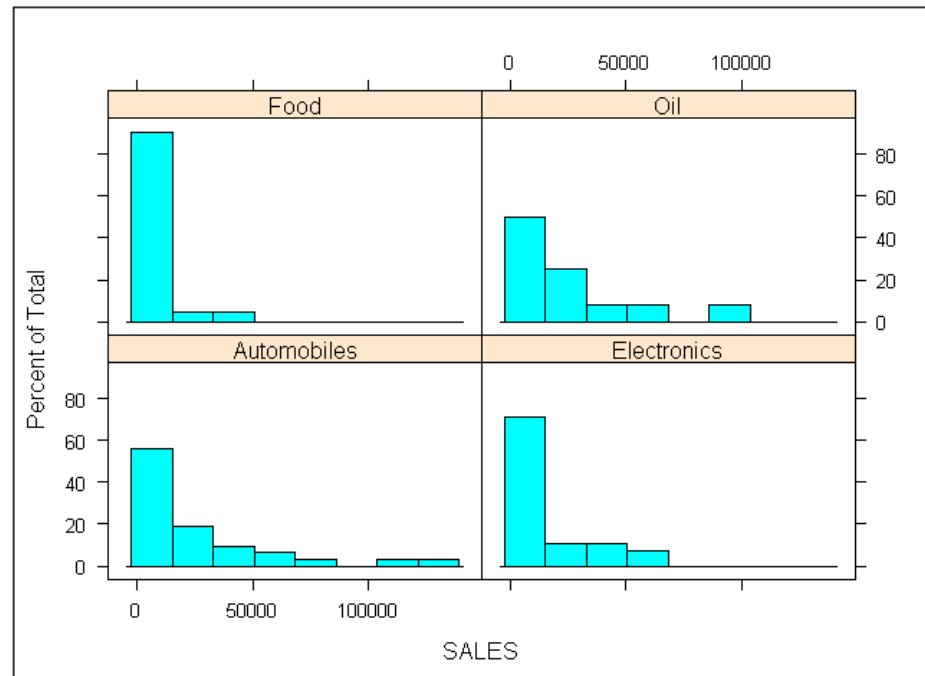
```
>summary(ビジネス$PROFITS)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
-2551.0	51.0	151.0	442.2	506.0	5280.0	2

```
>#②産業別の売上高の層別ヒストグラム
```

```
>library(lattice)
```

```
>histogram(~SALES|INDUSTRY, data=ビジネス)
```



演習問題：一変数の記述統計 3. 解答 (2/5)

8

>#②産業別の売上高の箱ひげ図

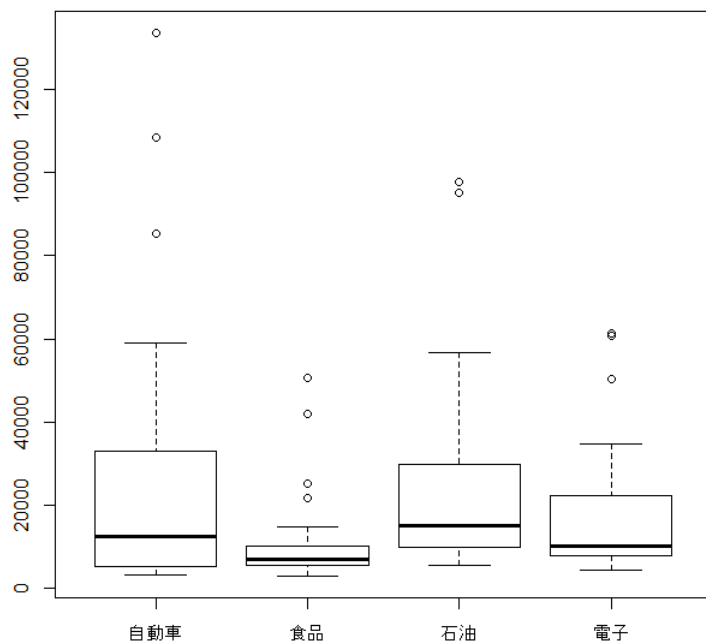
```
>自動車=subset(ビジネス, INDUSTRY=="Automobiles")$SALES
```

```
>食品=subset(ビジネス, INDUSTRY=="Food")$SALES
```

```
>石油=subset(ビジネス, INDUSTRY=="Oil")$SALES
```

```
>電子=subset(ビジネス, INDUSTRY=="Electronics")$SALES
```

```
>boxplot(自動車, 食品, 石油, 電子, names=c("自動車", "食品", "石油", "電子"))
```



演習問題：一変数の記述統計 3. 解答 (3/5)

9

>#③国籍および産業別に売上高の平均、標準偏差、変動係数を算出

>#subset関数で日本とアメリカを抽出

>ビジネス_JP=subset(ビジネス, NATION=="Japan")

>ビジネス_US=subset(ビジネス, NATION=="U. S.")

>#subset関数で日本の産業別（電子と食品）に抽出

>ビジネス_JP_E=subset(ビジネス_JP, INDUSTRY=="Electronics")

>ビジネス_JP_F=subset(ビジネス_JP, INDUSTRY=="Food")

>#subset関数でアメリカの産業別（電子と石油）に抽出

>ビジネス_US_E=subset(ビジネス_US, INDUSTRY=="Electronics")

>ビジネス_US_O=subset(ビジネス_US, INDUSTRY=="Oil")

演習問題：一変数の記述統計 3. 解答 (4/5)

10

```
>#日本の電子産業
>#平均の出力
>日本電子産業_平均=mean(ビジネス_JP_ESALES)
>日本電子産業_平均
[1] 18567.75
>#標準偏差の出力
>日本電子産業_標準偏差=sqrt(mean((ビジネス_JP_ESALES-日本電子産業_平均)^2))
>日本電子産業_標準偏差
[1] 16829.42
>#変動係数の出力
>日本電子産業_標準偏差/日本電子産業_平均 *100
[1] 90.63793

>#日本の食品産業
>#平均の出力
>日本食品産業_平均=mean(ビジネス_JP_FSALES)
>日本食品産業_平均
[1] 6438.182
>#標準偏差の出力
>日本食品産業_標準偏差=sqrt(mean((ビジネス_JP_FSALES-日本食品産業_平均)^2))
>日本食品産業_標準偏差
[1] 1869.507
>#変動係数の出力
>日本食品産業_標準偏差/日本食品産業_平均 *100
[1] 29.03781
```

演習問題：一変数の記述統計 3. 解答 (5/5)

11

```
>#アメリカの電子産業
>#平均の出力
>アメリカ電子産業_平均=mean(ビジネス_US_E$SALES)
>アメリカ電子産業_平均
[1] 14015
>#標準偏差の出力
>アメリカ電子産業_標準偏差=sqrt(mean((ビジネス_US_E$SALES-アメリカ電子産業_平均)^2))
>アメリカ電子産業_標準偏差
[1] 15091.4
>#変動係数の出力
>アメリカ電子産業_標準偏差/アメリカ電子産業_平均 *100
[1] 107.6804

>#アメリカの石油産業
>#平均の出力
>アメリカ石油産業_平均=mean(ビジネス_US_O$SALES)
>アメリカ石油産業_平均
[1] 24751.54
>#標準偏差の出力
>アメリカ石油産業_標準偏差=sqrt(mean((ビジネス_US_O$SALES-アメリカ石油産業_平均)^2))
>アメリカ石油産業_標準偏差
[1] 25006.2
>#変動係数の出力
>アメリカ石油産業_標準偏差/アメリカ石油産業_平均 *100
[1] 101.0289
```

演習問題：統計的仮説検定 解答（1/2）

12

1.ある工場では、1個の内容量が80グラムとなるようにヨーグルトを生産している。生産されたヨーグルトから無作為に選んだ5個のヨーグルトの内容量を調べたところ、79,80,81,81,81（単位：グラム）であった。この工場で生産されているヨーグルトの内容量が80グラムであるかどうかを検定しなさい。

> ヨーグルト内容量=c(79,80,81,81,81)

> ヨーグルト内容量

[1] 79 80 81 81 81

> t.test(ヨーグルト内容量,mu=80)

One Sample t-test

data: ヨーグルト内容量

t = 1, df = 4, p-value = 0.3739

alternative hypothesis: true mean is not equal to 80

95 percent confidence interval:

79.28942 81.51058

sample estimates:

mean of x

80.4

仮説を立てる H_0 ：等しい
 H_1 ：異なる

α を定める

p 値を計算する

$p \text{ 値} \geq \alpha$ 帰無仮説を採択
 $p \text{ 値} < \alpha$ 帰無仮説を棄却

統計的仮説検定のフロー

H_0 ：ヨーグルトの容量は80グラムと等しい

H_1 ：ヨーグルトの容量は80グラムと等しくない

有意水準 $\alpha = 0.05$ （両側検定）

p 値 = 0.3739

0.3739（p 値） > 0.05（有意水準 α ）より、帰無仮説を採択。

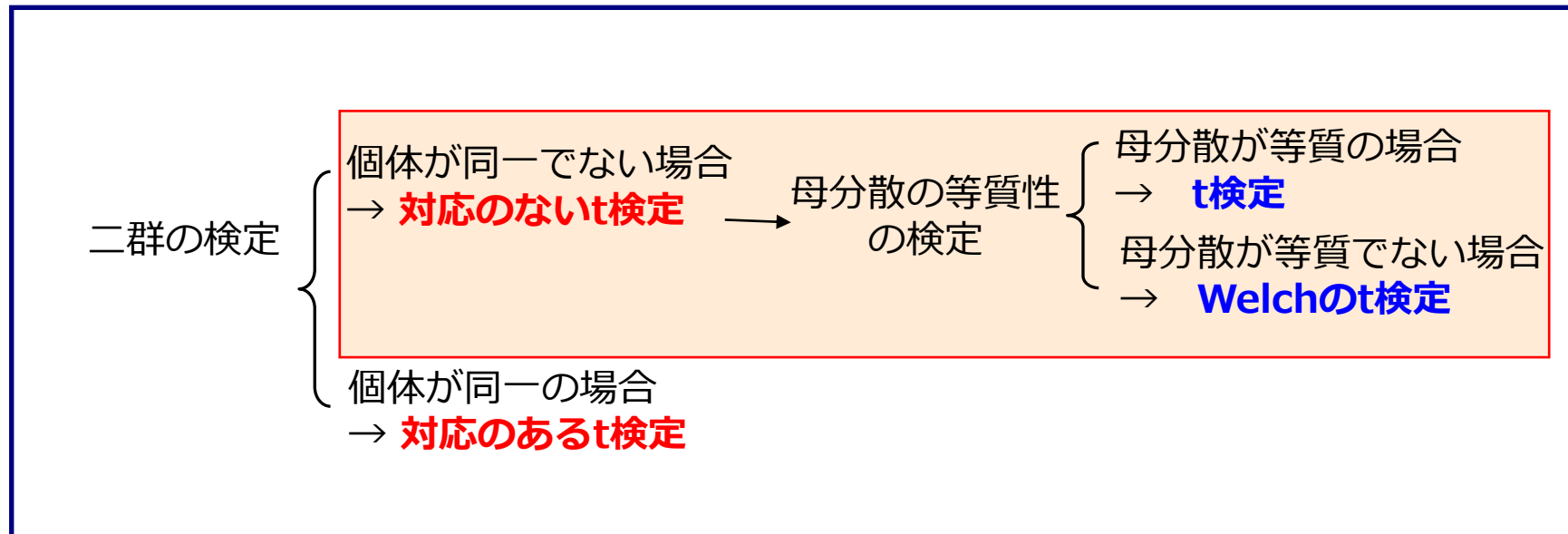
【結論】 工場で生産されるヨーグルトの容量は80グラムであるといえる。

演習問題：二群の検定 解答（1/5）

14

指導法データの「数学テスト」の得点について、男女で有意な差があるでしょうか。有意水準5%で確認してください。

二群の検定の全体像



演習問題：二群の検定 解答（2/5）

15

母分散の等質性を検定します。

```
>指導法_男=subset(指導法,性別=="男")  
>指導法_女=subset(指導法,性別=="女")  
>var.test(指導法_男$数学テスト,指導法_女$数学テスト)
```

出力結果

F test to compare two variances

data: 指導法_男\$数学テスト and 指導法_女\$数学テスト

F = 1.5379, num df = 11, denom df = 7, p-value = 0.583

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.3265503 5.7803295

sample estimates:

ratio of variances

1.537879

演習問題：二群の検定 解答（3/5）

16

仮説を立てる H_0 ：等しい
 H_1 ：異なる

α を定める

p 値を計算する

p 値 $\geq \alpha$ 帰無仮説を採択
p 値 $< \alpha$ 帰無仮説を棄却

統計的仮説検定のフロー

H_0 ：二群の母分散は等しい

H_1 ：二群の母分散は等しくない

有意水準 $\alpha = 0.05$

p 値 = 0.583

0.583 (p 値) > 0.05 (有意水準 α) より、帰無仮説を採択。

【結論】 二群の母分散は等しいといえる。

演習問題：二群の検定 解答（4/5）

17

母分散の等質性が満たされているため、t検定を行います。

```
>t.test(指導法_男$数学テスト,指導法_女$数学テスト,var.equal=TRUE)
```

出力結果

Two Sample t-test

data: 指導法_男\$数学テスト and 指導法_女\$数学テスト

t = 0.5238, df = 18, p-value = 0.6068

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2.508958 4.175625

sample estimates:

mean of x mean of y

10.33333 9.50000

演習問題：二群の検定 解答（5/5）

18

仮説を立てる H_0 ：等しい
 H_1 ：異なる

α を定める

p 値を計算する

p 値 $\geq \alpha$ 帰無仮説を採択
p 値 $< \alpha$ 帰無仮説を棄却

統計的仮説検定のフロー

H_0 ：二群の数学テストの平均値に差がない
 H_1 ：二群の数学テストの平均値に差がある

有意水準 $\alpha = 0.05$

p 値 = 0.6068

0.6068 (p 値) > 0.05 (有意水準 α) より、帰無仮説を採択。

【結論】二群の数学テストの平均値に差がないといえる。