



確率と統計

中山クラス
第12週



本日の内容

- ◆第3回レポートの成績評価
- ◆第11章 統計解析で分かること・分からないこと
- ◆第5回レポート課題の説明
- ◆第11章のコンピュータ演習
- ◆第4回レポート作成

第3回レポートの成績評価

- A+: 課題を全てクリアし, 用語説明, 考察が優れている.
- A: 課題を全てクリア
- A-: 課題を一応クリアしているが, 一部にミスが見られる.
- B: 1問分が不完全である.
- C: 2問分が不完全である.

第11章 統計解析で分かること・ 分からないこと

問題：子供のゲーム時間と親の認識

調査結果：表11.1 (x : 実際の時間, y : 親の認識)

(1) このデータについて以下の解析を行う

(1.1) x , y それぞれについて度数分布, 代表値, 散布図

(1.2) $y-x$ について度数分布, 代表値, 散布図

(1.3) x と y の相関係数

(1.4) x を独立変数, y を従属変数とした回帰式

(2) (1)の分析結果から分かること, 分からないこと

(3) (2)をふまえて, 実態と認識のずれを調べる分析方法

データ収集

```
> t11_1 <- read.csv("ch11.csv")
```

```
> t11_1
```

```
  親子  x  y  
1    1 290 350  
2    2  50  70  
3    3  80 100  
4    4 100 130  
5    5 200 250  
<以下, 省略>
```

```
> x<-t11_1[,2]
```

```
> x
```

```
[1] 290  50  80 100 200 350 430  80 210 110  70 260
```

```
> y<-t11_1[,3]
```

```
> y
```

```
[1] 350  70 100 130 250 430 520 100 260 140  90 320
```

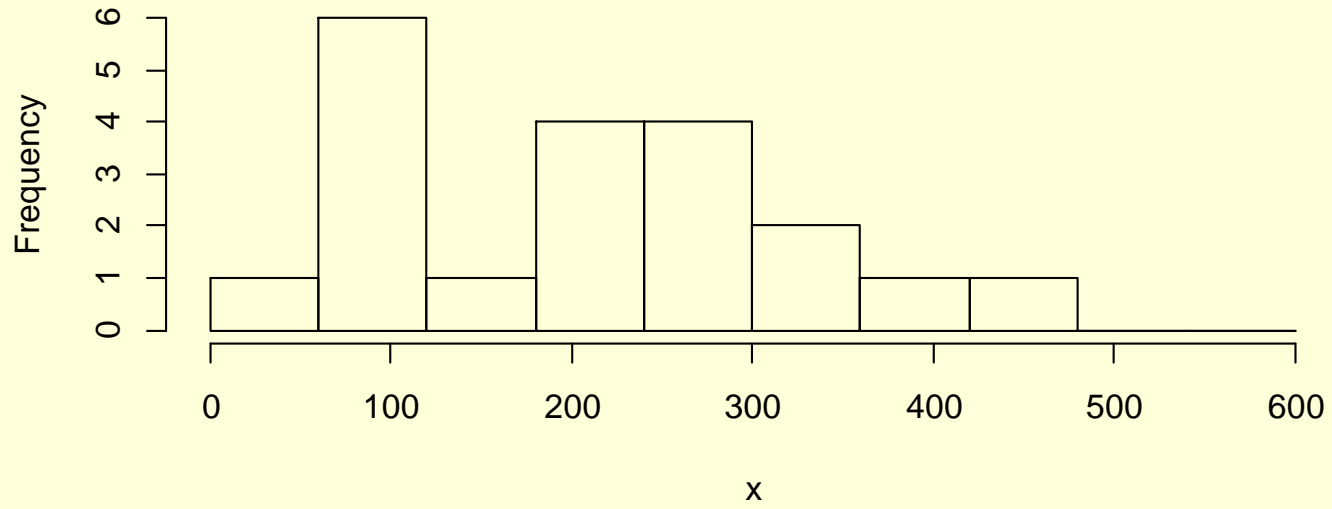
(1.1) xとyの度数分布, 代表値, 散布度

```
> par(mfrow=c(2,1))
```

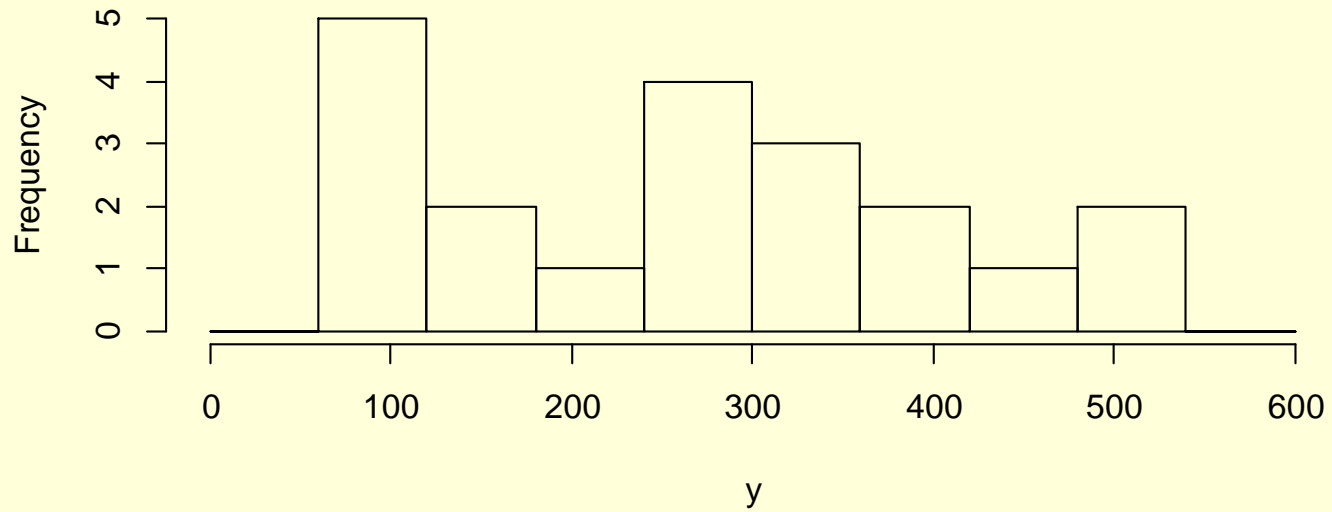
```
> hist(x, breaks=c(0,60,120,180,240,300,360,  
> 420,480,540,600))
```

```
> hist(y, breaks=c(0,60,120,180,240,300,360,  
> 420,480,540,600))
```

Histogram of x



Histogram of y



代表值：平均，中央值

```
> mean(x)
```

```
[1] 213.5
```

```
> median(x)
```

```
[1] 215
```

```
> mean(y)
```

```
[1] 262.5
```

```
> median(y)
```

```
[1] 265
```


散布度：標本分散，標準偏差

```
> varp <- function(x){  
+ 標本分散<-var(x)*(length(x)-1/length(x))  
+ 標本分散  
+ }
```

```
> source("varp.R")
```

```
> varp(x)  
[1] 13492.75
```

```
> varp(y)  
[1] 19328.75
```

```
> sqrt(varp(x))  
[1] 116.1583
```

```
> sqrt(varp(y))  
[1] 139.0279
```

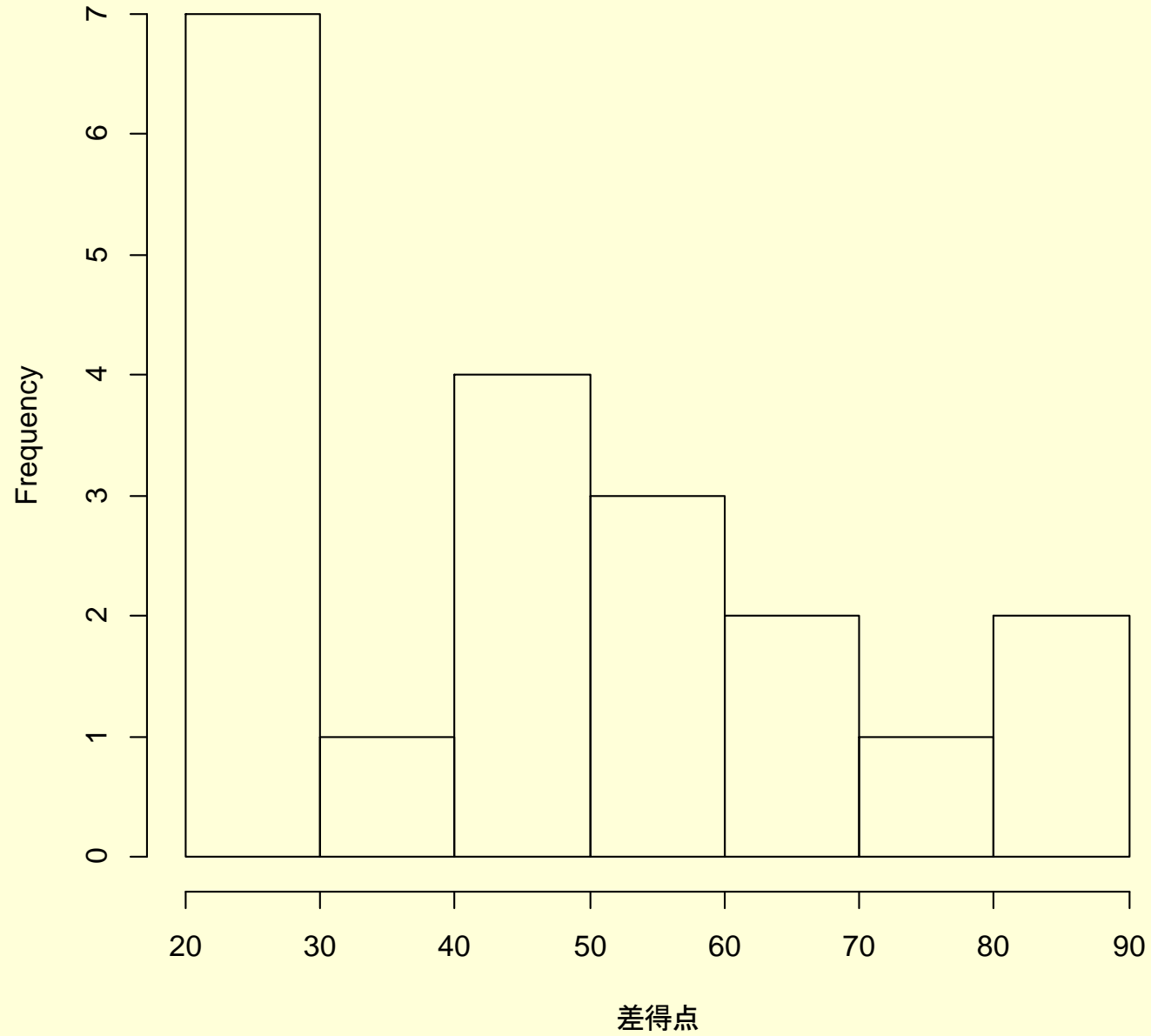
表11.2 x と y の代表値と散布度

	x	y
平均	213.5	262.5
中央値	215	265
標本分散	13492.75	19328.75
標準偏差	116.1583	139.0279

(1.2) $y-x$ の度数分布, 代表値, 散布度

```
> par(mfrow=c(1,1))
> 差得点<-y-x
> 差得点
[1] 60 20 20 30 50 80 90 20 50 30 20 60 50 70 40 90 20
70 60 50
> hist(差得点)
> mean(差得点)
[1] 49
> median(差得点)
[1] 50
> varp(差得点)
[1] 529
> sqrt(varp(差得点))
[1] 23
```

Histogram of 差得点



(1.3) xとyの相関係数

```
> cor(x,y)  
[1] 0.9998148
```

(1.4) x を独立変数, y を従属変数とした回帰式

```
> 単回帰結果<-lm(y~x)
> summary(単回帰結果)
```

Call:

```
lm(formula = y ~ x)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.0446	-2.3701	-0.3618	2.1800	4.1558

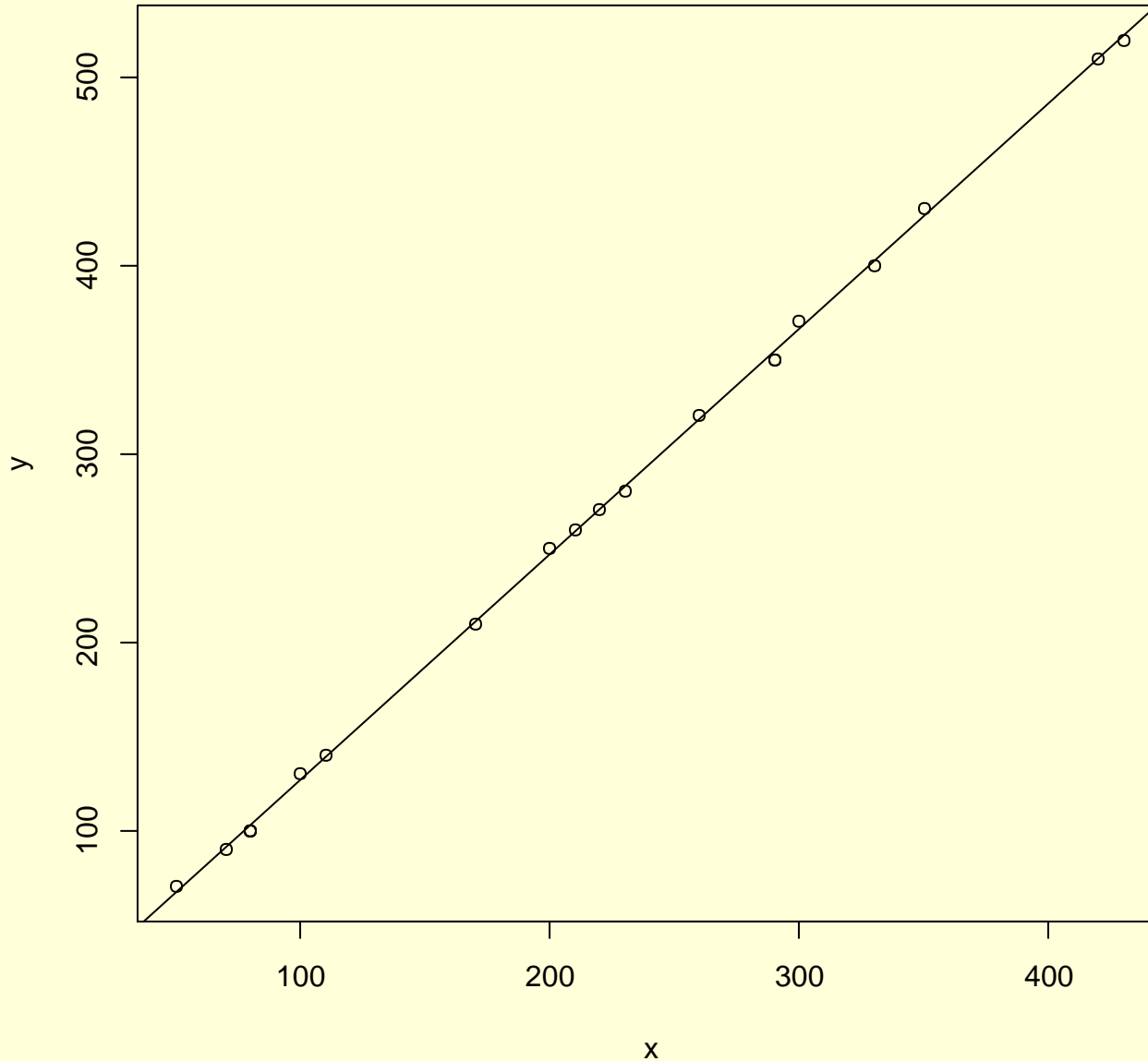
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.012840	1.319652	5.314	4.73e-05 ***
x	1.196661	0.005429	220.401	< 2e-16 ***

> plot(x,y)

> abline(単回帰結果)

$$y = 7.012840 + 1.196661x$$



問題(2)分析結果から分かること／分からないこと

- (1.1) ○ x, y個別の分布／平均値の差／xとyのずれ傾向
× 個々の親子のずれ
- (1.2) ○ ずれの全体の大きさ／ずれの個人差の分布
× ずれの個人差／どのようにずれているか
- (1.3) ○ 実態と認識に一定の傾向があるか否か
× ずれの大きさ
- (1.4) ○ 平均的なずれと倍数(親は約1.2倍と認識する)
× 相関が弱いときには不正確

問題(3) 実態と認識のずれを調べる分析方法

y-xの分布

親の認識が平均して50分ほどずれている。

相関係数

実態と認識の間にはずれがあるが、その間にはかなり密接な関係があって、一方から、他方を高い精度で予測できる。

回帰分析

1次式で近似できる場合

定数項→平均的なずれ

1次項 →何倍に見積もっているか

★1つの統計解析は一面しか見えない場合が多いので、複数の分析方法を組み合わせることも重要。

第5回レポート課題

(1) データを収集する

- ・ 大学における実験, 課題研究など
- ・ インターネットから収集
(例) ドル円の為替レート, 日本の株価
- ・ p.292の父母と娘の身長に関するデータ

(2) 統計的な性質及び相関係数またはファイ係数を解析

(3) 統計的仮説検定を行う

- ・ 母集団 $N(\mu, \sigma^2)$ からの無作為抽出であるか?
- ・ 相関があるか?
- ・ 連関があるか?

レポート作成について(1)

(2)に関して

第2章および第3章の内容を参考にする.

- 収集したデータについて説明する.
- 1つの変数を選び, 平均と標準偏差を求める.
- 2つの変数を選び, 散布図またはクロス集計表を作成
- 上記で使用した2つの変数の相関係数またはファイ係数を求める.
- 以上の分析結果から分かることを述べる.

レポート作成について(2)

(3)に関して

第5章の内容を参考にして統計的仮説検定を行う。

- データを取った目的とデータの意味を考え、適切な仮説を設定する。
- 帰無仮説と対立仮説を文章で述べる。
- 検定統計量を文章と数式で示す。
- 片側検定／両側検定，有意水準を示す。
- 検定統計量の実現値／棄却域を求め，明記する。
- 帰無仮説の棄却／採択を理由を付して述べる。
- 検定結果を文章で述べる。

今後の予定

◆7月17日(火) 金曜日の授業

◆第5回レポート締め切り

金曜クラス 7月24日(火)17:00

火曜クラス 7月24日(火)17:00

◆模擬問題演習

金曜クラス 7月20日(金)1限

火曜クラス 7月24日(火)2限

◆達成度確認試験

金曜クラス 7月27日(金)1限

火曜クラス 7月31日(火)2限