

# 重回帰分析 1

「予測」と「因果効果」

拓殖大学

浅野正彦 Ph.D.

1

**理論**

「人間の容姿は遺伝する」

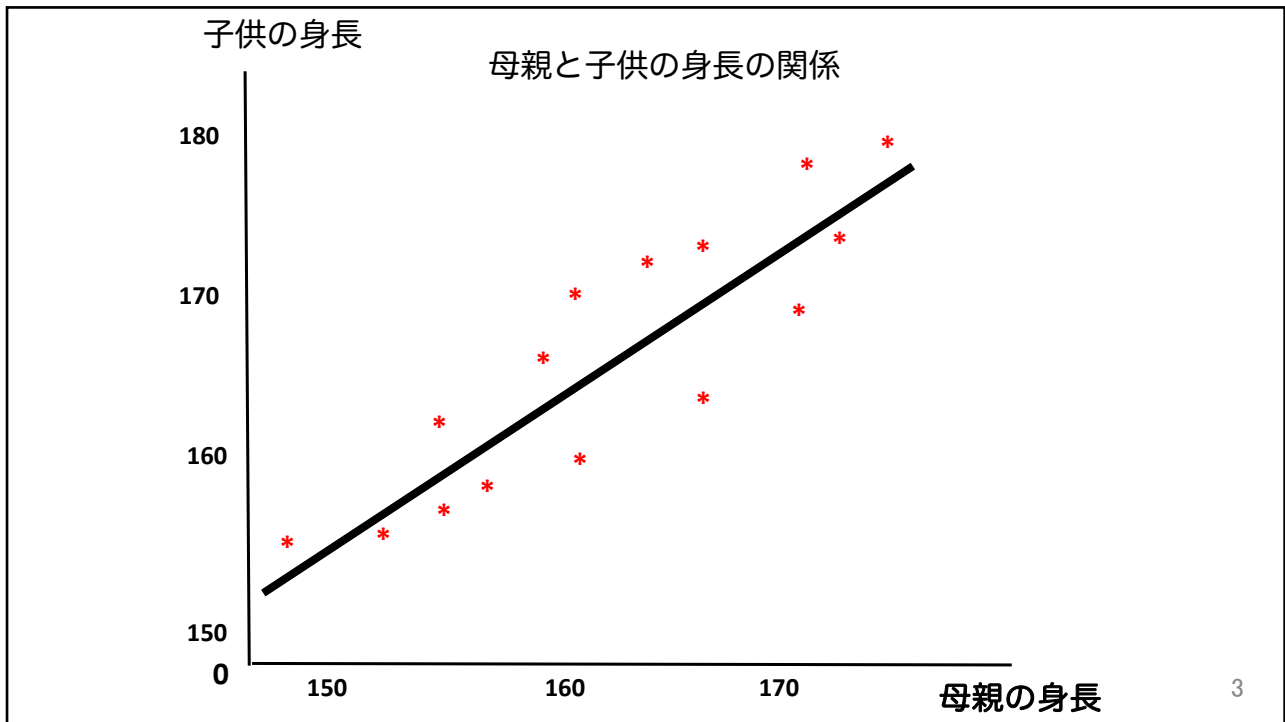


**仮説**

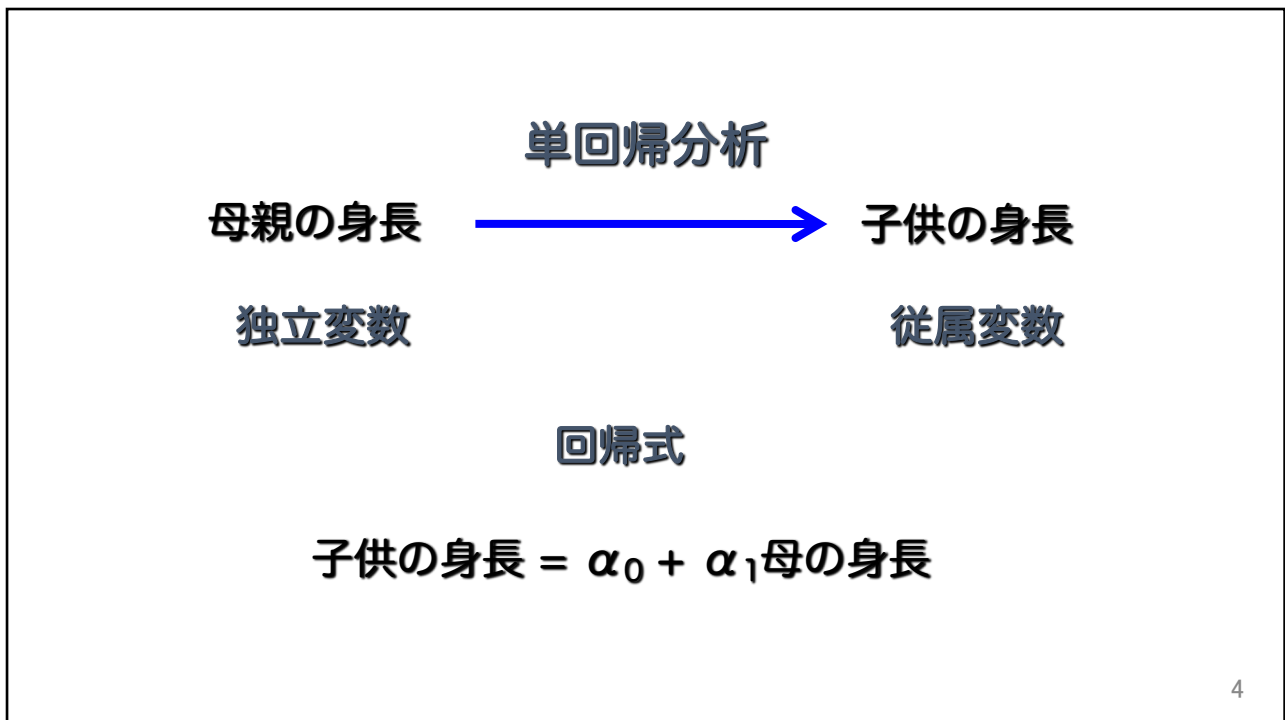
「背の高い母親の子供は背が高い」

2

2

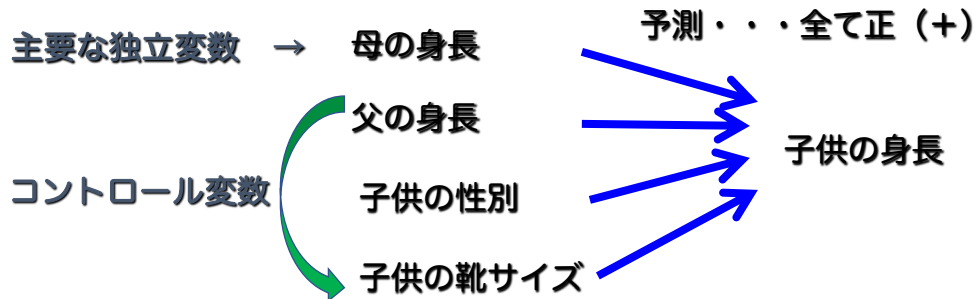


3



4

## 重回帰分析



## 重回帰式

$$\text{子供の身長} = \alpha_0 + \alpha_1 \text{母の身長} + \alpha_2 \text{父の身長} + \alpha_3 \text{子供の性別} + \alpha_4 \text{子供の靴サイズ}$$

5

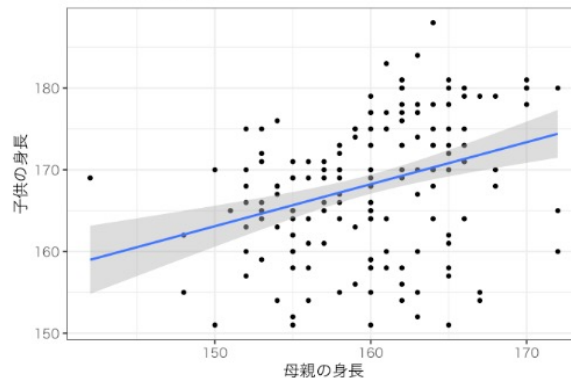
5

文字化けしないためのコマンド (マックユーザのみ)

```
theme_set(theme_bw(base_size = 14, base_family = "HiraKakuPro-W3"))

ggplot(df, aes(mom, height)) +
  geom_point() +
  labs(x = "母親の身長", y = "子供の身長",
       title = "母親と子供の身長") +
  stat_smooth(method = lm) # (method = lm, se = FALSE) → 95%
```

母親と子供の身長



6

6

```

```{r, results = "asis"}
stargazer(as.data.frame(df),
  type = "html",
  title = "親と子供の身長",
  digits = 2)
```

```

← チャンクオプション

親と子供の身長

| Statistic | N   | Mean   | St. Dev. | Min    | Pctl(25) | Pctl(75) | Max    |
|-----------|-----|--------|----------|--------|----------|----------|--------|
| height    | 169 | 168.22 | 8.17     | 151    | 163      | 175      | 188    |
| dad       | 167 | 171.83 | 6.05     | 140.00 | 168.00   | 176.00   | 185.00 |
| mom       | 169 | 159.98 | 5.29     | 142    | 156      | 164      | 172    |
| shoe      | 164 | 25.80  | 1.67     | 22.00  | 24.50    | 27.00    | 30.00  |
| parents   | 167 | 165.91 | 4.35     | 147.50 | 163.00   | 169.00   | 176.50 |

7

7

## 単回帰分析

```
model_1 <- lm(height ~ mom, data = df)
```

```
summary(model_1)
```

$$\text{height} = 86 + 0.5\text{mom}$$

```

Call:
lm(formula = height ~ mom, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-19.807  -4.723   1.193   7.249  17.707

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  85.9939    18.0437   4.766 4.07e-06 ***
mom          0.5140     0.1127   4.560 9.86e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 7.732 on 167 degrees of freedom
Multiple R-squared:  0.1107,    Adjusted R-squared:  0.1054
F-statistic: 20.79 on 1 and 167 DF,  p-value: 9.863e-06

```

サンプルでは  
母の身長が 1cm 高い  
→ 子供の身長が 0.5cm 高い

このことは母集団でもそうなのか?  
帰無仮説: 「母集団では係数は0」  
p値を確認

**0.0000098**

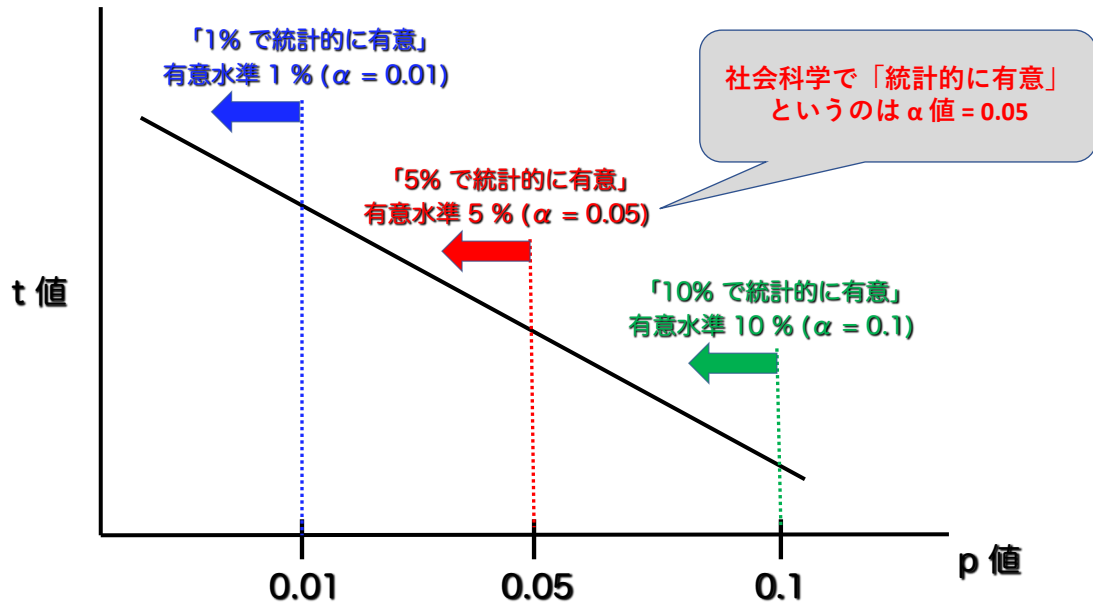
帰無仮説を棄却  
係数は0ではない

母親の身長は子供の身長と関係ある

8

8

## t 値・p 値・有意水準・統計的有意性の関係



9

## 重回帰分析

```
model_4 <- lm(height ~ mom + dad + male + shoe, data = df)
```

```
summary(model_4)
```

$$\text{height} = -3.4 + 0.42\text{mom} + \dots$$

Coefficients:

|             | Estimate | Std. Error | t value | Pr(> t )     |
|-------------|----------|------------|---------|--------------|
| (Intercept) | -3.36560 | 11.97926   | -0.259  | 0.796        |
| mom         | 0.42174  | 0.06277    | 6.719   | 3.19e-10 *** |
| dad         | 0.24454  | 0.05452    | 4.486   | 1.40e-05 *** |
| male男性      | 6.92812  | 1.05175    | 6.587   | 6.41e-10 *** |
| shoe        | 2.20870  | 0.29016    | 7.612   | 2.35e-12 *** |

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.125 on 157 degrees of freedom

(7 observations deleted due to missingness)

Multiple R-squared: 0.7545, Adjusted R-squared: 0.7482

F-statistic: 120.6 on 4 and 157 DF, p-value: < 2.2e-16

サンプルでは  
母の身長が 1cm 高い  
⇒ 子供の身長が 0.42cm 高い

このことは母集団でもそうなのか？

帰無仮説：「母集団では係数は0」

p値を確認

0.000000000319

帰無仮説を棄却

係数は0ではない

母親の身長は子供の身長と関係ある

10

10

最大の関心は mom なので  
どのような組み合わせでも、統計的に有意であることを示したい

mom が主要な独立変数

```
model_1 <- lm(height ~ mom, data = df)
model_2 <- lm(height ~ mom + dad, data = df)
model_3 <- lm(height ~ mom + dad + male, data = df)
model_4 <- lm(height ~ mom + dad + male + shoe, data = df)
```

```
```{r, results = "asis"}
stargazer(model_1, model_2, model_3, model_4,
  digits = 2,
  style = "ajps",
  title = "母親と子供の身長",
  type = "html")
```
```

11

11

母親と子供の身長

|                     | height                 |                        |                         |                         |
|---------------------|------------------------|------------------------|-------------------------|-------------------------|
|                     | Model 1                | Model 2                | Model 3                 | Model 4                 |
| mom                 | 0.51***<br>(0.11)      | 0.47***<br>(0.11)      | 0.51***<br>(0.07)       | 0.42***<br>(0.06)       |
| dad                 |                        | 0.27***<br>(0.10)      | 0.31***<br>(0.06)       | 0.24***<br>(0.05)       |
| male男性              |                        |                        | 12.96***<br>(0.83)      | 6.93***<br>(1.05)       |
| shoe                |                        |                        |                         | 2.21***<br>(0.29)       |
| Constant            | 85.99***<br>(18.04)    | 48.06**<br>(22.60)     | 24.17*<br>(14.38)       | -3.37<br>(12.98)        |
| N                   | 169                    | 167                    | 167                     | 162                     |
| R-squared           | 0.11                   | 0.15                   | 0.66                    | 0.75                    |
| Adj. R-squared      | 0.11                   | 0.14                   | 0.66                    | 0.75                    |
| Residual Std. Error | 7.73 (df = 167)        | 7.61 (df = 164)        | 4.82 (df = 163)         | 4.12 (df = 157)         |
| F Statistic         | 20.79*** (df = 1; 167) | 14.42*** (df = 2; 164) | 106.26*** (df = 3; 163) | 120.61*** (df = 4; 157) |

$p < .01$ ;  $p < .05$ ;  $p < .1$

どのモデルでも mom は安定して有意

主要な独立変数である mom は  
全てのモデルに投入

height の分散の 75% が  
Model 4によって説明できた

12

12

母親と子供の身長

|                     | height                  |
|---------------------|-------------------------|
| mom                 | 0.42***<br>(0.06)       |
| dad                 | 0.24***<br>(0.05)       |
| male男性              | 6.93***<br>(1.05)       |
| shoe                | 2.21***<br>(0.29)       |
| Constant            | -3.37<br>(12.98)        |
| N                   | 162                     |
| R-squared           | 0.75                    |
| Adj. R-squared      | 0.75                    |
| Residual Std. Error | 4.12 (df = 157)         |
| F Statistic         | 120.61*** (df = 4; 157) |

$p < .01$ ;  $p < .05$ ;  $p < .1$

( ) は標準誤差  
standard error

Model 4 の結果と解釈

$$\text{height} = -3.37 + 0.42\text{mom}^{***} + 0.24\text{dad}^{***} + 6.93\text{male}^{***} + 2.21\text{shoe}^{***}$$

解釈

- ・母親の身長が 1 cm 高い  
→ 子供身長が 0.42 cm 高い  
他の変数は全て平均値に固定

13

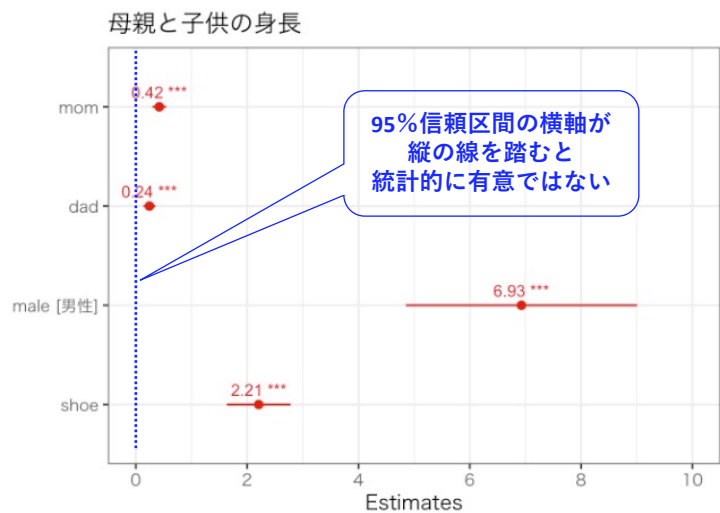
13

母親と子供の身長

重回帰分析の結果の可視化

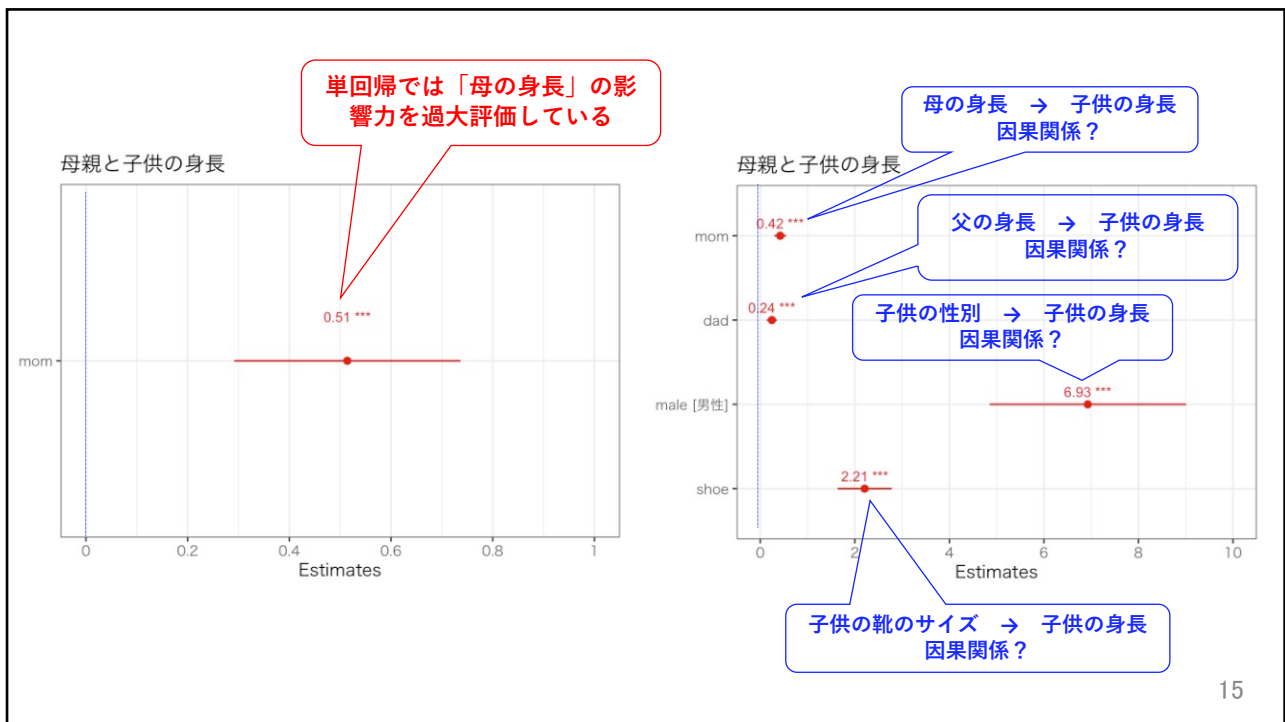
|                     | height                  |
|---------------------|-------------------------|
| mom                 | 0.42***<br>(0.06)       |
| dad                 | 0.24***<br>(0.05)       |
| male男性              | 6.93***<br>(1.05)       |
| shoe                | 2.21***<br>(0.29)       |
| Constant            | -3.37<br>(12.98)        |
| N                   | 162                     |
| R-squared           | 0.75                    |
| Adj. R-squared      | 0.75                    |
| Residual Std. Error | 4.12 (df = 157)         |
| F Statistic         | 120.61*** (df = 4; 157) |

$p < .01$ ;  $p < .05$ ;  $p < .1$



14

14



15

## Why Beauty Matters: Candidates' Facial Appearance and Electoral Success

**ONO, Yoshikuni**  
RIETI

**ASANO, Masahiko**  
Takushoku University

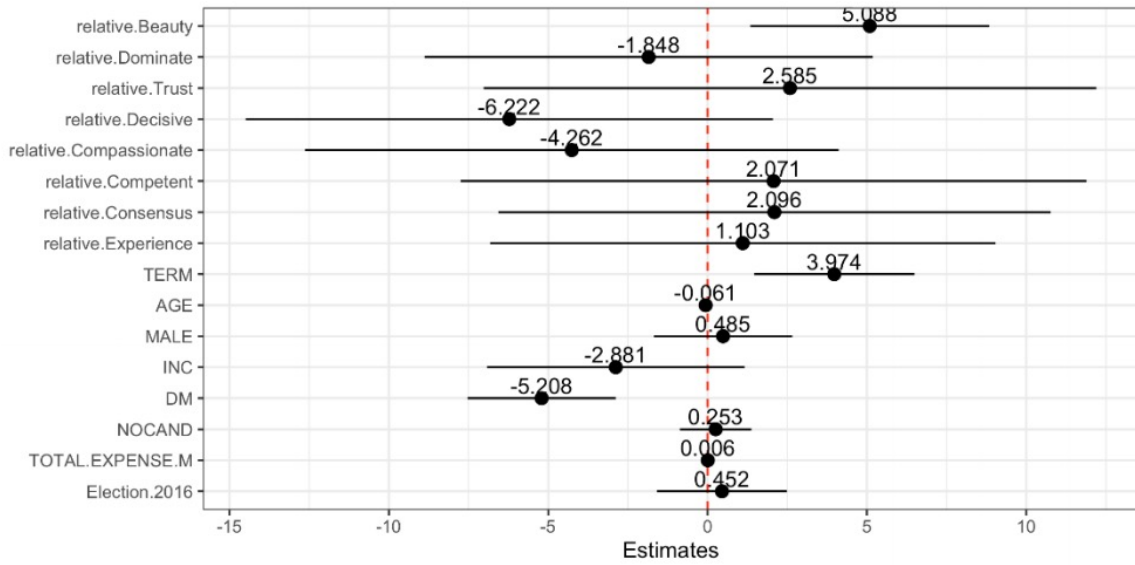
<https://www.rieti.go.jp/jp/publications/nts/20e072.html>

16

16



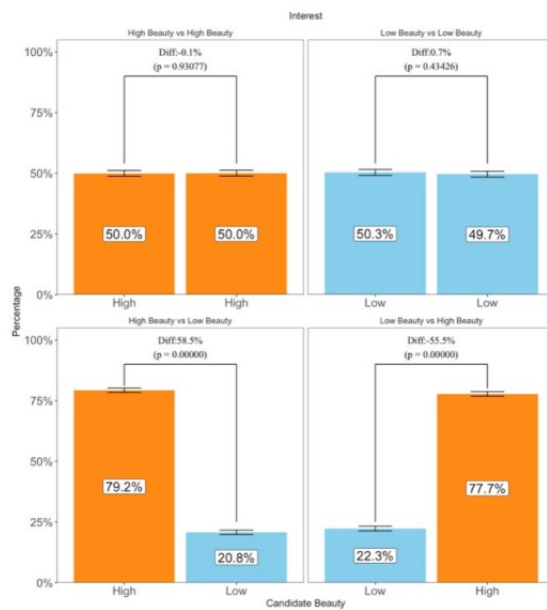
**Figure 4. Coefficient estimate plot (Model 3: Beauty and Impressions)**



17

17

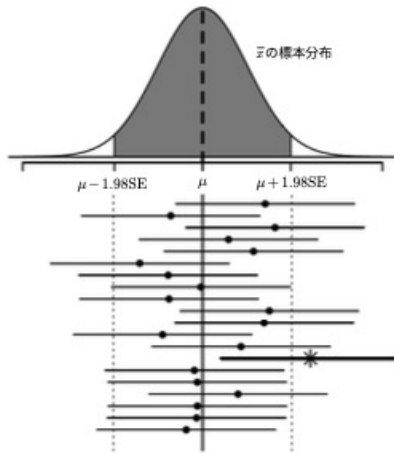
**Figure 6. Share of responses by conditions (interest)**



18

18

## 「統計的に有意」とはということか？



「95% 信頼区間」と聞くと、私たちが知ろうとしている母数の値がこの区間に含まれている確率が95%なのだろうと思うかもしれない。しかし、それは誤りである。あるデータセットから得られた 95% 信頼区間に母数が含まれる「確率」は、0% または 100% である。

浅野・矢内  
『Rによる計量政治学』pp.115-140

黒点が標本平均、水平方向に引かれた線分が信頼区間を表す。標本平均 (●) が  $\mu \pm 1.98SE$  の範囲に入るとき、信頼区間が垂直に引かれた実線と交わる。それらの信頼区間は母平均  $\mu$  を区間内に捉えている。運が悪いと、標本平均が  $\mu \pm 1.98SE$  の範囲外の値をとり (\*), 95% 信頼区間が母平均を捉え損なう (下から 7 番目の線分)。標本抽出を繰り返し、それぞれの標本について 95% 信頼区間を求めると、求めた信頼区間の 95% が母平均を区間内に捉える。

図 7.8 同じ母集団から抽出された 20 個の標本 ( $n = 100$ , すなわち  $t$  分布の自由度 99) から得られる 95% 信頼区間

19