

# 統計検定 (社会科学)

田淵進

November 12, 2025

## 第Ⅰ部 2021年

1

1.1

第  $h$  層の店舗の営業利益を  $X_{hi}, i = 1, \dots, N_i$  とおくと,

$$\begin{aligned}\mu &= \frac{1}{1000}(700 \times 100 + 200 \times 200 + 100 \times 500) = 160 \\ \sigma^2 &= \frac{1}{N} \sum_{h=1}^3 \sum_{i=1}^{N_h} (X_{hi} - \mu)^2 \\ &= \frac{1}{1000} \left[ \sum_h \sum_i (X_{hi} - \mu_h)^2 + \sum_h (\mu_h - \mu) \sum_i (X_{hi} - \mu_h) + \sum_h N_h (\mu_h - \mu)^2 \right] \\ &= 0.7 \times 10^2 + 0.2 \times 50^2 + 0.1 \times 100^2 \\ &\quad + 0 \\ &\quad + 0.7 \times 100^2 + 0.2 \times 200^2 + 0.1 \times 500^2 - 160^2 \\ &= 15970\end{aligned}$$

と計算できる。

1.2

条件から、有限母集団修正を考えないので、

$$\begin{aligned}E[\bar{X}] &= 160 \\ V[\bar{X}] &= 159.7\end{aligned}$$

である。

### 1.3

条件から、 $E[\hat{Y}] = 160$  である。分散を考えると、第  $h$  層の分散を  $\sigma_h^2$

$$V[\hat{Y}] = \sum_h \left( \frac{N_h}{N} \right)^2 \cdot \frac{\sigma_h^2}{n_h}$$

となる。比例配分では

$$\begin{aligned} V[\hat{Y}] &= \frac{0.7^2}{70} \times 10^2 + \frac{0.2^2}{20} \times 50^2 + \frac{0.1^2}{10} \times 100^2 \\ &= 15.7 \end{aligned}$$

である。問題文で与えられた条件からネイマン配分は

$$\begin{aligned} n_1 : n_2 : n_3 &= 700 \times 10 : 200 \times 50 : 100 \times 100 \\ &= 7 : 10 : 10 \end{aligned}$$

となるので、 $n_1 = 26, n_2 = n_3 = 37$  となる。このとき、

$$\begin{aligned} V[\hat{Y}] &= \frac{0.7^2}{26} \times 10^2 + \frac{0.2^2}{37} \times 50^2 + \frac{0.1^2}{37} \times 100^2 \\ &\approx 7.29 \end{aligned}$$

となる。

## 2

略

# 3

いま,  $J$  のランクは 1 で, 0 でない固有値は  $n$  である.  $J$  の固有値  $n$  のベクトル  $v$  に対して,  $Rv = (1 - \rho)Iv + \rho Jv = (1 + \rho(n - 1))v$  となり, 0 のベクトルに対しては  $Rv = (1 - \rho)v$  となる. 0 に対する固有ベクトルは  $n - 1$  の重複度を持つので,  $R$  の固有値は  $1 + \rho(n - 1)$  と  $1 - \rho$  の二つである. 以上から, 求める条件は

$$\frac{-1}{n-1} < \rho < 1$$

である.

## 3.1

前問を利用すると,  $1 - \rho$  に対応する固有空間の基底  $(v_1, \dots, v_{n-1})$  は  $J$  の固有値 0 に対応する固有空間の基底であることに注意すると,

$$\frac{1}{1-\rho}Iv_i - \frac{\rho}{(1-\rho)(1+(n-1)\rho)}Jv_i = \frac{1}{1-\rho}v_i$$

が成り立つ. また,  $1 + \rho(n - 1)$  に対応する固有ベクトル  $v_i$  は,  $J$  の固有値  $n$  の固有ベクトルであるので,

$$\frac{1}{1-\rho}Iv_n - \frac{\rho}{(1-\rho)(1+(n-1)\rho)}Jv_n = \frac{1}{1+(n-1)\rho}v_n$$

よって, 各固有ベクトル  $v_1, \dots, v_n$  について

$$R\left(\frac{1}{1-\rho}I - \frac{\rho}{(1-\rho)(1+(n-1)\rho)}J\right)v_n = v_n$$

であるので, 逆行列が示された.

## 3.2

定義から,

$$\begin{aligned} E[b] &= (X^T X)^{-1} X^T E[Y] = (X^T X)^{-1} X^T X b \\ &= b \end{aligned}$$

であり,  $\sum x_i = 0$  に注意すると,

$$\begin{aligned} X^T X &= \begin{pmatrix} n^2 & 0 \\ 0 & \sum x_i^2 \end{pmatrix} \\ (X^T X)^{-1} &= \begin{pmatrix} 1/n^2 & 0 \\ 0 & 1/(\sum x_i^2) \end{pmatrix} \\ X^T J X &= \begin{pmatrix} n^2 & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

となることに注意すると,

$$\begin{aligned}
V[b] &= (X^T X)^{-1} X^T V[Y] X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} X^T ((1 - \rho) I + \rho J) X (X^T X)^{-1} \\
&= \sigma^2 (1 - \rho) \begin{pmatrix} 1/n^2 & 0 \\ 0 & 1/(\sum x_i^2) \end{pmatrix} + \sigma^2 \rho \begin{pmatrix} 1/n^2 & 0 \\ 0 & 1/(\sum x_i^2) \end{pmatrix} \begin{pmatrix} n^2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/n^2 & 0 \\ 0 & 1/(\sum x_i^2) \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} \frac{1-\rho}{n} + \rho & 0 \\ 0 & \frac{1-\rho}{\sum x_i^2} \end{pmatrix}
\end{aligned}$$

と計算できる.

### 3.3

前問の結果から,  $\beta_1$  の推定は  $\rho$  が 1 に近いほど分散が小さくなる一方で,  $\beta_0$  の推定という観点からは,  $n \rightarrow \infty$  に置いても分散が 0 に収束しないため, 一致性がなくなる.

### 3.4

共変量  $u_i$  の仮定として, 観測範囲, 特に地域ごとに決まる確率変数  $u$  を取り,  $u_i \equiv U$  であるとすると, 最も単純な形でモデル化される. このとき前問までの推定は, 未知な  $u$  によって  $U = u$  と固定されたもとの推定

$$Y = (\beta_0 + u) + X\beta_1 + \varepsilon_i$$

に対応する. このとき,  $V[b_0]$  は, 本来の  $\beta_0$  の OLS の分散に加え, 本来評価すべき  $V[U]$  の加わった値になっていた.

$U$  を観測したうえで推定を行うと,  $V[U]$  の影響は取り除かれるので, OLS の精度が向上すると考えられる.

# 4

## 4.1

表から計算して， 23.6 となる.

## 4.2

詳細は省略.

$$y = \alpha + \beta_1 F_1 + \beta_2 F_2$$

と老いて， 一次方程式を溶け倍. このとき，  $\alpha = 24.7, \beta_1 = 4.0, \beta_2 = -0.5$  となる.

## 4.3

自由度調整済み決定係数  $R^2$  とは， 全変動  $S_T^2$ ， 回帰変動  $S_R^2$ ， 残差変動  $S_E^2$  を用いて， 決定係数

$$R^2 = \frac{S_R^2}{S_T^2} = 1 - \frac{S_E^2}{S_T^2}$$

に対して， 残差変動の自由度  $\phi_E$  と全変動の自由度  $\phi_T$  を用いて調整をかけたもので，

$$R^{*2} = 1 - \frac{S_E^2/\phi_E}{S_T^2/\phi_T}$$

とかける. よって表の値から

$$R^2 = \frac{409.50}{429.38} \approx 0.954$$
$$R^{*2} = 1 - \frac{19.88/47}{429.38/49} \approx 0.952$$

であり， 誤差分散は  $s^2 = 0.423$ . 標準誤差は  $s = \sqrt{0.423} = 0.650$  となる.

## 4.4

一般に， 回帰係数の推定値の標準誤差とは， OLS の分散  $V[\hat{\beta}]$  の  $\sigma^2$  を残差分散  $s^2$  で置き換えたものになる. 特に  $(X^T X)^{-1}$  の第  $J$  対角成分を  $C_j$  とおくと， 各係数の OLS の分散が  $\sigma^2 C_j$  となることに注意して，

$$\text{s.e.}(\hat{\alpha}) = \sqrt{s^2 C_1}$$
$$\text{s.e.}(\hat{\beta}_1) = \sqrt{s^2 C_2}$$
$$\text{s.e.}(\hat{\beta}_2) = \sqrt{s^2 C_3}$$

となる.

#### 4.5

前問で計算したように、各値は回帰変動と残差変動の値と自由度から求まる。ここで、二つのモデルの間で  $y_i$  に対する予測値  $\hat{y}_i$  が等しければ回帰変動  $\sum(y_i - \hat{y}_i)^2$  は同じ値が生じる。よって、残差変動も同じ値であり、自由度も変わらないため、 $R^2, s^2$  などは同じ値となる。

ここで、ダミー変数の取り方が変わっても、説明変数の行列から生成される空間は変わらないため、 $y_i$  の予測値としては同一なものが現れる。

## 5

略

## 第 II 部

# 2022 年

## 1

略

## 2

### 2.1

### 2.2

一般に、価格および数量ラスパイレス指数はそれぞれ

$$P_L = \frac{\sum_i^n p_{1i} n_{0i}}{\sum_i^n p_{0i} n_{0i}} \times 100$$
$$Q_L = \frac{\sum_i^n p_{0i} n_{1i}}{\sum_i^n p_{0i} n_{0i}} \times 100$$

で与えられ、パーシェ指数はそれぞれ

$$P_P = \frac{\sum_i^n p_{1i} n_{1i}}{\sum_i^n p_{0i} n_{1i}} \times 100$$
$$Q_P = \frac{\sum_i^n p_{0i} n_{0i}}{\sum_i^n p_{0i} n_{1i}} \times 100$$

で与えられる。よって表1からの計算によつて、ラスパイレス価格指数は

$$\begin{aligned} P_L &= \frac{10363 \times 43.12 + 2272 \times 118.17}{10363 \times 44.74 + 2272 \times 119.86} \times 100 \\ &= 0.9718 \times 100 = 97.18 \end{aligned}$$

となる。

## 2.3

前問の定義から

$$\begin{aligned} P_L &= \bar{x} \\ Q_L &= \bar{y} \end{aligned}$$

である。

## 2.4

いま、

$$\begin{aligned} s_{xy} &= \sum_{n=1}^n w_i(x_i - \bar{x})(y_i - \bar{y}) \\ &= \sum_{i=1}^n w_i x_i y_i - \bar{y} \sum_{i=1}^n w_i x_i - \bar{x} \sum_{i=1}^n w_i y_i + \bar{x} \bar{y} \sum_{i=1}^n w_i \\ &= \sum_{i=1}^n w_i x_i y_i - P_L Q_L \end{aligned}$$

とかける。ここで、

$$\begin{aligned} \sum_{i=1}^n w_i x_i y_i &= \frac{\sum_i^n p_{1i} n_{1i}}{\sum_i^n p_{0i} n_{0i}} \\ &= \frac{\sum_i^n p_{1i} n_{1i}}{\sum_i^n p_{0i} n_{1i}} \times \frac{\sum_i^n p_{0i} n_{1i}}{\sum_i^n p_{0i} n_{0i}} \\ &= P_P Q_L \end{aligned}$$

となるので、

$$\begin{aligned} s_{xy} &= P_P Q_L - P_L Q_L \\ &= Q_L (P_P - P_L) \end{aligned}$$

と示される。

## 2.5

$P_L$  と  $Q_L$  の大小関係は  $s_{xy}$  の符号と等しい。一般的に、価格と数量の間には負の相関がある場合が多いいため、 $P_P$  は  $P_L$  より大きくなる傾向がある。

# 3

## 3.1

$t = 1, T$  のみのデータを使った場合、モデルの LSE は

$$Y_t = \tilde{\alpha} + \tilde{\beta}X_t + \epsilon_t \quad (t = 1, T)$$

を満たすので、

$$\begin{aligned}\tilde{\beta} &= \frac{Y_T - Y_1}{X_T - X_1} \\ \tilde{\alpha} &= \frac{1}{2}(Y_1 + Y_T) - \tilde{\beta}\frac{1}{2}(X_1 + X_T) \\ &\qquad\qquad\qquad = \frac{X_1}{X_T - X_1}Y_1 - \frac{X_T}{X_T - X_1}Y_T\end{aligned}$$

となるので、

$$\begin{aligned}V[\tilde{\alpha}] &= \frac{X_1^2 + X_T^2}{(X_T - X_1)^2}\sigma^2 \\ V[\tilde{\beta}] &= \frac{2}{(X_T - X_1)^2}\sigma^2\end{aligned}$$

である。

## 3.2

いま、 $m = (X_1 + X_T)/2$  とおいて

$$\sum_{t=1}^T (X - \bar{X})^2 = \frac{(X_T - X_1)^2}{2} + 2(\bar{X} - m)^2 + \sum_{t=2}^{T-1} (X_t - \bar{X})^2$$

と変形できるので、条件は

$$\begin{aligned}\bar{X} &= m \\ X_2 &= \dots = X_{T-1} = m\end{aligned}$$

である。

## 3.3

前問の値を用いて、

$$\begin{aligned}V[\tilde{Y}_{T+1}] &= V[\tilde{\alpha} + \tilde{\beta}X_{T+1}] \\ &= V\left[\frac{X_{T+1} - X_T}{X_T - X_1}Y_1 + \frac{X_{T+1} - X_1}{X_T - X_1}Y_T\right] \\ &= \frac{(X_{T+1} - X_T)^2 + (X_1 - X_{T+1})^2}{(X_T - X_1)^2}\sigma^2\end{aligned}$$

### 3.4

モデルが正しいと仮定したとき、3点で測る際に最も  $V[\tilde{\beta}]$  を小さくするには、 $X_c$  を  $X_1$  または  $X_T$  と同じ値にするのが合理的である。

一方、単回帰であるかどうかを検証するためには、直線からのズレが大きくなる可能性が大きい  $X_c = (X_1 + X_T)/2$  が適当である。

# 4

## 4.1

## 4.2

$t$  期と  $t - 4$  期の関係を見て、 $A, B, C$  がそれぞれ  $a = 0.8, 0, -0.8$  である事がわかる。

## 4.3

AR モデルを MA モデルとして表現すると、

$$\begin{aligned} y_t &= ay_{t-4} + v_t \\ &= a(ay_{t-8} + v_{t-4}) + v_t \\ &= v_t + av_{t-4} + a^2v_{t-8} + a^3v_{t-12} + \dots \end{aligned}$$

となる。あるいは、恒等作用素を  $I$ 、ラグ作用素を  $L$  とおくと、 $|a| < 1$  のもとで  $I - L^4$  は可逆であり、

$$\begin{aligned} y_t &= (I - aL^4)^{-1}v_t \\ &= (I + aL^4 + a^2L^8 + \dots)v_t \end{aligned}$$

として、同じ結果を得る。

また、条件より  $y_t$  は定常であるので、 $\mu := E[y_t]$  とおくと、 $\mu = a\mu$  すなわち  $\mu = 0$  であり、 $\psi := V[y_t]$  とすると、

$$\begin{aligned} \psi &= a^2\psi + \sigma^2 \\ \psi &= \frac{\sigma^2}{1 - a^2} \end{aligned}$$

となる。また、共分散は

$$\begin{aligned} \text{Cov}(y_t, y_{t-k}) &= E\left[\left(\sum_{i=0}^{\infty} a^i v_{t-4i}\right)\left(\sum_{i=0}^{\infty} a^i v_{t-k-4i}\right)\right] \\ &= \begin{cases} \sum_{i=1}^{\infty} a^{2i+k/4} \sigma^2 & (k \in 4\mathbb{Z}) \\ 0 & (\text{otherwise}) \end{cases} \\ &= \begin{cases} \frac{a^{k/4} \sigma^2}{1 - a^2} & (k \in 4\mathbb{Z}) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned}$$

であるので、自己相関係数は、

$$\begin{aligned} \rho_k &= \text{Cov}(y_t, y_{t-k})/\psi \\ &= \begin{cases} a^{k/4} & (k \in 4\mathbb{Z}) \\ 0 & (\text{otherwise}) \end{cases} \end{aligned}$$

## 4.4

自己回帰モデルにおける SLE は、 $t = 5, \dots, T$  までの  $T - 4$  個の関係式

$$y_t = ay_{t-4} + v_t$$

から、平方和  $\sum_{t=5}^T (y_t - ay_{t-4})^2$  を最小化する値として求められる。これは通常の単回帰と同じで、SLE として

$$\hat{a} = \frac{\sum_{t=5}^T y_t y_{t-4}}{\sum_{t=5}^T y_{t-4}^2}$$

が求まる。このとき、残差の自由度は  $T - 5$  になるので、最小化された平方和から計算される分散  $\sigma^2$  の推定量としては、

$$\hat{\sigma}^2 = \frac{1}{T-5} \sum_{t=4}^T (y_t - \hat{a}y_{t-4})^2$$

と取れる。このとき、分母にも確率項があり、一般に推定量が不偏とは限らない。

一方で、大数の法則より<sup>\*1</sup>、

$$\frac{1}{T-4} \sum_{t=5}^T v_t y_{t-4} \rightarrow 0 \frac{1}{T-4} \sum_{t=5}^T y_{t-4} y_{t-4} \rightarrow a$$

が成り立ち、連続写像定理から  $\hat{a}$  は  $a$  に確率収束する。

同様に、 $e_t := y_t - \hat{a}y_{t-4} = v_t + (a - \hat{a})y_{t-4}$  として大数の法則を用いることで、一致推定量であることが示される。

## 4.5

$k = 4(n - 1) + i, i \in \{1, 2, 3, 4\}$  とおくと、

$$\hat{y}_{T+k} = \hat{a}^n y_{T-4+i}$$

と推定できる。このときの予測誤差は

$$y_{T+k} - \hat{y}_{T+k} = (a^n - \hat{a}^n)y_{T-4+i} + \sum_{j=0}^n a^j v_{T+k-4j}$$

となる。

---

<sup>\*1</sup> 大数の法則がちゃんと使えることを示すのはむずい気がする

# 5

## 5.1

分布の再生成より,  $D \sim N(0, 2\sigma^2)$  となるので,

$$\begin{aligned} E[D] &= 0 \\ V[D] &= 2\sigma_X^2 = 72 \\ P(D \leq -4) &= P(Z \leq -4/\sqrt{72}) && \approx 0.32 \end{aligned}$$

となる. ここで,  $Z$  は標準正規分布に従う確率変数とする.

## 5.2

条件より,  $Y | X \sim N(\alpha + \beta X, \sigma^2)$  となるので,  $\beta = 3/4, \alpha = 120 - 120\beta$  を用いて

$$\begin{aligned} E[Y | X = 132] &= \alpha + 132\beta \\ &= 129 \end{aligned}$$

となり,  $V[Y | X = 132] = \sigma^2 = \sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2 = 63$  である.

ここで, 平均への回帰効果とは, 母平均が同じであるとすれば, 一回目で 132 を記録しても, 2 回目の測定値の期待値は 132 よりも小さい値になることを指しており, 平均への回帰分は 3mmHg とみなされる.

## 5.3

Tower Property などから,

$$\begin{aligned} E[(X - \mu)^2] &= E[E[(X - \mu)^2 | \theta]] \\ &= E[E[(X - \theta)^2 + (\theta - \mu)^2 + 2(X - \mu)(X - \theta) | \theta]] \\ &= \psi^2 + \tau^2 \end{aligned}$$

と計算できる. 同様に  $V[Y] = \psi^2 + \tau^2$  であり,  $\text{Cov}(X, Y) = \tau^2$  と計算できる. これに  $\psi^2 + \tau^2 = 144, \tau^2 = 3(\psi^2 + \tau^2)/4$  を代入して,

$$\begin{aligned} \tau^2 &= 108 \\ \psi^2 &= 36 \end{aligned}$$

を得る.

## 5.4

条件から、 $\theta$  と  $\varepsilon$  はそれぞれ独立に正規分布に従うので、

$$\begin{pmatrix} \theta \\ X \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \theta \\ \varepsilon_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tau^2 & 0 \\ 0 & \psi^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\right)$$

となる。特に共分散行列は

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tau^2 & 0 \\ 0 & \psi^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \tau^2 + \psi^2 & \psi^2 \\ \psi^2 & \psi^2 \end{pmatrix}$$

である。このとき、 $\theta | X$  の分布は、 $X = 132$  とすると

$$\begin{aligned} E[\theta | X] &= \mu + \frac{\text{Cov}(\theta, X)}{V[X]}(X - \mu) \\ &= 129 \\ V[\theta | X] &= V[\theta] - \frac{\text{Cov}(\theta, X)^2}{V[X]} \\ &= 27 \end{aligned}$$

に従う正規分布となる。

## 5.5

$\varepsilon_1, \varepsilon_2$  は互いに独立であるので、 $\theta + \varepsilon_1 | X$  の分布は  $N(129, 63)$  となる。よって、

$$P(Y \leq 128 | X = 132) = P(Z \leq -1/\sqrt{63}) \approx 0.45$$

となる。ここで  $Z$  は標準正規分布に従う確率変数とする。

# 第 III 部

## 2023 年

1

1.1

いま、与えられた条件から

$$\bar{x} \approx 3.7327$$

$$\bar{y} \approx 27.49$$

$$\bar{x^2} \approx 16.15$$

$$\bar{y^2} \approx 855.27$$

$$\bar{xy} \approx 108.95$$

よって、 $n = 33$  であることから、不偏分散は

$$s_x^2 = \frac{33}{32}(\bar{x^2} - \bar{x}^2) \approx 2.33$$

$$s_y^2 = \frac{33}{32}(\bar{y^2} - \bar{y}^2) \approx \frac{33}{32}(855.27 - 27.49^2) \approx 102.68$$

$$s_{xy} = \frac{33}{32}(\bar{xy} - \bar{x}\bar{y}) \approx \frac{33}{32}(108.95 - 3.7327 \times 27.49) \approx 6.54$$

また、相関係数は

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x^2 s_y^2}} \approx \frac{6.54}{\sqrt{2.33 \times 102.68}} \approx 0.42$$

1.2

デルタ法とは、観測値の関数  $g(t_1, \dots, t_n)$  で表される統計量の分散を、 $g$  の各変数の平均値周りでの 1 次近似によって求める手法である。いま、

$$\begin{aligned} \hat{R} &= g(\bar{x}, \bar{y}) := \frac{\bar{y}}{\bar{x}} \\ &\approx g(\mu_X, \mu_X) + \frac{\partial g(\mu_X, \mu_X)}{\partial \mu_X}(\bar{x} - \mu_X) + \frac{\partial g(\mu_X, \mu_Y)}{\partial \mu_Y}(\bar{y} - \mu_Y) \\ &= \frac{\mu_Y}{\mu_X} - \frac{\mu_Y}{\mu_X^2}(\bar{x} - \mu_X) + \frac{1}{\mu_X}(\bar{y} - \mu_Y) \end{aligned}$$

と近似されるので、求める分散は

$$\begin{aligned} V[\hat{R}] &= \frac{1}{\mu_X^2} V[\bar{y} - R\bar{x}] \\ &= \frac{1}{n\mu_X^2} (E[(Y - RX)^2] - 0) \end{aligned}$$

となる。

### 1.3

前問の結果を用いて

$$\begin{aligned}
 V[\hat{\mu}_Y] &= \mu_X^2 V[\hat{R}] \\
 &= \frac{1}{n} (V[Y] + R^2 V[X] - 2R \text{Cov}(X, Y)) \\
 &= \frac{\mu_Y^2}{n} \left( \frac{\sigma_Y^2}{\mu_Y^2} + \frac{\sigma_X^2}{\mu_Y^2} \frac{\mu_Y^2}{\mu_X^2} - 2 \frac{\text{Cov}(X, Y)}{\mu_X \mu_Y} \right) \\
 &= \frac{\mu_Y^2}{n} (\text{cv}(Y)^2 + \text{cv}(X)^2 - 2\rho \text{cv}(X) \text{cv}(Y))
 \end{aligned}$$

と求まる。

### 1.4

$\hat{R} = \bar{y}/\mu_X$ としたときの分散は  $V_1 := \sigma_Y^2/\mu_X^2$  である。一方、 $\bar{y}/\bar{x}$ とした場合の分散は、前問での計算から

$$V[\hat{R}] := V_2 = \frac{1}{n\mu_X^2} (V[Y] + R^2 V[X] - 2R \text{Cov}(X, Y))$$

となるので、 $V_1 > V_2$ となるのは

$$R\sigma_X^2 < 2 \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

これを変形すると、

$$\frac{\text{cv}(X)}{2\text{cv}(Y)} < \rho$$

となるときである。よって、観測値から求まる  $\text{cv}(X), \text{cv}(Y), \rho$  を参考に、どちらを使うか切り替えるのが望ましい。

## 2

### 2.1

$X$  がパレート分布  $(a, b)$  に従うとき,

$$\begin{aligned} E[X] &= \int_n^\infty \frac{ab^a x}{x^{a+1}} dx \\ &= ab^a \left[ \frac{-1}{(a-1)x^{a-1}} \right]_b^\infty = \frac{ab}{a-1} \end{aligned}$$

となり, 累積分布関数は

$$\begin{aligned} F(x) &= \int_b^x \frac{ab^a}{t^{a+1}} dt \\ &= \left[ \frac{-b^a}{t^a} \right]_b^x \\ &= 1 - \left( \frac{b}{t} \right)^a \end{aligned}$$

となる. よって中央値は  $m = b \sqrt[a]{2}$  である.

### 2.2

$c > b$  として, 条件付き累積分布を考えると,

$$\begin{aligned} F(x \mid x > c) &= P(X \leq x \mid x > c) \\ &= \frac{P(c < X \leq x)}{P(X > c)} \\ &= \frac{(F(x) - F(c))}{F(c)} \\ &= \left( \frac{c}{b} \right)^a \left( \left( \frac{b}{c} \right)^a - \left( \frac{b}{x} \right)^a \right) \\ &= 1 - \left( \frac{c}{x} \right)^a \end{aligned}$$

となるので,  $X \mid X > c$  はパラメータ  $(a, c)$  のパレート分布に従う. よって, 下限で条件づけたパレート分布は再びパレート分布に従うことがわかる.

### 2.3

パレート分布における 80:20 の法則とは, 上位 2 割の領域における期待値が, 全体の期待値の 8 割を占める状態だと考えられる. すなわち,

$$P(X > x_0) = 0.2$$

となるような  $x_0$  に対して,

$$E[X, X > x_0] = \int_{x_0}^{\infty} \frac{ab^a}{x^a} dx = 0.8E[X]$$

が満たされれば良い. いま,  $b = 1$

$$P(X > x_0) = \left(\frac{1}{x_0}\right)^a = 0.2$$
$$E[X, X > x_0] = \left(\frac{1}{x_0}\right)^{a-1} = 0.8$$

なので,  $x_0 = 4$  であり,  $a = \log 5 / \log 4$

## 2.4

$b = 1, a = \log 5 / \log 4$  のとき, 上位 20% に入るには前問の結果から  $x = 4$ (百万) である. また, 上位 4% に入るには, 上位 20% の中に更に上位 20% のに入ればよいので, 前問の結果より,  $b = 4$  と取り直すことで  $x = 16$ (百万) が得られる.

# 3

## 3.1

条件より,

$$\begin{aligned} E[\exp(kU)] &= \frac{1}{\sqrt{2\pi\sigma_u^2}} \int_{\mathbb{R}} \exp\left(\frac{-u^2}{2\sigma_u^2} + ku\right) du \\ &= \frac{1}{\sqrt{2\pi\sigma_u^2}} \int_{\mathbb{R}} \exp\left(\frac{-u^2 + 2ku - k^2}{2\sigma_u^2} + \frac{k^2\sigma_u^2}{2}\right) du \\ &= \exp\left(\frac{k^2}{2}\sigma_u^2\right) \end{aligned}$$

## 3.2

$Y_t = e^{\alpha_0} Y_{t-1}^{\alpha_1} \exp(U_t)$  であることに注意すると,

$$\begin{aligned} E[Y_t | Y_{t-1}, \dots, Y_1] &= E[Y_t | Y_{t-1}] \\ &= e^{\alpha_0} Y_{t-1}^{\alpha_1} E[\exp(U_t)] \\ &= \exp(\alpha_0 + \sigma_u^2/2) Y_{t-1}^{\alpha_1} \end{aligned}$$

となるので,  $\alpha_0 = -\sigma^2/2, \alpha_1 = 1$  がマルチングールとなるための条件である.

## 3.3

$T$  までの値をもとに,  $Y_{T+1}$  の値を予測すると, 予測分散は,

$$\begin{aligned} V[\tilde{Y}_{T+1}] &= V[e^{\alpha_0} Y_T^{\alpha_1} \exp(U_{T+1})] \\ &= \exp(-\sigma^2) Y_T^2 (E[\exp(U_{T+1}^2)] - E[\exp(U_{T+1})]^2) \\ &= Y_T^2 (e^{\sigma_u^2} - 1) \end{aligned}$$

となる.

## 3.4

通常, 回帰分析では最小二乗法によって推定量  $\hat{\alpha}_1$  を求め, 中心極限定理より,  $\sqrt{\hat{\alpha}_1 - \alpha_1}$  が正規分布に近似できることを利用して検定を行う.

このとき,  $|\alpha_1| < 1$  においては, 極限分布は  $N(1, 1 - \alpha^2)$  となるが,  $\alpha_1 = 1$  においては正規分布に収束しないため, 通常の  $t$  検定で検定することはできない.

## 4

### 4.1

表から回帰係数を求めればいい。適当な二つの場合を持ってきて、男性の場合は

$$\begin{aligned} 7.2 &= a + 5b \\ 7.8 &= a + 6b \end{aligned}$$

となるので、 $a = 4.2, b = 0.6$  となり、女性の場合は

$$\begin{aligned} 6.6 &= a + 5b \\ 7.2 &= a + 6b \end{aligned}$$

となるので、 $a = 3.6, b = 0.6$  となる。よって、男性の方が同じ時間を働いたときの賃金が高いといえる。

### 4.2

いま、労働時間と賃金の平均値を考える。同じ労働時間での賃金の平均値は予測値と一致することに注意すると、

$$\begin{aligned} \bar{x}_M &= 5 \times 0.1 + 6 \times 0.15 + 7 \times 0.4 + 8 \times 0.35 = 7 \\ \bar{y}_M &= \hat{\bar{y}}_M = 7.2 \times 0.1 + 7.8 \times 0.15 + 8.4 \times 0.4 + 9.0 \times 0.35 = 8.4 \\ \bar{x}_F &= 5 \times 0.35 + 6 \times 0.4 + 7 \times 0.15 + 8 \times 0.1 = 6 \\ \bar{y}_F &= \hat{\bar{y}}_F = 6.6 \times 0.35 + 7.2 \times 0.4 + 7.8 \times 0.15 + 8.4 \times 0.1 = 7.2 \end{aligned}$$

よって  $y_M/x_M = y_F/x_F = 1.2$  となり、男女で等しくなる。

### 4.3

まず、 $y$  の分散を求める必要がある。いま、各データ  $(x_i, y_i)$  は

$$y_i = a + bx_i + u_i$$

と表されるとする。ここで、 $\hat{y}_i = a + bx_i$  において、全体のデータ数を  $N$  とすると、最小二乗法では

$$Ns_e^2 = \sum_{i=1}^N (y_i - \tilde{y}_i)^2$$

が最小化されるように  $a, b$  を決定する。ここで、 $x_i$  が固定されると  $\tilde{y}_i$  の値が決定されることに注意すると、

$$\begin{aligned} Ns_e^2 &= \sum (y_i - \bar{y} + \bar{y} - \tilde{y}_i)^2 \\ &= N(V[y] + V[\tilde{y}]) - 2 \sum_{j=5,\dots,8} (\tilde{y}_i - \bar{y}) \sum_{x_i \equiv j} (y_i - \bar{y}) \\ &= N(V[y] + V[\tilde{y}]) - 2N \cdot V[\tilde{y}] \\ &= N(V[y] - V[\tilde{y}]) \end{aligned}$$

となる。最小二乗推定を行った場合の  $s_e^2$  および  $V[\tilde{y}] = V[\hat{y}]$  は条件と与えられた表から求まり、

$$\begin{aligned} s_e^2 &= 0.972 \\ V[\tilde{y}_M^2] &= \bar{\tilde{y}}_M^2 - \bar{y}_M^2 = 0.324 \\ V[\tilde{y}_F^2] &= \bar{\tilde{y}}_F^2 - \bar{y}_F^2 = 0.324 \end{aligned}$$

となるので、 $V[y_M] = V[y_F] = 0.324 + 0.972 = 1.296$  と求まる。

また、共分散  $\text{Cov}(x, y)$  を考えると、 $b$  の最小二乗推定量が  $\hat{b} = \text{Cov } x, y / V[x]$  と置かれていたことに注意すると、

$$\begin{aligned} \text{Cov}(x_M, y_M) &= \hat{b}_M \cdot V[x_M] = 0.6(\bar{x}_M^2 - \bar{x}_M^2) = 0.54 \\ \text{Cov}(x_F, y_F) &= \hat{b}_F \cdot V[x_F] = 0.6(\bar{x}_F^2 - \bar{x}_F^2) = 0.54 \end{aligned}$$

と求まる。よって、最小二乗推定量を考えると、 $\hat{d} = \text{Cov}(x, y) / V[y]$  はともに  $5/12$  となる。一方で、

$$\begin{aligned} \hat{c}_M &= \bar{x}_M - \hat{d}\bar{y}_M = 3.5 \\ \hat{c}_F &= \bar{x}_F - \hat{d}\bar{y}_F = 3.0 \end{aligned}$$

となる。特に同じ賃金  $y = 7.2$  のもとでは、男性は 6.5 時間、女性は 6.0 時間の労働を行っている。

#### 4.4

前問で考察した通り。

#### 5

略。