

## 統計的分析方法

21世紀の情報革命によって、選手は様々な情報にアクセスできるようになった。これまでは一部の専門家しかアクセスできなかった情報に誰でも自由に、簡単にアクセスできるようになった意義は大きい。

例えばケニアのやり投げ選手であるジュリアス・イエゴ選手は *Youtube* でやり投げの勉強を行い、2015年の北京世界選手権で1位となった。これはこれまでの陸上界の常識からすれば革命的なことだった。

また、インターネットが普及したことで選手は多くのデータを手に入れて分析できるようになった。これまでは限定的なデータで経験的、主観的な考察しかできなかったものが、統計的分析によって客観的に合理的に説明できるようになった。これはつい最近のことで、まさに革命的な出来事だと言える。

今後、こうしたデータを使った研究が爆発的に世の中に広まることが予想される。これまで経済的な理由で知識にアクセスできなかった貧しい国の選手の技術力が大幅に向上するだろう。合わせて、世界規模の大会の記録水準もどんどん上がっていくと考えられる。

情報をうまく集め、分析していくことはこれからの選手や指導者にとって大切な作業になると思う。

第7章ではごくごく簡単に初歩的な統計的分析手法を説明する。これを読む選手や指導者が多くの情報を集めて分析し、新しい知識の発見することに役立ててほしい。

第7章では、特に説明がない場合は以下の表に従って記号を利用する。

記号	意味	記号	意味
$H_0$	帰無仮説	$\mu$	母平均
$H_1$	対立仮説	$\sigma^2$	母分散
$\alpha$	有意水準	$\sigma$	母標準偏差
$n$	サンプル数	$\bar{x}$	標本平均
$\phi$	自由度	$s^2$	標本分散
$F(\phi_1, \phi_2, \frac{\alpha}{2})$	F分布 ( $V_1 > V_2$ )	$s$	標本標準偏差
$t(\phi, \alpha)$	t分布	$\hat{\mu}$	母平均の不偏推定値
$p$	説明変数の個数	$\hat{\sigma}^2$	母分散の不偏推定値
$R$	相関係数	$\hat{\sigma}$	母標準偏差の不偏推定値
$R^2$	決定係数	$V$	不変分散 ( $\hat{\sigma}^2$ と同じ)
$E$	期待値	$S$	偏差平方和

表 7.1: 記号の意味

検定に用いる分布表を図 7.1~図 7.3 に示す。

分布表 (t-セント点)

	$\alpha$									
	0.5	0.4	0.3	0.2	0.1	0.05	0.02	0.01	0.001	
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619	
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.599	
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924	
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610	
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869	
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959	
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408	
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041	
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781	
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587	
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437	
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318	
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221	
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140	
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073	
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015	
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965	
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922	
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883	
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850	
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646	
31	0.682	0.853	1.054	1.309	1.696	2.040	2.453	2.744	3.633	
32	0.682	0.853	1.054	1.309	1.694	2.037	2.449	2.738	3.622	
33	0.682	0.853	1.053	1.308	1.692	2.035	2.445	2.733	3.611	
34	0.682	0.852	1.052	1.307	1.691	2.032	2.441	2.728	3.601	
35	0.682	0.852	1.052	1.306	1.690	2.030	2.438	2.724	3.591	
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496	

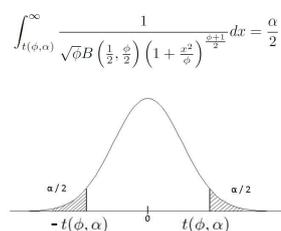


図 7.1: t 分布表 (パーセント点)

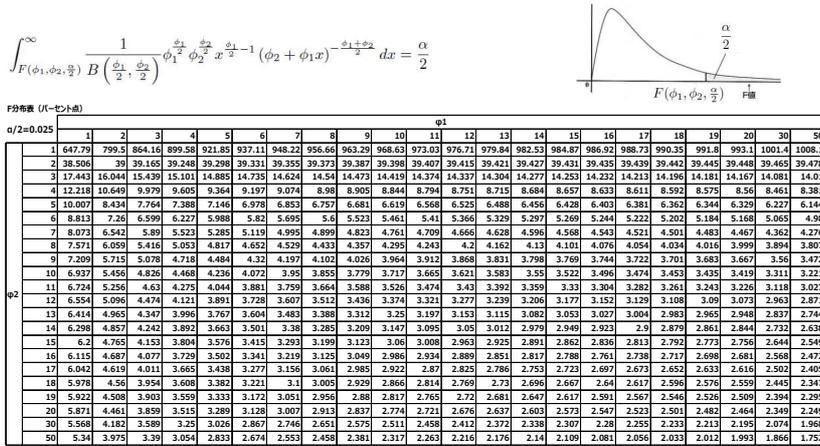


図 7.2:  $\alpha = 5\%$  の F 分布表 (パーセント点)

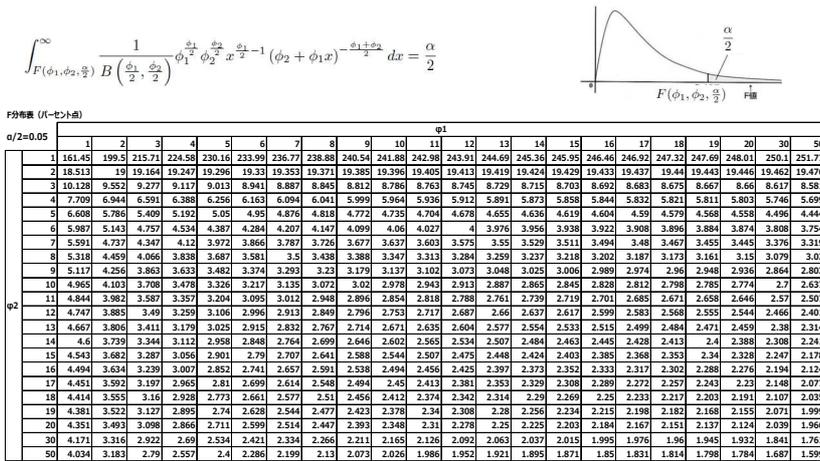


図 7.3:  $\alpha = 10\%$  の F 分布表 (パーセント点)

## 7.1 統計的検定

「スピードタイプの選手とパワータイプの選手の助走速度平均値に差があると言えるだろうか?」。こうした独立2群間の平均値の比較はスポーツ科学の分野ではよく行われる。こうした統計的仮説が正しいかどうかは統計的有意性の検定によって確かめる。

こうした検定は統計解析ソフトを使って行うことが多く、実際に手計算をして検定を行うという事はあまりない。本書の分析も統計解析ソフトを用いて独立2群間の平均値の比較を行っている。ここでは実際に統計解析ソフトで行われている計算手法を説明する(ここに紹介している手法は「2段階検定の多重性問題」から、近年はあまり利用されていない古い手法である。近年は等分散性の予備検定を行うことなく、直接ウェルチの  $t$  検定を用いて2群間の平均値の検定を行う手法が一般的であるが、統計解析ソフトの多くは古い手法がそのまま踏襲されている)。

2つのデータセット  $A, B$  の統計的検定は以下の2つの *STEP* で行う。

### STEP1 等分散の検定

2つのデータセット  $A, B$  の分散が等しい場合と異なる場合で  $t$  検定の方法が異なるため、まずはデータセットの分散を比較する  $F$  検定を行う。こうした検定は予備検定と呼ばれることもある。予備検定では通常は有意水準  $\alpha$  を10%や5%に設定して検定を行うことが多い。

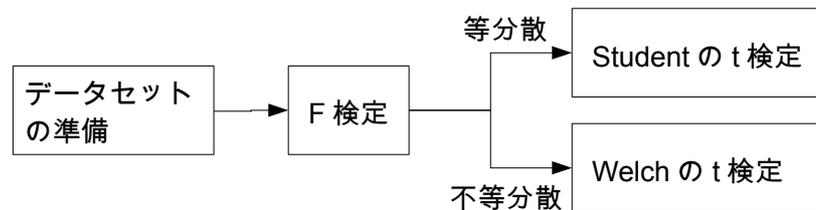


図 7.4: 検定の手順

$F$  検定ではデータセットの分散が等しいか等しくないかを検定したいので  
帰無仮説  $H_0$ : 2つのデータセットの分散に差がない  $\sigma_A^2 = \sigma_B^2$   
対立仮説  $H_1$ : 2つのデータセットの分散に差がある  $\sigma_A^2 \neq \sigma_B^2$   
とする。

$F$  検定の結果、帰無仮説を棄却できない場合はデータセットを等分散と仮定する。帰無仮説を棄却できる場合はデータセットは不等分散と仮定する。

$F$  検定では検定統計量  $F_0$  を以下のように計算する.

$$F_0 = \frac{V_1}{V_2} \quad (7.1)$$

ここで  $V_1$  はデータセット  $A, B$  の不変分散  $V_A, V_B$  のうちの大きいほうを表し,  $V_2$  はデータセット  $A, B$  の不変分散  $V_A, V_B$  のうちの小さいほうを表す. 不変分散  $V$  は  $x_i (i = 1, 2, \dots, n)$  のデータセットに対して, 以下の式で計算される.

$$V = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (7.2)$$

$F_0 < F(\phi_1, \phi_2, \frac{\alpha}{2})$  となれば帰無仮説  $H_0$  を棄却できないため,  $\sigma_A^2 = \sigma_B^2$  であり, データセットの分散は等しいと考える. また,  $F_0 \geq F(\phi_1, \phi_2, \frac{\alpha}{2})$  となれば帰無仮説  $H_0$  を棄却し,  $\sigma_A^2 \neq \sigma_B^2$  であり, データセットの分散は等しくないと考える. ここで  $\phi_S = n_S - 1$  はサンプル  $S$  の自由度を表している.

$F$  分布の確率密度関数は

$$f_{\phi_1, \phi_2}(x) = \frac{1}{B\left(\frac{\phi_1}{2}, \frac{\phi_2}{2}\right)} \phi_1^{\frac{\phi_1}{2}} \phi_2^{\frac{\phi_2}{2}} x^{\frac{\phi_1}{2}-1} (\phi_2 + \phi_1 x)^{-\frac{\phi_1+\phi_2}{2}} \quad (7.3)$$

で与えられる.

このとき  $n_S$  をサンプル  $S$  のデータ数を表し,  $B$  はベータ関数で

$$B(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx \quad (7.4)$$

となる.

$F(\phi_1, \phi_2, \frac{\alpha}{2})$  は最終的に以下の式を満たす値となる.

$$\int_{F(\phi_1, \phi_2, \frac{\alpha}{2})}^{\infty} \frac{1}{B\left(\frac{\phi_1}{2}, \frac{\phi_2}{2}\right)} \phi_1^{\frac{\phi_1}{2}} \phi_2^{\frac{\phi_2}{2}} x^{\frac{\phi_1}{2}-1} (\phi_2 + \phi_1 x)^{-\frac{\phi_1+\phi_2}{2}} dx = \frac{\alpha}{2} \quad (7.5)$$

通常,  $F(\phi_1, \phi_2, \frac{\alpha}{2})$  の値は  $F$  分布表を使って求める.

例えば図 7.5 に示す A 大学と B 大学の男子高跳び選手の記録に差があるかどうか検定する場合を考える。

A 大学	207	203	204	204	204	205	202
	205	205	202	203	205	204	202
	203	207	203				
B 大学	203	205	206	204	205	204	203
	207	205	206	207	205	205	205
	205	205	206				

図 7.5: 検定の例

A 大学のサンプルの不変分散は  $V_A = 2.38$ , B 大学のサンプルの不変分散は  $V_B = 1.31$  であり,  $V_A > V_B$  である. このときの検定統計量  $F_0$  は  $F_0 = \frac{V_A}{V_B} = 1.81$  となる. また, 自由度は  $\phi_A = 16$ ,  $\phi_B = 16$  である.

$\alpha = 5\%$  で  $F$  検定する場合は,  $F(\phi_A, \phi_B, \frac{\alpha}{2})$  (すなわち  $F(16, 16, 0.025)$ ) の値を図 7.2 を使って求める. 図 7.2 より  $F(16, 16, 0.025) = 2.761$  であり,  $F_0 < F(16, 16, 0.025)$  であるため, 帰無仮説  $H_0$  を棄却できない.  $F$  検定の結果から  $\sigma_A^2 = \sigma_B^2$  であり, データセットの分散は等しいと考える.

$F(\phi_A, \phi_B, \frac{\alpha}{2})$  の値が図 7.2 にない場合は, 自由度の逆数について線形補間を行って値を求める.

例えば  $\alpha = 5\%$  で,  $\phi_1 = 25$ ,  $\phi_2 = 30$  に対しては,  $\phi_1 = 20$  と  $\phi_1 = 30$  の間で補間して求める. 実際に計算すると

$$k = \frac{\left(\frac{1}{25} - \frac{1}{30}\right)}{\left(\frac{1}{20} - \frac{1}{30}\right)} = 0.4 \quad (7.6)$$

としたとき

$$F(25, 30, 0.025) = kF(20, 30, 0.025) + (1-k)F(30, 30, 0.025) = 2.122 \quad (7.7)$$

となる.

あるいは, *Excel* を使う場合は  $FINV(\frac{\alpha}{2}, \phi_A, \phi_B)$  を計算することで, 任意の自由度に対する  $F(\phi_A, \phi_B, \frac{\alpha}{2})$  の値を直接計算することができる.

## STEP2 $t$ 検定による平均値の比較

次に2つのデータセットの平均値に有意な差があるかどうかを検定するために  $t$  検定を行う。STEP1の結果によって  $t$  検定の手法は異なる。

2つの母平均の差を検定する場合は

帰無仮説  $H_0$ : 2つのデータセットの平均に差がない  $\mu_A = \mu_B$

対立仮説  $H_1$ : 2つのデータセットの平均に差がある  $\mu_A \neq \mu_B$

と設定して検定を行う。

STEP1の結果、等分散を仮定する場合は *Student* の  $t$  検定を行う。*Student* の  $t$  検定では検定統計量  $t_0$  を以下のように計算する [71]。

$$t_0 = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{S_A + S_B}{\phi_A + \phi_B} \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}} \quad (7.8)$$

ここで  $\bar{x}_S$  はサンプル  $S$  の平均、 $S_S$  はサンプル  $S$  の偏差平方和、 $n_S$  はサンプル  $S$  のデータ数、 $\phi_S = n_S - 1$  はサンプル  $S$  の自由度をそれぞれ表している。

$|t_0| < t(\phi_A + \phi_B, \alpha)$  となれば帰無仮説  $H_0$  を棄却できないため、 $\mu_A = \mu_B$  であり、データセットの平均は等しいと考える。また、 $|t_0| \geq t(\phi_A + \phi_B, \alpha)$  となれば帰無仮説  $H_0$  を棄却し、 $\mu_A \neq \mu_B$  であり、データセットの平均は等しくないと考え (有意差ありと考える)。

STEP1の結果、不等分散を仮定する場合は *Welch* の  $t$  検定を行う。*Welch* の  $t$  検定では検定統計量  $t_0$  を以下のように計算する [71]。 $V_S$  はサンプル  $S$  の不偏分散を表す。

$$t_0 = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\left( \frac{V_A}{n_A} + \frac{V_B}{n_B} \right)}} \quad (7.9)$$

$|t_0| < t(\phi^*, \alpha)$  となれば帰無仮説  $H_0$  を棄却できないため、 $\mu_A = \mu_B$  であり、データセットの平均は等しいと考える。また、 $|t_0| \geq t(\phi^*, \alpha)$  となれば帰無仮説  $H_0$  を棄却し、 $\mu_A \neq \mu_B$  であり、データセットの平均は等しくないと考え (有意差ありと考える)。

$\phi^*$  は以下のように計算する。

$$\phi^* = \frac{\left( \frac{V_A}{n_A} + \frac{V_B}{n_B} \right)^2}{\left( \frac{V_A}{n_A} \right)^2 \frac{1}{n_A - 1} + \left( \frac{V_B}{n_B} \right)^2 \frac{1}{n_B - 1}} \quad (7.10)$$

$t$  分布の確率密度関数は

$$f_{\phi}(x) = \frac{1}{\sqrt{\phi} B\left(\frac{1}{2}, \frac{\phi}{2}\right) \left(1 + \frac{x^2}{\phi}\right)^{\frac{\phi+1}{2}}} \quad (7.11)$$

で与えられる。自由度  $\phi$  は *Student* の  $t$  検定では  $\phi_A + \phi_B$  を使い、*Welch* の  $t$  検定では  $\phi^*$  を使う。B はベータ関数である。

$t(\phi, \alpha)$  は最終的に以下の式を満たす値となる。

$$\int_{t(\phi, \alpha)}^{\infty} \frac{1}{\sqrt{\phi} B\left(\frac{1}{2}, \frac{\phi}{2}\right) \left(1 + \frac{x^2}{\phi}\right)^{\frac{\phi+1}{2}}} dx = \frac{\alpha}{2} \quad (7.12)$$

通常、 $t(\phi, \alpha)$  の値は  $t$  分布表を使って求める。

例えば再び図 7.5 に示す A 大学と B 大学の男子高跳び選手の記録に差があるかどうか検定する場合を考える。F 検定の結果からデータセットの分散は等しいと考えられることから、*Student* の  $t$  検定を行う。

A 大学のサンプルの不変分散は  $V_A = 2.38$ 、B 大学のサンプルの不変分散は  $V_B = 1.31$  である。A 大学のサンプルの平均は  $\bar{x}_A = 204.00$ 、B 大学のサンプルの平均は  $\bar{x}_B = 205.06$  である。A 大学のサンプル数は  $n_A = 17$ 、B 大学のサンプル数は  $n_B = 17$  である。

このときの検定統計量  $t_0$  は  $t_0 = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{S_A + S_B}{\phi_A + \phi_B} \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = -2.27$  となる。

$\alpha = 5\%$  で  $t$  検定する場合は、 $t(\phi_A + \phi_B, \alpha)$  (すなわち  $t(32, 0.05)$ ) の値を図 7.1 を使って求める。あるいは、*Excel* を使う場合は  $TINV(\alpha, \phi_A + \phi_B)$  を計算することで、任意の自由度に対する  $t(\phi_A + \phi_B, \alpha)$  の値を直接計算することができる。図 7.1 より  $t(32, 0.05) = 2.037$  であり、 $|t_0| \geq t(32, 0.05)$  であるため、帰無仮説  $H_0$  を棄却する。 $t$  検定の結果から  $\mu_A \neq \mu_B$  であり、データセットの平均は等しくないと考える (有意差ありと考える)。

以上の検定結果より、「A 大学と B 大学の男子高跳び選手の記録には統計的に有意な差が認められる」と言える。

### 7.1.1 両側検定・片側検定

平均値の差の検定を行うときに「平均値が同じかどうか（差があるかどうか）」を確かめるときは両側検定を行い、「平均値が大きい（小さい）かどうか」を確かめるときには片側検定を行う。

ただし、片側検定の利用はその逆が起こることは理論的にありえない場合に限定される。こうした場合は通常ありえないので、一般的に検定では両側検定が利用される。本書でも当然ながら両側検定を用いて検定を実施している。

両側検定は片側検定よりも厳しい判定結果となるため、同じ分析対象の検定でも片側検定では有意差有りとなり、両側検定では有意差無しという結果になることがある。両側検定と片側検定の利用で迷った場合は、厳しい判定結果を出す両側検定を行うことが望ましい。

## 7.2 分散・標準偏差の考え方

標準偏差はデータの分布を示す指標の一つである。例えば正規分布の場合は平均値±標準偏差の範囲に全データの 68.27% が分布し、±標準偏差の 2 倍の範囲内に全データの 95.45% が分布する。

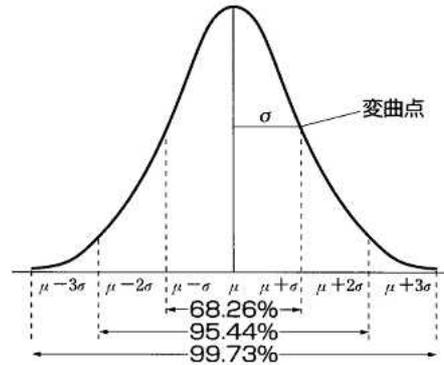


図 7.6: 正規分布と標準偏差

計測データの標本平均  $\bar{x}$ 、標本標準偏差  $s$  は以下のように計算される。

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7.13)$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.14)$$

しかしこれでは  $\sigma^2$  を母分散とした場合に、標本分散  $s^2$  は

$$E[s^2] = \frac{n-1}{n} \sigma^2 \quad (7.15)$$

となり、本来推定したかった母集団の母分散  $\sigma^2$  よりも標本分散  $s^2$  は  $\frac{n-1}{n}$  倍小さい値となる。

そこで本書では以下の式で計算される標準偏差の不偏推定値  $\hat{\sigma}$  と、分散の不偏推定値  $\hat{\sigma}^2$  を使って分析を行っている。

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.16)$$

このように  $\hat{\sigma}$  を定義すると

$$E[\hat{\sigma}^2] = \sigma^2 \quad (7.17)$$

となり、計測データから本来推定したかった母集団の母分散  $\sigma^2$  に近い値が得られる。科学論文などで利用される「標準偏差」や「分散」の値は、通常は  $\hat{\sigma}$  や  $\hat{\sigma}^2$  の値を使う。

### 7.3 統計データのグラフ表現

本書ではデータの標準偏差をエラーバー，検定結果を「\*」印でグラフ中に示している。「\*」は検定結果が5%有意であることを表し、「\*\*」は検定結果が1%有意であることを示している。

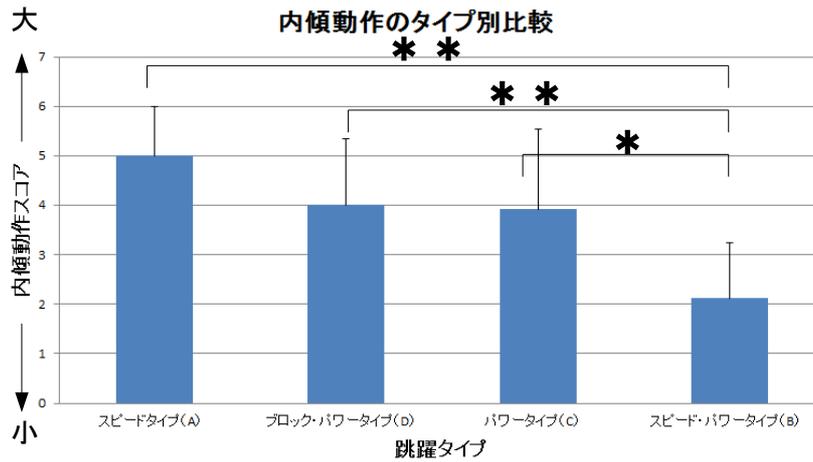


図 7.7: 統計データのグラフ表現

## 7.4 回帰分析

「助走速度と高跳びの記録の関係を式で表したい」。こうした式が得られれば、例えば220を跳ぶために必要な助走速度を予測したり、今の助走速度で跳べる高さの限界（潜在能力）を知ることができる。

このとき、「高跳びの記録」のように予測したい変数、注目している変数のことを目的変数（もしくは従属変数）と呼び、「助走速度」のようにその説明に用いる変数、要因となる変数のことを説明変数（もしくは独立変数）と呼ぶ。目的変数と説明変数の関係を表した関係式を回帰式と呼ぶ。

回帰式を求めることを回帰分析と言い、説明変数が1つの場合は単回帰分析、説明変数が2つ以上の場合には重回帰分析と呼ばれる。スポーツ科学の分野では単回帰分析は行われることがあるが、重回帰分析が行われることは稀である（重回帰分析を行えば、説明変数が目的変数に与える影響の大きさ、つまり説明変数の重要度のランキングを知ることができる。これから、ビッグデータの扱いが容易になれば、スポーツ科学の分野でも重回帰分析を使った分析が増えていくと考えられる）。

回帰式の作成はデータさえあれば可能である。実際の分析では統計解析ソフトを使って行うことが多く、実際に手計算をして検定を行うということはあまりない。ここでは実際に統計解析ソフトで行われている計算手法を説明する。

### 7.4.1 回帰式の作成方法

ここでは最もよく利用される、単回帰分析の方法について説明する。

目的変数を  $y$ ，説明変数を  $x$  とし，それぞれ  $n$  個のデータが測定されたとする。このとき， $i$  番目のデータの測定値を  $y_i$ ， $x_i$  で表し， $y_i$  を実績値と呼ぶ。最終的に求めたい  $x$  と  $y$  の関係式（回帰式）を

$$y = ax + b \quad (7.18)$$

とする。

$i$  番目の測定値  $x_i$  を使って回帰式  $y = ax + b$  で計算される  $y$  の値を  $\hat{y}_i$  と表現し， $\hat{y}_i$  を理論値と呼ぶ。 $\hat{y}_i$  は以下の式で表される。

$$\hat{y}_i = ax_i + b \quad (7.19)$$

実績値  $y_i$  と理論値  $\hat{y}_i$  の差を残差または誤差と呼び  $e_i$  で表す。

$$e_i = y_i - \hat{y}_i \quad (7.20)$$

である。残差は式  $y = ax + b$  の精度を表す。

回帰分析では全ての計測データに対する残差が最も小さくなる式  $y = ax + b$  を作る。具体的には残差の平方和  $S(a, b)$  を以下の式で定義するとき，残差平方和  $S(a, b)$  を最小にする  $(a, b)$  を求める。

$$S(a, b) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \{y_i - (ax_i + b)\}^2 \quad (7.21)$$

このような手法を最小 2 乗法と呼ぶ。

最小 2 乗法では残差平方和  $S(a, b)$  を最小にする  $(a, b)$  を求めるために， $S(a, b)$  を  $a$  と  $b$  で偏微分し，それが 0 となる  $(a, b)$  を求める。

$S(a, b)$  を  $b$  で偏微分すると

$$2nb - 2 \sum_{i=1}^n y_i + 2a \sum_{i=1}^n x_i = 0 \quad (7.22)$$

この式を  $b$  について整理すると

$$b = \frac{\sum_{i=1}^n y_i}{n} - a \frac{\sum_{i=1}^n x_i}{n} = \bar{y} - a\bar{x} \quad (7.23)$$

$S(a, b)$  を  $a$  で偏微分すると

$$2a \sum_{i=1}^n x_i^2 + 2b \sum_{i=1}^n x_i - 2 \sum_{i=1}^n x_i y_i = 0 \quad (7.24)$$

$b = \bar{y} - a\bar{x}$  を代入して， $a$  について整理すると

$$a = \frac{\frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x}\bar{y}}{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} \quad (7.25)$$

$a$  の式をもう少し整理する．分母を式変形すると

$$\begin{aligned}
 \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 &= \frac{\sum_{i=1}^n x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 \\
 &= \frac{\sum_{i=1}^n x_i^2}{n} - 2\bar{x} \frac{\sum_{i=1}^n x_i}{n} + \bar{x}^2 \\
 &= \frac{1}{n} \left( \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \right) \\
 &= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\
 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2
 \end{aligned} \tag{7.26}$$

分子を式変形すると

$$\begin{aligned}
 \frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x}\bar{y} &= \frac{\sum_{i=1}^n x_i y_i}{n} - 2\bar{x}\bar{y} + \bar{x}\bar{y} \\
 &= \frac{1}{n} \left( \sum_{i=1}^n x_i y_i - 2n\bar{x}\bar{y} + n\bar{x}\bar{y} \right) \\
 &= \frac{1}{n} \left( \sum_{i=1}^n x_i y_i - n\bar{y} \sum_{i=1}^n x_i - n\bar{x} \sum_{i=1}^n y_i + n\bar{x}\bar{y} \right) \\
 &= \frac{1}{n} \sum_{i=1}^n (x_i y_i - \bar{y}x_i - \bar{x}y_i + \bar{x}\bar{y}) \\
 &= \frac{1}{n} \sum_{i=1}^n \{(x_i - \bar{x})(y_i - \bar{y})\}
 \end{aligned} \tag{7.27}$$

従って最終的に  $a$  の値は以下の式で求めることができる．

$$a = \frac{\frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x}\bar{y}}{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} = \frac{\sum_{i=1}^n \{(x_i - \bar{x})(y_i - \bar{y})\}}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{7.28}$$

統計解析の教科書では  $x$  の標本分散を  $s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ ，共分散を  $s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$  として，回帰式の  $(a, b)$  の値を以下の式でシンプルに表現する．

$$a = \frac{s_{xy}}{s_x^2} \tag{7.29}$$

$$b = \bar{y} - a\bar{x} \tag{7.30}$$

$b$  の式より  $\bar{y} = a\bar{x} + b$  であるため，得られた回帰式は必ず  $(\bar{x}, \bar{y})$  を通る．

### 7.4.2 回帰式の評価

得られた回帰式の予測能力，説明能力は決定係数（寄与率と呼ばれることもある）で評価する．決定係数  $R^2$  は以下の式で計算する．式の第二項の分母は全変動と呼ばれ，データのばらつきの大きさを表す．第二項の分子は残差変動と呼ばれ，理論値と実績値のズレの大きさを表す．

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7.31)$$

決定係数  $R^2$  は 1 に近いほど，残差変動が全変動に比べて小さく，精度の高い予測式（回帰式）が得られたことを表している．

決定係数の値の良し悪しについては，統計的基準はないが，おおよそ図 7.8 の相場観で分析されることが多い．参考までに決定係数と分布の関係のイメージ図を図 7.9 に示す．

決定係数の値	精度
0.8以上	非常に良い
0.5以上～0.8未満	良い
0.25以上～0.5未満	あまり良くない
0.25未満	悪い

図 7.8: 決定係数と精度

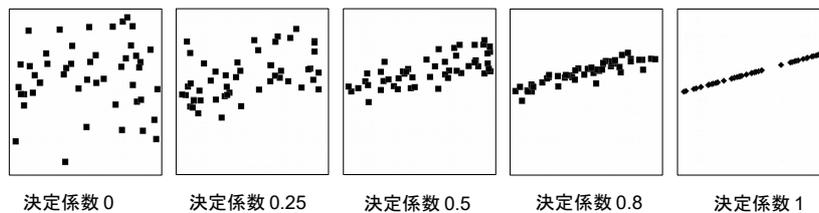


図 7.9: 決定係数とデータの分布

## 7.5 相関分析

回帰分析とは別に、2つの変数  $x$ ,  $y$  の間にある線形な関係の強弱（類似性の強弱）を測る指標として、相関係数  $r$  も分析ではよく用いられる。相関係数  $r$  は  $-1 \leq r \leq 1$  の値を取り、正のときには「正の相関を持つ」、負のときには「負の相関を持つ」と言う。また、 $r = 0$  のときは無相関であると言う。

相関係数は  $x$  の標本分散を  $s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ ,  $y$  の標本分散を  $s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ , 共分散を  $s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$  として以下の式で定義される。

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x s_y} \quad (7.32)$$

相関係数  $r$  はその絶対値が 1 に近いほど、2 変数間の相関が強い（直線的な関係性が強い、線形な関係性が強い）ことを表している。 $r$  の値が 1 に近い場合は「正の相関が強い」と表現し、 $r$  の値が  $-1$  に近い場合は「負の相関が強い」と表現する。

相関係数の値の良し悪しについては、統計的基準はないが、おおよそ図 7.10 の相場観で分析されることが多い。参考までに相関係数と分布の関係のイメージ図を図 7.11 に示す。

相関係数の絶対値	相関の強さの目安
$ r  < 0.3$	ほぼ相関なし
$0.3 \leq  r  < 0.5$	弱い相関あり
$0.5 \leq  r  < 0.7$	相関あり
$0.7 \leq  r  < 0.9$	強い相関あり
$0.9 \leq  r $	非常に強い相関あり

図 7.10: 相関係数と相関の強さ

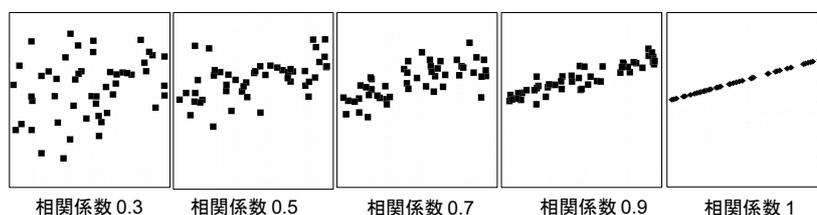


図 7.11: 相関係数とデータの分布

相関係数の値は必ず  $-1 \leq r \leq 1$  の範囲内となる。これはシュワルツの不等式  $(\sum_{i=1}^n a_i^2) (\sum_{i=1}^n b_i^2) \leq (\sum_{i=1}^n a_i b_i)$  に  $a_i = x_i - \bar{x}$ ,  $b_i = y_i - \bar{y}$  を代入して、式を整理することで簡単に証明できる。

### 7.5.1 決定係数 $R^2$ と相関係数 $r$ の関係

最小2乗法による直線フィッティング ( $y = ax + b$  による回帰) の場合は、決定係数  $R^2$  と相関係数  $r$  の間に  $R^2 = r^2$  の関係式が成り立つ (全く別々の意味で定義された統計量が偶然に一致する)。ここでは  $R^2 = r^2$  の関係を証明する。

$\sum_{i=1}^n (y_i - \bar{y})^2$  を測定値そのものの散らばり具合を表す指標として全変動  $TSS$  と呼ぶ。  $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$  を回帰式から計算される値の散らばり具合を表す指標として回帰変動  $ESS$  と呼ぶ。  $\sum_{i=1}^n (y_i - \hat{y}_i)^2$  を測定値と回帰式から計算される値のズレ具合を表す指標として残差変動  $RSS$  と呼ぶ。このとき  $R^2 = 1 - \frac{RSS}{TSS}$  となる。

$TSS$ ,  $ESS$ ,  $RSS$  の間には以下の関係式が成り立つ。

$$TSS = ESS + RSS \quad (7.33)$$

次に、この関係式を証明する。  $\alpha_i = \hat{y}_i - \bar{y}$ ,  $\beta_i = y_i - \hat{y}_i$  とすると

$$TSS - ESS - RSS = 2 \sum_{i=1}^n \alpha_i \beta_i \quad (7.34)$$

となる。ここで  $\hat{y}_i = ax_i + b$ ,  $b = \bar{y} - a\bar{x}$  を代入して  $\alpha_i \beta_i$  の式を整理すると

$$\begin{aligned} \alpha_i \beta_i &= (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) \\ &= (ax_i + b - \bar{y})(y_i - ax_i - b) \\ &= (ax_i - a\bar{x})(y_i - ax_i - \bar{y} + a\bar{x}) \\ &= a(x_i - \bar{x})(y_i - \bar{y}) - a^2(x_i - \bar{x})^2 \end{aligned} \quad (7.35)$$

従って、 $x$  の標本分散を  $s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ , 共分散を  $s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$  として、 $a = \frac{s_{xy}}{s_x^2}$  を代入して  $\sum_{i=1}^n \alpha_i \beta_i$  を整理すると

$$\begin{aligned} \sum_{i=1}^n \alpha_i \beta_i &= nas_{xy} - na^2 s_x^2 \\ &= na(s_{xy} - as_x^2) \\ &= 0 \end{aligned} \quad (7.36)$$

となる。よって  $TSS = ESS + RSS$  が示された。

$TSS = ESS + RSS$  のとき、 $R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$  となり、以下の式が成り立つ。

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7.37)$$

次に分子に  $\hat{y} = a\bar{x} + b$ 、分母に  $y$  の標本分散  $s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$  を使って、 $a = \frac{s_{xy}}{s_x^2}$  を代入して式を整理すると

$$\begin{aligned} R^2 &= \frac{1}{n s_y^2} \sum_{i=1}^n \{(ax_i + b) - (a\bar{x} + b)\}^2 \\ &= \frac{a^2}{n s_y^2} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{s_{xy}^2}{n s_y^2 s_x^4} n s_x^2 = \left( \frac{s_{xy}}{s_x s_y} \right)^2 = r^2 \end{aligned} \quad (7.38)$$

よって最小 2 乗法による直線フィッティング ( $y = ax + b$  による回帰) の場合は、決定係数  $R^2$  と相関係数  $r$  の間に  $R^2 = r^2$  の関係式が成り立つことが示された。

## 7.6 クラスタ分析

クラスタ分析とはデータ同士をその類似度に基づいていくつかのかたまり（クラスタ）に分類する手法である。適用分野は多岐にわたり、医学、生物学、考古学、心理学など対象をいくつかのグループに分けて分析する学問分野では幅広く利用されている。クラスタ分析は主に下記のステップで実施される。

1.  $n$  個データを全てグラフにプロットする。  
出発点は  $n$  個のクラスタと考える。
2. クラスタ間で最も類似度が高い（非類似度の低い）クラスタを一つのクラスタにまとめる
3. クラスタが1つになるまで上記の操作を繰り返す

クラスタ間の類似度はクラスタ内のデータ間の距離によって評価される。データ間の距離の評価にはユークリッド距離、ミンコフスキー距離、マハラノビス距離などが利用される。クラスタ同士の類似度の評価方法には最短距離法、最長距離法、郡平均法、重心法、ウォード法などがある。

例えば、ユークリッド距離を用いて最短距離法でクラスタ間の類似度を評価する場合は、2つのクラスタ内のデータ距離の最小値をクラスタ間の類似度とする。

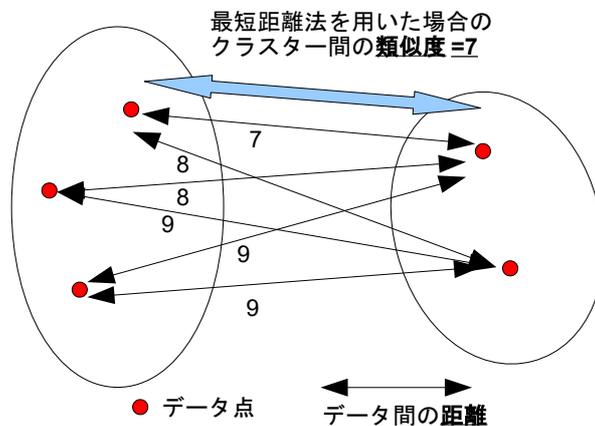


図 7.12: 距離と類似度

### 7.6.1 データ間の距離の計算方法

一般的に空間上の2点  $(x, y)$  の距離  $d(x, y)$  は以下の性質を持っている.

- 非負性

$$d(x, y) \geq 0 \quad (7.39)$$

- 非退化性

$$d(x, y) = 0 \Leftrightarrow x = y \quad (7.40)$$

- 対称性

$$d(x, y) = d(y, x) \quad (7.41)$$

- 三角不等式

$$d(x, z) \geq d(x, y) + d(y, z) \quad (7.42)$$

こうした性質を持つ  $d(x, y)$  は「距離」として考えることができる. クラスター分析では以下の距離の定義がよく用いられる.

- ユークリッド距離

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (7.43)$$

- ミンコフスキー距離

$$d(x, y) = \left( \sum_{k=1}^n |x_k - y_k|^r \right)^{\frac{1}{r}} \quad (7.44)$$

- マハラノビス距離

$$d(x, y) = \sqrt{(x - y)\sigma^{-1}(x - y)^T} \quad (7.45)$$

いずれも  $n$  はデータの次元数を表している.

ユークリッド距離は単純な幾何的距離を表しており最も一般的に利用される。ミンコフスキー距離はユークリッド距離を更に一般化したもので  $r$  が小さいほど次元軸方向を重視した距離になり、 $r$  が大きいほど次元軸にとらわれない方向を重視した距離となる。 $r = 1$  の場合はマンハッタン距離、 $r = 2$  の場合はユークリッド距離、 $r = \infty$  の場合はチェビシェフ距離となる。

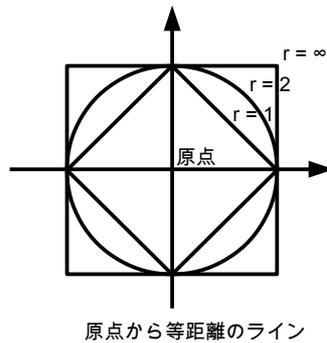


図 7.13: ミンコフスキー距離のイメージ

マハラノビス距離はデータに相関のある方向に平行な距離を相対的に短く、相関のある方向に垂直な距離を相対的に長くした距離になっている。 $\sigma$  は  $x$  と  $y$  の分散共分散行列を表している。

クラスター分析の距離として何を用いるかは分析者が「どんなデータを近いとみなし、どんなデータを遠いとみなすか」という方針を決めて選ぶ。本書の分析では最も一般的に利用されているユークリッド距離を用いてクラスター分析を行なっている。

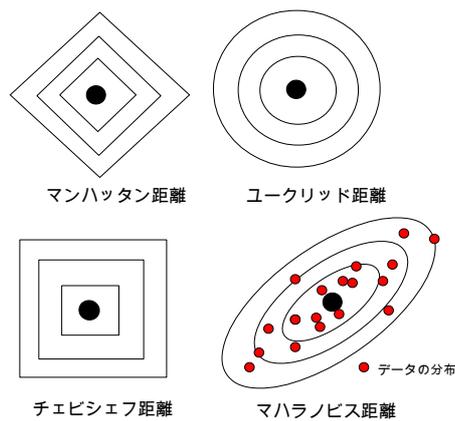


図 7.14: それぞれの距離の等距離線

## 7.6.2 クラスターの類似度の評価

クラスター間の類似度の評価には以下のような手法がある。

- 最短距離法

2つのクラスターのデータ間の距離の中で最小値をクラスターの類似度として定義する。データのちらばりが一方向に偏っている場合などに有効。

$$d(C_1, C_2) = \min_{x_1 \in C_1, x_2 \in C_2} d(x_1, x_2) \quad (7.46)$$

- 最長距離法

2つのクラスターのデータ間の距離の中で最大値をクラスターの類似度として定義する。データがいくつかの集団にかたまっている場合に有効。

$$d(C_1, C_2) = \max_{x_1 \in C_1, x_2 \in C_2} d(x_1, x_2) \quad (7.47)$$

- 郡平均法

クラスター内の全てのデータの組み合わせで距離を計算し、その平均をクラスターの類似度として定義する。データのちらばりが一方向に偏っている場合やいくつかの集団にかたまっている場合に有効。

$$d(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{x_1 \in C_1} \sum_{x_2 \in C_2} d(x_1, x_2) \quad (7.48)$$

- 重心法

ユークリッド距離を用いてクラスター内のデータの重心を計算し、クラスター間の重心距離をクラスターの類似度として定義する。データが同じ程度のデータ個数を持ついくつかの集団に分離できる場合に有効。

$$d(C_1, C_2) = \|M(C_1) - M(C_2)\|^2 \quad (7.49)$$

ただし  $M(C)$  はクラスター  $C$  の重心を表し  $M(C) = \sum_{x \in C} \frac{x}{|C|}$  で計算される。  $\|M(C_1) - M(C_2)\|^2$  はユークリッド距離の二乗を表す。

- ウォード法

クラスターの結合前後のクラスター内の変動増加量（重心まわりの偏差平方和）をクラスターの類似度として定義する。具体的には以下の数式で類似度を定義する。ウォード法では、クラスター内のデータからその重心までの距離を最小化するようにクラスターが結合されていくため、他の方法に比べて分類感度が高く、クラスター分析で利用されることが多い。

$$d(C_1, C_2) = E(C_1 \cup C_2) - E(C_1) - E(C_2) \quad (7.50)$$

ここで  $E(C) = \sum_{x \in C} \|x - M(C)\|^2$  で計算される。

どの類似度の評価手法を用いるかは、分析者が分析対象となるデータの性質を吟味し、どのようにクラスター分けを行うか方針を決めて選ぶ。本書の分析では最もバランスが取れ、一般的に広く利用されているウォード法を用いてクラスター分析を行なっている。

### 7.6.3 クラスター分析における順序尺度の扱い

クラスター分析は目的変数となるデータと説明変数となるデータの両方が量的データでなければ解析を行うことはできない。

アームアクションがダブルアームかシングルアームか、踏み切り位置が遠いか近いかといったデータは通常は質的データとして扱われる。こうした質的データはダミー変数と呼ばれる0または1をとる変数によって数量化が可能になり量的データとして扱うことができる。本書の分析ではアームアクションについてはダミー変数化することで量的データとして扱っている。

また、踏み切り位置が「非常に遠い」「遠い」「普通」「近い」「非常に近い」といった順序データは通常は質的データとして扱われる。しかし、順序尺度であっても4段階や5段階、7段階で評価され、測定される選択肢に等間隔性があると考えられる場合には量的データと同等の性質を持つため、厳密性を欠くが量的データとして扱って分析しても実用上十分な分析結果が得られる [72]。このため本書では多段階で評価された順序データは量的データとして扱い分析を行なっている。

本書では全てのデータを量的データに変換して、クラスター分析によって跳躍選手のグループ分けを行なっている。

## 7.6.4 デンドログラム

デンドログラムとはクラスター分析においてクラスターを結合する順番や、結合したクラスター同士の距離を樹形図で表現したものである。

類似度 ( 距離 )

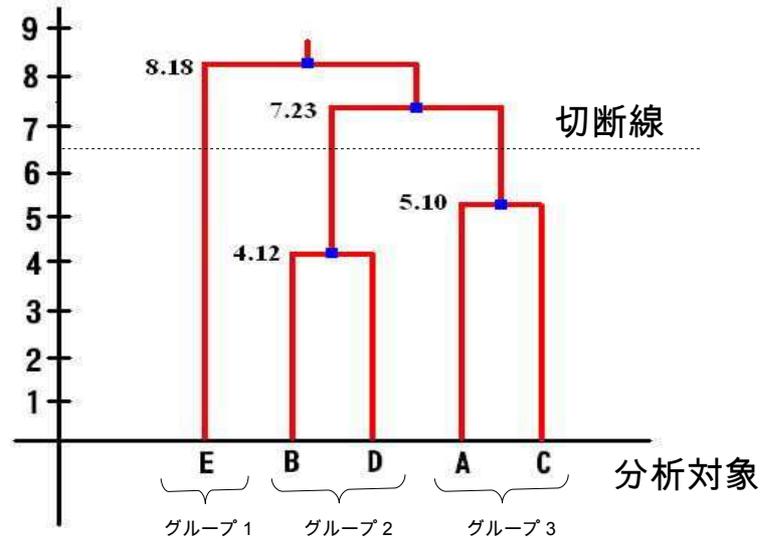


図 7.15: デンドログラム

デンドログラムの縦軸はクラスター分析を行う過程で結合したクラスター間の類似度 ( 距離 ) を表している。横軸は分析対象の各個体を表している。デンドログラムを見ることで各クラスターがどのように結合されていったかを知ることができる。

デンドログラムは縦軸を適当な高さで切ることによって、クラスターをいくつかのグループに分類することができる。デンドログラムを見て分かるように、どの高さで切断するかによってクラスターの分類結果は変化する。

最適なクラスター数を自動的に判定し、デンドログラムを自動的に切断するアルゴリズムは存在しないため、切断する高さは固有技術や、結果の説明し易さから分析者が判断して決める。本書の分析でも著者が分析対象となる選手の跳躍を丹念に観察することで、適切なクラスターの分類個数を判断している。

## 7.7 跳躍分類分析の詳細

ここでは第8章から第9章で、跳躍技術のクラスタリング分類に用いた各種スコア（順序データ）をどのように作り出したか詳細に説明する。

インターネットで公開されている動画の多くは、選手を撮影しているカメラの設置方法がバラバラであり、どの動画から選手の動作の正確な物理量を推定することは難しい。ここではこうした動画情報からどのようにして選手の動作を分析し、クラスタ分析に用いたスコア（順序データ）を作り出したか説明する。

### 7.7.1 助走の分析方法

#### 助走速度の分析方法

選手の基準ステップ長を  $L_{step}$  とする。ステップ長とは助走一歩分の長さであり、基準ステップ長は全選手の平均的なステップ長を表すものとする。

動画分析では目視確認でステップ長を「 $L$ （一歩の長さが短い）」「 $M$ （一歩の長さが普通）」「 $H$ （一歩の長さが長い）」の3グループに分類し、それぞれのステップ長を「 $L$  ( $0.97L_{step}$ )」「 $M$  ( $L_{step}$ )」「 $H$  ( $1.03L_{step}$ )」とした（図 7.16）。



図 7.16: ステップ長の判定例

また、動画のフレームレートから踏み切りを含む最後の3歩動作にかかった時間を計測し、上記のステップ長から各選手の助走速度指数を計算する（「助走速度＝助走速度指数×基準ステップ長」の関係式が成り立つ）。

分析では助走速度指数が3.48以上3.68未満の選手を助走速度スコア1（非常に遅い）に分類し、3.68以上3.88未満の選手を助走スコア2（遅い）、3.88以上4.08未満の選手を助走スコア3（普通）、4.08以上4.28未満の選手を助走スコア4（速い）、4.28以上4.48未満の選手を助走スコア5（非常に速い）、として五段階で評価した（図7.17）。

参考までに基準ステップ長を1.9mとした場合の各選手の推定助走速度も合わせて表に掲載しておく。

ランキン	名前	踏み切りを含む最後の3歩[sec]	ステップ長判定	ステップ長	助走速度指数(m/s/step)	助走速度スコア	ランキン	名前	基準ステップ長を1.9mとした場合の推定助走速度(m/s)
1	Bohdan Bondarenko	0.64	M	1	4.68	5	1	Bohdan Bondarenko	8.89
2	Mukar Essa Barshim	0.76	L	0.97	3.82	2	2	Mukar Essa Barshim	7.25
3	Derek Drouin	0.77	M	1	3.89	3	3	Derek Drouin	7.39
4	Erik Kynard	0.68	L	0.97	4.27	4	4	Erik Kynard	8.11
5	Ivan Ukhov	0.72	L	0.97	4.04	3	5	Ivan Ukhov	7.67
6	Konstadinos Baniotis	0.84	H	1.03	3.67	1	6	Konstadinos Baniotis	6.97
6	Yury Krymarenko	0.72	M	1	4.16	4	6	Yury Krymarenko	7.9
8	Yu Wang	0.76	M	1	3.94	3	8	Yu Wang	7.48
9	Donald Thomas	0.83	L	0.97	3.5	2	9	Donald Thomas	6.65
9	Aleksandr Shustov	0.88	H	1.03	3.51	1	9	Aleksandr Shustov	6.66
12	Michael Mason	0.73	L	0.97	3.98	3	12	Michael Mason	7.56
12	Jesse Williams	0.76	L	0.97	3.82	2	12	Jesse Williams	7.25
12	Dusty Jonas	0.83	L	0.97	3.5	2	12	Dusty Jonas	6.65
12	Silvano Chesani	0.76	L	0.97	3.82	2	12	Silvano Chesani	7.25
12	Robert Grabarz	0.68	M	1	3.57	1	12	Robert Grabarz	6.78
12	Sergey Mudrov	0.76	L	0.97	3.82	2	12	Sergey Mudrov	7.25
12	Michael Haranov	0.76	M	1	3.94	3	12	Michael Haranov	7.48
12	Mihai Donisan	0.72	L	0.97	4.04	3	12	Mihai Donisan	7.67
21	Aleksey Dmitrik	0.77	M	1	3.89	3	21	Aleksey Dmitrik	7.39
21	Daniil Tsyplov	0.72	L	0.97	4.04	3	21	Daniil Tsyplov	7.67
21	Ryan Ingraham	0.73	L	0.97	3.98	3	21	Ryan Ingraham	7.56
24	Kabele Kassemänge	0.72	L	0.97	4.04	3	24	Kabele Kassemänge	7.67
24	Liam Zamek-Pace	0.68	L	0.97	4.27	4	24	Liam Zamek-Pace	8.11
24	Suwel Zhané	0.72	M	1	4.16	4	24	Suwel Zhané	7.9
27	Keith Moffatt	0.8	L	0.97	3.63	1	27	Keith Moffatt	6.89
27	Edgar Rivera	0.67	M	1	4.47	5	27	Edgar Rivera	8.49
27	Tom Parsons	0.8	M	1	3.75	2	27	Tom Parsons	7.12
27	Diego Ferrin	0.8	H	1.03	3.66	1	27	Diego Ferrin	7.33
27	Piotr Skoboda	0.68	M	1	4.41	5	27	Piotr Skoboda	8.97
27	Douwe Amels	0.68	L	0.97	4.27	4	27	Douwe Amels	8.11
27	Roble Prezelj	0.68	M	1	4.41	5	27	Roble Prezelj	8.37
27	Raiyvdas Stanys	0.76	M	1	3.94	3	27	Raiyvdas Stanys	7.48
27	Nacoto Tobe	0.73	L	0.97	3.98	3	27	Nacoto Tobe	7.56
41	Marco Fassinotti	0.68	M	1	4.41	5	41	Marco Fassinotti	8.37
41	Jaroslav Baba	0.84	M	1	3.67	1	41	Jaroslav Baba	6.78
41	Takashi Eto	0.77	H	1.03	4.01	3	41	Takashi Eto	7.61
45	Jerron Robinson	0.7	L	0.97	4.15	4	45	Jerron Robinson	7.88
45	Adonios Mastoras	0.8	L	0.97	3.63	1	45	Adonios Mastoras	6.89

図 7.17: 助走速度スコアの分析

## 内傾動作の分析方法

動画分析では目視確認で6段階の内傾動作の評価を行った(図7.18)。また、分析対象となる選手の内傾が最も大きくなった瞬間の姿勢で評価を行った。

### 内傾が非常に小さい



Ryan Ingraham

Keith Moffatt

### 内傾がやや大きい



Bohdan Bondarenko

Tom Parsons

Jaroslav Bába

### 内傾が小さい



Michael Mason

Mutaz Essa Barshim

### 大きい



Jesse Williams

Edgar Rivera

Takashi Eto

### 内傾がやや小さい



Derek Drouin

Guowei Zhang

Nick Ross

### 非常に大きい



Ivan Ukhov

Konstadinos Banólis

Rožle Prezelj

図 7.18: 内傾動作の分析

## 助走歩数の分析方法

セット助走の選手については、助走開始から踏み切り動作までの歩数をカウントした。補助助走を用いる選手については補助助走から本助走に切り替わった後の助走歩数をカウントした。

## 7.7.2 踏み切り動作の分析方法

### アームアクション

踏み切り足が着地した瞬間に踏み切り足と逆側の腕を前方に残し、そのまま上昇姿勢、クリアランス動作へと移る選手はシングルアームに分類した。踏み切り動作の開始時点で両腕を引き、体幹部の後方に残している選手はダブルアームに分類した。

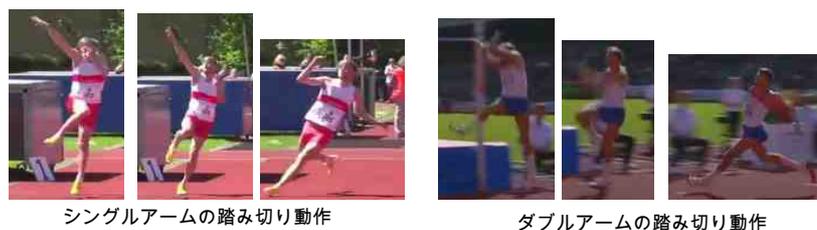


図 7.19: アームアクションの分類

### 踏み切り足接地方法

踏み切り足の接地方法を「踵接地」「土踏まず接地」「小指球接地」の3種類に分類した。映像分析では踵接地と土踏まず接地の判定が困難な場合があるが、判断に迷った場合は接地前の足運び動作でつま先を接地前に大きく持ち上げる選手は「踵接地」、足裏を地面に平行になるように素早く動かし接地動作を行っている選手は「土踏まず接地」に分類した。逆に接地前の足運び動作でつま先を下げた状態でそのまま接地している選手は「小指球接地」に分類した。



図 7.20: 踏み切り足接地方法の分類

## 踏み切り位置

踏み切り位置は動画に映るバーの高さを基準に推定している。踏み切り足が接地した瞬間の画像を用意し、支柱から踏み切り位置までの長さ（支柱からバーまでの長さ）と地面からバーまでの長さを比較することで踏み切り位置の距離を推定した。

分析ではなるべく選手の踏み切り位置を真横から捉えた画像のカットを利用し、支柱から接地足の中央（土踏まず）までの長さを踏み切り位置の距離として推定している。

分析では踏み切り位置が0.6m以上0.8m未満の選手を踏み切り位置スコア1（非常に近い）に分類し、0.8m以上1.0m未満の選手を踏み切り位置スコア2（近い）、1.0m以上1.2m未満の選手を踏み切り位置スコア3（普通）、1.2m以上1.4m未満の選手を踏み切り位置スコア4（遠い）、1.4m以上1.6m未満の選手を踏み切り位置スコア5（非常に遠い）、として五段階で評価した（図7.21）。

ランキング	名前	推定踏み切り位置[m]	踏み切り位置スコア
1	Bohdan Bondarenko	1.37	4
2	Mutaz Essa Barshim	1.1	3
3	Derek Drouin	1.17	3
4	Erik Kynard	1.5	5
5	Ivan Ukhov	0.91	2
6	Konstadinos Baniotis	1.2	4
6	Yuriy Krymareenko	1.07	3
8	Yu Wang	1.5	5
9	Donald Thomas	0.75	1
9	Aleksandr Shustov	1.3	4
12	Michael Mason	1.36	4
12	Jesse Williams	1.26	4
12	Dusty Jonas	1.47	5
12	Silvano Chesani	1.17	3
12	Robert Grabarz	0.72	1
12	Sergey Mudrov	1.27	4
12	Mickael Hanany	1.07	3
12	Mihai Donisan	1.5	5
21	Aleksey Dmitrik	1.44	5
21	Daniil Tsyplakov	0.78	1
21	Ryan Ingraham	1.01	3
24	Kabelo Kgosiemang	1.12	3
24	Liam Zamel-Paez	0.86	2
24	Guowei Zhang	1.08	3
27	Keith Moffatt	1.27	4
27	Edgar Rivera	1	3
27	Tom Parsons	1.1	3
27	Diego Ferrin	0.88	2
27	Piotr Sleboda	1.3	4
27	Douwe Amels	1.25	4
27	Rožle Prezelj	0.92	2
27	Raivydas Stanys	0.72	1
27	Naoto Tobe	0.77	1
41	Marco Fassinotti	0.93	2
41	Jaroslav Bába	1.1	3
41	Takashi Eto	1.15	3
45	Jerron Robinson	0.84	2
45	Adónios Mástoras	0.97	2

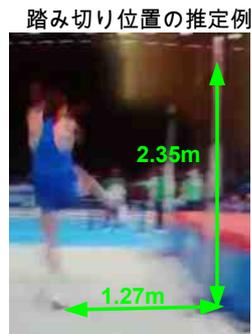


図 7.21: 踏み切り位置スコアの分析

## 振り上げ脚

動画分析では目視確認で5段階の振り上げ脚動作の評価を行った(図7.22)。5段階評価の主な評価基準は以下の通り。

- 評価1 (非常に小さい)  
踏み切り動作の最終局面で膝の高さが腰より低く、振り上げ脚動作が弱い
- 評価2 (小さい)  
踏み切り動作の最終局面で膝の高さが腰より低く、振り上げ脚動作は評価1の選手より明確に強い
- 評価3 (普通)  
踏み切り動作の最終局面で膝の高さが腰の高さ程度で、振り上げ脚動作が強い
- 評価4 (大きい)  
踏み切り動作の最終局面で膝の高さが腰の高さ程度で、振り上げ脚動作が非常に強く、高い膝位置をキープしたままクリアランス動作に移る
- 評価5 (非常に大きい)  
評価4の条件に加えて、振り上げ脚動作が評価4の選手より明確に強い

### 振り上げ脚動作が非常に小さい



Jesse Williams

Robert Grabarz

### 振り上げ脚動作が大きい



Derek Drouin

Ivan Ukhov

Ryan Ingraham

### 振り上げ脚動作が小さい



Bohdan Bondarenko

Donald Thomas

Erik Kynard

### 振り上げ脚動作が非常に大きい



Dusty Jonas

Liam Zamel-Paez

### 振り上げ脚動作が普通



Mutaz Essa Barshim

Aleksandr Shustov

Naoto Tobe

図 7.22: 振り上げ脚動作の分析

### 7.7.3 クリアランス動作の分析方法

動画分析では目視確認で5段階のクリアランス動作の評価を行った(図7.23)。また、分析対象となる選手のクリアランス中の反りが最も大きくなった瞬間の姿勢で評価を行った。5段階評価の主な評価基準は以下の通り。

- 評価1 (非常に小さい)  
ほとんど反っておらず、体幹より下の体のパーツが少ない
- 評価2 (小さい)  
小さく反っており、評価1の選手に比べて体幹より下の体のパーツが多い
- 評価3 (普通)  
平均的な体の反りでクリアランスを行っている
- 評価4 (大きい)  
評価3の選手に比べて明確に大きく反っている
- 評価5 (非常に大きい)  
評価4の選手に加えて、頭を下げ腰を大きく浮かした空中姿勢を作っている

クリアランスの反りが非常に小さい



クリアランスの反りが小さい



クリアランスの反りが普通



クリアランスの反りが大きい



クリアランスの反りが非常に大きい



図 7.23: クリアランス動作の分析

