

模擬探究 3

テゲバジャーロの戦い

— Battle of Tegevajaro —

ここでは、スポーツのデータをどうやって分析していくかという方法を学びましょう。スポーツでのデータ分析は、次の戦術を生み出し、そのスポーツの戦いを変えていく力を持っています。模擬探究3は「テゲバジャーロの戦い」。サッカーJ3のデータ分析をテキスト「理数探究基礎」を参考にしてやっていきます。

データ分析の方法

公開データの取得

その準備として、スポーツのデータの分析とはどういうことか、バスケットボールを例にして説明します。

バスケットの試合でシュートが外れた後のボールをとることをリバウンドと言います。バスケットのコーチは、いかにリバウンドの数を増やすかが勝敗に直結するといっています。これは本当でしょうか。いや、3ポイントシュートの成功率も大きく関わっているのではないかと、いろいろな意見があることでしょう。

バスケットの日本のプロリーグであるB1リーグでは、いろいろな試合データをHPに掲載しています。次のページにその一部を掲載しました。文字が小さくてよくわからないかもしれませんが、これは公開されているデータの一部です。

これらのデータから、先ほどのリバウンドと勝率の関係をグラフに表してみましょう。リバウンドには2種類あります。攻めているときに相手のゴールに入らなかったボールをとることを、オフENSリバウンド(OR)と言います。また相手が攻めてきてゴールに入らなかった後のボールをとることをディフェンスリバウンド(DR)と言います。

次のページのデータから、ORとDRを見つけ出します。ここでは分かりやすいように線で囲んであります。数字が並んでいるだけで、ちょっとわかりにくいですね。

それでも、どのチームもフェンスリバウンド(DR)のほう
が、オフェンスの時より 2~3 倍多いのが分かりますね。

2020-21		B1リーグ														
チーム名	チーム名	勝	負	勝率	G: 試合数	PTS: 総得点 数	PPG: 平均得 点	FGM: フィー ルドゴール成 功数	FGA: フィー ルドゴール試 投数	FG%: フィー ルドゴール成 功率	3FGM: 3P シュート成功 数	3FGA: 3P シュート試投 数	3FG%: 3P シュート成功 率	FTM: フリー スロー成功数	FTA: フリー スロー試投数	FT%: フリー スロー成功率
チーム	チーム				G	PTS	PPG	FGM	FGA	FG%	3FGM	3FGA	3FG%	FTM	FTA	FT%
1	富山グラウジーズ	39	21	0.650	60	5350	89.2	1866	4076	45.80%	486	1455	33.40%	862	1227	70.30%
2	サンロッカーズ渋谷	38	22	0.633	60	5080	84.7	1866	4076	45.80%	486	1455	33.40%	862	1227	70.30%
3	千葉ジェッツ	43	14	0.754	57	5073	89.0	1798	3762	47.80%	474	1358	34.90%	1003	1293	77.60%
4	川崎ブレイドサンダース	43	16	0.729	59	4931	83.6	1766	3744	47.20%	524	1554	33.70%	875	1199	73.00%
5	宇都宮ブレックス	49	11	0.817	60	4880	81.3	1827	3998	45.70%	476	1396	34.10%	750	1097	68.40%
6	島根スサノオマジック	28	32	0.467	60	4692	78.2	1769	3890	45.50%	382	1254	30.50%	772	1060	72.80%
7	シーホース三河	34	21	0.618	55	4663	84.8	1772	3594	49.30%	509	1337	38.10%	610	792	77.00%
8	アルバルク東京	32	24	0.571	56	4651	83.1	1707	3611	47.30%	488	1267	38.50%	749	990	75.70%
9	琉球ゴールデンキングス	40	16	0.714	56	4638	82.8	1706	3685	46.30%	476	1375	34.60%	750	1002	74.90%
10	滋賀レイクスターズ	23	36	0.390	59	4590	77.8	1681	3852	43.60%	538	1637	32.90%	690	968	71.30%
11	秋田ノーザンハピネッツ	28	31	0.475	59	4572	77.5	1691	3891	43.50%	440	1382	31.80%	750	1077	69.60%
12	大阪エヴェッサ	34	20	0.630	54	4561	84.5	1726	3574	48.30%	434	1170	37.10%	675	912	74.00%
13	名古屋ダイヤモンドドルフィン	32	24	0.571	56	4560	81.4	1631	3404	47.90%	520	1378	37.70%	778	1000	77.80%
14	京都ハンナリーズ	21	36	0.368	57	4499	78.9	1621	3518	46.10%	394	1110	35.50%	863	1121	77.00%
15	レバンガ北海道	14	45	0.237	59	4477	75.9	1622	3773	43.00%	488	1431	34.10%	745	970	76.80%
16	三遠(さんえん)ネオフェニックス	12	47	0.203	59	4466	75.7	1679	3846	43.70%	376	1181	31.80%	732	972	75.30%
17	横浜ビー・コルセアーズ	19	40	0.322	59	4320	73.2	1643	3658	44.90%	432	1335	32.40%	602	943	63.80%
18	広島ドラゴンフライ	9	46	0.164	55	4180	76	1539	3435	44.80%	398	1172	34.00%	704	939	75.00%
19	新潟アルビレックスBB	16	38	0.296	54	4131	76.5	1512	3327	45.40%	397	1205	32.90%	710	950	74.70%
20	信州ブレイドウォリアーズ	20	34	0.370	54	3881	71.9	1346	3232	41.60%	543	1697	32.00%	646	865	74.70%

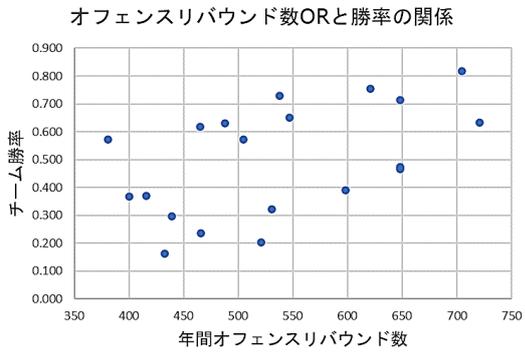
2020-21		B1リーグ														
チーム名	チーム名	OR: オフェ ンスリバウン ド数	DR: ディ フェンスリバ ウンド数	TR: リ バウンド差	トータル: リバウンド数	PPG: 平均リ バウンド獲得 数	AS: アンス ト数	APG: アンス ト数	TO: ターン オーバー数	ST: スティ ール数	BS: ブロウ ク数	ESR: 被ブ ロック数	F: ファウル 数	FD: 被ファ ウル数	EFF: 貢献度	MINPG: 平均 プレイ時間
チーム	チーム	OR	DR	TR												
1	富山グラウジーズ	547	1504	2051												201:00:00
2	サンロッカーズ渋谷	721	1266	1987												200:00:00
3	千葉ジェッツ	621	1495	2116												202:00:00
4	川崎ブレイドサンダース	538	1480	2018												200:00:00
5	宇都宮ブレックス	705	1449	2154												200:00:00
6	島根スサノオマジック	648	1447	2095												202:00:00
7	シーホース三河	465	1336	1801												201:00:00
8	アルバルク東京	505	1293	1798												202:00:00
9	琉球ゴールデンキングス	648	1412	2060												200:00:00
10	滋賀レイクスターズ	598	1345	1943												200:00:00
11	秋田ノーザンハピネッツ	648	1279	1927												201:00:00
12	大阪エヴェッサ	488	1437	1925												201:00:00
13	名古屋ダイヤモンドドルフィン	381	1464	1845												202:00:00
14	京都ハンナリーズ	400	1386	1786												200:00:00
15	レバンガ北海道	466	1298	1764												200:00:00
16	三遠(さんえん)ネオフェニックス	521	1302	1823												200:00:00
17	横浜ビー・コルセアーズ	531	1359	1890												200:00:00
18	広島ドラゴンフライ	433	1248	1681												200:00:00
19	新潟アルビレックスBB	439	1330	1769												201:00:00
20	信州ブレイドウォリアーズ	416	1310	1726												201:00:00



散布図を描く

それではこのデータ表から、年間オフェンスリバウンド数 OR とチーム勝率を各チーム毎にプロット（点を打つこと）してみたのが右の図です。このようなグラフを散布図と呼んでいます。このとき一方の値が増えるにつれて、もう一方の値も直線的に増える傾向があることを正の相関関係があるといえます。

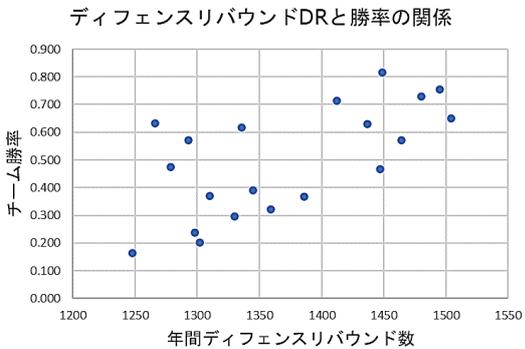
右の図では確かに年間オフェンスリバウンド数が多ければチームの勝率も上がっているような傾向が分かります。しかし、全体的に点が広がっており、強い相関関係があるわけではなさそうです。



【グラフ A】

続けて、同じデータ表から年間ディフェンスリバウンド数 DR とチームの勝率との関係を見てみましょう。なんとなく、ディフェンスリバウンドの方が、相関関係が強いようにも思いますし、あまり変わらないと言えば変わりませんね。

バスケットでは、オフェンスとディフェンスのどっちのリバウンドの方が重要なのでしょう。ここで、テキスト「理数探究基礎」の P54 を見てみましょう。「相関関係の強さを見た目以外で判断できないかな？」と考えている仲間の絵が左下に書かれていますね。確かに、そんな方法があると便利そうです。



【グラフ B】

相関関係の強さを表す相関係数

テキストの P55 を見てください。そこに相関関係の強さを数値で表すのに相関係数という値 r があることが紹介されています。この相関係数 r は

$r=0$ 相関関係はなし

$r=1$ 強い相関関係がある

ということを示しています。

通常データでは $-1 < r < 1$ の間をとり、負の相関の場合は右下下がりになります。この相関係数 r の計算式はテキスト P55 にあり、次のように書かれてあります。

$$r = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2\} \{(y_1 - \bar{y})^2 + \dots + (y_n - \bar{y})^2\}}}$$

n : データ値の個数 \bar{x}, \bar{y} : x, y の平均値

$(x_1, y_1), \dots, (x_n, y_n)$: データの値の組

皆さん解読できますか。ちょっと難しいですね。この式は、数学の分野の一つ「統計学」から導かれた式です。ここでは、まずこの式を使って計算し、相関係数を出すということを目指しましょう。なぜこんな式が導かれたのか

は、高校の数学で統計学の基礎を学び、大学で「データサイエンティスト」を目指す人や、医療や社会科学、生物学等のデータ分析をするようになり、改めて統計学を基礎から勉強するようになると思われる。

とはいっても大まかな説明だけしておきますね。

まずデータが n 個あるとします。データは例えば $n = 1$ のデータを (x_1, y_1) という組で表すとします。また (\bar{x}, \bar{y}) は x, y の平均値の組を表しているとしたら

$x_1 - \bar{x}$ というのは $n = 1$ のデータの x 座標 x_1 からデータ全体の x 成分の平均値 \bar{x} が引いてあります。 $y_1 - \bar{y}$ については、 y について同様なことが言えますね。このとき

$$(x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})$$

の全体を n で割った形は**共分散**と言われる量で、各データの平均座標値からの各成分の差をすべて足し合わせたものになっています。これは正になったり負になったりします。

次に分母の方を見てみましょう。

$\sqrt{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}$ の部分はルートを外し全体を n で割った形が「**分散**」と言われる量でそのルートをとると「**標準偏差**」と言われるデータの真の値からの散らばり具合を表す量になります。 y についても同様なことが言えます。

つまり式の形としては

$$\text{相関係数 } r = \frac{\text{共分散}}{(\text{xの標準偏差}) \times (\text{yの標準偏差})}$$

という形になっております。

ともかくこの計算はパソコンのエクセルといった表計算ソフトはある命令であつという間に計算してくれますが、この模擬探究ではあえてアナログでグループで協力してやっていくことにしましょう。

模擬探究のテゲバジャーロの分析のところでは、数表をつくり、この計算に電卓を持って、みんなで作戦を立てて挑むことにしましょう。それが競技と言ってもいいですね。

話をバスケットの方に戻すと、実は上のグラフ **A** の相関係数 r を上の式を使って計算してみると

$$r = 0.51$$

となります。そしてディフェンスリバウンドのグラフ **B** の相関係数 r は

$$r = 0.65$$

となりました。

つまり、相関係数は r が位置に近いほど相関が強いわけですから、どちらかというとディフェンスリバウンドつまり守っているとき、相手のシュートミスをとるリバ

ウンドの方が勝率を上げるのには有利であることが分かるのです。でも、このデータ分析はまだまだデータが足らず、たまたまなのかもしれません。つまり、ディフェンスリバウンドの方がより重要といっても、オフェンスリバウンドとそう確定的な差があるわけではありません。君たちが高校で探究活動をする上で、このレベルまででディフェンスリバウンドの有効性を主張しても、誰も耳を傾けてはくれないでしょう。

では、どうしたらいいのでしょうか。決め手は**仮説検定**という数学の論理です（テキスト P141～147）。

仮説検定とは

仮説検定という言葉は、数学の統計分野からとった言葉です。分かりやすく言うと数学的根拠を持ってそれはどのぐらいの割合で正しいのかを**推定する方法**です。

ここではいきなり P146 を見てもらいましょう。ここに**無相関検定**という方法が紹介されています。テキストの図 29 を見てください。先ほどのバスケットのディフェンスリバウンドの数とチームの勝率の関係は、たまたまこの年だけ正の相関が出てきただけで、「本当は相関関係なんてない」のかもしれません。そこで「相関関係はない」という仮説を立てて調べてみるという方法が**無相関検定**と言われるものです。

もし相関関係がなかったら、統計的な標本の分布はテキストの無相関検定の分布に従うことが分かっています。

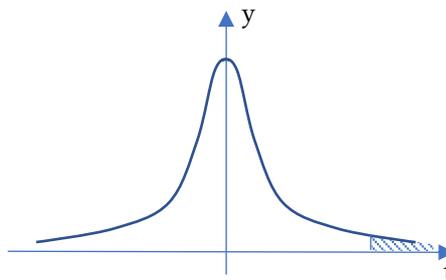
このときの検定統計量 t は

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

と表わされることが分かっています。つまりこの式を使えば、相関係数が $r = 0.65$ のときの t の値が計算できます。 n はデータの個数です。

t って突然出てきたけどナニモノ?! これは何の役に立つのかって? みんな落ち着いて説明を聞いてください。それは、相関がない ($r = 0$) としたときの t の分布と確率密度関数 $y = f(t)$ といわれる曲線がすでに求められており、このグラフを使うと、曲線と t 軸で囲まれた面積で相関がない ($r = 0$) とする確率が求まるわけなんです。このグラフを「 t 分布」と呼んでいます。

P159 を見てください。ここに「 t 分布表」とその見方が書かれてあります。つまり詳しいことを知らなくても計算ができるようになっています。



仮説検定の方法

実際にどのように行うのかやってみました。

ディフェンスリバウンド数と勝率の関係は

相関係数 r が $r=0.65$

データをとったチーム数は $n=20$ チーム

t 分布表から有意水準 1% (曲線の左右両側の外にいる必要がある)のでP159の表の

$\alpha = \frac{0.01}{2}$ で検証すると

$t_{\frac{0.01}{2}}(18) = 2.878$ の幅の中に入っていれば相関係数が $r_m = 0$

の場合の誤差の範囲内にあることになり相関関係があるとは言えなくなる。

そこで、 $r=0.65$ の場合の t 分布の値は

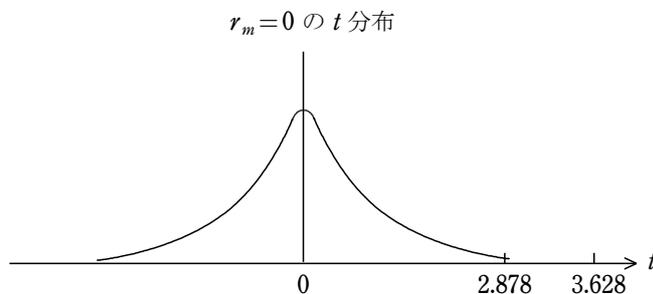
$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.65 \times \sqrt{20-2}}{\sqrt{1-0.65^2}} = 3.6288... > 2.878$$

となり $t=2.878$ よりもかなりはずれたところにあることになる。これは、相関関係がないとした誤差の範囲 99% から大きく外れている。つまり、相関がない、たまたまこうなったとは有意水準 1% で言えないことになる。これを「帰無仮説 $H_0: r_m = 0$ で相関はない」は棄却された

と統計学では表現する。つまりディフェンスリバウンドの数と勝率には有意水準 1% で相関関係があるといえる。

それでは、オフェンスリバウンドではどうだろうか。相関係数が $r=0.51$ である

これは $t=2.515...$ となり有意水準 1% では相関がないとする誤差の範囲内に入ってしまふ。ここで言えるのは、「ディフェンスリバウンドと勝率には相関関係がある」という結論になる。



探究 サッカーJ3の研究

背景

プロサッカーリーグの J リーグやバスケットボールの B リーグ、またアメリカの大リーグ (MLB) など多くのプロリーグの公式サイトには、ゲームの結果だけではなく、詳細な試合データが公開されるようになってきている。

これらのデータは貴重であり、統計的な分析によりそのスポーツの指導法や戦い方に大きな影響を与えている。例えば、バスケットボールの DR (ディフェンスリバウンド) の数は試合の勝率に影響を与えていることが分析結果から分かっている (宮西あさひ 2022)

目的

本研究では、宮崎のプロサッカーチームである J3 のテゲバジャーロ宮崎を中心とした 2021 年の試合データをもとに統計分析を行い、新しい試合の視点を探究することとした。

方法

データ分析 1

次ページのデータ表 1 は、2021 年度のサッカー J3 リーグ 15 チームの年間総合結果である。わがテゲバジャーロは、初めての J3 ながら堂々たる成績を残した。「理数探究基礎」の 54~55 ページを参考にしてこのデータ表 1 から以下の分析を行い、考察せよ。

- (1) 2021 年のサッカー J3 のデータ表 1 から得点数 (点) を横軸、勝った試合数を縦軸にして散布図 A を描け。
- (2) 「理数探究基礎」のテキストの 55 ページを参考に (1) の散布図の相関係数 r を求めよ。
- (3) 2021 年のサッカー J3 のデータ表 1 から失点数 (点) を横軸、負けた試合数を縦軸にして散布図 B を書け。
- (4) 同様にして (3) の散布図の相関係数 r を求めよ。
- (5) 以上の分析から得点数と勝った試合数の相関と失点数と負けた試合数の相関関係はどちらが強いかなを考察せよ。

※取り組む上での条件

各チーム協力して電卓のみで計算し、分布図は手書きで仕上げること。その際、グラフのタイトル、縦軸と横軸の軸の名前 (軸ラベル) などを分かるように描くこと。

		J3 2021年								
2022/03/27 更新		大会	総勝	PK勝	分	総敗	PK敗	得点	失点	現所属DIV
チーム										
ヴァンラーレ八戸	八戸	J3	7	0	8	13	0	24	44	J3
いわてグルージャ盛岡	岩手	J3	15	0	8	5	0	43	28	J2
福島ユナイテッドFC	福島	J3	13	0	6	9	0	41	32	J3
横浜スポーツ&カルチャークラブ	横浜	J3	11	0	7	10	0	31	33	J3
A C長野パルセイロ	長野	J3	8	0	12	8	0	35	28	J3
カターレ富山	富山	J3	13	0	7	8	0	40	34	J3
藤枝MYFC	藤枝	J3	8	0	8	12	0	42	42	J3
アスクラロ沼津	沼津	J3	7	0	6	15	0	32	44	J3
FC岐阜	岐阜	J3	12	0	5	11	0	38	35	J3
ガイナレ鳥取	鳥取	J3	9	0	2	17	0	36	53	J3
カマタマーレ讃岐	讃岐	J3	4	0	9	15	0	20	41	J3
FC今治	今治	J3	7	0	9	12	0	34	33	J3
ロアッソ熊本	熊本	J3	15	0	9	4	0	39	20	J2
テゲバジャーロ宮崎	宮崎	J3	16	0	5	7	0	44	31	J3
鹿児島ユナイテッドFC	鹿児島	J3	11	0	7	10	0	34	35	J3

データ表 1

データ分析 2

下のデータ表 2 と次ページのデータ表 3 は、2021 年度のサッカー J3 リーグ 15 チームのホームゲームにおける観客数データとホームゲームにおける勝率のデータである。「理数探究基礎」テキストの 146 ページを参考にして以下の分析を行い、考察せよ。

- (1) 2021 年シーズンにおける、チーム毎の 1 試合当たりの平均入場者数（ホーム試合）と勝った試合数を示した散布図を縦に勝った試合数、横軸に平均入場者数をとった散布図を描け。
- (2) 求めた散布図の相関係数 r を計算せよ。
- (3) 相関係数はかなり小さいと考えられるが、このときの母相関係数を r_m として r_m が 0 ではないと言えるかどうかを有意水準 5 % で検証せよ。

J3 2021年度 ホームゲーム観客数データ				
	平均観客数	最大	最小	試合数
テゲバジャーロ宮崎	1124.4	2101	421	14
いわてグルージャ盛岡	1394.0	2453	610	14
カターレ富山	2780.0	4131	1332	14
福島ユナイテッドFC	876.9	1641	264	14
ロアッソ熊本	3341.6	11314	1559	14
FC岐阜	3505.1	5472	2184	14
FC今治	1660.6	2236	1155	14
ヴァンラーレ八戸	1137.3	2114	315	14
藤枝MYFC	1322.6	2323	559	14
アスクラロ沼津	1482.4	2743	759	14
鹿児島ユナイテッドFC	3737.6	4957	1838	14
横浜スポーツ&カルチャークラブ	1156.9	1917	477	14
A C長野パルセイロ	2517.9	3336	1765	14
ガイナレ鳥取	1186.0	1861	739	14
カマタマーレ讃岐	1475.1	2441	835	14

データ表 2

J3 2021年度 ホームゲーム勝率データ									
	試合数	勝	引き分け	負	勝率	得点	失点	イエローカード	レッドカード
テゲバジャーロ宮崎	14	9	3	2	64.30%	25	18	10	0
いわてグルージャ盛岡	14	9	2	3	64.30%	23	15	17	0
カターレ富山	14	8	4	2	57.10%	20	11	17	0
福島ユナイテッドFC	14	7	4	3	50.00%	22	16	14	0
ロアッソ熊本	14	6	6	2	42.90%	19	13	9	0
FC岐阜	14	7	3	4	50.00%	21	17	18	0
FC今治	14	6	4	4	42.90%	22	14	12	1
ヴァンラーレ八戸	14	6	4	4	42.90%	14	15	10	0
藤枝MYFC	14	5	4	5	35.70%	21	17	16	1
アスクラロ沼津	14	5	4	5	35.70%	18	17	19	1
鹿児島ユナイテッドFC	14	5	2	7	35.70%	15	21	11	1
横浜スポーツ&カルチャークラブ	14	4	4	6	28.60%	12	19	18	3
AC長野パルセイロ	14	3	6	5	21.40%	16	17	11	0
ガイナレ鳥取	14	4	1	9	28.60%	15	32	8	1
カマタマーレ讃岐	14	1	3	10	7.10%	8	20	10	0

データ表3

分析結果

手書きで散布図を描こう。このとき縦軸、横軸の数値、単位、表している量をしっかりと記入すること。

また、相関係数についても計算結果を分かるように記入すること。

仮説検定については計算の過程までしっかりと書いておこう。

考察・結論

君たちの分析でどのようなことが分かったのかを話し合い、まとめてみよう。

参考文献

このテキストで示した様式に従って、参考にした文献やサイトをしっかりと書いておこう。