Projekt: ANOVA

Magdalena Smalarz

1. Wstęp

Poker to gra karciana, w której uczestnicy otrzymują karty, a zadaniem jest skompletowanie jak najsilniejszego układu. Wynik gry zależy nie tylko od przyjętej strategii gracza, ale też od strategii pozostałych graczy. Nie są również obojętne umiejętności gracza, a także zwykłe szczęście w postaci dobrze rozdanych kart. Pozostaje zadać sobie pytanie - czy poker jest bardziej grą umiejętności czy szczęścia? W trakcie badania postaram się odpowiedzieć na to pytanie, a konkretniej co ma większy wpływ na wynik gry - otrzymane karty, umiejętności gracza czy ich interakcja. W tym celu wykorzystana zostanie dwuczynnikowa analiza wariancji ANOVA. Przyjęty poziom istotności: 5%.

2. Wykorzystane narzędzia

Analiza wariancji (ANOVA) dla klasyfikacji podwójnej bada wpływ dwóch czynników klasyfikujących (najczęściej podzielonych na wiele poziomów) na wartości badanej cechy. Jeśli w każdej podgrupie utworzonej poprzez jednoczesny podział obu czynników na poziomy znajdują się przynajmniej 2 obserwacje, badany jest także wpływ interakcji rozpatrywanych dwóch cech na zmienność cechy.

Niezbędnymi założeniami analizy wariancji jest normalność oraz jednorodność wariancji w podgrupach wyznaczonych przez poziomy czynników klasyfikujących. Za pomocą dwuczynnikowej analizy wariancji każde źródło zmienności testuje się osobno.

H_{0A}: $m_1 = m_2 = \ldots = m_k$ (źródło zmienności A nie różnicuje wyników),

H_{1A}: nie wszystkie m_j są sobie równe ($j = 1,2,\ldots,k$).

H_{0B}: $m_1 = m_2 = \ldots = m_k$ (źródło zmienności B nie różnicuje wyników),

H_{1B}: nie wszystkie m_j są sobie równe ($j = 1,2,\ldots,k$).

H_{0AB}: $m_1 = m_2 = \ldots = m_k$ (źródło zmienności AB nie różnicuje wyników),

H_{1AB}: nie wszystkie m_j są sobie równe ($j = 1,2,\ldots,k$),

gdzie: m_i - średnie badanej zmiennej w populacjach, z których pobrano próby.

Sprawdzianem używanym w ANOVA jest wartość statystyki testowej F, która przy założeniu prawdziwości hipotezy zerowej ma rozkład F Snedecora o liczbie stopni swobody odpowiadających liczbie stopni swobody analizowanego źródła zmienności oraz błędu czynnika losowego.

3. Dane

Wykorzystano dane anova_poker.csv zwierające 300 obserwacji dotyczących wyników gry w pokera (*result*). Oprócz tego w danych znajdują się zmienne *skill* i *hand* charakteryzujące graczy, którzy otrzymali dany wynik:

- *skill* - poziom umiejętności gracza:
  - average - przeciętny,
  - expert - zaawansowany,
- *hand* - karty otrzymane podczas rozdania:
  - good - dobre rozdanie,
  - neutral - neutralne rozdanie,
  - bad - złe rozdanie.

Poniżej przedstawiono kilka losowo wybranych danych.

```
##       skill    hand result
## 64   expert neutral  12.29
## 249 average neutral  -4.47
## 212 average neutral  10.16
## 60   expert neutral  15.30
## 39   expert     bad   8.09
## 204 average neutral   9.19
## 258 average    good  13.72
## 300 average    good   1.38
## 184 average     bad   8.38
## 208 average neutral   9.55
```

Statystyki opisowe wyników gry w pokera przedstawiają się w następujący sposób:

- dla zmiennej *skill*

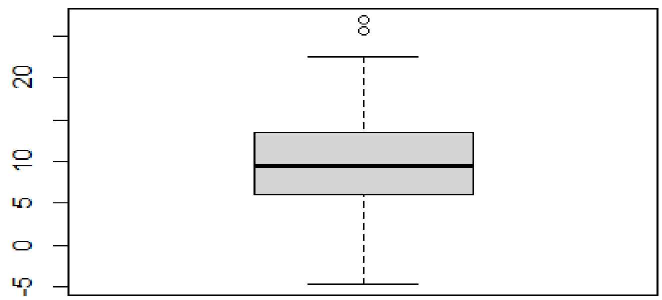
	item	group1	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X11	1	average	1	150	9.26	5.76	9.41	9.31	7.38	-4.47	22.48	26.95	-0.05	-0.92	0.47
X12	2	expert	1	150	10.07	4.64	9.59	9.93	3.91	-4.68	26.99	31.67	0.45	1.60	0.38

- dla zmiennej *hand*

	item	group1	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X11	1	bad	1	100	5.78	3.62	5.70	5.77	3.88	-4.68	16.68	21.36	0.02	0.09	0.36
X12	2	good	1	100	12.98	4.76	13.96	13.25	3.95	-1.61	26.99	28.60	-0.55	1.09	0.48
X13	3	neutral	1	100	10.24	4.50	10.16	10.24	3.47	-4.47	25.63	30.10	0.06	1.74	0.45

Średni wynik gracza o umiejętnościach przeciętnych wynosi 9.26 punktów z odchyleniem standardowym 5.76. W grupie tej co najmniej połowa graczy miała wynik nie większy niż 9.41. Dla umiejętnosci zaawansowanych, średnim wynikiem jest 10.07 z odchyleniem 4.64. Co najmniej połowa ekspertów osiągnęła poziom 9.59 punktów. Maksymalny wynik w grze: 26.99 pkt. Przeciętny wynik uzyskany z dobrym rozdaniem wynosi 12.89, a ze złym 5.78.

Rozkład wyników gry zaprezentowany został na wykresie pudełkowym poniżej. Wśród danych można zaobserwować dane odstające, dodatkowo test Grubbsa wskazał obecność outlierów. Ze względu na liczna grupę analizowanych danych, nieznaną istotność outlierów w badaniu oraz możliwość zaburzenia badania postanowiono je pozostawić.



4. Przebieg badania

4.1. Sprawdzenie założeń ANOVA

4.1.1. Testy normalności

W celu sprawdzenia pierwszego z założeń ANOVA dla każdej z grup w zmiennych *skill* i *hand* przeprowadzono testy normalności: Shapiro-Wilka, Shapiro-Francia i Andersona-Darlinga. Poniżej został zaprezentowany zestaw hipotez akutualny dla każdego z testów oraz wyniki testów dla zmiennych: *skill*, *hand* i interakcji (*average-bad*, *expert-bad*, *average-good*, *expert-good*, *average-netural*, *expert-neutral*).

H_0: rozkład badanej cechy jest rozkładem normalnym,

H_1: rozkład badanej cechy nie jest rozkładem normalnym.

Test Shapiro-Wilka jest uznawany za najlepszy test do sprawdzenia normalności rozkładu zmiennej losowej. Statystyka testowa jest oparta o iloczyn skalarny unormowanych uporządkowanych statystyk porządkowych z unormowanymi wartościami oczekiwanymi statystyk porządkowych. Głównym atutem tego testu jest jego duża moc, tzn. dla ustalonego poziomu istotności prawdopodobieństwo odrzucenia hipotezy H_0, jeśli jest ona fałszywa, jest większe niż w przypadku innych tego typu testów.

```
## $skill
##             p-value
## average 0.029578656
## expert  0.006303388
##
## $hand
##             p-value
## bad     0.816503002
## good    0.006549271
## neutral 0.028385310
##
## $interakcja
```

```
##           p-value
## [1,] 5.111806e-01
## [2,] 4.768006e-01
## [3,] 1.758863e-05
## [4,] 2.162713e-02
## [5,] 5.842708e-02
## [6,] 1.060196e-02
```

Test Shapiro-Francia to uproszczona wersja testu Shapiro-Wilka, w którym macierz kowariancji statystyk pozycyjnych zastąpiono przez macierz identycznościową. Tak jak wcześniej wspomniany test, należy do grupy testów opartych o statystykach pozycyjnych próby. Dla dużych prób test zachowuje się podobnie do testu Shapiro-Wilka.

```
## $skill
##             p-value
## average 0.056505492
## expert  0.003048825
## 
## $hand
##             p-value
## bad     0.497740735
## good    0.004484642
## neutral 0.009750896
## 
## $interakcja
##           p-value
## [1,] 5.642935e-01
## [2,] 1.921348e-01
## [3,] 5.204984e-05
## [4,] 1.273490e-02
## [5,] 4.233443e-02
## [6,] 6.223024e-03
```

Test Andersona-Darlinga jest jednym z testów statystycznych zgodności rozkładu z zadanym rozkładem wzorcowym. Mierzy stopień rozbieżności między dystrybuantą empiryczną, a teoretyczną.

```
## $skill
##            p-value
## average 0.01410220
## expert  0.02610723
## 
## $hand
##             p-value
## bad     0.688820919
## good    0.005011305
## neutral 0.028295124
## 
## $interakcja
##           p-value
## [1,] 3.615896e-01
```

```
## [2,] 6.313014e-01
## [3,] 7.368663e-07
## [4,] 2.137234e-01
## [5,] 1.028989e-01
## [6,] 8.767793e-03
```

Testy zgodnie wskazały na odrzucenie hipotezy zerowej w każdej grupie, oprócz cechy *bad* zmiennej *hand*. W przypadku *hand:bad* nie ma podstaw do odrzucenia hipotezy zerowej. Jeśli chodzi o interakcje cech nie we wszystkich kombinacjach można zaobserwować rozkład normalny. Założenia ANOVA w przypadku normalności rozkładu badanych cech nie są spełnione. Jednak, w przypadku analizowanych danych i ich ilości (powyżej 30 obserwacji) można założyć, że pochodzą one z rozkładu normalnego powołując się na Centralne Twierdzenie Graniczne.

Centralne Twierdzenie Graniczne (Twierdzenie Lindeberga — Levy'ego)

Zakładając, że x_i są niezależnymi zmiennymi podlegającymi (dowolnemu) rozkładowi o skończonej wartości oczekiwanej μ i wariancji σ^2. Dla $n \to \infty$, wielkość $y = \frac{(\sum_{i=1}^{n} x_i) - n\mu}{\sigma\sqrt{n}}$ podlega rozkładowi normalnemu $N(0,1)$.

4.1.2. Testy jednorodności wariancji

Drugim założeniem wymaganym do sprawdzenia jest jednorodność wariancji. W tym celu wykorzystano testy: Levene'a i Browna-Forsythe'a. Gdy wariancje w dwóch grupach różnią się między sobą, dodawanie ich nie jest właściwe i nie daje oszacowania wspólnej wariancji wewnątrzgrupowej. Poniżej został zaprezentowany zestaw hipotez akutualny dla każdego z testów oraz wyniki p-value testów dla zmiennych: *skill*, *hand* i interakcji.

H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$,

H_1: nie wszystkie σ_j^2 są sobie równe ($j = 1,2,\ldots,k$),

gdzie: σ_i^2 - wariancje badanej zmiennej w populacjach, z których pobrano próby.

Test Levene'a to test, w którym wariancje porównuje się poprzez badanie równości wartości przeciętnych modułów odchyleń od średnich poszczególnych grup.

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.

## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.

## $skill
## [1] 0.0001054724
##
## $hand
## [1] 0.160473
##
## $interakcja
## [1] 0.1578037
```

Test Browna-Forsythe'a jest modyfikacją testu Levene'a. Występujące we wzorach średnie arytmetyczne z prób zastępowane są medianami. Zwiększa to odporność testu na brak spełnienia warunku o normalności rozkładów, jednak nie rozwiązuje tego problemu całkowicie.

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.

## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.

## $skill
## [1] 0.0001003019
##
## $hand
## [1] 0.2435029
##
## $interakcja
## [1] 0.3960351
```

Testy zgodnie wskazały rozbieżne wyniki dla poszczególnych grup. W przypadku cechy *hand* oraz interakcji cech nie ma podstaw do odrzucenia hipotezy zerowej - można przyjąć założenie o homogeniczności wariancji. Dla *skill* należy przyjąć hipotezę alternatywną. Ciekawą sytuacją jest fakt, że osobno dla cech *hand* i *skill* otrzymujemy rozbieżne wyniki, a interakcja cech sprawia, że wariancje są jednorodne.

Podsumowując, liczność danych poddanych analizie (300) jest wystarczająco duża, wówczas odchylenia od rozkładu normalnego nie mają znaczenia ze względu na centralne twierdzenie graniczne, zgodnie z którym rozkład średnich z próby zmierza do rozkładu normalnego, niezależnie od rozkładu zmiennej w populacji. Dodatkowo, gdy kurtoza obliczona dla każdej z grup, jest większa od 0, wówczas wartość statystyki F zmierza do małych wartości i nie można odrzucić hipotezy zerowej, nawet jeśli nie jest prawdziwa. Powoduje to odporność ANOVA na założenie o rozkładzie normalnym. W przypadku drugiego założenia o jednorodności wariancji nie otrzymano jednoznacznej odpowiedzi, jednak statystyka F jest odporna na naruszenia tego założenia. Można zatem przyjąć, że wyniki otrzymane w analizie wariancji będą prawidłowe, mimo niespełnienia założeń.

4.2. Dwuczynnikowa ANOVA

Mając sprawdzone założenia, można przejść do wykonania analizy wariancji. Wykorzystano funkcję *aov()* dostępną w języku R. Tak jak zostało wspomniane w rozdziale 2. za pomocą dwuczynnikowej analizy wariancji każde źródło zmienności testuje się osobno. Poniżej został zaprezentowany zestaw hipotez oraz wyniki testów dla zmiennych: *skill*, *hand* i interakcji.

H_{0skill}: $m_1 = m_2 = \ldots = m_k$ (źródło zmienności *skill* nie różnicuje wyników),

H_{1skill}: nie wszystkie m_j są sobie równe ($j = 1,2,..,k$).

H_{0hand}: $m_1 = m_2 = \ldots = m_k$ (źródło zmienności *hand* nie różnicuje wyników),

H_{1hand}: nie wszystkie m_j są sobie równe ($j = 1,2,..,k$).

$H_{0hand*skill}$: $m_1 = m_2 =\ldots= m_k$ (źródło zmienności *hand* * *skill* nie różnicuje wyników),

$H_{1hand*skill}$: nie wszystkie m_j są sobie równe ($j = 1,2,\ldots,k$),

gdzie: m_i - średnie badanej zmiennej w populacjach, z których pobrano próby.

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## hand          2   2647  1323.3  73.712 < 2e-16 ***
## skill         1     49    49.2   2.739 0.09901 .
## hand:skill    2    219   109.5   6.100 0.00254 **
## Residuals   294   5278    18.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wyniki ANOVA wskazują na odrzucenie hipotezy zerowej w przypadku *hand* i interakcji cech. Oznacza to, że rozdanie kart oraz interakcja rozdania i umiejętności gracza wpływają na wynik gry. Wpływ czynnika *hand* zależy jednak od poziomu drugiego czynnika, jakim jest *skill*. W przypadku źródła zmienności czynnika *skill* nie ma podstaw do odrzucenia hipotezy zerowej na przyjętym poziomie istotności - umiejętności gracza nie wpływają na rezultaty gry w pokera.

4.3. Analiza post-hoc

ANOVA wskazała istotne różnice występujące dla *hand* i interakcji. Należy teraz sprawdzić, które z porównywalnych populacji są odpowiedzialne za odrzucenie hipotezy zerowej, czyli znaleźć kombinacje, które mają wpływ na wynik gry w pokera. W tym celu wykonano: test Tuckey'a, test Duncana i test Scheffe. Poniżej został zaprezentowany zestaw hipotez akutualny dla każdego z testów oraz wyniki p-value testów dla zmiennych: *skill*, *hand* i interakcji.

H_0: średnie w grupach różnią się nieistotnie,

H_1: średnie w grupach różnią się istotnie.

Test Tuckey'a to jeden z najczęściej używanych testów do porównywania par średnich. Jest on oparty na rozkładzie nazywanym statystyką rozstępu studentyzowanego. Poziom błędu doświadczenia dla wszystkich porównań parami pozostaje na poziomie błędu dla zbioru, co oznacza, że jeżeli założono dla testu ANOVA poziom istotności statystycznej 5%, to taki sam poziom istotności statystycznej będzie użyty podczas wszystkich porównań pomiędzy parami (próbkami).

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = result ~ hand * skill, data = poker)
##
## $hand
##                 diff       lwr       upr    p adj
## good-bad      7.2079  5.796394  8.619406 0.00e+00
## neutral-bad   4.4601  3.048594  5.871606 0.00e+00
## neutral-good -2.7478 -4.159306 -1.336294 1.99e-05
##
## $skill
##                     diff        lwr      upr     p adj
```

```
## expert-average 0.8096667 -0.1532014 1.772535 0.0990062
##
## $`hand:skill`
##                                   diff       lwr         upr     p adj
## good:average-bad:average        9.2698  6.838939 11.70066096 0.0000000
## neutral:average-bad:average     5.1800  2.749139  7.61086096 0.0000000
## bad:expert-bad:average          2.6642  0.233339  5.09506096 0.0224318
## good:expert-bad:average         7.8102  5.379339 10.24106096 0.0000000
## neutral:expert-bad:average      6.4044  3.973539  8.83526096 0.0000000
## neutral:average-good:average   -4.0898 -6.520661 -1.65893904 0.0000328
## bad:expert-good:average        -6.6056 -9.036461 -4.17473904 0.0000000
## good:expert-good:average       -1.4596 -3.890461  0.97126096 0.5180280
## neutral:expert-good:average    -2.8654 -5.296261 -0.43453904 0.0105193
## bad:expert-neutral:average     -2.5158 -4.946661 -0.08493904 0.0377155
## good:expert-neutral:average     2.6302  0.199339  5.06106096 0.0253430
## neutral:expert-neutral:average  1.2244 -1.206461  3.65526096 0.6994669
## good:expert-bad:expert          5.1460  2.715139  7.57686096 0.0000001
## neutral:expert-bad:expert       3.7402  1.309339  6.17106096 0.0002058
## neutral:expert-good:expert     -1.4058 -3.836661  1.02506096 0.5601012
```

Test Duncana używa statystyki opartej na studentyzowanym rozstępie. Jest procedurą krokową bazującą na intuicjach testu Tuckey'a. W jego przypadku im bardziej średnie są od siebie oddalone, tym łatwiej uzyskać wynik istotny statystycznie.

```
##
##   Posthoc multiple comparisons of means : Duncan's new multiple range test
##     95% family-wise confidence level
##
## $hand
##                 diff    lwr.ci    upr.ci    pval
## good-bad      7.2079  5.966462  8.449338 3.5e-13 ***
## neutral-bad   4.4601  3.280832  5.639368 1.8e-12 ***
## neutral-good -2.7478 -3.927068 -1.568532 6.7e-06 ***
##
## $skill
##                     diff     lwr.ci   upr.ci   pval
## expert-average 0.8096667 -0.1532014 1.772535 0.0990 .
##
## $`hand:skill`
##                                   diff     lwr.ci     upr.ci    pval
## good:average-bad:average        9.2698  7.3782558 11.1613442 1.5e-13 ***
## neutral:average-bad:average     5.1800  3.4243417  6.9356583 4.7e-09 ***
## bad:expert-bad:average          2.6642  0.9964636  4.3319364  0.0018 **
## good:expert-bad:average         7.8102  5.9525047  9.6678953 1.9e-13 ***
## neutral:expert-bad:average      6.4044  4.5900104  8.2187896 1.3e-12 ***
## neutral:average-good:average   -4.0898 -5.9041896 -2.2754104 4.4e-06 ***
## bad:expert-good:average        -6.6056 -8.4632953 -4.7479047 4.7e-13 ***
## good:expert-good:average       -1.4596 -3.1273364  0.2081364  0.0860 .
## neutral:expert-good:average    -2.8654 -4.6210583 -1.1097417  0.0012 **
## bad:expert-neutral:average     -2.5158 -4.1835364 -0.8480636  0.0032 **
## good:expert-neutral:average     2.6302  0.8745417  4.3858583  0.0030 **
```

```
## neutral:expert-neutral:average  1.2244 -0.4433364  2.8921364  0.1496
## good:expert-bad:expert          5.1460  3.3316104  6.9603896 7.7e-09 ***
## neutral:expert-bad:expert       3.7402  1.9845417  5.4958583 2.1e-05 ***
## neutral:expert-good:expert     -1.4058 -3.0735364  0.2619364  0.0982 .
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test Scheffe wykonuje równocześnie wszystkie możliwe łączne porównania parami pomiędzy wszystkimi możliwymi parami średnich. Wykorzystuje rozkład statystyki F. Najbardziej ogólna procedura, ale również najbardziej wymagająca.

```
##
##   Posthoc multiple comparisons of means: Scheffe Test
##     95% family-wise confidence level
##
## $hand
##                diff    lwr.ci     upr.ci    pval
## good-bad      7.2079  5.200484  9.2153156 < 2e-16 ***
## neutral-bad   4.4601  2.452684  6.4675156 8.6e-10 ***
## neutral-good -2.7478 -4.755216 -0.7403844  0.0010 **
##
## $skill
##                     diff     lwr.ci   upr.ci   pval
## expert-average 0.8096667 -0.8293813 2.448715 0.7400
##
## $`hand:skill`
##                                 diff     lwr.ci      upr.ci    pval
## good:average-bad:average        9.2698  6.4308856 12.10871443 < 2e-16 ***
## neutral:average-bad:average     5.1800  2.3410856  8.01891443 1.3e-06 ***
## bad:expert-bad:average          2.6642 -0.1747144  5.50311443 0.08198 .
## good:expert-bad:average         7.8102  4.9712856 10.64911443 9.3e-15 ***
## neutral:expert-bad:average      6.4044  3.5654856  9.24331443 4.3e-10 ***
## neutral:average-good:average   -4.0898 -6.9287144 -1.25088557 0.00042 ***
## bad:expert-good:average        -6.6056 -9.4445144 -3.76668557 1.0e-10 ***
## good:expert-good:average       -1.4596 -4.2985144  1.37931443 0.70508
## neutral:expert-good:average    -2.8654 -5.7043144 -0.02648557 0.04621 *
## bad:expert-neutral:average     -2.5158 -5.3547144  0.32311443 0.12039
## good:expert-neutral:average     2.6302 -0.2087144  5.46911443 0.08978 .
## neutral:expert-neutral:average  1.2244 -1.6145144  4.06331443 0.83643
## good:expert-bad:expert          5.1460  2.3070856  7.98491443 1.6e-06 ***
## neutral:expert-bad:expert       3.7402  0.9012856  6.57911443 0.00196 **
## neutral:expert-good:expert     -1.4058 -4.2447144  1.43311443 0.73796
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Każdy z przeprowadzonych testów wskazał na odrzucenie hipotezy zerowej dla zmiennej *hand*. W każdej parze z osobna dla zmiennej *hand* średnie są istotnie różne. Oznacza to, że każde rozdanie ma wpływ na wynik gry i nie ma od tego żadnych wyjątków. Jeśli chodzi o interakcje cech, można zauważyć, że istnieją grupy w których średnie różnią i nie różnią się istotnie. Większość kombinacji cech *hand* i *skill* wpływa na wynik gry, ale nie każda. Test Scheffe zwrócił 9 par dla których należy odrzucić hipotezę zerową, a test Duncana i Tuckey'a 12. Warto wspomnieć, że test Tuckey'a wskazał tylko dość spore różnice i dla większości par zaobserwowano p-value na poziomie 0. Wszystkie testy również zgodnie wskazały największe różnice: *good-bad* dla *hand* oraz *good:average-bad:average* dla interakcji cech.

4.4. Efekty eksperymentalne

Wielkość efektu jest ilościową miarą siły zjawiska obliczaną na podstawie danych. Stosuje się ją do mierzenia wpływu pewnego czynnika na wynik ogólny grupy. Wielkość efektu nie jest zależna od wielkości próby, a jego interpretacja opiera się na założeniu o normalności rozkładów wyników porównywanych grup. Wykorzystując dane zwrócone w analizie wariancji obliczono wskaźniki:

- cząstkowe η^2 - miara siły efektu. Jej wartość może znajdować się w przedziale od 0 do 1; wskaźnik ten pokazuje jaki procent zmienności w zakresie zmiennej zależnej jest wyjaśniany przez zmienną niezależną,
- ω^2 - estymator wariancji zmiennej zależnej wyjaśnionej przez zmienną niezależną w całej populacji.

```
##            eta-kwadrat omega-kwadrat
## hand       0.334006309   0.317988065
## skill      0.009198423   0.003775423
## hand*skill 0.039839913   0.022287176
```

Interpretacja:

- zmienna *hand* wyjaśnia 33,4% zmienności wyników gry i jest to silny efekt,
- dla zmiennej *skill* i interakcji otrzymano odpowiednio 0,9% i 3,9% - efekty słabe, niemające znaczącego wpływu na zmienną objaśnianą,
- zmienna *hand* bardziej różnicuje wynik gry - w największym stopniu wpływa na wielkość uzyskanych punktów,
- dla każdej grupy zachodzi nierówność $\eta^2 > \omega^2$.

4.5. Test Kruskala-Wallisa

W przypadku braku spełnienia założeń testu ANOVA alternatywą jest nieparametryczny test Kruskala–Wallisa. Można go wykorzystać również wtedy gdy zmienne mają charakter porządkowy. Test sprawdza czy n niezależnych próbek pochodzi z populacji z taką samą medianą. Minusem Kruskala-Wallisa jest fakt, że nie można przetestować wpływu dwóch cech. Wymagania testu:

- rozkład cechy jest ciągły,
- liczne grupy,
- maksymalne porównanie dziesięciu grup.

Analizowane dane spełniają wymagania. Hipotezy tesu dotyczą równości średnich rang dla kolejnych populacji lub są upraszczane do median. Poniżej przedstawiono wyniki testu odpowiednio dla *hand* i *skill* oraz zestaw hipotez.

$H_0: Me_1 = Me_2 = \ldots = Me_n$,

$H_1: Me_i \neq Me_j$ dla pewnych $i \neq j$, gdzie: Me_k - mediany badanej zmiennej w populacjach, z których pobrano próby.

```
##
##  Kruskal-Wallis rank sum test
##
## data:  result by hand
## Kruskal-Wallis chi-squared = 108.42, df = 2, p-value < 2.2e-16

##
##  Kruskal-Wallis rank sum test
##
## data:  result by skill
## Kruskal-Wallis chi-squared = 1.0437, df = 1, p-value = 0.307
```

Nie ma podstaw do odrzucenia H_0 dla cechy *skill* - czynnik ten nie różnicuje wyników gry w pokera. Dla *hand* test wskazał na odrzucenie hipotezy zerowej, co oznacza, że rozdane karty mają wpływ na grę. Otrzymano wyniki podobne do tych z dwuczynnikowej analizy wariancji.

4.6. Analiza post-hoc

Istotny statystycznie wynik testu Kruskala-Wallisa mówi tylko o tym, że co najmniej jedna z grup różni się od innej grupy. W celu sprawdzenia które grupy różnią się miedzy sobą wykonuje się test wielokrotnych porównań średnich rang dla wszystkich prób - **test Dunna**. Poniżej znajdują się wyniki testu cechy *hand* oraz hipotezy.

H_0: średnie w grupach różnią się nieistotnie,

H_1: średnie w grupach różnią się istotnie.

```
## Dunn (1964) Kruskal-Wallis multiple comparison

##   p-values adjusted with the Holm method.

##       Comparison          Z      P.unadj        P.adj
## 1     bad - good -10.314807 6.040228e-25 1.812068e-24
## 2  bad - neutral  -6.391121 1.646738e-10 3.293476e-10
## 3 good - neutral   3.923686 8.720430e-05 8.720430e-05
```

Podobnie jak w przypadku analizy post-hoc dla ANOVA, test Dunna wskazał istotne statystycznie różnice średnich w każdej parze.

5. Podsumowanie

Test F jest odporny na odchylenia od normalności oraz na naruszenie założenia o jednorodności wariancji, dlatego ANOVA zwróciła identyczne wyniki co nieparametryczny test Kruskala-Wallisa.

Badanie wykazało, że umiejętności gracza nie mają wpływu na wynik gry w pokera, natomiast ich interakcja z rozdaniem już tak. Największy wpływ mają jednak rozdane karty. Można zatem powiedzieć, że poker to gra szczęścia - rezulat gry w dużej mierze zależy od tego jakie karty otrzyma się w rozdaniu.

Bibliografia

1. Amir D. Aczel, *„Statystyka w zarządzaniu”*,
2. M. Czajkowski, *„Poker – gra szczęścia czy umiejętności? Przegląd analiz teoretycznych i empirycznych oraz wnioski dla regulacji”*,
3. Dane ze strony smarterpoland.pl/index.php/2013/04/wybrane-testy-normalnoci/ [Dostęp: 30.10.2020],
4. Dane ze strony brain.fuw.edu.pl/edu/index.php/WnioskowanieStatystyczne/CLT [Dostęp: 30.10.2020],
5. Dane ze strony pstconsulting.pl/po-godzinach/testy-porownan-wielokrotnych-post-hoc/ [Dostęp: 31.10.2020],
6. Dane ze strony ibm.com/support/knowledgecenter/pl/SSLVMB_sub/statistics_mainhelp_ddita/spss/base/idh_onew_post.html [Dostęp: 31.10.2020],
7. Dane ze strony nauka.metodolog.pl/wielkosc-efektu-podstawowe-miary-szacujace-sile-zwiazkow-i-wplywow/,
8. Dane ze strony statsoft.pl/textbook/stathome_stat.html [Dostęp: 31.10.2020],
9. Dane ze strony statystyka.az.pl/2-czynnikowa-analiza-wariancji-anova-uklad-czynnikowy.php [Dostęp: 31.10.2020].