

# Gry wideo

Damian Szopiński 185394

Maciej Pestka 170088

27 stycznia 2023

## 1 Wstęp

W analizie zakładam, że zmienne, które analizujemy są zmiennymi losowymi a dotyczące ich dane, którymi dysponujemy są próbkami z populacji.

W projekcie zajmiemy się analizą dwóch zmiennych Year to rok wydania gry video, natomiast Global Sale, JP Sale, EU Sale, NA Sale i Other Sale to liczba sprzedanych egzemplarzy danej gry w danym regionie (odpowiednio dla świata ogółem, Japonii, Europy, Ameryki Północnej oraz sumy wszystkich pozostałych regionów). Liczby podane są w milionach.

Na początku w naszej bazie danych musieliśmy odrzucić wszystkie te wiersze, w których przynajmniej jedna z korzystanych zmiennych wskazywała na "N/A", to znaczy, brak danych w danej kategorii. Nie były one brane pod uwagę w naszych testach.

Dane zostały zaczerpnięte ze strony:

<https://www.kaggle.com/datasets/gregorut/videogamesales>

W tej pracy wplotliśmy również fragmenty kodów z R studio na końcu każdego zadania.

Przygotowywanie danych w R:

```
11 sprzedaż_gier <- read_excel("D:/GitHub/Metody-Numeryczne/Statystyka/sprzedaż gier.xlsx")
12 view(sprzedaż_gier) # Numer kolumny oraz nazwa kolumny
13 #kasujemy wiersze z brakiem wystarczających danych
14 dane <- sprzedaż_gier[!grepl("N/A", sprzedaż_gier$Year),]
15
16 # 4. Year
17 # 11. Global_sales\
18 #10. Other_Sales
19 #9. JP_Sales
20 #8. EU_SALES
21 #7. NA_SALES
22 Year<-as.numeric(dane[4]$Year)
23 Global_Sales<-as.numeric(gsub(",", ".", dane[11]$Global_Sales))
24 Other_Sales<-as.numeric(gsub(",", ".", dane[10]$Other_Sales))
25 JP_Sales<-as.numeric(gsub(",", ".", dane[9]$JP_Sales))
26 EU_Sales<-as.numeric(gsub(",", ".", dane[8]$EU_Sales))
27 NA_Sales<-as.numeric(gsub(",", ".", dane[7]$NA_Sales))
28
```

## 2 Własności zmiennych (1)

Miary tendencji centralnej:

W powyższej tabelce najbardziej wydaje się wyróżniać średnia dla sprzedaży gier w Europie.

	Year	Global Sale	JP Sale	EU Sale	NA sale	Other Sale
Średnia	2006.406	0.5402315	0.07866111	0.1475544	0.07866111	0.04832547
Mediana	2007	0.17	0	0.02	0.08	0.01
Moda	2009	0.02	0	0	0	0

Jest ona zauważalnie większa od wszystkich pozostałych regionów. Pomimo tego, mediana Stanów Zjednoczonych jest 4 razy większa niż Europie. Być może, może to być spowodowane tym, że Europejczycy wolą grać w węższą listę tytułów. Nie jest to jednak sprawą naszych późniejszych testów.

Miary położenia:

Kwartył Year				
0%	25%	50%	75%	100%
1980	2003	2007	2010	2020
Kwartył Global Sale				
0%	25%	50%	75%	100%
0.01	0.06	0.17	0.48	82.74
Kwartył Other Sale				
0%	25%	50%	75%	100%
0	0	0.01	0.04	10.57
Kwartył JP Sale				
0%	25%	50%	75%	100%
0	0	0	0.04	10.22
Kwartył EU Sale				
0%	25%	50%	75%	100%
0	0	0.02	0.11	29.02
Kwartył NA Sale				
0%	25%	50%	75%	100%
0	0	0.08	0.24	41.49

Kwartył pierwszy (kolumna 25%) informuje, że 25% obserwacji nie jest większa niż podana pod nią liczba. Analogicznie dla pozostałych kwartyli.

# Miary Rozproszenia:

	Year	Global Sale
wariancja	33.97702	2.451516
odchylenie standardowe	5.828981	1.565732
wspolczynn timer zmienności	0.002905185	2.898261
rozstep kwartyłowy	7	0.42
Kwartyłowy współczynnik zmienności	0.003487793	2.470588

	Other Sale	JP Sale	EU Sale	NA Sale
wariancja	0.03605648	0.09706775	0.2588425	0.6750115
odchylenie standardowe	0.1898854	0.311557	0.5087657	0.8215909
wspolczynn timer zmienności	3.929303	3.96075	3.447988	3.095496
rozstep kwartyłowy	0.04	0.04	0.11	0.24
Kwartyłowy współczynnik zmienności	4	Inf	5.5	3

Wariancja wskazuje na to, że rozproszenie w każdym z regionów jest różne. Po raz kolejny zauważamy, że w Stanach Zjednoczonych rozproszenie tych zmiennych jest najbardziej widoczne. Odchylenie standardowe, podobnie jak wariancja, wskazuje na umiarkowane rozrzucenie obserwacji wokół średniej.

Kod w R:

```

26 # Zadanie 1
27 # Średnia
28 mean(Year)
29 mean(Global_Sales)
30 mean(Other_Sales)
31 mean(JP_Sales)
32 mean(EU_Sales)
33 mean(NA_Sales)
34 # mediana
35 median(Year)
36 median(Global_Sales)
37 median(Other_Sales)
38 median(JP_Sales)
39 median(EU_Sales)
40 median(NA_Sales)
41 # moda
42 getmode <- function(v)
43 {
44   uniqv<-unique(v)
45   uniqv[which.max(tabulate(match(v,uniqv)))]
46 }
47 moda_Year=getmode(Year)
48 print(moda_Year)
49 moda_Global_Sales=getmode(Global_Sales)
50 print(moda_Global_Sales)
51 moda_Other_Sales=getmode(Other_Sales)
52 print(moda_Other_Sales)
53 moda_JP_Sales=getmode(JP_Sales)
54 print(moda_JP_Sales)
55 moda_EU_Sales=getmode(EU_Sales)
56 print(moda_EU_Sales)
57 moda_NA_Sales=getmode(NA_Sales)
58 print(moda_NA_Sales)
59 # kwartyle
60 quantile(Year)
61 quantile(Global_Sales)
62 quantile(Other_Sales)
63 quantile(JP_Sales)
64 quantile(EU_Sales)
65 quantile(NA_Sales)
66
67 # wariancja
68 var(Year)
69 var(Global_Sales)
70 var(Other_Sales)
71 var(JP_Sales)
72 var(EU_Sales)
73 var(NA_Sales)
74

```

```

75 # odchylenie standardowe
76 sd(Year)
77 sd(Global_Sales)
78 sd(other_Sales)
79 sd(JP_Sales)
80 sd(EU_Sales)
81 sd(NA_Sales)
82
83 # współczynnik zmienności
84 wZ_Year=sd(Year)/mean(Year)
85 print(wZ_Year)
86 wZ_Global_Sale=sd(Global_Sales)/mean(Global_Sales)
87 print(wZ_Global_Sale)
88 wZ_other_Sale=sd(other_Sales)/mean(other_Sales)
89 print(wZ_other_Sale)
90 wZ_JP_Sale=sd(JP_Sales)/mean(JP_Sales)
91 print(wZ_JP_Sale)
92 wZ_EU_Sale=sd(EU_Sales)/mean(EU_Sales)
93 print(wZ_EU_Sale)
94 wZ_NA_Sales=sd(NA_Sales)/mean(NA_Sales)
95 print(wZ_NA_Sales)
96 # rozstęp kwartyłowy
97 IQR(Year)
98 IQR(Global_Sales)
99 IQR(other_Sales)
100 IQR(JP_Sales)
101 IQR(EU_Sales)
102 IQR(NA_Sales)
103
104 # Kwartyłowy współczynnik zmienności
105 KwZ_Year<-IQR(Year)/median(Year)
106 print(KwZ_Year)
107 KwZ_Global_Sales<-IQR(Global_Sales)/median(Global_Sales)
108 print(KwZ_Global_Sales)
109 KwZ_other_Sales<-IQR(other_Sales)/median(other_Sales)
110 print(KwZ_other_Sales)
111 KwZ_JP_Sales<-IQR(JP_Sales)/median(JP_Sales)
112 print(KwZ_JP_Sales)
113 KwZ_EU_Sales<-IQR(EU_Sales)/median(EU_Sales)
114 print(KwZ_EU_Sales)
115 KwZ_NA_Sales<-IQR(NA_Sales)/median(NA_Sales)
116 print(KwZ_NA_Sales)
117

```

### 3 Miary klasyczne i nieklasyczne (2)

	Year	Global Sales
Miary klasyczne		
Moment trzeci centralny ( $\mu_3$ )	-198.522	66.47233
Moment trzeci względny ( $\alpha_3$ )	-1.002376	17.31764
Moment czwarty centralny ( $\mu_4$ )	5595.161	3605.495
Moment czwarty względny ( $\alpha_4$ )	4.846653	599.9227
Miary nieklasyczne		
Kwartyl współczynnik skośności	-1.002376	17.31764

	Other Sales	JP Sales	EU Sales	NA Sales
Miary klasyczne				
Moment trzeci centralny ( $\mu_3$ )	0.1651514	0.3367088	2.474508	10.40093
Moment trzeci względny ( $\alpha_3$ )	24.12168	11.13376	18.79037	18.75449
Moment czwarty centralny ( $\mu_4$ )	1.321612	1.832858	50.2929	294.7414
Moment czwarty względny ( $\alpha_4$ )	1016.57	27.35631	194.5265	646.8726
Miary nieklasyczne				
Kwartyl współczynnik skośności	24.12168	11.13376	18.79037	18.75449

Zmienna Year:

Moment trzeci centralny jest mniejszy od 0, moment trzeci względny jest w przedziale -2 i 2, a moment czwarty względny jest większy niż 3. Czyli dla momentu 3 jest zatem mała skośność. Dla momentu 4 jest to rozkład wysmukły. Jest prawoskośny (współczynnik skośności ujemny oraz moment trzeci centralny).

Zmienne Global Sales, Other Sales, JP Sales, EU Sales, Na Sales wyglądają podobnie:

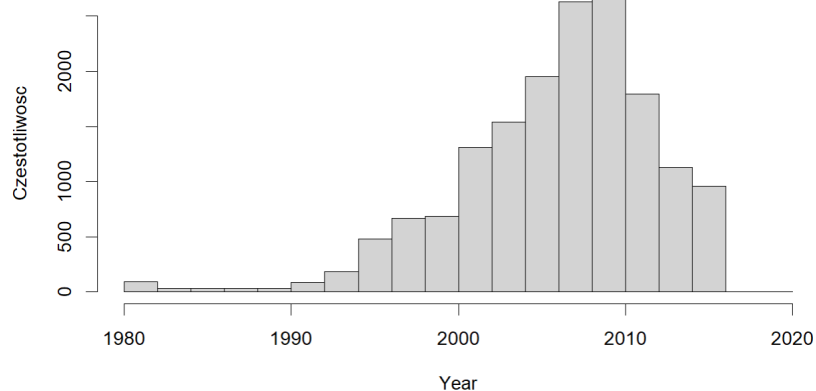
Moment trzeci centralny jest większy od 0, moment trzeci względny jest wysoki, wyższy od 2, a moment czwarty względny jest bardzo duży. Występuje duża skośność, oraz rozkład smukły. Są prawoskośne, ponieważ momenty trzecie centralne oraz współczynniki skośności są dodatnie.

Kod w R:

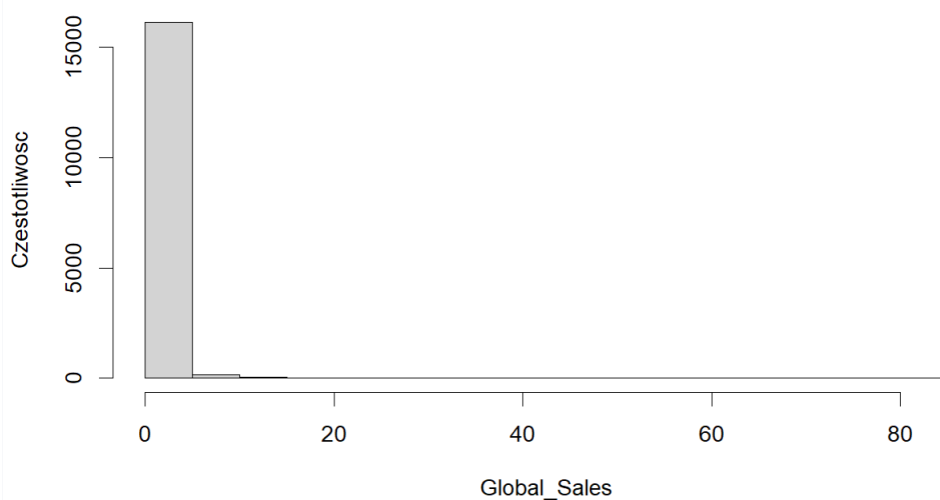
```
119 #Zadanie 2
120 #Miary klasyczne
121 #Moment trzeci centralny
122 m3Year<-moment(Year, order=3, center=TRUE)
123 print(m3Year)
124 m3Global_Sales<-moment(Global_Sales, order=3, center=TRUE)
125 print(m3Global_Sales)
126 m3Other_Sales<-moment(other_sales, order=3, center=TRUE)
127 print(m3Other_Sales)
128 m3JP_Sales<-moment(JP_Sales, order=3, center=TRUE)
129 print(m3JP_Sales)
130 m3EU_Sales<-moment(EU_Sales, order=3, center=TRUE)
131 print(m3EU_Sales)
132 m3NA_Sales<-moment(NA_Sales, order=3, center=TRUE)
133 print(m3NA_Sales)
134 #moment trzeci względny
135 print(m3Year/sd(Year)^3)
136 print(m3Global_Sales/sd(Global_Sales)^3)
137 print(m3Other_Sales/sd(other_sales)^3)
138 print(m3JP_Sales/sd(JP_Sales)^3)
139 print(m3EU_Sales/sd(EU_Sales)^3)
140 print(m3NA_Sales/sd(NA_Sales)^3)
141 #moment czwarty centralny
142 m4Year<-moment(Year, order=4, center=TRUE)
143 print(m4Year)
144 m4Global_Sales<-moment(Global_Sales, order=4, center=TRUE)
145 print(m4Global_Sales)
146 m4Other_Sales<-moment(other_sales, order=4, center=TRUE)
147 print(m4Other_Sales)
148 m4JP_Sales<-moment(JP_Sales, order=4, center=TRUE)
149 print(m4JP_Sales)
150 m4EU_Sales<-moment(EU_Sales, order=4, center=TRUE)
151 print(m4EU_Sales)
152 m4NA_Sales<-moment(NA_Sales, order=4, center=TRUE)
153 print(m4NA_Sales)
154 ##moment czwarty względny
155 print(m4Year/sd(Year)^4)
156 print(m4Global_Sales/sd(Global_Sales)^4)
157 print(m4Other_Sales/sd(other_sales)^4)
158 print(m4JP_Sales/sd(JP_Sales)^4)
159 print(m4EU_Sales/sd(EU_Sales)^4)
160 print(m4NA_Sales/sd(NA_Sales)^4)
161 #Miary nieklasyczne
162 #Kwartył współczynnik skosności
163 skewness(Year)
164 skewness(Global_Sales)
165 skewness(other_sales)
166 skewness(JP_Sales)
167 skewness(EU_Sales)
168 skewness(NA_Sales)
```

## 4 Histogramy (3)

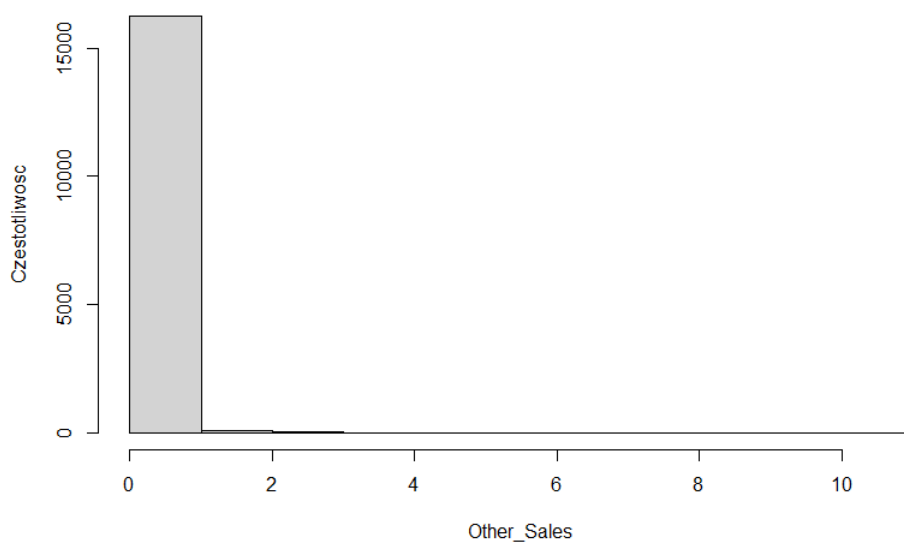
Histogram danych dotyczących roku



Histogram danych dotyczących sprzedaży gier

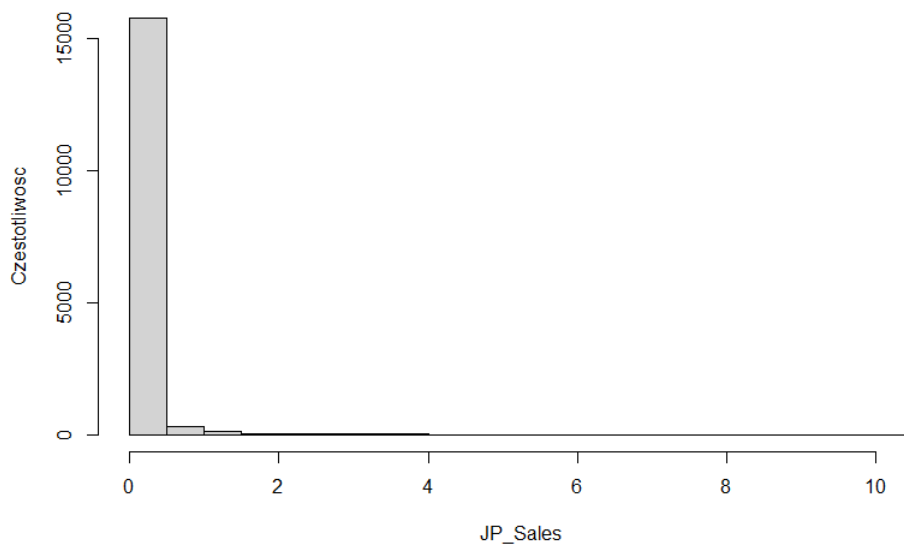


Histogram danych dotyczących sprzedaży gier w reszcie świata

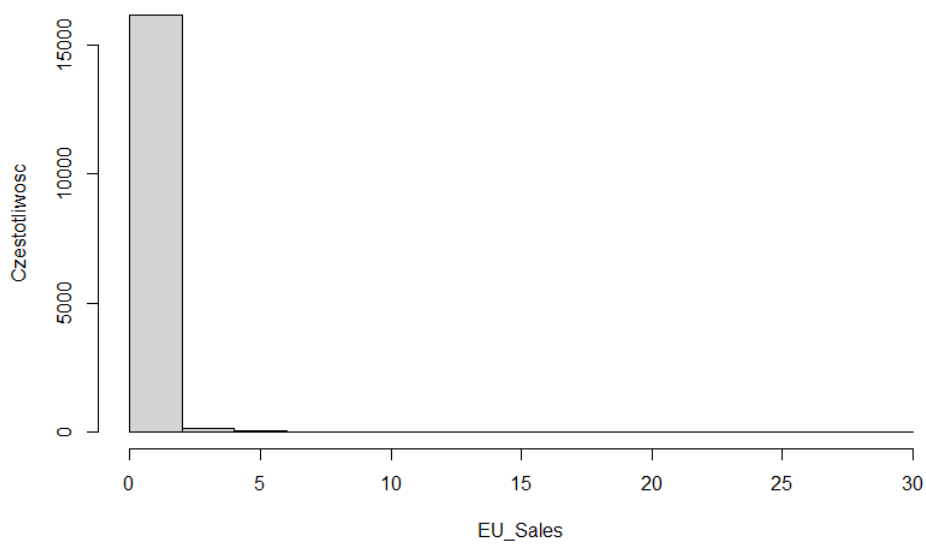




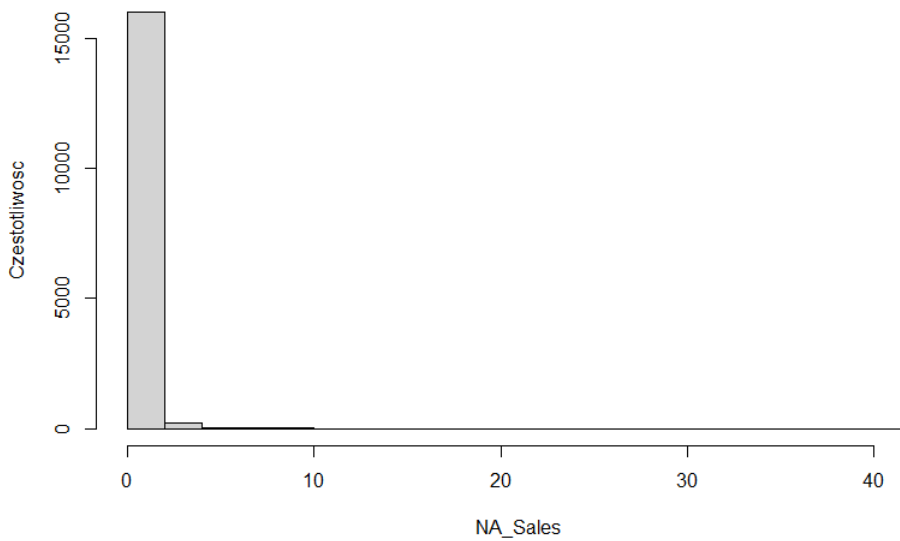
**Histogram danych dotyczących sprzedaży gier w Japoni**



**Histogram danych dotyczących sprzedaży gier w Europie**



Histogram danych dotyczących sprzedaży gier w Na



Kod w R:

```
170 #Zadanie 3
171 #Histogramy
172 hist(Year, main="Histogram danych dotyczących roku", ylab="Częstotliwość")
173 hist(Global_Sales, main="Histogram danych dotyczących sprzedaży gier", ylab="Częstotliwość")
174 hist(other_Sales, main="Histogram danych dotyczących sprzedaży gier w reszcie świata", ylab="Częstotliwość")
175 hist(JP_Sales, main="Histogram danych dotyczących sprzedaży gier w Japoni ", ylab="Częstotliwość")
176 hist(EU_Sales, main="Histogram danych dotyczących sprzedaży gier w Europie", ylab="Częstotliwość")
177 hist(NA_Sales, main="Histogram danych dotyczących sprzedaży gier w Na", ylab="Częstotliwość")
178
```

## 5 Testy zgodności z rozkładem normalnym (4)

$H_0$ -hipoteza zerowa, dane pochodzą z rozkładu normalnego.

$H_A$ -hipoteza alternatywna, dane nie pochodzą z rozkładu normalnego.

Zakładamy poziom istotności  $\alpha=0,05$ .

Nie udało się przeprowadzić testu "Shapiro test" oraz "Cramer-von Mises" ze względu na dane (R nie zwrócił nam ich wartości). Problemy być może mogła sprawiać zbyt duża liczba wierszy w bazie danych. (W kodzie R, skopiowałem treść napotkanych błędów).

Wartość teoretyczna testu Kolmogorov-Smirnov wynosi u nas około 0.0105556...

Wartość teoretyczna testu Anderson-Darling wynosi u nas około 1.225...

Data:	Year
p-value (dla każdego)	<2.2e-16
Kolmogorov-Smirnov	0.1044
Anderson-Darling	170.87
Data:	Global Sales
p-value (dla każdego)	<2.2e-16
Kolmogorov-Smirnov	0.36744
Anderson-Darling	3232.7
Data:	Other Sales
p-value (dla każdego)	<2.2e-16
Kolmogorov-Smirnov	0.39956
Anderson-Darling	3763.6
Data:	JP Sales
p-value (dla każdego)	<2.2e-16
Kolmogorov-Smirnov	0.40034
Anderson-Darling	4003
Data:	EU Sales
p-value (dla każdego)	<2.2e-16
Kolmogorov-Smirnov	0.3859
Anderson-Darling	3515
Data:	NA Sales
p-value (dla każdego)	<2.2e-16
Kolmogorov-Smirnov	0.37333
Anderson-Darling	3206

W każdym z testów uzyskaliśmy p-value mniejsze od  $2.2e-16$ , zatem  $p < \alpha = 0.05$  (czyli maksymalnego dopuszczenia błędu 1 rodzaju). Mamy zatem podstawy, żeby odrzucić hipotezę zerową. Rozkład danych nie jest normalny.

Kod w R:

```

179 #Zadanie 4
180 # Kolmogorov-Smirnov test
181 T_Year<-ks.test(Year,"pnorm",mean=mean(Year),sd=sd(Year))
182 print(T_Year)
183 # - wartości powtórzone nie powinny być obecne w teście kolmogorowa-Smirnowa
184 T_Sales<-ks.test(Global_Sales,"pnorm",mean=mean(Global_Sales),sd=sd(Global_Sale:
185 print(T_Sales)
186 T_Other_Sales<-ks.test(Other_Sales,"pnorm",mean=mean(Other_Sales),sd=sd(Other_S:
187 print(T_Other_Sales)
188 T_JP_Sales<-ks.test(JP_Sales,"pnorm",mean=mean(JP_Sales),sd=sd(JP_Sales))
189 print(T_JP_Sales)
190 T_EU_Sales<-ks.test(EU_Sales,"pnorm",mean=mean(EU_Sales),sd=sd(EU_Sales))
191 print(T_EU_Sales)
192 T_NA_Sales<-ks.test(NA_Sales,"pnorm",mean=mean(NA_Sales),sd=sd(NA_Sales))
193 print(T_NA_Sales)
194
195 # Cramer-von Mises test
196 # R nie zwraca jego wartości
197 #CM_Year<-cvm.test(Year)
198 #print(AD_Year)
199 #CM_Sales<-cvm.test(Global_Sales)
200 #print(AD_Sales)
201 #CM_Sales<-cvm.test(NA_Sales)
202 #CM_Sales<-cvm.test(EU_Sales)
203 #CM_Sales<-cvm.test(JP_Sales)
204
205 # Anderson-Darling test
206 AD_Year<-ad.test(Year)
207 print(AD_Year)
208 AD_Sales<-ad.test(Global_Sales)
209 print(AD_Sales)
210 AD_Other_Sales<-ad.test(Other_Sales)
211 print(AD_Other_Sales)
212 AD_JP_Sales<-ad.test(JP_Sales)
213 print(AD_JP_Sales)
214 AD_EU_Sales<-ad.test(EU_Sales)
215 print(AD_EU_Sales)
216 AD_NA_Sales<-ad.test(NA_Sales)
217 print(AD_NA_Sales)
218
219
220 # Nie mogliśmy wykonać testu Schapiro ze względu na poniższe problemy
221 # - długość argumentu 'x' musi być pomiędzy 3 a 5000
222 #ST_Sales<-shapiro.test(Global_Sales)
223 #print(ST_Sales)
224 #ST_Year<-shapiro.test(Year)
225 #print(ST_Year)
226 #ST_Sales<-shapiro.test(Other_Sales)
227 #ST_Sales<-shapiro.test(JP_Sales)
228 #ST_Sales<-shapiro.test(NA_Sales)
229 #ST_Sales<-shapiro.test(EU_Sales)|
230

```

## 6 Test Mardia (6)

$H_0$ -hipoteza zerowa, wektor losowy złożony z danych ma rozkład normalny.

$H_A$ -hipoteza alternatywna, wektor losowy nie ma rozkładu normalnego.

Przyjmujemy taki sam poziom istotności jak w poprzednich testach.

Mardia test			
	Beta-hat	kappa	p-val
Year, Global Sales			
Skewness	303.2542	825205.149	0
Kurtosis	612.9963	9663.089	0
Year, Other Sales			
Skewness	585.515	1593283.93	0
Kurtosis	1026.503	16267.67	0
Year, JP Sales			
Skewness	128.1869	348817.940	0
Kurtosis	208.8988	3208.785	0
Year, EU Sales			
Skewness	354.0930	963546.20	0
Kurtosis	757.0341	11963.68	0
Year, NA Sales			
Skewness	354.9592	965903.1	0
Kurtosis	660.8681	10427.7	0

Otrzymano  $p = 0 < \alpha = 0,05$ , zatem wektor losowy nie ma rozkładu normalnego. Hipoteza zerowa została odrzucona, zatem obliczamy współczynnik korelacji rang.

Kod w R:

```

232 # zadanie 6
233 #Mardia test
234 dataMT<-data.frame(Year,Global_Sales)
235 print(mult.norm(dataMT)$mult.test)
236 dataMT_Other_Sales<-data.frame(Year,Other_Sales)
237 print(mult.norm(dataMT_Other_Sales)$mult.test)
238 dataMT_JP_Sales<-data.frame(Year,JP_Sales)
239 print(mult.norm(dataMT_JP_Sales)$mult.test)
240 dataMT_EU_Sales<-data.frame(Year,EU_Sales)
241 print(mult.norm(dataMT_EU_Sales)$mult.test)
242 dataMT_NA_Sales<-data.frame(Year,NA_Sales)
243 print(mult.norm(dataMT_NA_Sales)$mult.test)
244

```

## 7 Współczynnik korelacji rang Spearmana (8)

$H_0$ -hipoteza zerowa,  $r_s = 0$ .

$H_A$ -hipoteza alternatywna,  $r_s \neq 0$ .

Im bliżej współczynnik zbliża się do 1, tym bardziej powiązane są zmienne. Wartość teoretyczna testu Spearmana wyniosła u nas około 0.07... (miałem problem, by otrzymać dokładną liczbę dla tak wielu próbek)

Dla sprzedarzy w Japoni:

Wynik testu Spearmana:

$$S = 7.1842e + 11$$

$$p - value = 0.2197$$

Otrzymano współczynnik korelacji rang Spearmana  $\rho = 0.0096053322 \in (-1; 1)$ .

**Oznacza to, że dla zmiennych Year oraz JP\_Sales nie odrzucamy hipotezy zerowej! (ponieważ  $p - value > \alpha$ )**

Dla pozostałych zmiennych nie udało się nam uzyskać  $p - value > \alpha = 0,05$ .

Dla zmiennej Global Sale:

Wynik testu Spearmana:

$$S = 8.351e + 11$$

$$p - value < 2.2e - 16$$

Otrzymano współczynnik korelacji rang Spearmana  $\rho = -0.1512482 \in (-1; 1)$ .

Dla sprzedarzy w NA:

Wynik testu Spearmana:

$$S = 8.2192e + 11$$

$$p - value < 2.2e - 16$$

Otrzymano współczynnik korelacji rang Spearmana  $\rho = -0.133088 \in (-1; 1)$ .

Dla sprzedarzy w Europy:

Wynik testu Spearmana:

$$S = 7.6726e + 11$$

$$p - value = 1.558e - 13$$

$$\rho : -0.05772944$$

Otrzymano współczynnik korelacji rang Spearmana  $\rho = -0.05772944 \in (-1; 1)$ .

Dla sprzedarzy na reszcie świata:

Wynik testu Spearmana:

$$S = 6.8496e + 11$$

$$p - value = 1.037e - 12$$

$$\rho : 0.05572604$$

Otrzymano współczynnik korelacji rang Spearmana  $\rho = 0.05572604 \in (-1; 1)$ .

Dla tych zmiennych Hipoteza alternatywna:  $\rho$  nie jest równe 0.

Test korelacji dla nich wyliczył wartość p, będącą mniejszą od założonej wartości  $\alpha$ . Oznacza to, że odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej.

Kod w R:

```

244
245 # Zadanie 8
246 #Spearman's rank correlation
247 SCT=cor.test(x=dataMT$Year, y=dataMT$Global_Sales, method='spearman', exact=FALSE)
248 print(SCT)
249 SCT_Other_Sales=cor.test(x=dataMT_Other_Sales$Year, y=dataMT_Other_Sales$Other_Sales, method='spearman', exact=FALSE)
250 print(SCT_Other_Sales)
251 SCT_JP_Sales=cor.test(x=dataMT_JP_Sales$Year, y=dataMT_JP_Sales$JP_Sales, method='spearman', exact=FALSE)
252 print(SCT_JP_Sales)
253 SCT_EU_Sales=cor.test(x=dataMT_EU_Sales$Year, y=dataMT_EU_Sales$EU_Sales, method='spearman', exact=FALSE)
254 print(SCT_EU_Sales)
255 SCT_NA_Sales=cor.test(x=dataMT_NA_Sales$Year, y=dataMT_NA_Sales$NA_Sales, method='spearman', exact=FALSE)
256 print(SCT_NA_Sales)
257

```

## 8 Test Chi-kwadrat oraz współczynnik kontyngencji V Cramera (11)

Zakładamy poziom istotności  $\alpha=0,05$ . Wartość teoretyczna testu Chi-kwadrat wyniosła u nas około 16482...

$H_0$ -hipoteza zerowa, dwie zmienne klasyfikujące są wzajemnie zależne.

$H_A$ -hipoteza alternatywna, dwie zmienne klasyfikujące nie są wzajemnie zależne.

Wynik testu Chi-Squared (Reszta świata):

$X - squared = 5758.6$

$df = 5928$

$p - value = 0.9412$

**Oznacza to, że dla zmiennych Year oraz Other\_Sales nie odrzucamy hipotezy zerowej! (ponieważ  $p - value > \alpha$ )** Jest to prawdopodobne, ponieważ regiony mniej rozwinięte, mogą mieć mniejsze możliwości na zakup najnowszych gier.

Dla pozostałych zmiennych:

Test chi-kwadrat wyliczył wartość  $p < 2.2e - 16$ . Odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej.

Wynik testu Chi-Squared (Cały Świat):

$X - squared = 49801$

$df = 23560$

$p - value < 2.2e - 16$

Wynik testu Chi-Squared (Japonia):

$X - squared = 47852$

$df = 9234$

$p - value < 2.2e - 16$

Wynik testu Chi-Squared (Europa):

$X - squared = 15680$

$df = 11552$

$p - value < 2.2e - 16$

Wynik testu Chi-Squared (NA):

$df = 15466$

$p - value < 2.2e - 16$

Współczynnik kontyngencji V Cramera wyniósł u nas  $X - squared = 36672$  0.283317 dla całego świata, natomiast dla poszczególnych regionów:

- Japonia: 0.2777191
- Europa: 0.1589726
- NA: 0.2431211
- Reszta świata: 0.09634113

Współczynnik V Cramera daje w wyniku wartości pomiędzy 0 a 1. Im wynik jest bliżej 0, tym słabszy jest związek między badanymi cechami, a im bliżej jest 1, tym silniejszy jest związek między badanymi cechami.

Stwierdza się więc, że powiązanie między zmiennymi jest na niskim poziomie, ale nie zerowe. Nie należy jednak na tej podstawie wyciągać zbyt daleko idących wniosków.

Kod w R:

```
258 # Zadanie 11
259 #Chi-squared test
260 CT=chisq.test(x=Year,y=Global_Sales)
261 print(CT)
262 CT_Other_Sales=chisq.test(x=Year,y=Other_Sales)
263 print(CT_Other_Sales)
264 CT_JP_Sales=chisq.test(x=Year,y=JP_Sales)
265 print(CT_JP_Sales)
266 CT_EU_Sales=chisq.test(x=Year,y=EU_Sales)
267 print(CT_EU_Sales)
268 CT_NA_Sales=chisq.test(x=Year,y=NA_Sales)
269 print(CT_NA_Sales)|
270
271 #Zadanie 11b
272 cramersV(x=Year,y=Global_Sales)
273 cramersV(x=Year,y=Other_Sales)
274 cramersV(x=Year,y=JP_Sales)
275 cramersV(x=Year,y=EU_Sales)
276 cramersV(x=Year,y=NA_Sales)
```

269:19 | (Top Level) ↕