Zaawansowane Modele Liniowe - Lista 2

Klaudia Weigel

1 Zadanie 1

Mamay zbiór danych zawierający informacje, o liczbie klientów przychodzących do pewnego sklepu w okresie około trzech miesięcy.

```
df_sklep = read.csv("sklep")
df_sklep = df_sklep[,2:5]
df_sklep[1:3,]
```

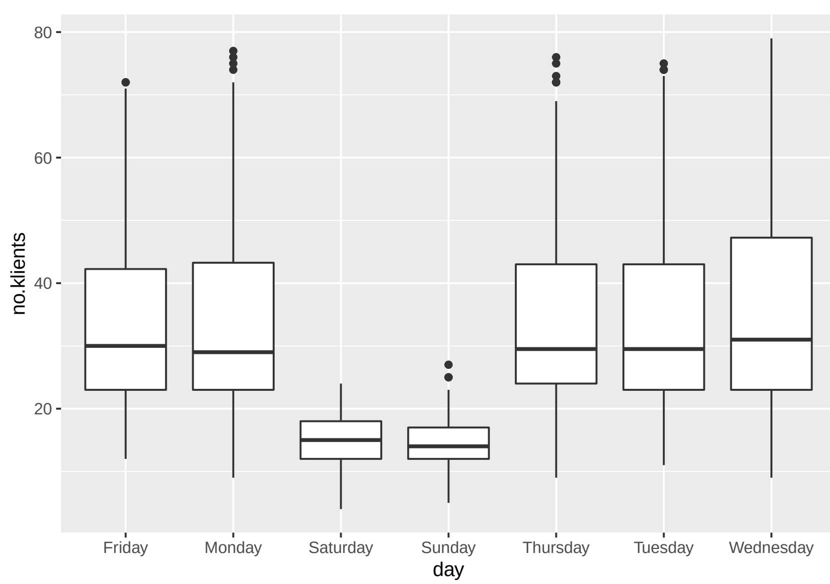
```
##   hour    day events no.klients
## 1    8 Monday      0         26
## 2    9 Monday      0         37
## 3   10 Monday      0         36
```

Dane przeanalizujemy za pomocą reresji Poissona traktując liczbę obsłużonych klientów jako zmienną objaśnianą, a pozostałe zmienne jako potencjalne predyktory.

2 Zadanie 2

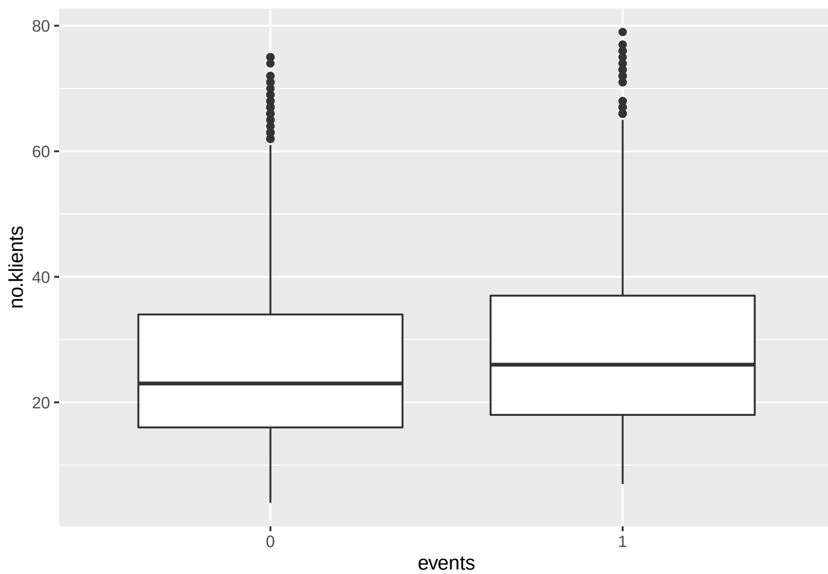
Spójrzmy najpierw na boxploty zmiennej objaśnianej w zależności od każdego z predyktorów oddzielnie.

2.1 Day



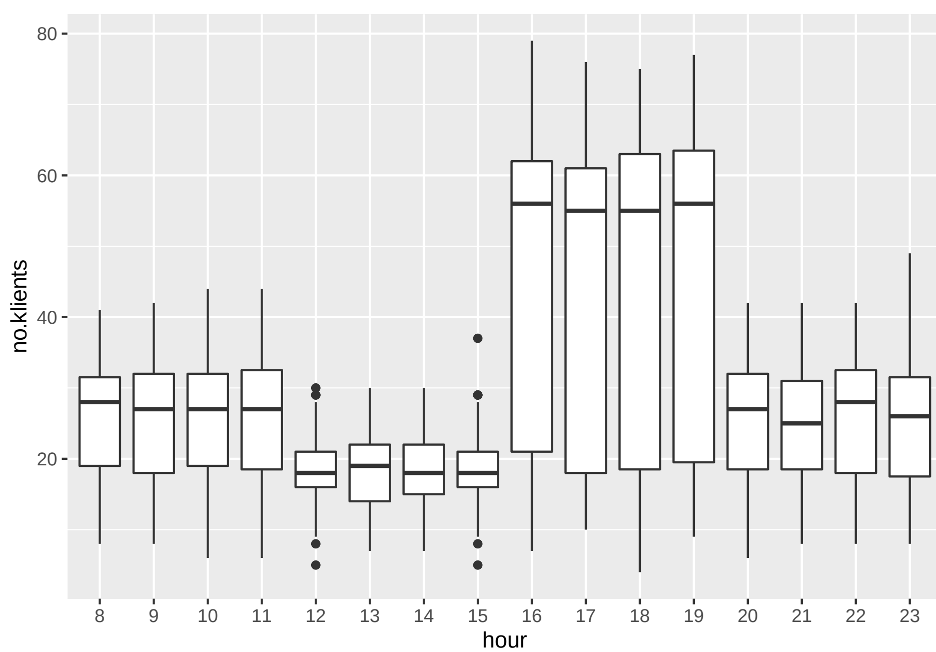
Widzimy, że w sobotę oraz niedzielę liczba klientów jest zdecydowanie niższa niż w pozostałych dniach. Mediana w obu przypadkach wynosi około 15, zatem połowa wartości dla tych dni jest mniejsza niż 15. Rozkłady dla dni roboczych sa bardzo podobne i są niesymetryczne. Sugeruje to, że mamy do czynienia ze skośnych rozkładem. Ponieważ dużo więcej wartości jest powyżej mediany to rozkład jest prawostronnie skośny.

2.2 Events



Zróżnicowanie ilości klientów ze względu na zmienną *events* jest bardzo nieznaczne. Prawdopodobnie ta zmienna nie będzie istotnie wpływać na zmienną objaśnianą.

2.3 Hour



Możemy podzielić dane ze względu na cztery przedziały czasowe $[8, 11]$, $[12, 15]$, $[16, 19]$, oraz $[20, 23]$. Najmniej klientów jest w godzinach $[12, 15]$. Maksymalna ilość klientów jest osiągana dla godzin $[16, 19]$. Rozkład dla tych godzin jest też mocno asymetryczny, większeość obserwacji leży poniżej mediany.

2.4 Wykresy

Przeanalizujemy teraz związek pomiędzy zmienną objaśnianą i predyktorami przy użyciu dodatkowych wykresów.

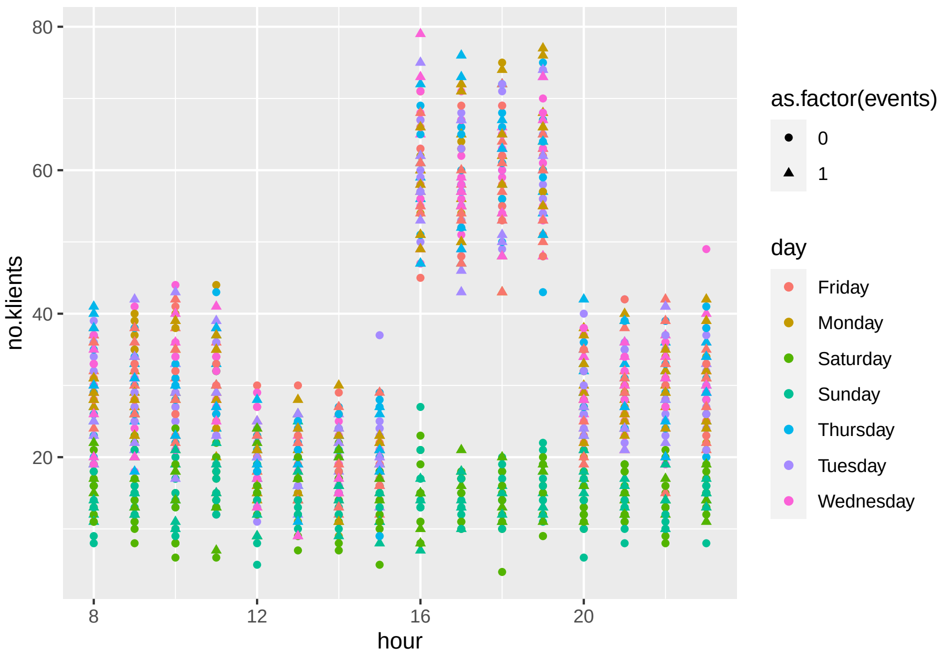


Figure 1: Zależność y od godziny w rozbiciu na *dzień* i *wydarzenie sportowe*.

Z wykresu widzimy, że w weekend do sklepu przychodzi najmniej klientów. Dane dla tych dni rozłożone są równomiernie ze względu na godziny.

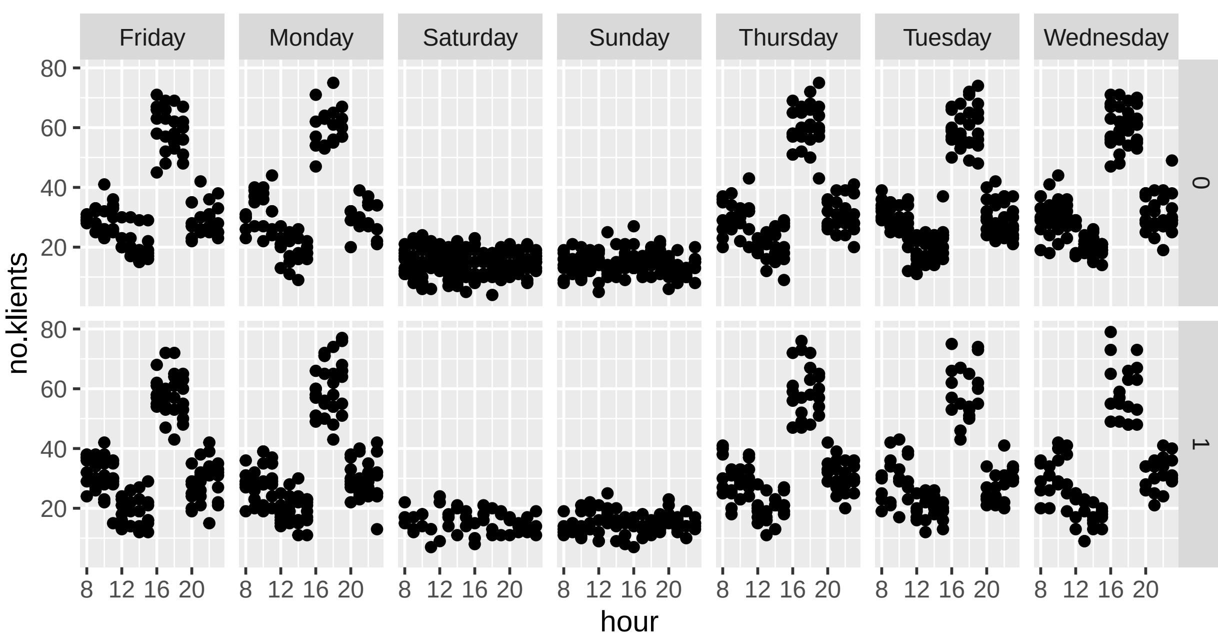


Figure 2: Rozbicie na podgrupy ze względu na *wydarzenie sportowe*.

Rozproszenie danych jest bardzo podobne pomiędzy różnymi dniami roboczymi, jak i pomiędzy dniami weekendowymi. Nie widać znaczących róznic ze względu na zmienną *events*.

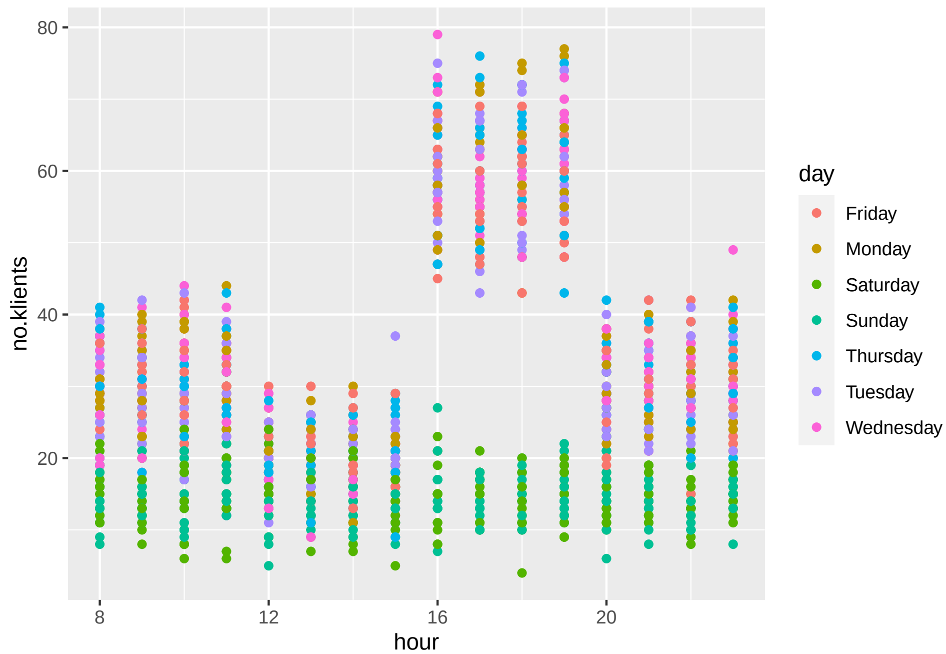
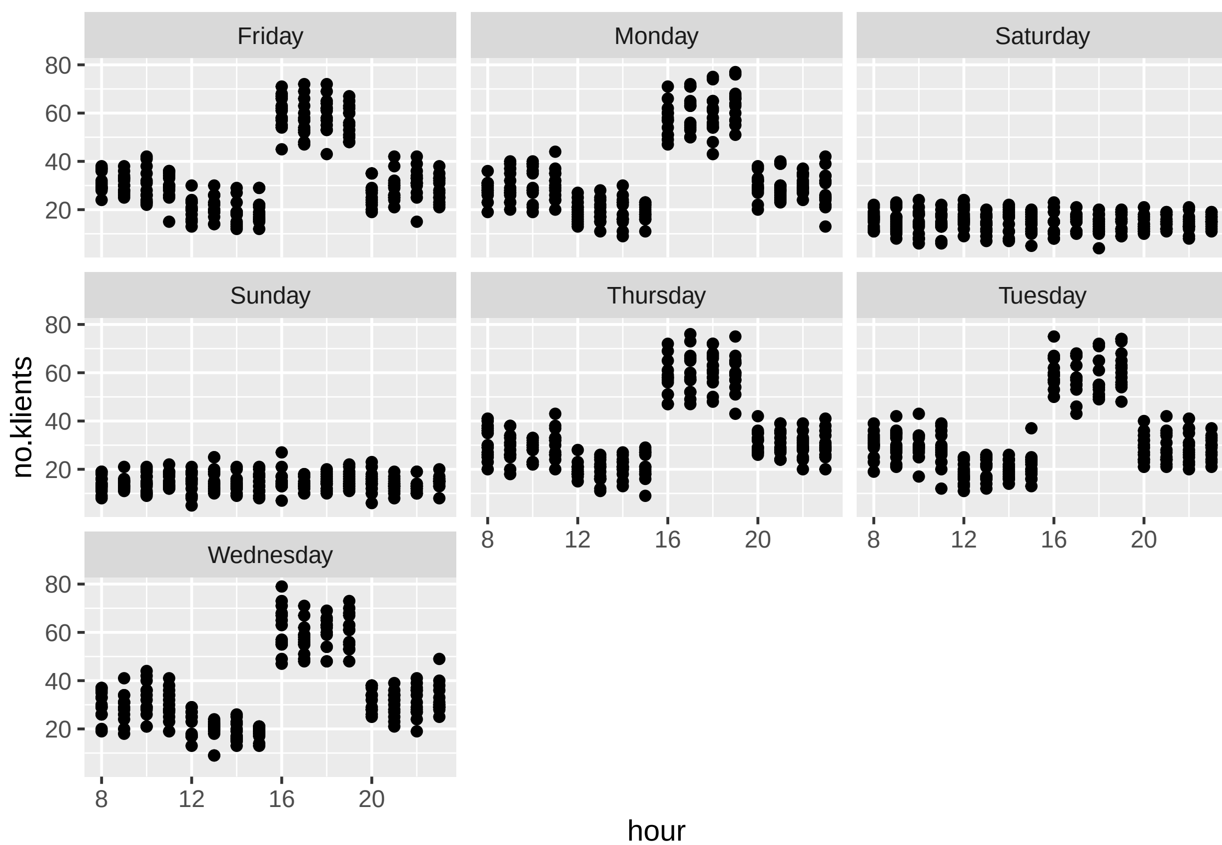


Figure 3: Zależność y od godziny w rozbiciu na *dzień tygodnia*.



Z powyższych wykresów widać, że wśród dni roboczych możemy wyraźnie wyróżnić cztery grupy, te same które zaobserwowaliśmy już przy boxplocie $y \sim hour$, czyli $[8, 11]$, $[12, 15]$, $[16, 19]$, oraz $[20, 23]$. Sugeruje to, że zamiast analizować dane ze względu na każdą z godzin osobno możemy podzielić dane na cztery grupy godzinowe. W dni weekendowe rozkład klientów jest jednostajny i nie ma istotnej różnicy pomiędzy sobotą i niedzielą.

3 Zadanie 3

Regresja Poissona zakłada, że zmienne objaśniane y_i sa realizacjami niezależnych zmiennych losowych z rozkładu Poissona ze średnią λ_i.

$$\log(\lambda_i) = \beta_0 + x_{i,1}\beta_1 + \cdots + x_{i,p-1}\beta_{p-1}.$$

3.1 Model z podwójnymi interakcjami

Rozważymy najpierw model zawierający jako zmienne pojedyncze predyktory oraz podwójne interakcje pomiędzy nimi (bez interakcji pomiędzy wszystkimi trzema predyktorami).

```
df_sklep_factors = df_sklep
df_sklep_factors$day = as.factor(df_sklep_factors$day)
df_sklep_factors$events = as.factor(df_sklep_factors$events)
df_sklep_factors$hour = as.factor(df_sklep_factors$hour)

m1_2 = glm(no.klients ~ day*events + hour*day + hour*events, data = df_sklep_factors,
           family = "poisson")
```

Ilość zmiennych możemy łatwo obliczyć znając liczbę stanów każdego regresora. Każdy dodatkowy predyktor przyjumujący k wartości zwiększy liczbę zmiennych o $k-1$. Ilość zmiennych odpowiadających interakcji jest iloczynem ilości zmiennych odpowiadających każdemu składnikowi interakcji. Ponieważ predyktorów mamy 3 - *events*, *hour* oraz *day*, przyjmujących odpowiednio 2, 16 i 7 różnych wartości to ilość zmiennych (oprócz interceptu) wynosi:

$$1 + 6 + 15 + 1 * 6 + 1 * 15 + 6 * 15 = 133.$$

Dodając intercept, dostajemy 134 zmienne.

Dana zmienna jest istotna jeżeli odpowidający jej współczynnik regresji jest różny od zera. Istotność współczynników regresji badamy za pomocą statystyki:

$$T_i = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}.$$

Przy H_0 statystyka ma asymptotycznie rozklad $N(0,1)$. Odrzucamy hipotezę zerową gdy wartość $|T_i|$ jest większa od kwantyla rzędu $1-\alpha/2$ z rozkładu normalnego $N(0,1)$, gdzie α jest ustalonym poziomem istotności. W zadaniach bedziemy rozważać $\alpha = 0.05$.

Sprawdźmy ile współczynników jest istotnych dla poziomu istotności 0.05.

```
nrow(summary(m1_2)$coef[summary(m1_2)$coef[,4] <= 0.05,])
```

```
## [1] 30
```

Spośród 134 współczynników, 30 jest istotnych.

Table 1: Tabela regresji Poissona (podwójne interakcje), tylko istotne współczynniki.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.4438728	0.0557087	61.819303	0.0000000
daySaturday	-0.6863823	0.0876124	-7.834309	0.0000000

	Estimate	Std. Error	z value	Pr(>|z|)
daySunday	-0.8252589	0.0914279	-9.026336	0.0000000
hour12	-0.4125183	0.0865061	-4.768663	0.0000019
hour13	-0.3976116	0.0857165	-4.638684	0.0000035
hour14	-0.5089965	0.0883271	-5.762630	0.0000000
hour15	-0.4400057	0.0873782	-5.035647	0.0000005
hour16	0.6645594	0.0672830	9.877080	0.0000000
hour17	0.6279190	0.0676710	9.278997	0.0000000
hour18	0.6720847	0.0673497	9.979032	0.0000000
hour19	0.5882147	0.0678951	8.663586	0.0000000
daySaturday:hour12	0.4769619	0.1265038	3.770337	0.0001630
daySunday:hour12	0.4282533	0.1317429	3.250676	0.0011513
daySaturday:hour13	0.2621523	0.1294571	2.025013	0.0428660
daySunday:hour13	0.5279855	0.1282979	4.115308	0.0000387
daySaturday:hour14	0.4812752	0.1291272	3.727139	0.0001937
daySunday:hour14	0.5278856	0.1323509	3.988530	0.0000665
daySaturday:hour15	0.3690918	0.1298651	2.842118	0.0044815
daySunday:hour15	0.5047653	0.1315935	3.835792	0.0001252
daySaturday:hour16	-0.7707284	0.1189548	-6.479170	0.0000000
daySunday:hour16	-0.5461296	0.1191513	-4.583498	0.0000046
daySaturday:hour17	-0.6412689	0.1171575	-5.473562	0.0000000
daySunday:hour17	-0.6070263	0.1214696	-4.997350	0.0000006
daySaturday:hour18	-0.8126462	0.1200392	-6.769840	0.0000000
daySunday:hour18	-0.5780717	0.1200739	-4.814299	0.0000015
dayMonday:hour19	0.2165046	0.0879449	2.461822	0.0138233
daySaturday:hour19	-0.6743823	0.1187003	-5.681388	0.0000000
daySunday:hour19	-0.4706907	0.1192880	-3.945834	0.0000795
dayMonday:hour20	0.2088813	0.1029688	2.028588	0.0425003
dayWednesday:hour20	0.2315919	0.1016440	2.278462	0.0226991

Ani jedna istotna zmienna nie zależy od *events*, mimo że łączna ilość zmiennych zależnych of tego regresora to $1 + 6 + 15 = 23$. Możemy sprawdzić, czy regresor *events* wnosi coś do modelu za pomocą testu chi-kwadrat opartego na statystyce deviance. Hipoteza zerowa odpowiada modelowi zredukowanemu (M_R), w którym niektóre współczynniki zostały wyzerowane, a hipoteza alternatywna związana jest z modelem bez żadnych warunków (M_F) czyli

$$H_0 : \forall(i \in A)\ \beta_i = 0 \quad H_1 : \exists(i \in A)\ \beta_i \neq 0.$$

Statystyka testowa to

$$\chi^2 = D(M_R) - D(M_F).$$

Przy hipotezie zerowej statystyka ma rozkład χ^2 ze stopniami swobody równymi ilości wyzerowanych współczynników w modelu zredukowanym ($df = |A|$).

```
m1_noevents <- glm(no.klients ~ hour*day, data = df_sklep_factors, family = "poisson")
```

```
anova(m1_noevents, m1_2, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: no.klients ~ hour * day
## Model 2: no.klients ~ day * events + hour * day + hour * events
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1344     1475.7
## 2      1322     1457.7 22   18.017    0.705
```

P-wartość jest większa niż ustalony poziom istotności, nie mamy zatem podstaw aby odrzucić hipotezę zerową. Przyjmujemy model zredukowany bez regresora *events*.

3.2 Model z potrójnymi interakcjami

Poszerzymy teraz model rozważany w poprzednim punkcie o potrójne interakcje pomiędzy wszystkimi trzema regresorami.

Ilość zmiennych w tym przypadku wynosi:

$$1 + 6 + 15 + 1 * 6 + 1 * 15 + 6 * 15 + 1 * 6 * 15 = 223.$$

Nasz model ma 223 współczynniki, 224 włącznie z interceptem.

Istotność współczynników regresji badamy tak samo jak wyżej, za pomocą statystki

$$T_i = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}.$$

```
m1_3 = glm(no.klients ~ day*events*hour, data = df_sklep_factors, family = "poisson")
```

Sprawdźmy ile współczynników jest istotnych dla poziomu istotności 0.05.

```
nrow(summary(m1_3)$coef[summary(m1_3)$coef[,4] <= 0.05,])
```

```
## [1] 34
```

Spośród 224 współczynników, tylko 34 są istotne.

Table 2: Tabela regresji Poissona, tylko istotne zmienne.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	3.3787245	0.0753778	44.823846	0.0000000
daySaturday	-0.6443570	0.1103419	-5.839643	0.0000000
daySunday	-0.7307782	0.1205461	-6.062228	0.0000000
hour12	-0.2801349	0.1148939	-2.438204	0.0147604
hour13	-0.3106716	0.1159041	-2.680420	0.0073530
hour14	-0.3997994	0.1189806	-3.360205	0.0007788
hour15	-0.3746934	0.1180944	-3.172831	0.0015096
hour16	0.7430190	0.0915670	8.114482	0.0000000
hour17	0.7016338	0.0921885	7.610860	0.0000000
hour18	0.7128384	0.0920181	7.746718	0.0000000
hour19	0.6701577	0.0926758	7.231208	0.0000000
dayTuesday:events1	-0.3765239	0.1469301	-2.562606	0.0103890
daySaturday:hour12	0.3430487	0.1605979	2.136072	0.0326735
dayTuesday:hour12	-0.3867444	0.1562380	-2.475355	0.0133104
daySaturday:hour14	0.3734821	0.1652772	2.259731	0.0238380
daySunday:hour14	0.3728919	0.1791560	2.081382	0.0373990
daySunday:hour15	0.4835875	0.1753064	2.758527	0.0058062
daySaturday:hour16	-0.7495337	0.1463351	-5.122035	0.0000003
daySunday:hour16	-0.5651321	0.1569761	-3.600116	0.0003181
daySaturday:hour17	-0.7346401	0.1473215	-4.986644	0.0000006
daySunday:hour17	-0.7194914	0.1623493	-4.431750	0.0000093
daySaturday:hour18	-0.8519512	0.1497448	-5.689352	0.0000000
daySunday:hour18	-0.6039443	0.1589132	-3.800467	0.0001444
daySaturday:hour19	-0.7442656	0.1485776	-5.009272	0.0000005

	Estimate	Std. Error	z value	Pr(>|z|)
daySunday:hour19	-0.4997057	0.1578006	-3.166690	0.0015418
dayTuesday:events1:hour11	0.6099298	0.2077120	2.936420	0.0033202
dayTuesday:events1:hour12	0.6929291	0.2334871	2.967740	0.0030000
daySunday:events1:hour13	0.5083964	0.2592082	1.961343	0.0498390
dayTuesday:events1:hour14	0.6103763	0.2316257	2.635184	0.0084092
dayTuesday:events1:hour16	0.4708986	0.1789005	2.632182	0.0084838
dayTuesday:events1:hour17	0.3558529	0.1806334	1.970028	0.0488351
dayTuesday:events1:hour19	0.4595039	0.1793810	2.561609	0.0104189
daySunday:events1:hour20	0.6145479	0.2553729	2.406473	0.0161074
dayTuesday:events1:hour23	0.5010963	0.2081664	2.407191	0.0160757

Tym razem istotnych jest 10 zmiennych zależących od regresora *events* na poziomie istotności 0.05. Łącznie zmiennych zależnych od *events* jest $1 + 6 + 15 + 6 * 15 = 112$.

Możemy porównać model z modelami rozważanymi w poprzednim punkcie, czyli z modelem bez regresora *events* oraz modelem tylko z podwójnymi interakcjami.

```
anova(m1_noevents, m1_3, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: no.klients ~ hour * day
## Model 2: no.klients ~ day * events * hour
##   Resid. Df Resid. Dev  Df Deviance Pr(>Chi)
## 1      1344     1475.7
## 2      1232     1359.6 112   116.13   0.3755
```

```
anova(m1_2, m1_3, test="Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: no.klients ~ day * events + hour * day + hour * events
## Model 2: no.klients ~ day * events * hour
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      1322     1457.7
## 2      1232     1359.6 90   98.117    0.262
```

W obu przypadkach p-wartość jest większa od ustalonego poziomu istotności. Odrzucamy model z potrójnymi interakcjami.

4 Zadanie 4

Stworzymy dwie nowe zmienne. Pierwsza, opisująca to czy dzień jest dniem roboczym czy weekendowym. Druga grupująca godziny każdego dnia w bloki cztero-godzinne.

```
df_sklep_new = df_sklep
# workingday - 1 if the day is a working day
df_sklep_new$workingday = sapply(df_sklep_new$day,
                                 function(d) 1 - (d %in% c("Saturday", "Sunday"))*1 )
df_sklep_new$timeslot = cut(df_sklep_new$hour, c(8,12,16,20,23),
                            right = FALSE, include.lowest = TRUE)
df_sklep_new$timeslot = as.factor(df_sklep_new$timeslot)
df_sklep_new$workingday = as.factor(df_sklep_new$workingday)
```

```r
m2 = glm(no.klients~timeslot*workingday, data = df_sklep_new, family = "poisson")
```

Ilość współczynników w tym modelu to

```r
length(m2$coefficients)
```

```
## [1] 8
```

W tym ilość istotnych współczynników dla $\alpha = 0.05$:

```r
nrow(summary(m2)$coef[summary(m2)$coef[,4] <= 0.05,])
```

```
## [1] 4
```

Table 3: Tabela regresji Poissona dla modelu z pogrupowanymi danymi.

term	estimate	std.error	statistic	p.value	signif
(Intercept)	2.6938473	0.0254989	105.6455438	0.0000000	***
timeslot[12,16)	0.0109927	0.0359622	0.3056728	0.7598538	
timeslot[16,20)	0.0051881	0.0360142	0.1440563	0.8854560	
timeslot[20,23]	-0.0283567	0.0363193	-0.7807603	0.4349435	
workingday1	0.7076065	0.0278992	25.3629382	0.0000000	***
timeslot[12,16):workingday1	-0.4312423	0.0402067	-10.7256419	0.0000000	***
timeslot[16,20):workingday1	0.6817233	0.0385964	17.6628588	0.0000000	***
timeslot[20,23]:workingday1	0.0274591	0.0396932	0.6917822	0.4890741	

Sprawdzimy teraz czy nowy model statystycznie różni się od modelu z zadania 4. Ponieważ modele są zagnieżdżone, to do ich porównania możemy użyć testu chi-kwadrat opartego na statystyce deviance, opisanego w poprzednim zadaniu. Statystyka alternatywna odpowiada modelowi z zadania 3, rozważamy modele z podwójnymi i potrójnymi interakcjami.

```r
anova(m2, m1_2, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: no.klients ~ timeslot * workingday
## Model 2: no.klients ~ day * events + hour * day + hour * events
##   Resid. Df Resid. Dev  Df Deviance Pr(>Chi)
## 1      1448     1552.5
## 2      1322     1457.7 126   94.735   0.9829
```

```r
anova(m2, m1_3, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model 1: no.klients ~ timeslot * workingday
## Model 2: no.klients ~ day * events * hour
##   Resid. Df Resid. Dev  Df Deviance Pr(>Chi)
## 1      1448     1552.5
## 2      1232     1359.6 216   192.85   0.8694
```

W obu przypadkach p-wartość jest większa od ustalonego poziomu istotności, nie mamy zatem podstaw do odrzucnia hipotezy zerowej. Przyjmujemy prostszy model z pogrupowanymi i kategorycznymi danymi.

5 Zadanie 5

Model można zapisać następująco

$$log(\lambda_i) = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + \beta_5 x_{i,5} + \beta_6 x_{i,6} + \beta_7 x_{i,7},$$

gdzie

$x_{i,1} = 1$ jeśli i-ta obserwacja należy do przedziału czasowego $[12, 16), 0$ w przeciwnym przypadku,
$x_{i,2} = 1$ jeśli i-ta obserwacja należy do przedziału czasowego $[16, 20)$, 0 wpp,
$x_{i,3} = 1$ jeśli i-ta obserwacja należy do przedziału czasowego $[20, 23)$, 0 wpp,
$x_{i,4} = 1$ jesli i-ta obserwacja została zmierzona w dzień roboczy, 0 wpp,
$x_{i,5} = 1$ jeśli i-ta obserwacja należy do przedziału czasowego $[12, 16)$ i została zmierzona w dzień roboczy, 0 wpp,
$x_{i,6} = 1$ jeśli i-ta obserwacja należy do przedziału czasowego $[16, 20)$ i została zmierzona w dzień roboczy, 0 wpp,
$x_{i,7} = 1$ jeśli i-ta obserwacja należy do przedziału czasowego $[20, 23]$ i została zmierzona w dzień roboczy, 0 wpp,

Checemy znaleźć informacje dla grup

```r
s1 = expand.grid(c('0','1'), unique(df_sklep_new$timeslot))
colnames(s1) = c("workingday", "timeslot")
s1
```

```
##   workingday timeslot
## 1          0   [8,12)
## 2          1   [8,12)
## 3          0  [12,16)
## 4          1  [12,16)
## 5          0  [16,20)
## 6          1  [16,20)
## 7          0  [20,23]
## 8          1  [20,23]
```

Do znalezienia średniej i wartości η_i możemy użyć funkcji `predict`

```r
# Średnia dla danych grup
predict(m2, s1, type = "response")
```

```
##        1        2        3        4        5        6        7        8
## 14.78846 30.00769 14.95192 19.71154 14.86538 59.64231 14.37500 29.98077
```

```r
predict(m2, s1)
```

```
##        1        2        3        4        5        6        7        8
## 2.693847 3.401454 2.704840 2.981204 2.699035 4.088365 2.665491 3.400556
```

Table 4: Grupy.

	Dzień roboczy	Czas	Średnia	Postać η_i	Wartość η_i
1	0	[8,12)	14.7885	$\hat{\beta}_0$	2.6938
3	0	[12,16)	14.9519	$\hat{\beta}_0 + \hat{\beta}_1$	2.7048
5	0	[16,20)	14.8654	$\hat{\beta}_0 + \hat{\beta}_2$	2.6990
7	0	[20,23]	14.3750	$\hat{\beta}_0 + \hat{\beta}_3$	2.6655
2	1	[8,12)	30.0077	$\hat{\beta}_0 + \hat{\beta}_4$	3.4015
4	1	[12,16)	19.7115	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_4 + \hat{\beta}_5$	2.9812
6	1	[16,20)	59.6423	$\hat{\beta}_0 + \hat{\beta}_2 + \hat{\beta}_4 + \hat{\beta}_6$	4.0884
8	1	[20,23]	29.9808	$\hat{\beta}_0 + \hat{\beta}_3 + \hat{\beta}_4 + \hat{\beta}_7$	3.4006

Możemy zaobserwować, że średnia liczba klientów w dni weekendowe jest bardzo podobna, niezależnie od godziny. W dni robocze najwięcej klientów jest w godzinach $[16, 20)$.

6 Zadanie 6

Chcemy sprawdzić czy predyktory liniowe odpowidające podgrupom godzin weekendowych są takie same. Testujemy:

$$H_0 : \beta_0 = \beta_0 + \beta_1 = \beta_0 + \beta_2 = \beta_0 + \beta_3 \quad H_1 : \text{ Przynajmniej jedna równość nie zachodzi.}$$

Równoważnie

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \qquad H_1 : \exists_{i,j}\ \beta_i \neq \beta_j.$$

Użyjemy testu Walda. Zdefiniujmy macierz $A_{3\times 8}$:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Statystyka Walda dla naszego problemu testowania ma postać:

$$(A\hat{\beta})'(A\,\widehat{\mathrm{Cov}}(\hat{\beta})A')^{-1}(A\hat{\beta}).$$

Przy hipotezie zerowej rozkład statystyki zbiega do rozkladu χ^2 z trzema stopniami swobody. Test Walda odrzuca H_0 gdy statystyka W przyjmuje odpowiednio duże wartości. Dokładniej H_0 odrzucamy gdy W jest większe niż kwantyl rzędu $1 - \alpha$ z rozkładu χ^2 z trzema stopniami swobody. Za poziom istotności α obierzemy 0.05.

```
A = matrix(c(0,1,0,0,0,0,0,0,  0,0,1,0,0,0,0,0,  0,0,0,1,0,0,0,0), nrow = 3, byrow = T)
coefs_m2 = m2$coefficients
W = t(A%*%coefs_m2) %*% solve(A%*%vcov(m2)%*%t(A)) %*% (A%*%coefs_m2)
W > qchisq(1-0.05, df=3)
```

```
##       [,1]
## [1,] FALSE
```

Możemy też policzyć p-wartość.

```
1 - pchisq(W, df=3)
```

```
##           [,1]
## [1,] 0.7095304
```

P-wartość jest większa od ustalonego poziomu istotności, nie mamy zatem podstaw aby odrzucić hipotezę zerową. Uznajemy więc, że rozważane predyktory liniowe sa równe.

7 Zadanie 7

Na podstawie wyników tabeli z zadania 5 zaplanujemy grafik pracowników przyjmując, że jeden pracownik może obsłużyć 20 klientów na godzinę. Ponieważ grafik mam byc na podstawie tabeli to zakładamy, że jedna zmiana trwa 4 godziny i podział dnia na zmiany jest taki sam jak w tabeli. Przyjmiemy dodatkowo, że pracownicy nie mogą pracować więcej niż 8 godzin dziennie, oraz nie więcej niż 40 godzin w tygodniu.

Zobaczmy najpierw ilu pracowników musi pracować w danych godzinach dnia roboczego oraz weekendowego

Godzina	Pracownicy (roboczy)	Pracownicy (weekend)
[8, 12)	2	1
[12, 16)	1	1
[16, 20)	3	1
[20, 24)	2	1

Możemy teraz obliczyć ile minimalnie pracowników musi pracować w sklepie. W dniu roboczym mamy całkowitą ilość 8 zmian, natomiast w weekend całkowita ilosć zmian to 4. Łącznie w tygodniu mamy więc $5 * 8 + 4 * 2 = 48$ zmian. Jeden pracownik może pracować na nie więcej niż 10 zmianach. Potrzebujemy zatem minimalnie 5 pracowników. Wystarczy teraz ułożyć grafik dla 5 pracowników, spełniający wszystkie założenia. Oznaczmy pracowników przez A, B, C, D, E.

Godzina	Pon	Wt	Śr	Czw	Pt	Sob	Nd
[8, 12)	A, B	A, B	A, B	C, D	C, D	A	A
[12, 16)	E	E	E	C	C	A	A
[16, 20)	C, D, E	C, D, E	C, D, E	D, A, E	D, A, E	B	B
[20, 24)	C, D	C, D	C, D	A, E	B, E	B	B