analaysis for categorical data

kiana mirbaghestan

2022-07-12

داده های یک مطالعه در انگلستان در دو دوره از نوامبر 1969 تا اکتبر 1971 و نوامبر 1971 تا اکتبر 1973را در نظر گرفتیم. یک اقدام ایمنی اجباری جدید برای کامیون ها در اکتبر 1971 معرفی شد.پس، سؤال این است که آیا این اقدام ایمنی بر تعداد تصادفات کامیون ها اثر داشته یانه.

24مشاهده و 5متغیر داریم.

منبع داده ها:

E. B. Andersen (1991), The Statistical Analysis of Categorical Data, Table 6.8.

رفرنس دیتاها:

E. B. Andersen (1991), The Statistical Analysis of Categorical Data. 2nd edition. Springer-Verlag, Berlin.

خودم هم با کمک ویدیوهای تی ای و لینک زیر پروژه را به اتمام رساندم:

<https://youtu.be/j9YFazcAjB4>

طبقه بندی قبل و بعد از 1971:A

تصادم در عقب ماشین یا جلوی ان بوده است:B

کامیون در پارک بوده یا خیر:C

شرایط نوری (روشنایی روز،شب در جاده دارای روشنایی،شب در جاده بدون روشنایی): D

پکیج زیر را برای فراخوانی داده های رسته ای فراخوانی میکنیم:

##installing vcd package for categorical dataset ##find data type

library(vcd)

## Loading required package: grid

Trucks

## Freq period collision parked light  
## 1 712 before back yes daylight  
## 2 613 after back yes daylight  
## 3 192 before forward yes daylight  
## 4 179 after forward yes daylight  
## 5 2557 before back no daylight  
## 6 2373 after back no daylight  
## 7 10749 before forward no daylight  
## 8 9768 after forward no daylight  
## 9 634 before back yes night, illuminate  
## 10 411 after back yes night, illuminate  
## 11 95 before forward yes night, illuminate  
## 12 55 after forward yes night, illuminate  
## 13 325 before back no night, illuminate  
## 14 283 after back no night, illuminate  
## 15 1256 before forward no night, illuminate  
## 16 987 after forward no night, illuminate  
## 17 345 before back yes night, dark  
## 18 179 after back yes night, dark  
## 19 46 before forward yes night, dark  
## 20 39 after forward yes night, dark  
## 21 579 before back no night, dark  
## 22 494 after back no night, dark  
## 23 1018 before forward no night, dark  
## 24 885 after forward no night, dark

با استفاده از دستور زیر نوع داده مان مشخص می شود:

class(Trucks)

## [1] "data.frame"

جدول داده ها در صفحه دیگر باز می شود:

View(Trucks)  
str(Trucks)

## 'data.frame': 24 obs. of 5 variables:  
## $ Freq : num 712 613 192 179 2557 ...  
## $ period : Factor w/ 2 levels "before","after": 1 2 1 2 1 2 1 2 1 2 ...  
## $ collision: Factor w/ 2 levels "back","forward": 1 1 2 2 1 1 2 2 1 1 ...  
## $ parked : Factor w/ 2 levels "yes","no": 1 1 1 1 2 2 2 2 1 1 ...  
## $ light : Factor w/ 3 levels "daylight","night, illuminate",..: 1 1 1 1 1 1 1 1 2 2 ...

جدول توافقی چندبعدی مان:

##Agreement table

Trucks\_data<-ftable(Trucks)  
Trucks\_data

## light daylight night, illuminate night, dark  
## Freq period collision parked   
## 39 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 1  
## no 0 0 0  
## 46 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 1  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 55 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 1 0  
## no 0 0 0  
## 95 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 1 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 179 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 1  
## no 0 0 0  
## forward yes 1 0 0  
## no 0 0 0  
## 192 before back yes 0 0 0  
## no 0 0 0  
## forward yes 1 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 283 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 1 0  
## forward yes 0 0 0  
## no 0 0 0  
## 325 before back yes 0 0 0  
## no 0 1 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 345 before back yes 0 0 1  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 411 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 1 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 494 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 1  
## forward yes 0 0 0  
## no 0 0 0  
## 579 before back yes 0 0 0  
## no 0 0 1  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 613 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 1 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 634 before back yes 0 1 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 712 before back yes 1 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 885 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 1  
## 987 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 1 0  
## 1018 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 1  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 1256 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 1 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 2373 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 1 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 2557 before back yes 0 0 0  
## no 1 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## 9768 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 1 0 0  
## 10749 before back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 1 0 0  
## after back yes 0 0 0  
## no 0 0 0  
## forward yes 0 0 0  
## no 0 0 0

جدول توافقی های دو بعدی را تشکیل میدهیم:

##Tables

Freq\_period<-table(Trucks$Freq,Trucks$period)  
Freq\_collision<-table(Trucks$Freq,Trucks$collision)  
Freq\_parked<-table(Trucks$Freq,Trucks$parked)  
Freq\_light<-table(Trucks$Freq,Trucks$light)  
period\_collision<-table(Trucks$period,Trucks$collision)  
period\_parked<-table(Trucks$period,Trucks$parked)  
period\_light<-table(Trucks$period,Trucks$light)  
collision\_parked<-table(Trucks$collision,Trucks$parked)  
collision\_light<-table(Trucks$collision,Trucks$light)  
parked\_light<-table(Trucks$parked,Trucks$light)

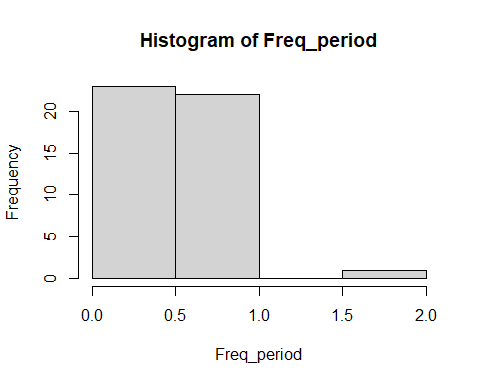
نمودار هیستوگرام برای داده ها:

نمودار هیستوگرام، مجموعه‌ای از ستون‌های در کنار هم است که ارتفاع هر ستون با سایرین تفاوت بوده و این ارتفاع در هر ستون بیانگر میزان فراوانی دسته آن ستون است. نمودارهای ستونی را می‌توان پرکاربردترین و پر استفاده‌ترین انواع نمودارها در علم ریاضی و آمار دانست. در صنایع مختلف هم از این نمودارها برای تعریف و دسته بندی داده‌ها استفاده می‌شود.

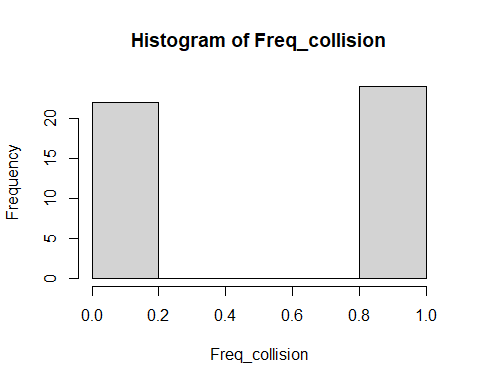
histogram در واقع یک نمودار ستونی است؛ در این نمودار هر محور بیانگر اطلاعاتی از داده‌ها است. برای مثال در یک دسته بندی از داده‌های آماری، هر ستون را می‌توان به عنوان یک دسته از این داده‌ها در نظر گرفت. ارتفاع هر ستون نیز بیانگر میزان فراوانی هر دسته در داده‌های موجود است.

##Histogram

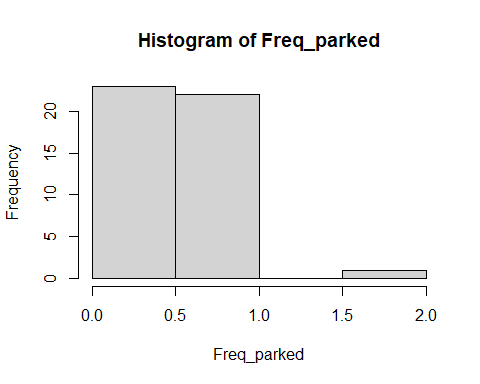
hist(Freq\_period)



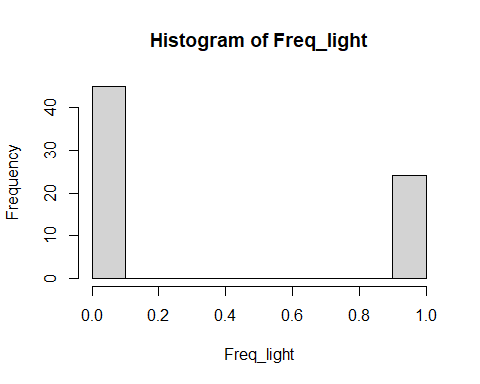
hist(Freq\_collision)



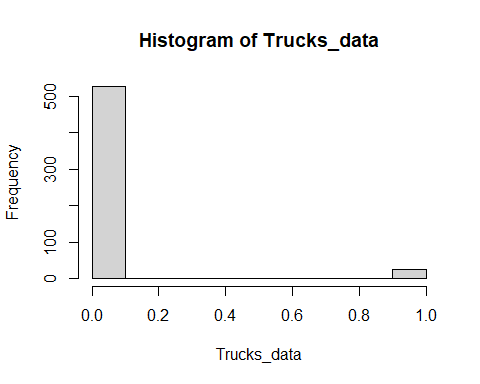
hist(Freq\_parked)



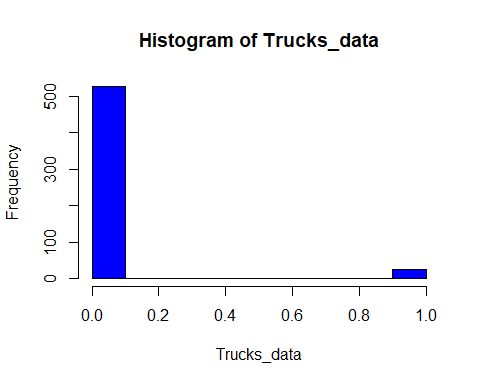
hist(Freq\_light)



hist(Trucks\_data)



hist(Trucks\_data,col="blue")



برای احتمال سلول ها داریم:(دو به دو در مقابل هم)

##Prob tables

prop.table(period\_collision)

##   
## back forward  
## before 0.25 0.25  
## after 0.25 0.25

prop.table(period\_parked)

##   
## yes no  
## before 0.25 0.25  
## after 0.25 0.25

prop.table(period\_light)

##   
## daylight night, illuminate night, dark  
## before 0.1666667 0.1666667 0.1666667  
## after 0.1666667 0.1666667 0.1666667

prop.table(collision\_parked)

##   
## yes no  
## back 0.25 0.25  
## forward 0.25 0.25

prop.table(collision\_light)

##   
## daylight night, illuminate night, dark  
## back 0.1666667 0.1666667 0.1666667  
## forward 0.1666667 0.1666667 0.1666667

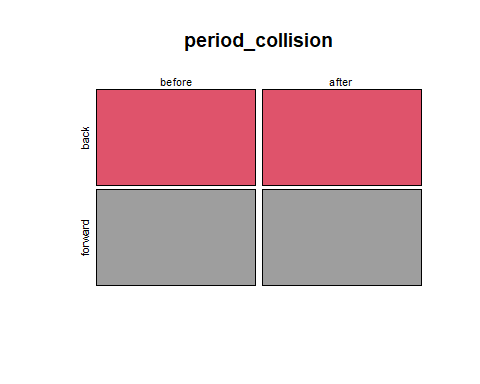
prop.table(parked\_light)

##   
## daylight night, illuminate night, dark  
## yes 0.1666667 0.1666667 0.1666667  
## no 0.1666667 0.1666667 0.1666667

نمودارهای موزائیکی را به صورت زیر می کشیم:

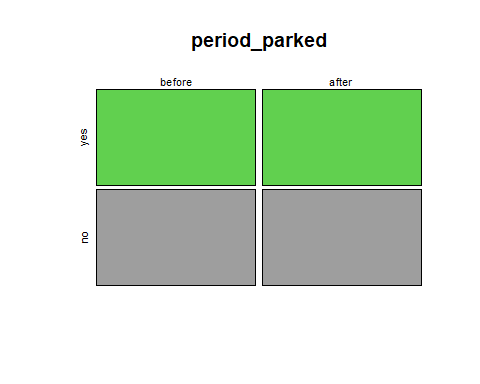
##mosaic plot and qplot

mosaicplot(period\_collision,color=c(2,8))

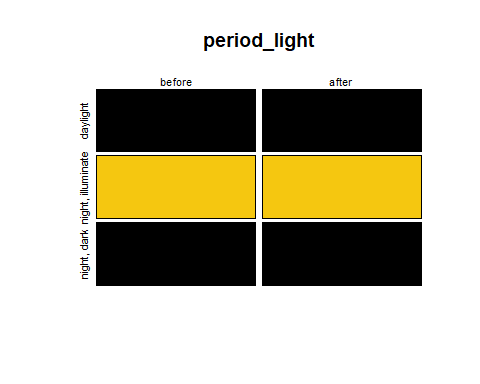


که نمودار بالا نشان می دهد کامیون هایی که قبل و بعد قانون گذاری هستند تعدادشان دقتی از جلو یا عقب تصادف می کنند تقریبا برابر هساتد.(در بقیه نمودار هم باتوجه به متغیرهای دیگر همین تفسیر را دارا هستند.)

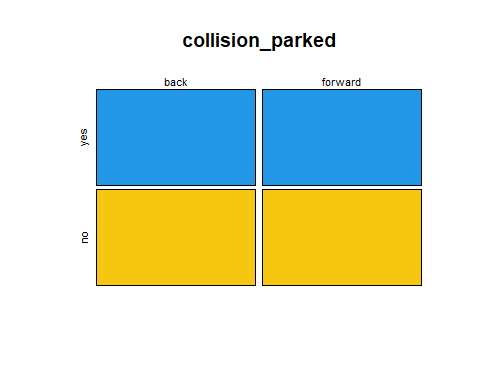
mosaicplot(period\_parked,color=c(3,8))



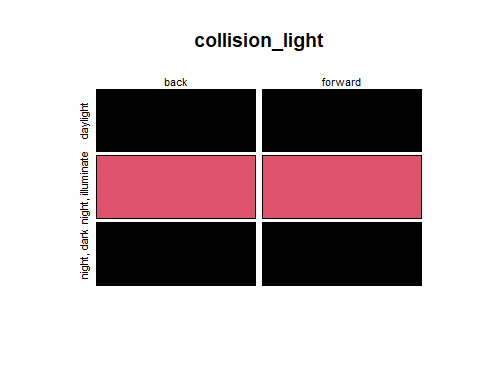
mosaicplot(period\_light,color=c(1,7))



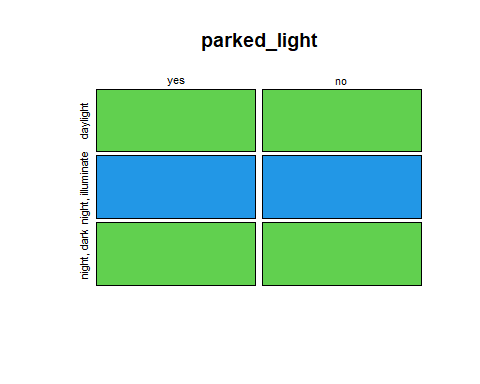
mosaicplot(collision\_parked,color=c(4,7))



mosaicplot(collision\_light,color=c(1,10))



mosaicplot(parked\_light,color=c(11,12))

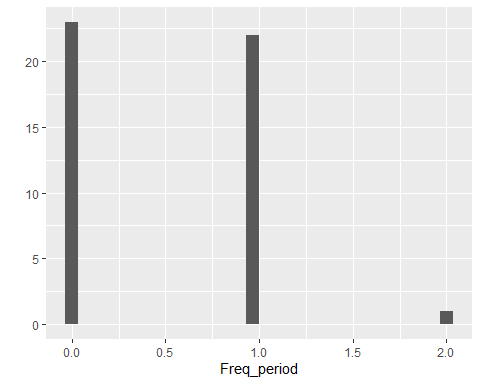


برای نمودارهای ستونی هم دستورات زیر را اجرا میکنیم: (برای متغیرهای عددی در برابر یک متغیر دیگر بکار می رود)

library(ggplot2)  
qplot(Freq\_period)

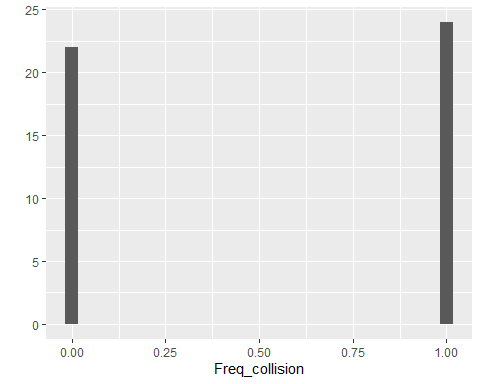
## Don't know how to automatically pick scale for object of type table. Defaulting to continuous.

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



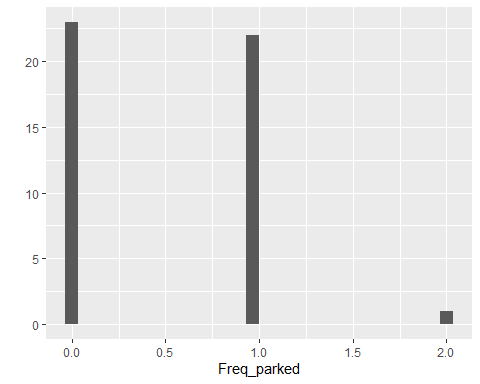
qplot(Freq\_collision)

## Don't know how to automatically pick scale for object of type table. Defaulting to continuous.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



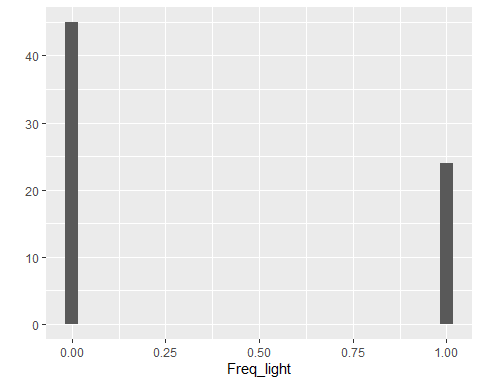
qplot(Freq\_parked)

## Don't know how to automatically pick scale for object of type table. Defaulting to continuous.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



qplot(Freq\_light)

## Don't know how to automatically pick scale for object of type table. Defaulting to continuous.  
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



طبق نمودارهای پایینتر متوجه میشویم خروجی این پکیج ها مثل نمودارهای بالا است تنها برای مصور سازی و زیبایی از پکیج زیر استفاده میکنیم:

##ggmosaic and ggplot

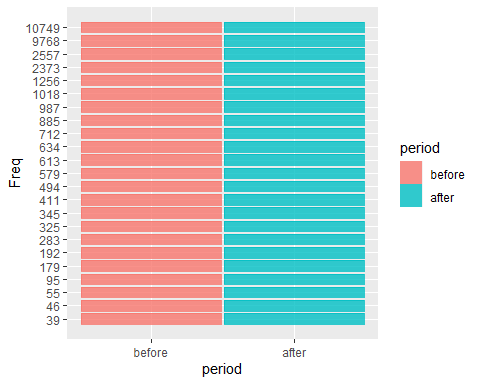
library(ggmosaic)

##   
## Attaching package: 'ggmosaic'

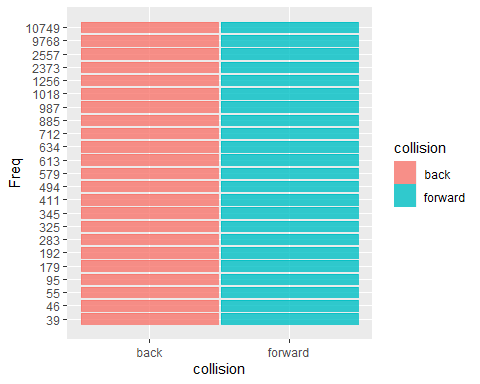
## The following objects are masked from 'package:vcd':  
##   
## mosaic, spine

s<-as.data.frame(Trucks\_data)  
ggplot(data=s)+  
 geom\_mosaic(aes(x=product(Freq,period),na.rm=T,fill=period))

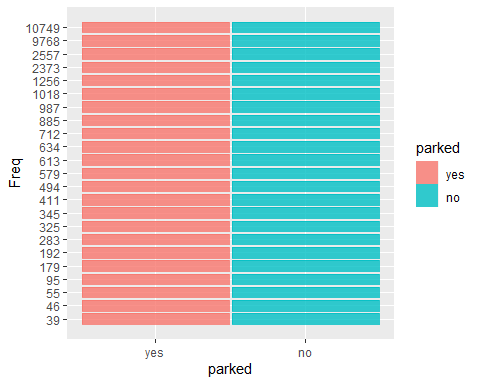
## Warning: `unite\_()` was deprecated in tidyr 1.2.0.  
## Please use `unite()` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was generated.



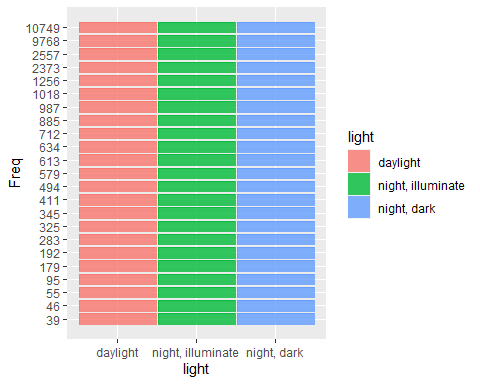
ggplot(data=s)+  
 geom\_mosaic(aes(x=product(Freq,collision),na.rm=T,fill=collision))



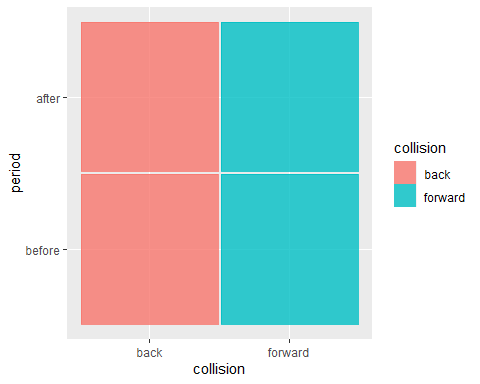
ggplot(data=s)+  
 geom\_mosaic(aes(x=product(Freq,parked),na.rm=T,fill=parked))



ggplot(data=s)+  
 geom\_mosaic(aes(x=product(Freq,light),na.rm=T,fill=light))

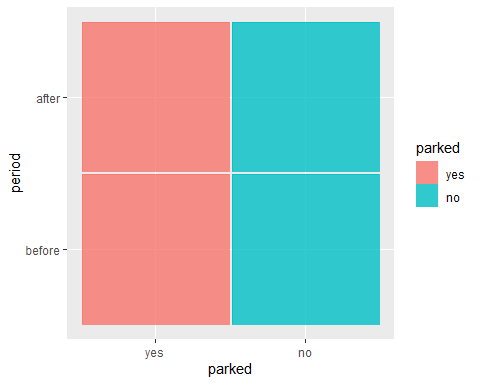


ggplot(data=s)+  
 geom\_mosaic(aes(x=product(period,collision),na.rm=T,fill=collision))

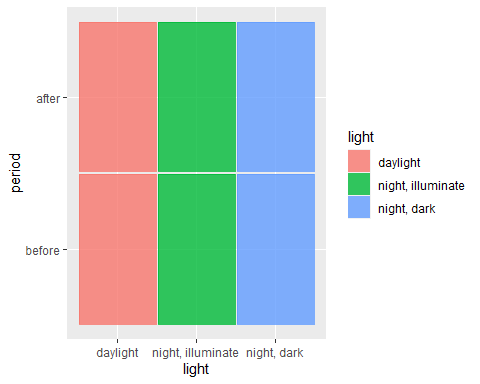


مثلا از نمودار زیر متوجه می شویم شمارش تصادفات در جلو وعقب کامیون،قبل و بعد از اجرای قوانین یکسان هستند.

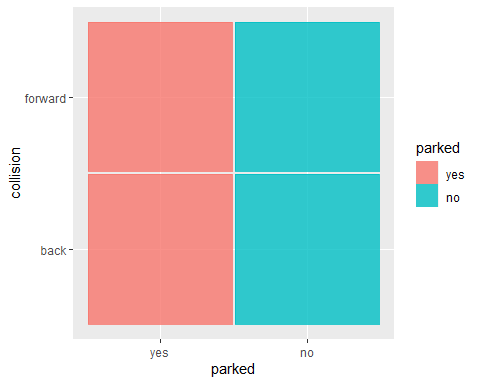
ggplot(data=s)+  
 geom\_mosaic(aes(x=product(period,parked),na.rm=T,fill=parked))



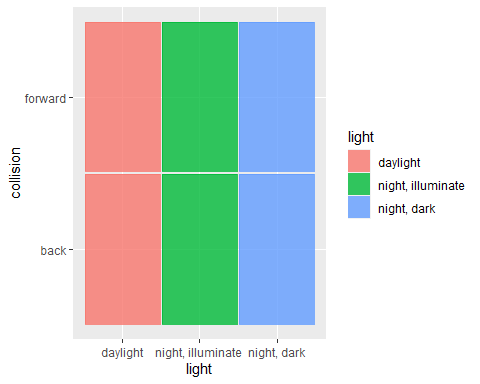
ggplot(data=s)+  
 geom\_mosaic(aes(x=product(period,light),na.rm=T,fill=light))



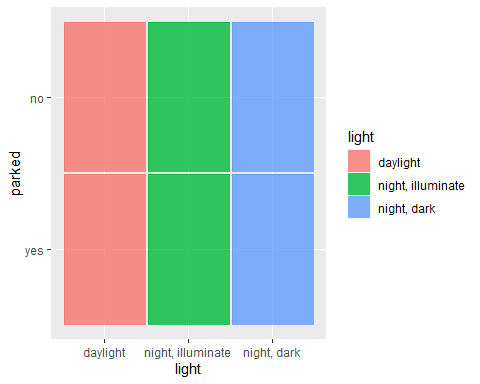
ggplot(data=s)+  
 geom\_mosaic(aes(x=product(collision,parked),na.rm=T,fill=parked))



ggplot(data=s)+  
 geom\_mosaic(aes(x=product(collision,light),na.rm=T,fill=light))



ggplot(data=s)+  
 geom\_mosaic(aes(x=product(parked,light),na.rm=T,fill=light))



حال به بررسی استقلال ها می پردازیم(بعنوان مثال فقط یک متغیر با دیگر متغیرها)

:مستقل اند و رابطه ای با هم ندارند.H0

مستقل نیستند و با یکدیگر رابطه دارند.:H1

##chisq.test

chisq.test(Freq\_period)

## Warning in chisq.test(Freq\_period): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: Freq\_period  
## X-squared = 24, df = 22, p-value = 0.3472<0.05

بین متغیرها استقلال داریم.پی اچ صفر پذیرفته می شود.

chisq.test(Freq\_collision)

## Warning in chisq.test(Freq\_collision): Chi-squared approximation may be  
## incorrect

##   
## Pearson's Chi-squared test  
##   
## data: Freq\_collision  
## X-squared = 22, df = 22, p-value = 0.4599

chisq.test(Freq\_parked)

## Warning in chisq.test(Freq\_parked): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: Freq\_parked  
## X-squared = 24, df = 22, p-value = 0.3472

chisq.test(Freq\_light)

## Warning in chisq.test(Freq\_light): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: Freq\_light  
## X-squared = 45, df = 44, p-value = 0.4298

برای مدل لگ خطی آزمایش میکنیم که از نظر منطقی کدام متغیر به بقیه وابسته و کدام مستق است ابتدا فرض میکنیم فقط تصادم از بقیه مستقل است داریم:

##fitting model1

library(MASS)  
data("Trucks")  
tab1 <- xtabs(Freq ~ period + collision + light + parked, data = Trucks)  
Trucks\_model1<-loglm(~ (period+parked+light)\*collision,data=tab1)  
coef(Trucks\_model1)

## $`(Intercept)`  
## [1] 6.119097  
##   
## $period  
## before after   
## 0.0707138 -0.0707138   
##   
## $collision  
## back forward   
## 0.2505351 -0.2505351   
##   
## $light  
## daylight night, illuminate night, dark   
## 1.2024527 -0.5462554 -0.6561973   
##   
## $parked  
## yes no   
## -1.133069 1.133069   
##   
## $period.collision  
## collision  
## period back forward  
## before 0.01354606 -0.01354606  
## after -0.01354606 0.01354606  
##   
## $collision.light  
## light  
## collision daylight night, illuminate night, dark  
## back -0.3037715 0.1141473 0.1896242  
## forward 0.3037715 -0.1141473 -0.1896242  
##   
## $collision.parked  
## parked  
## collision yes no  
## back 0.720021 -0.720021  
## forward -0.720021 0.720021

residuals(Trucks\_model1)

## Re-fitting to get frequencies and fitted values

## , , light = daylight, parked = yes  
##   
## collision  
## period back forward  
## before -10.564986 -4.704863  
## after -9.275747 -3.887397  
##   
## , , light = night, illuminate, parked = yes  
##   
## collision  
## period back forward  
## before 18.62577 9.358293  
## after 10.69626 4.706283  
##   
## , , light = night, dark, parked = yes  
##   
## collision  
## period back forward  
## before 4.786794 3.710315  
## after -3.031840 3.152414  
##   
## , , light = daylight, parked = no  
##   
## collision  
## period back forward  
## before 4.039856 -0.256736  
## after 8.274540 1.592777  
##   
## , , light = night, illuminate, parked = no  
##   
## collision  
## period back forward  
## before -13.16058 0.6103321  
## after -11.64706 -3.5011482  
##   
## , , light = night, dark, parked = no  
##   
## collision  
## period back forward  
## before -0.9461254 -0.2364723  
## after -0.6546648 -0.9893835

در مرحله بعد پارک شدن یا نشدن را به نور مستقل میگیریم:

##fitting model2

library(MASS)  
data("Trucks")  
tab1 <- xtabs(Freq ~ period + collision + light + parked, data = Trucks)  
Trucks\_model2<-loglm(~ (parked+light)\*period\*collision,data=tab1)  
coef(Trucks\_model2)

## $`(Intercept)`  
## [1] 6.112338  
##   
## $period  
## before after   
## 0.1247328 -0.1247328   
##   
## $collision  
## back forward   
## 0.2477694 -0.2477694   
##   
## $light  
## daylight night, illuminate night, dark   
## 1.2078030 -0.5505389 -0.6572640   
##   
## $parked  
## yes no   
## -1.137235 1.137235   
##   
## $period.collision  
## collision  
## period back forward  
## before 0.02149411 -0.02149411  
## after -0.02149411 0.02149411  
##   
## $period.light  
## light  
## period daylight night, illuminate night, dark  
## before -0.05579803 0.04353362 0.0122644  
## after 0.05579803 -0.04353362 -0.0122644  
##   
## $collision.light  
## light  
## collision daylight night, illuminate night, dark  
## back -0.3012966 0.1142621 0.1870344  
## forward 0.3012966 -0.1142621 -0.1870344  
##   
## $period.parked  
## parked  
## period yes no  
## before 0.04159345 -0.04159345  
## after -0.04159345 0.04159345  
##   
## $collision.parked  
## parked  
## collision yes no  
## back 0.7175299 -0.7175299  
## forward -0.7175299 0.7175299  
##   
## $period.collision.light  
## , , light = daylight  
##   
## collision  
## period back forward  
## before -0.02074981 0.02074981  
## after 0.02074981 -0.02074981  
##   
## , , light = night, illuminate  
##   
## collision  
## period back forward  
## before -0.003646614 0.003646614  
## after 0.003646614 -0.003646614  
##   
## , , light = night, dark  
##   
## collision  
## period back forward  
## before 0.02439642 -0.02439642  
## after -0.02439642 0.02439642  
##   
##   
## $period.collision.parked  
## , , parked = yes  
##   
## collision  
## period back forward  
## before 0.01999318 -0.01999318  
## after -0.01999318 0.01999318  
##   
## , , parked = no  
##   
## collision  
## period back forward  
## before -0.01999318 0.01999318  
## after 0.01999318 -0.01999318

residuals(Trucks\_model2)

## Re-fitting to get frequencies and fitted values

## , , light = daylight, parked = yes  
##   
## collision  
## period back forward  
## before -11.744711 -5.168676  
## after -7.743744 -3.369800  
##   
## , , light = night, illuminate, parked = yes  
##   
## collision  
## period back forward  
## before 15.79280 8.623827  
## after 13.71464 5.434676  
##   
## , , light = night, dark, parked = yes  
##   
## collision  
## period back forward  
## before 2.3438132 3.419955  
## after -0.5158831 3.462451  
##   
## , , light = daylight, parked = no  
##   
## collision  
## period back forward  
## before 7.504831 0.7811831  
## after 4.493469 0.4960769  
##   
## , , light = night, illuminate, parked = no  
##   
## collision  
## period back forward  
## before -13.91816 -1.7027535  
## after -10.66651 -0.9803802  
##   
## , , light = night, dark, parked = no  
##   
## collision  
## period back forward  
## before -1.6939597 -0.6064329  
## after 0.3160379 -0.5952208

دوره و چراغ را مستقل بگیریم:

##fitting model3

library(MASS)  
data("Trucks")  
tab1 <- xtabs(Freq ~ period + collision + light + parked, data = Trucks)  
Trucks\_model3<-loglm(~ (period+light)\*parked\*collision,data=tab1)  
coef(Trucks\_model3)

## $`(Intercept)`  
## [1] 6.206276  
##   
## $period  
## before after   
## 0.09319963 -0.09319963   
##   
## $collision  
## back forward   
## 0.1417974 -0.1417974   
##   
## $light  
## daylight night, illuminate night, dark   
## 0.9794376 -0.3828927 -0.5965448   
##   
## $parked  
## yes no   
## -0.9471106 0.9471106   
##   
## $period.collision  
## collision  
## period back forward  
## before 0.01546456 -0.01546456  
## after -0.01546456 0.01546456  
##   
## $collision.light  
## light  
## collision daylight night, illuminate night, dark  
## back -0.1822923 0.01488052 0.1674118  
## forward 0.1822923 -0.01488052 -0.1674118  
##   
## $period.parked  
## parked  
## period yes no  
## before 0.04159345 -0.04159345  
## after -0.04159345 0.04159345  
##   
## $collision.parked  
## parked  
## collision yes no  
## back 0.6922777 -0.6922777  
## forward -0.6922777 0.6922777  
##   
## $light.parked  
## parked  
## light yes no  
## daylight -0.3887394 0.3887394  
## night, illuminate 0.4021098 -0.4021098  
## night, dark -0.0133704 0.0133704  
##   
## $period.collision.parked  
## , , parked = yes  
##   
## collision  
## period back forward  
## before 0.01999318 -0.01999318  
## after -0.01999318 0.01999318  
##   
## , , parked = no  
##   
## collision  
## period back forward  
## before -0.01999318 0.01999318  
## after 0.01999318 -0.01999318  
##   
##   
## $collision.light.parked  
## , , parked = yes  
##   
## light  
## collision daylight night, illuminate night, dark  
## back -0.02004867 0.1168641 -0.09681541  
## forward 0.02004867 -0.1168641 0.09681541  
##   
## , , parked = no  
##   
## light  
## collision daylight night, illuminate night, dark  
## back 0.02004867 -0.1168641 0.09681541  
## forward -0.02004867 0.1168641 -0.09681541

residuals(Trucks\_model3)

## Re-fitting to get frequencies and fitted values

## , , light = daylight, parked = yes  
##   
## collision  
## period back forward  
## before -2.266918 -0.8393465  
## after 2.603222 0.9073253  
##   
## , , light = night, illuminate, parked = yes  
##   
## collision  
## period back forward  
## before 0.9407571 1.351855  
## after -1.1327260 -1.581197  
##   
## , , light = night, dark, parked = yes  
##   
## collision  
## period back forward  
## before 2.173993 -0.1038465  
## after -2.714907 0.1140513  
##   
## , , light = daylight, parked = no  
##   
## collision  
## period back forward  
## before -0.4723671 -0.8153457  
## after 0.4935320 0.8600454  
##   
## , , light = night, illuminate, parked = no  
##   
## collision  
## period back forward  
## before 0.3741772 2.060350  
## after -0.3951140 -2.226494  
##   
## , , light = night, dark, parked = no  
##   
## collision  
## period back forward  
## before 0.7246203 0.4137382  
## after -0.7677673 -0.4396547

##Fitting the best model(linear log model)\_coef\_residuals

library(vcd)  
library(MASS)  
data("Trucks")  
Trucks

## Freq period collision parked light  
## 1 712 before back yes daylight  
## 2 613 after back yes daylight  
## 3 192 before forward yes daylight  
## 4 179 after forward yes daylight  
## 5 2557 before back no daylight  
## 6 2373 after back no daylight  
## 7 10749 before forward no daylight  
## 8 9768 after forward no daylight  
## 9 634 before back yes night, illuminate  
## 10 411 after back yes night, illuminate  
## 11 95 before forward yes night, illuminate  
## 12 55 after forward yes night, illuminate  
## 13 325 before back no night, illuminate  
## 14 283 after back no night, illuminate  
## 15 1256 before forward no night, illuminate  
## 16 987 after forward no night, illuminate  
## 17 345 before back yes night, dark  
## 18 179 after back yes night, dark  
## 19 46 before forward yes night, dark  
## 20 39 after forward yes night, dark  
## 21 579 before back no night, dark  
## 22 494 after back no night, dark  
## 23 1018 before forward no night, dark  
## 24 885 after forward no night, dark

بهترین مدل براساس مانده ها و از نظر منطقی مدل زیر است که دوره و تصادم از هم مستقل اند:

tab <- xtabs(Freq ~ period + collision + light + parked, data = Trucks)  
Trucks\_model=loglm(~ (collision + period) \* parked \* light, data = tab)  
coef(Trucks\_model)

## $`(Intercept)`  
## [1] 6.198649  
##   
## $period  
## before after   
## 0.135117 -0.135117   
##   
## $collision  
## back forward   
## 0.144055 -0.144055   
##   
## $light  
## daylight night, illuminate night, dark   
## 0.9909545 -0.3851637 -0.6057908   
##   
## $parked  
## yes no   
## -0.9528365 0.9528365   
##   
## $period.light  
## light  
## period daylight night, illuminate night, dark  
## before -0.07914331 0.03151516 0.04762816  
## after 0.07914331 -0.03151516 -0.04762816  
##   
## $collision.light  
## light  
## collision daylight night, illuminate night, dark  
## back -0.1822923 0.01488052 0.1674118  
## forward 0.1822923 -0.01488052 -0.1674118  
##   
## $period.parked  
## parked  
## period yes no  
## before 0.05887773 -0.05887773  
## after -0.05887773 0.05887773  
##   
## $collision.parked  
## parked  
## collision yes no  
## back 0.6947688 -0.6947688  
## forward -0.6947688 0.6947688  
##   
## $light.parked  
## parked  
## light yes no  
## daylight -0.37941497 0.37941497  
## night, illuminate 0.40258267 -0.40258267  
## night, dark -0.02316769 0.02316769  
##   
## $period.light.parked  
## , , parked = yes  
##   
## light  
## period daylight night, illuminate night, dark  
## before -0.04871746 -0.001765867 0.05048333  
## after 0.04871746 0.001765867 -0.05048333  
##   
## , , parked = no  
##   
## light  
## period daylight night, illuminate night, dark  
## before 0.04871746 0.001765867 -0.05048333  
## after -0.04871746 -0.001765867 0.05048333  
##   
##   
## $collision.light.parked  
## , , parked = yes  
##   
## light  
## collision daylight night, illuminate night, dark  
## back -0.02004867 0.1168641 -0.09681541  
## forward 0.02004867 -0.1168641 0.09681541  
##   
## , , parked = no  
##   
## light  
## collision daylight night, illuminate night, dark  
## back 0.02004867 -0.1168641 0.09681541  
## forward -0.02004867 0.1168641 -0.09681541

residuals(Trucks\_model)

## Re-fitting to get frequencies and fitted values

## , , light = daylight, parked = yes  
##   
## collision  
## period back forward  
## before 0.2160732 -0.4108989  
## after -0.2315181 0.4344653  
##   
## , , light = night, illuminate, parked = yes  
##   
## collision  
## period back forward  
## before -0.1384994 0.3629394  
## after 0.1728234 -0.4614727  
##   
## , , light = night, dark, parked = yes  
##   
## collision  
## period back forward  
## before 0.4654382 -1.193071  
## after -0.6308281 1.488650  
##   
## , , light = daylight, parked = no  
##   
## collision  
## period back forward  
## before -0.4112346 0.2012466  
## after 0.4292984 -0.2108239  
##   
## , , light = night, illuminate, parked = no  
##   
## collision  
## period back forward  
## before -0.6663739 0.3442760  
## after 0.7335614 -0.3855325  
##   
## , , light = night, dark, parked = no  
##   
## collision  
## period back forward  
## before 0.1332304 -0.1001873  
## after -0.1436621 0.1076945

جداول را جدا جدا و تیکه تیکه با هم بررسی کنیم:

##slice tables

library(gmodels)  
CrossTable(Trucks$Freq)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 24   
##   
##   
## | 39 | 46 | 55 | 95 | 179 |   
## |-----------|-----------|-----------|-----------|-----------|  
## | 1 | 1 | 1 | 1 | 2 |   
## | 0.042 | 0.042 | 0.042 | 0.042 | 0.083 |   
## |-----------|-----------|-----------|-----------|-----------|  
##   
##   
## | 192 | 283 | 325 | 345 | 411 |   
## |-----------|-----------|-----------|-----------|-----------|  
## | 1 | 1 | 1 | 1 | 1 |   
## | 0.042 | 0.042 | 0.042 | 0.042 | 0.042 |   
## |-----------|-----------|-----------|-----------|-----------|  
##   
##   
## | 494 | 579 | 613 | 634 | 712 |   
## |-----------|-----------|-----------|-----------|-----------|  
## | 1 | 1 | 1 | 1 | 1 |   
## | 0.042 | 0.042 | 0.042 | 0.042 | 0.042 |   
## |-----------|-----------|-----------|-----------|-----------|  
##   
##   
## | 885 | 987 | 1018 | 1256 | 2373 |   
## |-----------|-----------|-----------|-----------|-----------|  
## | 1 | 1 | 1 | 1 | 1 |   
## | 0.042 | 0.042 | 0.042 | 0.042 | 0.042 |   
## |-----------|-----------|-----------|-----------|-----------|  
##   
##   
## | 2557 | 9768 | 10749 |   
## |-----------|-----------|-----------|  
## | 1 | 1 | 1 |   
## | 0.042 | 0.042 | 0.042 |   
## |-----------|-----------|-----------|  
##   
##   
##   
##

CrossTable(Trucks$period)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 24   
##   
##   
## | before | after |   
## |-----------|-----------|  
## | 12 | 12 |   
## | 0.500 | 0.500 |   
## |-----------|-----------|  
##   
##   
##   
##

CrossTable(Trucks$collision)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 24   
##   
##   
## | back | forward |   
## |-----------|-----------|  
## | 12 | 12 |   
## | 0.500 | 0.500 |   
## |-----------|-----------|  
##   
##   
##   
##

CrossTable(Trucks$parked)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 24   
##   
##   
## | yes | no |   
## |-----------|-----------|  
## | 12 | 12 |   
## | 0.500 | 0.500 |   
## |-----------|-----------|  
##   
##   
##   
##

CrossTable(Trucks$light)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 24   
##   
##   
## | daylight | night, illuminate | night, dark |   
## |-------------------|-------------------|-------------------|  
## | 8 | 8 | 8 |   
## | 0.333 | 0.333 | 0.333 |   
## |-------------------|-------------------|-------------------|  
##   
##   
##   
##

مدل لوجیت:

##logit model\_confidence interval

logit\_fit<-glm(collision ~ period + Freq + light + parked, data = Trucks,family = "binomial")  
logit\_fit

##   
## Call: glm(formula = collision ~ period + Freq + light + parked, family = "binomial",   
## data = Trucks)  
##   
## Coefficients:  
## (Intercept) periodafter Freq   
## -0.8239359 0.0566087 0.0004219   
## lightnight, illuminate lightnight, dark parkedno   
## 0.9889937 1.0139482 -0.8150930   
##   
## Degrees of Freedom: 23 Total (i.e. Null); 18 Residual  
## Null Deviance: 33.27   
## Residual Deviance: 30.5 AIC: 42.5

confint(logit\_fit)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -3.263757e+00 1.260579063  
## periodafter -1.656422e+00 1.782948677  
## Freq -6.166879e-05 0.001712197  
## lightnight, illuminate -1.327631e+00 3.541598208  
## lightnight, dark -1.316981e+00 3.580835542  
## parkedno -2.882292e+00 1.074590759

عدد ستاره داری نداریم پس همگی باید در مدل باشند.

یک چالش در این مثال داریم که گویا اکثر اثرات متقابل سه تایی هستند و اگر بخواهیم اندازه نمودارها برابر با هم در نیایند باید بطور سه بعدی آن ها را بکشیم که کار دشواری است.

عواملی که تعداد تصادفات را بطور معنی دار افزایش می دهد :روشنایی،پارک شدن یا نشدن وبعدوقبل نوامبر.

کامیون های پارک شده در جاده های با روشنایی و همچنین کامیونهای در حال حرکت در جاده تاریک بیشتر از غقب ضربه دیدند،کامیون های در حال حرکت در جاده با روشنایی و پارک شده در خیابان تاریک بیشتر از جلو ضربه دیده بودند.