

UNIVERSIDAD DE COSTA RICA

DISTRIBUCIONES DE PÉRDIDAS

MINOR HEADING

---

# Frecuencia y severidad de default crediticio

---

*Autores*

David Zumbado  
Leonardo Blanco  
Ignacio Barrantes

31 de agosto de 2022

# Table of contents

<b>Bitácora 1</b>	<b>3</b>
Punto 1 . . . . .	3
1- Nombres de los integrantes . . . . .	3
2- Idea . . . . .	3
<b>Intro</b>	<b>4</b>
Sección 5. . . . .	4
Sección 8. Literatura . . . . .	5
<b>Referencias para escribir en Quarto</b>	<b>7</b>
<b>Metodología</b>	<b>8</b>
<b>Referencias</b>	<b>9</b>

# List of Figures

1	City and highway mileage for 38 popular models of cars. . . . .	8
---	---	---

## List of Tables

```
library(ggplot2)
```

Warning: package 'ggplot2' was built under R version 4.1.3

```
library(janitor)
```

Warning: package 'janitor' was built under R version 4.1.3

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

```
library(tidyverse)
```

Warning: package 'tidyverse' was built under R version 4.1.3

-- Attaching packages ----- tidyverse 1.3.2 --

v tibble	3.1.8	v dplyr	1.0.9
v tidyr	1.2.0	v stringr	1.4.1
v readr	2.1.2	v forcats	0.5.2
v purrr	0.3.4		

Warning: package 'tibble' was built under R version 4.1.3

Warning: package 'tidyr' was built under R version 4.1.3

Warning: package 'readr' was built under R version 4.1.3

Warning: package 'dplyr' was built under R version 4.1.3

Warning: package 'stringr' was built under R version 4.1.3

Warning: package 'forcats' was built under R version 4.1.3

-- Conflicts ----- tidyverse\_conflicts() --

x dplyr::filter() masks stats::filter()  
x dplyr::lag() masks stats::lag()

## *List of Tables*

```
library(lubridate)
```

Attaching package: 'lubridate'

The following objects are masked from 'package:base':

date, intersect, setdiff, union

```
library(kableExtra)
```

Warning: package 'kableExtra' was built under R version 4.1.3

Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

group\_rows

# Bitácora 1

## Punto 1

### 1- Nombres de los integrantes

El grupo de trabajo estará integrado por:

- Ignacio Barrantes Valerio
- Leonardo Blanco Villalobos
- David Zumbado Fernández

### 2- Idea

# Intro

## Sección 5.

### Fuente de Información:

Los datos se obtuvieron del Department of Homeland Security, un organismo del gobierno de Estados Unidos y se puede encontrar en Homeland Security (2015)

```
#Se cargan y depuran los datos
datos <- read.csv('tsa_claims.csv') %>%
  clean_names() #se limpian nombres columnas
datos <- datos %>% mutate(date_received = dmy(date_received),
                          incident_date = mdy_hm(incident_date),
                          claim_amount = as.numeric(gsub("\\$", "", claim_amount)),
                          close_amount = as.numeric(gsub("\\$", "", close_amount)))
```

Warning: 17651 failed to parse.

Warning in mask\$eval\_all\_mutate(quo): NAs introduced by coercion

Warning in mask\$eval\_all\_mutate(quo): NAs introduced by coercion

### Contexto temporal y espacial de los datos:

La base registra la ocurrencia de reclamos entre 2002 y 2015 en 466 aeropuertos alrededor de Estados Unidos.

### Facilidad de obtener la información:

La base fue extraída de la página oficial del departamento de seguridad nacional la cual es accesible por cualquier persona por lo que se considera fácil de obtener.

### Población de estudio:

Los aeropuertos

### Muestra observada:



Aeropuertos estadounidenses donde se presentaron reclamos por daños ocasionados por seguridad.

**Unidad estadística o individuos:**

La unidad estadística es el registro de una ocurrencia de un reclamo.

**Descripción de las variables de la tabla:**

Los datos se conforman por 13 variables: `claim_number` es el identificador del reclamo, `date_received` es la fecha que se registró el reclamo, `incident_date` es la fecha que ocurrió el incidente que ameritó el reclamo, `airport_code` son las 3 letras que identifican el aeropuerto donde ocurrieron el incidente, y `airport_name` es el nombre del aeropuerto. `Claim_type` es el tipo de daño ocasionado (daño a propiedad, daño a personas, entre otras), `claim_site` es el lugar dentro del aeropuerto donde sucedió el incidente. `Item` es el ítem que sufrió el daño, `claim_amount` es la cantidad en dólares que la persona pide, `status` es el estado del reclamo (se llegó a un acuerdo, se negó, etc. . . ), y `close_amount` fue el monto que efectivamente se pagó.

## Sección 8. Literatura

1. **Título:** Modelling Dependencies in Airport Passenger Claim Data Using Copulas Flores (2022)

**Autor:** Roberto Carcache Flores

**Nombre del tema:** Modelación del riesgo utilizando cópulas

**Forma de organizarlo:**

- Cronológico: Febrero 2022
- Metodológico: Cópulas bivariadas y multivariadas y simulaciones
- Temático: Funciones de distribución y dependencia de variables aleatorias
- Teoría: Probabilidad y estadística

**Resumen en una oración:** Se encuentra la mejor distribución para la severidad y frecuencia de cada reclamo y luego estas distribuciones marginales se incorporan en diferentes modelos de cópulas

**Argumento central:** En la metodología tradicional del modelamiento del riesgo se asume independencia entre frecuencia y severidad, lo cual no se hace en esta investigación. Además, se utiliza un proceso de eliminación de la tendencia con respecto al tiempo para mejorar los resultados.

**Problemas con el argumento o el tema:** Las medidas de riesgo utilizando cópulas resultan en medidas de riesgo más altas que en los datos históricos.

**Resumen en un párrafo:** Se eliminan los reclamos que fueron negados justificando el hecho de que el punto de la investigación es cuantificar los pagos que efectivamente fueron hechos, además del gran volumen de los datos. La agregación de los datos se hace por mes y con suma para la severidad y por frecuencia de los reclamos. El autor nota que hay una tendencia negativa de la frecuencia y severidad con respecto al tiempo por lo que procede a eliminar la tendencia. Luego determina la mejor marginal para cada variable utilizando MLE. Se encuentra que la binomial negativa se ajusta mejor a las frecuencias. Por otro lado la Log-Laplace se ajusta mejor a los reclamos por daños a la propiedad y la lognormal se ajusta mejor a los reclamos por pérdidas de los bienes por lo que se utilizan estas dos para modelar la severidad. Luego se procede a hacer algo similar con los resultados de eliminar la tendencia. Se encuentra que el proceso de eliminación de la tendencia facilita la búsqueda de una distribuciónes. Se encuentra que todas las variables pares muestran algún tipo de dependencia en las colas. Finalmente, las cópulas multivariadas se comparan utilizando log verosimilitud y se obtiene que las cópulas elípticas (Gaussiana y t-Student) se ajustan mejor que las arquimedianas (Clayton y Gumbel).

# Referencias para escribir en Quarto

Según Knuth (1984) el comportamiento mono sí existe

# Metodología

```
ggplot(mpg, aes(x = hwy, y = cty, color = cyl)) +  
  geom_point(alpha = 0.5, size = 2) +  
  scale_color_viridis_c() +  
  theme_minimal()
```

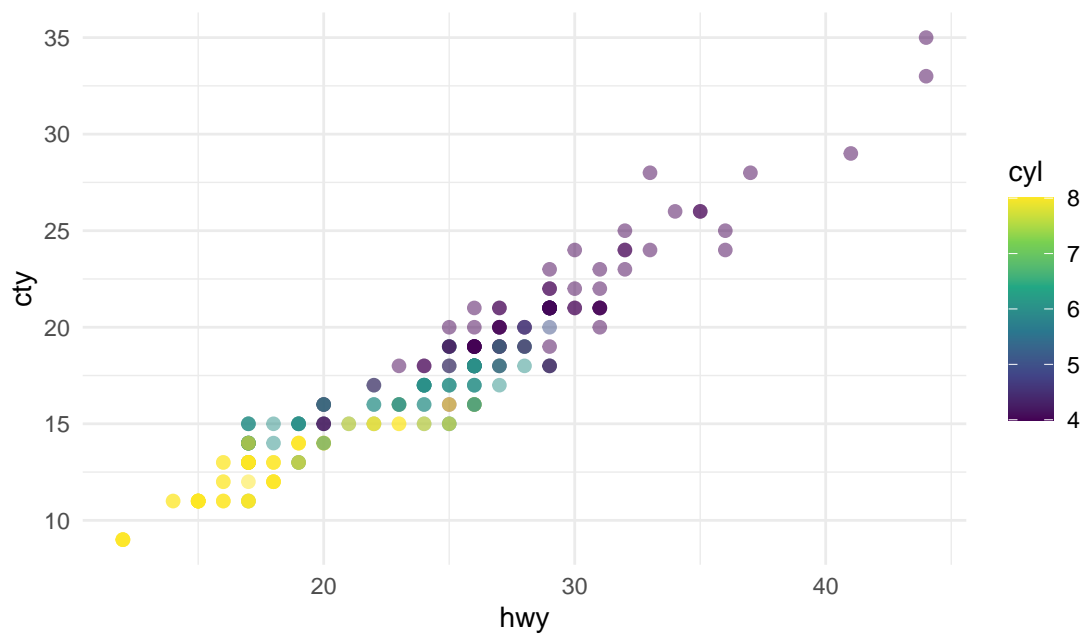


Figure 1: City and highway mileage for 38 popular models of cars.

La figura Figure 1 muestra un tendencia positiva en el comportamiento mono

## Referencias

- Flores, Roberto Carcache. 2022. “Modelling Dependencies in Airport Passenger: Claim Data Using Copulas.” PhD thesis, Instituto Superior de Economia e Gestão.
- Homeland Security, Department of. 2015. “TSA Claims Data.” <https://www.dhs.gov/tsa-claims-data>.
- Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.