

# Examen Parcial - Segunda Versión de Reporte

**Profesor:** Luis Enrique Nieto Barajas.

**Alumno:** César Zamora Martínez

**Fecha:** 11 de Noviembre de 2019

## Examen parcial

### Consideraciones generales

De acuerdo a las instrucciones, debemos realizar un análisis estadístico inferencial de la siguiente base de datos, mismas que contiene datos de características de mujeres embarazadas y del producto. Para ello procedemos a cargar los datos en cuestión:

#### Lectura de los datos

```
#-Lectura de datos-  
df <- read_csv("/media/Box/Modelos_Lineales_Generalizados/Examen_midterm/Gestacional.csv")
```

```
## Parsed with column specification:  
## cols(  
##   Anestesia = col_double(),  
##   Complica = col_double(),  
##   Edad = col_double(),  
##   Escolaridad = col_double(),  
##   Estado_civil = col_double(),  
##   Forma_nac = col_double(),  
##   Hospital = col_double(),  
##   Ind_cesarea = col_double(),  
##   Monitoreo = col_double(),  
##   Mortalidad = col_double(),  
##   Peso_madre = col_double(),  
##   Peso_prod = col_double()  
## )
```

```
#colnames(salarios) <- c("salario", "indice_qw", "experiencia", "indice_pubs")
```

Al respecto, cabe destacar que la base de datos se integra de 12 variables, que describen características de mujeres embarazadas y del producto con un total 535 casos.

En este sentido, las variables en comento se refieren a lo siguiente:

- *Anestesia* : Indicador si durante el parto se usó anestesia (1-no, 2-sí).
- *Complica* : Indicador si la madre tuvo algún tipo de complicación durante su embarazo (1-no, 2-sí).
- *Edad* : Edad de la madre en años.
- *Escolaridad* : Escolaridad de la madre (1-primaria, 2-secundaria, 3-preparatoria, 4-universidad).
- *Estado\_civil* : Estado civil de la madre (1-casada, 2-divorciada, 3-soltera, 4-unión libre).
- *Forma\_nac* : Tipo de parto (1-cesárea, 2-vaginal).
- *Hospital* : Nombre del hospital en donde se atendió el embarazo (1-Ángeles, 2- Español, 3-General, 4-Médica Sur, 5-Oaxaca, 6-Pemex Sur, 7-Tlaxcala) .
- *Ind\_cesarea* : Motivo por el que se llevó acabo la cesárea (1-DCP (desproporción céfalo pélvica), 2-electiva, 3-emergencia fetal, 4-emergencia materna, 5-iterativa, 6 -no hubo).
- *Monitoreo* : Indicador si la madre recibió un monitoreo fetal durante el embarazo (1-no, 2-sí).
- *Mortalidad* : Mortalidad del producto (1-no, 2-sí).
- *Peso\_madre* : Peso de la madre al inicio del embarazo (en Kgs).
- *Peso\_prod* : Peso al nacer del producto (en Kgs).

En primera procederemos a realizar un análisis exploratorio de los datos para comprender la información contenida en la base en estudio. Comenzaremos primero con un análisis univariado

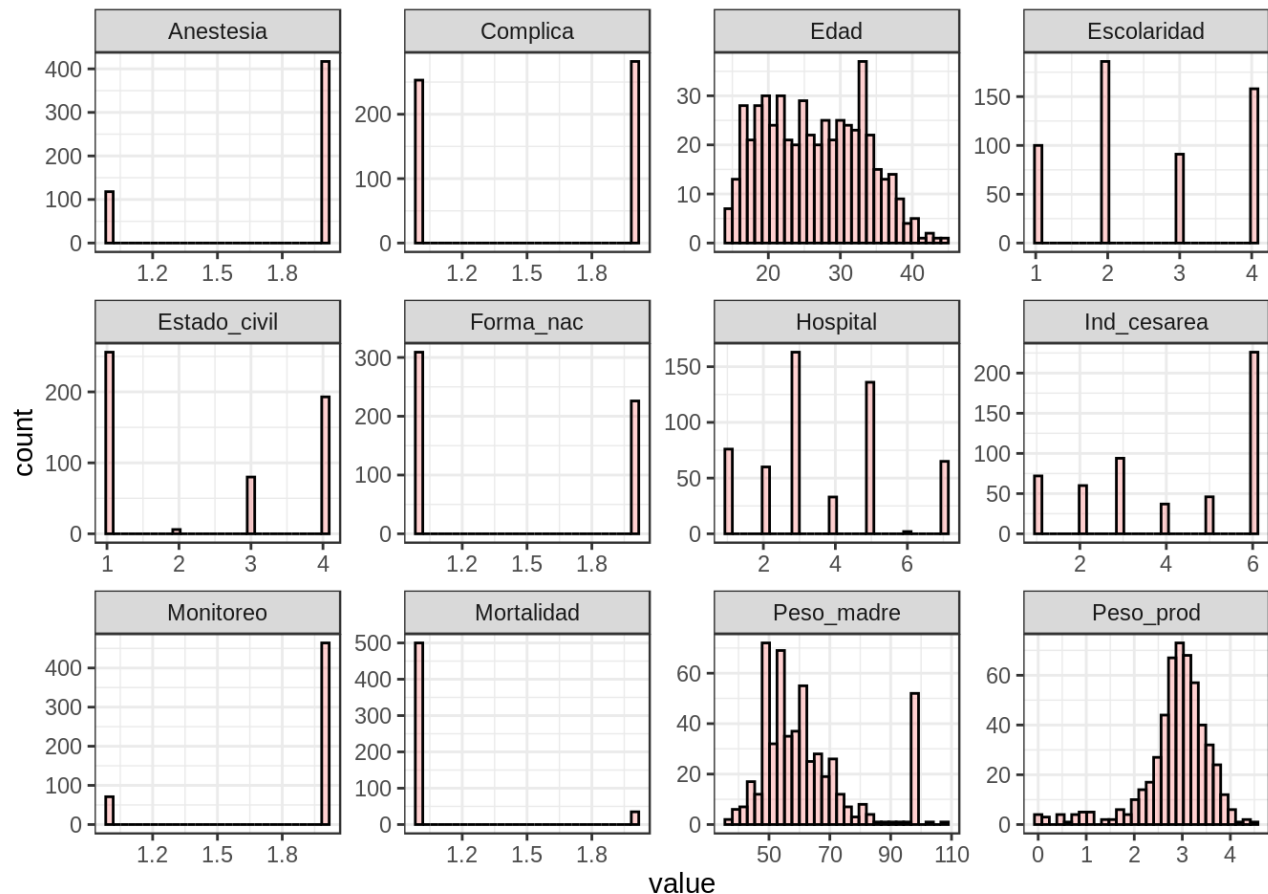
## Análisis exploratorio univariado

A efecto de realizar la exploración de cada una de las variables y aprovechando las herramientas del paquete *ggplot2*, a continuación de muestran histogramas de frecuencia (relativa) de las 12 variables que integran la base de datos.

**Nota:** En la parte superior de cada histograma se ha colocado el nombre de la variable con respecto a la cual se ha construido cada gráfica.

```
suppressWarnings(
df %>%
  gather(key = "var", value = "value") %>% # Se alarga la base de datos
  ggplot(aes(x = value)) + # se pasan datos de base alargada
  geom_histogram(colour="black", fill="#FF6666",alpha=0.3) + # Genera histogram
a
  facet_wrap(~ var, scales = "free") +
  theme_bw()
)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



De este gráfico, se destacan los siguientes hallazgos:

- En la mayor parte de los partos (más del 80% de los casos), el procedimiento médico requirió anestesia.
- Para la mayoría de los embarazos (un poco más del 50% de los casos) se presentaron complicaciones.
- La distribución de la edad de las madres presenta dos modas, la primera cerca de los 20 años de edad la segunda cerca de los 34 años. Se observa un ligero sesgo hacia la derecha de la distribución, indicando que aunque infrecuentes, existe embarazos en mujeres que sobrepasan los 35 años de edad.
- Por lo que hace a la escolaridad de las madres, destacan dos categorías como las que tienen más observaciones; en la primera para madres con educación a nivel secundaria (poco más del 30% de los casos) mientras que en el segundo (poco más del 28% de los casos) las madres cursaron la universidad.
- Respecto al estado civil de las madres, en poco mas del 46% de los casos se observa que se encuentran casadas.
- Asimismo, en más de 300 casos (cerca de 56% de todos los casos documentados) en los procedimientos de parto se empleo la técnica de cesárea.
- La mayoría de los partos se atendieron en el Hospital General y en el Hospital Oaxaca.
- En complemento, también se desprende que en la mayoría de los casos no hubo cesárea. Sin embargo, en los casos donde este procedimiento si fue requerido para el parto, el motivo que se

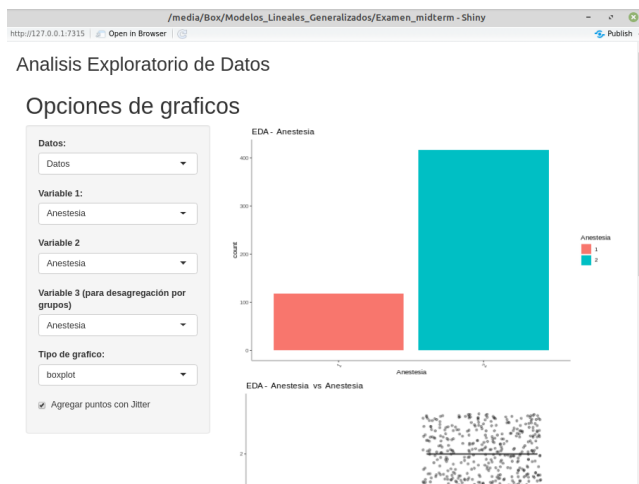
presento en mayor medida (cerca de 100 de los casos documentados) refiere a emergencia fetal, que como se sabe pone el riesgo tanto de la madre como del producto.

- Adicionalmente, en más de 450 casos (es decir, un poco más del 84% de los embarazados documentados) existió seguimiento de monitoreo fetal.
- Respecto a si existió mortalidad del producto (es decir, a si el bebé de la madre nació muerto o no) se observa que más de 500 casos (poco más del 93% de los que conforman a la base) se observa que este fenómeno estuvo ausente. Ello equivale a decir que la mayoría de los casos documentados en la base corresponde a productos vivos tras el parto.
- En lo tocante al peso de la madre, de manera general se puede afirmar que se presentan dos modas, la primera alrededor de los 55 kg de peso y la segunda cerca de 100 kg. Ello sugiere que la distribución del peso de las madres es segada hacia la derecha.
- Por lo que hace al peso del producto, se observa que este parece seguir una distribución que se agrupa alrededor de los 3 kg de pesos, y que es sesgada hacia la izquierda, lo cual implica que aunque poco frecuentes existen nacimiento de bebes que presentan peso menor a los dos kilogramos.

## Análisis bivariado.

Para esta parte, se recurrió a la implementación de un paquete interactivo de R que permite generar plots de la base seleccionando a través de un menú con pestañas a las variables en estudio y el tipo de plot que se desean explorar. La implementación corresponde al archivo de R denominado *mlg\_shiny.R* (ver anexo al final del documento)

La interfaz diseñada se muestra a continuación:

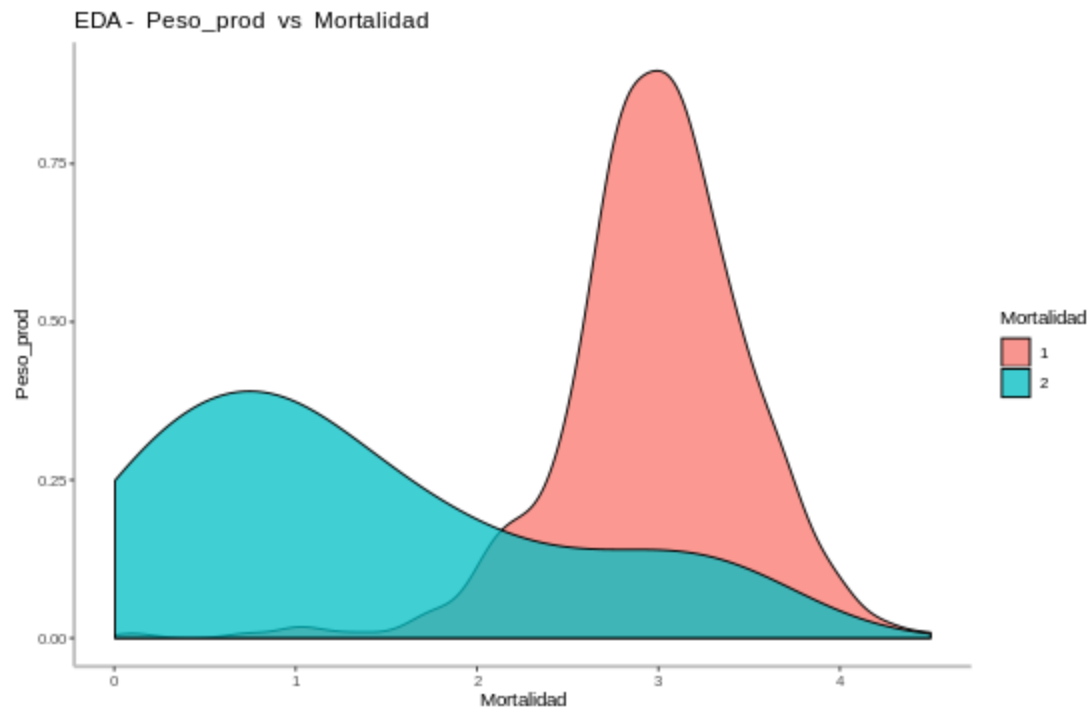


Interfaz del dashboard de shiny generada para esta parte del análisis exploratorio

### Relación del peso del producto y de la madre

A continuación mostramos la relación entre la mortalidad y el peso del producto.

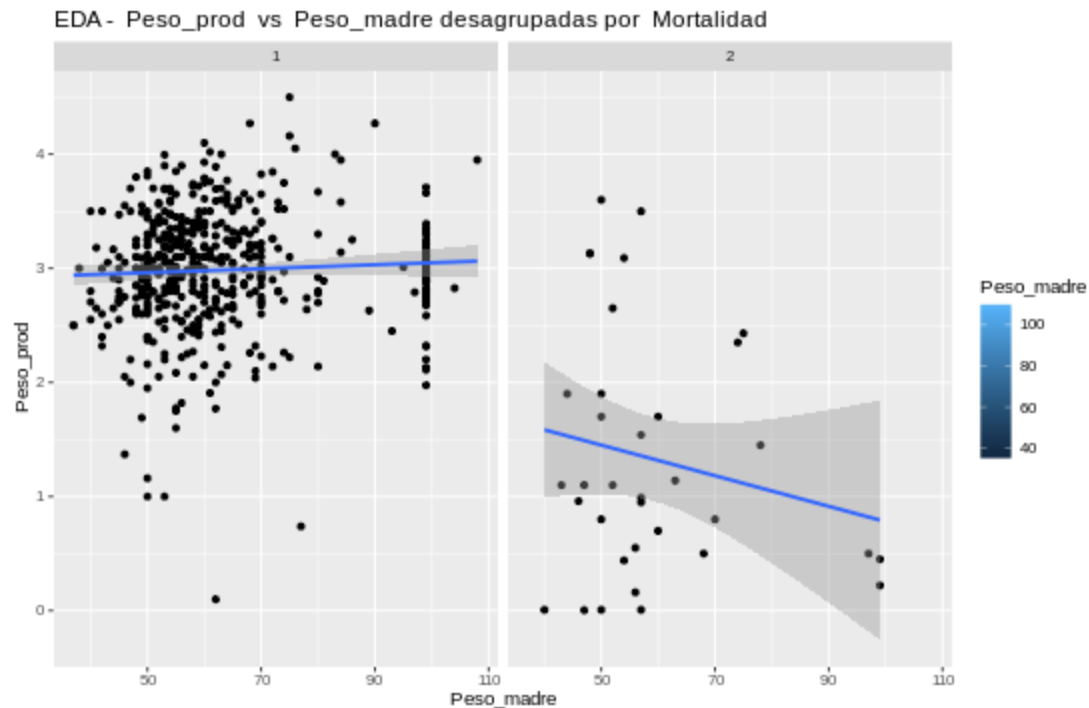
### Relación del peso de mortalidad y peso del producto



Scatterplot del peso del producto y mortalidad

En tales casos, la figura muestra como nuevamente hay dos distribuciones diferentes del peso de los productos para aquellos en los que fallecieron y los que no. En el primer caso, se observa una distribución que se acumula alrededor de 900 gr, que es sesgada hacia la derecha. En el segunda caso la distribución se agrupa alrededor de los tres kilogramos, y es sesgada hacia la izquierda.

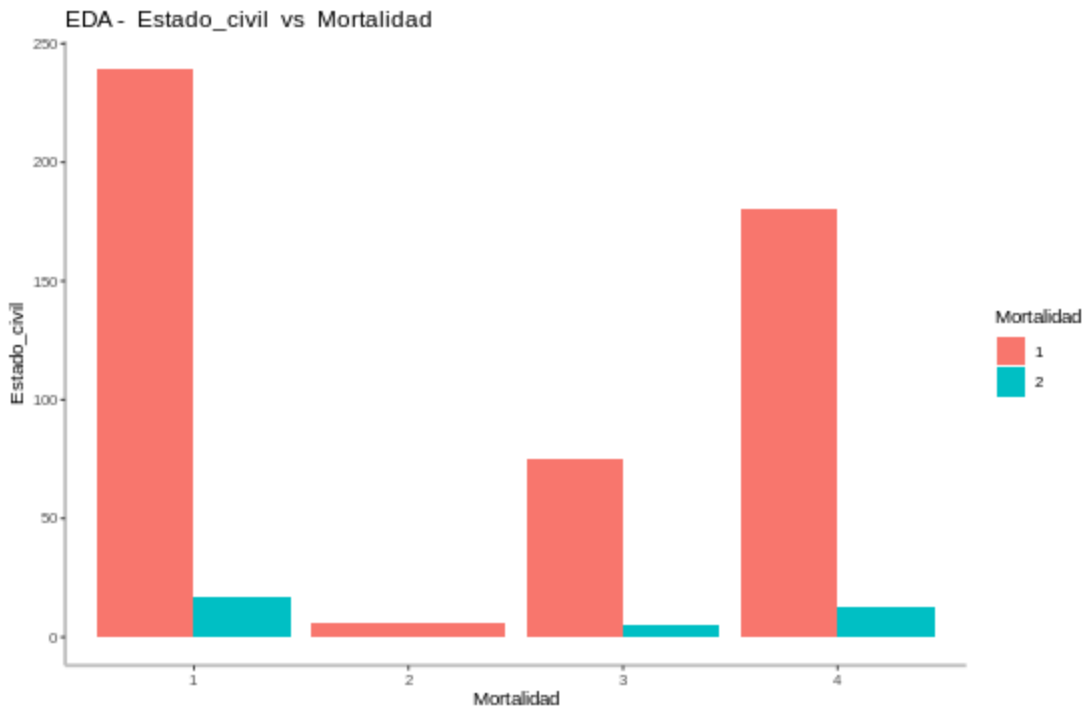
### Relación del peso del producto y de la madre, por mortalidad del producto



Scatterplot del peso del producto y de la madre, según estatus de mortalidad del producto

En la figura se ilustra como es que en caso de presentarse mortalidad en general hay una relación negativa entre el peso de la madre y del producto; la cual se invierte a su vez en el caso que no existiese mortalidad del producto (es decir, que el bebé nació vivo).

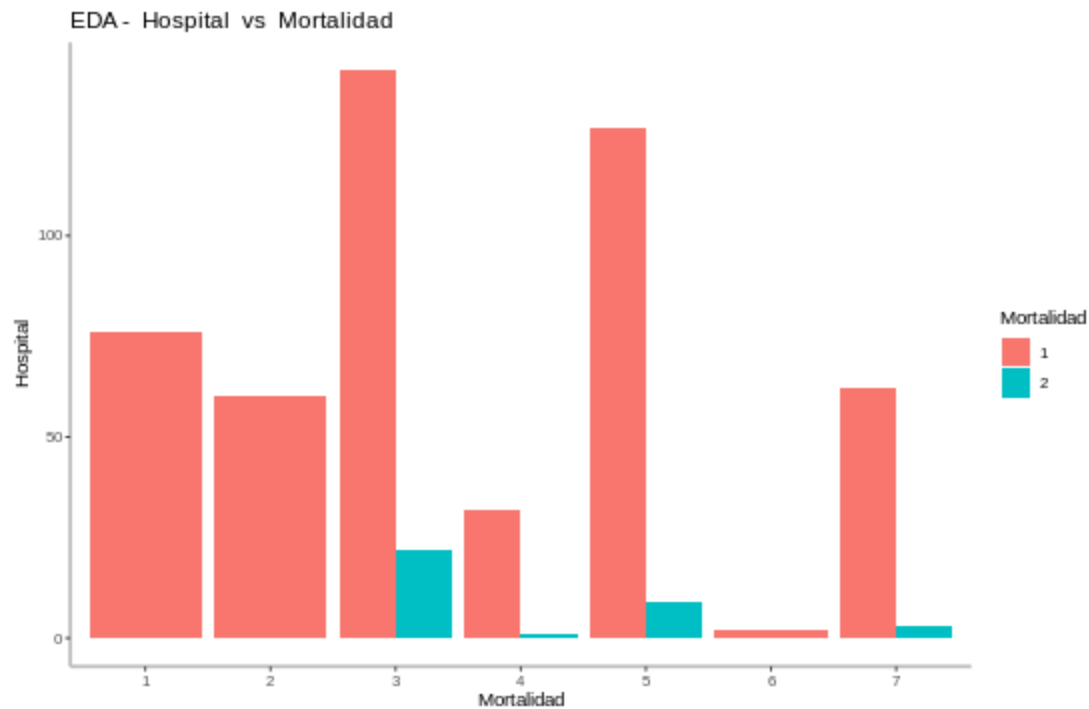
**Relación entre la mortalidad del producto y estado civil de la madre**



Distribución de mortalidad de acuerdo al estado civil de la madre

En la figura se ilustra la mortalidad del producto de acuerdo al estado civil de la madre. Resulta notable el hecho de que el mayor número observado de mortalidad se presente en mujeres que son casadas, seguido de aquellas que viven en unión libre.

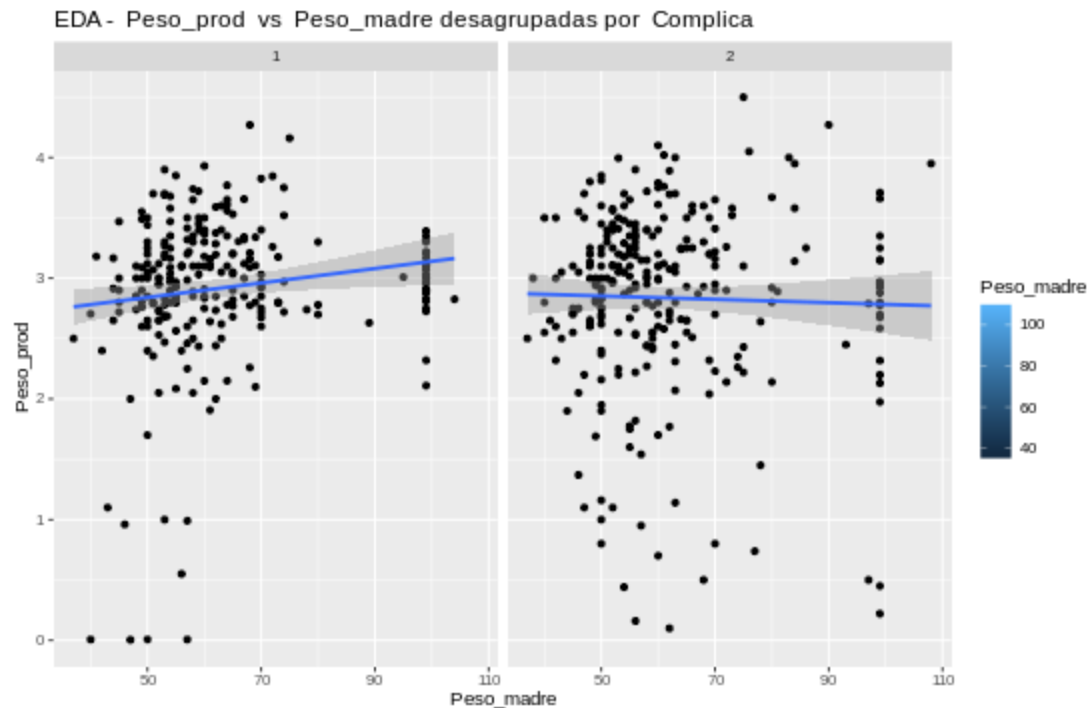
**Relación entre la mortalidad del producto y el hospital donde se atendió la madre**



Distribución de mortalidad de acuerdo al hospital donde se atendió la madre

En la figura se aprecia como en los hospitales 1) General, 2) Oaxaca y 3) Tlaxacala existió el mayor número de casos de mortalidad del producto. En adición, el Hospital Médica Sur reportó una cantidad pequeña de este fenómeno, mientras que el resto de nosocomios (a saber, Hospital Ángeles, Hospital Español, y el Hospital Pemex Sur) no registraron incidencias de mortalidad.

**Relación del peso del producto y de la madre, por presencia de complicaciones en el embarazo**

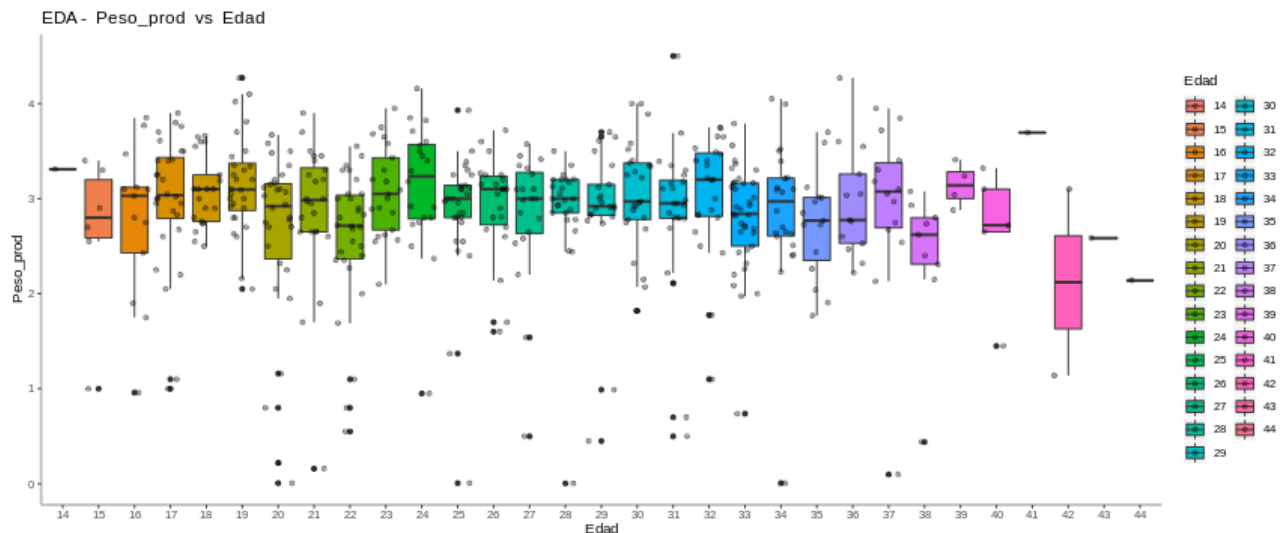


Scatterplot del peso del producto y de la madre, según la existencia de complicaciones en el embarazo

En este caso, la figura permite observar que en caso de no existir complicaciones en el parto, la relación entre el peso del producto y de la madre es positiva. Sin embargo, para aquellos casos en donde en efecto existieron complicaciones se observa en general una relación negativa entre el peso de la madre y del producto.

### Relación del peso del producto y edad de la madre

Otro punto relevante es explorar la relación entre el peso del producto y la edad de la madre. Esto se puede vislumbrar en la figura 6, la cual muestra que, en media, los producto de mujeres con al menos 42 años tienden a ser mas bajos que el resto



Relación de peso del producto y de la edad de la madre

## Ajuste de una familia de modelos para la mortalidad en la base de datos

Tal como hemos comentado en las secciones previas, la base de datos guarda una serie de campos que describen información de datos gestacionales del sistema de salud del país. De entre las variables contenidas en esta, destaca la mortalidad de producto en los casos de embarazos documentados a través de ésta.

De hecho, tal como se ha evidenciado en el análisis exploratorio de la sección previa, existen diversos factores que parecen incidir sobre la mortalidad de un neo-nato. Es por ello que para entender en mejor medidas el fenómeno de la mortalidad del producto.

Con este propósito, a la variable *Mortalidad* se aplicará una transformación para que sus valores indiquen con un 1 a si el producto no nació con vida y con 0 en otro caso. Dicho procedimiento nos permite plantear la relación de las demás variables de la base para con esta a través de una serie de modelos lineales generalizados.

En concreto, se puede pensar que si  $Y_i$  denota el estatus de mortalidad del  $i$ -ésimo individuo (es decir,



$Y_i = Mortalidad_i$ ), de manera que para cada  $i = 1, \dots, 535$ :

$$Y_i | \mu_i \sim Ber(\mu_i)$$

Consideraremos a una función liga:

$$g(\mu_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} \\ + \beta_8 + \beta_8 X_{8i} + \beta_9 X_{9i} + \beta_{10} X_{10i} + \beta_{11} X_{11i}$$

En donde los términos involucrado tienen el significado descrito a continuación:

- $Y_i$  : Estatus de mortalidad del producto en el  $i$ -ésimo embarazado documentado en la base,  $i \in \{1, \dots, 535\}$ .
- $\beta_0$  : ordenada.
- El parámetro  $\beta_{iX_{li}}$  refiere a una variable dummy asociada a:
  - Para  $l = 1$ , a un indicador del uso de anestesia en el procedimiento de parto (variable *Anestesia*), donde  $i = 1, 2$ .
  - Para  $l = 2$ , a un indicador de la presencia de complicaciones en el embarazo (variable *Complica*), donde  $i = 1, 2$ .
  - Para  $l = 4$ , al nivel de escolaridad de la madre (variable *Escolaridad*), donde  $i = 1, 2, 3, 4$ .
  - Para  $l = 5$ , al estado civil de la madre (variable *Estadocivil*), en donde donde  $i = 1, 2, 3, 4$ .
  - Para  $l = 6$ , si un indicador de si el nacimiento se dió por cesárea (variable *Forma\_nac*), donde  $i = 1, 2$ .
  - Para  $l = 7$ , a hospital en el que se atendió el embarazo (variable *Hospital*), con  $i = 1, 2, 3, 4, 5, 6, 7$ .
  - Para  $l = 8$ , a un indicador del motivo por el que se dio la cesárea (variable *Ind\_cesarea*), donde  $i = 1, 2, 3, 4, 5, 6$ .
  - Para  $l = 9$ , a un indicador de si hubo monitoreo en el embarazo (variable *Monitoreo*), con  $i = 1, 2$ .
- En complemento, para  $i = 3$  el parámetro  $\beta_{iX_{li}}$  corresponde a la edad de la madre (variable *Edad*),
- Asimismo, para  $i = 10$  el parámetro  $\beta_{iX_{li}}$  corresponde al peso de la madre (variable *Peso\_madre*)
- Para  $i = 11$  el parámetro  $\beta_{iX_{li}}$  corresponde al peso del producto (variable *Peso\_producto*)

Cabe destacar que para que el modelo se encuentre bien definido, se deben incorporar condiciones de estimabilidad para que la matriz de diseño. Ello se traduce en ella definición del ajuste de sus coeficientes, a saber:

- $\beta_0^* < -\beta_0 - \sum_{l=1,2,4,5,6,7,8,9} \bar{\beta}_{iX_{li}}$
- $\beta_{iX_{li}}^* < -\beta_{iX_{li}} - \bar{\beta}_{iX_{li}}$  para  $l = 1, 2, 4, 5, 6, 7, 8, 9$

Ahora bien, de acuerdo a lo expuesto en clase el modelo Bernoulli puede abordarse a través de

diferentes funciones ligas tales como 1) logit, 2) probit, 3) log-log y 4) log-log complementario. En concreto, recordemos dichas funciones tiene la expresión siguiente:

- **Logística:**  $g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$
- **Probit:**  $g(\mu_i) = \Phi^{-1}(\mu_i)$
- **Log-log:**  $g(\mu_i) = \log(-\log(\mu_i))$
- **Log-log complementaria:**  $g(\mu_i) = \log(-\log(1 - \mu_i))$

En vista de lo anterior, la estrategia para ajustar los datos al modelo de descrito será aprovechar el conjunto de ligas recién descritas, de modo que podamos realizar aproximaciones numéricas de sus parámetros empleando el método MCMC, a través de JAGS.

Por otra parte, en JAGS para el ajuste del modelo Bernoulli con las ligas mencionadas previamente se puede consultar tanto en el Anexo que acompaña al presente documento. Cabe destacar que la implementación se basa a su vez en las siguientes premisas:

- Se ha considerado un número de 50,000 simulaciones, considerando un periodo de calentamiento de 10% de ellas.
- Para cada coeficiente involucrado  $\beta_i$  se ha asumido que  $\beta_i \sim \mathcal{N}(0, 0.01)$
- Se ha considerado adelgazamiento de 1-en-1 puntos
- Cabe destacar que la convergencia de las cadenas en los modelos en cuestión se evaluó a través del comando *traceplot*, así como de *summary* que provee JAGS. Para mayor referencia véase el código que se anexa al presente documento (en concreto, aquella que se refiere los archivo *models\_preagncy.R*, "*model\_ber.txt*", "*model\_ber\_probit.txt*", "*model\_ber\_loglog.txt*" y "*model\_ber\_cloglog.txt*").

## Comparación de ajuste del modelo Bernoulli para las diferentes ligas

A través de la implementación realizada, se obtuvieron los siguientes valores para el DIC y la Pseudo  $R^2$ :

#	Liga del modelo	DIC	Pseudo $R^2$
1	Logística	132.8331	0.6483502
2	Probit	136.2807	0.619516
3	Log-log	133.0388	0.5788448
4	Log-log complementaria	130.8704	0.6472392

De ello se desprende que el modelo que mejor se ajusta, en términos de un valor menor para el DIC es aquel que posee la liga log-log complementaria. Por otra parte, que observa el mayor valor de la pseudo  $R^2$  es el modelo con liga logística, seguido del modelo con una liga de tipo log-log complementaria.

En consecuencia, el modelo que tiene un mejor ajuste a los datos dadas las hipótesis consideradas previamente es aquel que tiene a la liga log-log complementaria. A continuación se presentan los principales resultados del mismo.

## Resultados para el modelo ajuste del modelo Bernoulli con liga log-log complementaria

A continuación presentamos el valor medio de cada uno de los parámetros estimados con el modelo Bernoulli con liga log-log complementaria. Se presenta además la desviación estándar correspondiente, junto con los límites izquierdo y derecho del intervalo de confianza (al 95%) alrededor de su valor medio.

**Nota:** Tales valores se han extraído con el comando *summary*

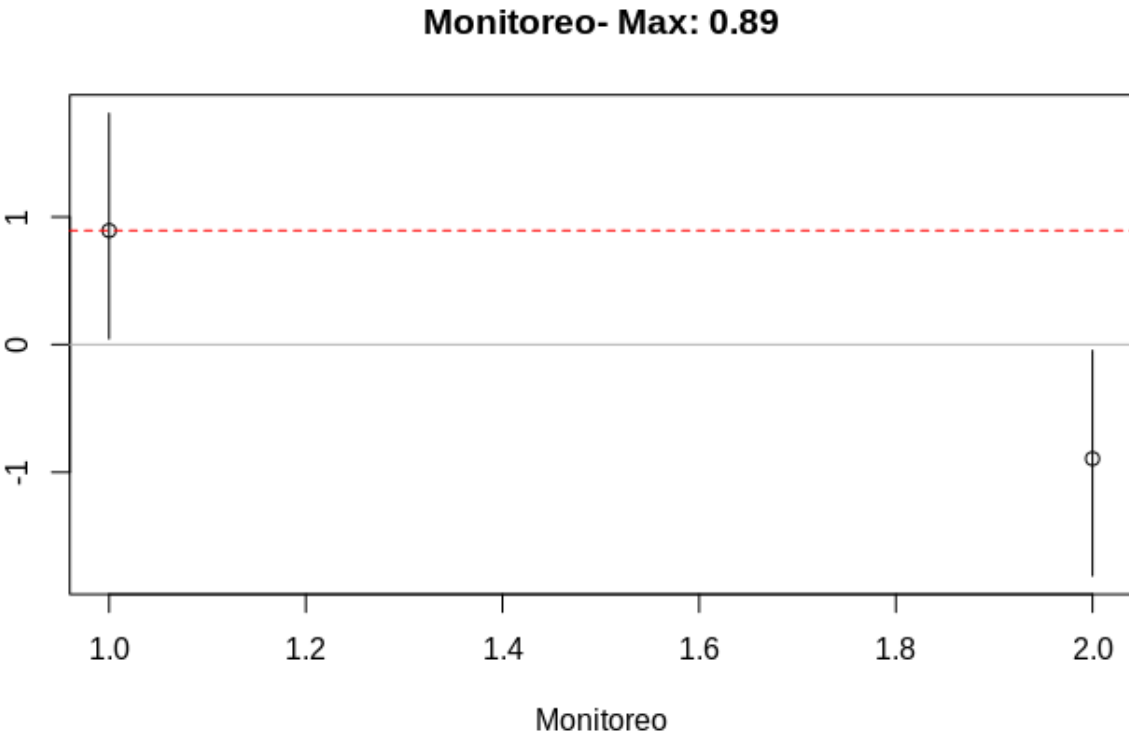
Parámetro	Valor medio	Desviación estándar	Límite izquierdo	Límite derecho
beta0.est	-3.340829e+00	8.028844e+00	-1.987927e+01	1.226441e+01
beta1.est[1]	-4.949201e-01	5.515235e-01	-1.657546e+00	5.415444e-01
beta1.est[2]	4.949201e-01	5.515235e-01	-5.415444e-01	1.657546e+00
beta2.est[1]	1.063146e-01	3.258782e-01	-5.479406e-01	7.548964e-01
beta2.est[2]	-1.063146e-01	3.258782e-01	-7.548964e-01	5.479406e-01
beta3	-4.025560e-03	4.363326e-02	-9.275642e-02	8.090774e-02
beta4.est[1]	9.663994e-01	5.760313e-01	-8.174542e-02	2.139183e+00
beta4.est[2]	-5.482488e-01	4.974288e-01	-1.520350e+00	4.081204e-01
beta4.est[3]	-1.110186e+00	6.527851e-01	-2.431221e+00	7.587208e-02
beta4.est[4]	6.920357e-01	9.407923e-01	-1.295643e+00	2.371928e+00
beta5.est[1]	2.795521e+00	1.909690e+00	7.804501e-02	7.275604e+00
beta5.est[2]	-5.463143e+00	5.600455e+00	-1.848545e+01	2.321456e+00
beta5.est[3]	1.339687e+00	1.963492e+00	-1.683882e+00	5.782715e+00
beta5.est[4]	1.327934e+00	1.897121e+00	-1.400809e+00	5.689123e+00
beta6.est[1]	-6.074358e-01	4.047601e+00	-8.363664e+00	7.599564e+00
beta6.est[2]	6.074358e-01	4.047601e+00	-7.599564e+00	8.363664e+00
beta7.est[1]	-8.530332e+00	5.396362e+00	-2.059686e+01	-5.059152e-02
beta7.est[2]	-6.486729e+00	5.949268e+00	-2.009271e+01	2.238984e+00

Parámetro	Valor medio	Desviación estándar	Límite izquierdo	Límite derecho
beta7.est[3]	6.480875e+00	2.087299e+00	2.732098e+00	1.096233e+01
beta7.est[4]	1.223689e+00	2.434924e+00	-3.340029e+00	6.237179e+00
beta7.est[5]	3.997013e+00	2.093478e+00	3.679943e-01	8.613699e+00
beta7.est[6]	-3.136051e+00	6.875616e+00	-1.910715e+01	7.357965e+00
beta7.est[7]	6.451534e+00	2.168480e+00	2.565039e+00	1.115988e+01
beta8.est[1]	4.798101e-01	2.157767e+00	-3.726291e+00	4.905401e+00
beta8.est[2]	2.568727e+00	2.286765e+00	-1.995509e+00	7.073279e+00
beta8.est[3]	1.586956e+00	1.985704e+00	-2.060778e+00	5.807381e+00
beta8.est[4]	1.989062e+00	2.005850e+00	-1.690070e+00	6.298552e+00
beta8.est[5]	-7.266184e+00	5.650868e+00	-2.044422e+01	9.630456e-01
beta8.est[6]	6.416292e-01	6.718081e+00	-1.213620e+01	1.412495e+01
beta9.est[1]	8.932816e-01	4.484159e-01	4.647354e-02	1.810533e+00
beta9.est[2]	-8.932816e-01	4.484159e-01	-1.810533e+00	-4.647354e-02
beta10	-1.296104e-02	1.984103e-02	-5.031676e-02	2.575749e-02
beta11	-2.733680e+00	4.262541e-01	-3.633241e+00	-1.954164e+00

Por otra parte, a continuación se presenta un gráfico de los puntos más relevantes encontrados respecto a los parámetros presentados en dicha tabla:

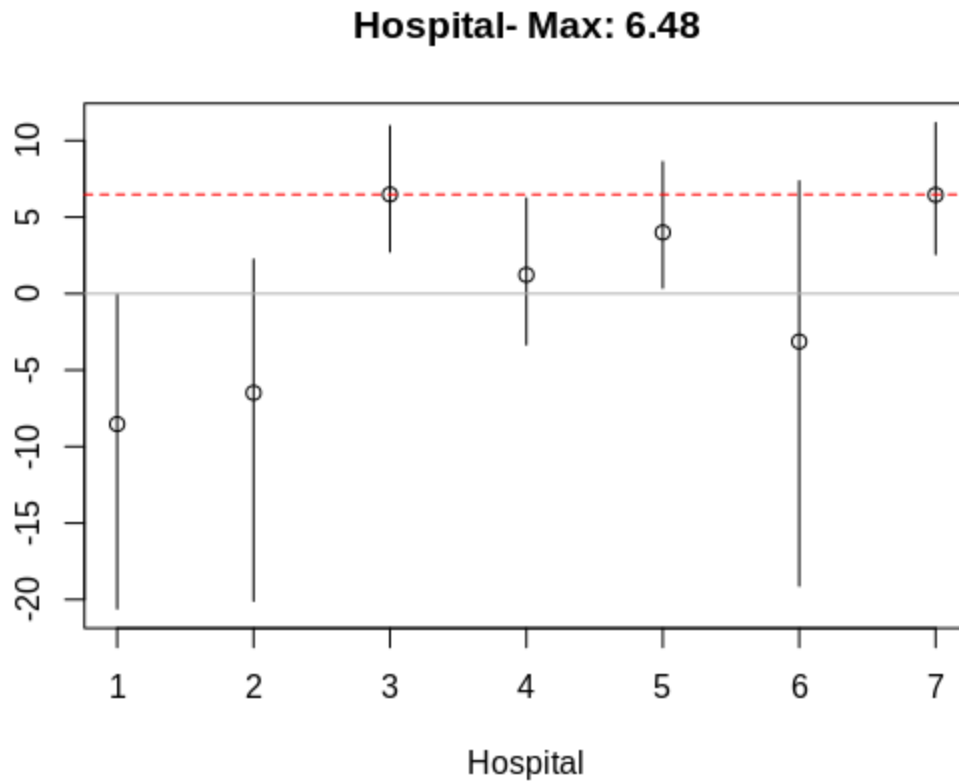
### Monitoreo

Al respecto de la variable asociada al monitoreo del embarazo, se encontró que tanto el monitoreo y la ausencia de este en el embarazo son significativas (dado que el intervalo de confianza a su alrededor no contiene al cero). En este sentido, se aprecia que el monitoreo del embarazo es un factor que favorece la probabilidad de vida del producto, mientras que la ausencia del monitoreo resulta un factor en su contra.



**Hospital**

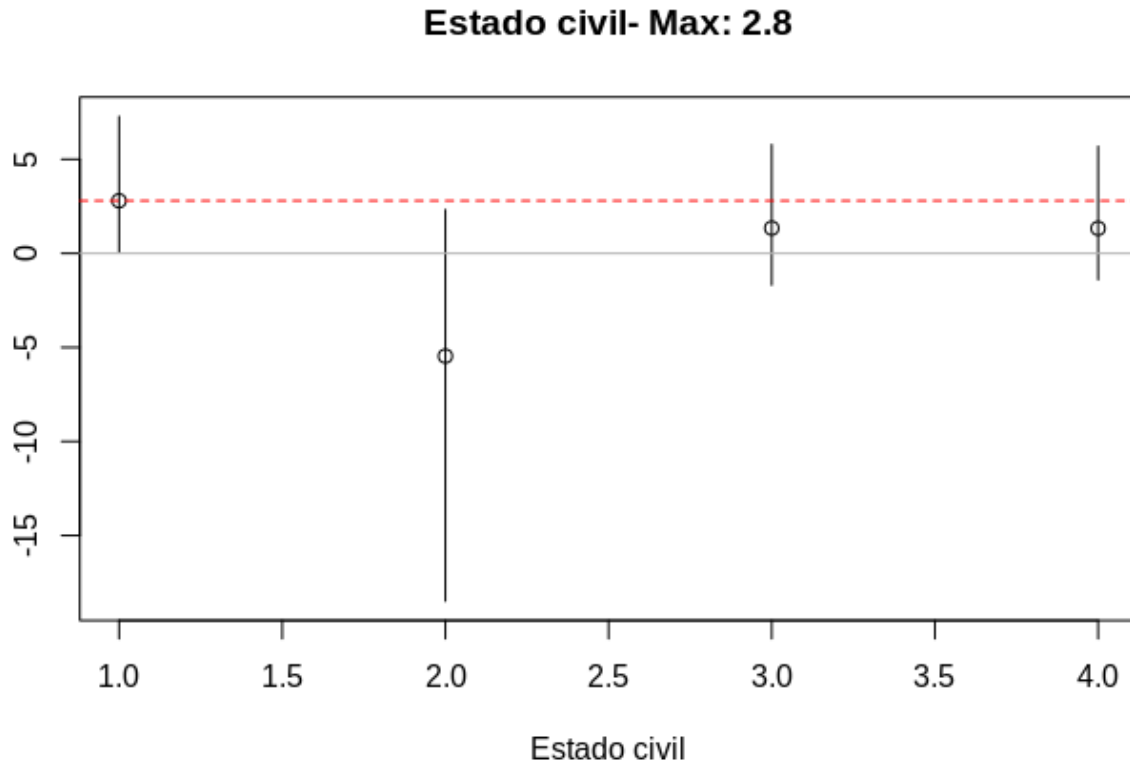
Al respecto de la variable asociada al hospital donde se atendió el embarazo, se encontró que tanto la atención en hospitales 1) General, 2) Oaxaca y 3) Tlaxcala son significativos. En este sentido, se aprecia que la atención en estos favorece la probabilidad de muerte en el producto.



Coeficientes de variable asociada a hospitales

### Estado Civil

Al respecto de la variable asociada al estado civil de la madre, se encontró que tanto la atención el estado civil “casada” es significativo. En este sentido, se aprecia que el estar casada favorece la probabilidad de muerte en el producto.



Coeficientes de variable asociada al estado civil

## Conclusiones

- En este documento se ha expuesto el ajuste realizado un análisis a un conjunto de datos gestacionales del sistema de salud del país, a través de análisis exploratorio de datos.
- A partir de dicha informacion se realizó un análisis estadístico inferencia con base en la teoría bayesiana de dicha base de datos, en el cual se ajustó una familia de modelos Bernoulli donde la variable explicativa es la mortalidad de feto y las covariables son el resto de características de los casos de embarazo documentados, entre las que se incluyen datos de la educación, estado civil, factores de salud y de atención médica del embarazo.
- Para el ajuste del modelo de tipo Bernoulli se exploraron variantes del mismo empleando diferentes funciones liga: a saber, 1) logit, 2) probit, 3) log-log y 4) log-log complementario; encontrándose que ésta última es la que mejor se ajusta al modelo al considerar el DIC.
- Finalmente, se encontraron diversos factores que en términos de probabilidad favorecen la mortalidad del producto, tales como la ausencia de monitoreo en el embarazo.

## Anexo de códigos de R empleados para este documento

# Implementación del código para Jags - Modelo Bernoulli con liga logística

**Archivo: model\_ber.txt**



```

model
{
  #Likelihood
  for (i in 1:n) {
    y[i] ~ dbern(mu[i])
    eta[i] <- beta0+beta1[Anestesia[i]]+beta2[Complicacion[i]]+beta3*Edad[i] +beta4[Escolaridad[i]]+beta5[EstadoCivil[i]]+beta6[Forma_nac[i]]+beta7[Hospital[i]]+
      beta8[Ind_cesarea[i]]+beta9[Monitoreo[i]]+beta10*PesoMadre[i]+beta11*PesoProd[i]
    mu[i] <- exp(eta[i])/(1+exp(eta[i]))
  }
  #Priors
  beta0 ~ dnorm(0,0.001) # Intercepto
  for (i in 1:2) {beta1[i] ~ dnorm(0,0.001)} # Anestesia
  for (i in 1:2) {beta2[i] ~ dnorm(0,0.01)} # Complica
  beta3 ~ dnorm(0,100) # Edad
  for (i in 1:4) {beta4[i] ~ dnorm(0,0.01)} # Escolaridad
  for (i in 1:4) {beta5[i] ~ dnorm(0,0.01)} # Estado civil
  for (i in 1:2) {beta6[i] ~ dnorm(0,0.01)} # Forma_nac
  for (i in 1:7) {beta7[i] ~ dnorm(0,0.01)} # Hospital
  for (i in 1:6) {beta8[i] ~ dnorm(0,0.01)} # Ind_cesarea
  for (i in 1:2) {beta9[i] ~ dnorm(0,0.01)} # Monitoreo
  beta10 ~ dnorm(0,0.01) # Peso madre
  beta11 ~ dnorm(0,0.01) # Peso producto

  # Condiciones de estimabilidad
  beta0.est <- beta0+mean(beta1[])+mean(beta2[])+mean(beta3[])+mean(beta4[])+
    mean(beta5[])+mean(beta6[])+mean(beta6[])+mean(beta7[])+mean(beta8[])+mean(beta9[])
  for (i in 1:2) {beta1.est[i] <- beta1[i]-mean(beta1[])} # Anestesia
  for (i in 1:2) {beta2.est[i] <- beta2[i]-mean(beta2[])} # Complica
  for (i in 1:4) {beta4.est[i] <- beta4[i]-mean(beta4[])} # Escolaridad
  for (i in 1:4) {beta5.est[i] <- beta5[i]-mean(beta5[])} # Estado civil
  for (i in 1:2) {beta6.est[i] <- beta6[i]-mean(beta6[])} # Forma_nac
  for (i in 1:7) {beta7.est[i] <- beta7[i]-mean(beta7[])} # Hospital
  for (i in 1:6) {beta8.est[i] <- beta8[i]-mean(beta8[])} # Ind_cesarea
  for (i in 1:2) {beta9.est[i] <- beta9[i]-mean(beta9[])} # Monitoreo

  #Predictive
  for (i in 1:n) {
    yf[i] ~ dbern(mu[i])
  }
}

```

# Implementación del código para Jags - Modelo Bernoulli con liga probit

**Archivo:** model\_ber\_probit.txt

```

model
{
  #Likelihood
  for (i in 1:n) {
    y[i] ~ dbern(mu[i])
    eta[i] <- beta0+beta1[Anestesia[i]]+beta2[Complicacion[i]]+beta3*Edad[i] +beta4[Escolaridad[i]]+beta5[EstadoCivil[i]]+beta6[Forma_nac[i]]+
      beta7[Hospital[i]]+beta8[Ind_cesarea[i]]+beta9[Monitoreo[i]]+
      beta10*PesoMadre[i]+beta11*PesoProd[i]
    mu[i] <- phi(eta[i])
  }

  #Priors
  beta0 ~ dnorm(0,0.001) # Intercepto
  for (i in 1:2) {beta1[i] ~ dnorm(0,0.001)} # Anestesia
  for (i in 1:2) {beta2[i] ~ dnorm(0,0.01)} # Complica
  beta3 ~ dnorm(0,100) # Edad
  for (i in 1:4) {beta4[i] ~ dnorm(0,0.01)} # Escolaridad
  for (i in 1:4) {beta5[i] ~ dnorm(0,0.01)} # Estado civil
  for (i in 1:2) {beta6[i] ~ dnorm(0,0.01)} # Forma_nac
  for (i in 1:7) {beta7[i] ~ dnorm(0,0.01)} # Hospital
  for (i in 1:6) {beta8[i] ~ dnorm(0,0.01)} # Ind_cesarea
  for (i in 1:2) {beta9[i] ~ dnorm(0,0.01)} # Monitoreo
  beta10 ~ dnorm(0,0.01) # Peso madre
  beta11 ~ dnorm(0,0.01) # Peso producto

  # Condiciones de estimabilidad
  beta0.est <- beta0+mean(beta1[])+mean(beta2[])+mean(beta3[])+mean(beta4[])+
    mean(beta5[])+mean(beta6[])+mean(beta6[])+mean(beta7[])+mean(beta8[])+mean(beta9[])
  for (i in 1:2) {beta1.est[i] <- beta1[i]-mean(beta1[])} # Anestesia
  for (i in 1:2) {beta2.est[i] <- beta2[i]-mean(beta2[])} # Complica
  for (i in 1:4) {beta4.est[i] <- beta4[i]-mean(beta4[])} # Escolaridad
  for (i in 1:4) {beta5.est[i] <- beta5[i]-mean(beta5[])} # Estado civil
  for (i in 1:2) {beta6.est[i] <- beta6[i]-mean(beta6[])} # Forma_nac
  for (i in 1:7) {beta7.est[i] <- beta7[i]-mean(beta7[])} # Hospital
  for (i in 1:6) {beta8.est[i] <- beta8[i]-mean(beta8[])} # Ind_cesarea
  for (i in 1:2) {beta9.est[i] <- beta9[i]-mean(beta9[])} # Monitoreo

  #Predictive
  for (i in 1:n) {
    yf[i] ~ dbern(mu[i])
  }
}

```

# Implementación del código para Jags - Modelo Bernoulli con liga probit

**Archivo: model\_ber\_loglog.txt**

```

model
{
  #Likelihood
  for (i in 1:n) {
    y[i] ~ dbern(mu[i])
    eta[i] <- beta0 + beta1[Anestesia[i]] + beta2[Complicacion[i]] + beta3*Edad[i] + be
ta4[Escolaridad[i]] + beta5[EstadoCivil[i]] + beta6[Forma_nac[i]] +
    beta7[Hospital[i]] + beta8[Ind_cesarea[i]] + beta9[Monitoreo[i]] +
    beta10*PesoMadre[i] + beta11*PesoProd[i]
    mu[i] <- exp(-exp(eta[i]))
  }
  #Priors
  beta0 ~ dnorm(0,0.001) # Intercepto
  for (i in 1:2) {beta1[i] ~ dnorm(0,0.001)} # Anestesia
  for (i in 1:2) {beta2[i] ~ dnorm(0,0.01)} # Complica
  beta3 ~ dnorm(0,100) # Edad
  for (i in 1:4) {beta4[i] ~ dnorm(0,0.01)} # Escolaridad
  for (i in 1:4) {beta5[i] ~ dnorm(0,0.01)} # Estado civil
  for (i in 1:2) {beta6[i] ~ dnorm(0,0.01)} # Forma_nac
  for (i in 1:7) {beta7[i] ~ dnorm(0,0.01)} # Hospital
  for (i in 1:6) {beta8[i] ~ dnorm(0,0.01)} # Ind_cesarea
  for (i in 1:2) {beta9[i] ~ dnorm(0,0.01)} # Monitoreo
  beta10 ~ dnorm(0,0.01) # Peso madre
  beta11 ~ dnorm(0,0.01) # Peso producto

  # Condiciones de estimabilidad
  beta0.est <- beta0 + mean(beta1[]) + mean(beta2[]) + mean(beta3[]) +
    mean(beta4[]) + mean(beta5[]) + mean(beta6[]) + mean(beta6[]) +
    mean(beta7[]) + mean(beta8[]) + mean(beta9[])
  for (i in 1:2) {beta1.est[i] <- beta1[i] - mean(beta1[])} # Anestesia
  for (i in 1:2) {beta2.est[i] <- beta2[i] - mean(beta2[])} # Complica
  for (i in 1:4) {beta4.est[i] <- beta4[i] - mean(beta4[])} # Escolaridad
  for (i in 1:4) {beta5.est[i] <- beta5[i] - mean(beta5[])} # Estado civil
  for (i in 1:2) {beta6.est[i] <- beta6[i] - mean(beta6[])} # Forma_nac
  for (i in 1:7) {beta7.est[i] <- beta7[i] - mean(beta7[])} # Hospital
  for (i in 1:6) {beta8.est[i] <- beta8[i] - mean(beta8[])} # Ind_cesarea
  for (i in 1:2) {beta9.est[i] <- beta9[i] - mean(beta9[])} # Monitoreo

  #Predictive
  for (i in 1:n) {
    yf[i] ~ dbern(mu[i])
  }
}

```

## Implementación del código para Jags - Modelo Bernoulli

con liga probit

Archivo: model\_ber\_cloglog.txt

```

model
{
  #Likelihood
  for (i in 1:n) {
    y[i] ~ dbern(mu[i])
    eta[i] <- beta0+beta1[Anestesia[i]]+beta2[Complicacion[i]]+beta3*Edad[i] +beta4[Escolaridad[i]]+beta5[EstadoCivil[i]]+beta6[Forma_nac[i]]+
      beta7[Hospital[i]]+beta8[Ind_cesarea[i]]+beta9[Monitoreo[i]]+
      beta10*PesoMadre[i]+beta11*PesoProd[i]
    cloglog(mu[i])<-eta[i]
  }
  #Priors
  beta0 ~ dnorm(0,0.001) # Intercepto
  for (i in 1:2) {beta1[i] ~ dnorm(0,0.001)} # Anestesia
  for (i in 1:2) {beta2[i] ~ dnorm(0,0.01)} # Complica
  beta3 ~ dnorm(0,100) # Edad
  for (i in 1:4) {beta4[i] ~ dnorm(0,0.01)} # Escolaridad
  for (i in 1:4) {beta5[i] ~ dnorm(0,0.01)} # Estado civil
  for (i in 1:2) {beta6[i] ~ dnorm(0,0.01)} # Forma_nac
  for (i in 1:7) {beta7[i] ~ dnorm(0,0.01)} # Hospital
  for (i in 1:6) {beta8[i] ~ dnorm(0,0.01)} # Ind_cesarea
  for (i in 1:2) {beta9[i] ~ dnorm(0,0.01)} # Monitoreo
  beta10 ~ dnorm(0,0.01) # Peso madre
  beta11 ~ dnorm(0,0.01) # Peso producto
  phy ~ dgamma(0.01,0.01)
  # Condiciones de estimabilidad
  beta0.est <- beta0+mean(beta1[])+mean(beta2[])+mean(beta3[])+
    mean(beta4[])+mean(beta5[])+mean(beta6[])+mean(beta6[])+
    mean(beta7[])+mean(beta8[])+mean(beta9[])
  for (i in 1:2) {beta1.est[i] <- beta1[i]-mean(beta1[])} # Anestesia
  for (i in 1:2) {beta2.est[i] <- beta2[i]-mean(beta2[])} # Complica
  for (i in 1:4) {beta4.est[i] <- beta4[i]-mean(beta4[])} # Escolaridad
  for (i in 1:4) {beta5.est[i] <- beta5[i]-mean(beta5[])} # Estado civil
  for (i in 1:2) {beta6.est[i] <- beta6[i]-mean(beta6[])} # Forma_nac
  for (i in 1:7) {beta7.est[i] <- beta7[i]-mean(beta7[])} # Hospital
  for (i in 1:6) {beta8.est[i] <- beta8[i]-mean(beta8[])} # Ind_cesarea
  for (i in 1:2) {beta9.est[i] <- beta9[i]-mean(beta9[])} # Monitoreo

  #Predictive
  for (i in 1:n) {
    yf[i] ~ dbern(mu[i])
  }
}

```

## Implementación del código para Jags - para

# correr los modelos Bernoulli con diferentes ligas

Archivo: models\_preagnancy.R



```
#--- Carga de paquetes necesarios ---
library(R2openBUGS)
library(R2jags)
library(tidyverse)

#--- Funciones utiles ---
prob<-function(x){
  out<-min(length(x[x>0])/length(x),length(x[x<0])/length(x))
  out
}

#-Working directory-
wdir<-"/media/Box/Modelos_Lineales_Generalizados/Examen_midterm"
setwd(wdir)

#-Reading data-

# Para ejectuar este archivo, por favor situarse en la ruta donde esta el arc
hivo csv con los datos
# Cargamos los datos limpios
df <- read_csv("/media/Box/Modelos_Lineales_Generalizados/Examen_midterm/Gest
acional.csv")

##### ----- Se convierten a variables caracteres las variables categoricas
----- #####
df$Anestesia <- as.character(df$Anestesia)
df$Complica <- as.character(df$Complica)
df$Escolaridad <- as.character(df$Escolaridad)
df$Estado_civil <- as.character(df$Estado_civil)
df$Forma_nac <- as.character(df$Forma_nac)
df$Hospital <- as.character(df$Hospital)
df$Ind_cesarea <- as.character(df$Ind_cesarea)
df$Monitoreo <- as.character(df$Monitoreo)

# Recorremos la variable de mortalidad
df$Mortalidad<- df$Mortalidad-1

# Creamos una nueva variable para el modelo gamma
df$Mortalidad_Mod<- df$Mortalidad
for (i in 1:n){
  df$Mortalidad_Mod[i] = if_else(df$Mortalidad_Mod[i]==0,0.1,0.9)
}

excluded_vars <- c("Mortalidad")
df_filter <- select(df, -one_of(excluded_vars))
```

```
##### ----- Parametros de la simulacion ----- #####

# Parametros de la simulacion
s= 50000 # numero de puntos a obtenerse en la cadena
umbral_calentamiento = 0.1 # % puntos a quemar en periodo de calentamiento
s_cal = s*umbral_calentamiento # numero de puntos a quemar en periodo de cale
ntamiento

#-Defining data-
data<-list("n"=n,"y"=df$Mortalidad,"Anestesia"=df_filter$Anestesia,"Complicac
ion"=df_filter$Complica,
          "Edad"=df_filter$Edad,"Escolaridad"=df_filter$Escolaridad,"EstadoC
ivil"=df_filter$Estado_civil,
          "Forma_nac"=df_filter$Forma_nac,"Hospital"=df_filter$Hospital,"Ind
_cesarea"=df_filter$Ind_cesarea,
          "Monitoreo"=df_filter$Monitoreo,"PesoMadre"=df_filter$Peso_madre,"
PesoProd"=df_filter$Peso_prod)

data2<-list("n"=n,"y"=df$Mortalidad_Mod,"Anestesia"=df_filter$Anestesia,"Comp
licacion"=df_filter$Complica,
          "Edad"=df_filter$Edad,"Escolaridad"=df_filter$Escolaridad,"EstadoC
ivil"=df_filter$Estado_civil,
          "Forma_nac"=df_filter$Forma_nac,"Hospital"=df_filter$Hospital,"Ind
_cesarea"=df_filter$Ind_cesarea,
          "Monitoreo"=df_filter$Monitoreo,"PesoMadre"=df_filter$Peso_madre,"
PesoProd"=df_filter$Peso_prod)

#-Defining inits-
inits_bernoulli<-function(){list(beta0=0, beta1=rep(0,2),beta2=rep(0,2),beta
3=0,
                                beta4=rep(0,4),beta5=rep(0,4),beta6=rep(0,
2),beta7=rep(0,7),
                                beta8=rep(0,6),beta9=rep(0,2),beta10=0,beta1
1=0,yf=rep(1,n))}

#-Selecting parameters to monitor-
pars_bernoulli<-c("beta0.est","beta1.est","beta2.est","beta3","beta4.est","be
ta5.est",
                  "beta6.est","beta7.est","beta8.est","beta9.est","beta10","b
eta11","mu","yf")

##### ----- Running code ----- #####
```

```

#--- OpenBUGS
# Familia de modelos binomiales con diferentes ligas
model.ber_log<-jags(data,init_bernoulli, pars_bernoulli,model.file="model_ber_log.txt",n.iter=s,n.chains=2,n.burnin=s_cal)
model.ber_probit<-jags(data,init_bernoulli, pars_bernoulli,model.file="model_ber_probit.txt",n.iter=s,n.chains=2,n.burnin=s_cal)
model.ber_loglog<-jags(data,init_bernoulli, pars_bernoulli,model.file="model_ber_loglog.txt",n.iter=s,n.chains=2,n.burnin=s_cal)
model.ber_cloglog<-jags(data,init_bernoulli, pars_bernoulli,model.file="model_ber_cloglog.txt",n.iter=s,n.chains=2,n.burnin=s_cal)

# ---- Traceplot

# Monitoreamos la convergencia de las cadenas

traceplot(model.ber_log)
traceplot(model.ber_probit)
traceplot(model.ber_loglog)
traceplot(model.ber_cloglog)

#model.beta_log<-jags(data2,init_beta,pars_beta,model.file="model_beta_logit.txt",n.iter=s,n.chains=2,n.burnin=s_cal)

saveRDS(model.ber_log,"model.ber_log")
saveRDS(model.ber_probit,"model.ber_probit")
saveRDS(model.ber_loglog,"model.ber_loglog")
saveRDS(model.ber_cloglog,"model.ber_cloglog")

# DIC de los modelos
model.ber_log$BUGSoutput$DIC
model.ber_probit$BUGSoutput$DIC
model.ber_loglog$BUGSoutput$DIC
model.ber_cloglog$BUGSoutput$DIC

#- Pseudo R2 - cuadrada
# log
out.yf <- model.ber_log$BUGSoutput$summary[grep("yf",rownames(model.ber_log$BUGSoutput$summary)),]
R2<-(cor(df$Mortalidad, out.yf[,1]))^2
print(R2)

# probit
out.yf <- model.ber_probit$BUGSoutput$summary[grep("yf",rownames(model.ber_probit$BUGSoutput$summary)),]
R2 <-(cor(df$Mortalidad, out.yf[,1]))^2

```

```

print(R2)

# loglog
out.yf <- model.ber_loglog$BUGSoutput$summary[grep("yf",rownames(model.ber_lo
glog$BUGSoutput$summary)),]
R2<-(cor(df$Mortalidad, out.yf[,1]))^2
print(R2)

# cloglog
out.yf <- model.ber_cloglog$BUGSoutput$summary[grep("yf",rownames(model.ber_c
loglog$BUGSoutput$summary)),]
R2<-(cor(df$Mortalidad, out.yf[,1]))^2
print(R2)

#-- Summarys de los modelos
model.ber_log$BUGSoutput$summary
model.ber_probit$BUGSoutput$summary
model.ber_loglog$BUGSoutput$summary
model.ber_cloglog$BUGSoutput$summary

# Imprimimos el summary del modelo Bernoulli con liga cloglog
options(max.print = 1000000000)
sink("summary_ber_cloglog.txt")
print(model.ber_cloglog$BUGSoutput$summary[,c(1,2,3,7)])
sink()

#--- Plots del efecto de las variables en la mortalidad

# Anestesia
plot1 <- model.ber_cloglog$BUGSoutput$summary[grep("beta1.est",rownames(mode
l.ber_cloglog$BUGSoutput$summary)),] # devuelve la tabla de

k<-2
ymin<-min(plot1[,c(1,3,7)])
ymax<-max(plot1[,c(1,3,7)])
plot(1:k,plot1[,1],xlab="Anestesia",ylab="",ylim=c(ymin,ymax))
segments(1:k,plot1[,3],1:k,plot1[,7])
abline(h=0,col="grey70")
abline(h=max(plot1[,c(1)]) ,col="red",lwd=1, lty=2)
title(paste0("Precio: efecto Anestesia", "- Max: ", round(max(plot1[,c(1)]),
2) ))

# Complica
plot1 <- model.ber_cloglog$BUGSoutput$summary[grep("beta2.est",rownames(mode

```

```

l.ber_cloglog$BUGSoutput$summary)),] # devuelve la tabla de

k<-2
ymin<-min(plot1[,c(1,3,7)])
ymax<-max(plot1[,c(1,3,7)])
plot(1:k,plot1[,1],xlab="Complica",ylab="",ylim=c(ymin,ymax))
segments(1:k,plot1[,3],1:k,plot1[,7])
abline(h=0,col="grey70")
abline(h=max(plot1[,c(1)]),col="red",lwd=1, lty=2)
title(paste0("Precio: Complica", "- Max: ", round(max(plot1[,c(1)]),2) ))

# Escolaridad
plot1 <- model.ber_cloglog$BUGSoutput$summary[grep("beta4.est",rownames(mode
l.ber_cloglog$BUGSoutput$summary)),] # devuelve la tabla de

k<-4
ymin<-min(plot1[,c(1,3,7)])
ymax<-max(plot1[,c(1,3,7)])
plot(1:k,plot1[,1],xlab="Escolaridad",ylab="",ylim=c(ymin,ymax))
segments(1:k,plot1[,3],1:k,plot1[,7])
abline(h=0,col="grey70")
abline(h=max(plot1[,c(1)]),col="red",lwd=1, lty=2)
title(paste0("Escolaridad", "- Max: ", round(max(plot1[,c(1)]),2) ))

# Estado civil
plot1 <- model.ber_cloglog$BUGSoutput$summary[grep("beta5.est",rownames(mode
l.ber_cloglog$BUGSoutput$summary)),] # devuelve la tabla de

k<-4
ymin<-min(plot1[,c(1,3,7)])
ymax<-max(plot1[,c(1,3,7)])
plot(1:k,plot1[,1],xlab="Estado civil",ylab="",ylim=c(ymin,ymax))
segments(1:k,plot1[,3],1:k,plot1[,7])
abline(h=0,col="grey70")
abline(h=max(plot1[,c(1)]),col="red",lwd=1, lty=2)
title(paste0("Estado civil", "- Max: ", round(max(plot1[,c(1)]),2) ))

# Forma_nac
plot1 <- model.ber_cloglog$BUGSoutput$summary[grep("beta6.est",rownames(mode
l.ber_cloglog$BUGSoutput$summary)),] # devuelve la tabla de

k<-2
ymin<-min(plot1[,c(1,3,7)])
ymax<-max(plot1[,c(1,3,7)])
plot(1:k,plot1[,1],xlab="Forma de Nacimiento",ylab="",ylim=c(ymin,ymax))
segments(1:k,plot1[,3],1:k,plot1[,7])

```

```
abline(h=0,col="grey70")
abline(h=max(plot1[,c(1)]),col="red",lwd=1, lty=2)
title(paste0("Forma de Nacimiento", "- Max: ", round(max(plot1[,c(1)]),2) ))

# Hospitales
plot1 <- model.ber_cloglog$BUGSoutput$summary[grep("beta7.est",rownames(model.ber_cloglog$BUGSoutput$summary)),] # devuelve la tabla de

k<-7
ymin<-min(plot1[,c(1,3,7)])
ymax<-max(plot1[,c(1,3,7)])
plot(1:k,plot1[,1],xlab="Hospital",ylab="",ylim=c(ymin,ymax))
segments(1:k,plot1[,3],1:k,plot1[,7])
abline(h=0,col="grey70")
abline(h=max(plot1[,c(1)]),col="red",lwd=1, lty=2)
title(paste0("Hospital", "- Max: ", round(max(plot1[,c(1)]),2) ))

# Ind_cesarea
plot1 <- model.ber_cloglog$BUGSoutput$summary[grep("beta8.est",rownames(model.ber_cloglog$BUGSoutput$summary)),] # devuelve la tabla de

k<-6
ymin<-min(plot1[,c(1,3,7)])
ymax<-max(plot1[,c(1,3,7)])
plot(1:k,plot1[,1],xlab="Indicador cesarea",ylab="",ylim=c(ymin,ymax))
segments(1:k,plot1[,3],1:k,plot1[,7])
abline(h=0,col="grey70")
abline(h=max(plot1[,c(1)]),col="red",lwd=1, lty=2)
title(paste0("Indicador cesarea", "- Max: ", round(max(plot1[,c(1)]),2) ))

# Monitoreo
plot1 <- model.ber_cloglog$BUGSoutput$summary[grep("beta9.est",rownames(model.ber_cloglog$BUGSoutput$summary)),] # devuelve la tabla de

k<-2
ymin<-min(plot1[,c(1,3,7)])
ymax<-max(plot1[,c(1,3,7)])
plot(1:k,plot1[,1],xlab="Monitoreo",ylab="",ylim=c(ymin,ymax))
segments(1:k,plot1[,3],1:k,plot1[,7])
abline(h=0,col="grey70")
abline(h=max(plot1[,c(1)]),col="red",lwd=1, lty=2)
title(paste0("Monitoreo", "- Max: ", round(max(plot1[,c(1)]),2) ))
```

## Plantilla del documento Shiny para realizar análisis exploratorio de datos

**Archivo: mlg\_shiny.R**

```
library(shiny)
library(plotly)
library(ggplot2)
library(dplyr)
library(tidyverse)

# Para ejecutar este archivo, por favor situarse en la ruta donde esta el archi
vo csv con los datos
# Cargamos los datos limpios
df1 <- read_csv("/media/Box/Modelos_Lineales_Generalizados/Examen_midterm/Gesta
cional.csv")
df1$Anestesia <- as.character(df1$Anestesia)
df1$Complica <- as.character(df1$Complica)
df1$Escolaridad <- as.character(df1$Escolaridad)
df1$Estado_civil <- as.character(df1$Estado_civil)
df1$Forma_nac <- as.character(df1$Forma_nac)
df1$Hospital <- as.character(df1$Hospital)
df1$Ind_cesarea <- as.character(df1$Ind_cesarea)
df1$Monitoreo <- as.character(df1$Monitoreo)
df1$Mortalidad <- as.character(df1$Mortalidad)
df1$Edad <- as.character(df1$Edad)

# Especificaciones de tema grafico para ggplot
.theme<- theme(
  axis.line = element_line(colour = 'gray', size = .95),
  panel.background = element_blank(),
  plot.background = element_blank()
)

# nombres de variables categoricas
categorical_names <- c("Anestesia","Complica","Edad","Escolaridad","Forma_na
c","Hospital","Ind_cesarea","Monitoreo","Mortalidad")

#####----- Diseno de la interfaz de usuario -----#####

ui<-fluidPage(
  # Application title
  titlePanel("Análisis Exploratorio de Datos"),

  # Creamos el titulo del panel
  headerPanel("Opciones de graficos"),

  #Especificaciones de panel lateral izquierdo
  sidebarPanel
```



```

( # Seleccion de inputs para el analisis: variables y tipo de grafico a despl
egar
  selectInput("dataset","Datos:", choices =list(Datos = "df1"), selected=NUL
L), # Datos en que se basa el analisis
  selectInput("variable1","Variable 1:", names(df)), # Campo para seleccionar
variable 1 a cruzar
  selectInput("variable2","Variable 2", names(df)), # Campo para seleccionar
variable 2 a cruzar
  selectInput("variable3","Variable 3 (para desagregación por grupos)", choic
es = NULL), # Campo para seleccionar variable 3 a cruzar de manera que se desag
reguen los plots con ella
  selectInput("plot.type","Tipo de grafico:", # Tipos de gráficas que permiti
rá definir el dashboard
    list(boxplot = "boxplot",
          histograma = "histogram",
          densidad = "density",
          barras = "bar",
          scatterplot = "scatterplot")
  ),
  checkboxInput("show.points", "Agregar puntos con Jitter", TRUE) # Campo par
a decidir si agregan a los boxplots los puntos con algun jitter
),

# Llamaos a los graficos al panel del dashboard
mainPanel(
  plotOutput("distPlot"), # Dasboard bivariado
  plotOutput("p"), # Dasboard bivariado
  plotOutput("q") # Dashboard multivariado
)
)

##### ----- Diseno del servidor para montar el dashboard ----- #####

server<-(function(input, output, session){

  #Actualizacion de las variables elegidas por el usuario en el panel lateral i
zquierdo
  observe({
    if(!exists(input$dataset)) return() # Revisamos que si se hallan cargado lo
s datos
    var.opts<-colnames(get(input$dataset))
    updateSelectInput(session, "variable1", choices = var.opts) # usuario varia
ble 1 - de entre todas las variables de la base
    updateSelectInput(session, "variable2", choices = var.opts) # usuario varia
ble 2 - de entre todas las variables de la base

```

```

    # usuario elige unicamente entre las variables categoricas para desglosar l
os graficos con face_wrap
    updateSelectInput(session, "variable3", choices = c("Anestesia","Complic
a","Escolaridad","Forma_nac","Hospital","Ind_cesarea","Monitoreo","Mortalidad")
    )
  })

  # Campo para mostrar etiquetas del tipo de plot mostrado en funcion de grafica
seleccionada en el panel
  output$caption<-renderText({
    switch(input$plot.type,
      "boxplot"      = "Grafica de cajas",
      "histogram"    = "Histograma",
      "density"      = "Densidad",
      "bar"          = "Grafica de Barras",
      "scatterplot"  = "Scatterplot")
  })

  #get data object
  get_data<-reactive({

    if(!exists(input$dataset)) return() # if no upload

    check<-function(x){is.null(x) || x==""}
    if(check(input$dataset)) return()

    obj<-list(data=get(input$dataset),
              variable1=input$variable1,
              variable2=input$variable2,
              variable3=input$variable3
    )

    #require all to be set to proceed
    if(any(sapply(obj,check))) return()
    #make sure choices had a chance to update
    check<-function(obj){
      !all(c(obj$variable1,obj$variable2,obj$variable3) %in% colnames(obj$dat
a))
    }

    if(check(obj)) return()

    return(obj)
  })

```

```

}))

#####----- Univariado -----#####
output$distPlot <- renderPlot({
  plot.obj<-get_data()

  # revisamos que se hayan agregado los datos especificados para hacer las gr
  aficas, sino paramos
  if(is.null(plot.obj)) return()

  # verificamos que el usuario ha establecido la variable a cruzar
  if(plot.obj$variable1 == "") return()

  if(!plot.obj$variable1 %in% categorical_names){
    #p<- ggplot(plot.obj$data, aes_string(plot.obj$variable1))+ geom_histogra
    m(color="darkblue", fill="lightblue") #geom_histogram(aes(y=..density..), colou
    r="black", fill="white") #+ geom_density(alpha=.2, fill="#FF6666")
    p<- ggplot(plot.obj$data, aes_string(plot.obj$variable1))+ geom_histogram
    (aes(y=..density..), color="darkblue", fill="lightblue") + geom_density(alpha=.
    2, fill="#FF6666")

    p<-p+labs(
      x      = input$variable1,
      title = paste("EDA - Variable - ", plot.obj$variable1)
    )
    p + .theme # agrega tema personalizado de ggplot
  }
  else{
    p<- ggplot(plot.obj$data, aes_string(plot.obj$variable1, fill=plot.obj$va
    riable1))+ geom_bar()
    p<- p + theme(axis.text.x = element_text(angle = 60, hjust = 1))
    p<-p+labs(
      # #fill    = as.factor(input$variable1),
      x      = input$variable1,
      title = paste("EDA - ", plot.obj$variable1)
    )
    p + .theme # agrega tema personalizado de ggplot
  }
})

#####----- Bivariado -----#####
output$p <- renderPlot({
  plot.obj<-get_data()

```

```

    # revisamos que se hayan agregado los datos especificados para hacer las gr
    aficas, sino paramos
    if(is.null(plot.obj)) return()

    # verificamos que el usuario ha establecido las variables a cruzar
    if(plot.obj$variable1 == "" | plot.obj$variable2 == "") return()

    # Variable para indicarle a ggplot el tipo de grafico que ha seleccionado el
    usuario
    plot.type<-switch(input$plot.type,
                      "boxplot"      = geom_boxplot(),
                      "histogram"    = geom_histogram(alpha=0.65,position="identit
y"),
                      "density"      = geom_density(alpha=.75),
                      "bar"          = geom_bar(position="dodge"),
                      "scatterplot"  = geom_point()
    )

    # Boxplots
    # Nota: si el usuario marca la casilla de añadir jitter se muestran los pun
    tos con ruido para evitar overplotting
    if(input$plot.type=="boxplot") {
      p<-ggplot(plot.obj$data,
                aes_string(
                  x      = plot.obj$variable2,
                  y      = plot.obj$variable1,
                  fill   = plot.obj$variable2 # llenamos grafica con variable 2
                )
      ) + plot.type

      if(input$show.points==TRUE) # agregamos ruido si el usuario marco la casi
      lla correspondiente
      {
        p<-p+ geom_point(color='black',alpha=0.35, position = 'jitter')
      }
    } else {
      # Scatterplots
      if(input$plot.type=="scatterplot") {
        p<-ggplot(plot.obj$data,
                  aes_string(
                    x      = plot.obj$variable2,
                    y      = plot.obj$variable1,
                    fill   = plot.obj$variable2
                  )
        ) + plot.type + geom_smooth(method = "lm") # agrega linea de tendencia

```

con suavizado

```

    }else{
      # Sintaxis para generar el resto de plots
      p<-ggplot(plot.obj$data,
               aes_string(
                 x      = plot.obj$variable1,
                 fill    = plot.obj$variable2,
                 group   = plot.obj$variable2
               )
      ) + plot.type
    }
  }

  # Se agregan descripciones de los ejes y codigo de colores
  p<-p+labs(
    fill = input$variable2,
    x     = input$variable2,
    y     = input$variable1,
    title = paste("EDA - ", plot.obj$variable1, " vs ", plot.obj$variable2)
  ) +
    .theme # agrega tema personalizado de ggplot
  print(p)
})

#####----- Multivariado ----- #####
output$q <- renderPlot({
  plot.obj<-get_data()

  # revisamos que se hayan agregado los datos especificados para hacer las graficas, sino paramos
  if(is.null(plot.obj)) return()

  # verificamos que el usuario ha establecido las variables a cruzar
  if(plot.obj$variable1 == "" | plot.obj$variable2 == "" | plot.obj$variable3 == "") return()

  # Variable para indicarle a ggplot el tipo de grafico que ha seleccionado el usuario
  plot.type<-switch(input$plot.type,
    "boxplot"      = geom_boxplot(),
    "histogram"    = geom_histogram(alpha=0.45,position="identity"),
    "density"      = geom_density(alpha=.75),
    "bar"          = geom_bar(position="dodge"),
    "scatterplot"  = geom_point()
  )

```

```

)

# Boxplots
# Nota: si el usuario marca la casilla de añadir jitter se muestran los pun
tos con ruido para evitar overplotting
if(input$plot.type=="boxplot") {
  p<-ggplot(plot.obj$data,
    aes_string(
      x      = plot.obj$variable2,
      y      = plot.obj$variable1,
      fill   = plot.obj$variable2 # Llenamos con la segunda variable
    )
  ) + plot.type + facet_wrap(~plot.obj$variable3) # se hace un face_wrat co
n la variable 3 seleccionada

  if(input$show.points==TRUE)
  {
    p<-p+ geom_point(color='black',alpha=0.5, position = 'jitter')      # No
ta: si el usuario marca la casilla de añadir jitter se muestran los puntos con
ruido para evitar overplotting

  }

} else {
  # Scatterplot
  if(input$plot.type=="scatterplot") {
    p<-ggplot(plot.obj$data,
      aes_string(
        x      = plot.obj$variable2,
        y      = plot.obj$variable1,
        fill   = plot.obj$variable2 # llenamos con la segunda vari
ables
      )
    ) + plot.type + geom_smooth(method = "lm") + facet_wrap( ~ get(plot.ob
j$variable3), ncol=3) # se hace un face_wrat con la variable 3 seleccionada

  }else{

    p<-ggplot(plot.obj$data,
      aes_string(
        x      = plot.obj$variable1,
        fill   = plot.obj$variable2,
        group  = plot.obj$variable2
      )
    ) + plot.type + facet_wrap( ~ get(plot.obj$variable3), ncol=3) # se hac
e un face_wrat con la variable 3 seleccionada

```

```
    }  
  }  
  
  # Se agregan descripciones de los ejes y código de colores  
  p<-p+labs(  
    fill = input$variable2,  
    x     = input$variable2,  
    y     = input$variable1,  
    title = paste("EDA - ", plot.obj$variable1, " vs ", plot.obj$variable2, "  
desagrupadas por ", plot.obj$variable3)  
  )  
  print(p)  
})  
  
}))  
  
# Corremos la aplicación con la interfaz de usuario y el servidor descrito previamente  
shinyApp(ui = ui, server = server)
```