

Tarea 11

Guillermo Segura Gómez

Problema 3

Considera el siguiente modelo:

$$Y = 1 + \beta \log(x) + \epsilon, \epsilon \sim \mathcal{N}(0, 0.1 * x), \quad x > 1$$

Calcula a mano el estimador de Máxima Verosimilitud de β usando la muestra $\{(y_i, x_i)\}$. ¿Qué distribución tiene? (resolver hasta el martes) Supongamos que $n = 30$ y $x_i = i, 2 \leq i \leq n + 1$; genera diferentes muestras de este modelo; dibuja el modelo estimado; visualiza la distribución empírica de $\hat{Y}(10)$. Compáralo con lo que da el método 1-KNN. ¿Qué observas?

Solución

La verosimilitud se define a partir de la distribución de los errores ϵ , que en este caso es una distribución normal con media cero y varianza dependiente de x . Dado que $\epsilon = Y - 1 - \beta \log(x)$, la verosimilitud para una observación individual es:

$$L_i(\beta) = \frac{1}{\sqrt{2\pi \cdot 0.1x_i}} \exp\left(-\frac{(y_i - 1 - \beta \log(x_i))^2}{2 \cdot 0.1x_i}\right)$$

Encontramos la la log-verosimilitud para facilitar el cálculo. Es la suma de los logaritmos de las verosimilitudes individuales, se expresa como:

$$\ell(\beta) = -\sum_i \left(\frac{(y_i - 1 - \beta \log(x_i))^2}{2 \cdot 0.1x_i} + \frac{1}{2} \log(2\pi \cdot 0.1x_i) \right)$$

Necesitamos encontrar la derivada de la función de verosimilitud. Al derivar la log-verosimilitud respecto a β y establecerla igual a cero para encontrar el máximo, obtenemos:

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_i \frac{(y_i - 1 - \beta \log(x_i)) \log(x_i)}{0.1x_i}$$

Ahora solucionamos para $\hat{\beta}$. Despejando β de la ecuación anterior, se llega a la fórmula del estimador de Máxima Verosimilitud:

$$\hat{\beta} = \frac{\sum_i \frac{(y_i - 1) \log(x_i)}{x_i}}{\sum_i \frac{\log^2(x_i)}{x_i}}$$

Sustituyendo $y_i = 1 + \beta \log(x_i) + \epsilon$ en la fórmula de $\hat{\beta}$, se obtiene:

$$\hat{\beta} = \beta + \epsilon \frac{\sum_i \frac{\log(x_i)}{x_i}}{\sum_i \frac{\log^2(x_i)}{x_i}}$$

De aquí, se concluye que $\hat{\beta}$ sigue una distribución normal con media β y varianza dependiente de x_j y de los valores de x_i en la muestra:

$$\hat{\beta} \sim \mathcal{N} \left(\beta, 0.1 x_j \frac{\sum_i \frac{\log(x_i)}{x_i}}{\sum_i \frac{\log^2(x_i)}{x_i}} \right)$$

```
# Cargar las librerías necesarias
library(ggplot2)
library(class) # para KNN
library(FNN)

##
## Attaching package: 'FNN'

## The following objects are masked from 'package:class':
##
##      knn, knn.cv

# Definir el número de muestras y el tamaño de cada muestra
n_muestras <- 1000
n <- 30

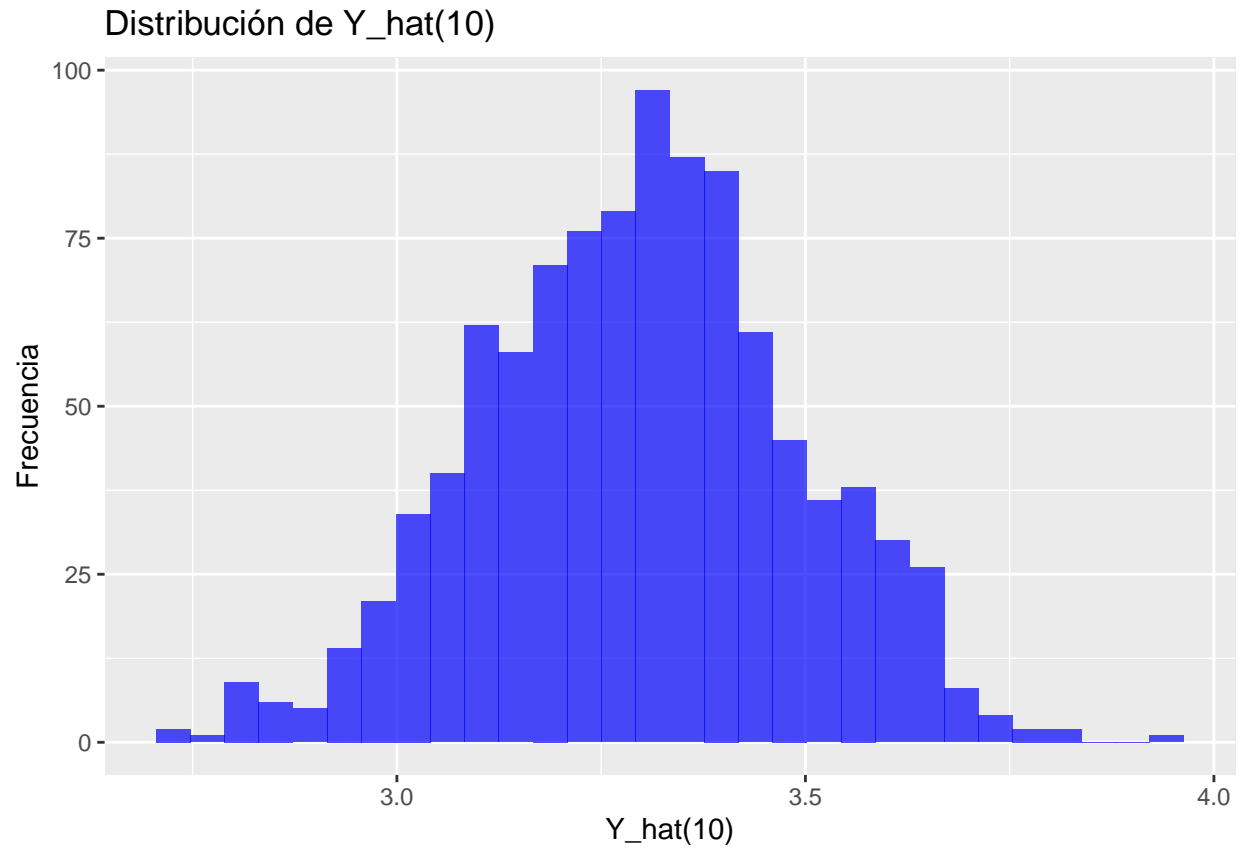
# Inicializar un vector para almacenar los valores de Y_hat(10)
y_hat_10 <- numeric(n_muestras)

# Bucle para generar muestras y calcular Y_hat(10)
set.seed(123) # Para reproducibilidad
for (i in 1:n_muestras) {
  x <- 2:(n + 1) # x_i
  beta_real <- 1 # Valor real de beta
  epsilon <- rnorm(n, mean = 0, sd = sqrt(0.1 * x))
  y <- 1 + beta_real * log(x) + epsilon

  # Estimar beta
  beta_estimado <- sum((y - 1) * log(x) / x) / sum(log(x)^2 / x)

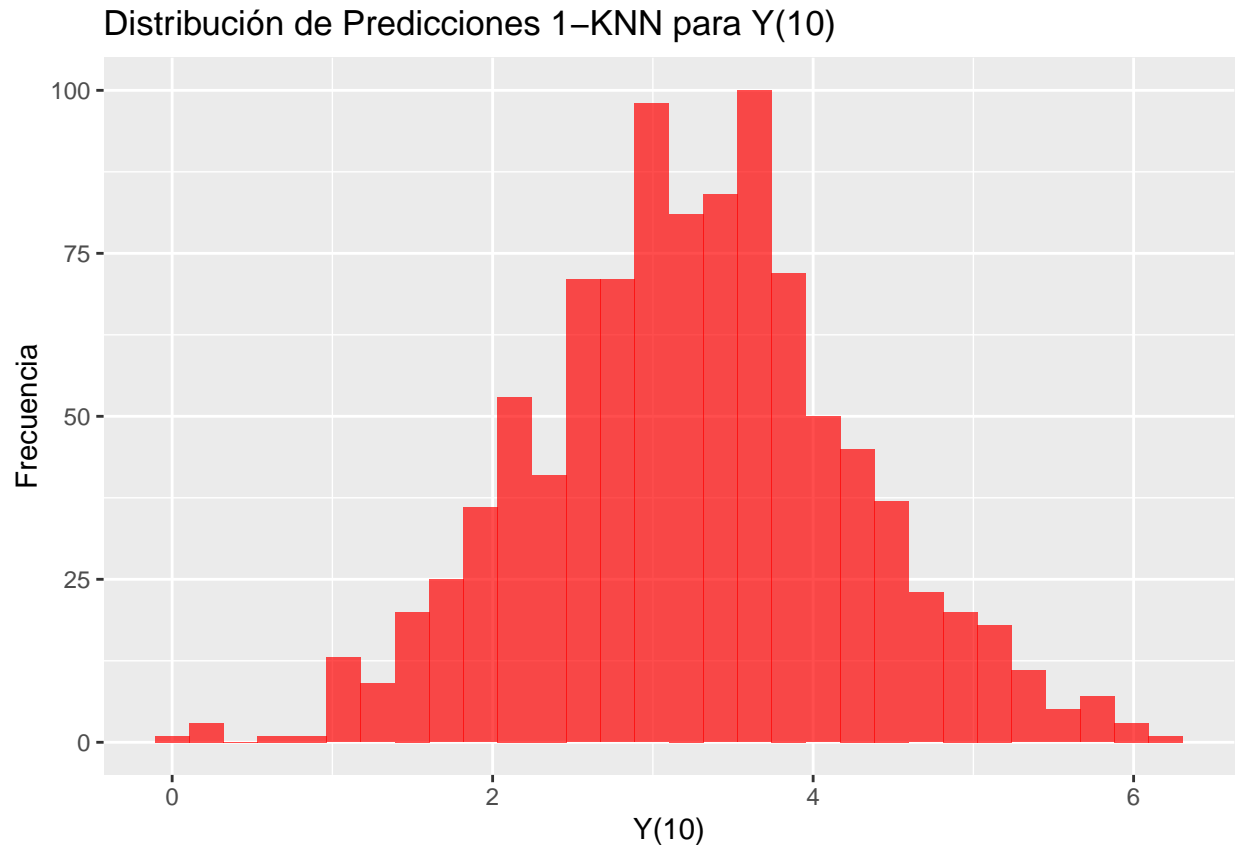
  # Calcular Y_hat(10)
  y_hat_10[i] <- 1 + beta_estimado * log(10)
}

# Visualizar la distribución de Y_hat(10)
ggplot(data.frame(y_hat_10), aes(x = y_hat_10)) +
  geom_histogram(bins = 30, fill = "blue", alpha = 0.7) +
  labs(title = "Distribución de Y_hat(10)", x = "Y_hat(10)", y = "Frecuencia")
```



```
# Comparación con 1-KNN
knn_pred <- numeric(n_muestras)
for (i in 1:n_muestras) {
  x <- 2:(n + 1)
  y <- 1 + beta_real * log(x) + rnorm(n, mean = 0, sd = sqrt(0.1 * x))
  knn_pred[i] <- knn.reg(train = data.frame(x), test = data.frame(x = 10), y = y, k = 1)$pred
}

# Visualizar la distribución de las predicciones de 1-KNN
ggplot(data.frame(knn_pred), aes(x = knn_pred)) +
  geom_histogram(bins = 30, fill = "red", alpha = 0.7) +
  labs(title = "Distribución de Predicciones 1-KNN para Y(10)", x = "Y(10)", y = "Frecuencia")
```



Ambos histogramas siguen una distribución normal, esto indica que, en promedio, ambos métodos están convergiendo hacia una predicción similar para $\hat{Y}(10)$.

Problema 4

Considera los datos en `datos.csv` relacionados con elecciones pasadas en EE.UU. Contiene para muchas elecciones, el porcentaje de votos que el partido del presidente saliente obtuvo (Y) y el porcentaje de crecimiento económico medido en ingreso personal en el año previo a las elecciones (X). Busca y discute un modelo de regresión lineal adecuado para Y en función de X . Haz una predicción del porcentaje de votos para los demócratas en las elecciones entre Hillary y Trump, si sabes que la economía había crecido en 2015 2 %.

Solución

Vamos a utilizar los datos de las elecciones pasadas para entrenar un modelo que sea capaz de hacer una predicción. La regresión que se va a calcular, busca encontrar una relación entre el crecimiento económico y el porcentaje de votos del partido saliente.

Primero necesitamos importar los datos.

```
datos <- read.csv("datos(1).csv", sep = " ")
datos
```

```
##   year growth  vote inc_party_candidate other_candidate
## 1  1952   2.40 44.60          Stevenson      Eisenhower
## 2  1956   2.89 57.76          Eisenhower      Stevenson
```

## 3	1960	0.85	49.91	Nixon	Kennedy
## 4	1964	4.21	61.34	Johnson	Goldwater
## 5	1968	3.02	49.60	Humphrey	Nixon
## 6	1972	3.62	61.79	Nixon	McGovern
## 7	1976	1.08	48.95	Ford	Carter
## 8	1980	-0.39	44.70	Carter	Reagan
## 9	1984	3.86	59.17	Reagan	Mondale
## 10	1988	2.27	53.94	Bush, Sr.	Dukakis
## 11	1992	0.38	46.55	Bush, Sr.	Clinton
## 12	1996	1.04	54.74	Clinton	Dole
## 13	2000	2.36	50.27	Gore	Bush, Jr.
## 14	2004	1.72	51.24	Bush, Jr.	Kerry
## 15	2008	0.10	46.32	McCain	Obama
## 16	2012	0.95	52.00	Obama	Romney

Ahora vamos a generar un modelo de regresión lineal para los datos.

```
# Ajustar el modelo de regresión lineal
modelo <- lm(vote ~ growth, data = datos)

# Resumen del modelo para ver los coeficientes
summary(modelo)
```

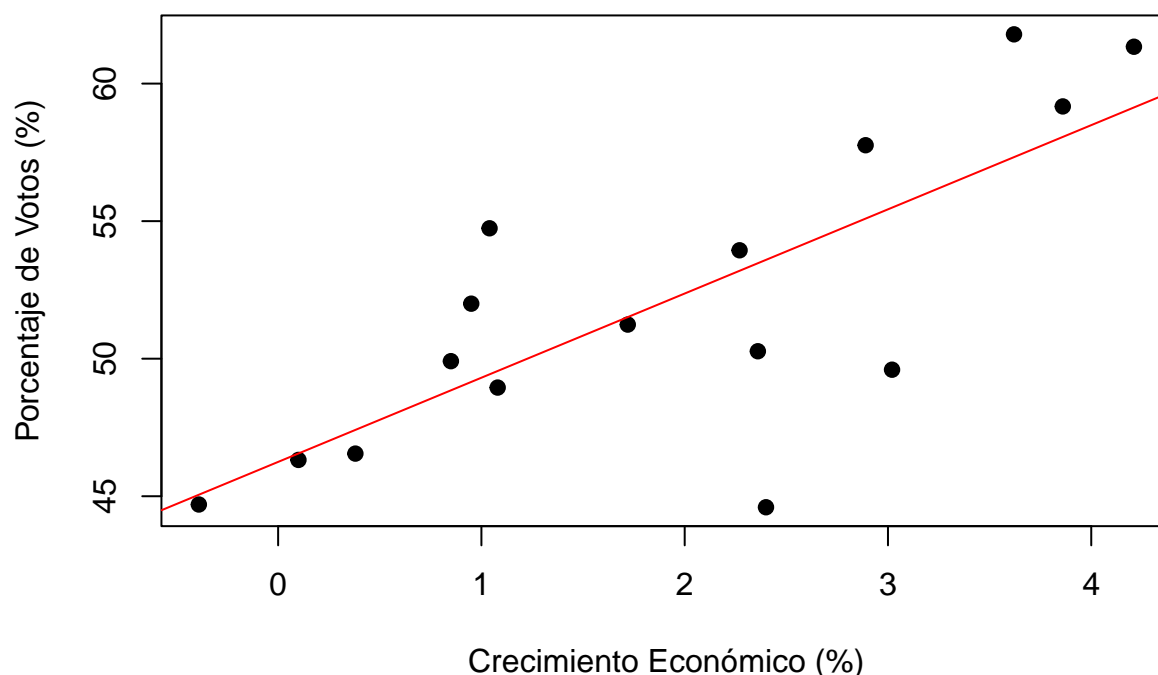
```
##
## Call:
## lm(formula = vote ~ growth, data = datos)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.9929 -0.6674  0.2556  2.3225  5.3094
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  46.2476     1.6219  28.514 8.41e-14 ***
## growth       3.0605     0.6963   4.396 0.00061 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.763 on 14 degrees of freedom
## Multiple R-squared:  0.5798, Adjusted R-squared:  0.5498
## F-statistic: 19.32 on 1 and 14 DF,  p-value: 0.00061
```

Antes de realizar el análisis vamos a graficar el dataset para ver el comportamiento de los datos.

```
# Gráfico de dispersión
plot(datos$growth, datos$vote, main = "Crecimiento Económico vs Porcentaje de Votos",
      xlab = "Crecimiento Económico (%)", ylab = "Porcentaje de Votos (%)", pch = 19)

# Añadir la línea de regresión al gráfico
abline(modelo, col = "red")
```

Crecimiento Económico vs Porcentaje de Votos



Es evidente una tendencia positiva del porcentaje de votos en función del crecimiento económico.

El modelo de regresión lineal se puede expresar como

$$Y = \beta_0 + \beta_1 X + \epsilon$$

donde Y es el porcentaje de votos, X es el crecimiento económico. Los coeficientes calculados por el modelo son

$$\beta_0 = 46.2476$$

$$\beta_1 = 3.0605$$

β_0 también llamado intercepto, es el valor de Y cuando X es 0. Esto significa que si el crecimiento económico fuera 0%, el modelo predice que el porcentaje de votos para el partido del presidente saliente sería aproximadamente 46.25%.

β_1 es la pendiente y es el cambio en Y por cada unidad de cambio en X . Esto indica que por cada 1% de aumento en el crecimiento económico, el modelo predice un aumento de aproximadamente 3.06% en el porcentaje de votos para el partido del presidente saliente.

Para hacer una predicción, utilizamos el modelo.

```
# Hacer una predicción
prediccion <- predict(modelo, newdata = data.frame(growth = 2))

# Imprimir la predicción
print(prediccion)
```

```
##          1
## 52.3687
```

El resultado obtenido para la predicción es de 52.3687. Esto se puede interpretar como lo siguiente: con un crecimiento económico del 2%, el partido del presidente saliente recibiría aproximadamente un 52.37% de los votos. Esta predicción asume que la relación entre el crecimiento económico y el porcentaje de votos observada en el pasado se mantiene constante, y de forma lineal.

Problema 5

Consideramos el conjunto de datos `datos2.csv`, el cual surge en un estudio sobre factores que pueden influir en el peso de recién nacidos. Este conjunto contiene un total de 189 observaciones y presenta las siguientes variables:

1. **age**: Edad de la madre.
2. **lwt**: Peso de la madre en el último ciclo de menstruación.
3. **race**: Etnicidad de la madre (1=blanca, 2=negra, 3=otra).
4. **smoke**: Indicador de si la madre fumaba durante el embarazo.
5. **ptd**: Indicador de si la madre ha tenido bebés prematuros antes o no.
6. **ht**: Indicador de si la madre ha tenido hipertensión o no.
7. **ui**: Nivel de irritabilidad de la uterina.
8. **ftv**: Número de visitas al médico durante los primeros tres meses del embarazo.
9. **bwt**: Peso del recién nacido (=Y).

El objetivo es buscar visualizaciones informativas de los datos y ajustar y discutir modelos de regresión lineal para predecir el peso del recién nacido en función de las demás variables.

Es importante destacar que la variable **race** no tiene una escala métrica, sino que es nominal. Los valores son meramente etiquetas, por lo que no tiene sentido calcular “race x race”. La solución consiste en introducir dos variables binarias en lugar de la variable original. Una variable binaria que indica si la raza es blanca o no, y otra si es negra o no. Esto se puede lograr automáticamente en R convirtiendo la variable en un tipo factor: `factor(race)`. En Python, se puede realizar utilizando la función `C(race)` con la “C” de categoría.

Solución

Para realizar la correcta exploración y análisis, primero se deben importar los datos.

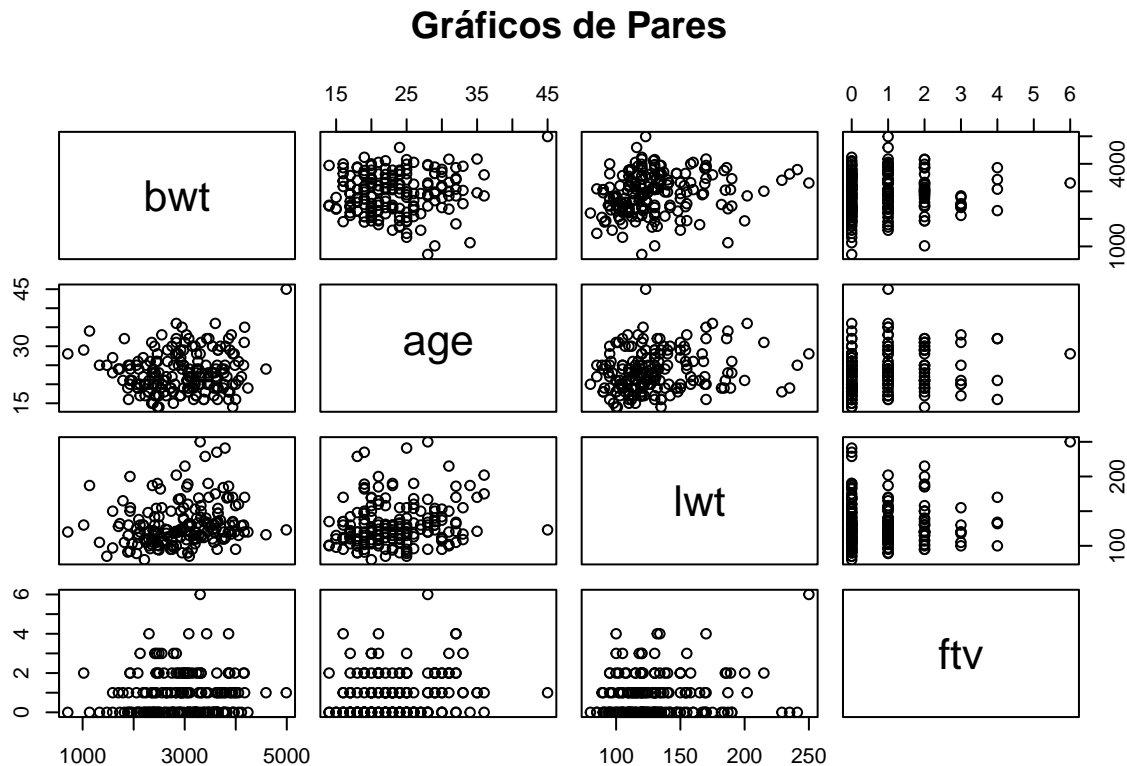
```
datos2 <- read.csv("datos2(1).csv")
```

Ahora, debido a la consideración de que el campo es race es nominal y no métrica, es necesario hacer la conversión de la variable en tres categóricas diferentes.

```
# Convertir 'race' en factor y luego en variables dummy
datos2$race <- factor(datos2$race)
datos2 <- cbind(datos2, model.matrix(~race - 1, data = datos2))
```

Visualizaciones Vamos a realizar una exploración visual de los datos. Al ser una gran cantidad de variables una buena alternativa es utilizar gráficos que engloben varias variables. Para coparar las variables cuantitativas se utiliza un gráfico de pares.

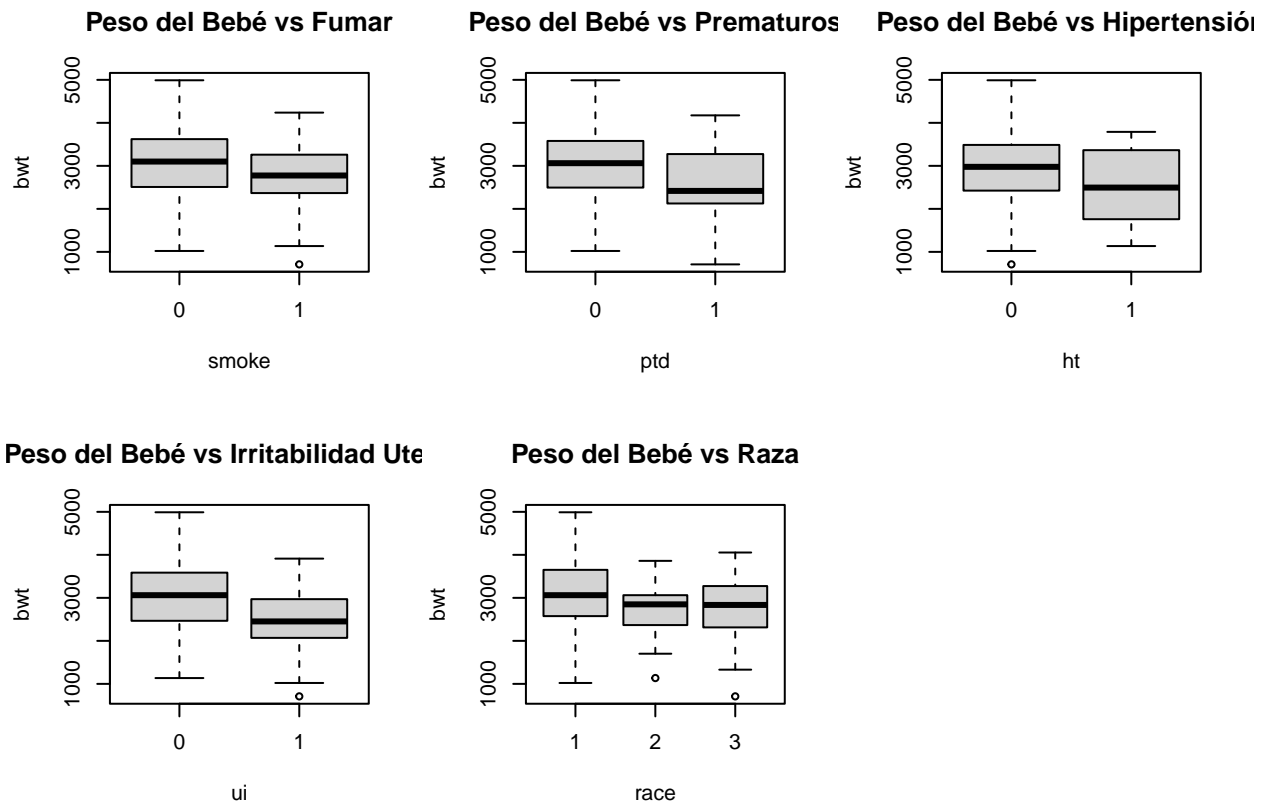
```
# Gráfico de pares
pairs(~ bwt + age + lwt + ftv, data = datos2, main = "Gráficos de Pares")
```



En este gráfico la primera columna representa la variable **bwt** que es el peso del bebé, comparada en dispersión con las variables de edad, peso de la madre en el último ciclo y visitas al médico. Es claro que ninguna de estas variables está correlacionada con el peso del bebé, las gráficas no muestran correlación de ningún tipo según se puede observar a simple vista.

Para las variables cualitativas, podemos realizar diagramas de caja para compararlas.

```
# Diagramas de caja
par(mfrow = c(2, 3)) # Ajustar el layout para múltiples gráficos
boxplot(bwt ~ smoke, data = datos2, main = "Peso del Bebé vs Fumar")
boxplot(bwt ~ ptd, data = datos2, main = "Peso del Bebé vs Prematuros")
boxplot(bwt ~ ht, data = datos2, main = "Peso del Bebé vs Hipertensión")
boxplot(bwt ~ ui, data = datos2, main = "Peso del Bebé vs Irritabilidad Uterina")
boxplot(bwt ~ race, data = datos2, main = "Peso del Bebé vs Raza")
```

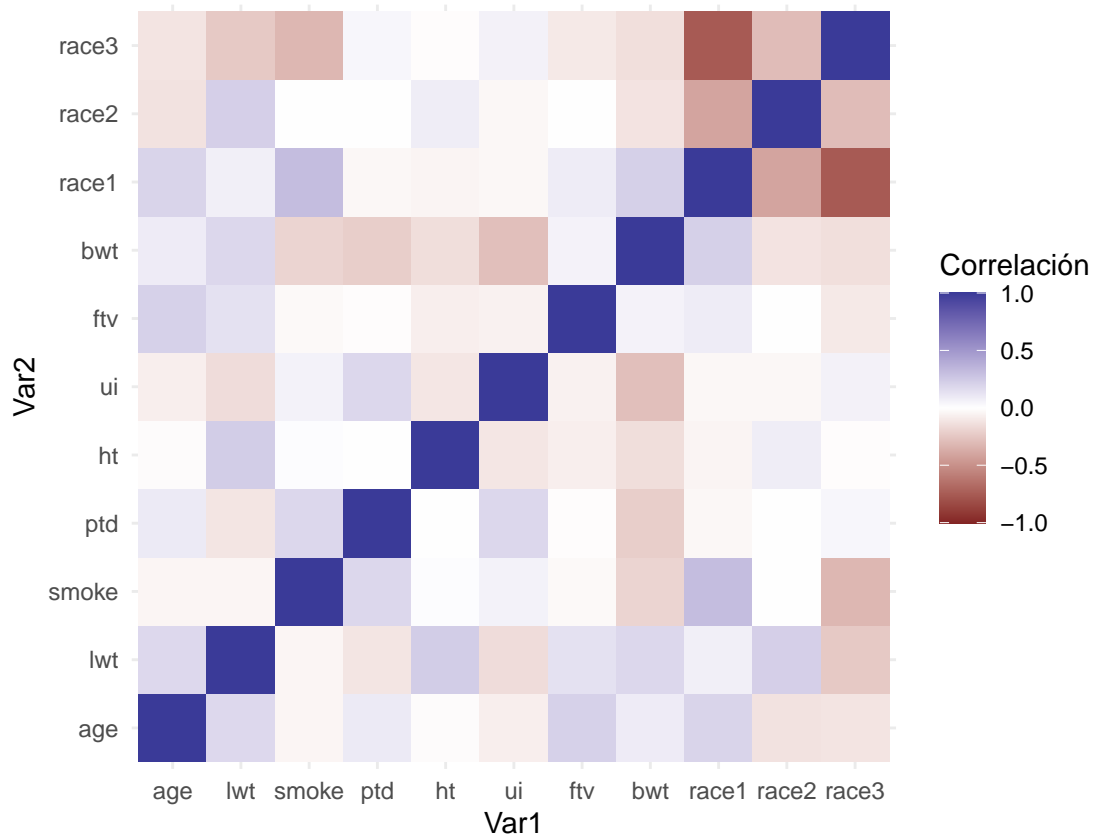



Para estas variables, si podemos realizar algunas conclusiones a simple vista. Las variables en las que existe una diferencia evidente en cuanto al peso del bebe son la irritabilidad uterina, la hipertensión, bebes prematuros y en menor medida fumar. En donde no es posible apreciar una diferencia significativa es en la raza del bebe, de hecho las medias parecen estar casi en el mismo sitio.

Una mejor práctica puede ser la visualización por mapa de calor. Realizamos un cálculo de correlación para determinar la relación existente entre las variables.

```
# Calcular la matriz de correlación
cor_matriz <- cor(datos2[, sapply(datos2, is.numeric)])

# Visualizar la matriz de correlación como un mapa de calor
library(reshape2)
ggplot(data = melt(cor_matriz), aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(midpoint = 0, limit = c(-1, 1), space = "Lab",
    name="Correlación") +
  theme_minimal() +
  coord_fixed()
```



En el mapa anterior, si comparamos todas las variables con el peso del bebe. La correlación existente en su mayoría es muy *tenue*, es decir podemos categorizar la correlación menor a ± 0.5 de todas las variables con el peso del bebe. La variable mas correlacionada con el peso del bebe es la irritabilidad uterina.

Modelo de regresión lineal Ahora vamos a ejecutar un modelo de regresión lineal con las variables mas correlacionadas. Tomemos la irritación uterina, y el fumar de la madre. Vamos a preguntarnos, que tan bien podemos predecir el peso en función de si la madre fuma o tiene irritación uterina.

```
# Ajuste del Modelo de Regresión Lineal
modelo1 <- lm(bwt ~ ui + smoke, data = datos2)
summary(modelo1)

##
## Call:
## lm(formula = bwt ~ ui + smoke, data = datos2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1798.63  -460.16   46.37   500.37  1861.37
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3128.63     67.08  46.644 < 2e-16 ***
## ui            -559.20    141.84  -3.942 0.000114 ***
## smoke         -258.48    103.24  -2.504 0.013150 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 691.4 on 186 degrees of freedom
## Multiple R-squared:  0.1106, Adjusted R-squared:  0.101
## F-statistic: 11.56 on 2 and 186 DF,  p-value: 1.847e-05
```

El modelo de regresión lineal ajustado para predecir el peso del recién nacido (*bwt*) en función de la irritabilidad uterina (*ui*) y si la madre fuma (*smoke*) es el siguiente:

$$bwt = 3128.63 - 559.20 \times ui - 258.48 \times smoke$$

Podemos interpretar los coeficientes de la siguiente manera: - **Intercepto (3128.63)**: Este valor representa el peso promedio estimado de un recién nacido cuando tanto *ui* como *smoke* son 0. Es decir, para madres sin irritabilidad uterina y que no fuman. - **Coefficiente de ui (-559.20)**: Por cada unidad que aumenta la irritabilidad uterina, el peso promedio del recién nacido disminuye en 559.20 gramos, manteniendo constante el efecto de fumar. - **Coefficiente de smoke (-258.48)**: Si la madre fuma, el peso promedio del recién nacido disminuye en 258.48 gramos, manteniendo constante la irritabilidad uterina.

Podemos ver en los resultados del modelo que tan significativo es el mismo:

- El p-valor asociado con el coeficiente de *ui* es 0.000114, lo que indica que la irritabilidad uterina es un predictor estadísticamente significativo del peso del recién nacido.
- El p-valor para *smoke* es 0.013150, lo que también indica una relación estadísticamente significativa entre fumar durante el embarazo y el peso del recién nacido.

La calidad del modelo esta determinada por el R ajustado y el error estandar residual.

- El **R-cuadrado ajustado** es 0.101, lo que sugiere que aproximadamente el 10.1% de la variabilidad en el peso del recién nacido es explicada por este modelo. Aunque esto indica alguna relación, hay una gran cantidad de variabilidad en el peso del recién nacido que no es explicada por estos dos factores.
- El **error estándar residual** es 691.4, indicando la variabilidad típica de las predicciones del modelo respecto a los valores reales.

La irritabilidad uterina y el hábito de fumar en la madre tienen un impacto negativo significativo en el peso del recién nacido. Aunque el modelo proporciona cierta información útil, hay otros factores no incluidos en el modelo que también afectan el peso del recién nacido, lo que se refleja en el R-cuadrado ajustado relativamente bajo.

Podemos considerar un segundo modelo que incluya las otras variables mas correlacionadas con el peso del bebe. Tomemos la irritación uterina, el fumar de la madre, la hipertensión y bebes prematuros.

```
# Ajuste del Modelo de Regresión Lineal
modelo2 <- lm(bwt ~ ui + smoke + ptd + ht, data = datos2)
summary(modelo2)
```

```
##
## Call:
## lm(formula = bwt ~ ui + smoke + ptd + ht, data = datos2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1618.92  -465.58   15.42   463.42  1802.42
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3187.58      67.99  46.881 < 2e-16 ***
## ui          -547.66     141.77  -3.863 0.000155 ***
## smoke       -216.72     102.54  -2.113 0.035910 *
## ptd         -279.59     139.14  -2.009 0.045956 *
## ht          -513.85     202.88  -2.533 0.012153 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 675.8 on 184 degrees of freedom
## Multiple R-squared:  0.1594, Adjusted R-squared:  0.1412
## F-statistic: 8.726 on 4 and 184 DF, p-value: 1.799e-06
```

Discusión y Conclusiones del Modelo de Regresión Lineal Mejorado

Modelo Utilizado El modelo de regresión lineal ajustado para predecir el peso del recién nacido (*bwt*) en función de la irritabilidad uterina (*ui*), si la madre fuma (*smoke*), si ha tenido bebés prematuros antes (*ptd*) y si ha tenido hipertensión (*ht*) es el siguiente:

$$bwt = 3187.58 - 547.66 \times ui - 216.72 \times smoke - 279.59 \times ptd - 513.85 \times ht$$

- **Intercepto (3187.58):** Este valor representa el peso promedio estimado de un recién nacido cuando todas las variables predictoras son 0.
- **Coefficiente de ui (-547.66):** Por cada unidad que aumenta la irritabilidad uterina, el peso promedio del recién nacido disminuye en 547.66 gramos.
- **Coefficiente de smoke (-216.72):** Si la madre fuma, el peso promedio del recién nacido disminuye en 216.72 gramos.
- **Coefficiente de ptd (-279.59):** Si la madre ha tenido bebés prematuros antes, el peso promedio del recién nacido disminuye en 279.59 gramos.
- **Coefficiente de ht (-513.85):** Si la madre ha tenido hipertensión, el peso promedio del recién nacido disminuye en 513.85 gramos.
- Todos los coeficientes son estadísticamente significativos, como lo indican sus p-valores. Esto sugiere que cada uno de estos factores tiene un impacto significativo en el peso del recién nacido.
- El **R-cuadrado ajustado** es 0.1412, lo que indica que aproximadamente el 14.12% de la variabilidad en el peso del recién nacido es explicada por este modelo. Esto es una mejora respecto al modelo anterior, pero aún hay una gran cantidad de variabilidad no explicada.
- El **error estándar residual** es 675.8, que es ligeramente menor que en el modelo anterior, indicando una mejor precisión en las predicciones.

Podemos concluir lo siguiente: La irritabilidad uterina, el hábito de fumar, tener antecedentes de bebés prematuros y la hipertensión en la madre son factores significativos que afectan negativamente el peso del recién nacido. Aunque el modelo mejorado explica una mayor proporción de la variabilidad en el peso del recién nacido en comparación con el modelo anterior, aún hay otros factores no considerados que pueden influir en el peso al nacer. Este modelo puede ser útil para identificar y cuantificar los riesgos asociados con el bajo peso al nacer, pero debe ser utilizado como parte de un enfoque más amplio que incluya otros factores y consideraciones clínicas.

Problema 6

Para una muestra de 14 muchachas elegidas al azar se tienen las mediciones del ancho de su cara a la edad de 5 años y 6 años. Haz una prueba de hipótesis para verificar si en promedio la diferencia entre las dos mediciones es 0.2 pulgadas $\alpha = 0.1$.

Solución

Para realizar una prueba de hipótesis para verificar si en promedio la diferencia entre las dos mediciones, los 5 años y los 6 años, es 0.2 pulgadas, podemos utilizar una prueba t de muestras emparejadas. En este caso, estamos interesados en probar la hipótesis nula de que la diferencia media es 0.2 contra la hipótesis alternativa de que no es 0.2.

Es decir

$$H_0 : \mu_d = 0.2 \quad H_a : \mu_d \neq 0.2$$

Donde μ_d representa la diferencia media entre las mediciones a los 5 años y a los 6 años. La hipótesis nula establece que la diferencia media es exactamente 0.2 pulgadas. La prueba t para muestras emparejadas que se propone realizar busca probar esta hipótesis.

Primero necesitamos cargar los datos

```
# Datos de las mediciones a los 5 y 6 años
mediciones_5 <- c(7.33, 7.49, 7.27, 7.93, 7.56, 7.81, 7.46, 6.94, 7.49, 7.44, 7.95, 7.04, 7.10, 7.64)
mediciones_6 <- c(7.53, 7.70, 7.46, 8.21, 7.81, 8.01, 7.72, 7.13, 7.68, 7.66, 8.11, 7.20, 7.25, 7.79)
```

Hay que calcular las diferencias entre las mediciones

```
# Calcular la diferencia entre las mediciones
diferencias <- mediciones_6 - mediciones_5
```

Ahora realizamos la prueba de hipótesis con un nivel de significancia $\alpha = 0.1$

```
# Realizar la prueba t para muestras emparejadas
t_test_result <- t.test(diferencias, mu = 0.2, alternative = "two.sided", conf.level = 0.9)

# Decidir si rechazar o no la hipótesis nula
t_test_result
```

```
##
## One Sample t-test
##
## data:  diferencias
## t = 0.06573, df = 13, p-value = 0.9486
## alternative hypothesis: true mean is not equal to 0.2
## 90 percent confidence interval:
##  0.1814695 0.2199591
## sample estimates:
## mean of x
## 0.2007143
```

El valor-p es significativamente alto (0.9486), lo cual está muy por encima del nivel de significancia elegido ($\alpha = 0.1$). Esto indica que no hay suficiente evidencia para rechazar la hipótesis nula.

El intervalo de confianza al 90% para la diferencia media está entre 0.1814695 y 0.2199591 pulgadas. Esto incluye el valor de 0.2 pulgadas, lo cual está en línea con no rechazar la hipótesis nula.

La media estimada de las diferencias es 0.2007143, lo cual está muy cerca del valor hipotético de 0.2 pulgadas.

Con base en estos resultados, se concluye que no hay evidencia estadística suficiente para afirmar que la diferencia promedio en el ancho de la cara entre los 5 y 6 años es diferente de 0.2 pulgadas. Esto sugiere que, en promedio, el ancho de la cara de las muchachas en la muestra aumentó en 0.2 pulgadas entre estas edades.