



Trabajo Práctico N°2

Análisis Univariado

Profesores: Victoria Santa María
Eugenia Camila Berrino

Alumnas: Iara Sofía RAVAGNI, 59492
Valentina GONZALEZ, 59128

Fecha de entrega: 25-09-2022

Índice

1. Primera Parte	4
1.1. Leer CSV	4
1.2. Clasificación de variables	4
1.3. Tipos de Datos	4
1.4. Datos faltantes	5
1.5. Análisis datos faltantes	6
1.6. Histograma	6
1.7. Boxplot	6
1.8. QQ-Plot	6
1.9. Análisis de normalidad	7
1.9.1. Grasas	7
1.9.2. Alcohol	8
1.9.3. Calorías	8
2. Segunda Parte	9
2.1. Conversión a factor	9
2.2. Ajuste Histograma	9
2.3. Ajuste Boxplot	10
2.4. Ajuste QQ-Plot	11
2.5. Test de Shapiro-Wilk	11
2.6. Test de Lilliefors	12
2.7. Test de Levene	12

2.8. Análisis de normalidad	13
2.9. T-test	14
3. Tercera Parte	15
3.1. Tabla de contingencia	15
3.2. Valores esperados y residuos	16
3.3. OR sin ajustar y OR de Mantel-Haenzel ajustado por año de nacimiento, de madres adolescentes vs. mayores para muerte	17
3.4. Único OR entre los estratos	21
3.5. OR sin ajustar y OR de Mantel-Haenzel ajustado por cuidado perinatal, de madres adolescentes vs. mayores para muerte	21

1. Primera Parte

1.1. Leer CSV

Como respuesta a este ejercicio se generó el siguiente código:

```
1 read_file = function(datapath,header,sep){  
2  
3   file <- read.csv(datapath, sep=sep, header=header)  
4   return(file)  
5  
6 }
```

1.2. Clasificación de variables

Las variables que componen a la base de datos son las siguientes: *Grasas*, *Alcohol*, *Calorías* y *Dieta*. Cada una de ella se clasifica según lo indicado en el Cuadro 1.

Variable	Clasificación	Subclasificación
Grasas	Cuantitativa	Continua
Alcohol	Cuantitativa	Continua
Calorías	Cuantitativa	Discreta
Dieta	Cualitativa	Nominal

Cuadro 1: Clasificación de variables.

1.3. Tipos de Datos

Inicialmente, como respuesta a este ejercicio se generó el siguiente código:

```
1 data_type = function(df){  
2  
3   df_type <- data.frame(Variable = colnames(df),  
4                         Tipo = sapply(df, class))  
5  
6   rownames(df_type)<-NULL
```

```

7 |
8 |   print(df_type)
9 |
10| }

```

Luego de correr el código en la aplicación se generó la tabla que se observa en la Figura 1.

	Variable	Tipo
1	Grasas	numeric
2	Alcohol	numeric
3	calorias	integer
4	Dieta	integer

Figura 1: Tabla con tipos de datos.

En la imagen se puede apreciar que para todas las variables coincide la clasificación teórica con la obtenida, a excepción de la variable *Dieta*. El programa la identifica como una variable cuantitativa discreta en lugar de cualitativa nominal. Para corregir este error, se debería convertir dicha variable a factor.

1.4. Datos faltantes

Como respuesta a este ejercicio se generó el siguiente código:

```

1 | na_counts = function(df){
2 |
3 |   df_NA <- data.frame(Variable = colnames(df),
4 |                       Cant_NA = colSums(is.na(df)))
5 |
6 |   rownames(df_NA)<-NULL
7 |
8 |   print(df_NA)
9 | }

```

1.5. Análisis datos faltantes

A partir del punto anterior, se puede observar que existe un único dato faltante que corresponde a la variable *calorías*. Esto se podría corregir reemplazándolo por la medida de tendencia central (media, mediana o moda; dependiendo de la distribución de la variable). Dicha modificación no invalida el estudio ya que el dato faltante representa tan solo un 0.9% de la muestra total.

1.6. Histograma

Como respuesta a este ejercicio se generó el siguiente código:

```
1 histogram = function(df,colname,inpbins){  
2  
3   ggplot(df, aes(df[,colname])) + geom_histogram(aes(y = ..density..),  
4     bins = inpbins) + xlab(colname) +  
5     stat_function(fun = dnorm, args = list(mean =  
6     mean(na.omit(df[,colname])), sd = sd(na.omit(df[,colname]))))  
}
```

1.7. Boxplot

Como respuesta a este ejercicio se generó el siguiente código:

```
1 my_boxplot = function(df,colname){  
2  
3   ggplot(df, aes(df[,colname])) + geom_boxplot() + coord_flip() +  
4     xlab(colname)  
5 }
```

1.8. QQ-Plot

Como respuesta a este ejercicio se generó el siguiente código:

```

1 my_qqplot = function(df,colname){
2
3   ggplot(df, mapping = aes(sample = df[,colname])) + stat_qq_band() +
4     stat_qq_line() + stat_qq_point() + labs(x = "Theoretical
5     Quantiles", y = "Sample Quantiles")
6 }

```

1.9. Análisis de normalidad

Con el fin de concluir si las distintas variables poseen distribución normal o no, se analizaron los histogramas, boxplots y qq-plots para cada caso.

1.9.1. Grasas

A partir de los gráficos de la Figura 2, se puede concluir que la variable *Grasas* no tiene una distribución normal. Si bien a primera vista los primeros dos gráficos parecieran ajustarse a este tipo de distribución, al analizar el qq-plot, este sugiere que existe una asimetría hacia la derecha.

Por lo tanto, en cuanto a la medida de tendencia central y dispersión, se utiliza la mediana y el rango intercuartil, respectivamente. El beneficio de elegir la mediana es la robustez frente a los *outliers*

Beneficio???

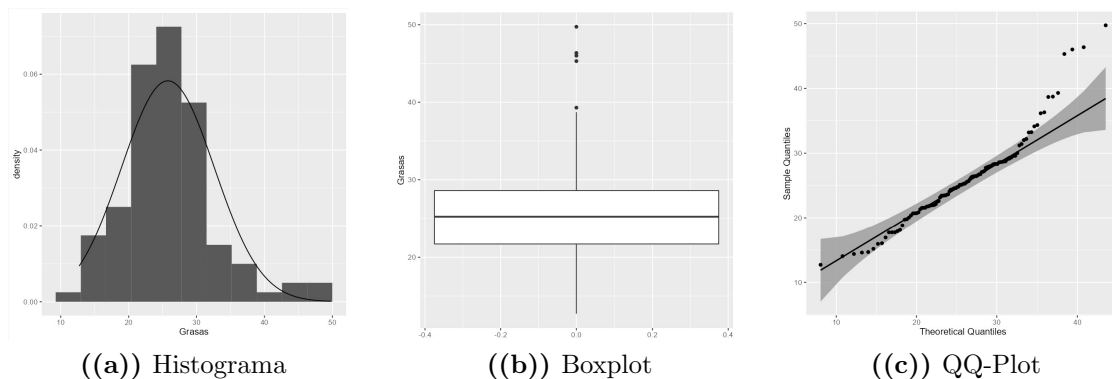


Figura 2: Gráficos para *Grasas*.

1.9.2. Alcohol

A partir de los gráficos de la Figura 3, se puede concluir que la variable *Alcohol* no tiene distribución normal. Se puede observar que tiene una asimetría derecha, tanto por el histograma desplazado hacia la izquierda, el boxplot desplazado hacia abajo y el poco ajuste de los puntos en los extremos del qq-plot.

Por lo tanto, en cuanto a la medida de tendencia central y dispersión, se utiliza la mediana y el rango intercuartil, respectivamente.

Beneficio???

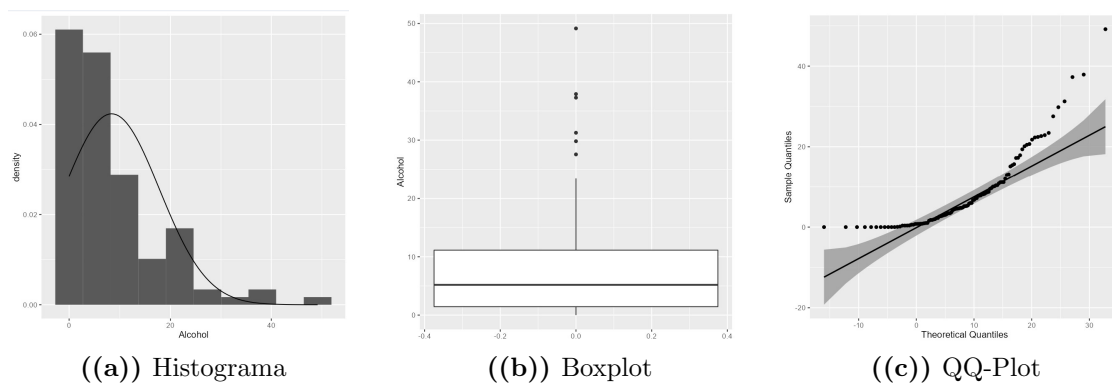


Figura 3: Gráficos para *Alcohol*.

1.9.3. Calorías

A partir de los gráficos de la Figura 4, se puede concluir que la variable *Calorías* tiene distribución normal ya que todos los gráficos se comportan como tal. Sin embargo, sería prudente verificarlo con un método analítico.

Por lo tanto, en cuanto a la medida de tendencia central y dispersión, se utiliza la media y el desvío estándar, respectivamente.

Beneficio???

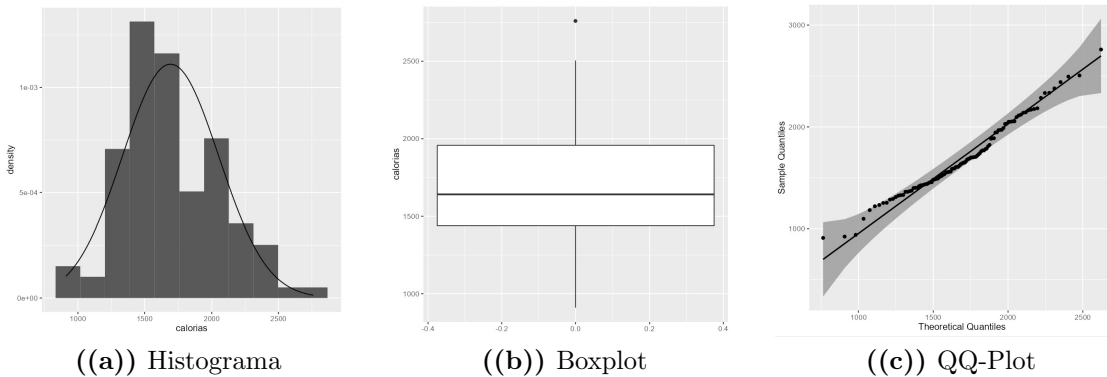


Figura 4: Gráficos para *Calorías*.

2. Segunda Parte

2.1. Conversión a factor

Como respuesta a este ejercicio se generó el siguiente código:

```

1 | convert_to_factor = function(df,colname){
2 |
3 |   df[,colname] <- as.factor(df[,colname])
4 |
5 |   return(df)
6 |
7 | }
```

2.2. Ajuste Histograma

Como respuesta a este ejercicio se generó el siguiente código:

```

1 | histogram = function(df,groupcol,colname,inpbins){
2 |   if (groupcol == 'No agrupar'){
3 |
4 |     ggplot(df, aes(df[,colname])) +
5 |       geom_histogram(aes(y = ..density..), bins = inpbins) +
```

```
6       xlab(colname) +
7       stat_function(fun = dnorm, args = list(mean =
mean(na.omit(df[,colname])), sd = sd(na.omit(df[,colname]))))
8
9   } else {
10
11     ggplot(df, aes(df[,colname], fill=df[,groupcol])) +
12     geom_histogram(aes(y = ..density..), alpha = 0.6, bins = inpbins)
13     +
14     xlab(colname)
15   }
16 }
17 }
```

2.3. Ajuste Boxplot

Como respuesta a este ejercicio se generó el siguiente código:

```
1 my_boxplot = function(df,colname,groupcol){
2
3   if (groupcol == 'No agrupar'){
4
5     ggplot(df, aes(df[,colname])) +
6     geom_boxplot() +
7     coord_flip() +
8     xlab(colname)
9
10  } else {
11
12    ggplot(df, aes(df[,colname], fill=df[,groupcol])) +
13    geom_boxplot(alpha = 0.6) +
14    xlab(colname)
15  }
16 }
```

2.4. Ajuste QQ-Plot

Como respuesta a este ejercicio se generó el siguiente código:

```
1 my_qqplot = function(df,colname,groupcol){
2
3   if (groupcol == 'No agrupar'){
4
5     ggplot(df, mapping = aes(sample = df[,colname])) +
6       stat_qq_band() + stat_qq_line() + stat_qq_point() +
7       labs(x = "Theoretical Quantiles", y = "Sample Quantiles")
8
9   } else {
10
11     ggplot(df, aes(sample=df[,colname], fill=df[,groupcol])) +
12       stat_qq_band(alpha = 0.5) + stat_qq_line() + stat_qq_point() +
13       xlab(colname)
14   }
15 }
```

2.5. Test de Shapiro-Wilk

Como respuesta a este ejercicio se generó el siguiente código:

```
1 shapiro_w = function(df,colname,groupcol){
2
3   if (groupcol == 'No agrupar'){
4
5     return(shapiro.test(df[,colname]))
6
7   } else {
8
9
10    data = split(df, df$Preparado)
11    dataG1 = data[[1]]
12    dataG2 = data[[2]]
13
14    shapiro1 = shapiro.test(as.numeric(dataG1$Betacaroteno))
15    shapiro2 = shapiro.test(as.numeric(dataG2$Betacaroteno))
16 }
```

```
17     return(list(shapiro1 = shapiro1, shapiro2 = shapiro2))
18   }
19 }
20 }
```

2.6. Test de Lilliefors

Como respuesta a este ejercicio se generó el siguiente código:

```
1  lilliefors = function(df,colname,groupcol){
2
3    if (groupcol == 'No agrupar'){
4
5      return(shapiro.test(df[,colname]))
6
7    } else {
8
9      data = split(df, df$Preparado)
10     dataG1 = data[[1]]
11     dataG2 = data[[2]]
12
13     lillie1 = lillie.test(as.numeric(dataG1$Betacaroteno))
14     lillie2 = lillie.test(as.numeric(dataG2$Betacaroteno))
15
16     return(list(lillie1 = lillie1, lillie2 = lillie2))
17   }
18 }
19 }
```

2.7. Test de Levene

Como respuesta a este ejercicio se generó el siguiente código:

```
1  levene = function(df, colname, groupcol){
2
3    return(leveneTest(y = df$Betacaroteno, group = df$Preparado))
```

4 || }

El test de Levene se utiliza para analizar si hay homocedasticidad o no, por lo tanto, las hipótesis son:

- H_0 : Las varianzas de ambas variables son homogéneas.
- H_0 : Las varianzas de ambas variables no son homogéneas.

2.8. Análisis de normalidad

Cargamos el archivo *cancer.csv* en *Shiny* y observamos mediante el análisis descriptivo de los datos que tiene dos tipos de variable en su interior. Específicamente, notamos que se trata de una variable llamada *Betacaroteno*, la cual contiene niveles plasmáticos de esta vitamina, y otra llamada *Preparado*, la cual indica que preparado se le suministró al paciente.

A partir de esta información, notamos que se debe modificar el tipo de la variable *Preparado* para poder trabajar con la misma, ya que es dicotómica y se trata de una variable de tipo factor. Por ende, realizamos este cambio en la parte de *Preprocesamiento*.

Una vez realizado esto y el análisis exploratorio, se procedió con el análisis de normalidad. Para ello, se hizo uso de las funciones para los test de Shapiro-Wilk y Lilliefors descriptos anteriormente cuyas hipótesis nula y alternativa son las siguientes:

- H_0 : La muestra proviene de una distribución normal
- H_1 : La muestra no proviene de una distribución normal.

Analizando los resultados obtenidos para ambos grupos, se concluye que los dos tienen distribución normal en sus datos. Esto es así ya que, los valores del p-valor obtenidos fueron:

	Shapiro/Wilk	Lilliefors
Preparado 1	p-value = 0.6362	p-value = 0.9119
Preparado 2	p-value = 0.5299	p-value = 0.4959

Como se puede observar, en ambos preparados, los tests estadísticos (Shapiro y Kolmogorov) obtuvieron un p-valor mayor a 0.05, rechazando la hipótesis alternativa y asegurando que la distribución es significativamente similar a la normal.

2.9. T-test

Como respuesta a este ejercicio se generó el siguiente código:

```
1 my_test = function(df,colname,groupcol){
2
3   data = split(df, df$Preparado)
4   dataG1 = data[[1]]
5   dataG2 = data[[2]]
6
7   return(t.test(dataG1$Betacaroteno, dataG2$Betacaroteno,
8                 paired = FALSE, alternative = "two.sided"))
9
10 }
```

Las variables en este ensayo no están apareadas, el número de sujetos es pequeño y se demostró un comportamiento normal y por ende se eligió aplicar el t-test. Las hipótesis son:

- H_0 : La media de las dos muestras (grupos 1 y 2) son iguales.
- H_1 : La media de los dos grupos (el que tomo ibuprofeno y el que no) son diferentes.

Previo a realizar el t-test, se debe realizar el test de Levene para analizar si hay homocedasticidad. Como resultado del test de Levene, se obtiene un p-valor de 0,0061 ($> 0,05$). Por lo tanto, se rechaza la hipótesis alternativa y se acepta la nula. Es decir, las variables tienen varianzas homogéneas y se puede aplicar el t-test.

El p-valor obtenido fue de 0,0002059, con lo cual se puede asegurar que hay una diferencia significativa entre las medias de ambos grupos. Es decir, que se rechaza la hipótesis nula.

En conclusión, un preparado eleva significativamente más los niveles de betacaroteno plasmáticos frente al otro y esta va a ser la mejor opción terapéutica para los pacientes con cáncer.

3. Tercera Parte

3.1. Tabla de contingencia

Como respuesta a este ejercicio se generó el siguiente código:

```
1 contingency = function(df,colname1,colname2){
2
3   tabla <- table(df[,colname1],df[,colname2])
4   tabla
5 }
```

Las tablas de contingencia obtenidas para la relación entre *madre* y *cuidado* y *madre* y *estado* se observan en la Figura 5

	0	1
0	3484	272
1	1428	440

((a)) Relación entre *madre*(columnas) y *cuidado*(filas)

	0	1
0	3507	249
1	1610	258

((b)) Relación entre *madre*(columnas) y *estado*(filas)

Figura 5: Tablas de contingencia.

Para determinar si las variables *cuidado* y *madre* y *estado* y *madre* son independientes se utiliza el Test de Chi Cuadrado. Se puede emplear este método ya que el tamaño de la muestra es grande, es decir $np > 5$ en todos los casos. Estos valores se pueden observar en los elementos de las tablas de valores esperados de las Figuras 6 y 7 (todos mayores a 5).

Por lo tanto, se definen los siguientes pares de hipótesis, según el caso:

- H_0 : El cuidado perinatal en las primeras 48hs del nacimiento y la edad de la madre son variables independientes.

- H_1 : El cuidado perinatal en las primeras 48hs del nacimiento y la edad de la madre no son variables independientes.
- H_0 : El estado del recién nacido y la edad de la madre son variables independientes.
- H_1 : El estado del recién nacido y la edad de la madre no son variables independientes.

Luego de realizado el Test de Chi Cuadrado para cada caso, se obtuvo en ambos el siguiente resultado: $p - \text{valor} < 2,2e - 16$. Dado que es menor a 0.05 se acepta la hipótesis alternativa y se rechaza la hipótesis nula. Por lo tanto, se puede concluir que el cuidado perinatal antes de las 48hs del nacimiento y la edad de la madre no son variables independientes, al igual que el estado del recién nacido y la edad de la madre.

3.2. Valores esperados y residuos

Como respuesta a esta ejercicio se generó el siguiente código:

```
1 | chisq_expected = function(df,colname1,colname2){  
2 |   res = chisq.test(df[,colname1],df[,colname2], correct=FALSE)  
3 |   res$expected  
4 | }  
5 |  
6 | chisq_residuals = function(df,colname1,colname2){  
7 |   res = chisq.test(df[,colname1],df[,colname2], correct=FALSE)  
8 |   res$stdres  
9 | }
```

En consecuencia, los valores esperados para cada celda y los residuos de la tabla de frecuencia para cada caso se muestran en las Figuras 6 y 7.

	0	1
0	3280.49	475.51
1	1631.51	236.49

((a)) Valores esperados

	0	1
0	17.33	-17.33
1	-17.33	17.33

((b)) Residuos

Figura 6: Tablas para *madre*(columnas) y *cuidado*(filas).

	0	1
0	3417.40	338.60
1	1699.60	168.40

((a)) Valores esperados

	0	1
0	8.86	-8.86
1	-8.86	8.86

((b)) Residuos

Figura 7: Tablas para *madre*(columnas) y *estado*(filas).

3.3. OR sin ajustar y OR de Mantel-Haenzel ajustado por año de nacimiento, de madres adolescentes vs. mayores para muerte

Como respuesta a esta ejercicio se generó el siguiente código:

```

1 Odds_r = function(df,colname1,colname2){
2
3   tabla <- table(df[,colname1],df[,colname2])
4   tabla = tabla[2:1, 2:1]
5
6
7   Odds.ratio(tabla)
8
9 }
10 Odds_r_mh = function(df,colname1,colname2, groupcol){

```

```
11
12   outcome <- df[,colname1]
13   exposure <- df[,colname2]
14   stratification <- df[,groupcol]
15   mhor(outcome, exposure, stratification)
16
17 }
18
19 cont_t_s = function(df,colname1,colname2,groupcol){
20
21   tabla = table(df)
22   tablas_parciales = margin.table(tabla,c(colname1,colname2,groupcol))
23   tablas_parciales
24
25 }
26
27
28 Odds_s = function(df,colname1,colname2,groupcol){
29
30
31   tabla = table(df)
32
33   tablas_parciales = margin.table(tabla,c(colname1,colname2,groupcol))
34   ods_estrat = as.vector(apply(tablas_parciales,3,Odds.ratio))
35   ods_estrat
36
37 }
38
39 bdt = function(df,colname1,colname2,groupcol){
40
41   tabla = table(df)
42   tablas_parciales <- margin.table(tabla,c(colname1,colname2,groupcol))
43   BreslowDayTest(x = tablas_parciales)
44 }
```

En primer lugar, se calculó el Odds Ratio crudo, es decir, sin ajustar y se obtuvo un resultado de $OR = 2.26$ e $IC = [1.88, 2.71]$ del 95 %. Por lo tanto, como el OR es > 1 , se puede afirmar que ser madre adolescente es un factor de riesgo que aumenta el riesgo de tener un bebé que no sobreviva en 2.26 veces en comparación a madres mayores.

Luego, se procedió a calcular el Odds Ratio de Mantel-Haenzel de madres adolescentes vs mayores para muerte, ajustado por año de nacimiento. Para ello, se utilizó una tercera variable llamada *añonac* (año de nacimiento). De esta manera, mediante el método mencionado, se busca probar la asociación condicional de las dos variables binarias (*estado* y *madre*) en presencia de una tercera variable (*añonac*).

Específicamente, para poder calcular el Odds Ratio de Mantel-Haenzel, se comienza calculando la tabla de contingencias para cada año y su Odds Ratio, las cuales se observan en la Figura 8.

madre	estado	añonac	Freq				
0	0	1986	306				
1	0	1986	112				
0	1	1986	28	0	0	1991	495
1	1	1986	14	1	0	1991	35
0	0	1987	275	0	1	1991	217
1	0	1987	109	1	1	1991	38
0	1	1987	28	0	0	1992	429
1	1	1987	13	1	0	1992	28
0	0	1988	261	0	1	1992	222
1	0	1988	109	1	1	1992	36
0	1	1988	16	0	0	1993	408
1	1	1988	22	1	0	1993	26
0	0	1989	432	0	1	1993	213
1	0	1989	178	1	1	1993	33
0	1	1989	30	0	0	1994	338
1	1	1989	27	1	0	1994	19
0	0	1990	419	0	1	1994	184
1	0	1990	190	1	1	1994	30
0	1	1990	29	0	0	1995	144
1	1	1990	35	1	0	1995	10
0	0	1991	495	0	1	1995	76
1	0	1991	217	1	1	1995	10
0	1	1991	35				
1	1	1991	38				

Figura 8: Tablas de contingencias para cada año.

Al observar los resultados obtenidos para el Odds Ratio para cada año, se observa que el mismo es mayor a 1 en todos los años (Figura 9).

\$`1986`	OR	2.5 %	97.5 %	p
Fisher's test	1.36509	0.63974	2.7962	0.3678
\$`1987`	OR	2.5 %	97.5 %	p
Fisher's test	1.17090	0.53589	2.4393	0.7168
\$`1988`	OR	2.5 %	97.5 %	p
Fisher's test	3.2815	1.5780	6.9648	0.0007765 ***

Signif. codes:	0	****	0.001	*** 0.01 ** 0.05 ., 0.1 ' ' 1
\$`1989`	OR	2.5 %	97.5 %	p
Fisher's test	2.1815	1.2098	3.919	0.006496 **

Signif. codes:	0	****	0.001	*** 0.01 ** 0.05 ., 0.1 ' ' 1
\$`1990`	OR	2.5 %	97.5 %	p
Fisher's test	2.6573	1.5283	4.6518	0.0002531 ***

Signif. codes:	0	****	0.001	*** 0.01 ** 0.05 ., 0.1 ' ' 1
\$`1991`	OR	2.5 %	97.5 %	p
Fisher's test	2.4735	1.4772	4.1531	0.0003341 ***

Signif. codes:	0	****	0.001	*** 0.01 ** 0.05 ., 0.1 ' ' 1
\$`1992`	OR	2.5 %	97.5 %	p
Fisher's test	2.4812	1.4308	4.342	0.0006017 ***

Signif. codes:	0	****	0.001	*** 0.01 ** 0.05 ., 0.1 ' ' 1
\$`1993`	OR	2.5 %	97.5 %	p
Fisher's test	2.4278	1.3682	4.3481	0.001602 **

Signif. codes:	0	****	0.001	*** 0.01 ** 0.05 ., 0.1 ' ' 1
\$`1994`	OR	2.5 %	97.5 %	p
Fisher's test	2.8947	1.5287	5.6062	0.0005597 ***

Signif. codes:	0	****	0.001	*** 0.01 ** 0.05 ., 0.1 ' ' 1
\$`1995`	OR	2.5 %	97.5 %	p
Fisher's test	1.88945	0.67309	5.3077	0.2226

Figura 9: Odds ratio para cada año.

Luego de realizadas las tablas de contingencia para cada año y su respectivo Odds Ratio, se realiza análisis de la homogeneidad de los Odds Ratio calculados utilizando el Test de Breslow-Day. Este test tiene como objetivo determinar si el Odds Ratio es el mismo para todos los estratos y sus hipótesis son:

- H_0 : Los Odds Ratio entre los estratos son iguales (que el año de nacimiento es una variable confusora).
- H_1 : Al menos un Odds Ratio es distinto (el año de nacimiento es una variable modificadora de efecto).

En los resultados del test se obtuvo un p-valor de 0.495 el cual es mayor a 0.05, por lo que se acepta la hipótesis nula (y se rechaza la alternativa) y se considera que la variable año de nacimiento podría estar actuando como una variable confusora. En consecuencia se introduce un sesgo o error sistemático en la estimación de la medida de asociación.

Por último, se procede a realizar el test de Mantel-Haenszel Chi2 para estudiar si la relación entre las variables *madre* y *estado* sigue siendo estadísticamente significativa, aún si se separa el análisis por estratos. Específicamente, las hipótesis del mismo son:

- H_0 : El Odds Ratio global es igual a 1.
- H_1 : El Odds Ratio global es distinto a 1.

En los resultados del test se obtuvo un p-valor de $2.2e-16$ el cual es menor a 0.05, por lo que se rechaza la hipótesis nula (y se acepta la alternativa) y se considera que el Odds Ratio global es distinto de 1. Específicamente, se obtuvo un Odds Ratio global de 2.2798 con un IC = $[1.895 - 2.742]$ del 95 %, donde no está incluido el 1.

Podemos concluir por lo tanto que, aún analizando por estratos (años), la asociación entre las variables *madre* y *estado* sigue siendo estadísticamente significativa. Además, al realizar el cociente entre el Odds Ratio crudo sobre el Odds Ratio global, obtenemos una diferencia de 0.9913 (diferencia de aproximadamente 1 %). Esto quiere decir que no estamos en presencia de un confundidor ya que la diferencia no fue mayor al 20 %.

3.4. Único OR entre los estratos

En este caso, un único Odds Ratio entre los estratos es coherente con los resultados, ya que, según el Test de Breslow-Day-Wolff no hay evidencia estadísticamente significativa de que al menos uno de los OR sea distinto. Además tampoco hay evidencia estadística de que el año de nacimiento actúe como variable confusora (Test de Mantel-Haenszel Chi2 - relación entre ORs).

Por lo tanto, al reportar los resultados no es necesario ajustar por año de nacimiento, sino que se puede tomar el OR crudo.

3.5. OR sin ajustar y OR de Mantel-Haenszel ajustado por cuidado perinatal, de madres adolescentes vs. mayores para muerte

Para este ejercicio se utilizó el mismo código que en el ejercicio 3.3 pero se seleccionó en la interfaz un ajuste por cuidado perinatal en lugar de año de nacimiento.

El Odds Ratio crudo será el mismo ya que no depende de la tercera variable (ajuste).

En cuanto al Odds Ratio de Mantel-Hanzel ajustado por cuidado perinatal, se utilizó como tercera variable a *cuidado* (cuidado perinatal). Para poder calcularlo, se realizaron las tablas de contingencia por estrato así como también el Odds Ratio, las cuales se observan en las Figuras 10 y 11.

	madre	estado	cuidado	Freq
0	0	0	0	3265
1	0	0	0	1244
0	1	0	0	219
1	1	0	0	184
0	0	1	1	242
1	0	1	1	366
0	1	1	1	30
1	1	1	1	74

Figura 10: Tabla de contingencia por estratos.

```

$`0`
      OR   2.5 % 97.5 %      p
Fisher's test 2.2048 1.7827 2.7247 1.919e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

$`1`
      OR   2.5 % 97.5 %      p
Fisher's test 1.6299 1.0177 2.6634 0.03783 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Figura 11: Odds ratio.

Podemos ver que el Odds Ratio obtenido para los dos tipos de cuidado (Sin cuidado y Con cuidado) es mayor a 1 y distintos al Odds Ratio crudo.

Además, se realizó el análisis de homogeneidad utilizando el Test de Breslow-Day y se obtuvo un p-valor de 0.2347 ¿0.05, con lo cual se acepta la H_0 y se concluye

que los Odds Ratios son homogéneos. Por esta razón, se considera que la variable cuidado perinatal podría estar actuando como variable confusora.

Finalmente, se realizó el test de Mantel-Haenszel χ^2 y se obtuvo un p valor menor a 0.05, por lo que se rechaza la hipótesis nula (y se acepta la alternativa) y se considera que el Odds Ratio global es distinto de 1. Específicamente, se obtuvo un Odds Ratio global de 2.08 con un IC = [1.72;2.51] del 95 %, donde no está incluido el 1.

Podemos concluir por lo tanto que, aún analizando por estratos (cuidado perinatal), la asociación entre las variables *madre* y *estado* sigue siendo estadísticamente significativa. Además, al realizar el cociente entre el Odds Ratio crudo sobre el Odds Ratio global, obtenemos una diferencia de 1.09 (diferencia de aproximadamente 1 %). Esto quiere decir que no estamos en presencia de un confundidor ya que la diferencia no fue mayor al 20 %.

Por lo tanto, al igual que en el caso del año de nacimiento, al reportar los resultados no es necesario ajustar por cuidado, sino que se puede tomar el OR crudo.