

Trabajo Grupal 01

Jesus Mauricio Huayhua Flores

2024-04-26

Apertura de archivos

```
library(readxl)

## Warning: package 'readxl' was built under R version 4.3.3

library(DescTools)
library(MASS)

## Warning: package 'MASS' was built under R version 4.3.3

library(ggplot2)
```

Lectura de Archivos

```
#Consideraciones se eliminaron los 3 valores encontrados a la derecha ya que estos
#se podian calcular usando R.
abulones_data <- read_xlsx(path = "abulones.xlsx")
# Impresion de los valores de file
head(abulones_data)

## # A tibble: 6 x 9
##   sex   length Diameter Height WholeHeight ShuckedWeight VisceraWeight
##   <chr>   <dbl>    <dbl>  <dbl>      <dbl>        <dbl>        <dbl>
## 1 M      0.455    0.365  0.095      0.514        0.224        0.101
## 2 M      0.35     0.265  0.09       0.226        0.0995       0.0485
## 3 F      0.53     0.42   0.135     0.677        0.256        0.142
## 4 M      0.44     0.365  0.125     0.516        0.216        0.114
## 5 I      0.33     0.255  0.08      0.205        0.0895       0.0395
## 6 I      0.425    0.3    0.095     0.352        0.141        0.0775
## # i 2 more variables: ShellWeight <dbl>, Rings <dbl>
```

Pregunta01

Calcule la tabla de distribución con las frecuencias absolutas y con las frecuencias esperadas que se obtendrían bajo la hipótesis de que el diámetro de los abulones sigue la distribución normal. (Para determinar el número de intervalos, no se olvide de usar la regla de Sturges.)

Desarrollo

```
# Calcular el número de intervalos utilizando la regla de Sturges
n <- nrow(abulones_data)
```

```

k <- ceiling(1 + log2(n))

# Calcular el ancho de cada intervalo
rango_diametro <- range(abulones_data$Diameter)
ancho_intervalo <- (rango_diametro[2] - rango_diametro[1]) / k

# Crear los límites de los intervalos
limites_intervalo <- seq(rango_diametro[1], rango_diametro[2] + ancho_intervalo,
                        by = ancho_intervalo)

# Calcular las frecuencias absolutas
frecuencia <- cut(abulones_data$Diameter, breaks = limites_intervalo, right = FALSE)
tabla_frecuencia <- table(frecuencia)

# Calcular las frecuencias esperadas asumiendo una distribución normal
# Ajustar una distribución normal a los datos de diámetro
normal_ajustada <- fitdistr(abulones_data$Diameter, "normal")

# Obtener los parámetros de la distribución normal ajustada
mu <- normal_ajustada$estimate[1]
sigma <- normal_ajustada$estimate[2]

# Calcular las probabilidades acumuladas en cada intervalo
probabilidades_acumuladas <- pnorm(limites_intervalo, mean = mu, sd = sigma)

# Calcular las frecuencias esperadas
frecuencia_esperada <- diff(probabilidades_acumuladas) * n

# Mostrar la tabla de distribución
tabla_distribución <- data.frame(Intervalo = paste(round(limites_intervalo[-length(limites_intervalo)]),
                                                  Frecuencia_Absoluta = as.vector(tabla_frecuencia),
                                                  Frecuencia_Esperada = as.vector(frecuencia_esperada))
print(tabla_distribución)

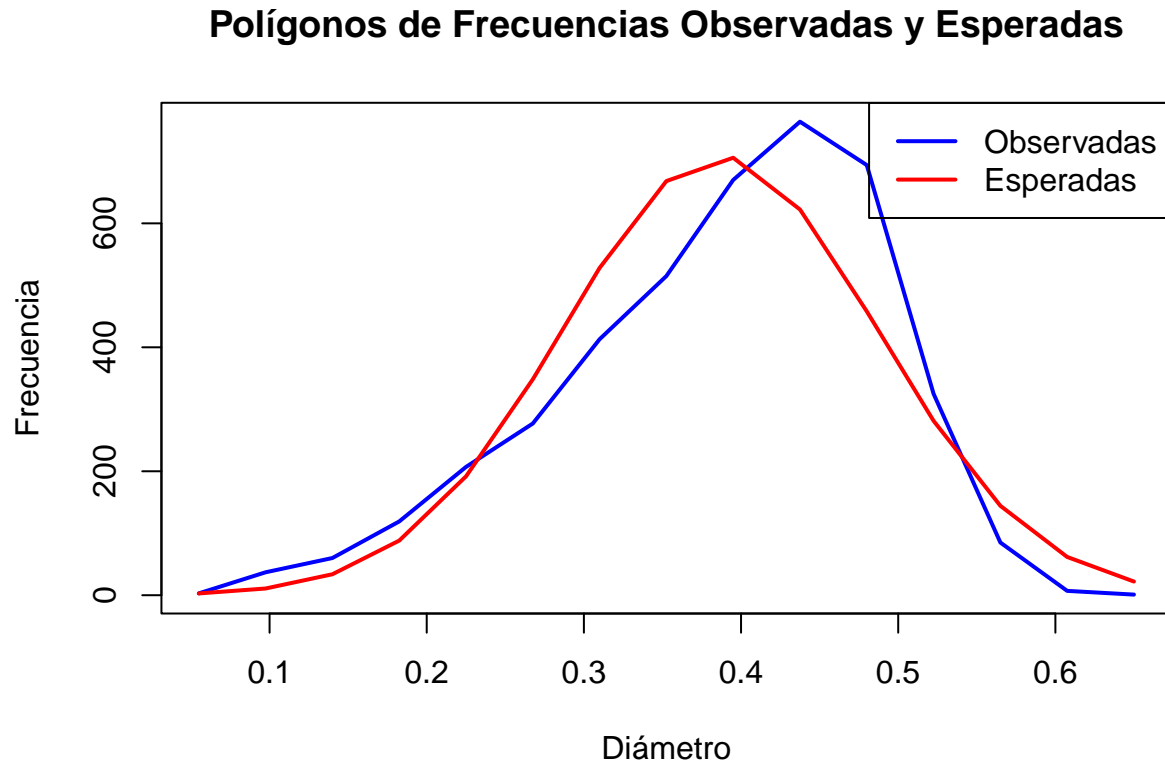
```

##	Intervalo	Frecuencia_Absoluta	Frecuencia_Esperada
## 1	0.06 - 0.1	3	2.890641
## 2	0.1 - 0.14	37	10.820353
## 3	0.14 - 0.18	60	33.801919
## 4	0.18 - 0.22	119	88.128587
## 5	0.22 - 0.27	207	191.773055
## 6	0.27 - 0.31	277	348.313388
## 7	0.31 - 0.35	413	528.052929
## 8	0.35 - 0.39	515	668.218630
## 9	0.39 - 0.44	670	705.827791
## 10	0.44 - 0.48	764	622.327792
## 11	0.48 - 0.52	694	458.012266
## 12	0.52 - 0.57	325	281.362182
## 13	0.57 - 0.61	85	144.269459
## 14	0.61 - 0.65	7	61.743154
## 15	0.65 - 0.69	1	22.054191

Pregunta02

Con las distribuciones observadas y las esperadas, desarrolle 2 gráficos de polígonos de frecuencias superpuestos en el mismo plano. Comente si le parece que la distribución parece o no ser la normal.

Desarrollo



Para evaluar si la distribución parece ser normal, observamos la superposición de las frecuencias observadas y esperadas en el gráfico. Si las curvas se superponen de manera cercana, sugiere que la distribución observada es similar a la esperada bajo una distribución normal.

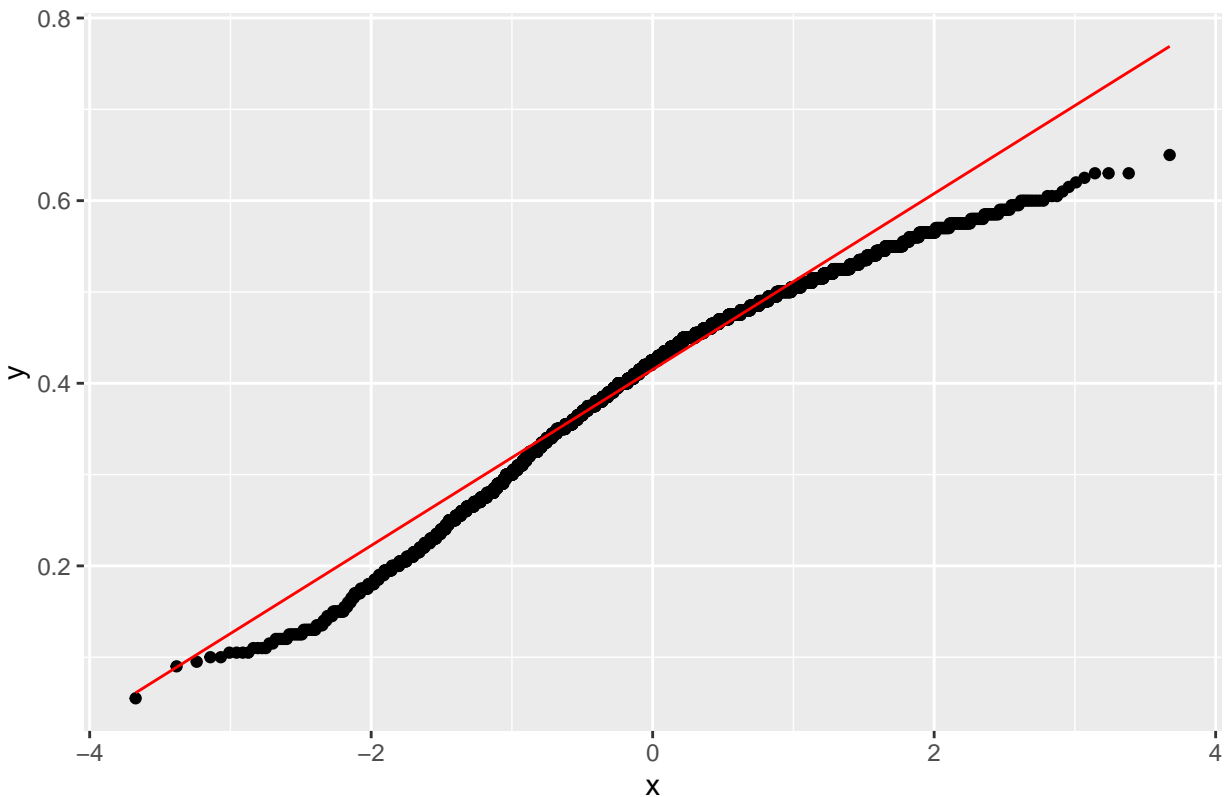
Pregunta03

Usando el RStudio (o el Rcommander), obtenga la gráfica de normalidad. Comente si se puede decir que la distribución del diámetro de los abulones parece seguir la distribución normal.

Desarrollo

```
# Crear el gráfico de normalidad
ggplot(data = abulones_data, aes(sample = Diameter)) +
  geom_qq() +
  geom_qq_line(color = "red") +
  labs(title = "Gráfico Q-Q de Normalidad para el Diámetro de los Abulones")
```

Gráfico Q–Q de Normalidad para el Diámetro de los Abulones



Después de observar la gráfica, podemos verificar que los puntos siguen la línea diagonal; por lo tanto se puede indicar que los datos están distribuidos normalmente.

Pregunta04

De acuerdo al objetivo de la investigación, establezca las hipótesis nula y alternativa. Calcule manualmente de la tabla la estadística de prueba chi-cuadrado, el valor crítico correspondiente y el p-valor (si es el caso, no se olvide de: juntar intervalos con menos de 5 casos esperados y considerar el número de grados de libertad adecuado). ¿A cuál conclusión se puede llegar?

Desarrollo

Para establecer las hipótesis nula y alternativa, primero necesitamos definir el objetivo de la investigación. Dado que estamos analizando si el diámetro de los abulones sigue una distribución normal, una posible hipótesis nula y alternativa podrían ser las siguientes:

- Hipótesis nula(H_0): El diámetro de los abulones sigue una distribución normal.
- Hipótesis alternativa(H_1): El diámetro de los abulones no sigue una distribución normal.

Para probar estas hipótesis, podemos utilizar la prueba de bondad de ajuste Chi-cuadrado, que compara las frecuencias observadas con las frecuencias esperadas bajo la distribución normal.

Para calcular la estadística de prueba Chi-cuadrado manualmente, primero necesitamos la frecuencia observada (O_i) y la frecuencia esperada (E_i) para cada intervalo. Luego, calcularemos la diferencia entre las frecuencias observadas y esperadas al cuadrado, dividiremos esta diferencia por las frecuencias esperadas y sumaremos estos valores para obtener la estadística de prueba Chi-cuadrado.

Después, necesitaremos determinar el valor crítico y el p-valor correspondiente. El valor crítico se encuentra en la tabla de la distribución Chi-cuadrado con un nivel de significancia dado y el número adecuado de grados

de libertad. El p-valor se calcula utilizando la distribución Chi-cuadrado y representa la probabilidad de obtener un valor de prueba al menos tan extremo como el observado bajo la hipótesis nula.

Finalmente, compararemos la estadística de prueba Chi-cuadrado con el valor crítico y el p-valor para tomar una decisión sobre la hipótesis nula.

Ahora, procedamos a calcular manualmente la estadística de prueba Chi-cuadrado, el valor crítico y el p-valor.

La fórmula para la estadística de prueba Chi-cuadrado es:

$$X^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Valor crítico Chi-cuadrado($X^2_{\alpha, df}$)

- $\alpha = 0.5$
- $df = k - 1 = 14 - 1 = 13$