1. El número de horas por semana que 200 estudiantes de MACC gastan procrastinando (en redes sociales, videojuegos, etc.) está agrupada en las clases 0 a 3, 4 a 7, 8 a 11, 12 a 15, 16 a 19, 20 a 23, y 24 a 27, con las respectivas frecuencias observadas 12, 25, 36, 45, 34, 31 y 17. La media y la desviación estándar agrupadas se pueden calcular a partir de los datos. *La hipótesis nula es que los datos se encuentran distribuidos normalmente*. Usando la media y la desviación estándar que se encuentran a partir de los datos agrupados, se determinaron las siguientes frecuencias esperadas (correspondientes): 10, 30, 40, 50, 36, 28 y 6. Realice una prueba de bondad de ajuste para determinar si hay sustento estadístico para la hipótesis planteada.

    a) Calcule el estadístico de prueba.

    b) Determine la región de rechazo para un nivel de significancia del 5 % y concluya.

    c) Determine el *valor p* y concluya.

2. La tabla 1 proporciona valores experimentales de la presión P de cierta masa de gas que corresponde a varios valores del volumen V. De acuerdo a la termodinámica, existe una relación entre las variables de la forma $PV^{\gamma} = C$, donde γ y C son constantes. Realice todos los cálculos usando las expresiones que dependen de S_{xx}, S_{xy} y S_{yy}.

Volumen (m^3)	54.3	61.8	72.4	88.7	118.6	194.0
Presión (N/m^2)	61.2	49.5	37.6	28.4	19.2	10.1

Cuadro 1: Datos para el punto 2

    a) Estime los valores de γ y C, y escriba la ecuación que relaciona a P y V.

    b) Estime el valor de P cuando $V = 100\,m^3$ y su intervalo al 95 % de confianza.

    Sugerencia: Dado que $PV^{\gamma} = C$, se puede tomar logaritmos (base 10, por ejemplo) a ambos lados de la ecuación produciendo: $\log P = \log C - \gamma \log V$. Puede utilizar $x = \log V$ y $y = \log P$ y realizar una regresión lineal simple para solucionar el problema escribiendo la ecuación como $y = \beta_0 + \beta_1 x$.

3. Un coeficiente de correlación para una muestra de 24 se calculó igual a $r = 0{,}75$. ¿Se puede rechazar la hipótesis de que el coeficiente de correlación poblacional es tan pequeño como los valores de ρ indicados, a un nivel de significancia de 0.05?

    a) $\rho = 0{,}60$.

    b) $\rho = 0{,}50$.

4. Como parte del estudio de una especie de insectos se han capturado 10 especímenes y se les ha medido su longitud en centímetros. A continuación se muestran los datos recolectados.

0.2	0.3	0.6	0.7	0.9	1.3	1.4	1.5	1.7	1.9

Al observar estos datos, la persona a cargo de la investigación ha propuesto que la longitud de los insectos sigue una distribución uniforme entre 0 y 2 centímetros. Realice una prueba de bondad de ajuste para determinar si hay sustento estadístico para la hipótesis planteada.

a) Construya un histograma de los datos con 5 cajas de igual tamaño entre 0 y 2.

b) Calcule el estadístico de prueba.

c) Determine la región de rechazo y concluya.

5. Se busca estudiar la relación entre la longitud y el ancho del sépalo de una especie de flores. Para este fin se han recolectado 150 muestras y se ha propuesto un modelo lineal de la forma

$$\text{LongSepalo} = \beta_0 + \beta_1 \text{AnchoSepalo} + \epsilon.$$

El resultado de correr el modelo en R se muestra continuación:

```
Call:
lm(formula = Sepal.Length ~ Sepal.Width)

Residuals:
    Min      1Q  Median      3Q     Max
-1.5561 -0.6333 -0.1120  0.5579  2.2226

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.5262     0.4789   13.63   <2e-16 ***
Sepal.Width  -0.2234     0.1551   -1.44    0.152
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8251 on 148 degrees of freedom
Multiple R-squared:  0.01382,   Adjusted R-squared:  0.007159
F-statistic: 2.074 on 1 and 148 DF,  p-value: 0.1519
```

a) ¿Qué puede concluir sobre β_1? ¿Es significativamente diferente de 0?

b) ¿Qué puede concluir sobre la relación propuesta entre estas dos variables?

6. Ahora se busca estudiar la relación entre la longitud del sépalo y el ancho del **pétalo** de la misma especie de flores del punto anterior. Usando los mismos datos del punto anterior se ha propuesto un modelo lineal de la forma

$$\text{LongSepalo} = \beta_0 + \beta_1 \text{AnchoPetalo} + \epsilon.$$

El resultado de correr el modelo en R se muestra continuación:

```
Call:
lm(formula = Sepal.Length ~ Petal.Width)

Residuals:
     Min        1Q    Median        3Q       Max
-1.38822  -0.29358  -0.04393   0.26429   1.34521

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.77763    0.07293   65.51   <2e-16 ***
Petal.Width  0.88858    0.05137   17.30   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.478 on 148 degrees of freedom
Multiple R-squared:  0.669,     Adjusted R-squared:  0.6668
F-statistic: 299.2 on 1 and 148 DF,  p-value: < 2.2e-16
```

a) ¿Qué puede concluir sobre β_1? ¿Es significativamente diferente de 0?

b) Al comparar el R^2 de este modelo con el del modelo del punto anterior, ¿qué puede concluir?

7. Se han tomado 100 muestras del peso (en gramos) y el número de días de vida de pollos de cierta especie. Con estos datos se ha estimado un modelo lineal que explica la variable dependiente Peso a través de la variable independiente Días, así

$$\text{Peso} = \beta_0 + \beta_1 \text{Dias} + \epsilon.$$

Se ha estimado el valor de los parámetros con el método de cuadrados mínimos, obteniendo $\hat{\beta}_0 = 20$ y $\hat{\beta}_1 = 10$. También se sabe que la suma de cuadrados del error (SSE) del modelo es 1500, mientras la media muestral del número de días de vida es $\bar{x} = 15$, y la suma de las diferencias al cuadrado en el número de días de vida es $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 1000$.

a) Determine un intervalo al 95 % de confianza para el valor esperado del peso de un pollo de 10 días de vida.

b) Determine un intervalo al 95 % de confianza para el peso de un pollo de 10 días de vida.

c) Interprete el valor de $\hat{\beta}_0$ y $\hat{\beta}_1$ dentro del contexto del problema.

8. En R cargue los datos `mtcars` (mediante `data(mtcars)`) y utilícelos para realizar los siguientes ejercicios. Realice todos los cálculos usando las expresiones que dependen de S_{xx}, S_{xy} y S_{yy} (es decir, sin utilizar comandos directos de R, como `lm`).

a) Estudie y asegúrese de entender todos los campos.

b) Calcule estadísticas descriptivas de cada campo.

c) Considere los campos `mpg` y `hp`.

1) Grafique un diagrama de dispersión de estas dos variables.
2) Estime un modelo de regresión lineal entre estas dos variables, dejando a la variable `hp` como variable independiente. Determine el valor de $\hat{\beta}_0$ y $\hat{\beta}_1$.
3) Realice una prueba de hipótesis para determinar si β_1 es diferente de 0 o no.
4) Construya un intervalo de confianza para β_1.
5) Concluya sobre el valor de β_1 en el contexto del problema.
6) Realice una prueba de hipótesis para determinar si β_0 es diferente de 0 o no.
7) Construya un intervalo de confianza para β_0.
8) Concluya sobre el valor de β_0 en el contexto del problema.
9) Estime la correlación entre estas dos variables.

d) Considere los campos `qsec` y `hp`.

1) Estime la correlación entre estas dos variables.
2) Grafique un diagrama de dispersión de estas dos variables.
3) Estime un modelo de regresión lineal entre estas dos variables, dejando a la variable `qsec` como variable independiente. Determine el valor de $\hat{\beta}_0$ y $\hat{\beta}_1$.
4) Realice una prueba de hipótesis para determinar si β_1 es diferente de 0 o no.
5) Construya un intervalo de confianza para β_1.
6) Concluya sobre el valor de β_1 en el contexto del problema.
7) Realice una prueba de hipótesis para determinar si β_0 es diferente de 0 o no.
8) Construya un intervalo de confianza para β_0.
9) Concluya sobre el valor de β_0 en el contexto del problema.