

Ejercicio 5

```
In [6]: 1 import pandas as pd
        2 import seaborn as sns
        3 import matplotlib.pyplot as plt
        4
        5 # Cargar el conjunto de datos
        6 df = pd.read_csv('datos2(1).csv')
        7
        8
        9
       10 df
```

```
Out[6]:
```

	age	lwt	race	smoke	ptd	ht	ui	ftv	bwt
0	19	182	2	0	0	0	1	0	2523
1	33	155	3	0	0	0	0	3	2551
2	20	105	1	1	0	0	0	1	2557
3	21	108	1	1	0	0	1	2	2594
4	18	107	1	1	0	0	1	0	2600
...
184	28	95	1	1	0	0	0	2	2466
185	14	100	3	0	0	0	0	2	2495
186	23	94	3	1	0	0	0	0	2495
187	17	142	2	0	0	1	0	0	2495
188	21	130	1	1	0	1	0	3	2495

189 rows × 9 columns

```
In [7]: 1 sns.pairplot(df, hue="bwt")
        2
        3 # Mostrar el gráfico
        4 plt.show()
```

```
/home/jl_deleon/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:306: UserWarning: Dataset has 0 variance; skipping density estimate.
```

```
warnings.warn(msg, UserWarning)
```

```
/home/jl_deleon/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:306: UserWarning: Dataset has 0 variance; skipping density estimate.
```

```
warnings.warn(msg, UserWarning)
```

```
/home/jl_deleon/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:306: UserWarning: Dataset has 0 variance; skipping density estimate.
```

```
warnings.warn(msg, UserWarning)
```

```
/home/jl_deleon/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:306: UserWarning: Dataset has 0 variance; skipping density estimate.
```

```
warnings.warn(msg, UserWarning)
```

```
/home/jl_deleon/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:306: UserWarning: Dataset has 0 variance; skipping density estimate.
```

```
warnings.warn(msg, UserWarning)
```

```
/home/jl_deleon/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:306: UserWarning: Dataset has 0 variance; skipping density estimate.
```

```
warnings.warn(msg, UserWarning)
```

```
/home/jl_deleon/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:306: UserWarning: Dataset has 0 variance; skipping density estimate.
```

```
warnings.warn(msg, UserWarning)
```

```
/home/jl_deleon/anaconda3/lib/python3.8/site-packages/seaborn/distributions.py:306: UserWarning: Dataset has 0 variance; skipping density estimate.
```

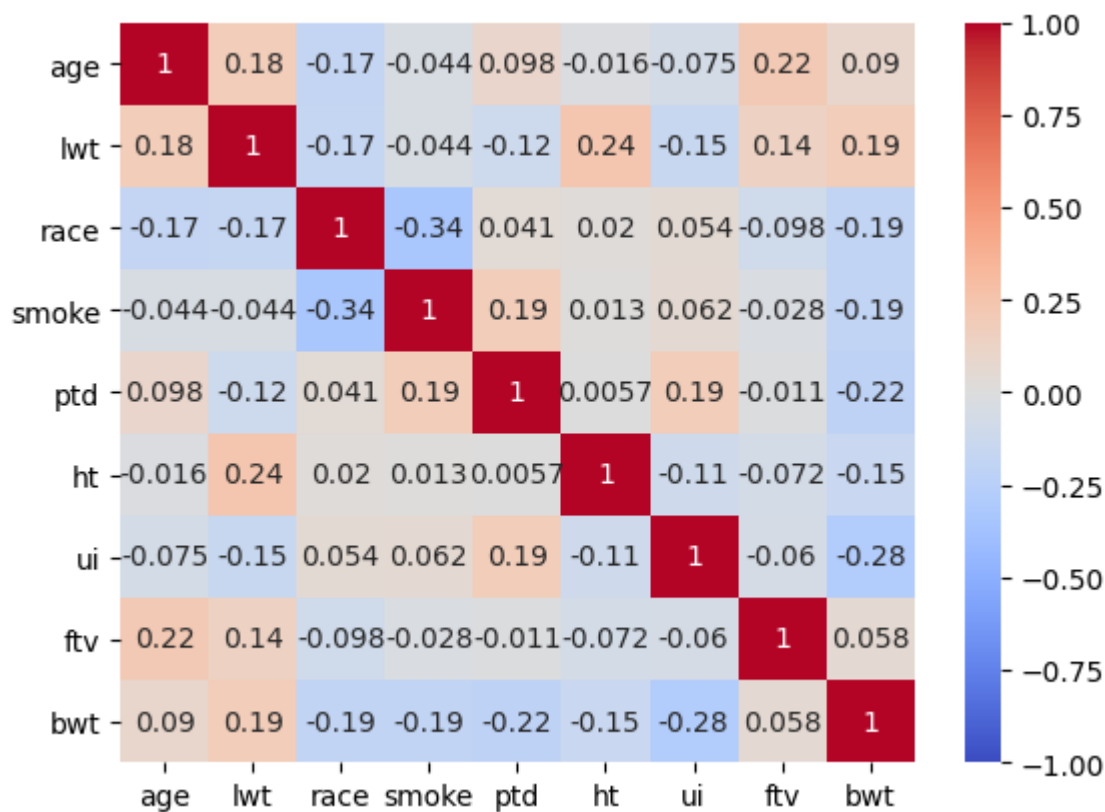
```
warnings.warn(msg, UserWarning)
```

In [8]:

```

1 correlation_matrix = df.corr()
2
3 sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", vmin=-
4
5 plt.show()
6

```



In [14]:

```
1 import statsmodels.api as sm
2
3 # Crear variables binarias para la raza
4 df['race_white'] = (df['race'] == 1).astype(int)
5 df['race_black'] = (df['race'] == 2).astype(int)
6
7 # Seleccionar las variables predictoras (X) y la variable dependien
8 X = df[['age', 'lwt', 'race_white', 'race_black', 'smoke', 'ptd', '
9 y = df['bwt']
10
11 # Agregar una constante al conjunto de variables predictoras (inter
12 X = sm.add_constant(X)
13
14 # Ajustar el modelo de regresión lineal con variables categóricas
15 model = sm.OLS(y, X).fit()
16
17 # Obtener un resumen del modelo
18 print(model.summary())
19
20 # Obtener los residuos del modelo
21 residuals = model.resid
22
23 # También puedes obtener otras estadísticas relacionadas con los re
24 print("Estadísticas sobre los residuos:")
25 print("Media de los residuos:", residuals.mean())
26 print("Desviación estándar de los residuos:", residuals.std())
```

OLS Regression Results

```
=====
Dep. Variable:          bwt    R-squared:
0.251
Model:                  OLS    Adj. R-squared:
0.217
Method:                 Least Squares    F-statistic:
7.528
Date:                  Fri, 24 Nov 2023    Prob (F-statistic):
1.18e-08
Time:                  21:04:50    Log-Likelihood:
-1486.3
No. Observations:      189    AIC:
2991.
Df Residuals:          180    BIC:
3020.
Df Model:               8
Covariance Type:       nonrobust
=====
```

	coef	std err	t	P> t	[0.025
const	2580.7070	280.969	9.185	0.000	2026.289
age	-2.3633	9.434	-0.251	0.802	-20.979

lwt	4.1270	1.711	2.413	0.017	0.752
7.503					
race_white	338.5422	114.120	2.967	0.003	113.357
563.727					
race_black	-135.1935	158.028	-0.856	0.393	-447.020
176.633					
smoke	-328.1404	105.957	-3.097	0.002	-537.217
-119.063					
ptd	-202.2497	135.722	-1.490	0.138	-470.060
65.561					
ht	-576.5515	199.823	-2.885	0.004	-970.848
-182.255					
ui	-493.5903	136.661	-3.612	0.000	-763.254
-223.926					

```

=====
=====
Omnibus:                1.467    Durbin-Watson:
0.613
Prob(Omnibus):          0.480    Jarque-Bera (JB):
1.552
Skew:                   -0.178    Prob(JB):
0.460
Kurtosis:               2.736    Cond. No.
820.
=====
=====

```

Notes:

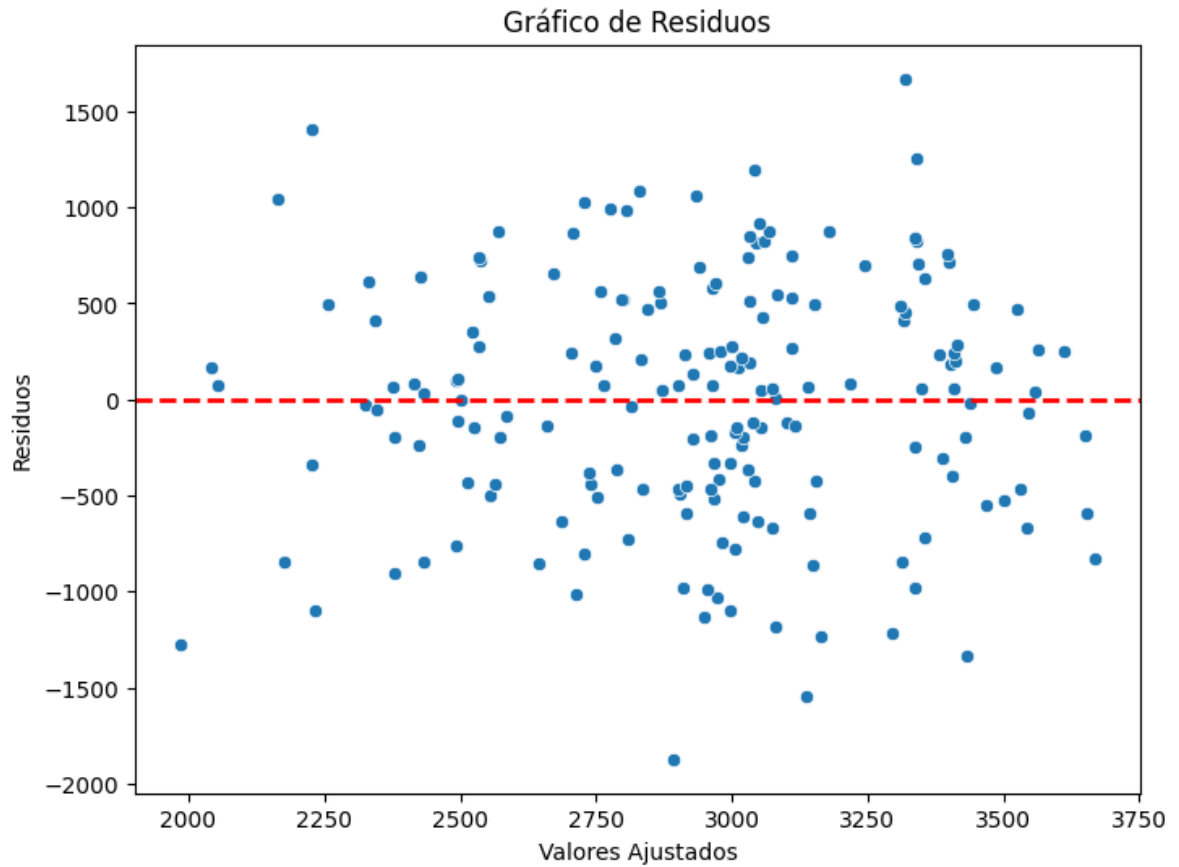
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Estadísticas sobre los residuos:

Media de los residuos: 2.9835275931175424e-12

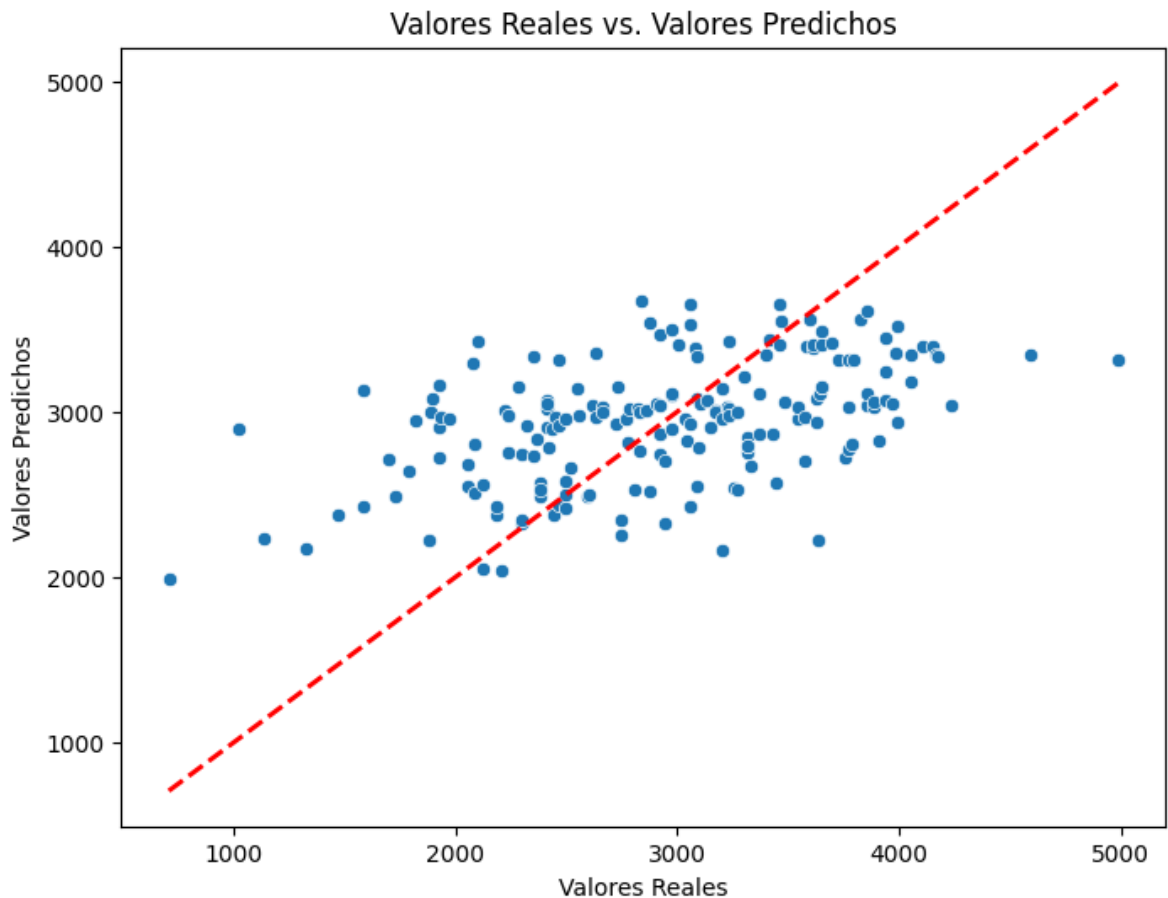
Desviación estándar de los residuos: 631.2283273593328

```
In [15]: 1 # Crear un gráfico de dispersión de los residuos
2 plt.figure(figsize=(8, 6))
3 sns.scatterplot(x=model.fittedvalues, y=residuals)
4 plt.axhline(y=0, color='red', linestyle='--', linewidth=2) # Línea
5 plt.title('Gráfico de Residuos')
6 plt.xlabel('Valores Ajustados')
7 plt.ylabel('Residuos')
8 plt.show()
```



In [16]:

```
1 # Obtener los valores ajustados por el modelo
2 predicted_values = model.fittedvalues
3 # Crear un gráfico de dispersión de valores reales vs. valores pred
4 plt.figure(figsize=(8, 6))
5 sns.scatterplot(x=y, y=predicted_values)
6 plt.plot([y.min(), y.max()], [y.min(), y.max()], linestyle='--', co
7 plt.title('Valores Reales vs. Valores Predichos')
8 plt.xlabel('Valores Reales')
9 plt.ylabel('Valores Predichos')
10 plt.show()
```



Comentarios

Después de analizar con varias variables del data set el modelo que mejor se ajusta es quitando la variable del ftv: visitas al médico durante los primeros tres meses del embarazo.

Esto tambien se puede ver en que es de las variables que menos correlación tiene en nuestra matriz de correlación, hablando de razones reales puede deberse a que visitas al medico durante los primeros tres meses del embarazo no tiene alguna relación proporcional con el peso, ya que se puede interpretar que muchas visitas al médico puede significar un estado delicado del embarazo lo que puede resultar en un peso bajo, pero se puede hacer la misma interpretación de pocas visitas.

