```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

# Load dataset
df = pd.read_csv("/content/archive (3).zip")

# Descriptive statistics
df.describe()
```

	Athlete_ID	Age	Session_Duration	Heart_Rate_Avg	Speed_Avg	Distance_Covered	Endurance_Score	Technique_Score
count	10.00000	10.000000	10.000000	10.00000	10.000000	10.000000	10.000000	10.000000
mean	5.50000	21.800000	48.600000	137.80000	7.465927	9335.000000	75.260188	66.048600
std	3.02765	1.813529	16.507237	10.71655	1.734120	3544.659238	17.512748	10.130394
min	1.00000	19.000000	30.000000	122.00000	5.325258	5021.000000	52.035228	52.982026
25%	3.25000	20.250000	37.250000	129.50000	5.838411	7092.750000	58.443645	60.949816
50%	5.50000	22.000000	43.000000	137.50000	7.338324	7893.000000	80.893238	63.641673
75%	7.75000	23.500000	53.000000	145.50000	8.886781	12058.250000	91.180063	70.167655
max	10.00000	24.000000	80.000000	156.00000	9.828160	14805.000000	93.631308	83.149500

Statistical Analysis of Descriptive Data

Interpretation

The principal characteristics of the dataset were summarized with the help of descriptive statistics. All the results show that, athletes mostly engaged in moderate length training sessions with an average session length of 48.6 minutes. The average result of endurance that stands at 75.26 indicates reasonably good performers among sportspeople. Standard deviation of the time spent on the session is moderate and this means that the training time was different among the individuals. Generally, the descriptive statistics will give a basic knowledge on the training behavior and performance prior to performing the inferential analysis.

```
corr, p_value = stats.pearsonr(df['Session_Duration'], df['Endurance_Score'])
corr, p_value
```

```
(np.float64(-0.04222307888724895), np.float64(0.9078015025301417))
```

Inferential Statistical Analysis (Pearson Correlation)

Interpretation

The correlation analysis that Pearson undertook was aimed at studying the correlation of the duration of training sessions and the endurance performance. The correlation coefficient (r = -.04) shows that there is a very weak negative association of the two variables. The p-value of (.91) is also significantly higher than the significance level of (.05) implying that the observed correlation is non-statistically significant. This finding indicates that variations in the training session duration are not coupled with the significant variation in endurance performance in the considered data set.

```
df.isnull().sum()
```

	0
Athlete_ID	0
Age	0
Gender	0
Sport_Type	0
Session_ID	0
Date	0
Session_Duration	0
Heart_Rate_Avg	0
Speed_Avg	0
Distance_Covered	0
Endurance_Score	0
Technique_Score	0
Performance_Level	0

dtype: int64

Diagnostic Measures

Interpretation

The analysis of the missing values indicated that none of the variables had a null value. This ensures that the dataset is comprehensive and it is applicable in the statistical analysis. Incidentally, missing data minimizes the chances of bias and increases the credibility of the findings.

```
df.duplicated().sum()
```

```
np.int64(0)
```

Duplicate Records Check

Interpretation

The check of duplicate records proved that no repetition was taken in the dataset. This implies that every training session will be an independent piece of data, such that the statistical analysis will not be confounded by repeated data.

```
Q1 = df['Session_Duration'].quantile(0.25)
Q3 = df['Session_Duration'].quantile(0.75)
IQR = Q3 - Q1

df_filtered = df[
    (df['Session_Duration'] >= Q1 - 1.5 * IQR) &
    (df['Session_Duration'] <= Q3 + 1.5 * IQR)
]
```

Outlier Detection (IQR Method)

Interpretation

An IQR was used to determine the outliers that may occur in the training session duration. The dataset was not changed significantly by the process of filtering, which is the implication that there were few extreme and unrealistic training periods. This indicates that there is some stable and reasonable training behavior in the athletes and it is going to be reasonable to carry out the subsequent analyses.

Double-click (or enter) to edit

```
if p_value < 0.05:
    print("Reject H₀: Training duration significantly affects endurance performance.")
else:
    print("Fail to reject H₀: No significant relationship found.")
```

```
Fail to reject H₀: No significant relationship found.
```

Hypothesis Decision and Rejecting

Interpretation

The significance level that was used to test the hypothesis is a =.05. The p-value found in the Pearson correlation analysis was more than.05, which states that the correlation between the duration of the training session and the endurance performance was not significant. Thus, the null hypothesis was not not rejected. This result indicates that, to the best of the current data, there is no statistically significant relationship between training duration and the endurance performance.

It is possible that the small sample size or the availability of other factors that were not measured such as recovery, nutrition or training intensity led to lack of statistical significance. It is suggested that future researches be done on a larger population and more variables of performance as an alternative to further examine this relationship.