In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [ ]:
```
df=pd.read_csv('/content/drive/MyDrive/datasets/weight-height.csv')
df.sample(5)
```

Out[ ]:

	Gender	Height	Weight
4986	Male	71.644419	190.063286
480	Male	69.736087	185.524260
4880	Male	72.153511	196.538474
9405	Female	68.238296	146.865810
9203	Female	63.683083	131.593043

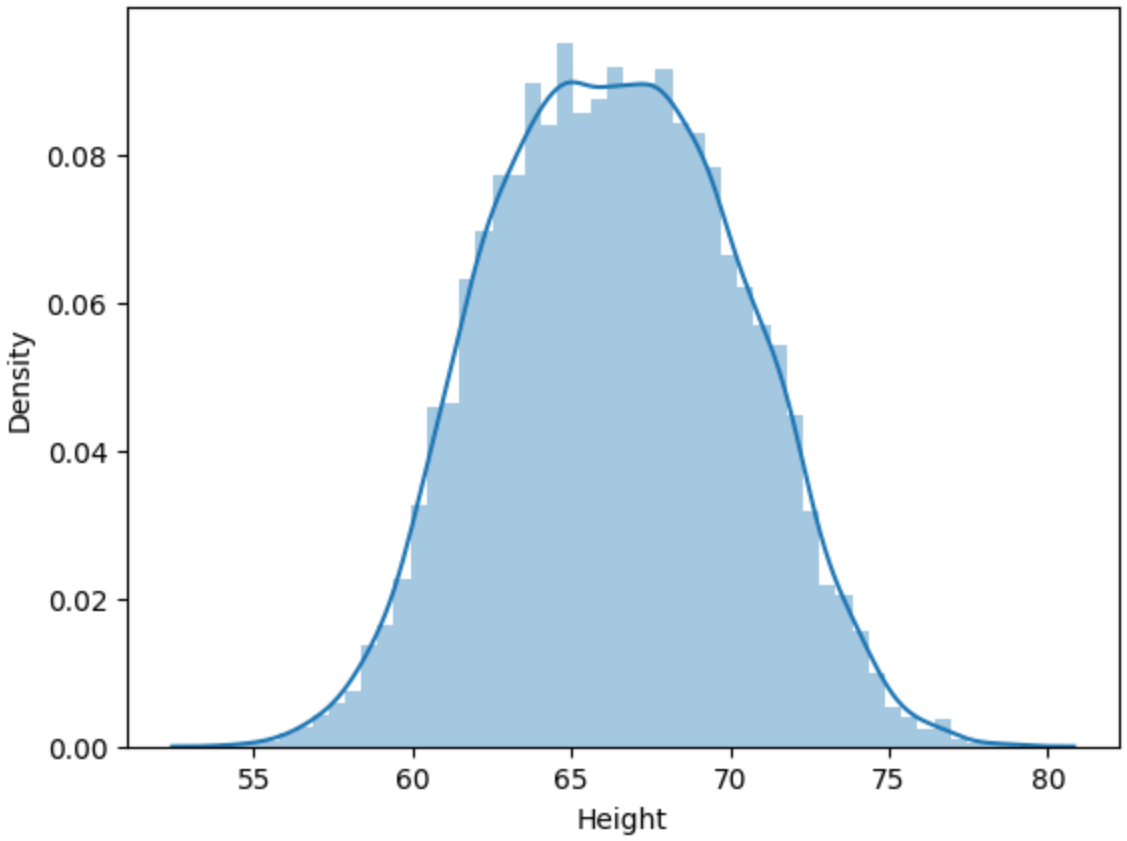
In [ ]:
```
df.shape
```

Out[ ]:
```
(10000, 3)
```

In [ ]:
```
sns.distplot(df['Height'])
```

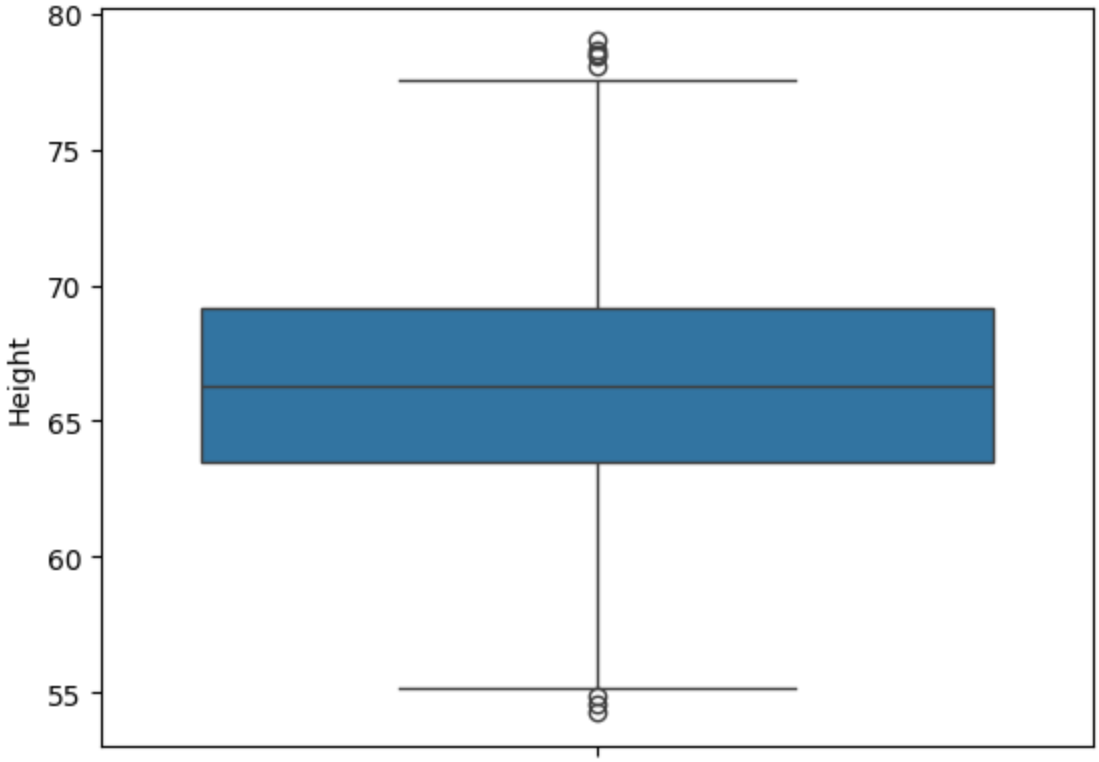
```
<ipython-input-21-26dbcd94f059>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['Height'])
```

Out[ ]:
```
<Axes: xlabel='Height', ylabel='Density'>
```



In [ ]:
```
sns.boxplot(df['Height'])
```

Out[ ]:
```
<Axes: ylabel='Height'>
```



In [ ]:
```
df.describe()
```

Out[ ]:

	Height	Weight
count	10000.000000	10000.000000
mean	66.367560	161.440357
std	3.847528	32.108439
min	54.263133	64.700127
25%	63.505620	135.818051
50%	66.318070	161.212928
75%	69.174262	187.169525
max	78.998742	269.989699

In [ ]:
```
print('upperlimit'),df['Height'].mean() + 3*df['Height'].std()
```

```
upperlimit
```

Out[ ]:
```
(None, 77.91014411714093)
```

In [ ]:
```
print('lowerlimit'),df['Height'].mean() - 3*df['Height'].std()
```

```
lowerlimit
```

Out[ ]:
```
(None, 54.82497539250156)
```

In [ ]:
```
df[(df['Height']>77.9) | (df['Height']<54.8)]
```

Out[ ]:

	Gender	Height	Weight
994	Male	78.095867	255.690835
1317	Male	78.462053	227.342565
2014	Male	78.998742	269.989699
3285	Male	78.528210	253.889004
3757	Male	78.621374	245.733783
6624	Female	54.616858	71.393749
9285	Female	54.263133	64.700127

In [ ]:
```
new_df=df[(df['Height']<77.9) & (df['Height']>54.8)]
new_df
```

Out[ ]:

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801
...	...	...	...
9995	Female	66.172652	136.777454
9996	Female	67.067155	170.867906
9997	Female	63.867992	128.475319
9998	Female	69.034243	163.852461
9999	Female	61.944246	113.649103

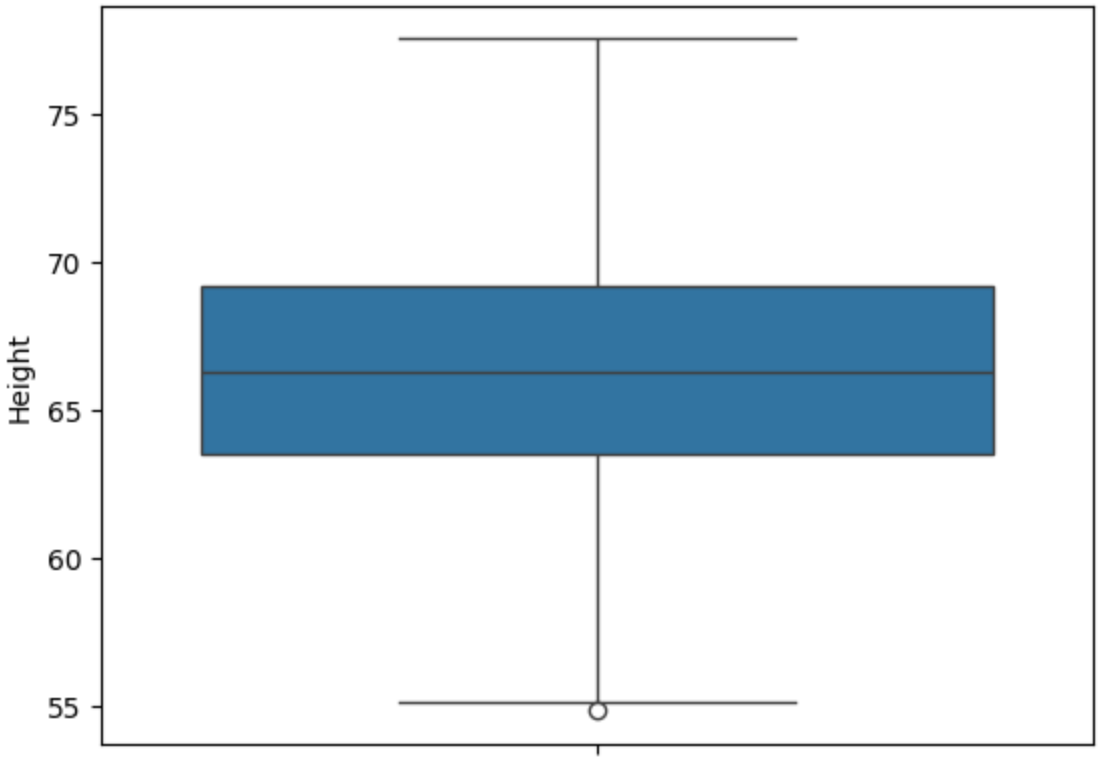
9993 rows × 3 columns

In [ ]:
```
new_df.shape
```

Out[ ]:
```
(9993, 3)
```

In [ ]:
```
sns.boxplot(new_df['Height'])
```

Out[ ]:
```
<Axes: ylabel='Height'>
```



In [58]:
```
#Z-score Method
df["z_score"]=(df['Height']-df['Height'].mean()) / (df['Height'].std())
df.head()
```

Out[58]:

	Gender	Height	Weight	z_score
0	Male	73.847017	241.893563	1.943964
1	Male	68.781904	162.310473	0.627505
2	Male	74.110105	212.740856	2.012343
3	Male	71.730978	220.042470	1.393991
4	Male	69.881796	206.349801	0.913375

In [61]:
```
df[df['z_score']>3]
```

Out[61]:

	Gender	Height	Weight	z_score
994	Male	78.095867	255.690835	3.048271
1317	Male	78.462053	227.342565	3.143445
2014	Male	78.998742	269.989699	3.282934
3285	Male	78.528210	253.889004	3.160640
3757	Male	78.621374	245.733783	3.184854

In [62]:
```
df[df['z_score']<-3]
```

Out[62]:

	Gender	Height	Weight	z_score
6624	Female	54.616858	71.393749	-3.054091
9285	Female	54.263133	64.700127	-3.146027

In [66]:
```
df[(df['z_score']>3) | (df['z_score']<-3)]
```

Out[66]:

	Gender	Height	Weight	z_score
994	Male	78.095867	255.690835	3.048271
1317	Male	78.462053	227.342565	3.143445
2014	Male	78.998742	269.989699	3.282934
3285	Male	78.528210	253.889004	3.160640
3757	Male	78.621374	245.733783	3.184854
6624	Female	54.616858	71.393749	-3.054091
9285	Female	54.263133	64.700127	-3.146027

In [71]:
```
last_df=df[(df['z_score']<3) & (df['z_score']>-3)]
last_df
```

Out[71]:

	Gender	Height	Weight	z_score
0	Male	73.847017	241.893563	1.943964
1	Male	68.781904	162.310473	0.627505
2	Male	74.110105	212.740856	2.012343
3	Male	71.730978	220.042470	1.393991
4	Male	69.881796	206.349801	0.913375
...	...	...	...	...
9995	Female	66.172652	136.777454	-0.050658
9996	Female	67.067155	170.867906	0.181830
9997	Female	63.867992	128.475319	-0.649655
9998	Female	69.034243	163.852461	0.693090
9999	Female	61.944246	113.649103	-1.149651

9993 rows × 4 columns

In [77]:
```
Upperlimit=df['Height'].mean() + 3*df['Height'].std()
Lowerlimit=df['Height'].mean() - 3*df['Height'].std()
```

In [78]:
```
Upperlimit
```

Out[78]: 77.91014411714093

In [79]:
```
Lowerlimit
```

Out[79]: 54.82497539250156

In [86]:
```
df['Height']=np.where(
    df['Height']>Upperlimit,
    Upperlimit,
    np.where(
        df['Height']<Lowerlimit,
        Lowerlimit,
        df['Height']
    )
)
```

In [83]:
```
last_df.shape
```

Out[83]:
```
(9993, 4)
```

In [84]:
```
df['Height'].describe()
```

Out[84]:
```
count    10000.000000
mean        66.367321
std          3.846314
min         54.824975
25%         63.505620
50%         66.318070
75%         69.174262
max         77.910144
Name: Height, dtype: float64
```