# NORMAL DISTRIBUTION

In [1]:

```
from scipy.stats import norm
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns



#Generate random numbers from N(0,1)

data_normal=norm.rvs(size=10000,loc=0,scale=1)
```

In [4]:

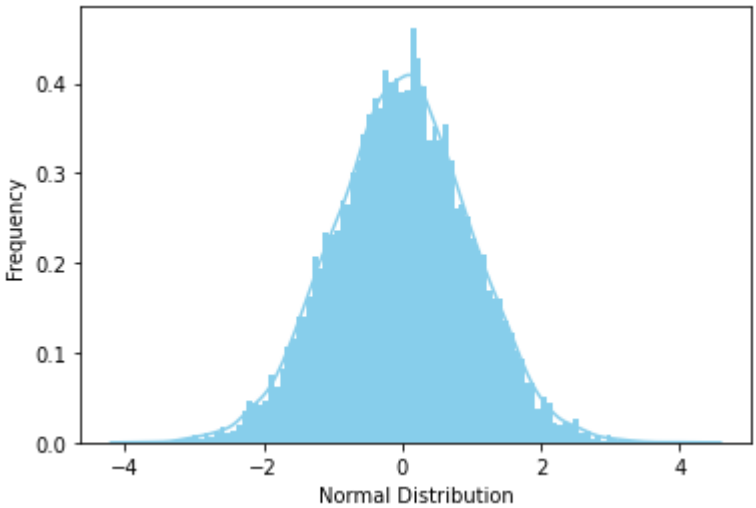
```
ax=sns.distplot(data_normal,
                bins=100,
                kde=True,
                color='skyblue',
                hist_kws={'linewidth': 15,'alpha':1})

ax.set(xlabel='Normal Distribution',ylabel='Frequency')
```

```
C:\Users\MSCIT\anaconda3\lib\site-packages\seaborn\distributions.py:2619: Fu
tureWarning: `distplot` is a deprecated function and will be removed in a fu
ture version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

Out[4]:

```
[Text(0.5, 0, 'Normal Distribution'), Text(0, 0.5, 'Frequency')]
```



In [6]:

```
df=pd.read_csv('weight-height.csv')
df
```

Out[6]:

| | Gender | Height |
|---|---|---|
| **0** | Male | 73.847017 |
| **1** | Male | 68.781904 |
| **2** | Male | 74.110105 |
| **3** | Male | 71.730978 |
| **4** | Male | 69.881796 |
| **...** | ... | ... |
| **9995** | Female | 66.172652 |
| **9996** | Female | 67.067155 |
| **9997** | Female | 63.867992 |
| **9998** | Female | 69.034243 |
| **9999** | Female | 61.944246 |

10000 rows × 2 columns

In [7]:

```
df=pd.read_csv('weight-height.csv')
df.head(5)
```

Out[7]:

| | Gender | Height |
|---|---|---|
| **0** | Male | 73.847017 |
| **1** | Male | 68.781904 |
| **2** | Male | 74.110105 |
| **3** | Male | 71.730978 |
| **4** | Male | 69.881796 |

In [9]:

```
df.Height.describe()
```

Out[9]:

```
count    10000.000000
mean        66.367560
std          3.847528
min         54.263133
25%         63.505620
50%         66.318070
75%         69.174262
max         78.998742
Name: Height, dtype: float64
```

In [10]:

```
df.describe()
```

Out[10]:

| | Height |
|---|---|
| **count** | 10000.000000 |
| **mean** | 66.367560 |
| **std** | 3.847528 |
| **min** | 54.263133 |
| **25%** | 63.505620 |
| **50%** | 66.318070 |
| **75%** | 69.174262 |
| **max** | 78.998742 |

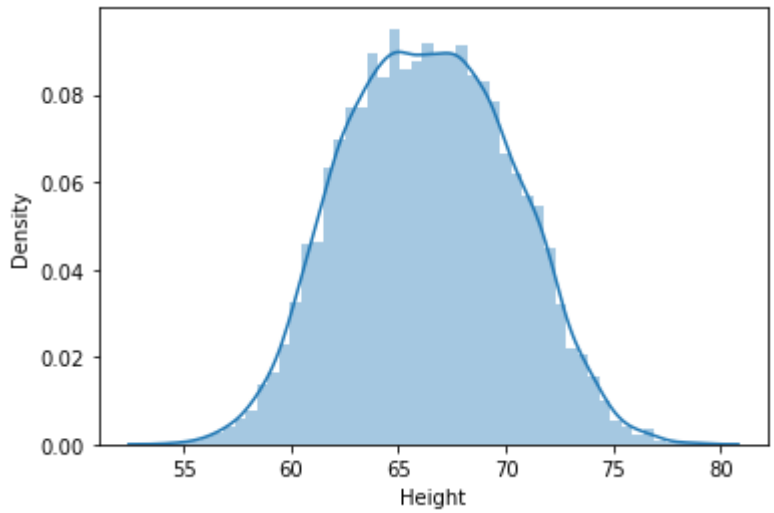
In [11]:

```
#to verify it is normal distribution or not
sns.distplot(df.Height,kde=True)
```

```
C:\Users\MSCIT\anaconda3\lib\site-packages\seaborn\distributions.py:2619: Fu
tureWarning: `distplot` is a deprecated function and will be removed in a fu
ture version. Please adapt your code to use either `displot` (a figure-level
function with similar flexibility) or `histplot` (an axes-level function for
histograms).
  warnings.warn(msg, FutureWarning)
```

Out[11]:

```
<AxesSubplot:xlabel='Height', ylabel='Density'>
```




In [12]:

```
mean=df.Height.mean()
mean
```

Out[12]:

```
66.3675597548656
```

In [13]:

```
std_deviation=df.Height.std()
std_deviation
```

Out[13]:

```
3.847528120795573
```

In [14]:

```
mean-3*std_deviation
```

Out[14]:

```
54.824975392478876
```

In [15]:

```
mean+3*std_deviation
```

Out[15]:

```
77.91014411725232
```

In [16]:

```
df[(df.Height<54.82)|(df.Height>77.91)]
```

Out[16]:

| | Gender | Height |
|---|---|---|
| **994** | Male | 78.095867 |
| **1317** | Male | 78.462053 |
| **2014** | Male | 78.998742 |
| **3285** | Male | 78.528210 |
| **3757** | Male | 78.621374 |
| **6624** | Female | 54.616858 |
| **9285** | Female | 54.263133 |

In [17]:

```
df_no_outlier=df[(df.Height<77.91)&(df.Height>54.82)]
df_no_outlier.shape
```

Out[17]:

```
(9993, 2)
```

In [18]:

```
df['zscore']=(df.Height-df.Height.mean())/df.Height.std()
df.head(5)
```

Out[18]:

| | Gender | Height | zscore |
|---|---|---|---|
| **0** | Male | 73.847017 | 1.943964 |
| **1** | Male | 68.781904 | 0.627505 |
| **2** | Male | 74.110105 | 2.012343 |
| **3** | Male | 71.730978 | 1.393991 |
| **4** | Male | 69.881796 | 0.913375 |

In [19]:

```
df.Height.mean()
```

Out[19]:

66.3675597548656

In [20]:

```
df.Height.std()
```

Out[20]:

3.847528120795573

In [21]:

```
(73.84-66.37)/3.84 # cross verify z-score
```

Out[21]:

1.9453124999999998

In [25]:

```
df[df['zscore']>3]
```

Out[25]:

| | Gender | Height | zscore |
|---|---|---|---|
| **994** | Male | 78.095867 | 3.048271 |
| **1317** | Male | 78.462053 | 3.143445 |
| **2014** | Male | 78.998742 | 3.282934 |
| **3285** | Male | 78.528210 | 3.160640 |
| **3757** | Male | 78.621374 | 3.184854 |

In [24]:

```
df[df['zscore']<-3]
```

Out[24]:

| | Gender | Height | zscore |
|---|---|---|---|
| **6624** | Female | 54.616858 | -3.054091 |
| **9285** | Female | 54.263133 | -3.146027 |

In [ ]:

```

```