

Normal Probability Plots

About

In Section 5 of Unit 6, you have seen how to construct a probability plot to check whether a normal distribution is a plausible model for the variation in a dataset.

Computer activity 9

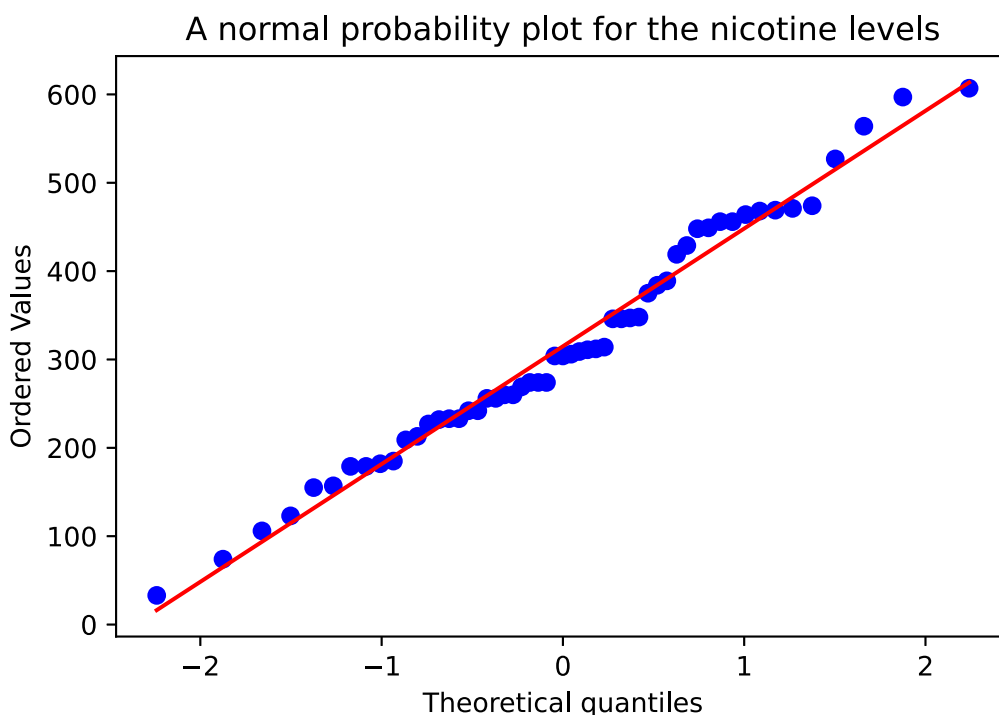
```
from scipy import stats
import pandas as pd
import matplotlib.pyplot as plt
```

Data on the blood plasma nicotine levels of 55 smokers are contained in the worksheet *plasma.mwx*. Obtain a normal probability plot for these data. Is a normal distribution a plausible model for the variation in blood plasma nicotine levels?

```
# import the dataset
data = pd.read_csv("plasma.csv")
```

We will use `scipy.stats.probplot` to produce the Normal probability plot. (See [documentation](#)).

```
fig = plt.figure()
ax = fig.add_subplot()
res = stats.probplot(data["Level"], plot=ax)
ax.set_title("A normal probability plot for the nicotine levels")
plt.show()
```



Notes

The function returns a tuple of arrays. The data used to produce the plot is the 0th element of the tuple

The function differs in some ways to the plot produced by Minitab

- The x and y axis are flipped, compared to how they are presented in **M248**
- `probp1ot()` uses least-squares regression (best-fit) to produce the regression line

These differences are tolerable, given that the method:

- has the same utility
 - The unit judges appropriateness by eye
- removes a need to produce and maintain a bespoke function.