Python Projects

On IPE-205

Developed by: Tanmoy Das

Table of Contents

[Exploratory Data Analysis (EDA) 3](#_Toc527915532)

[Probability Distribution 4](#_Toc527915533)

[Correlation & Regression 5](#_Toc527915534)

[ANOVA 6](#_Toc527915535)

[Design of Experiment 7](#_Toc527915536)

# Exploratory Data Analysis (EDA)

# Probability Distribution

## Normal Distribution

### Math Problem 1

|  |  |
| --- | --- |
| Suppose the lengths of telephone calls form a normal distribution with a mean length of 8.0 min and a standard deviation of 2.5 min. The probability that a telephone call selected at random will last more than 15.5 min is most nearly: | |
| **Python Code:**  import scipy.stats  mean\_normal = 8  standard\_deviation\_normal = 2.5  probability\_norm\_gt = scipy.stats.norm.sf(15.5, mean\_normal,standard\_deviation\_normal)  print(probability\_norm\_gt)  # SF = Survival Function (1-CDF)  **Output:**  0.0013 | **Explanation:**  When ***more than*** condition is involved, we apply survival function, sf, from Scipy package  <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html> |
|  | |

|  |  |
| --- | --- |
|  | |
| **Python Code:**  **Output:** | **Explanation:** |
|  | |

### Poison Distribution

|  |  |
| --- | --- |
| Assume baggage is rarely lost by Delta Airlines. Most flights do not experience any mishandled bags; some have one bag lost; a few have two bags lost; rarely a flight will have three lost bags; and so on. Suppose a random sample of 1,000 flights shows a total of 300 bags were lost. Determine the probability of losing no bag. Source: P 208, Chapter 6, Lind | |
| **Python Code:**  import scipy.stats mean\_poisson = 300/1000 *# prob = poisson.cdf(x, mu); x= random variable; mu = mean* probability\_poisson = scipy.stats.poisson.cdf(0, mean\_poisson)  **Output:** | **Explanation:** |
|  | |

### Exponential Distribution

|  |  |
| --- | --- |
|  | |
| **Python Code:**  **Output:** | **Explanation:** |
|  | |

### Uniform Distribution

|  |  |
| --- | --- |
|  | |
| **Python Code:**  **Output:** | **Explanation:** |
|  | |

More example can be found on [Github](https://github.com/tanmoyie/Applied-Statistics/blob/master/Python%20Projects%20%26%20Dataset%20on%20Applied%20Statistics/Probability%20Distribution%20related%20problems/Distribution%20related%20problems%20and%20their%20solutions%20_%20Kaggle.pdf) or Kaggle [[1]](#footnote-1)

# Correlation & Regression

## Simple Linear Regression

## Multiple Linear Regression

# ANOVA

## One way ANOVA

## Two way ANOVA

# Design of Experiment

# Predictive Data Analysis

## Fit the MODEL to the DATA

1. Distribution Related problems & their solutions.pdf [↑](#footnote-ref-1)