

Risk and Decision-Making for Data Science and AI

Week 4 Lab 3 Exercise

Question 1

Team X:

Average Completion Time for Team X:

Mean = $9+11+10+12+10+11+11+10+12+10$ / $10 = 11.6$ weeks

Median Completion Time for Team X:

- First, sort the completion times in ascending order: [9, 10, 10, 10, 11, 11, 11, 12, 12]
- The median is the average of the 5th and 6th values Median = 11 weeks

Team X exhibits a normal distribution with the mean and median close to each other, so both metrics are appropriate for presenting the central tendency.

Team Y:

Average Completion Time for Team Y:

Mean = $8+9+10+11+12+13+14+15+16+30$ / $10 = 13.8$ weeks

Median Completion Time for Team Y:

- First, sort the completion times in ascending order: [8, 9, 10, 11, 12, 13, 14, 15, 16, 30]
- The median is the average of the 5th and 6th values Median = 12.5 weeks

Team Y shows a skewed distribution, with the mean influenced by the presence of an outlier, so median is more appropriate for presenting the central tendency.

Question 2

Linear regression is a statistical method that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

Question 3

a) Chronic Absenteeism = $21 - 0.54 \times \text{Average Temperature}$

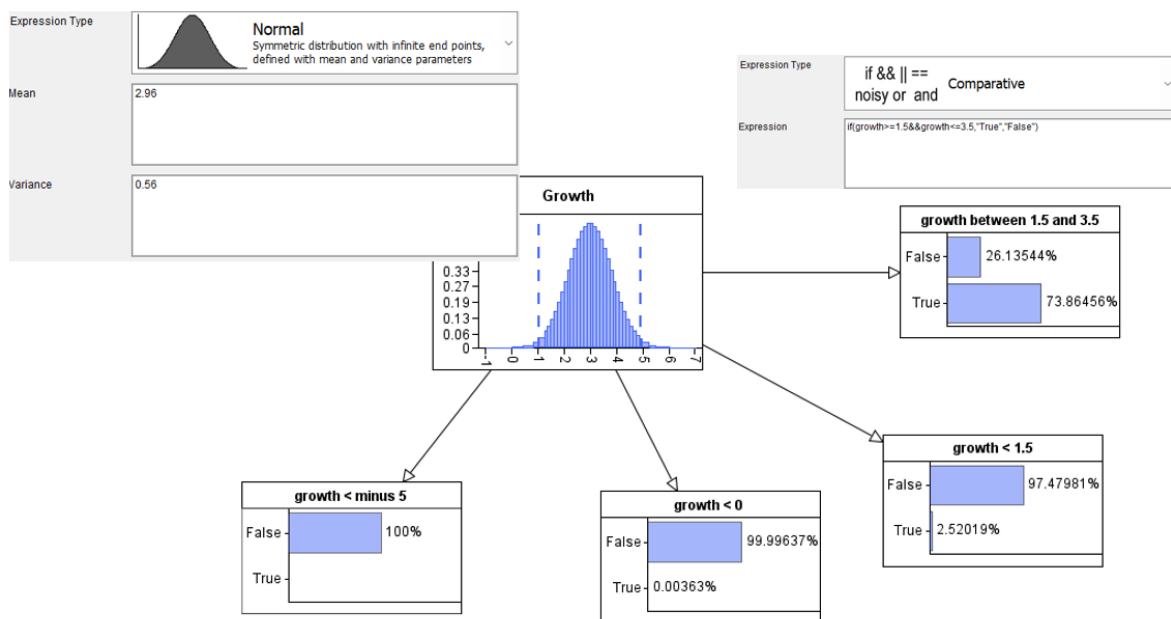
The average chronic absenteeism is 21%. For any 1C increase in temperature the average chronic absenteeism is decreased 0.54%, holding all other factors constant.

Risk and Decision-Making for Data Science and AI

Week 4 Lab 3 Exercise

b) On the surface, it seems as though there might be a slight correlation between average monthly temperature and chronic absenteeism according to the p-value metric. However, there are far more likely explanations for this trend. Winter temperatures do not necessarily cause students to mysteriously not show up to school. Perhaps the winter weather causes road blockages that parents feel uncomfortable navigating, or that in winter months illness is more common among the student body. It is even likely a contributing reason is that more engaging activities are provided to students later in the school year, resulting in increased morale and incentive to attend in April/May.

Question 4:



Using R:

a)

```
> pnorm(3.5,2.96,0.75) - pnorm(1.5,2.96,0.75)
```

```
[1] 0.7384501
```

b)

```
> pnorm(1.5,2.96,0.75)
```

```
[1] 0.02578736
```

c)

Risk and Decision-Making for Data Science and AI

Week 4 Lab 3 Exercise

```
> pnorm(0,2.96,0.75)
```

```
[1] 3.962338e-05
```

d)

```
> pnorm(-5,2.96,0.75)
```

```
[1] 1.29182e-26
```

Using Python:

```
[2] ✓ 1.6s Python
from scipy.stats import norm

[3] ✓ 0.0s Python
prob_3_5 = norm.cdf(3.5, loc=2.96, scale=0.75)
prob_1_5 = norm.cdf(1.5, loc=2.96, scale=0.75)
probability = prob_3_5 - prob_1_5
print(probability)
... 0.7384501470381348

[4] ✓ 0.0s Python
probability = norm.cdf(1.5, loc=2.96, scale=0.75)
print(probability)
... 0.025787355182614012

[5] ✓ 0.0s Python
probability = norm.cdf(0, loc=2.96, scale=0.75)
print(probability)
... 3.9623381955842555e-05

[6] ✓ 0.0s Python
probability = norm.cdf(-5, loc=2.96, scale=0.75)
print(probability)
... 1.2918197478485516e-26
```