**Probability and Statistics**

Project 2



|  |  |
| --- | --- |
| 학번 | 2016025469 |
| 이름 | 서건식 |
| 수업 | 확률통계론 |
| 교수님 | 배완덕 교수님 |
| 제출일 | 2017.05.23 |

1. Question studied

First, I state the population and variable.

**The population**: Wooden beams produced a Canadian company

**The variable**: Bending strength of wooden beams

1. the population of interest

I think I am more likely to believe the customers reviews. The reason is below at sample data distribution. At that part, I state the reason. By the way, my questions are as follows:

Q1) The average bending strength of the wooden beams produced by the company is below 50N/mm2.

Q2) The average bending strength of wooden beams produced in Eastern Canada is higher than that of wooden beams produced in Western Canada.

Q3) The variance of bending strength of wooden beams produced in Eastern Canada is greater than that of wooden beams produced in Western Canada.

1. variable of interest

Bending strength of wooden beams.

1. Sample datesets

I use 4 datasets as follows:

• douglas data 50.csv - a sample of size 50 from the population of wooden beams produced in 360 locations, in Eastern and Western Canada.

• douglas data 20.csv - a sample of size 20 from the population of wooden beams produced in 360 locations, in Eastern and Western Canada.

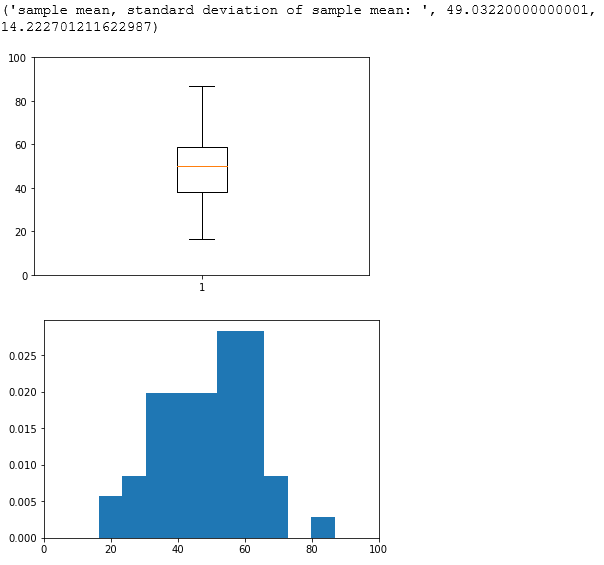
• douglas eastern data 42.csv - a sample of size 42 from the population of wooden beams produced in 210 locations in Eastern Canada.

• douglas western data 30.csv - a sample of size 30 (from the population of wooden beams produced in 150 locations in Western Canada.

1. Sample data distribution

I provide result of sample data using python.

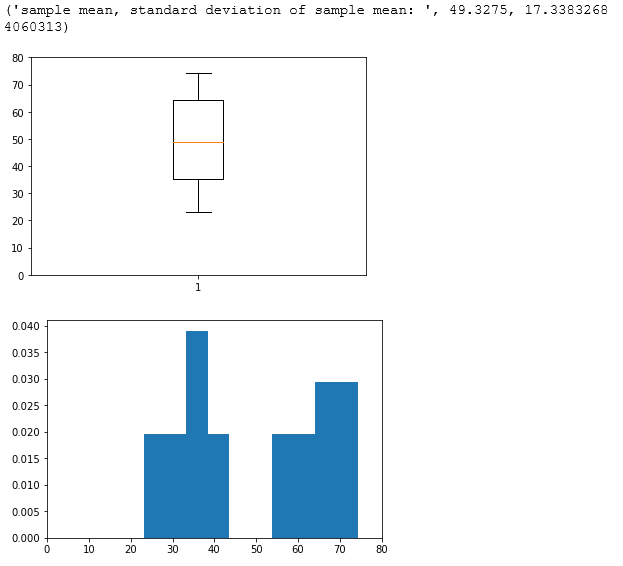
1. In douglas data 50.csv
2. Graphical analysis of the sample data & sample mean, standard deviation of sample mean



1. discussion of the shape of the data

This shape of data is almost left skewed. And has no outlier. So if the data have no outlier, the data is not symmetrical.

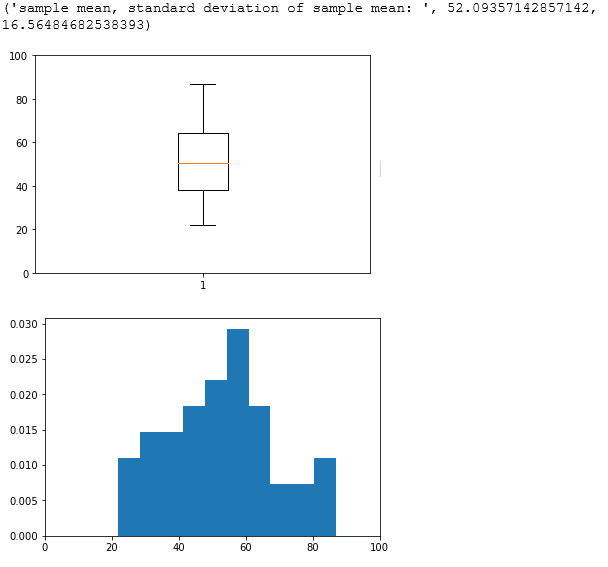
1. In douglas data 20.csv
2. Graphical analysis of the sample data & sample mean, standard deviation of sample mean



1. discussion of the shape of the data

I think this shape of the graph can’t provide enough meaningful statistical value. I think because of the sample size. Sample size is too small. And this do not have any skewness or outliers present.

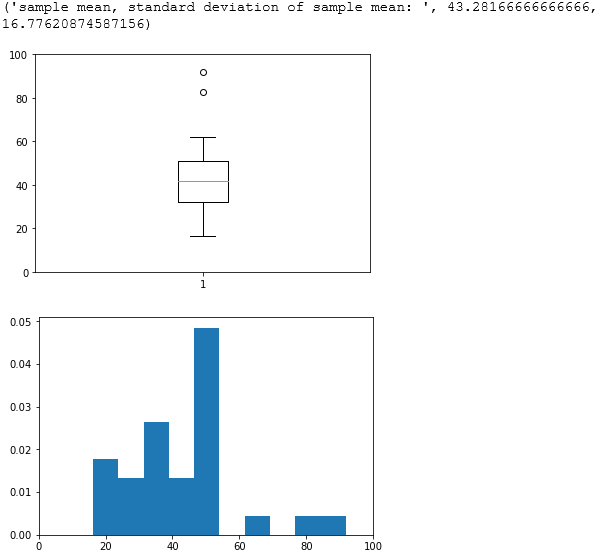
1. In douglas eastern data 42
2. Graphical analysis of the sample data & sample mean, standard deviation of sample mean



1. discussion of the shape of the data

this shape is almost symmetric. So, I think in eastern Canada distribution of bending strength is almost normal distribution.

1. In douglas western data 30
2. Graphical analysis of the sample data & sample mean, standard deviation of sample mean



1. discussion of the shape of the data

일단 남김.

1. The reason I choose the customer’s claim.

In douglas eastern data 42, the shape of data is almost symmetric. So, In eastern Canada, distribution of bending strength is almost normal distribution. It is matched with the company’s claim. But in douglas western data 30, almost data is below 50N/mm2 I think the reason most of reviews is written from western Canada is in here. So I think I am more likely to believe the customers reviews.

1. Hypotheses test results and my findings

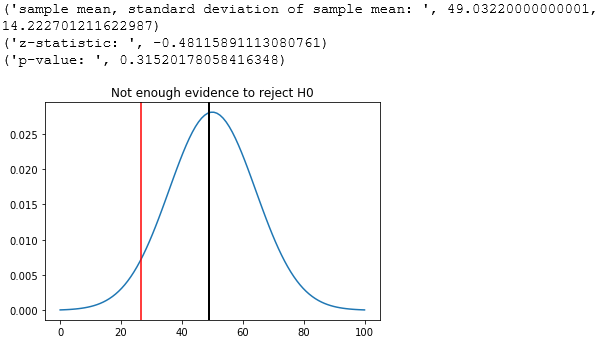
In this project, we assume that the population is normally distributed and you have no way of knowing the population standard deviation. So, Let’s perform a hypothesis test on the given sample dataset.

1. The significant level: α = 0.05
2. hypothesis test
3. Q1: Is The average bending strength of the wooden beams produced by the company below 50N/mm2?

**null hypothesis(H0)**: μ = 50N/mm2

**alternative hypothesis(Ha)**: μ < 50N/mm2

>> z-test on large sample data (douglas data 50.csv)

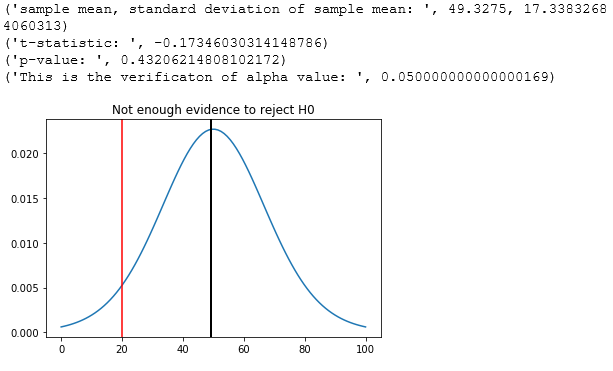


In this picture, I state sample mean, standard deviation of sample mean, p-value, z-statistic. In this dataset, p-value is larger than significant level α. So, I can’t have enough evidence to reject H0.

* What this means about this population

Consumer dissatisfaction is mostly in western Canada. But This sample dataset is from the population of wooden beams produced in 360 locations, in Eastern and Western Canada. Hence in this case, H0 can be true as the company has said.

>> t-test on small sample data (douglas data 20.csv)



In this picture, I state sample mean, standard deviation of sample mean, p-value, t-statistic. In this dataset, p-value is larger than significant level α. So, I can’t have enough evidence to reject H0.

* What this means about this population

I think It will be same result with z-test on large sample data. Because this dataset is also from the population of wooden beams produced in 360 locations, in Eastern and Western Canada. Hence in this case, H0 can be true as the company has said.

🡪Conclusion of Q1:

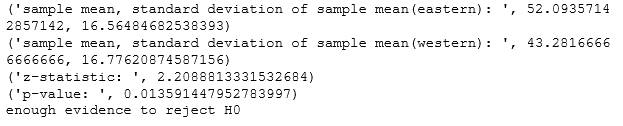
There is no evidence that The average bending strength of the wooden beams produced by the company is below 50N/mm2. I think this result is reasonable. But It is important to note that most of the complaints appeared in western Canada. Maybe the company’s announcement use this result, and may say there is no problem. So next part, We should make sure that there is a difference in the bending strength between the West and the East Canada. If Or not, customer’s claim is not reasonable.

1. Q2: Is The average bending strength of wooden beams produced in Eastern Canada higher than that of wooden beams produced in Western Canada? ( To find the diﬀerence between the averages of bending strength of Eastern wooden beams and Western wooden beams)

**null hypothesis(H0)**: μ1 = μ2

**alternative hypothesis(Ha)**: μ1 > μ2

>> z-test on sample datasets from two populations between douglas eastern data 42 and douglas western data 30



In this picture, I state sample mean, standard deviation of sample mean, p-value, z-statistic each datasets. In this case, p-value is smaller than significant level α. So, It has enough evidence to reject H0.

* What this means about this population

Rejecting H0 is not mean that the probability of accepting Ha is 95%, but if H0 is true and all [hypothesis](http://endic.naver.com/search.nhn?query=hypothesis) are effective, the probability of getting same result with current result is 5%. So, It has enough evidence to accept Ha.

🡪Conclusion of Q2

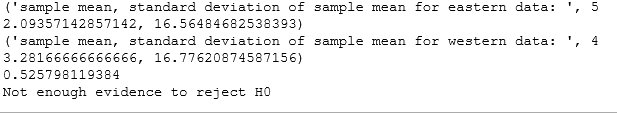
This means that the probability that the average of the bending strength in eastern Canada is higher than in western Canada is higher. So in this case, customer’s claim is reasonable.

1. Q3: Is The variance of bending strength of wooden beams produced in Eastern Canada greater than that of wooden beams produced in Western Canada? ( To find the diﬀerence between the variances of bending strength of Eastern wooden beams and Western wooden beams.)

**null hypothesis(H0)**: σ12 = σ22

**alternative hypothesis(Ha)**: σ12 > σ22

>> F-test on sample datasets from two populations between douglas eastern data 42 and douglas western data 30.



In this picture, I state sample mean, standard deviation of sample mean, p-value, each datasets. In this case, p-value is larger than significant level α. So, I can’t have enough evidence to reject H0

* What this means about this population

This means There is insufficient evidence to suggest a difference in standard deviation.

🡪Conclusion of Q2

In this case, It does not have enough evidence that there is difference in standard deviation between two population. So, the customer’s claim can’t be reasonable.