Chapter 4. Testing on Interval Data (II)

4.1. Fuzzy Normal distribution

Given n sample data $I_i = [a_i, b_i]$, $i = 1, 2, \cdots, n$, we firstly test two sets of crisp data $\boldsymbol{a_1, a_2, ..., a_n}$ and $\boldsymbol{b_1, b_2, ..., b_n}$ to see what distribution they are. In this chapter we also assume they fit normal. Under this situation we say the population is fuzzy normally distributed with the same idea as the above Chapter discussion.

4.2. Method of Computing Sample Mean and sample variance for Interval Data

Given an interval $I = [X, Y]$, X and Y are independent normally distributed random variables, then we say I is an *interval-valued* normally distributed random variable.

4.2.1. Method of Computing Sample Mean in Interval Form

Let's take the desire of a job seeker as an example first. The boss asks a job seeker how much you would like for your salary. Suppose the answer is between 20000 to 22000 dollars, it means the job seeker will be satisfied if his salary is between 20000 to 22000. Let $I_i = [a_i, b_i]$, $i = 1, 2, \cdots, n$ be a set of sample interval data. Therefore, an acceptable sample mean of n sample data I_i, $i = 1, 2, \cdots, n$ can be chosen from a value of $\frac{1}{n}\sum_{i=1}^{n} x_i$, where $x_i \in I_i$. It is obvious that the minimum and the maximum of $\frac{1}{n}\sum_{i=1}^{n} x_i$ are given as $\frac{1}{n}\sum_{i=1}^{n} a_i$ and $\frac{1}{n}\sum_{i=1}^{n} b_i$, respectively. Hence the sample mean of I_i, $i = 1, 2, \cdots, n$ can be intuitively defined by using the minimum and maximum values of $\frac{1}{n}\sum_{i=1}^{n} x_i$.

Definition 4.2.1. Interval Sample Mean

Given a set of interval sample data I_i , $i = 1, 2, \cdots, n$ *where* $I_i = [a_i, b_i]$. *The interval sample mean* $\overline{X}_I$ *is defined as* $\boldsymbol{\overline{X}_I = [\frac{1}{n}\sum_{i=1}^{n} a_i, \frac{1}{n}\sum_{i=1}^{n} b_i]}$.

4.2.2. Method of Computing Sample Variance of Interval Data

Given a set of sample interval data $I_i = [a_i, b_i]$, $i = 1, 2, \cdots, n$. By choosing a point x_i from I_i, it is easy to compute the conventional sample mean and sample variance, namely

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Hence, an acceptable sample variance of I_i , $i = 1, 2, \cdots, n$ may be chosen from a value of $\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ where $x_i \in I_i$ and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. By the same argument the interval variance can be defined as the minimum and maximum value of s^2 .

Definition 4.2.2. Interval Sample Variance

Gives a set of interval sample data I_i , $i = 1, 2, \cdots, n$ *where* $I_i = [a_i, b_i]$. *The interval sample variance* $S_I^2 = [s_1^2, s_2^2]$ *is defined as*

$$s_1^2 = \min_{x_i \in I_i, \forall i}\left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\right\}, \quad s_2^2 = \max_{x_i \in I_i, \forall i}\left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\right\}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

It is not easy to compute the interval variance $S_I^2 = [s_1^2, s_2^2]$. Finding the minimum and maximum value of s^2 can be recast as the following *quadratic programming* problems:

$$s_1^2 = \min \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$s_2^2 = \max \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\text{s.t.} \qquad a_i \le x_i \le b_i \qquad i = 1, 2, \cdots, n$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

These problems can be easily solved by any bookshelf operation research software package. However, some special case may be computed by hand. We give several theorems for computing interval sample variance.

Lemma 4.2.1. Let $f(x, y) = (x - \frac{x+y}{2})^2 + (y - \frac{x+y}{2})^2$ such that $x \in [x_m, x_M]$ and $y \in [y_m, y_M]$, then the extreme values happen at $x = y$ or at boundary points. (Note that minimum value of $f(x, y)$ is zero when $x = y$ and $[x_m, x_M]$, $[y_m, y_M]$ are not disjoint.)

Proof: Let $\frac{\partial f}{\partial x} = 0$ and $\frac{\partial f}{\partial y} = 0$, we have

$$2(x - \frac{x+y}{2})(1 - \frac{1}{2}) + 2(y - \frac{x+y}{2})(-\frac{1}{2}) = 0 \qquad (1)$$

$$2(x-\frac{x+y}{2})(-\frac{1}{2})+2(y-\frac{x+y}{2})(1-\frac{1}{2})=0 \qquad (2)$$

Solving for (1) and (2), we get $x=y$. Obviously, we have the minimum value when $x=y$. The other extreme value will be happened at the boundary points. ❑

Remark: If $[x_m, x_M]$ and $[y_m, y_M]$ are disjointed, i.e. $x_m \le x_M \le y_m \le y_M$, then minimal value of $f(x,y)$ is $f(x_M, y_m)$ and maximal value of $f(x,y)$ is $f(x_m, y_M)$. It is easily verified as follows:

$$(x_M-\frac{x_M+y_m}{2})^2+(y_m-\frac{x_M+y_m}{2})^2=\frac{(x_M-y_m)^2}{2}\le\frac{(x-y)^2}{2}\le\frac{(x_m-y_M)^2}{2}$$

where $x_m \le x \le x_M$ and $y_m \le y \le y_M$.

Similarly, we can easily extend Lemma 5.2.1 to consider a function with n variables as follows:

Lemma 4.2.2. Let $f(x_1, x_2, ..., x_n)=\sum_{i=1}^{n}(x_i-\bar{x})^2$ where $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ and $x_i \in [a_i, b_i]$ then the extreme values occur at $x_1=x_2=...=x_n$ and boundary point.

Next we would like to solve the following problem: How to find the extreme value of $\sum_{i=1}^{n}(x_i-\bar{x})^2$ subject to $x_i \in [a_i, b_i]$, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$, $i=1,2,\cdots,n$. Also we note that it is clear to get minimum if $x_1=x_2=...=x_n$. Let $x_0=m$ and $x_{n+1}=M$ where $m \le M$. Firstly we have the following theorems.

Theorem 4.2.1. Let a real value function $f(x_1, x_2, \cdots, x_n)$ defined as

$$f(x_1, x_2, \cdots, x_n) = \sum_{i=0}^{n+1} (x_i - \hat{x})^2$$

where $\hat{x} = \dfrac{m + x_1 + \cdots + x_n + M}{n+2}$. Then f is a convex function.

Proof: The Hessian matrix H of this function is given as

$$H = \begin{bmatrix} \frac{2n+2}{n+2} & \frac{-2}{n+2} & \cdot & \cdot & \frac{-2}{n+2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{-2}{n+2} & \cdot & \cdot & \cdot & \frac{2n+2}{n+2} \end{bmatrix}$$

where the diagonal entries of H are $\dfrac{2n+2}{n+2}$ and the off-diagonal entries of H are $\dfrac{-2}{n+2}$. Since $\dfrac{2n+2}{n+2} > 0$ and $\Delta_i = \dfrac{2^i(n-i+2)}{n+2} > 0$, where

$$\Delta_i = \begin{vmatrix} \frac{2n+2}{n+2} & \cdot & \cdot & \cdot & \frac{-2}{n+2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \frac{-2}{n+2} & \cdot & \cdot & \cdot & \frac{2n+2}{n+2} \end{vmatrix}$$

is an $i \times i$ determinant with $i \le n$. The diagonal entries of Δ_i are $\dfrac{2n+2}{n+2}$ and off-diagonal entries of Δ_i are $\dfrac{-2}{n+2}$. Therefore $\det H = \dfrac{2^{n+1}}{n+2} > 0$. This completes the proof. ❑

Theorem 4.2.2. Suppose $m < M$ and $x_i \in [m, M]$, $i = 1, 2, \cdots, n$ the function

$$f(x_1, x_2, \cdots, x_n) = \sum_{i=0}^{n+1} (x_i - \hat{x})^2$$

has minimal value at $x_1 = x_2 = \ldots = x_n = \dfrac{m+M}{2}$. The maximal value occurs at the point which is half of $x_1, x_2, \cdots, x_n$ taking the value of m and half of them taking the value at M.

Proof: Let $\frac{\partial f}{\partial x_i} = 0$, $i = 1, 2, \cdots, n$, then we have the following equations:

$$2(x_i - \hat{x}) - \frac{2}{n+2}\sum_{i=0}^{n+1}(x_i - \hat{x}) = 0$$

Solving these equations yields a critical point with $x_1 = x_2 = ... = x_n = \frac{m+M}{2}$. By theorem 5.2.1 f has minimal value at this critical point. Now, we consider two special cases. First, let $x_1 = x_2 = ... = x_n = m$ then we have

$$\begin{aligned} & f(m, m, \cdots, m) \\ & = (n+1)(m - \frac{(n+1)m + M}{n+2})^2 + (M - \frac{(n+1)m + M}{n+2})^2 \\ & = (n+1)\frac{(M-m)^2}{n+2} \end{aligned}$$

Second, let $x_1 = x_2 = ... = x_n = M$, then we have

$$\begin{aligned} & f(M, M, \cdots, M) \\ & = (n+1)(M - \frac{(n+1)m + M}{n+2})^2 + (m - \frac{(n+1)m + M}{n+2})^2 \\ & = (n+1)\frac{(M-m)^2}{n+2} \end{aligned}$$

Let p of $x_1, x_2, \cdots, x_n$ equal to m and q of $x_1, x_2, \cdots, x_n$ equal to M where $1 < p < n$, $1 < q < n$, and $p + q = n$. By the same argument, we have

$$\begin{aligned} f(x_1, x_2, \cdots, x_n) & = (p+1)\left[m - \frac{(p+1)m + (q+1)M}{n+2} \right]^2 \\ & \quad + (q+1)\left[M - \frac{(p+1)m + (q+1)M}{n+2} \right]^2 \\ & = (p+1)\left[\frac{(q+1)m - (q+1)M}{n+2} \right]^2 \\ & \quad + (q+1)\left[\frac{(p+1)m - (p+1)M}{n+2} \right]^2 \end{aligned}$$

$$= (p+1)(q+1)\frac{(M-m)^2}{n+2} \qquad (3)$$

It is clear, when $p+1 = q+1$, i.e. $p = \frac{n}{2}$, (3) yield the maximal value. ❑

As a matter of fact, in searching the maximal value and minimal value of

$$f(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

where $\bar{\boldsymbol{x}} = \frac{\boldsymbol{1}}{\boldsymbol{n}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{x_i}$, subject to $x_i \in [a_i, b_i]$, I = 1, 2, …, n is a quadratic programming problem which is a well known problem in operation research and can be solved by any bookshelf software. On the other hand, the above theorems show that under some easy conditions, we can calculate the minimal value and maximal value directly. For example, in computing minimal value:

Case1: If all $[a_i, b_i]$*, i=1,2,…,n have common element then the minimal value is zero.*

Case2: Looking for two points m and M, the most left point and the most right point respectively, such that they belong to distinct interval, say

$$m \in [a_1, b_1] \;\; \textit{and} \;\; M \in [a_n, b_n]$$

and as close as possible.

If the rest $n-2$ *intervals have common element* $\frac{m+M}{2}$ *then the minimal value is*

$$f(m, \frac{m+M}{2}, \frac{m+M}{2}, \cdots, \frac{m+M}{2}, M).$$

As looking for maximal value, we only consider endpoints of each interval.

4.3. Testing Hypotheses on Mean and Variance with Interval Data

In this part, we test interval-valued sample data with conventional methods, say *t*-test (for mean) and *F*-test (for variance), on two ends simultaneously. When we attempt to test population means (in interval form), we compute sample interval data to get sample mean by the above discussion at first. Then we compute sample variance of left and right crisp data individually.

The null hypothesis is set up to be equal. If both sides of testing results suggest to accept "equal" then we accept null hypothesis otherwise we reject the null hypothesis.

Similarly, we test two population means by the same way. As for testing sample variance, we use χ^2-test on population variance and *F*-test on two population variances. By the same process as the above, we test sample variance on two ends simultaneously.

4.3.1. Testing hypothesis for interval sample mean

In testing, we suppose the population is normal, otherwise nonparametric method should be used in classical statistics. In this paper, we extend traditional *t*-testing method to test data in interval form. So, we first suppose the population is fuzzy normally distributed.

Until now, there is no standard way to test interval data. We propose extended classical testing method to do this work. We simultaneously test two ends, the left end and the right end, under certain significant level. We will reject the null hypothesis if one of the two ends is rejected.

For testing mean of population of interval data, we suggest the following process:

1. *Null hypothesis*:

$$H_0: \mu_I = \mu_{I_0}$$

$$H_1: \mu_I \neq \mu_{I_0}$$

where $\mu_{I_0} = [x_0, y_0]$.

2. *Testing Statistics*: $[\bar{x}, \bar{y}]$.

3. Under the significant level $\boldsymbol{\alpha_I}$, we use two tail *t*-test for $\bar{x}$, $\bar{y}$

If

$$\bar{x} \in (x_0 - t_{n-1,\ \alpha/2} \frac{S_x}{\sqrt{n}}, x_0 + t_{n-1,\ \alpha/2} \frac{S_x}{\sqrt{n}})$$

and

$$\bar{y} \in (y_0 - t_{n-1,\ \alpha/2} \frac{S_y}{\sqrt{n}}, y_0 + t_{n-1,\ \alpha/2} \frac{S_y}{\sqrt{n}})$$

then we accept null hypothesis $\mu_I = [x_0, y_0]$, where

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

and

$$S_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

and $I_i = [x_i, y_i]$, $i = 1, 2, \cdots, n$ be the sample data randomly collected from the population.

For testing the difference of mean for the interval data $\{I_{x1}, I_{x2}, ..., I_{xn}\}$ and $\{I_{y1}, I_{y2}, ..., I_{ym}\}$ on two populations, we suggest the following process.

1. *Null hypothesis*:

$$H_0 : \mu_{I_x} = \mu_{I_y}$$

$$H_1 : \mu_{I_x} \neq \mu_{I_y}$$

where $\mu_{I_x} = [\mu_{1x}, \nu_{1y}]$ and $\mu_{I_y} = [\mu_{2x}, \nu_{2y}]$

2. *Testing Statistics*: t_l , t_r

where

$$t_l = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_{x1}^2}{n} + \frac{S_{x2}^2}{m}}} \quad \text{and} \quad t_r = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_{y1}^2}{n} + \frac{S_{y_2}^2}{m}}}$$

3. Under the significant level $\boldsymbol{\alpha_I}$, we use two tail t-test for $\mu_{I_x} = \mu_{I_y}$. If $t_l \in (-t_{k,\alpha/2},\ t_{k,\alpha/2})$ and $t_r \in (-t_{k,\alpha/2},\ t_{k,\alpha/2})$, then we accept null hypothesis $\mu_{I_x} = \mu_{I_y}$ where

$$I_{xi} = [x_i^{(1)},\ y_i^{(1)}],\quad i = 1, 2, \cdots, n$$

and

$$I_{yj} = [x_j^{(2)},\ y_j^{(2)}],\quad j = 1, 2, \cdots, m$$

as well as

$$\bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i^{(1)} ,\ \ \bar{y}_1 = \frac{1}{n}\sum_{i=1}^{n} y_i^{(1)} ,\ \ \bar{x}_2 = \frac{1}{m}\sum_{j=1}^{m} x_j^{(2)} ,\ \ \bar{y}_2 = \frac{1}{m}\sum_{j=1}^{m} y_j^{(2)} ,$$

$$S_{x1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i^{(1)} - \bar{x}_1)^2 ,\ \ S_{y1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i^{(1)} - \bar{y}_1)^2 ,$$

and

$$S_{x2}^2 = \frac{1}{m-1}\sum_{j=1}^{m}(x_j^{(2)} - \bar{x}_2)^2 ,\ \ S_{y2}^2 = \frac{1}{m-1}\sum_{j=1}^{m}(y_j^{(2)} - \bar{y}_2)^2 ,$$

with k degree of freedom, $k = \min\{n-1, m-1\}$.

4.3.2. Testing hypothesis for interval sample variance

Similarly, let us observe $1-\alpha$ confidence interval

$$(F_{n_1-1,\, n_2-1,\, 1-\alpha/2},\, F_{n_1-1,\, n_2-1,\, \alpha/2})$$

of two populations' variance σ_1^2, σ_2^2 where $F = S_1^2 \big/ S_2^2$. We accept null hypothesis $\sigma_1^2 = \sigma_2^2$ if

$$S_2^2 F_{n_1-1,\, n_2-1,\, 1-\alpha/2} < S_1^2 < S_2^2 F_{n_1-1,\, n_2-1,\, \alpha/2}\,.$$

Extends this method. We set the following rules:

For testing variance of population of interval data, we suggest the following process:

1. *Null hypothesis*:

$$H_0 : \sigma_I = \sigma_{I_0}$$

$$H_1 : \sigma_I \neq \sigma_{I_0}$$

where $\sigma_{I_0} = [p_0, q_0]$.

2. *Testing Statistics*: $S_I^2 = [S_1^2, S_2^2]$

where

$$S_1^2 = \min_{t_i \in I_i} \frac{1}{n-1} \sum_{i=1}^{n} (t_i - \bar{t})^2 ,$$

$$S_2^2 = \max_{t_i \in I_i} \frac{1}{n-1} \sum_{i=1}^{n} (t_i - \bar{t})^2$$

and

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

3. Under the significant level $\boldsymbol{\alpha_I}$, we use two tail χ^2-test for S_1^2, S_2^2. If

$$\frac{(n-1)S_1^2}{p_0^2} \in (\chi^2_{n-1,1-\alpha/2}, \chi^2_{n-1,\alpha/2})$$

and

$$\frac{(n-1)S_2^2}{p_0^2} \in (\chi^2_{n-1,1-\alpha/2}, \chi^2_{n-1,\alpha/2})$$

then we accept the null hypothesis $H_0 : \sigma_I = \sigma_{I_0}$.

For testing the difference of variance for the interval data $\{I_{x1}, I_{x2}, ..., I_{xn}\}$ and $\{I_{y1}, I_{y2}, ..., I_{ym}\}$ on two populations, we suggest the following process:

1. *Null hypothesis*:

$$H_0 : \sigma_{I_x} = \sigma_{I_y}$$

$$H_0 : \sigma_{I_x} \neq \sigma_{I_y}$$

where $\sigma_{I_x} = [p_x, q_x]$ and $\sigma_{I_y} = [p_y, q_y]$.

2. *Testing Statistics*: $F_l = \dfrac{S_{1x}^2}{S_{1y}^2}$ and $F_r = \dfrac{S_{2x}^2}{S_{2y}^2}$

where

$$S_{1x}^2 = \min_{t_i \in I_{xi}} \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{t})^2 , \quad S_{2x}^2 = \max_{t_i \in I_{xi}} \frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{t})^2 ,$$

and

$$\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

as well as

$$S_{1y}^2 = \min_{t_j \in I_{yj}} \frac{1}{m-1}\sum_{j=1}^{m}(t_j - \bar{t})^2 , \quad S_{2y}^2 = \max_{t_j \in I_{yj}} \frac{1}{m-1}\sum_{j=1}^{m}(t_j - \bar{t})^2$$

and

$$\bar{t} = \frac{1}{m}\sum_{j=1}^{m} t_j$$

3. Under the significant level $\boldsymbol{\alpha_I}$, we use two tail F-test for $F_l = \frac{S_{1x}^2}{S_{1y}^2}$ and $F_r = \frac{S_{2x}^2}{S_{2y}^2}$. If

$$F_l \in (F_{n-1,\, m-1,\, 1-\alpha/2},\quad F_{n-1,\, m-1,\, \alpha/2})$$

and

$$F_r \in (F_{n-1,\, m-1,\, 1-\alpha/2},\ F_{n-1,\, m-1,\, \alpha/2})$$

then we accept the null hypothesis $H_0 : \sigma_{I_x} = \sigma_{I_y}$.

4.4. Empirical studies

Here we give some examples for testing hypothesis under the above assumption.

Example 4.4.1. Government assumes the demanding salary of a fresh college graduate is [1000, 6000] (unit: dollars) in average per month.

$$H_0 : \boldsymbol{\mu_I} = \mathbf{[1000, 6000]} \text{ and } H_1 : \boldsymbol{\mu_I} \neq \mathbf{[1000, 6000]}.$$

Under the significant level $\boldsymbol{\alpha_I}$ = 0.05 and suppose the population is interval normally distributed. We randomly collect 25 persons who just graduated from college and survey their demanding for salary. After simple calculation we obtain sample mean is [1350,1850] by the above method. Now,

$$\bar{x} = 1350,\quad \bar{y} = 1850$$

and

$$x_0 = 1000,\quad y_0 = 6000,$$

as well as

$$S_x^2 = \frac{1}{24}\sum_{i=1}^{25}(x_i - \bar{x})^2 = 9, \quad S_y^2 = \frac{1}{24}\sum_{i=1}^{25}(y_i - \bar{y})^2 = 16 .$$

Since

$$\bar{x} \notin (1000 - 2.06\frac{3}{\sqrt{25}}, 1000 + 2.06\frac{3}{\sqrt{25}}) \ = (998.76, 1001.24)$$

and

$$\bar{y} \notin (6000 - 2.06\frac{3}{\sqrt{25}}, 6000 + 2.06\frac{3}{\sqrt{25}}) \ = (5998.76, 6001.24),$$

hence we reject null hypothesis $\mu_I = [x_0, y_0]$. That is, the government's claim is not true.

Example 4.4.2. The manger of Brand **A** car claims that their car can drive 9-11 miles per liter in average at highway. One consumer magazine is requested to check this data. They randomly choose 36 persons and ask them to drive this car six times at highway. Then they record the miles after driving at the highway. Finally, they make an interval data with the minimum mile and maximum mile on the left and right hand side of points respectively. For example, if the record data is

9, 9.1, 9.5, 11, 10.2, 10.1

miles each liter, then we get an interval data [9, 11]. Under the significant level $\alpha = 0.05$ and suppose the population is interval normally distributed. By using the above method we compute the sample mean and the sample variance for the 36 persons' experiment as follows:

$$[\bar{x}, \bar{y}] = [8, 12], \quad S_I^2 = [12, 28]$$

We are performing the following test:

$$H_0 : \ \mu_I = [9, 11]$$

$H_1: \ \mu_I \neq [9,11]$

Now, we have

$$\bar{x} = 8 , \quad \bar{y} = 12$$

$$x_0 = 9 , \quad y_0 = 11$$

and

$$S_x^2 = \frac{1}{35}\sum_{i=1}^{36}(x_i - \bar{x})^2 = 100 , \quad \boldsymbol{S_y^2 = \frac{1}{35}\sum_{i=1}^{36}(y_i - \bar{y})^2 = 144}$$

Since

$$\bar{x} \in (9 - 2.03\frac{10}{\sqrt{36}}, \ 9 + 2.03\frac{10}{\sqrt{36}}) \ = (7.78, 10.22)$$

and

$$\bar{y} \in (11 - 2.03\frac{12}{\sqrt{36}}, \ 11 + 2.03\frac{12}{\sqrt{36}}) \ = (6.94, 15.06)$$

hence we accept the null hypothesis $\mu_I = [9,11]$. That is, the manger's claim is true. On the other hand, the sample variance = [80, 200], which is quite large comparing to sample mean. Hence we feel that the quality of Brand **A** car is very unstable. Therefore, we will give a negative suggestion for buying the car.

Example 4.4.3. There are two communities *X* and *Y*, we want to compare their income level to set up a sale strategy. Suppose we choose 36 data randomly from *X* and *Y* community and assume those data comes from two interval normally l distributed populations. After simple calculation, we get their sample means

$$\bar{x}_I = [3.7, \ 4.3] , \quad \bar{y}_I = [3.7, \ 5.1]$$

and their sample variances

$$S_{x1}^2 = 0 , \quad S_{y1}^2 = 0.52 , \quad S_{x2}^2 = 3.2 , \quad S_{y2}^2 = 10$$

The test is as follows:

$H_0: \ \mu_{I_x} = \mu_{I_y}$

$H_1: \ \mu_{I_x} \neq \mu_{I_y}$

where $\mu_{I_x} = [\mu_{1x}, \ \nu_{1y}]$, $\mu_{I_y} = [\mu_{2x}, \ \nu_{2y}]$, and $\mu_{I_y} = [x_2, \ y_2]$.

The testing statistics t_l and t_r are obtained as follows:

$$t_l = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_{x1}^2}{n} + \frac{S_{x2}^2}{m}}} = 0, \quad t_r = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{S_{y1}^2}{n} + \frac{S_{y2}^2}{m}}} = -1.5$$

Under the significant level $\boldsymbol{\alpha_I = 0.05}$, we use two tail *t*-test for $\mu_{I_x} = \mu_{I_y}$. Since $t_l \in (-2.03, 2.03)$ and $t_r \in (-2.03, 2.03)$, then we accept null hypothesis $\mu_{I_x} = \mu_{I_y}$.

Next we test the Null hypothesis

$$H_0: \sigma_{I_x} = \sigma_{I_y}$$

$$H_1: \sigma_{I_x} \neq \sigma_{I_y}$$

where $\sigma_{I_x} = [p_x, q_x]$ and $\sigma_{I_y} = [p_y, q_y]$.

By simple calculations, we have

$$S_{1x}^2 = 2.5, \quad S_{2x}^2 = 3.3$$

and

$$S_{1y}^2 = 10.2, \quad S_{2y}^2 = 15$$

Then the two testing statistics are obtained as follows:

$$F_l = \frac{S_{1x}^2}{S_{1y}^2} = 0.25 \text{ and } F_r = \frac{S_{2x}^2}{S_{2y}^2} = 0.22$$

Under the significant level $\boldsymbol{\alpha_I = 0.05}$, we use two tail *F*-test. Since $F_l \notin (0.5, 1.96)$ and $F_r \notin (0.5, 1.96)$. Hence we reject the null hypothesis $H_0: \sigma_{I_x} = \sigma_{I_y}$. We conclude that the two community variances are not equal. Clearly, $\sigma_{ix}^2 < \sigma_{iy}^2$, we find community *Y*'s variance is larger than *X*'s. This situation shows that people living in community *Y* have unstable income but people's income in community *X* is stable. Therefore, we suggest selling middle price goods in community *X* and selling low price on most kinds of goods but high price on a few kinds of goods.