A CASE STUDY ON CLASS JOINING PATTERN OF STUDENTS DURING COVID

Authors:

Souhardya Mitra

(Department of Statistics, St. Xavier's College , Kolkata)

INTRODUCTION:

2020 started as was expected. We were back into the loop of alarm – running to college – classes – exhausted – dozing off – alarm. The *CORONA VIRUS* had started to spread around the world and soon was declared a pandemic by **WHO** in March 2020.

On March 14, our college declared a two week holiday owing to the ongoing *CORONA VIRUS* pandemic. We were quite excited to have a break to the loop of the monotonous day and also we would get some time for our upcoming end semester examination.

This two week holiday has now been extended to almost a year. Physical classes had to be replaced by online classes.

We have collected data on the number of students who have joined 5 minutes before the start of the first class of the day and the first class of the second half (i.e. post lunch break) for the three years of our department **(STATISTICS**) and the third years of two other departments, viz. **MICROBIOLOGY** and **COMPUTER SCIENCE** for the third week of February, 2021.

In our article, we have worked on the following:

- Comparison between daily average number of students joining 5 minutes before the 1st class of the day and 5 minutes before the 1st class of second half (i.e. post lunch break)
    - Students of 3rd year of the Statistics department
    - Students of the Statistics department (all the three years taken together)
    - Students of 3rd year of the three departments considered in our experiment (viz. Statistics, Microbiology and Computer Science)
- Test for homogeneity over the three years of Statistics department
    - Joining time for the 1st class of the day
    - Joining time for the 1st class of 2nd half (i.e. post lunch break)
- Test for homogeneity over three departments (viz. Statistics, Microbiology and Computer Science)
    - Joining time for the 1st class of the day
    - Joining time for the 1st class of 2nd half (i.e. post lunch break)
- Test for independence of year and joining time in online class
    - Joining time for the 1st class of the day
    - Joining time for the 1st class of 2nd half (i.e. post lunch break)
- Test for independence of department and joining time in online class
    - Joining time for the 1st class of the day
    - Joining time for the 1st class of 2nd half (i.e. post lunch break)

TESTING PROBLEMS:

1.1 Comparison between daily average number of students of 3rd year of the Statistics department joining 5 minutes before the 1st class of the day and 5 minutes before the 1st class of second half (i.e. post lunch break):

Let X be the random variable denoting the number of students who have joined 5 minutes before the 1st class of the day and let Y be that for the 1st class of 2nd half (i.e. post lunch break).

Let us assume that $X \sim$ Poisson (λ_1) independently of $Y \sim$ Poisson (λ_2).

To test, $H_0 : \lambda_1 = \lambda_2$ against $H_1 : \lambda_1 \neq \lambda_2$.

Let random samples of sizes n_1 and n_2 be drawn respectively from the distributions of X and Y, independent of each other.

We have,

$X_i \sim$ Poisson (λ_1), i = 1 (1) n_1, independently of

$Y_i \sim$ Poisson (λ_2), i = 1 (1) n_2.

Define, $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ and $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$.

By Variance Stabilizing Transformation (*Square Root Transformation of Poisson mean*), we have,

$\sqrt{n_1}\left\{\sqrt{\bar{X}} - \sqrt{\lambda_1}\right\} \xrightarrow{d} N\left(0, \frac{1}{4}\right)$, independently of

$\sqrt{n_2}\left\{\sqrt{\bar{Y}} - \sqrt{\lambda_2}\right\} \xrightarrow{d} N\left(0, \frac{1}{4}\right)$, for moderately large n_1, n_2.

Test statistic: Define $T = \sqrt{\bar{X}} - \sqrt{\bar{Y}}$

$E(T) = \sqrt{\lambda_1} - \sqrt{\lambda_2}$

$V(T) = \frac{1}{4n_1} + \frac{1}{4n_2}$

$\therefore \frac{(\sqrt{\bar{X}} - \sqrt{\bar{Y}}) - (\sqrt{\lambda_1} - \sqrt{\lambda_2})}{\sqrt{\frac{1}{4n_1} + \frac{1}{4n_2}}} \xrightarrow{d} N(0,1)$, for moderately large n_1, n_2.

Under H_0, $Z = \frac{(\sqrt{\bar{X}} - \sqrt{\bar{Y}})}{\sqrt{\frac{1}{4n_1} + \frac{1}{4n_2}}} \sim N(0,1)$ asymptotically for moderately large n_1, n_2.

Test rule: Reject H_0 iff $|Z_{obs}| > \tau_{\frac{\alpha}{2}}$, where α is the level of significance of the test.

Computation:

$n_1 = 6, n_2 = 6, \bar{X} = 25.67, \bar{Y} = 20.67, \alpha = 0.05, \tau_{\frac{\alpha}{2}} = 1.95996$

$Z_{obs} = 1.8019$

Clearly, $|Z_{obs}| < \tau_{\frac{\alpha}{2}}$

Therefore, we accept H_0 at 5% level of significance.

Based on the data that we have collected it seems that the daily average number of students of 3rd year of the Statistics department joining 5 minutes before the 1st class of the day and 5 minutes before the 1st class of second half (i.e. post lunch break) are equal.

1.2 Comparison between daily average number of students of the Statistics department (all the three years taken together) joining 5 minutes before the 1st class of the day and 5 minutes before the first class of second half (i.e. post lunch break):

Let X be the random variable denoting the number of students who have joined 5 minutes before the 1st class of the day and let Y be that for the 1st class of 2nd half (i.e. post lunch break).

Let us assume that $X \sim$ Poisson (λ_1) independently of $Y \sim$ Poisson (λ_2).

To test, $H_0 : \lambda_1 = \lambda_2$ against $H_1 : \lambda_1 \neq \lambda_2$.

Proceeding similarly as in **1.1**, we arrive at the following conclusions.

Computation:

$n_1 = 6, n_2 = 6, \bar{X} = 79.50, \bar{Y} = 67.67, \alpha = 0.05, \tau_{\frac{\alpha}{2}} = 1.95996$

$Z_{obs} = 2.3913$

Clearly, $|Z_{obs}| > \tau_{\frac{\alpha}{2}}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the daily average number of students of the Statistics department (all the three years taken together) joining 5 minutes before the 1st class of the day and 5 minutes before the first class of second half (i.e. post lunch break) are equal.

1.3 Comparison between daily average number of students of 3rd year of the three departments considered in our experiment (viz. Statistics, Microbiology and Computer Science) joining 5 minutes before the 1st class of the day and 5 minutes before the 1st class of second half (i.e. post lunch break):

Let X be the random variable denoting the number of students who have joined 5 minutes before the 1st class of the day and let Y be that for the 1st class of 2nd half (i.e. post lunch break).

Let us assume that $X \sim$ Poisson (λ_1) independently of $Y \sim$ Poisson (λ_2).

To test, $H_0 : \lambda_1 = \lambda_2$ against $H_1 : \lambda_1 \neq \lambda_2$.

Proceeding similarly as in **1.1**, we arrive at the following conclusions.

Computation:

$n_1 = 6, n_2 = 6, \bar{X} = 75.17, \bar{Y} = 64.67, \alpha = 0.05, \tau_{\frac{\alpha}{2}} = 1.95996$

$Z_{obs} = 2.1765$

Clearly, $|Z_{obs}| > \tau_{\frac{\alpha}{2}}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the daily average number of students of 3rd year of the three departments considered in our experiment (viz. Statistics, Microbiology and Computer Science) joining 5 minutes before the 1st class of the day and 5 minutes before the 1st class of second half (i.e. post lunch break) are equal.

2.1 Test for homogeneity over the three years of Statistics department:

2.1.1 Joining time for the 1st class of the day

Consider the three years of Statistics department as 3 independent populations, each classified into 2 classes based on joining time.

Let p_{ij} denote the population proportion of members in the i^{th} class of the j^{th} population, i =1,2, j = 1,2,3.

To test, $H_0 : p_{i1} = p_{i2} = p_{i3}$, i = 1,2 against H_1 : not H_0

Let a random sample of size n_j be drawn from the j^{th} population, j = 1,2,3.

Let f_{ij} denote the number of members in the i^{th} class of the j^{th} population, i = 1,2, j = 1,2,3.

$\therefore \ \sum_{i=1}^{2} f_{ij} = n_j$.

Test statistic: Define, $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(f_{ij}- n_j p_{ij})^2}{n_j p_{ij}}$

Under H₀, $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(f_{ij}- n_j p_{io})^2}{n_j p_{io}}$, where p_{io} is the common value of p_{i1}, p_{i2}, p_{i3} under H₀, i = 1,2.

We estimate p_{i0} as, $\widehat{p_{io}} = \frac{\sum_{j=1}^{3} f_{ij}}{\sum_{j=1}^{3} n_j}$.

$\therefore \ \chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(f_{ij}- n_j \widehat{p_{io}})^2}{n_j \widehat{p_{io}}} \sim \chi^2_{(3-1)(2-1)}$.

Test rule: Reject H₀ iff $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$, where α is the level of significance of the test.

Computation:

Joining time \ Year	1st	2nd	3rd	Total
Joined 5 mins before	135	188	154	477
Not joined 5 mins before	195	124	206	525
Total	330	312	360	1002

$\alpha = 0.05, \chi^2_{\alpha;(3-1)(2-1)} = 5.99146$

$\chi^2_{obs} = 29.3166$

Clearly, $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$

Therefore, we reject H₀ at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the three years of Statistics department are homogeneous with respect to their joining time for the 1st class of the day.

2.1.2 Joining time for the 1st class of 2nd half (i.e. post lunch break)

Proceeding similarly as in **2.1.1**, we arrive at the following conclusions.

Computation:

Joining time \ Year	1st	2nd	3rd	Total
Joined 5 mins before	98	184	124	406
Not joined 5 mins before	232	128	236	596
Total	330	312	360	1002

$\alpha = 0.05, \chi^2_{\alpha;(3-1)(2-1)} = 5.99146$

$\chi^2_{obs} = 65.6395$

Clearly, $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the three years of Statistics department are homogeneous with respect to their joining time for the 1st class of the 2nd half (i.e. post lunch break).

Based on the data that we have collected it seems that there is not enough evidence to say that the three years of Statistics department are homogeneous with respect to their joining time in class.

2.2 Test for homogeneity over three departments (viz. Statistics, Microbiology and Computer Science) considered in our experiment:

2.2.1 Joining time for the 1st class of the day

Consider the three departments as 3 independent populations, each classified into 2 classes based on joining time.

Let p_{ij} denote the population proportion of members in the i^{th} class of the j^{th} population, i = 1,2, j = 1,2,3.

To test, $H_0 : p_{i1} = p_{i2} = p_{i3}$, i = 1,2 against H_1 : not H_0

Proceeding similarly as **2.1.1**, we arrive at the following conclusions.

Computation:

Joining time \ Dept.	STSA	MCBA	CMSA	Total
Joined 5 mins before	154	97	200	451
Not joined 5 mins before	206	185	208	599
Total	360	282	408	1050

$\alpha = 0.05, \chi^2_{\alpha;(3-1)(2-1)} = 5.99146$

$\chi^2_{obs} = 14.5574$

Clearly, $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the three departments considered in our experiment are homogeneous with respect to their joining time for the 1st class of the day.

2.2.2 Joining time for the 1st class of 2nd half (i.e. post lunch break)

Proceeding similarly as **2.2.1**, we arrive at the following conclusions.

Computation:

Joining time \ Dept.	STSA	MCBA	CMSA	Total
Joined 5 mins before	124	75	189	388
Not joined 5 mins before	236	207	219	662
Total	360	282	408	1050

$\alpha = 0.05, \chi^2_{\alpha;(3-1)(2-1)} = 5.99146$

$\chi^2_{obs} = 29.3349$

Clearly, $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the three departments considered in our experiment are homogeneous with respect to their joining time for the 1st class of the 2nd half (i.e. post lunch break).

(Based on the data that we have collected it seems that there is not enough evidence to say that the three departments considered in our experiment are homogeneous with respect to their joining time in class.)

3.1 Test for independence of year and joining time in online class:

3.1.1 Joining time for the 1st class of the day

Let the population of the students of Statistics department of the current year (2020 - 2021) be classified according to two attributes based on their year of study (1st, 2nd and 3rd years) (attribute A, say) and based on their joining time (attribute B, say).

Let A_1, A_2 and A_3 denote classes corresponding to attribute A and let B_1 and B_2 be those corresponding to attribute B.

Let p_{ij} denote the population proportion of members who belong to the i^{th} class of A and the j^{th} class of B, i = 1,2,3, j = 1,2.

Define, $p_{io} = \sum_{j=1}^{2} p_{ij}$: proportion of members in the population who belong to the i^{th} class of A, i = 1,2,3, and,

$p_{oj} = \sum_{i=1}^{3} p_{ij}$: proportion of members in the population who belong to the j^{th} class of B, j = 1,2.

To test, $H_0 : p_{ij} = p_{io} * p_{oj}, \forall (i,j)$ against H_1 : not H_0

Let a random sample of size n be drawn from the population. Let f_{ij} denote the number of members in the sample who belong to the i^{th} class of A and the j^{th} class of B, i = 1,2,3, j = 1,2.

Define, $f_{io} = \sum_{j=1}^{2} f_{ij}$, and, $f_{oj} = \sum_{i=1}^{3} f_{ij}$.

Test statistic: Define, $\chi^2 = \sum_{i=1}^{3} \sum_{j=1}^{2} \frac{(f_{ij} - n\, p_{ij})^2}{n\, p_{ij}}$

Under H_0, $\chi^2 = \sum_{i=1}^{3} \sum_{j=1}^{2} \frac{(f_{ij} - n\, p_{io} p_{oj})^2}{n\, p_{io} p_{oj}}$

We estimate p_{io} and p_{oj} as,

$\widehat{p_{io}} = \frac{f_{io}}{n}$, i = 1,2,3, and, $\widehat{p_{oj}} = \frac{f_{oj}}{n}$, j=1,2.

$\therefore\ \chi^2 = \sum_{i=1}^{3} \sum_{j=1}^{2} \frac{(f_{ij} - n\, \widehat{p_{io}}\widehat{p_{oj}})^2}{n\, \widehat{p_{io}}\widehat{p_{oj}}} \sim \chi^2_{(3-1)(2-1)}.$

Test rule: Reject H₀ iff $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$, where α is the level of significance of the test.

Computation:

Year \ Joining time	Joined 5 mins before	Not joined 5 mins before	Total
1^{st}	135	195	330
2^{nd}	188	124	312
3^{rd}	154	206	360
Total	477	525	1002

$\alpha = 0.05, \chi^2_{\alpha;(3-1)(2-1)} = 5.99146$

$\chi^2_{obs} = 29.3163$

Clearly, $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the attributes 'year of study (Statistics department)' and 'joining time for the 1^{st} class of the day' are independent.

3.1.2 Joining time for the 1^{st} class of 2^{nd} half (i.e. post lunch break)

Proceeding similarly as in **3.1.1**, we arrive at the following conclusions.

Computation:

Dept. \ Joining time	Joined 5 mins before	Not joined 5 mins before	Total
1^{st}	98	232	330
2^{nd}	184	128	312
3^{rd}	124	236	360
Total	406	596	1002

$\alpha = 0.05, \chi^2_{\alpha;(3-1)(2-1)} = 5.99146$

$\chi^2_{obs} = 65.6401$

Clearly, $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the attributes ‘Year of Study (Statistics department)’ and ‘joining time for the 1st class of the 2nd half (i.e. post lunch break)’ are independent.

Thus, based on the data that we have collected it seems that there is not enough evidence to say that the attributes ‘year of study (Statistics department)’ and ‘joining time in class’ are independent.

3.2 Test for independence of department and joining time in online class:

3.2.1 Joining time for the 1st class of the day

Let the population of the students of the three departments considered in our experiment, of the current year (2020 - 2021) be classified according to two attributes based on their department (viz. Statistics, Microbiology and Computer Science) (attribute A, say) and on their joining time (attribute B, say).

Let A_1, A_2 and A_3 denote classes corresponding to attribute A and let B_1 and B_2 be those corresponding to attribute B.

Let p_{ij} denote the population proportion of members who belong to the i^{th} class of A and the j^{th} class of B, i = 1,2,3, j = 1,2.

Define, $p_{io} = \sum_{j=1}^{2} p_{ij}$: proportion of members in the population who belong to the i^{th} class of A, i = 1,2,3, and,

$p_{oj} = \sum_{i=1}^{3} p_{ij}$: proportion of members in the population who belong to the j^{th} class of B, j = 1,2.

To test, $H_0 : p_{ij} = p_{io} * p_{oj}, \forall (i,j)$ against H_1 : not H_0

Proceeding similarly as in **3.1.1**, we arrive at the following conclusions.

<u>Computation</u>:

Dept. \ Joining time	Joined 5 mins before	Not joined 5 mins before	Total
STSA	154	206	360
MCBA	97	185	282
CMSA	200	208	408
Total	451	599	1050

$\alpha = 0.05, \chi^2_{\alpha;(3-1)(2-1)} = 5.99146$

$\chi^2_{obs} = 14.5572$

Clearly, $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the attributes 'department' and 'joining time for the 1st class of the day' are independent.

3.2.2 Joining time for the 1st class of 2nd half (i.e. post lunch break)

Proceeding similarly as in **3.2.1**, we arrive at the following conclusions.

Computation:

Dept. \ Joining time	Joined 5 mins before	Not joined 5 mins before	Total
STSA	124	236	360
MCBA	75	207	282
CMSA	189	219	408
Total	388	662	1050

$\alpha = 0.05, \chi^2_{\alpha;(3-1)(2-1)} = 5.99146$

$\chi^2_{obs} = 29.3342$

Clearly, $\chi^2_{obs} > \chi^2_{\alpha;(3-1)(2-1)}$

Therefore, we reject H_0 at 5% level of significance.

Based on the data that we have collected it seems that there is not enough evidence to say that the attributes 'department' and 'joining time for the 1st class of the 2nd half (i.e. post lunch)' are independent.

Thus, based on the data that we have collected it seems that there is not enough evidence to say that the attributes 'department' and 'joining time in class' are independent.

CONCLUSION:

From the above tests based on our data we arrive at the following conclusions:

- The daily average number of students joining 5 minutes before the 1^{st} class of the day and 5 minutes before the 1^{st} class of second half (i.e. post lunch break)
    - Seems to be equal for the students of 3^{rd} year of the Statistics department
    - Seems to be unequal for the students of the Statistics department (all the three years taken together)
    - Seems to be unequal for the students of 3^{rd} year of the three departments considered in our experiment (viz. Statistics, Microbiology and Computer Science)

- The three years of Statistics department
    - Does not seem to be homogeneous with respect to the joining time for the 1^{st} class of the day
    - Does not seem to be homogeneous with respect to the joining time for the 1^{st} class of 2^{nd} half (i.e. post lunch break)

Thus, the three years of Statistics department does not seem to be homogeneous with respect to the joining time in class.

- The three departments (viz. Statistics, Microbiology and Computer Science)
    - Does not seem to be homogeneous with respect to the joining time for the 1^{st} class of the day
    - Does not seem to be homogeneous with respect to the joining time for the 1^{st} class of 2^{nd} half (i.e. post lunch break)

Thus, the three departments considered here does not seem to be homogeneous with respect to the joining time in class.

- The year of study (Statistics department) seems to be dependent on the
    - Joining time for the 1^{st} class of the day
    - Joining time for the 1^{st} class of 2^{nd} half (i.e. post lunch break)

Thus, the year of study (Statistics department) seems to be dependent on the joining time in class.

- The attribute department seems to be dependent on the
    - Joining time for the 1st class of the day
    - Joining time for the 1st class of 2nd half (i.e. post lunch break)

Thus, the attribute department seems to be dependent on the joining time in class.

REFERENCES:

1) Nonparametric Tests for Complete Data by Vilijandas Bagdonavicius, Julius Kruopis, Mikhail S. Nikulin.
2) Multiway Contingency Tables Analysis for the Social Sciences by Thomas D. Wickens.
3) Statistical Methods for Meta-Analysis by Larry V. Hedges, Ingram Olkin