香 港 資 優 教 育 學 苑
The Hong Kong Academy for Gifted Education

Functional Data Analysis (FDA) of Longitudinal Data of Alumni's Quarter Completions of Courses and Workshops

Introduction

SLPs (**S**tudent **L**earning **P**rofiles) of students capture their details of progressions in the Academy. From these pieces of information, we may be able to extract the characteristics of their learning behaviors and explore any regular patterns.

In this study, we focus at those alumni who joined the Academy when they were in junior secondary level. Specifically, we focus at the completed histories of those who joined the Academy at the age of 14. It is because a larger number of completed records could be located for this cohort, as compared with other alternatives. The number of progression records totally amounts to around 800, after excluding those (inactive) alumni, who never participated at all.

Methodology

In the analysis of these historical progression data, a novel approach, called functional data analysis is deployed. **F**unctional **D**ata **A**nalysis (FDA) is a branch of statistics that analyses data providing information about curves, surfaces or anything else varying over a continuum. Under an FDA framework each sample element is, in general, considered to be a function. The physical continuum over which these functions are defined is often time, but may also be other types of continuum, such as spatial location.

In our study, the number of completions in the activities of high engagement (e.g., courses and workshops) in each quarter of an alumnus is concerned. Thus, for each alumnus, there would be a series of 20 numbers, comprising his/her history of progressions. However, following the perspective of FDA, it is viewed that 800 sampled functions are the primary focus of our investigation; instead of 16 000 individual data points. From these 800 sampled functions reflecting development trends of these alumni, we attempt to explore the characteristics of student learning behaviors.

Some Demographic Information of Targets

Amongst these some 800 respondents, their distribution pattern, in terms of gender and study domain is tabulated below. Amongst these (active) alumni, it seems that more males from Science/Math domains could be located. By applying chi-square test, it is confirmed that there exists a strong dependence between domain and gender (chi-square test statistics = 82.768, $p<0.001$).

Table 1: Domain and Gender Distribution of the (Active) Alumni in the Study

Domain/Gender	Female	Male	Total
Math/Science	149	449	598
(row-wise percentage)	24.90%	75.10%	-
(column-wise percentage)	56.90%	85.20%	75.79%
Hum & Lang/ Leadership	97	56	153
(row-wise percentage)	63.40%	36.60%	-
(column-wise percentage)	37.00%	10.60%	19.39%
Multi-discipline	16	22	38
(row-wise percentage)	42.10%	57.90%	-
(column-wise percentage)	6.10%	4.20%	4.82%
Total	262	527	789
(row-wise percentage)	33.21%	66.79%	-

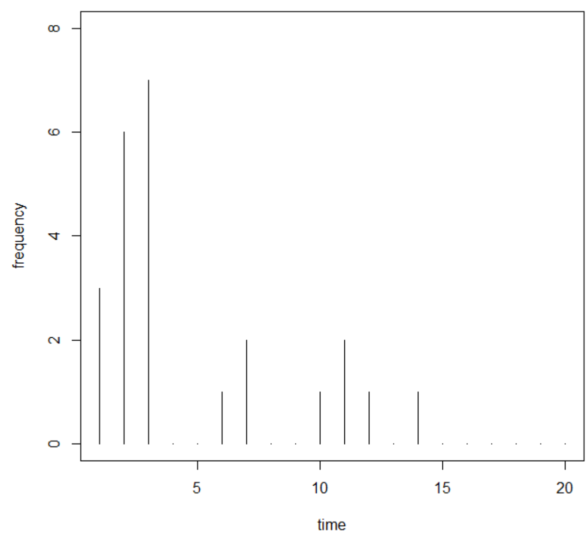
Visualizations of Raw Data

As mentioned above, for each alumnus a series of 20 numbers is available to reflect his/her study behavior. For a particular student, the corresponding data are shown below.

Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
3	6	7	0	0	1	2	0	0	1	2	1	0	1	0	0	0	0	0	0

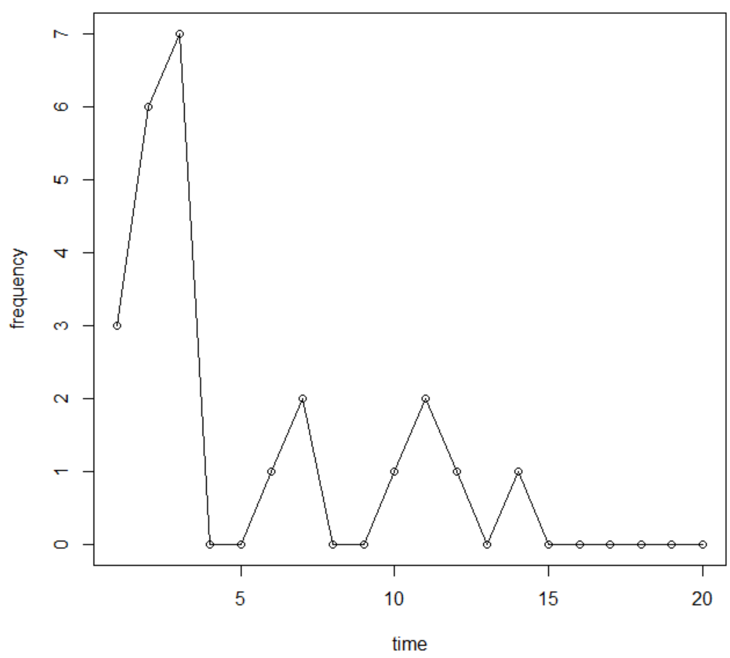
The data series could be graphically represented as below.

Fig.1a: Plot of a Particular Data Series



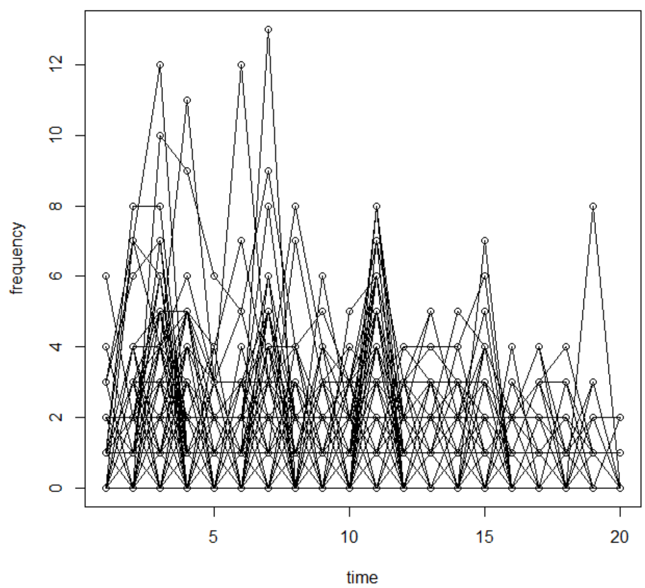
We study data on the alumni of around 800 individuals over several years of their studies in the Academy. For each of them a function is constructed over the time interval [1, 20], representing the person's level of study activity between ages 14 and 18 (i.e., one observation per quarter). Fig. 1a shows a typical record of an alumnus. We might link up the points in the figure using straight lines as a first attempt to represent them as a single function (see Fig. 1b).

Fig.1b: Plot of a Particular Data Series Joined using Line Segments



Altogether we consider around 800 records like this one. The records are all plotted in Fig. 2.

Fig.2: Plot of ALL Data Series Joined using Line Segments



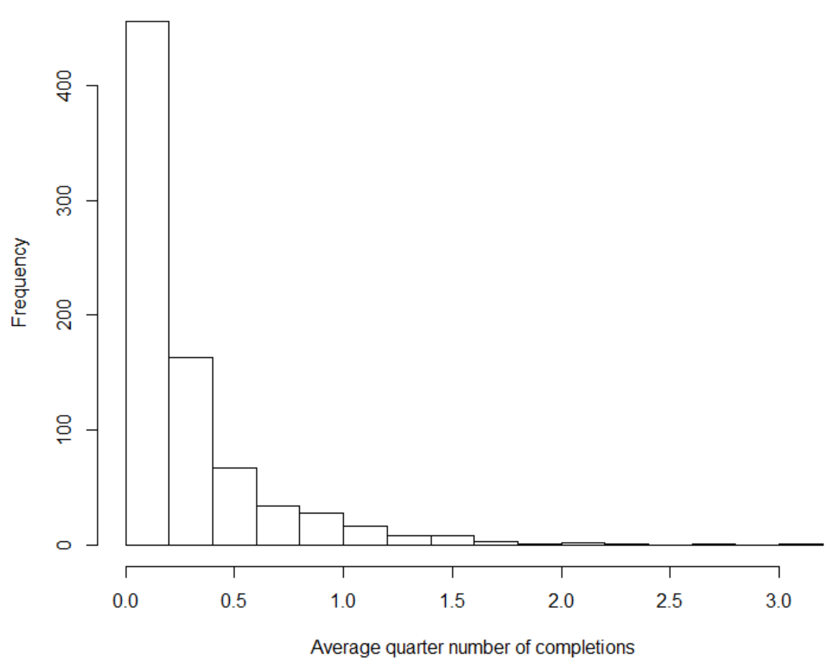
This figure demonstrates little more than the need for careful methods of summarizing and analyzing collections of functional data. Data of this kind are the simplest kind of functional data: we have a number of independent individuals, for each of whom we observe a single function. In standard statistics, we are accustomed to the notion of a sequence of independent numerical observations. This is the functional equivalent: a sequence of independent functional observations.

The questions we address in the next section include the following:

- How to make raw data on an individual record into a continuous functional observation?
- How could we estimate the mean of a population such as that shown in Fig. 2, and how can we investigate its variability?
- Are there distinct groups of alumni, in terms of their studying patterns?
- How does our analysis point to salient features of particular data?

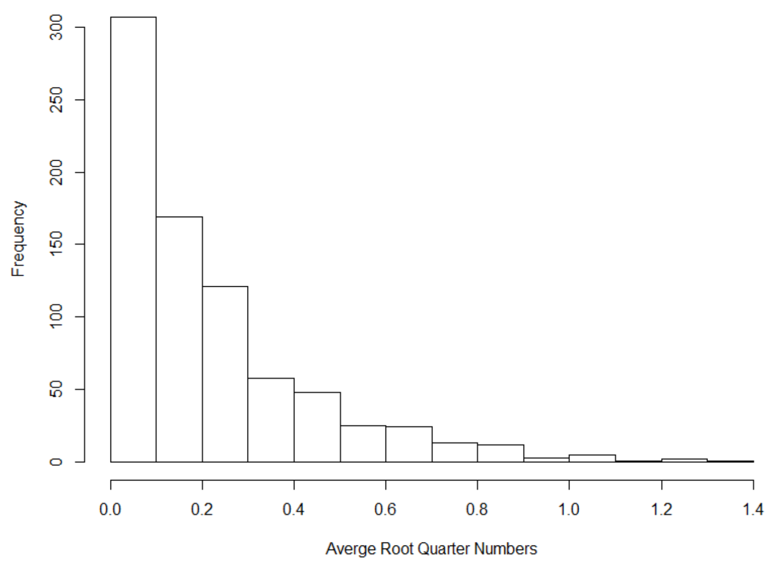
Before addressing the abovementioned questions, we would like to explain that it is more appropriate to track the square root of the number of completions per quarter; instead of the original ones. It is because of the skewness of the quarter counts (see Fig. 3; skewness = 2.85), giving inordinate weight to high values in the original data.

Fig.3: Histogram of Average Quarter Numbers of Completions of ALL Alumni



Taking the square root of the number of completions each quarter could stabilize the variability somewhat. If we plot a histogram of the averages across time of these square roots we see from Fig. 4 that the skewness is somewhat reduced (skewness =1.87).

Fig.4: Histogram of Average Root Quarter Numbers of Completions for ALL Alumni



Preliminary Phase of Data Analysis: Mean Function and Smoothing

As a preliminary step in the analysis of the data, it is to estimate the mean function of the functional data. The natural estimator to begin with is simply the sample average defined in this case by the function

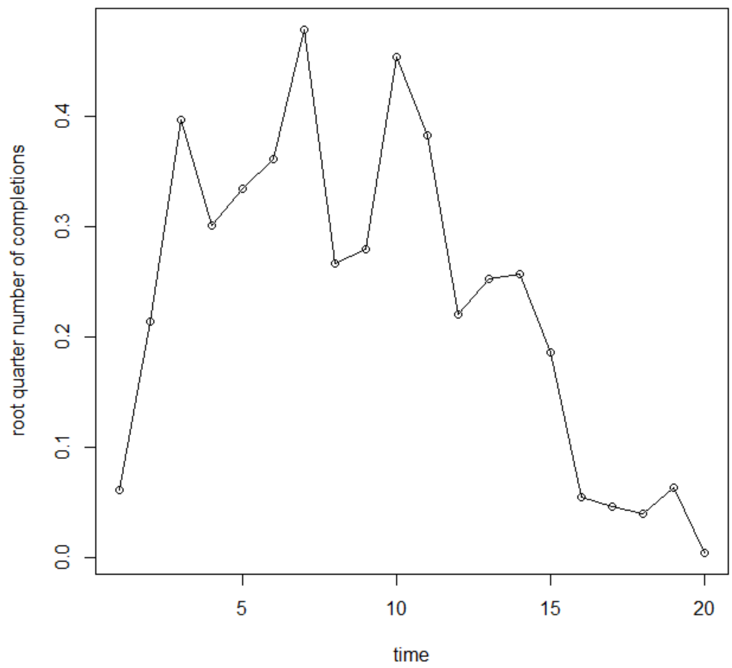
$$\frac{1}{N}\sum_{i=1}^{N} Y_i(t)$$

where N is the total number of observations, and

t is time point running from 1 to 20.

The mean function is plotted in Fig. 5. It can be seen that, despite the large number of functions on which the mean is based, there is still some fluctuation in the result of a kind that is clearly not relevant to the problem at hand. This particular set of data has high local variability.

Fig.5: Sample Mean Function of the Completion Functional Data



Smoothing: Computing Curves from Noisy Data

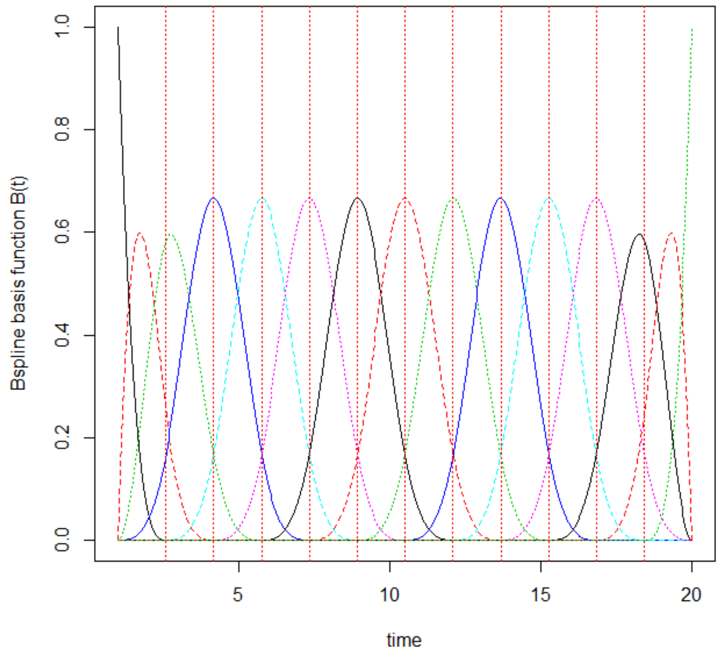
Smoothing function *x(t)* is defined using basis function expansion as below:

$$x(t)=\sum_{k=1}^{K} c_k \phi_k (t)$$

A set of functional building blocks ϕ_k; k = 1,…,K called basis functions, which are combined linearly. The parameters c_1, c_2, …, c_K are the coefficients of the expansion. The commonly used basis functions are splines, which are piecewise polynomials. Spline bases are more flexible. They are defined by the range of validity, the (interior) knots, and the order. There are many different kinds of splines. B-splines, short for basis splines, are considered here. Splines are constructed by dividing the interval of observation into subintervals, with boundaries at points called break points (interior knots). Over any subinterval, the spline function is a polynomial of fixed degree, but the nature of the polynomial changes as one passes into the next subinterval. We use the term degree to refer the highest power in the polynomial. The order of a polynomial is one higher than its degree.

We generate the basis functions (15 basis functions with order being equal to 4) using the R package, 'fda' (see Fig. 6).

Fig.6: 15 Bspline Basis Functions with 11 Interior Knots



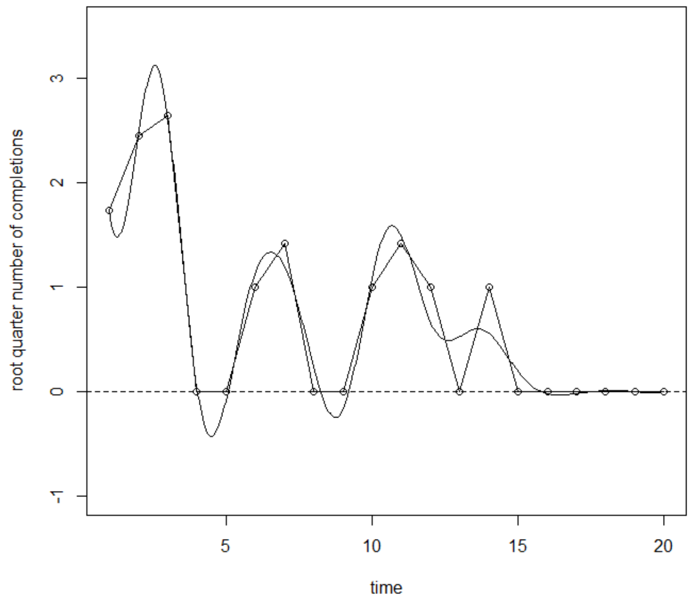
The polynomial segments are cubic or order four polynomials, and at each knot the polynomial values and their first two derivatives are required to match.

In this study, elementary smoothing, called regression smoothing is adopted without applying any roughness penalty. Simply speaking, the coefficients are determined by minimizing the sum of squared errors as shown below.

$$SSE(x) = \sum_j^n [y_j - x(t_j)\ ]^2$$
$$= \sum_j^n [y_j - \emptyset(\boldsymbol{t_j})'\boldsymbol{c}]^2$$

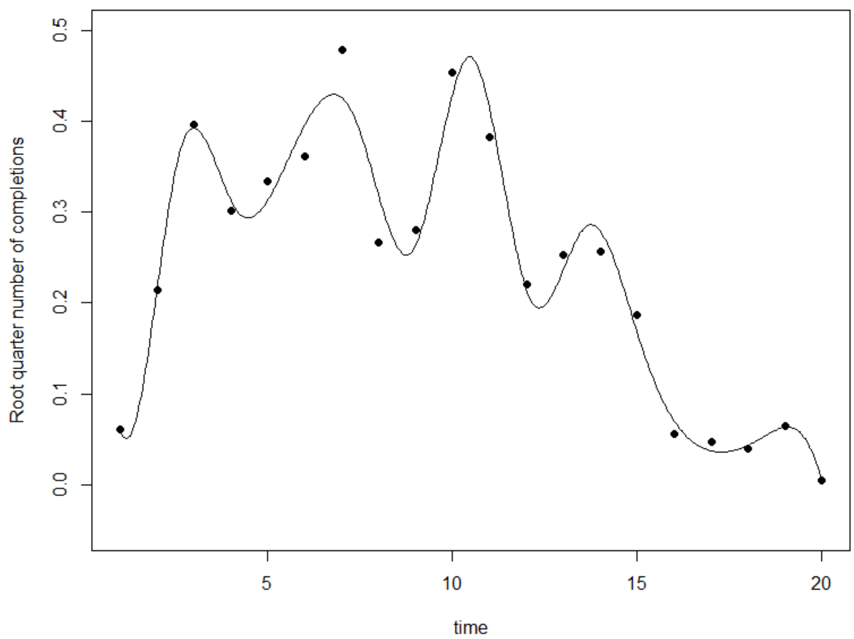
The smoothed curve of a particular data series for an alumnus is shown in Fig. 7.

Fig.7: Plot of a Particular Data Series After Smoothing with BSplines



The mean of all these smoothed functions is shown in Fig. 8 below. It is no surprise that there are 3 prominent peaks, each of which appears in around the summer term. Most likely, due to the study pressure of preparing for public exam, the frequency of completions dies down abruptly in the last two years.

Fig.8: Sample Mean Function of the Completion Functional Data After Smoothing



From Fig.9a and Fig. 9b, it can be observed that there are still variations amongst individuals with respect to the reference of overall mean function. The largest variation amongst individuals, in terms of numbers of completions, occurs in the early period of study.

Fig.9a: ALL Smoothed Data Series Using BSplines

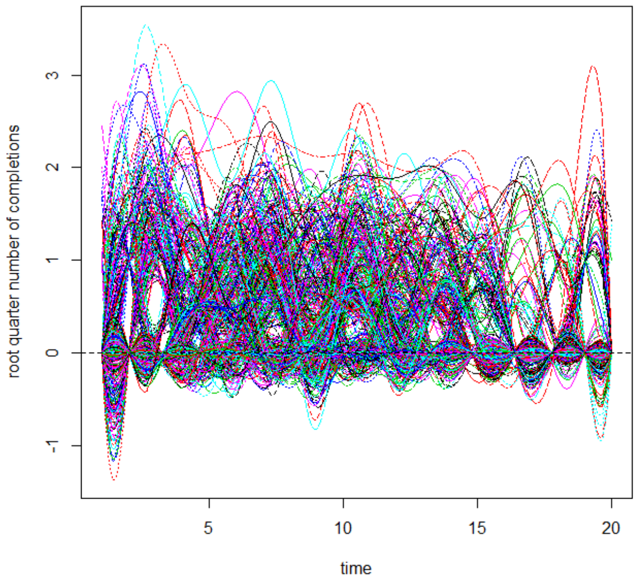
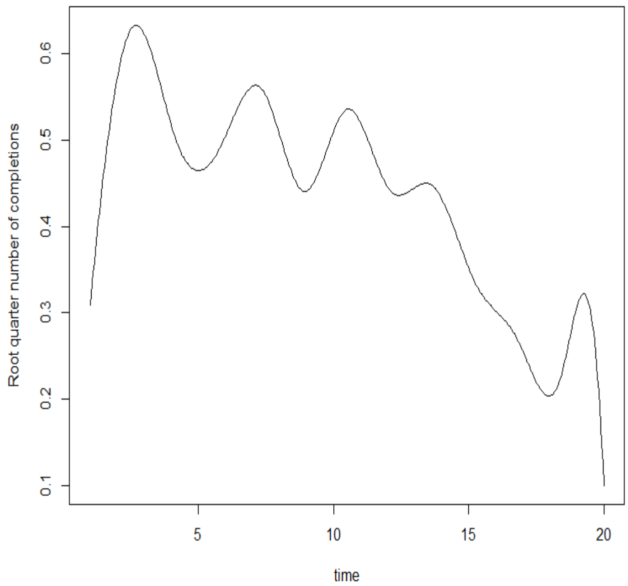
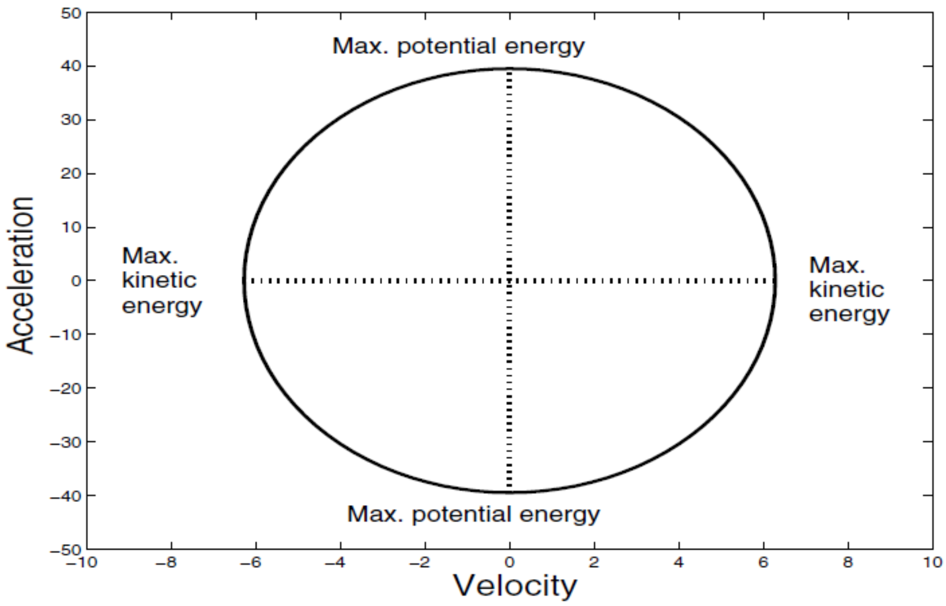


Fig.9b: Standard Deviation Function for ALL Smoothed Curves of Completion



We will study the characteristics and (regular) patterns of individual variations with respect to the reference of overall mean function in the next section. Beforehand, as we have derivatives at our disposal after smoothing individual data series, we can view dynamic behavior of an alumnus from a new perspective by studying how 1st and 2nd derivatives relate to each other; i.e., our tool is the plot of acceleration against velocity of a smoothed curve. Fig. 10 shows the plot, called phase-plane plot, for the function sin(2πt).

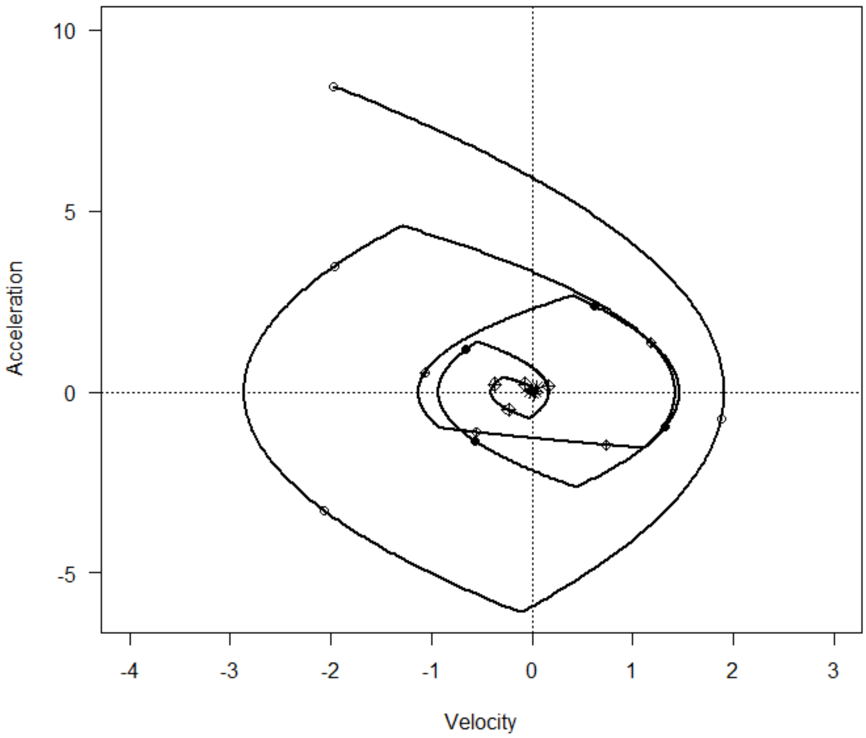
Fig.10: **A phase-plane plot of the simple harmonic function sin(2π*t*)**



The amount of energy in the system is related to the width and height of the ellipse in the figure. Kinetic energy is maximized when acceleration is 0, and potential energy is maximized when velocity is 0.

For the smoothed curve for an alumnus shown in Fig. 7 above, the corresponding phase-plane plot is generated (see Fig.11).

Fig.11: Phase-Plane Plot of a Smoothed Curve for a Particular Alumnus



From the figure, it can be observed that the largest cycle begins in the 1st quarter of the first year, starting with a large value of acceleration; i.e. a high value of potential energy. The velocity is increasing from this point. Then the cycle moves clockwise from 2nd quarter to 3rd quarter, passing through the horizontal zero acceleration. It implies that the maximum value of the function has been reached. From 3rd to 4th quarter the number of completions is decreasing, but the corresponding decreasing rate (i.e. velocity) is slowing down. Around the first quarter of the next year, the kinetic energy or velocity is near zero. The potential energy or acceleration is high, and it returns to the positive kinetic/zero potential phase in around the second quarter. To summarize, the progression process has one large cycle and two small ones swinging widely around zero; while the system energy is decreasing as time goes by.

Exploratory Data Analysis: Functional Principal Component Analysis (FPCA)

The motivation of Functional Principal components analysis is to search for a probe, $\xi(t)$ that reveals the most important type of variation in the functional data. That is, we ask, "For what weight function ξ would the probe scores

$$\rho_\xi(x_i) = \int \xi(t)\, x_i(t) dt$$

have the largest possible variation?" In order for the question to make sense, we have to impose a size restriction on ξ, and it is mathematically natural to require that: $\int \xi(t)^2 dt$=1.

As discussed in the previous section, the mean curve by definition is a mode of variation that tends to be shared by most curves, and we already know how to estimate this. Consequently, we usually remove the mean first and then probe the functional residuals $x_i - \bar{x}$. Thus, the probe score is defined as $\int \xi(t) \left(x_i(t) - \bar{x}(t)\right) dt$ instead.

The maximum probe score variance μ and the corresponding ξ are referred to as the largest eigenvalue and eigenfunction (or called harmonic) respectively. As in multivariate PCA, a non-increasing sequence of eigenvalues $\mu_1 \geq \mu_2 \ldots \geq \mu_k$ can be constructed stepwise by requiring each new eigenfunction, computed in step ℓ, to be orthogonal to those computed on previous steps.

$$\int \xi_j(t) \xi_l(t) dt = 0 \qquad j = 1,2,3,\ldots \ell-1 \text{ and } \int \xi_l(t)^2\, dt = 1$$

Similar to multivariate settings, we calculate eigenfunctions ξ_j of the bivariate covariance function *v(s, t)* as solutions of the following functional eigenequation.

$$\int v(s,t) \xi_j(t) dt = \mu_j \xi_j(s)$$

Suppose, then, that our software can present us with a number of positive eigenvalue/ eigenfunction pairs (μ_j, ξ_j). What do we do next? For each choice of ℓ leading eigenfunctions, it defines a basis system that can be used to approximate the sampled functions $x_i(t) - \bar{x}(t)$ These basis functions are orthogonal to each other and are normalized in the sense that $\int \xi_l(t)^2 dt = 1$. They are therefore referred to as an orthonormal basis. They are also the most efficient basis possible of size ℓ in the sense that the total error sum of squares

$$\text{PCASSE} = \sum_{i=1}^{N} \int [x_i(t) - \bar{x}(t) - c_i' \xi(t)]^2 dt$$

is the minimum achievable with only ℓ basis functions. Of course, other ℓ dimensional systems certainly exist that will do as well, but none will do better. These optimal basis functions ξ_j are often referred to as empirical orthogonal functions. It turns out that there is a simple relationship between the optimal total squared error and the eigenvalues that are discarded, namely that

PCASSE $=\sum_{j=l+1} u_j$ i.e., the total error sum of squares is equal to the summation of eigenvalues from l+1 onwards

The coefficient vectors ***C**$_i$'* contain the coefficients *Ci j* that define the optimal fit to $x_i(t)$ - $\bar{x}(t)$ and are referred to as *principal component scores*. They are given by the following:

$$Ci\,j = \int \xi_j(t)[x_i(t) - \bar{x}(t)]dt$$

They can be quite helpful in interpreting the nature of the variation identified by the PCA. It is also common practice to treat these scores as "data" to be subjected to a more conventional multivariate analysis.

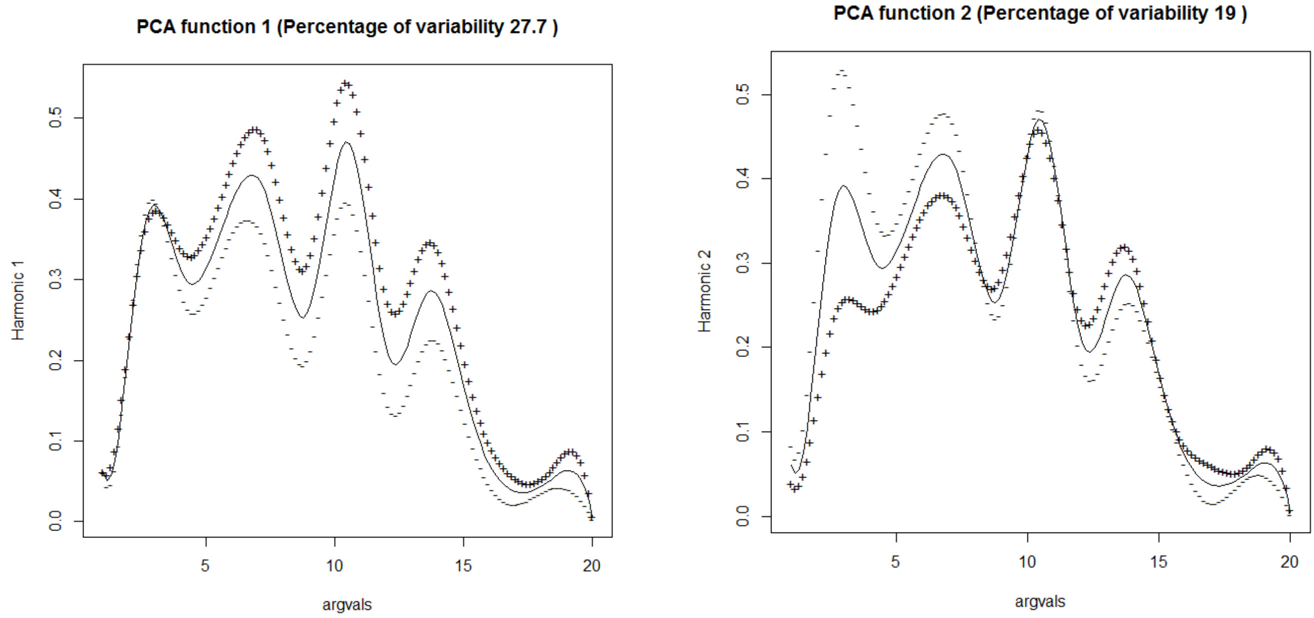
The eigenfunction basis was optimal but not unique. In fact, for any nonsingular square matrix **T** of order l, the system $\phi = \mathbf{T}\xi$ is also optimal and spans exactly the same functional subspace as that spanned by the eigenfunctions. Moreover, if $\mathbf{T}' = \mathbf{T}^{-1}$, such matrices being often referred to as *rotation matrices*, the new system ϕ is also orthonormal. Thus, for $l > 1$, there is nothing to prevent us from searching among the infinite number of alternative systems $\phi = \mathbf{T}\xi$ to find one where all of the orthonormal basis functions ϕj are seen to have some substantive interpretation. In the social sciences,
where this practice is routine, a number of criteria for optimizing the chances of interpretability have been devised for choosing a rotation matrix **T**, and we will demonstrate the usefulness of the popular *VARIMAX* criterion in the next section. In factor analysis, the rotation intends to maximize the number of loadings which are either close to 0 or 1 so as to simplify the interpretation.

Eigenfunctions, Eigenvalues and Principal Component Scores

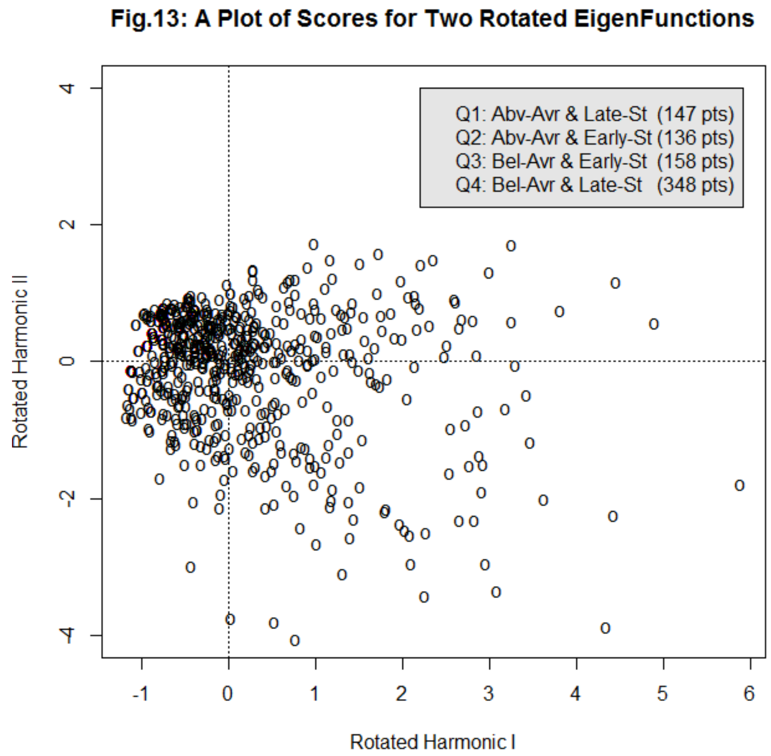
As discussed above, the linear combination of a few number of eigenfunctions $\xi_j(t)$ could capture the major components of variations of $x_i(t)$ from the overall mean $\bar{x}(t)$. The corresponding coefficients C_{ij} are called principal component scores. The proportion of variations that could be explained is reflected by the corresponding eigenvalues. To obtain a simplified version for easy interpretation, VARIMAX rotation is applied.

Fig. 12a and Fig. 12b shows the two principal component functions by displaying the mean curve along +'s and -'s indicating the consequences of adding and subtracting a small amount of each principal component. We do this because a principal component represents variation around the mean, and therefore is naturally plotted as such. We see that the first harmonic accounts for 27.7% of the variation, largely representing a vertical shift in the mean throughout the whole period except the first year. The second accounts for 19.0% of the variation, representing a prominent above-average development trend mainly at the first year when the corresponding principal component score is negative. Totally, they reflect 46.7% of the variation.

Fig. 12a and Fig. 12b: The First Two Eigenfunctions for the Data Set on Root Quarter Number of Completions after Applying VARIMAX Rotation



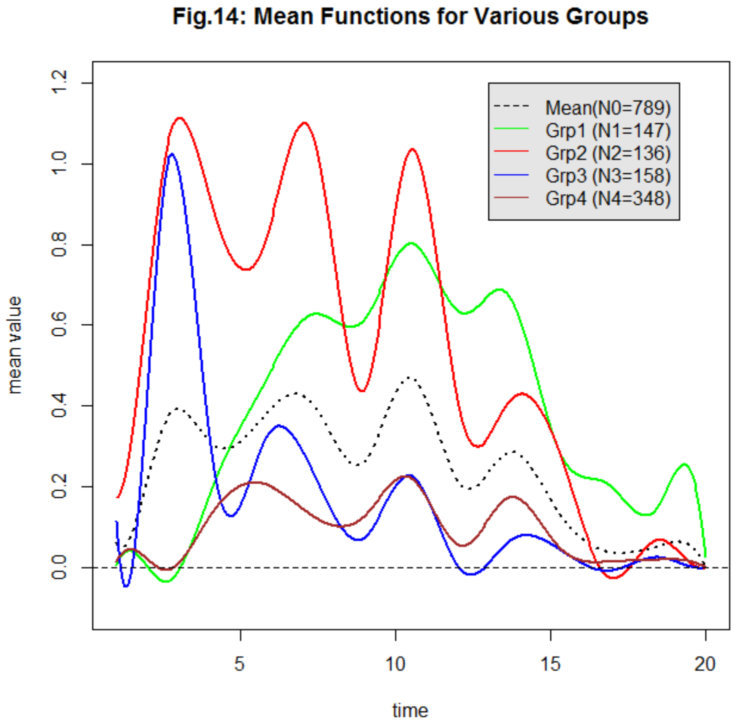
If the corresponding first principal component score of a data set is positive, it implies that the corresponding function would be largely above the overall mean. Similarly, if the corresponding second principal component score of a data set is negative, it implies that the corresponding function will be above the overall mean at the very beginning (i.e., the alumnus is an early-starter). It may be profitable to plot the principal component scores for these two harmonics to see how observations cluster and distribute themselves within the subspace spanned by the eigenfunctions. Fig. 13 displays these pairs of values.



As the means of these two principal component scores are respectively zeros, we could classify all the observations in Fig. 13 into four groups as follows:

(i) Group 1: The observations in the 1st quadrant; i.e., above-average & late-starter
(ii) Group 2: The observations in the 2nd quadrant; i.e., above-average & early-starter
(iii) Group 3: The observations in the 3rd quadrant; i.e., below-average & early-starter
(iv) Group 4: The observation in the 4th quadrant; i.e., below-average & late-starter

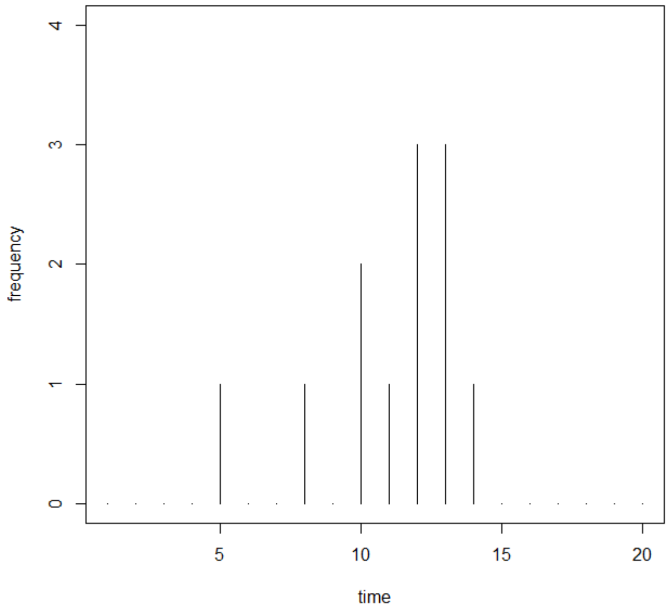
One way to capture the characteristics of these four groups is to plot the mean function for each group as shown in Fig. 14 below.



By comparing individual groups' mean functions vs. the overall one, we could visualize the group characteristics from the figure above. Besides, some typical individual examples in various groups are displayed below to further illustrate their characteristics.

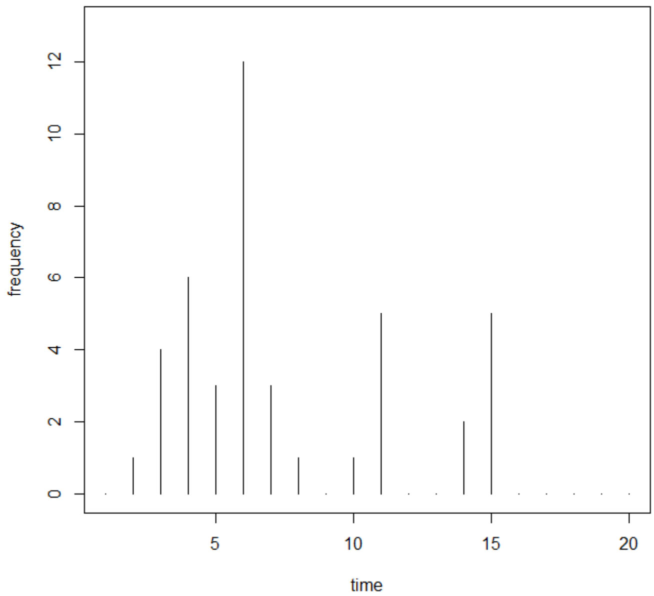
A typical student from <u>*Group 1*</u>: Above-average and late-starter (N=147; 18.6% of total):
Score of PCA 1= 1.71, Score of PCA 2= 1.01
The original data series:

Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
0	0	0	0	1	0	0	1	0	2	1	3	3	1	0	0	0	0	0	0



A typical student from *__Group 2__*: Above-average and early-starter (N=136; 17.2% of total):
Score of PCA 1= 3.08, Score of PCA 2= -3.34
The original data series:

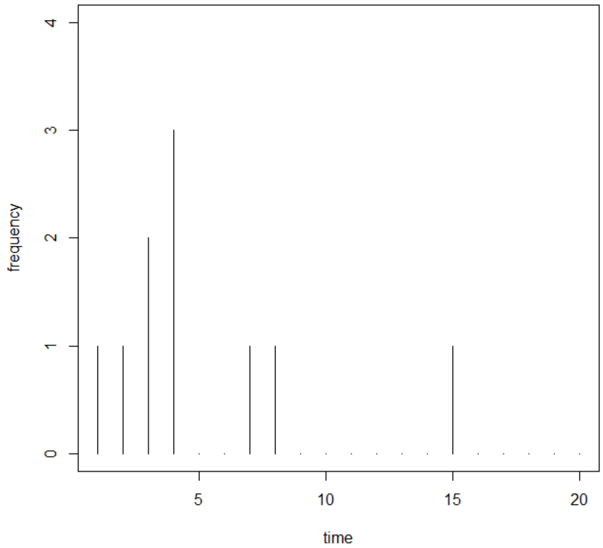
Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
0	1	4	6	3	12	3	1	0	1	5	0	0	2	5	0	0	0	0	0



A typical student from *__Group 3__*: Below-average and early-starter (N=158; 20.0% of total):
Score of PCA 1= -0.04, Score of PCA 2= -1.71

The original data series:

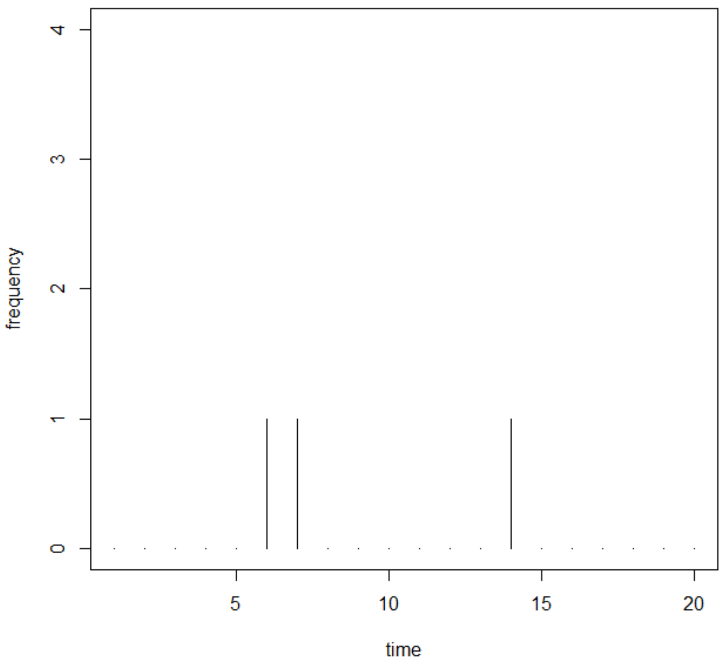
Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
1	1	2	3	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0



A typical student from *__Group 4__*: Below-average and late-starter (N=348; 44.1% of total): Score of PCA 1= -0.27, Score of PCA 2= 0.28

The original data series:

Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20
0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0



Association amongst Different Factors

In the following, we study the association amongst different factors, including starting pace, performance level, performance group, study domain, and gender.

Relationship between Starting Pace and Performance Level

First, we examine the relationship between starting pace (late-start or early-start) with performance level (below or above average). The corresponding two-way table is shown below.

Table 2: Two-way Table: Starting Pace vs. Performance Level

Start/Level	Below-Avr	Abv-Avr	Total
Late-Start	348	147	495
(row-wise percentage)	70.30%	29.70%	-
(column-wise percentage)	68.80%	51.90%	62.74%
Early-Start	158	136	294
(row-wise percentage)	53.70%	46.30%	-
(column-wise percentage)	31.20%	48.10%	37.26%
Total	506	283	789
(row-wise percentage)	64.13%	35.87%	-

The dependency between starting pace (late or early) and performance level (above or below average) is tested using chi-square test. A significant association is found (chi-square test statistics=21.279, p-value<0.001).

From the table, it can be observed that the importance of being an early-starter. Provided that an alumnus is a later-starter, he/she has a chance of 70.30% that his/her performance would be below average. On the other hand, if he/she is an early-starter, the chance of his/her performance being below average is approx. half-and-half.

Relationship between Performance Group and Study Domain

Next, we examine the relationship between performance group (group 1, 2, 3 and 4) and study domain (Math/Science, Humanities & Languages/Leadership and Multi-discipline). The corresponding two-way table is shown below.

Table 3: Two-way Table: Performance Group vs. Study Domain

Performance Grp/Domain	Math/Sci	Hum & Lang/ Leadership	Multi-Discipline	Total
Abv-avr & Late-Starter -Grp1	133	11	3	147
(row-wise percentage)	90.40%	7.50%	2.00%	-
(column-wise percentage)	22.20%	7.20%	7.90%	18.63%
Abv-avr & Early-Starter- Grp2	106	17	13	136
(row-wise percentage)	77.90%	12.50%	10.00%	-
(column-wise percentage)	17.70%	11.10%	34.20%	17.24%
Below-avr & Early-Starter -Grp3	106	41	11	158
(row-wise percentage)	67.10%	25.90%	7.00%	-
(column-wise percentage)	17.70%	26.80%	28.90%	20.03%
Below-avr & Late-Starter -Grp4	253	84	11	348
(row-wise percentage)	72.70%	24.10%	3.20%	-
(column-wise percentage)	42.30%	54.90%	28.90%	44.11%
Total	598	153	38	789
(row-wise percentage)	75.79%	19.39%	4.82%	-

The dependency between performance group and study domain is tested using chi-square test. A significant association is found (chi-square test statistics=40.104, p-value<0.001).

From the table, it can be observed that an alumnus in the study domain, Humanities & Languages/ Leadership was at greatest 'risk', in terms of his/her performance as compared with those in other domains. Provided that he/she is in Humanities & Languages/ Leadership, he/she has the greatest chance (54.90%) of being in the group of below-average and late-starter (42.30% for Math/Science and 28.90% for Multi-discipline). Besides, if he/she is in the group of late-starter, he/she has the greatest chance (88.4%) of having below-average performance (65.5% for Math/Science and 78.6% for Multi-discipline). A Humanities & Languages/ Leadership alumnus has a chance of 37.9% of being an early starter, which is similar to that of a Math/ Science alumnus (35.4%; 63.1% for Multi-discipline). However, even he/she is an early-starter, he/she would still have the greatest chance (70.7%) of having below-average performance (50.0% for Math/Science and 45.8% for Multi-discipline).

Relationship between Performance Group and Gender

Finally, we examine the relationship between performance group (group 1, 2, 3 and 4) and gender (Male, Female). The corresponding two-way table is shown below.

Table 4: Two-way Table: Performance Group vs. Gender

Performance Grp/Gender	Female	Male	Total
Abv-avr & Late-Starter -Grp1	44	103	147
(row-wise percentage)	29.90%	70.10%	-
(column-wise percentage)	16.80%	19.50%	18.63%
Abv-avr & Early-Starater - Grp2	36	100	136
(row-wise percentage)	26.50%	73.50%	-
(column-wise percentage)	13.70%	19.00%	17.24%
Below-avr & Early-Starter - Grp3	60	98	158
(row-wise percentage)	38.00%	62.00%	-
(column-wise percentage)	22.90%	18.60%	20.03%
Below-avr & Late-Starter - Grp 4	122	226	348
(row-wise percentage)	35.10%	64.90%	-
(column-wise percentage)	46.60%	42.90%	44.11%
Total	262	527	789
(row-wise percentage)	33.21%	66.79%	-

The dependency between performance group and gender is tested using chi-square test. No significant association is found (chi-square test statistics=5.65, p-value=0.13).

Summary and Conclusion

In this study, we investigate the cohort of alumni who joined and completed their studies in the Academy starting from the age of 14. Totally, it amounts to 789 completion records, after excluding those (inactive) ones, who never participated at all. The following observations could be concluded.

(i) From the overall mean function of functional completion data, it is no surprise that there are 3 prominent peaks, each of which appears in around the summer term. Most likely, due to the study pressure of preparing for public exam, the frequency of completions dies down abruptly in the last two years.

(ii) From the functional principal component analysis, we could classify all the completion records) in into four groups as follows, accounting for 46.7% of variation:

Group 1: Above-average & late-starter (N=147; 18.63% of the total)
Group 2: Above-average & early-starter (N=136; 17.24% of the total)
Group 3: Below-average & early-starter (N=158; 20.03% of the total)
Group 4: Below-average & late-starter (N= 348; 44.11% of the total)

(iii) It is found that overall speaking, an early-starter has a chance of around 50% of having above-average performance. On the other hand, a late-starter has a chance of only around 30% of having above-average performance.

(iv) The alumni in Humanities & Languages/ Leadership domains were in the 'worst' situation, as compared with those in other domains. They had a chance of 37.9% of being an early-starter, which was similar to the ones in Math/Science domains. However, the study results indicate that a large proportion (70.7%) of them lost their momentum later on. Besides, when they were a late-starter, a large proportion (88.4%) of them lost their momentum thereafter. Therefore, this group of alumni had the largest percentage (81.7%) of having below-average performance (60.0% for Math/Science and 57.8% for Multi-discipline).

Recommendations and Limitations of the Study

From the findings of the study, two recommendations are proposed for consideration.

(i) Orientation activities and promotion events to all new comers of the Academy are essential so as to arouse newly admitted students' interests in programmes and services provided by the Academy. This aspect of works should be further strengthened. It is because early participation could induce a much higher chance of having high performance level later on.

(ii) Further measures on promotion for newly admitted students belonging to Humanities and Languages/ Leadership domains should be considered, as the study results indicate that they could suffer from the greatest risks of dropping out as compared with those in other domains. The curriculum and nature of programmes and services provided in these domains could be reviewed so as to maintain these students' momentum of studying in the Academy.

The study examines only quantitative data on the number of completions of alumni after they have completed their studies in the Academy. It is difficult to recruit them for an interview and ask them to recall the (qualitative) reasons behind about their study patterns.