Чирчик С. В. Математичні передумови регресійного моделювання в оптимізаційній моделі «інтеграл компетентності дизайнера» // Проблеми освіти: Наук-метод. зб. / Інститут інноваційних технологій і змісту освіти МОН України. – Київ, 2015. – Вип. 85. – С. 217-223.

УДК 378.14:72.012

ЧИРЧИК С. В.,
канд. фіз.-мат. наук, доцент, докторант
(Житомирський державний університет імені Івана Франка)

МАТЕМАТИЧНІ ПЕРЕДУМОВИ РЕГРЕСІЙНОГО МОДЕЛЮВАННЯ В ОПТИМІЗАЦІЙНІЙ МОДЕЛІ «ІНТЕГРАЛ КОМПЕТЕНТНОСТІ ДИЗАЙНЕРА»

У роботі висвітлені проблеми діагностики якості вищої освіти на прикладі аналізу професійної підготовки у вищих навчальних закладах України освітньо-кваліфікаційного рівня «бакалавр», напряму 6.020207 «дизайн», кваліфікації «дизайнер інтер'єру». Описані математичні передумови регресійного моделювання при побудові запропонованої оптимізаційної моделі «Інтеграл компетентності дизайнера»

***Ключові слова**: регресійне моделювання, дизайн-освіта, оптимізаційна модель, діагностика*

В работе освещены проблемы диагностики качества высшего образования на примере анализа профессиональной подготовки в высших учебных заведениях Украины образовательно-квалификационного уровня «бакалавр», направления 6.020207 «дизайн», квалификации «дизайнер интерьера». Описаны математические предпосылки регрессионного моделирования при построении предложенной оптимизационной модели «Интеграл компетентности дизайнера»

***Ключевые слова:** регрессионное моделирование, дизайн-образование, оптимизационная модель, диагностика*

The problems of diagnostic the quality of high education with example of analysis of professional preparation in higher educational institutions of Ukraine in level "Bachelor", the faculty "Design" and qualifications "Interior Designer" were represented in this work. Mathematical backgrounds of regression modeling were described in this article using the building of proposal model "The integral of the competence of designer".

***Key words** : regression modeling, design education, an optimization model, diagnostics.*

Аналізуючи освітні програми провідних вузів за напрямом «дизайн», приходимо до висновку, що дизайн-освіта в Україні перейшла з стадії становлення в стадію оптимізації, що не може не вплинути на методи оцінки рівня професійної компетентності випускника. Підтвердженням цього є задекларовані засоби діагностики якості вищої освіти в методичних рекомендаціях з розробки складових галузевих стандартів вищої освіти.

При формуванні професійної компетентності майбутніх дизайнерів інтер'єру потрібно провести її диференціацію на компоненти, встановивши їх функціональні зв'язки і таким чином провести структуризацію цього поняття.

Для дослідження різних педагогічних явищ доцільно використовувати їх спрощений формальний опис - математичну модель. Так нами в роботі [1] запропонована оптимізаційна модель «Інтеграл компетентності дизайнера». Запропонований алгоритм оптимізаційного методу оцінки становлення професійної компетентності дизайнера, за функціональними компонентами передбачає оцінку рівня формування найбільш значущих, з професійної точки зору, компетенцій дизайнера інтер'єру (соціально-особистісних, інструментальних, загальнонаукових і професійних).

Методологічну оптимізації структури становлення професійної компетентності дизайнера, з урахуванням виробничих функцій і типових завдань діяльності, доцільно проводити методом порівняльної комплексної рейтингової оцінки за допомогою системи показників за такими функціональними компонентами [1, с.66]:

Перший блок – «*Мотиваційно-особистісний компонент*» – потребує аналізу показників:

1) рівня потреби власного удосконалення та здійснення професійної діяльності;
2) мети, мотивів, потреби, ціннісних орієнтирів актуалізації в професійній діяльності;
3) наявності професійного інтересу, що аргументує прагнення до професійно-особистісного зростання;
4) вольової полярності професійної самореалізації.

У другому блоці – «*Творчо-інноваційний компонент*» – розглядається:

1) відповідності знань, умінь, навичок до їх реалізації на практиці;
2) система засобів та прийомів професійної діяльності, що необхідна випускнику для проектування власної технології функціонування у професійній площині.

У третьому блоці – «*Реалізаційно-діяльнісний компонент*» – необхідно проаналізувати:

1) рівень здатності до генерування оригінальних ідей;
2) вміння реалізовувати сучасні проекти; здатності до прогнозування модних стилів та тенденцій у професійній площині;
3) рівень просторово-уявного мислення;
4) здатності до критичної самооцінки;
5) здатності до систематизування аналогів та прототипів;
6) вмінні узагальнювати та систематизувати;
7) рівень творчо-інноваційного прояву особистості у професійній площині.

Четвертий блок — «*Рефлексивний компонент*» – передбачає аналіз:

1) психологічних особливостей та природних професійних задатків;
2) уміння самодостатньо контролювати якісні і кількісні показники своєї діяльності та рівень власного розвитку у проекції на особистісні досягнення;
3) рівня сформованості таких якостей і властивостей як: ініціативність, послідовність, схильність до самоаналізу, здатність до передбачення та прогнозування результатів своєї діяльності.

Розглянемо поетапно математичний алгоритм оптимізаційної моделі «Інтеграл компетентності дизайнера», який базується на методі регресійного моделювання.

Математичні передумови регресійного моделювання при побудові оптимізаційної моделі

Регресійна модель застосовується для кількісного опису залежності деякої величини (в нашому випадку – рівень сформованості певної компетенції) від ряду факторів, що на неї впливають (ступеневий рівень освіти, рік навчання, методик навчання та ін.). Як правило, врахування усіх факторів неможливе у зв'язку з їх великою кількістю, або є занадто складним. Тому виділяються найбільш значущі фактори, що вважаються пояснюючими (незалежними) змінними, а вплив інших факторів враховується за допомогою деякої випадкової величини, що входить у модель. Внаслідок цього і залежну змінну необхідно розглядати як випадкову.

Нехай маємо n факторів, яким відповідають незалежні змінні $X1, X2 \ldots Xn$ і залежну змінну Y (рівень сформованості певної компетенції). Випадкову величину Y вважаємо неперервною з певним законом розподілу.

У традиційній постановці завдання регресійного моделювання передбачається, що умовні розподіли Y при кожному допустимому значенні незалежних змінних *нормальні*. Однак, це припущення не використовується при побудові моделі, а лише при її аналізі. Змінні $Xi\ (i = 1 \ldots n)$ можуть бути як випадковими так і невипадковими, в такому випадку їх значення задані чи відомі точно.

У досить загальному випадку, регресійна модель приймає вид [2, 3]:

$$Y = f(X_1, X_2, \ldots X_n) + \varepsilon,$$

де ε - випадкова величина, яка виражає вплив неврахованих факторів.

Природним вибором функціональної частини випадкової величини Y виступає її середнє значення, тобто умовне математичне сподівання $\mathrm{M_x(Y)}$ отримане при заданому наборі значень функціональних змінних $(\mathrm{x_1, x_2 \ldots x_n})$.

Таким чином, при природному виборі функціональної частини рівняння регресійної моделі прийме вигляд:

$$\mathrm{Y = M_x(Y)} + \varepsilon,$$

Передбачається, також, некорельованість випадкових помилок і змінних X і це дозволяє провести аналіз одержаної регресійної моделі.

Функціональна і кореляційна залежність випадкових величин.

Як відомо, залежність, при якій кожному значенню однієї змінної відповідає певний розподіл іншої змінної, називається статистичною [2, 3].

Введення поняття статистичного зв'язку аргументується тим, що вимірювання змінних супроводжується певними випадковими помилками, а також тим що на залежну змінну впливає ряд неврахованих факторів.

У зв'язку з неоднозначністю статистичної залежності між Y і X, для вченого представляє інтерес усереднене по X математичне сподівання $\mathrm{M_x(Y)}$. Наприклад, середній рівень сформованості професійної компетентності дизайнера інтер’єру у певній експериментальній групі.

Як відомо з курсу математичної статистики [2, 3], кореляційною залежністю між двома змінними називається функціональна залежність між значеннями однієї з них і умовним математичним очікуванням інший.

Кореляційна залежність представимо у вигляді:

$$\mathrm{M_x(Y)} = \varphi(\mathrm{x}).$$

При цьому залежна змінна Y називається функцією відгуку, результуюча змінна, а незалежна змінна X – вхідна, пояснююча. Рівняння, розглянуте вище, називається модельним рівнянням регресії, функція $\varphi(\mathrm{x})$ – модельної функцією регресії, а її графік – модельної лінією регресії.

Для опису результатів моделювання і їх аналізу необхідно знати умовний закон розподілу залежної змінної за умови виконання $X = x$. Завдяки тому, що дослідник має лише вибірку з пар значень $(x; y)$ обмеженої кількості n, у статистиці таку інформацію отримати, як правило, не вдається. За таких обставин оцінюються функція регресії оцінюється за вибіркою.

Цією оцінкою виступає вибіркове рівняння регресії:

$$\hat{y} = \hat{\varphi}(x, a_0, a_1, \ldots a_0)$$

де $\hat{y}$ – групова середня змінної Y (при $X = x$); $a_0, a_1, \ldots a_0$ – параметри функції.

Експериментальні дані в регресійному моделюванні

Щоб отримати досить достовірні і інформативні дані про формування професійної компетентності дизайнерів інтер'єру, необхідно мати вибірку спостережень досить великого обсягу (велика кількість експериментальних груп з достатньою кількістю студентів у кожній). Вибірка спостережень залежної змінної формування компетенцій і як результат становлення професійної компетентності дизайнера є реперної точкою нашого оптимізаційного дослідження.

Наші вибірки являють собою набори значень $(x_{i1}; x_{i2} \ldots x_{in}, y_i)$, $(i = 1 \ldots m)$, де n – кількість змінних, m – число спостережень.

Число спостережень m, в нашому випадку, велика і значно перевищує число незалежних змінних. При цьому виконується умова некорельованості їх збурень:

$$r(\varepsilon_i,\varepsilon_j)=0 \text{ при } i=j,$$

де $r(\varepsilon_i,\varepsilon_j)$ – коефіцієнт кореляції між збуреннями ε_i і ε_j.

Ця умова істотно спрощує модель і її статистичний аналіз.

Зокрема, у випадку однієї пояснюючої змінної, регресійна модель, побудована на основі вибірки експериментальних даних (x_i, y_i), має вигляд:

$$y_i = f(x_i)+\varepsilon_i, \quad i=1...m.$$

Припускається, також, що випадкові складові задовольняють умовам:

$$M(\varepsilon_i)=0,$$

$$r(\varepsilon_i,\varepsilon_j)=0,$$

$$D(\varepsilon_i)=\sigma^2,$$

де остання умова - це умова гомоскедастичності (постійність дисперсій помилок регресії). У цьому випадку розподілу величин Y_j відрізняються лише значенням математичного сподівання.

Якщо пояснююча змінна - рік навчання, то експериментальні данні можна розглядати і як динамічний ряд. Це дозволяє описати динаміку формування компетенції щодо вирішення типових задач професійної діяльності. В теорії статистичного аналізу [2, 3], динамічний ряд – це вибірка спостережень, у якій важливі не тільки самі спостережувані значення функціональних величин, але і порядок їх слідування. Найчастіше впорядкованість обумовлена тим, що експериментальні дані представляють собою серію спостережень однієї і тієї ж функціональної величини у впорядковані часові моменти (тимчасовий ряд). При цьому передбачається, що тип розподілу випадкової величини залишається одним і тим же (наприклад, нормальним), але параметри його змінюються в залежності від часових проміжків.

Побудова лінійної регресійної моделі

Нехай визначено характер експериментальних даних і відокремлено певний набір пояснюючих змінних (наприклад незалежна змінна – рік навчання, а залежна – рівень сформованості компетенції щодо вирішення типових задач професійної діяльності).

Для того, щоб знайти пояснену частину залежної змінної, тобто величину $M_x(Y)$, будується вибіркове рівняння регресії.

Ця процедура складається з двох етапів [2, 3]:

1) визначається параметричне сімейство, до якого належить шукана функція $M_x(Y)$, тобто

$$M_x(Y)=\hat{\varphi}(x,a_0,a_1...a_n),$$

де $a_0,a_1...a_n$ – параметри, що оцінюються;

2) знаходяться оцінки параметрів цієї функції за допомогою методів математичної статистики (як правило – метод найменших квадратів).

Відповідно до методу найменших квадратів невідомі параметри a_0 і a_1 вибираються з умови, що сума квадратів відхилень їх емпіричних значень y_i від значень $\hat{y}_i$, знайдених за рівнянням

$$\hat{y}=a_0+a_1(x),$$

була мінімальною:

$$S=\sum_{i=1}^{m}(\hat{y}_i-y_i)^2=\sum_{i=1}^{m}(a_0+a_1x_i-y_i)^2 \rightarrow \min$$

На підставі необхідної умови екстремуму функції двох змінних $S = S(a_0, a_1)$ прирівнюємо до нуля її частинні похідні:

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 2\sum_{i=1}^{n}(a_0 + a_1 x_i - y_i) = 0 \\ \dfrac{\partial S}{\partial a_1} = 2\sum_{i=1}^{n}(a_0 + a_1 x_i - y_i)x_i = 0 \end{cases}$$

після перетворень отримаємо систему нормальних рівнянь для визначення параметрів лінійної регресії:

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases}$$

Розділивши обидві частини рівнянь на n, одержимо систему рівнянь:

$$\begin{cases} a_0 + a_1 \bar{x} = \bar{y} \\ a_0 \bar{x} + a_1 \overline{x^2} = \overline{xy} \end{cases};$$

Уточнимо, що середні значення визначаються за формулами:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},\ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n},\ \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n},\ \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}.$$

Шляхом підстановки значення з першого рівняння системи у рівняння регресії отримаємо:

$$\hat{y} - \bar{y} = a_1(x - \bar{x}),$$

де a_1 - вибірковий коефіцієнт регресії Y по X (коеф. регресії Y по X показує, на скільки у середньому одиниць змінюється змінна Y при збільшенні змінної X на 1).

Розв'язуючи систему нормальних рівнянь, отримаємо:

$$b_1 = \frac{\overline{xy} - \overline{x}\overline{y}}{\overline{x^2} - \bar{x}^2} = \frac{\overline{\mathrm{cov}}(X,Y)}{S_x^2},$$

де S_x^2 – вибіркова дисперсія змінної X: $S_x^2 = \dfrac{\sum_{i=1}^{n} x_i^2}{n} - (\bar{x})^2$,

$\overline{\mathrm{cov}}(X,Y)$ – вибіркова коваріація.

Оцінка значущості рівняння регресії в оптимізаційної моделі.

Як відомо з курсу математичної статистики [2, 3] перевірити значущість рівняння регресії, означає виявити відповідність математичної моделі експериментальним даним.

Така перевірка значущості рівняння регресії проводиться на основі дисперсійного аналізу. У математичній статистиці дисперсійний аналіз виступає як самостійний метод. У нашому випадку він використовується у якості допоміжного для якісної оцінки регресійної моделі.

У відповідності до дисперсійного аналізу (див., наприклад, [2, 3]) повна сума квадратів відхилень окремих значень залежної змінної від середнього розкладається на два доданки:

$$Q = \sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n}[(\hat{y} - \bar{y}) + (y_i - \hat{y})]^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y})^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \bar{y})(y_i - \hat{y}).$$

Продемонструємо, що

$$2\sum_{i=1}^{n}(\hat{y}_i-\overline{y})(y_i-\hat{y})=0 .$$

$\hat{y}_i-\overline{y}=a_i(x_i-\overline{x})$; $\hat{y}_i-\overline{y}=y_i-a_0-a_1x_i=y_i-(\overline{y}-a_1\overline{x})-a_1x_i=(y_i-\overline{y})-a_1(x_i-\hat{x})$.

Тобто:

$$2\sum_{i=1}^{n}(\hat{y}_i-\overline{y})(y_i-\hat{y})=2a_1\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})-2a_1{}^2\sum_{i=1}^{n}(x_i-\overline{x})^2=0 .$$

Отже:

$$Q=Q_R+Q_E$$

де Q – загальна сума квадратів відхилень залежної змінної від середньої, Q_R – сума квадратів, що обумовлена регресією, Q_E – залишкова сума квадратів, яка характеризує вплив факторів, що не були враховані.

Схема дисперсійного аналізу має вигляд [див., наприклад 2, 3]:

Компоненти дисперсії	Сума квадратів	Число степенів свободи	Середні квадрати
Регресія	$Q_R=\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2$	$k-1$	$S_R^2=\dfrac{Q_R}{k-1}$
Залишкова	$Q_E=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$n-k$	$S^2=\dfrac{Q_E}{n-k}$
Загальна	$Q=\sum_{i=1}^{n}(y_i-\overline{y})^2$	$n-1$	

k – число параметрів рівняння регресії, що підлягають оцінці; n – число спостережень.

Величину Q можна обрахувати і за формулою:

$$Q=\sum_{i=1}^{n}(y_i)^2-\frac{(\sum_{i=1}^{n}y_i)^2}{n} .$$

При відсутності лінійної залежності між залежною і пояснюючими змінними випадкові величини S_R^2 і S^2 мають χ^2-розподіл відповідно з k - 1 і n - k степенями вільності, а їх відношення – F-розподіл з такими самими степенями вільності. Це відношення виступає в ролі тестової статистики:

$$F=\frac{Q_R(n-k)}{Q_E(k-1)}=\frac{S_R^2}{S^2} .$$

Для перевірки вибирається рівень значимості α – це ймовірнисть помилки першого роду, тобто ймовірність відхилення правильної гіпотези, в даному випадку – гіпотези про значущість знайденого рівняння. Як правило $\alpha=0.01...0.05$.

Значення, отримане при спостереженні F порівнюється із критичним F_{α,m_1,m_2} : якщо $F>F_{\alpha,m_1,m_2}$, то рівняння регресії є значущим на рівні α .

де F_{α,m_1,m_2} – табличне значення F-критерію Фішера-Снедекора, визначене на рівні значущості α при $m_1=k-1$ і $m_2=n-k$ степенях вільності,

У випадку лінійної парної регресії $k=2$, і вибіркове рівняння регресії є значущим на рівні α , якщо

$$F=\frac{Q_R(n-2)}{Q_E}>F_{\alpha,1,n-2} .$$

Відзначимо, що значущість вибіркового рівняння парної лінійної регресії може бути перевірена і шляхом оцінки значущості коефіцієнта a_1, який, має *t*-розподіл Ст'юдента з $m = n - 2$ степенями вільності.

Рівняння парної лінійної регресії або коефіцієнт регресії α_1 значущі на рівні α, якщо спостережуване значення статистики критерію Ст'юдента

$$t = \frac{a_1 - 0}{s}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

за абсолютною величиною більше критичного ($|t| > t_{1-\alpha;n-2}$). Тут $s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$ незміщена оцінка дисперсії обурення.

Можна показати, що для парної лінійної моделі обидва способи перевірки значимості з використанням F- і t- критеріїв рівносильні, тому що вони пов'язані співвідношенням $F = t^2$.

Однією з найбільш змістовних оцінок якості рівняння регресії є коефіцієнт детермінації, який можна визначити за формулою:

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_E}{Q}.$$

Його величина показує, яка частка варіації залежної змінної обумовлена варіацією пояснюючої змінної ($0 \le R^2 \le 1$).

Чим ближче значення R^2 до 1, тим вища якість апроксимації емпіричних даних при регресії (при $R^2 = 1$ – емпіричні точки $(x_i; y_i)$, що лежать на лінії регресії; при $R^2 = 0$ – варіація залежної змінної повністю обумовлена впливом неврахованих у моделі змінних, і лінія регресії паралельна до осі абсцис).

При відомому коефіцієнті детермінації R^2 статистику критерію значущості рівняння регресії можна подати у вигляді:

$$F = \frac{Q_R(n-k)}{(1-R^2)(k-1)} > F_{\alpha,m_1,m_2}.$$

Зауважимо, що коефіцієнт детермінації дорівнює квадрату коефіцієнта кореляції ($R^2 = r^2$) при побудові парної лінійної регресійної моделі.

Наприклад, в нашій моделі достатньо високе значення R^2 свідчить про стійку динаміку формування компетенції щодо вирішення типових задач професійної діяльності.

Використання цієї моделі дає можливість кількісно оцінити динаміку формування компетенцій, щодо вирішення типових задач професійної діяльності (соціально-особистісних, інструментальних, загальнонаукових і професійних).

Для порівняння ефективності впровадження запропонованої нами концептуальної моделі професійної підготовки майбутніх бакалаврів з дизайну інтер'єру [4] із існуючими нами буде використано дисперсійний аналіз, результати якого буде висвітлено у наступних роботах.

Література

1. Чирчик С.В. Оптимізаційна модель «Інтеграл компетентності дизайнера» / Вища школа – 2014. – № 2. – с. 59-95.
2. Ферстер Э., Ренц Б, Методы корреляционного и регрессионного анализа: Пер. с нем. — М.: Финансы и статистика, 1982. - 304 с.
3. Кремер Н.Ш. Теория вероятностей и математическая статистика. – 2-е изд., перераб. и доп. - М.: Юнити-Дана, 2004. — 573 с.
4. Chyrchyk S. Conceptual model of professional training of future interior designers / Збірник наукових доповідей. «Педагогіка. Актуальні наукові дослідження. Теорія, практика» (30.03.15 – 31.03.15) – Познань – С.77-85