1) Вычислите предел

$$\lim_{x\to 0}\left(1+\frac{\ln(1+\arcsin^2 x)}{tg(\pi x)}\right)^{\frac{2arctg\frac{1}{|x|}}{x}}$$

Решение

Обозначим L = $\lim_{x\to 0}\left(1+\frac{\ln(1+\arcsin^2 x)}{tg(\pi x)}\right)^{\frac{2arctg\frac{1}{|x|}}{x}}$.

Используя так называемую таблицу эквивалентностей, а, по сути, линейный член разложения функций в ряд Маклорена и свойства функции arctg, получаем

$$L=\lim_{x\to 0}\left(1+\frac{x^2}{\pi x}\right)^{\frac{2\cdot\frac{\pi}{2}}{x}}=\lim_{x\to 0}\left(1+\frac{x}{\pi}\right)^{\frac{\pi}{x}}=e$$

---

2) Найти собственные числа и собственные векторы матрицы.

$$\begin{pmatrix}1 & -3 & 4\\ 4 & -7 & 8\\ 6 & -7 & 7\end{pmatrix}$$

Ортогонализировать систему собственных векторов этой матрицы и дополнить ее до ортогонального базиса в исходном линейном пространстве.

Решение

Этап 1. Найдем собственные числа и собственные векторы исходной матрицы.

Обозначим исходную матрицу за A и составим для неё характеристическое уравнение

$$\det(A-\lambda E)=\begin{vmatrix}1-\lambda & -3 & 4\\ 4 & -7-\lambda & 8\\ 6 & -7 & 7-\lambda\end{vmatrix}=$$

$$=(1-\lambda)((-7-\lambda)(7-\lambda)+56)+3(4(7-\lambda)-48)+4(-28-6(-7-\lambda))=-\lambda^3+\lambda^2+5\lambda+3=0.$$

Подбором находим корень характеристического уравнения -1. По теореме Безу характеристический многочлен делится на $\lambda+1$.

Разделив, получаем

$$-\lambda^3+\lambda^2+5\lambda+3=(\lambda+1)(-\lambda^2+2\lambda+3)=0$$

Корни квадратного трехчлена $-\lambda^2+2\lambda+3$ равны -1 и 3.

Таким образом, собственные числа матрицы А равны $\lambda_1=-1$ (кратности 2) и $\lambda_2=3$.

Найдем собственные векторы/собственный вектор матрицы А, решив уравнения $(A-\lambda E)u=0$ (где u и 0 – трехмерные вектора) соответственно для $\lambda_1=-1$ и $\lambda_2=3$.

1) Пусть $\lambda_1=-1$, тогда

$$\begin{pmatrix}2 & -3 & 4\\ 4 & -6 & 8\\ 6 & -7 & 8\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

Первая и вторая строчки матрицы $(A+E)$ пропорциональны, поэтому оставляем только первую и третью. Получаем систему уравнений

$\begin{cases}2x-3y+4z=0\\ 6x-7y+8z=0\end{cases}$, размерность базиса решений которой равно 1.

Одно из нетривиальных решений и будет собственным вектором для $\lambda_1=-1$:

$$v_1=\begin{pmatrix}1\\ 2\\ 1\end{pmatrix}$$

2) Пусть $\lambda_1=3$, тогда

$$\begin{pmatrix}-2 & -3 & 4\\ 4 & -10 & 8\\ 6 & -7 & 4\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

Вторая строчка матрицы $(A-3E)$ является суммой первой и третьей, поэтому оставляем только первую и третью. Получаем систему уравнений

$\begin{cases}-2x-3y+4z=0\\ 6x-7y+4z=0\end{cases}$, размерность базиса решений которой равно 1.

Одно из нетривиальных решений будет собственным вектором для $\lambda_2=3$:

$$v_2=\begin{pmatrix}1\\ 2\\ 2\end{pmatrix}$$

Этап 2. Ортогонализируем систему собственных векторов $v_1=\begin{pmatrix}1\\ 2\\ 1\end{pmatrix}$ и $v_2=\begin{pmatrix}1\\ 2\\ 2\end{pmatrix}$.

Применим к системе векторов $v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ и $v_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ процедуру ортогонализации Грамма – Шмидта, согласно которой

$v_1' = v_1$, $\quad v_2' = v_1 + \lambda v_2$, причем вектор v_2' будет ортогонален вектору v_1' при

$\lambda = -\dfrac{(v_1, v_1)}{(v_1, v_2)} = -\dfrac{6}{7}$, т.е. ортогональная система векторов имеет вид

$$v_1' = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \quad v_2' = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - \frac{6}{7}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = -\frac{1}{7}\begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}$$

Этап 3. Дополним систему до ортогонального базиса.

Вектором, ортогональным к $v_1'\, u\, v_2'$, будет их векторное произведение.

Технические удобнее векторно перемножить $v_1'\, u - 7v_2'$:

$$v_1' \times -7v_2' = \begin{vmatrix} i & j & k \\ 1 & 2 & 1 \\ 1 & 2 & -5 \end{vmatrix} = -12i + 6j \sim \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}.$$

Т.е. в качестве вектора, ортогонального $v_1'\, u\, v_2'$, можно выбрать вектор $\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$.

Ответ.

Собственные числа исходной матрицы равны $\lambda_1 = -1$ (кратности 2) и $\lambda_2 = 3$, соответствующие собственные векторы $v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ и $v_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$. Ортогональный базис в исходном линейном пространстве (один из): $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -5 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$.

---

3) Исследуйте на экстремум следующую функцию: F(x,y)=3x^3+21y^3-18xy+5

Решение

Найдем точки, подозрительные на экстремум, решая следующую систему уравнений:

$$\begin{cases} \dfrac{\partial F}{\partial x} = 9x^2 - 18y = 0 \\ \dfrac{\partial F}{\partial y} = 63y^2 - 18x = 0 \end{cases}.$$

Получим точки со следующими координатами: $(0;0), \left(1;\frac{1}{2}\right)$.

Далее необходимо проверить выполнение условий второго порядка. Для этого найдем матрицу вторых производных исследуемой функции:

$$\begin{pmatrix} 18x & -18 \\ -18 & 126y \end{pmatrix}.$$

Проверим знакоопределенность этой матрицы в каждой из найденных подозрительных точек.

Для точки с координатами $(0;0)$ имеем следующую матрицу: $\begin{pmatrix} 0 & -18 \\ -18 & 0 \end{pmatrix}$. Очевидно, $0 - 18^2 < 0$. Следовательно, точка с координатами $(0;0)$ не является точкой экстремума.

Для точки с координатами $\left(1;\frac{1}{2}\right)$ имеем следующую матрицу: $\begin{pmatrix} 18 & -18 \\ -18 & 63 \end{pmatrix}$. Очевидно, $18 > 0$, $18 \cdot 63 - 18^2 > 0$. Следовательно, точка с координатами $\left(1;\frac{1}{2}\right)$ является точкой минимума. Значение функции в этой точке: F(x,y)=1.625

---

4) Пусть $Q(x,y) = 4x^2 - 9y^2$, аргументы $Q(x,y)$ удовлетворяют условию $F(x,y) = c + ax + by = 0$.

При каких значениях параметров $'a,b,c'$ функция $Q(x,y)$:

a) будет иметь ровно одну условную стационарную точку, определите, является ли данная точка экстремумом;
b) будет иметь более одной условной стационарной точки, определите, являются ли данные точки экстремумами;
c) не будет иметь стационарных точек.

<u>Указание</u>. Для нахождения условных стационарных точек использовать метод множителей Лагранжа. Дополнительных исследований проводить не надо.

<u>Решение</u>

A) Если a=0, b=0, c=0 ограничению удовлетворяют все значения аргументов. В этом случае условная стационарная точка совпадает с безусловной стационарной точкой. Она определяется из условий $\begin{cases} \dfrac{\partial}{\partial x} Q(x,y) = 0 \\ \dfrac{\partial}{\partial y} Q(x,y) = 0 \end{cases}$. Решением данной системы уравнений является точка $x_k = 0,\ y_k = 0$. Определим ее тип. Для этого используем критерий Сильвестра и проанализируем знаки главных миноров матрицы Гессе

$G = \begin{pmatrix} 8 & 0 \\ 0 & -18 \end{pmatrix}$. Поскольку второй главный минор отрицателен, в критической точке $x_k = 0,\ y_k = 0$ данная стационарная точка не является экстремумом.

B) Если $a = 0, b = 0, c \neq 0$, то этому ограничению не удовлетворяет ни одно значение аргументов, и стационарных точек нет.

C) Пусть хотя бы один из параметров (a или b) не равен нулю. В этом случае используем метод множителей Лагранжа. Функция Лагранжа имеет вид $L(x, y, \lambda) = 4x^2 - 9y^2 + \lambda(c + ax + by)$. Безусловные стационарные точки функции Лагранжа совпадают с условными стационарными точками в постановке задачи. Стационарная точка определяется из условий

(*) $\begin{cases} \dfrac{\partial}{\partial x} L(x, y, \lambda) = 8x + \lambda a = 0 \\ \dfrac{\partial}{\partial y} L(x, y, \lambda) = -18y + \lambda b = 0 \\ \dfrac{\partial}{\partial \lambda} L(x, y, \lambda) = c + ax + by = 0 \end{cases}$. Решение данной системы уравнений существует не всегда.

C.1) При условии, что $9a^2 \neq 4b^2$ единственным решением является точка

$\begin{cases} x_k = \dfrac{-9ac}{9a^2 - 4b^2} \\ y_k = \dfrac{4bc}{9a^2 - 4b^2} \end{cases}, \lambda = \dfrac{72c}{9a^2 - 4b^2}$. Определим ее тип. Достаточным условием существования условного экстремума в условной стационарной точке является постоянство знака второго дифференциала функции Лагранжа при учете условия. Второй дифференциал имеет вид:

$d^2L(x, y, \lambda) = \dfrac{\partial^2}{\partial x^2} L(x, y, \lambda) dx^2 + 2\dfrac{\partial^2}{\partial x \partial y} L(x, y, \lambda) dxdy + \dfrac{\partial^2}{\partial y^2} L(x, y, \lambda) dy^2$ Из ограничения следует, что $dy = -\left(\dfrac{\partial}{\partial x} F(x, y) \Big/ \dfrac{\partial}{\partial y} F(x, y) \right) dx$ Таким образом, тип стационарной точки определяется знаком выражения

$$A = \frac{\partial^2}{\partial x^2} L(x_k, y_k, \lambda_k) + 2\frac{\partial^2}{\partial x \partial y} L(x_k, y_k, \lambda_k)\left(\frac{\partial}{\partial x} F(x_k, y_k) \Big/ \frac{\partial}{\partial y} F(x_k, y_k) \right) + $$
$$+ \frac{\partial^2}{\partial y^2} L(x, y, \lambda)\left(\frac{\partial}{\partial x} F(x_k, y_k) \Big/ \frac{\partial}{\partial y} F(x_k, y_k) \right)^2$$

. В данном случае

$A = 8 - 18\left(\dfrac{a}{b}\right)^2 = 2\dfrac{4b^2 - 9a^2}{b^2}$. Если $9a^2 < 4b^2$, то стационарная точка является точкой минимума, если $9a^2 > 4b^2$ $9a^2 < 4b^2$, то стационарная точка является точкой максимума.

C.2) При условии, что $9a^2 = 4b^2$ возможны два случая. Если *'c=0'* любое значение λ удовлетворяет системе уравнений (*). В этом случае стационарных точек бесконечное количество, и они расположены на прямой $ax + by = 0$, проходящей через начало координат. С учетом связи между параметрами, множество стационарных точек имеет вид $y = \pm\dfrac{2}{3}x$. На данной прямой $Q(x, y) \equiv 0$. В этом случае величина 'A' равна нулю и требуются дополнительные исследования для определения типа стационарных точек. Если $c \neq 0$, то решений системы уравнений (*), т.е. стационарных точек, нет.

Таким образом:
а) Единственная стационарная точка существует, если выполняются условия:
а.1) $a=0, b=0, c=0$, условная стационарная точка совпадает с безусловной стационарной точкой, которая экстремумом не является;
а.2) бы один из параметров 'a' и 'b' не равен нулю и $9a^2 < 4b^2$, стационарная точка является точкой минимума;
а.3) бы один из параметров 'a' и 'b' не равен нулю и $9a^2 > 4b^2$, стационарная точка является точкой максимума.
b) Бесконечное количество стационарных точек существует при выполнении условий $9a^2 = 4b^2$ и $'c=0'$. Для определения их типа требуются дополнительные исследования.
c) Стационарные точки отсутствуют, если выполняются условия $9a^2 = 4b^2, c \neq 0$.

5)

Решите дифференциальное уравнение $y'''(x) - 3y''(x) + y'(x) - 3y(x) + 6x + 13 + 5e^{2x} = 0$.

1. (2 балла) Характеристическое уравнение $\lambda^3 - 3\lambda^2 + \lambda - 3 = 0$
2. (2 балла) Корни, $\lambda = i$, $\lambda = -i$, $\lambda = 3$
3. (2 балла) Общее решение однородного $y(x) = c_1 \cos(x) + c_2 \sin(x) + c_3 e^{3x}$
4. (2 балла) Подбор частного решения вида $y(x) = ax + b + ce^{2x}$
5. (2 балла) Получаем частное решение $y(x) = 2x + 5 + e^{2x}$
6. (бесценно!) Ответ, $y(x) = c_1 \cos(x) + c_2 \sin(x) + c_3 e^{3x} + 2x + 5 + e^{2x}$

---

6) На курсе учатся две группы.

В первой группе – 20 юношей и 10 девушек.
Во второй группе – 15 юношей и 15 девушек.
Мы выбираем одного юношу из всего курса равновероятно наугад. Затем из группы, в которой учится этот юноша, выбираем наугад равновероятно еще одного человека, кроме этого юноши.
А) Какова вероятность того, что будут выбраны люди из первой группы?
Б) Какова вероятность того, что оба выбранных человека будут юношами?
В) Какова вероятность того, что оба - юноши, если известно, что они из первой группы?
Д) Какова вероятность того, что оба из первой группы, если известно, что это юноши?

Решение

1. $P(A) = {}^{20}/_{35}$
2. $P(B) = \frac{20}{35} \cdot \frac{19}{29} + \frac{15}{35} \cdot \frac{14}{29} = \frac{118}{203} \approx 0.5813$
3. $P(A \cap B) = \frac{20}{35} \cdot \frac{19}{29} = \frac{76}{203} \approx 0.3744$
4. $P(A|B) = \frac{76}{118} \approx 0.6441$

5. $P(B|A)=\frac{76}{116}\approx 0.6552$

---

7) Совместная функция плотности случайных величин X и Y имеет вид:

$$f(x,y)=\begin{cases}(x^2+4xy)/34, \{1\le x\le 2; 0\le y\le 3\}\\ 0, \textit{иначе}\end{cases}$$

а) Найдите вероятности $P(Y\le 2)$ и $P(X\le 1.5\,|\,Y\le 2)$.

б) Найдите E(X) и Cov(X,Y).

<u>Решение</u>

а) Найдите вероятности $P(Y\le 2)$ и $P(X\le 1.5\,|\,Y\le 2)$.

Найдем маржинальную плотность распределения для случайной величины Y:

$$f_y(y)=\int_1^2\frac{(x^2+4xy)}{34}dx=\frac{1}{34}\left(\frac{x^3}{3}+2x^2y\right)_1^2=\frac{1}{34}\left(\frac{7}{3}+6y\right).$$

Тогда $P(Y\le 2)=\frac{1}{34}\int_0^2\left(\frac{7}{3}+6y\right)dy=\frac{1}{34}\left(\frac{7}{3}y+3y^2\right)_0^2=\frac{50}{102}\approx 0.49$

Рассчитаем условную вероятность $P(X\le 1.5\,|\,Y\le 2)$ по формуле:

$$P(X\le 1.5\,|\,Y\le 2)=\frac{P(X\le 1.5\cap Y\le 2)}{P(Y\le 2)}.$$

Знаменатель уже рассчитан, осталось найти числитель по формуле:

$$P(X\le 1.5\cap Y\le 2)=\int_0^2 dy\int_1^{1.5}\frac{(x^2+4xy)}{34}dx=\frac{1}{34}\int_0^2\left(\frac{2.375}{3}+2.5y\right)dy=\frac{1}{34}\left(\frac{2.375}{3}y+\frac{2.5y^2}{2}\right)_0^2=\frac{19.75}{102}$$

Тогда $P(X\le 1.5\,|\,Y\le 2)=\frac{P(X\le 1.5\cap Y\le 2)}{P(Y\le 2)}=\frac{19.75/102}{50/102}\approx 0.395$.

б) Найдите E(X) и Cov(X,Y).

Найдем маржинальную плотность распределения для случайной величины X:

$$f_x(x)=\int_0^3\frac{(x^2+4xy)}{34}dy=\frac{1}{34}\left(x^2y+2xy^2\right)_0^3=\frac{1}{34}\left(3x^2+18x\right).$$

Тогда $E(X)=\frac{1}{34}\int_1^2 x\left(3x^2+18x\right)dx=\frac{1}{34}\left(\frac{3x^4}{4}+6x^3\right)_1^2=\frac{213}{136}\approx 1.57$.

Ковариацию будем искать по формуле:

$$Cov(X,Y)=\int_0^3\int_1^2(x-E(x))(y-E(y))\frac{(x^2+4xy)}{34}dxdy=$$

$$E(Y)=\frac{1}{34}\int_0^3 y\left(\frac{7}{3}+6y\right)dy=\frac{1}{34}\left(\frac{7y^2}{6}+6\frac{y^3}{3}\right)_0^3=\frac{129}{68}\approx 1.9$$

Тогда

$$Cov(X,Y)=\int_0^3\int_1^2(x-1.57)(y-1.9)\frac{(x^2+4xy)}{34}dxdy=\frac{1}{34}\int_0^3(15.08+\frac{9.008y}{6}-9.42y^2)dy\approx -0.96.$$

8) Производитель одежды хочет узнать, какой цвет футболок предпочитает целевая группа: малиновый или салатовый. В выборке из 225 человек 90 высказались в пользу малинового цвета, а 135 – в пользу салатового.

а) Рассчитайте 95% доверительный интервал для доли предпочитающих салатовый цвет.

б) Строгий начальник хочет, чтобы ширина доверительного интервала (разница между его верхней и нижней границей) была не больше 0.1. Какой доверительной вероятности можно добиться в таком случае?

Решение

Пользуемся доверительным интервалом для доли при большом объёме выборки:

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Выборочная доля $\hat{p} = 135/225 = 0.6$.

Табличное значение $z_{\frac{\alpha}{2}}$ для уровня доверия $(1-\alpha) = 0.95$: $z_{\frac{\alpha}{2}} = 1.96$.

Получаем:

$$0.6 - 1.96\sqrt{\frac{0.6 \cdot 0.4}{225}} < p < 0.6 + 1.96\sqrt{\frac{0.6 \cdot 0.4}{225}}$$

$0.536 < p < 0.664$ - это ответ на вопрос (а).

Длина доверительного интервала равна $2z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, так что из требования начальника следует неравенство $2z_{\frac{\alpha}{2}}\sqrt{\frac{0.6 \cdot 0.4}{225}} \le 0.1$. Его решение:

$$z_{\frac{\alpha}{2}}\sqrt{\frac{0.6 \cdot 0.4}{225}} \le \frac{0.1}{2\sqrt{\frac{0.6 \cdot 0.4}{225}}} = 1.53.$$

Значение $z_{\frac{\alpha}{2}} = 1.53$ соответствует доверительной вероятности $(1-\alpha) = 0.874$. Таким образом, доверительная вероятность, удовлетворяющая требованиям начальника, не превышает 0.874.

9) При отлаженном процессе упаковки чая в одну упаковку в среднем помещается 125 граммов чая, при этом дисперсия массы чая в упаковке не должна превышать 9 (граммов в квадрате). Отдел контроля качества отобрал 25 упаковок и рассчитал несмещённую оценку дисперсии $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = 13.5$.

Предполагается, что масса чая в упаковке имеет нормальное распределение.

а) Есть ли основания считать, что дисперсия массы чая превышает допустимый предел? Используйте уровень значимости 1%.

б) Гипотезу о равенстве средней массы 125 граммам решено проверять против двусторонней альтернативы с помощью следующего критерия: гипотеза о равенстве не отвергается, если средняя масса чая в выборке лежит в пределах $[123.452;126.548]$. В противном случае основная гипотеза отвергается, процесс упаковки останавливается для переналадки.
Предположим, что настоящая дисперсия массы чая в упаковке равна 9. С какой вероятностью хорошо отлаженный процесс упаковки будет остановлен для переналадки (иначе говоря, произойдёт ошибка первого рода)?

Решение

а) Проверим гипотезу $H_0:\sigma^2=9$ против альтернативы $H_A:\sigma^2>9$. Для этого рассчитаем критическую статистику: $\chi^2=\frac{(n-1)\hat{\sigma}^2}{\sigma_0^2}=\frac{24\cdot 13.5}{9}=36$.

Критическое значение: $\chi^2_{n-1,\alpha}=\chi^2_{24,0.01}=42.98$.

Основная гипотеза отвергается при $\chi^2>\chi^2_{n-1,\alpha}$. В нашем случае этого не происходит, так что проверка не даёт веских оснований считать, что дисперсия массы чая превышает допустимый предел в 9 граммов в квадрате.

б) Пусть процесс упаковки налажен, так что математическое ожидание массы чая в упаковке равно 125 граммам. В этом случае $\bar{X}\sim N\left(125,\frac{9}{25}\right)$. Нам нужно найти вероятность того, что выборочное среднее выйдет за установленные пределы $[123.452;126.548]$.

Центрируем и нормируем выборочное среднее: $Z=\frac{\bar{X}-125}{3/5}\sim N(0,1)$.

Вероятность того, что отдел контроля качества <u>не</u> будет требовать переналадки:

$$P(123.452\le\bar{X}\le126.548)=P\left(\frac{123.452-125}{3/5}\le Z\le\frac{126.548-125}{3/5}\right)=$$

$$=P\left(-2.58\le Z\le 2.58\right)=0.9901.$$

Значит, вероятность того, что отдел ошибочно потребует переналадки, равна $1-0.9901=0.0099$ (приблизительно 1%).

---

10) Директору по персоналу П. Друкеру доступны некоторые данные о <u>семидесяти одном</u> работнике на рынке труда (Таблица 1).

Данные	Переменная	Комментарии
Годовой заработок	salary	$, непрерывная шкала измерений
Опыт работы	experience	годы, непрерывная шкала измерений
Пол	gender	фиктивная (dummy) переменная, женский пол кодируется единицей
Наличие уровня квалификации	level1	фиктивная (dummy) переменная, первый, самый низкий уровень квалификации, наличие кодируется единицей.
	level2	фиктивная (dummy) переменная, средний уровень квалификации, наличие кодируется единицей.

	level3	фиктивная (dummy) переменная, высокий уровень квалификации, наличие кодируется единицей.

Таблица 1. Доступные данные.

<u>Во-первых</u>, он хочет узнать, существует ли на рынке гендерная дискриминация, чтобы оптимизировать издержки по найму нового персонала женского пола. Напомним, что гендерная дискриминация на рынке труда означает наличие неравных возможностей, в частности, в области оплаты труда у сотрудников разного пола, обладающих одинаковыми характеристиками в остальном.

a) Какое из уравнений A – D в <u>наибольшей степени</u> подходит для решения этой задачи Друкера?

b) Каким образом Друкер получил ответ на свой вопрос, если он верит, что случайная составляющая в наилучшей модели имеет нормальное распределение?

A. $salary_t = a_0 + a_1 experience_t + a_2 (experience_t)^2 + a_3 gender_t + v_t$

B. $salary_t = a_0 + a_1 experience_t + a_2 gender_t + v_t$

C. $salary_t = a_0 + a_1 experience_t + a_2 experience_t gender_t + a_3 gender_t + v_t$

D. $salary_t = a_0 + a_1 experience_t + a_2 experience_t gender_t + v_t$

<u>Во-вторых</u>, Друкер хочет узнать, в каких пределах должна находиться справедливая (рыночная) прибавка к годовому заработку женщин, достигших третьего уровня квалификации, по отношению к заработной плате женщин первого уровня квалификации, чтобы удержать их в компании и не переплатить им.

Для решения задачи Друкер построил следующую модель:

$$salary_t = c(1) + c(2) level2_t + c(3) level3_t + c(4) gender_t + c(5) gender_t level2_t + c(6) gender_t level3_t + v_t$$

Друкер верит, что распределение случайной составляющей в модели нормальное. Решение задачи он хочет представить в виде 95% доверительного интервала.

Результаты МНК оценивания значений коэффициентов приведены в Таблице 2

Переменная	МНК оценка	Стандартная ошибка оценки
C	45989,2	1480,0
level2	6951,6	2750,7
level3	8250,2	3744,1
gender	-17197,8	2750,7
gender*level2	8791,1	4182,3
gender*level3	16494,7	5846,3

Таблица 2. МНК оценки коэффициентов модели

В Таблице 3 приведены оценки значений коэффициентов ковариации полученных оценок Ω.

	C	level2	level3	gender	gender*level2	gender*level3
C	2190307,0	-2190307,0	-2190307,0	-2190307,0	2190307,0	2190307,0
level2	-2190307,0	7566515,0	2190307,0	2190307,0	-7566515,0	-2190307,0
level3	-2190307,0	2190307,0	14017966,0	2190307,0	-2190307,0	-14017966,0
Gender	-2190307,0	2190307,0	2190307,0	7566515,0	-7566515,0	-7566515,0
gender*level2	2190307,0	-7566515,0	-2190307,0	-7566515,0	17491823,0	7566515,0
gender*level3	2190307,0	-2190307,0	-14017966,0	-7566515,0	7566515,0	34178747,0

Таблица 3. Оценки коэффициентов ковариации МНК оценок коэффициентов модели.

a) Почему Друкер использовал в модели только два уровня квалификации?

b) Помогите Друкеру найти интересующие его границы надбавки.

Решение

Первая проблема, вопрос (a)

Гендерная дискриминация на рынке труда означает наличие неравных возможностей, в частности, в области оплаты труда у сотрудников разного пола, обладающих одинаковыми характеристиками в остальном. В качестве единственной характеристики Друкеру доступны данные об опыте работы. Таким образом, ему необходимо ответить на вопрос – «Будут ли сотрудники с одинаковым опытом работы, но разного пола, получать одинаковую заработную плату?» Отметим, прежде всего, что все модели A – D содержат константу, значение которой, согласно описанию переменных, равно заработной плате сотрудника мужского пола без опыта работы. Уравнения (A) и (B) содержат слагаемое вида $a * gender_t$. Коэффициент 'a' в этом слагаемом характеризует надбавку или скидку в заработной плате сотрудника женского пола без опыта работы. Однако, структура этих моделей не позволяет описать влияние пола на скорость изменения заработной платы с ростом опыта работы. В тоже время, уравнение (D) может быть преобразовано к виду $salary_t = a_0 + (a_1 + a_2 gender_t) experience_t + v_t$, который позволяет изучить этот вопрос. Однако, в этой модели не учтена возможная надбавка или скидка в заработной плате сотрудника женского пола без опыта работы. Только модель (C) позволяет изучить оба вопроса одновременно. Она наилучшим образом подходит для решения этой проблемы Друкера.

Первая проблема, вопрос (b)

Выбрав модель (C), Друкер получит МНК оценки значений ее параметров, а также стандартные ошибки оценок. Далее он перепишет эту модель в виде $salary_t = a_0 + (a_1 + a_2 gender_t) experience_t + a_3 gender_t + v_t$, из которого следует, что для ответа на его вопрос следует проанализировать значимость оценок значений параметров a_2 и a_3. Значимость оценок он проверит, используя t-критерий Стьюдента, поскольку он верит в то, что случайная составляющая имеет нормальное распределение.

Вторая проблема, вопрос (a)

Друкер включил в модель константу, таким образом, к набору независимых переменных фактически добавляется еще одна, тождественно равная единице.. Три уровня исчерпывают все возможные варианты квалификации, поэтому сумма фиктивных переменных level1, level2 и level3 равна в точности единице. Таким образом, одновременное включение в модель константы и всех трех фиктивных переменных приведет к возникновению мультиколлинеарности и МНК оценки параметров невозможно будет подсчитать.

Вторая проблема, вопрос (b)

Идея решения. Необходимо построить доверительный интервал для истинной (рыночной) величины прибавки к годовому заработку женщин (F3) при достижении третьего уровня квалификации по отношению к первому уровню.

1. Друкер верит, что случайная составляющая в его модели распределена нормально, им получено семьдесят одно измерение и в соответствии со структурой модели необходимо оценить шесть параметров. Таким образом, границы доверительного интервала имеют вид $P\left(F3 \in \left[\hat{F}3 - t_{\frac{1+\varepsilon}{2}}(65) Se\left[\hat{F}3\right], \hat{F}3 + t_{\frac{1+\varepsilon}{2}}(65) Se\left[\hat{F}3\right]\right]\right) = \varepsilon$, где $\hat{F}3$ - оценка искомой прибавки, $Se\left[\hat{F}3\right]$ - стандартная ошибка оценки прибавки, а $t_{\frac{1+\varepsilon}{2}}(65)$ – квантиль распределения Стьюдента с 65 степенями свободы порядка $\frac{1+\varepsilon}{2} = 0{,}975$. В качестве аппроксимации данной величины можно использовать квантиль нормального распределения того же порядка $u_{\frac{1+\varepsilon}{2}} \approx 1{,}96$.

2. Модель следует переписать в виде $salary_t = (c(1) + c(4) gender_t) + (c(2) + c(5) gender_t) level2_t + (c(3) + c(6) gender_t) level3_t + v_t$. Исходя из структуры модели и определения участвующих в ней переменных, величина $(c(1) + c(4) gender_t)$ равна годовому заработку сотрудника первого уровня квалификации, величина $(c(3) + c(6) gender_t)$ равна прибавке к годовому заработку сотрудника при достижении третьего уровня квалификации по отношению к первому уровню. Таким образом, прибавка к заработной плате женщин, достигших третьего уровня квалификации, по отношению к заработной плате женщин первого уровня квалификации определяется как $F3 = c(3) + c(6)$.

3. Исходя из результатов МНК оценивания (Таблица 1), точечная оценка искомой прибавки составляет 24744,9 долларов США.

4. Для построения доверительного интервала потребуется оценка дисперсии данной оценки. Ее можно подсчитать по следующей формуле: $D\left[\hat{F}3\right] = w'\Omega w,\ w \in R^6$, $w(k) = \begin{cases} 1, k = 3,6 \\ 0, k \neq 3,6 \end{cases}$ или более кратко $D\left[\hat{F}3\right] = D[\hat{c}(3)] + 2\operatorname{cov}[\hat{c}(3), \hat{c}(6)] + D[\hat{c}(6)]$.

5. Исходя из результатов МНК оценивания (Таблица 2), оценкой искомой дисперсии служит величина D3 = 14017966,00 - 2*14017966,00 + 34178747,00 = 20160781. Оценка стандартной ошибки равна квадратному корню из дисперсии Se3=4490,08.

6. Таким образом, границы искомого доверительного интервала приближенно равны: нижняя 24744,9 – 1,96*4490,08 = 15944,35, верхняя 24744,9 + 1,96*4490,08 = 33545,45