Задание по теории вероятностей

Инвестор покупает ценные бумаги за счет кредита, взятого на год под 10%. Доходность бумаг моделируется случайной величиной ξ с математическим ожиданием 15% и стандартным отклонением $2\sqrt{3}\%$ годовых.

1. Через год инвестор продает указанные ценные бумаги и планирует погасить кредит из доходов от их продажи. Оцените вероятность того, что он не сможет этого сделать:
   a. предполагая, что доходность распределена нормально;
   b. предполагая, что доходность распределена равномерно;
   c. не имея никаких сведений о характере распределения доходности;
   d. не имея никаких сведений о характере распределения доходности, но зная, что ее распределение непрерывно и симметрично относительно математического ожидания.
2. Пусть доходность ξ распределена равномерно. Взнос инвестора на благотворительные цели в у.е. η представляет собой случайную величину, чье распределение связано с распределением доходности ξ соотношением $\eta = \xi^{0,5} - 2$. Оцените математическое ожидание взноса на благотворительность.

Решение задания по теории вероятностей

1. ξ – доходность ценных бумаг (%).

   $M\xi = 15$

   $D\xi = \left(2\sqrt{3}\right)^2 = 12$

   Искомая вероятность равна $\mathrm{P}\left(1 + \frac{\xi}{100} < 1 + \frac{10}{100}\right) = \mathrm{P}(\xi < 10)$, то есть инвестор не сможет погасить кредит из доходов от продажи ценных бумаг, если доходы будут меньше 10%.

   a. $\xi \sim \mathrm{N}(15; 12)$

   $\mathrm{P}(\xi < 10) = \mathrm{P}\left(\frac{\xi - 15}{2\sqrt{3}} < -1{,}44\right) = 0{,}5 + \Phi_0(-1{,}44) = 0{,}0745$

   (здесь Φ_0 – функция Лапласа)

   b. $\xi \sim \mathrm{U}(15; 12)$

   Найдем параметры a и b, решив систему:

   $$\begin{cases} \xi = \dfrac{a+b}{2} = 15 \\ D\xi = \dfrac{(b-a)^2}{12} = \left(2\sqrt{3}\right)^2 \end{cases} \Rightarrow \begin{cases} a = 9 \\ b = 21 \end{cases}$$

   Итак, $\xi \sim \mathrm{U}(9; 21)$

   $$\mathrm{P}(\xi < 10) = \frac{1}{12} = 0{,}0833$$

   c. Мы не знаем распределение ξ, поэтому не можем посчитать точное значение вероятности. Но оценить ее можно, например, используя неравенство Чебышева:

   $$\mathrm{P}(\xi < 10) = P(\xi - M\xi < -5) \le P(|\xi - M\xi| > 5) \le P(|\xi - M\xi| \ge 5) \le \frac{D\xi}{25} = 0{,}48$$

   d. Решение аналогично предыдущему пункту, только дополнительно нужно учесть, что $P(\xi - M\xi < -5) = P(\xi - M\xi > 5) = \frac{1}{2}P(|\xi - M\xi| > 5)$, а также тот факт, что для непрерывного распределения $P(|\xi - M\xi| > 5) = P(|\xi - M\xi| \ge 5)$. Окончательно выкладки имеют следующий вид:

$$\mathrm{P}(\xi < 10) = P(\xi - M\xi < -5) = \frac{1}{2}P(|\xi - M\xi| > 5) = \frac{1}{2}P(|\xi - M\xi| \geq 5) \leq \frac{D\xi}{50} = 0{,}24$$

2. Как мы определили в пункте b, доходность имеет распределение $\xi \sim U(9;21)$. Найдем функцию распределения случайной величины $\eta = \xi^{0,5} - 2$. Поскольку η принимает значения от 1 до $\sqrt{21} - 2$, то для $y \in [1; \sqrt{21} - 2]$ верно:
$$F_\eta(y) = P(\eta < y) = P(\xi^{0,5} - 2 < y) = P(\xi < (2+y)^2) = F_\xi((2+y)^2)$$
Найдем плотность η:
$$p_\eta(y) = \begin{cases} 2(2+y)p_\xi((2+y)^2), \text{если } y \in [1; \sqrt{21} - 2] \\ 0, \text{иначе} \end{cases}$$
Окончательно имеем:
$$p_\eta(y) = \begin{cases} \dfrac{2+y}{6}, \text{если } y \in [1; \sqrt{21} - 2] \\ 0, \text{иначе} \end{cases}$$
Найдем математическое ожидание:
$$M\eta = \int_1^{\sqrt{21}-2} \frac{(2+y)y}{6} dy \approx 1{,}85 \text{ (у.е.)}$$

Задание по математической статистике

Аналитик страховой компании хочет оценить правильность компенсации оплаты медицинских услуг. Для проведения анализа из генеральной совокупности им были извлечены 100 документов, при этом 18 из них оказались неверно оплаченными. Средний размер компенсации среди верно оплаченных документов составил 116 у.е. с исправленным среднеквадратическим отклонением 45 у.е.

1. Можно ли на уровне значимости 1% утверждать, что доля неверно оплаченных компенсаций больше 15%?
2. Как изменится ответ на вопрос 1, если число неверно оплаченных документов вырастет до 24 единиц?
3. Предположим, что размер компенсации среди верно оплаченных документов распределен нормально. Можно ли на уровне значимости 5% утверждать, что средний размер такой компенсации меньше 120 у.е.?
4. Как изменится ответ на вопрос 3, если выборочное среднее увеличится?
5. Теперь предположим, что размер верно оплаченной компенсации распределен показательно с параметром $\lambda = \theta^2$. Оцените параметр θ методом максимального правдоподобия. Будет ли полученная Вами оценка состоятельной?

Решение задания по математической статистике

В задаче фигурируют две генеральные совокупности.

Первая генеральная совокупность $\left(\xi = \begin{cases} 1, \text{если оплачено неверно} \\ 0, \text{если оплачено верно} \end{cases}\right)$ имеет распределение Бернулли с параметром p, где p – вероятность того, что документ оплачен неверно. Из этой ГС взята выборка $x_1, \ldots, x_{100}$ объема $n = 100$, сумма элементов выборки равна числу неверно оплаченных документов, т.е. равна $\sum x_i = 18$, поэтому $\bar{x} = 0{,}18$.

Вторая генеральная совокупность (η – размер верно оплаченной компенсации) имеет неизвестное нам распределение. Из этой ГС взята выборка $x_1, \ldots, x_{82}$ объема $n = 82$, среднее значение по выборке равно среднему размеру компенсации среди верно оплаченных документов: $\bar{x} = 116$. Кроме того, нам дано исправленное среднеквадратическое отклонение $s = 45$.

1. В пункте один спрашивают о доле неверно оплаченных компенсаций, т.е. речь идет о первой генеральной совокупности. Сформулируем гипотезы:

$H_0: p = 0{,}15$
$H_1: p > 0{,}15$
При условии, что верна нулевая гипотеза, распределение статистики

$$u_{\text{расчетное}} = \frac{\bar{x} - 0{,}15}{\sqrt{0{,}15 * (1 - 0{,}15)}}\sqrt{n}$$

является асимптотически нормальным:
$u_{\text{расчетное}} \xrightarrow{d} N(0{,}1)$
Поскольку объем выборки достаточно велик, мы можем воспользоваться этим и принять решение, сравнив $u_{\text{расчетное}}$ с односторонней критической точкой стандартного нормального распределения при уровне значимости $\alpha = 1\%$:

$$u_{\text{критическое}} = \Phi_0^{-1}\left(\frac{1}{2} - \alpha\right)$$

Имеем:

$$\left.\begin{aligned} u_{\text{расчетное}} &= \frac{0{,}18 - 0{,}15}{\sqrt{0{,}15 * (1 - 0{,}15)}}\sqrt{100} = 0{,}84 \\ u_{\text{критическое}} &= 2{,}33 \end{aligned}\right\} \Longrightarrow \text{принимаем } H_0$$

Итак, на уровне значимости 1% нельзя утверждать, что доля неверно оплаченных компенсаций больше 15%.

2. В этом случае $u_{\text{расчетное}} = \frac{0{,}24 - 0{,}15}{\sqrt{0{,}15*(1-0{,}15)}}\sqrt{100} = 2{,}52$, это больше, чем $u_{\text{критическое}} = 2{,}33$, и теперь на уровне значимости 1% можно утверждать, что доля неверно оплаченных компенсаций превышает 15%.

3. В пункте три спрашивают о размере верно оплаченной компенсации, т.е. речь идет о второй генеральной совокупности. Мы предполагаем, что она распределена нормально с параметрами $\eta \sim N(a, \sigma^2)$, и формулируем гипотезы относительно ее математического ожидания:
$H_0: a = 120$
$H_1: a < 120$
При условии, что верна нулевая гипотеза, распределение статистики

$$u_{\text{расчетное}} = \frac{\bar{x} - 120}{s}\sqrt{n}$$

является асимптотически нормальным:
$u_{\text{расчетное}} \xrightarrow{d} N(0{,}1)$
Поскольку объем выборки достаточно велик, мы можем воспользоваться этим и принять решение, сравнив $u_{\text{расчетное}}$ с односторонней критической точкой стандартного нормального распределения при уровне значимости $\alpha = 5\%$:

$$u_{\text{критическое}} = \Phi_0^{-1}\left(\frac{1}{2} - \alpha\right)$$

Имеем:

$$\left.\begin{aligned} u_{\text{расчетное}} &= \frac{116 - 120}{45}\sqrt{82} = -0{,}8 \\ -u_{\text{критическое}} &= -1{,}65 \end{aligned}\right\} \Longrightarrow \text{принимаем } H_0$$

Итак, на уровне значимости 5% нельзя утверждать, что средний размер такой компенсации меньше 120 у.е.

4. В этом случае (при росте выборочного среднего) $u_{\text{расчетное}}$ тоже вырастет и, конечно, останется больше, чем $(-1{,}65)$, поэтому вывод не изменится.
5. В пункте пять мы предполагаем, что размер верно оплаченной компенсации распределен показательно с параметром $\lambda = \theta^2$, таким образом, плотность элементов выборки имеет вид:

$$p_\eta(x) = \theta^2 e^{-\theta^2 x}, x > 0$$

Чтобы оценить параметр θ методом максимального правдоподобия, составим функцию правдоподобия, прологарифмируем ее и решим задачу ее максимизации:

$$L(x_1, \dots, x_n, \theta) = \theta^{2n} e^{-\theta^2 \sum x_i}, \text{все } x_i \geq 0$$
$$lnL(x_1, \dots, x_n, \theta) = 2nln\theta - \theta^2 \sum x_i, \text{все } x_i \geq 0$$
$$lnL(x_1, \dots, x_n, \theta) \longrightarrow \max_\theta$$
$$\frac{\partial lnL}{\partial \theta} = \frac{2n}{\theta} - 2\theta \sum x_i = 0 \Longrightarrow \hat{\theta}_{\text{ММП}} = \sqrt{\frac{n}{\sum x_i}} = \sqrt{\frac{1}{116}} \approx 0{,}09$$

Это точка максимума, поскольку

$$\frac{\partial^2 lnL}{\partial \theta^2} = \frac{-2n}{\theta^2} - 2\sum x_i < 0 \text{ (т. к. } x_i \geq 0)$$

Задание по эконометрике

В вашем распоряжении имеются следующие данные о 540 работниках (270 мужчин и 270 женщин):
l_EARNINGS — логарифм текущего часового заработка в долларах США,
S — продолжительность обучения (число полных лет обучения),
EXP — общий стаж работы после окончания учебы,
FEMALE — пол респондента (0 — для мужчин, 1 — для женщин).
Ваша цель состоит в том, чтобы выявить влияние опыта работы и образования на доход индивида.
Ниже представлены результаты оценивания некоторой модели:

Модель 1: МНК, использованы наблюдения 1-540
Зависимая переменная: l_EARNINGS
Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1

	Коэффициент	*Ст. ошибка*	*t-статистика*	*P-значение*
const	0,7863	0,2204	3,567	0,0004
EXP	**??? (A)**	0,0081	3,637	0,0003
S	0,1200	0,0093	12,78	0,0001
FEMALE	−0,2258	0,1736	−1,301	0,1938
EXP*FEMALE	−0,0258	**??? (B)**	**??? (C)**	0,0100

Среднее зав. перемен	2,791993	Ст. откл. зав. перемен	0,588554
Сумма кв. остатков	126,2388	Ст. ошибка модели	**??? (D)**
R-квадрат	0,323869	Испр. R-квадрат	**??? (E)**

Тест Вайта (White) на гетероскедастичность -
Тестовая статистика: LM = 8,02854
p-значение = P(Хи-квадрат(10) > 8,02854) = 0,626049

(a) Заполните отмеченные знаками «???» пропуски в таблице. Обоснуйте ответ.

(б) Дайте содержательную интерпретацию коэффициента при переменной S.

(в) Дайте содержательную интерпретацию коэффициента при переменной EXP.

(г) Интерпретируйте представленные результаты теста Уайта. Опишите процедуру проведения этого теста.

(д) Ожидаете ли вы, что коэффициент при переменной S оценен состоятельно? Аргументируйте свой ответ, описав какой-либо возможный источник эндогенности переменной S в рассматриваемой модели.

Решение задания по эконометрике

(а) Во-первых, $A = 3{,}637 * 0{,}0081 = 0{,}03$.
Отметим, что в нашем случае выборка достаточно велика, чтобы использовать асимптотические результаты.
В строчке EXP*FEMALE P-значение равно в точности 0,01. Следовательно, в соответствии с таблицей стандартного нормального распределения t-статистика по модулю равна 2,58. Таким образом, $\text{C} = -2{,}58$ (не забываем поставить минус, так как коэффициент отрицательный).
Тогда стандартная ошибка коэффициента:

$$B = -\frac{0{,}0258}{\text{C}} = -\frac{0{,}0258}{-2{,}58} = 0{,}01$$

$$D = SEE = \sqrt{\frac{1}{n-k} * \sum_{i=1}^{n} e_i^2} = \sqrt{\frac{1}{540-5} * 126{,}2388} = 0{,}49$$

$$E = R_{adj}^2 = R^2 - \frac{k-1}{n-k} * (1 - R^2) = 0{,}323869 - \frac{5-1}{540-5} * (1 - 0{,}323869) = 0{,}319$$

(б) Каждый дополнительный год образования в среднем при прочих равных условиях увеличивает доход на $(e^{0{,}12} - 1) * 100\% = 13\%$.

(в) Каждый дополнительный год опыта работы в среднем при прочих равных условиях увеличивает доход ***мужчины*** на 3%.

(г) При проведении теста оказалось, что $\text{p} - \text{значение} = 0{,}626$, что больше любого разумного уровня значимости. Следовательно, принимается нулевая гипотеза, которая говорит об отсутствии гетероскедастичности.
Процедура теста Уайта:
H_0: $\sigma_1^2 = \cdots = \sigma_n^2$ (отсутствие гетероскедастичности)
H_1: *наличие гетероскедастичности.*
Шаг 1: оцениванием исходную модель $y_i = \beta_1 + \beta_2 * x_i^{(2)} + \cdots + \beta_k * x_i^{(k)} + \varepsilon_i$ по n наблюдениям и сохраняем из нее e_i^2 (квадраты остатков модели, где $e_i = y_i - \hat{y}_i$).
Шаг 2: строим регрессию e_i^2 по всем регрессорам исходной модели, их квадратам, попарным произведениям и константе, даже если ее не было в исходной модели. Сохраняем R^2.
Шаг 3: случайная величина nR^2 имеет асимптотическое распределение $\chi^2(p)$, где p – число переменных в регрессии e_i^2 без учета константы.

(д) В данном случае оценка может быть несостоятельной из-за пропуска существенной переменной. Например, талант индивида может влиять на доход и при этом быть коррелирован с образованием.