

Семинар 1

Задание 1.

Совместный закон распределения случайных величин (далее – сл.в.) X и Y задан следующей таблицей.

$X \backslash Y$	0	1
1	0.23	0.17
2	0.32	0.09
3	0.12	...

Определите

1. условное математическое ожидание $E(X|Y = 1)$
2. условную вариацию $Var(X|Y = 1)$
3. условное математическое ожидание $E(Y|X = 1)$
4. условную вариацию $Var(Y|X = 1)$

Задание 2.

Представьте, что Вы хотите выявить, как утренняя гимнастика влияет на сонливость днем. Участники исследования были разделены на 2 группы: группу воздействия (им нужно было выполнять гимнастику по утрам) и контрольную (они не выполняли утреннюю гимнастику). Среди участников были и взрослые, и дети. Ниже дана таблица совместного распределения по итогам исследования, где $Y = 1$ – отсутствие сонливости днем, $Y = 0$ – испытывал сонливость днем.

Группа / Подвыборка	Взрослые		Дети	
	$Y = 1$	$Y = 0$	$Y = 1$	$Y = 0$
Группа воздействия	0.15	0.225	0.1	0.025
Контрольная группа	0.0375	0.0875	0.2625	0.1125

Найдите

1. средний эффект воздействия утренней гимнастики на дневную сонливость по всей выборке в целом
2. средний эффект воздействия утренней гимнастики на дневную сонливость для взрослых
3. средний эффект воздействия утренней гимнастики на дневную сонливость для детей
4. сопоставьте полученные результаты, заметили ли Вы какие-нибудь расхождения? Если да, то с чем они связаны? Критически порассуждайте о процедуре деления участников

Задание 3. В таблице ниже представлены данные показателя оценки негарантированности занятости (job insecurity) и возраста (в годах) по 10 работникам предприятия. Первый показатель представлен в шкале от 0 до 10, где 10 – наиболее высокий уровень неуверенности работника в сохранении своей занятости.

1. Оценивается парная линейная регрессия восприятия job insecurity на возраст работника предприятия. Рассчитайте оценки константы и коэффициента при предикторе (объясняющей переменной)

Работник предприятия	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>	<i>H</i>	<i>I</i>	<i>J</i>
Восприятие job insecurity	4	2	1	9	4	6	8	2	5	9
Возраст	40	36	32	45	50	60	62	25	42	58

2. Запишите с помощью полученных оценок спецификацию модели и проинтерпретируйте оценки коэффициентов
3. Определите, являются ли статистически значимыми оценки коэффициентов в указанной модели, приняв 5%-ный фиксированный уровень значимости. Известно, что стандартная ошибка оценки константы равна 2.437, а оценка вариации выборочного коэффициента при предикторе составляет 0.0027. Запишите исходную формулировку нулевой гипотезы. Используйте двустороннюю альтернативу. Реализуйте последовательные шаги проверки гипотезы, сопроводите каждый шаг пояснениями. Определите границы доверительной области. Сделайте на основе полученных результатов вывод
4. Выполните предыдущий пункт, но на основе значений p-value. Сделайте вывод
5. Постройте 95%-ые доверительные интервалы для коэффициентов данной модели и проинтерпретируйте их
6. Рассчитайте коэффициент детерминации (R^2) и проинтерпретируйте полученное значение

Задание 4. В регрессионной модели оценка коэффициента при предикторе составляет 0.15, а оценка константы равна 4. Ниже также дана ковариационная матрица оценок коэффициентов для указанной модели.

	(Intercept) x	
(Intercept)	1.6111	-0.2068
x	-0.2068	0.0323

В общем виде проинтерпретируйте оценки коэффициентов. На основе значений элементов этой ковариационной матрицы проверьте, значимы ли коэффициенты при предикторе и константа на основании как фиксированного уровня значимости 1%, так и p-value. Используйте нормальную аппроксимацию. Для этого выпишите нулевую гипотезу и альтернативу для каждого коэффициента, статистики для проверки гипотез, распределение статистик при верной нулевой гипотезе. Сделайте вывод.

Задание 5. Построена регрессия индекса потребительских цен на уровень безработицы. Для оценки модели использовались данные 50 стран мира. Несмещенная выборочная оценка дисперсии индекса потребительских цен равна 800, а сумма квадратов остатков регрессии равна 25000. Рассчитайте коэффициент детерминации.