**Диалог №10 (Итог по Лекции 6)**

55 **Приведены несколько высказываний. Для каждого из них укажите, чем оно является – одним из Предположений модели I-V или одним из возможных несоответствий этим предположениям. Укажите, что означает каждое из высказываний (некоторые комментарии уже приведены):**

$$M(\varepsilon_i)=0,\ i=1,...n;$$

**– это …**

(1) что означает, (2) предположение из числа I-V или несоответствие?

$$Var(\varepsilon_i)=\sigma_i^2=\sigma^2=const,\ i=1,...n;$$

**– это …**

(1) что означает, (2) предположение из числа I-V или несоответствие?

$$Cov(\varepsilon_i,\varepsilon_j)=0,\ i\neq j,\ i,j=1,...n.$$

**– это …**

(1) что означает, (2) предположение из числа I-V или несоответствие?

$$Var(\varepsilon_i)=\sigma_i^2\neq const,\ i=1,...n,$$

**– это гетероскедастичность (другое название – неоднородная дисперсия ошибок), это …**

(предположение из числа I-V или несоответствие?)

**Существует такая пара** $i\neq j,\ i,j\in\{1,...n\}$**, что** $Cov(\varepsilon_i,\varepsilon_j)\neq 0,$

**– это означает, что модель содержит ………. ошибки (хотя бы пара случайных величин линейно статистически зависит друг от друга), это …………**

(предположение из числа I-V или несоответствие?)

**Для любых** $i\neq j,\ i,j=1,...n,\ Cov(\varepsilon_i,\varepsilon_j)\neq 0$

**– это ……………… ошибок (расшифровка – линейная статистическая зависимость каждой ошибки от всех остальных ошибок), это …………**

(предположение из числа I-V или несоответствие?)

56 **Укажите, в каких случаях можно встретить гетероскедастичные ошибки (то есть ошибки с разными дисперсиями), и в каких случаях – коррелированные ошибки (то есть линейно статистически зависящие друг от друга):**

**Если разброс изучаемого признака (отклика) зависит от объясняющих переменных (факторов), как правило, наблюдаются …….**
(так работают измерительные приборы, гарантирующие постоянную относительную погрешность)

**В экономических приложениях: если значения фактора и отклика регистрируются по времени, как правило, наблюдаются ……**

**57** **Чтобы выявить свойства ошибок, используют два основных приема (они показаны ниже). Укажите, какой из приемов служит для выявления неоднородной дисперсии ошибок, и какой – для выявления автокорреляции:**

**И в том, и в другом случае модель $Y = \beta_0 + \beta_1 X + \varepsilon$ сначала оценивается методом наименьших квадратов. Затем:**

**1) Квадраты остатков $\hat{\varepsilon}^2$ используются как отклик, а фактор $X$ как фактор, и при различных значениях $\eta$ на базе пар значений $(X_i, \hat{\varepsilon}_i)$, $i=1,...n,$ строят уравнения вида $\hat{\varepsilon}^2 = c_0 + c_1 X^{\eta}$**

**Значимое уравнение из этого класса с наибольшим (за счет подбора $\eta$) коэффициентом детерминации $R^2$ используют для того, чтобы…**

(сформулируйте, как это используется)

**2) На основе значений остатков (наблюдения должны быть предварительно упорядочены) исследуются частные коэффициенты корреляции**

$r_{\varepsilon(t)\varepsilon(t\text{-}s)\,/\,\varepsilon(t\text{-}1)\varepsilon(t\text{-}2)...\varepsilon(t\text{-}s+1)}$

**то есть такие зависимости остатков $\varepsilon(t)$ и $\varepsilon(t\text{-}s)$, удаленных друг от друга на $s$ шагов, из которых влияние «промежуточных» остатков $\varepsilon(t\text{-}1)$, $\varepsilon(t\text{-}2)$, $...\varepsilon(t\text{-}s+1)$ исключено.**

**Частные коэффициенты корреляции используют для того, чтобы построить и определить……**

(сформулируйте, как это используется)

**58** **Далее представлена модель автокорреляции 1-го порядка. Для модели линейной регрессии (с помощью индекса $t = 1, ...n$ перечислены все точки замеров)**

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t=1, ...n,$$

**заявлена линейная связь ошибок на соседних шагах:**

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + \nu_t, \quad t=1, ...n.$$

**Здесь $\rho$ является постоянной величиной, причем $|\rho| < 1$.**
**Величины $\varepsilon_0$ и $\nu_t$, $t=1, ...n,$ являются случайными.**
**Указанные величины так же, как и $\varepsilon_t$, $t=1, ...n,$ называют ошибками.**
**Требования к ошибкам $\varepsilon_0$ и $\nu_t$ сформулированы в Предположениях А-Е. Свойства ошибок $\varepsilon_t$, $t=1, ...n$ доказаны в Утверждении 6.1.**

**Запишите, в каком случае говорят о положительной автокорреляции ошибок, отрицательной автокорреляции ошибок, об отсутствии автокорреляции?**

**Если $\rho$ ……….автокорреляция (1-го порядка) отсутствует**

**Если $\rho$ ……….автокорреляция (1-го порядка) положительная**

**Если $\rho$ ……….автокорреляция (1-го порядка) отрицательная**

**59** **Если модель линейной регрессии**

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t=1, \ldots n,$$

**содержит ошибки $\varepsilon_t, \quad t=1, \ldots n,$ которые соответствуют модели автокорреляции 1-го порядка, но сама модель**

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t=1, \ldots n,$$

**сначала была оценена на основании «обычного» МНК, коэффициент автокорреляции $\rho$ можно оценить, используя остатки:**

$$\hat{\rho} = \frac{\sum_{t=2}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=2}^{n} \hat{\varepsilon}_t^2}$$

**Если значение оценочного $\hat{\rho}$ больше нуля, ставится нулевая гипотеза об отсутствии автокорреляции против альтернативы о положительной автокорреляции.**
**Если значение оценочного $\hat{\rho}$ меньше нуля, ставится нулевая гипотеза об отсутствии автокорреляции против альтернативы об отрицательной автокорреляции.**

**И в том, и в другом случае тест проводят по статистике Дарбина-Уотсона, обозначение $DW$ (формула есть в лекции)**

**Какое решение следует принять (на некотором заранее выбранном уровне значимости $\alpha$) в следующих трех случаях?**

$DW \approx 2$
**Нулевая гипотеза (на уровне значимости $\alpha$)…………..…….,**
**автокорреляция (1-го порядка) ….**

$DW \approx 4$
**Нулевая гипотеза (на уровне значимости $\alpha$)…………………,**
**автокорреляция (1-го порядка) ….**

$DW \approx 0$
**Нулевая гипотеза (на уровне значимости $\alpha$)…………………,**
**автокорреляция (1-го порядка) ….**

**Потому что статистика Дарбина-Уотсона $DW$ и оценочный коэффициент автокорреляции $\hat{\rho}$ связаны соотношением**

(далее дополнить текст)

$DW \approx \ldots\ldots\ldots\ldots\ldots\ldots\ldots.$

**Если $\hat{\rho} \approx 1$, тогда $DW \approx \ldots$**
**Если $\hat{\rho} \approx -1$, тогда $DW \approx \ldots$**
**Если $\hat{\rho} \approx 0$, тогда $DW \approx \ldots$**

**Правило принятия решения соответствует ответу на вопрос 58.**

**60** **Рассмотрим Практикум 4. Методом пошагового включения (с исключением) была построена и оценена модель, объясняющая объемы продаж на основе 4-х факторов:**

$$sales = \beta_0 + \beta_1 \times r\_seller + \beta_2 \times r\_sector + \beta_3 \times r\_byers + \beta_4 \times price + \varepsilon$$

**Приведем вид МНК-уравнения без указания значений МНК-коэффициентов:**

$$sales(\text{с крышкой}) = b_0 + b_1 \times r\_seller + b_2 \times r\_sector + b_3 \times r\_byers + b_4 \times price$$

**В протоколе SPSS (Таблица 7.15) для данной модели вычислена статистика Дарбина Уотсона:** $DW = 2{,}31 \approx 2$.

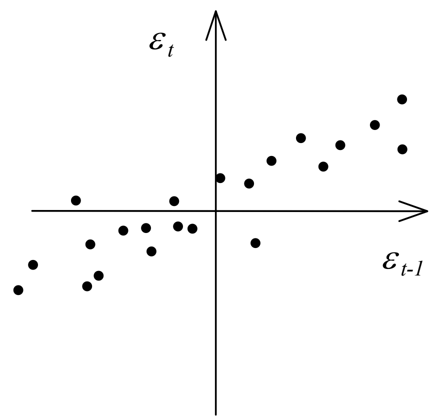
**Это означает, что в данной модели…..**

(сформулируйте соответствующее свойство ошибок (остатков))

**61** **Укажите, какие графики остатков свидетельствуют о наличии коррелированных ошибок, и какие – о неоднородной дисперсии ошибок**


А)


Б)




В)

**Что делать:** **прочесть текст, заполнить позиции 55-61, получить представление о позициях 62-69** (отвечать на 62-69 не требуется)

**62** **Пусть для модели вида**

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i=1,...n,$$

**выполнены Предположения I, II, IV и V, но не выполнено III, т.е.**

$$Var(\varepsilon_i) = \sigma_i^2 \neq const,\ i = 1,...n.$$

**Тогда для оценочных уравнений**

$$\hat{Y}_i = b_0 + b_1 X_i,\ i=1,...n$$

**коэффициенты** $b_0$ и $b_1$ **подбирают таким образом, чтобы**

$$S(b_0, b_1) = \sum_{i=1}^{n} \left(\frac{Y_i - \hat{Y}_i}{\sigma_i}\right)^2 \rightarrow min \text{ при } b_0 \in R,\ b_1 \in R.$$

**Как называется этот метод?**

**Как называется ситуация, когда он применяется?**

**Как узнать неизвестные** $\sigma_i,\ i=1,...n$ **?**

**Какая замена переменных (факторов и (или) отклика) в основе этого метода?**

**63** **Пусть для отклика** $Y$ **и факторов** $X^{(j)}, j=1,...K,$ **предложена мультипликативная модель**

$$Y = \gamma_0 [X^{(1)}]^{\gamma(1)} [X^{(2)}]^{\gamma(2)} ... [X^{(K)}]^{\gamma(K)}\ \nu,$$

**где** $\nu$ **- мультипликативная ошибка, то есть случайная величина, для которой при фиксированных значениях** $X^{(j)}, j=1,...K$

$$M(\nu) = 1.$$

**Как называется прием, который позволит оценить параметры модели?**

**Каким требованиям должна соответствовать случайная величина**

$$\varepsilon = ln\,\nu$$

**чтобы параметры** $\gamma(1),\ \gamma(2),\ \dots\ \gamma(K)$ **можно было оценить с помощью МНК?**

**Каким способом будет получен параметр** $\gamma(0)$?

**Какая замена переменных (факторов и отклика) является основой этого метода?**

**64 Рассмотрим модель вида**

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t=1,\ldots n,$$

**где ошибки** $\varepsilon_t, \quad t=1,\ldots n,$ **описываются уравнением автокорреляции 1-го порядка**

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + \nu_t, \quad t=1,\ldots n,$$

**и Предположения А-Е выполнены.**

**Оценкой параметров** $\beta_0$ **и** $\beta_1$ **являются (не-МНК) коэффициенты** $b_0$ **и** $b_1$**, определяемые следующим образом:**

$$b_0 = b_0^* /(1 - \rho),$$

$$b_1 = b_1^*.$$

**В этих формулах коэффициенты со звёздочками** $b_0^*$ **и** $b_1^*$**, являются МНК-оценкой параметров другой модели, с другим фактором и другим откликом.**

**Другая модель имеет вид**

$$Y_t^* = \beta_0^* + \beta_1^* X_t^* + \nu_t, \quad t=2,\ldots n,$$

**ее МНК-оценка, то есть уравнение вида**

$$\hat{Y}^* = b_0^* + b_1^* X^*$$

**строится по данным для нового фактора и нового отклика со звёздочками**

$$(X_t^*, Y_t^*),\ t=2,\ldots n.$$

**Как называется ситуация, в которой применяется данный метод?**

**Какая замена переменных (фактора и отклика) в основе этого метода?**

**Как узнать неизвестное значение** $\rho$ **?**

**65 Для проверки нулевой гипотезы о «равенстве дисперсии ошибок» против альтернативы о том, что «при малых и больших значениях фактора дисперсии ошибок различны: одна из двух дисперсий больше другой», используют проверку Голдфелда-Квандта.**

**Для этого совокупность объектов (наблюдений) упорядочивают по возрастанию фактора.**

**Затем исключают из рассмотрения некоторое число так называемых центральных наблюдений.**

**Получают две оценки неизвестной дисперсии ошибок: на основе МНК-уравнения, построенного по объектам (наблюдениям) с низким значениям фактора, и МНК уравнения, построенного по объектам (наблюдениям) с высокими значениями фактора.**

**Статистика проверки (если нулевая гипотеза верна) имеет распределение Фишера с числом степеней свободы** $\left(\frac{n-C}{2}-2\,;\frac{n-C}{2}-2\right)$,

**где** $n$ **— объем совокупности (число объектов или наблюдений),** $C$ **— количество центральных наблюдений, изъятых из анализа,** $\frac{n-C}{2}$ **– одинаковые объемы совокупностей с низким и высоким значением фактора.**

**Как вычисляется значение статистики проверки?**

(формулы и текст)

**Как нужно формулировать решающее правило?**

(формулы и текст)

**При каких условиях применяют проверку?**

(текст)

66 **В общем случае, если модель включает несколько факторов** $X^{(j)},\ j=1,\ldots K,$ **проверка равенства дисперсии ошибок проводится по каждому фактору отдельно.**

**В случае нескольких факторов (а также в рассмотренном выше случае с одним фактором, когда** $K$=1**) отобранные для проверки совокупности могут иметь разные объемы.**

**Пусть в совокупности с индексом 1 оценка дисперсии ошибок оказалась больше аналогичной оценки по совокупности 2.**

**Статистика проверки (в случае, если нулевая гипотеза верна) имеет распределение Фишера с числом степеней свободы**

$(n_1-K-1;\, n_2-K-1)$

**где** $n_1$ **и** $n_2$ **— объемы совокупностей с индексами 1 и 2,** $K$ **— число факторов.**

**Как вычисляется значение статистики проверки?**

(формулы и текст с пояснением)

**Как нужно формулировать решающее правило?**

(формулы и текст)

**Что поменяется, если оценка дисперсии ошибок по совокупности с индексом 2 оказалась больше аналогичной оценки по совокупности с индексом 1?**

(формулы и текст)

**67** **Объясните, как вычисляется частный коэффициент корреляции** $r_{\varepsilon(t)\varepsilon(t\text{-}s)\,/\,\varepsilon(t\text{-}1)\varepsilon(t\text{-}2)\ldots\varepsilon(t\text{-}s+1)}$

**Строят регрегрессию (МНК-уравнение) остатка $\varepsilon(t)$ на регрессоры** $\varepsilon(t\text{-}1),\ \varepsilon(t\text{-}2),\ \ldots\varepsilon(t\text{-}s+1)$

**Строят регрегрессию (МНК-уравнение) остатка $\varepsilon(t\text{-}s)$ на те же самые регрессоры** $\varepsilon(t\text{-}1),\ \varepsilon(t\text{-}2),\ \ldots\varepsilon(t\text{-}s+1)$

**Чтобы вычислить указанный выше частный коэффициент корреляции, нужно вычислить ……,**

(укажите, что нужно вычислить)

**68** **Вернемся к модели автокорреляции 1-го порядка.**
**Для модели линейной регрессии (с помощью индекса $t = 1, \ldots n$ перечислены все точки замеров)**

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, \quad t=1, \ldots n,$$

**заявлена линейная связь ошибок на соседних шагах:**

$$\varepsilon_t = \rho\, \varepsilon_{t\text{-}1} + \nu_t\ , \quad t=1, \ldots n.$$

**Коэффициент $\rho$ является постоянной величиной, $|\rho| < 1$ и называется коэффициентом автокорреляции.**

**Ошибки $\varepsilon_0$ и $\nu_t$ , $t=1, \ldots n,$ есть случайные величины.**
**Требования к ним сформулированы в Предположениях А-Е.**

**Сформулируйте свойства ошибок $\varepsilon_t, \quad t=1, \ldots n$ .**

(формулы и текст)

**69** **Как соотнесены модель простой линейной регрессии с Предположениями I-V и модель автокорреляции 1-го порядка?**

(аргументированный ответ)