Примеры задач с решениями

1. По данным с 1961 по 2005 гг. была оценена (статическая) кривая Филлипса, связывающая уровень инфляции inf_t и уровень безработицы $unem_t$:

$$\widehat{inf}_t = 2.34 - 0.23 unem_t, \quad s_1 = 0.04 \quad R^2 = 0.12$$

Протестируйте гипотезу $H_0 : \theta_1 = 0$ против альтернативы $H_1 : \theta_1 < 0$ при уровне значимости 1%.

Решение. *Гипотеза $H_0 : \theta_1 = 0$ говорит о том, что безработица является незначимым фактором для инфляции. Для проверки гипотезы требуется вычислить значение статистики распределения Стьюдента: $t = \frac{\widehat{\theta}_1}{s(\widehat{\theta}_1)} = \frac{-0.23}{0.04} = -5.75$. Теперь сравним данное значение с критическим значением: $t(\alpha, n-k-1) = t(0.01, 45-1-1) = t(0.01; 43) = -2.423$. Так как значение статистики t лежит в критической области, то гипотеза о том, что безработица не является значимым фактором, отвергается.*

2. На основе квартальных данных с 2001 по 2008 год было получено следующее уравнение регрессии, описывающее зависимость цены y_t на товар от нескольких факторов:

$$\widehat{y}_t = 9.5 + \underset{(0.01)}{0.2}\, x_t + \underset{(0.002)}{0.6}\, w_t + \underset{(0.02)}{0.7}\, z_t; \quad \text{ESS} = 62.5 \quad \text{RSS} = 44.7.$$

Когда в уравнение были добавлены фиктивные переменные, соответствующие первым трем кварталам года, величина ESS выросла до 76.4. Напишите спецификацию уравнения регрессии с учетом сезонности. Вычислите скорректированный коэффициент детерминации для обеих моделей регрессии. Улучшило ли включение сезонности качество модели?

Решение. *Новая модель имеет вид $\hat{y}_t = \beta_0 + \beta_1 x_t + \beta_2 w_t + \beta_3 z_t + \beta_4 \cdot d_1 + \beta_5 \cdot d_2 + \beta_6 \cdot d_3$, где d_i - это бинарная переменная, принимающая значения 1, если наблюдение соответствует сезону i, и 0 - в противном случае. Обратите внимание, что принадлежность к четвёртому сезону определяется равенством $d_1 = d_2 = d_3 = 0$.*

Для тестирования улучшения необходимо проверить гипотезу: $H_0 = \beta_4 = \beta_5 = \beta_6 = 0$, вычислим статистику: $F = \frac{(ESS_R - ESS_{UR})/q}{ESS_{UR}/(n-m)} \sim F(\alpha; q; n-m)$, q-количество фиктивных переменных, m-количество параметров, n-количество наблюдений.

$$F = \frac{(ESS_R - ESS_{UR})/q}{ESS_{UR}/(n-m)} = \frac{(76.4 - 62.5)/3}{76.4/(32-7)} \approx 1.516.$$

Так как критическое значение равно $F_{кр}(\alpha; q; n-m) = F_{кр}(0.05; 3; 25) \approx 2.991$, то гипотеза об отсутствии сезонности не отвергается.

3. Была оценена регрессионная модель зависимости издержек $COST$ энергетических компаний от объемов производства KWH (100 кВт. в ч), географического положения $South$ (бинарная переменная) и от цен на рабочую силу PL, топливо PF и капитал PK. Зависимая переменная: $\ln(COST)$; объем выборки 23.

Y	$const$	KWH	$\ln(PL)$	$\ln(PF)$	$\ln(PK)$	$South$
$coeff$	2.23	0.002	0.21	0.31	0.34	-0.03
$Std.Err$	1.45	0.0008	0.07	0.13	0.07	0.0014

$$ESS = 250.4, \quad RSS = 230.6$$

Предполагается, что ошибка модели регрессии удовлетворяет условиям Гаусса-Маркова. Уровень значимости 0.05.

a) Вычислите коэффициент детерминации.

b) Проверьте значимость регрессии в целом и сформулируйте проверяемую статистическую гипотезу. Вычислите p-значение.

Решение. *a) Вычислим коэффициент детерминации*

$$R^2 = \frac{ESS}{TSS} = \frac{ESS}{ESS + RSS} = \frac{250.4}{250.4 + 230.6} = 0.52$$

b) Основная гипотеза о незначимости регрессии в целом имеет вид

$$H_0 : \beta_1 = \ldots = \beta_5 = 0.$$

Алтернативная гипотеза $H_1 : \beta_1^2 + \ldots \beta_5^2 \neq 0$.

Для проверки данной гипотезы используется следующая статистика:

$$F = \frac{RSS}{TSS} \cdot \frac{n-k-1}{k} \Leftrightarrow F = \frac{R^2}{1-R^2} \cdot \frac{n-k-1}{k} = \frac{0.52}{1-0.52} \cdot \frac{23-5-1}{5} = 2.8$$

Сравним с критическим значением $F(m-1; n-m) = F(5; 17) = 2.21$ $F_{кр} < F_{расчетное}$, *поэтому гипотеза о незначимости регрессии отвергается.*

$p - value = 1 - F^{-1}(2.21) \approx 1 - 0.97 = 0.03.$

4. Была оценена регрессионная модель зависимости издержек COST энергетических компаний от объемов производства KWH (100 кВт. в ч), географического положения South (бинарная переменная) и от цен на рабочую силу PL, топливо PF и капитал PK. Зависимая переменная: $\ln(COST)$; объем выборки 23.

Y	$const$	KWH	$\ln(PL)$	$\ln(PF)$	$\ln(PK)$	$South$
$coeff$	2.23	0.002	0.21	0.31	0.34	-0.03
$Std.Err$	1.45	0.0008	0.07	0.13	0.07	0.0014

$$ESS = 250.4, \quad RSS = 230.6$$

Предполагается, что ошибка модели регрессии удовлетворяет условиям Гаусса-Маркова. Уровень значимости 0.05.

а) Напишите спецификацию регрессии с ограничениями для проверки статистической гипотезы $H_0 : \theta_2 = \theta_3 = \theta_4$. Дайте интерпретацию нулевой гипотезе.

б) Была оценена регрессия с ограничениями и ее остаточная сумма квадратов равна 258.2. Тестируйте гипотезу из предыдущего пункта.

Решение. *а) Гипотеза:* $H_0 : \beta_2 = \beta_3 = \beta_4$ *означает, что эластичности по каждому фактору равны.*

Регрессия с ограничениями имеет вид

$$\ln Y = \beta_0 + \beta_1 KWH + \beta_2(\ln(PL) + \ln(PF) + \ln(PK)) + \beta_5 South + \varepsilon$$

б) Для проверки гипотезы используется статистика: $F = \frac{(R^2_{ur} - R^2_r)(n-m)}{(1-R^2_{ur})q}$, *где* q *– количество ограничений,* m *– количество параметров,* n *– количество наблюдений.*

$$R^2_{ur} = \frac{ESS_{ur}}{TSS} = \frac{250.4}{250.4 + 230.6} \approx 0.52$$

$$R^2_r = \frac{ESS_r}{TSS} = \frac{TSS - RSS_r}{TSS} = \frac{481 - 258.2}{481} = \frac{222.8}{481} \approx 0.46.$$

Теперь находим значение F*-статистики* $F = \frac{(0.52-0.46)(23-6)}{(1-0.52)\cdot 2} \approx 1.06$.

По таблице Фишера находим $F_{кр}(0.05; 2; 17) \approx 3.592$, *поэтому гипотеза не отвергается.*

Дополнительно можно обратиться к пособию

Ивин Е. А., Артамонов Н. В., Курбацкий А. Н. Методическое пособие по эконометрике: для социально-экономических специальностей. — ИСЭРТ РАН Вологда, 2016. — С. 184.