

ДЗ №3

Ластовецкий Дмитрий Александрович

10 февраля 2021 г.

1.2 Нулевая гипотеза: средние результаты по математическому анализу не отличаются между группами. Т.е.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Находим F-статистику и p_F (с помощью R):

$$F_{value} = 2.973, p_F = 0.0346$$

Тестируем p_F на 5% уровне значимости:

$$p_F = 0.0346 < 0.05$$

Так как p_F меньше 5%, мы отвергаем нулевую гипотезу с данным уровнем значимости. Влияние такого фактора, как группа, на успеваемость по математике не доказано.

1.3 Нулевая гипотеза: средние результаты по микроэкономике не отличаются между группами. Т.е.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Находим F-статистику и p_F (с помощью R):

$$F_{value} = 0.677, p_F = 0.568$$

Тестируем p_F на 5% уровне значимости:

$$p_F = 0.568 > 0.05$$

Так как p_F больше 5%, мы не отвергаем нулевую гипотезу с данным уровнем значимости. Мы можем говорить о влиянии такого фактора, как группа, на успеваемость по микроэкономике.

2.1

$$Micro = 4.1 + 0.35 * math + \varepsilon$$

2.2 Интерпретация коэффициентов (в предположении о верной модели): при оценке "0" по матану мы наблюдаем примерно 4.1 по микре. При увеличении оценки по матану на 1 оценка по микре растет в среднем на 0.35.

2.3

$$t_0 = \frac{\overline{Const}}{st.error(Const)} = 6.236$$

$$t_1 = \frac{\overline{\beta_1}}{st.error(\beta_1)} = 3.95$$

$$Pr(> |t_0|) = 2 * p_t(6.236, df = 118) = 7.225868e^{-09}$$

$$Pr(> |t_1|) = 2 * p_t(3.95, df = 118) = 0.00013$$

Оценки коэффициентов являются значимыми (5%); $t - scores > 2$, $p_{val} < 0.05$

2.4 Доверительный интервал для интерцепта:

$$4.10664 - (0.65854 * qt(0.975, 118)) = 2.8026$$

$$4.10664 + (0.65854 * qt(0.975, 118)) = 5.4107$$

(2.8026, 5.4107) - 95% доверительный интервал для интерцепта

При проведении многократного сэмплинга в 95% случаев интерцепт будет находиться в интервале (2.8026, 5.4107).

Доверительный интервал для коэффициента при предикаторе:

$$0.34747 - (0.08796 * qt(0.975, 118)) = 0.1732853$$

$$0.34747 + (0.08796 * qt(0.975, 118)) = 0.5216547$$

(0.1732853; 0.5216547) - 95% доверительный интервал для коэффициента

При проведении многократного сэмплинга в 95% случаев коэффициент при предикаторе будет находиться в интервале (0.1732853; 0.5216547)

Оба доверительных интервала НЕ включают в себя 0, так что константа и коэффициент при предикаторе значимы.

2.5

Multiple R-squared: 0.1168

Adjusted R-squared: 0.1093

Можно сделать вывод о том, что модель объясняет 11.68% или 10.93% возмущений (в зависимости от того, используем мы обычный или скорректированный коэффициент детерминации).

2.6 Искренне и честно признаюсь, что не знаю, какую именно статистику использовали на семинаре для проверки значимости коэффициента детерминации. На майноре мы использовали t-критерии, поэтому я буду использовать их также. Очень надеюсь, что выводы при использовании t-критерия и F-критерия совпадут.

Для проверки значимости коэффициента детерминации используем Т-критерий уровня значимости 0.05.

$$H_0 : R^2 = 0$$

$$t_{real} = \sqrt{0.1168} * \frac{\sqrt{118}}{\sqrt{1-0.1168}} = 3.95$$

$$t_{table} = qt(0.975, 118) = 1.98$$

$t_{real} > t_{table}$, следовательно, мы отклоняем нулевую гипотезу. Другими словами, коэффициент детерминации статистически значим.