Домашнее задание по эконометрике

Выполнил Авилов Илья, Бэк 181

Задание 1

Взаимосвязь между объемом продаж йогуртов и затратами на рекламу.

Исследовательский вопрос: Как затраты на рекламу влияют на объем продажи йогуртов?

Гипотеза: Увеличение затрат на рекламу приводит к уеличению объемов продаж йогуртов

Механизм, лежащий в основе данной взаимосвязи:

БОльшие затраты на рекламу в большинстве случаев свидетельствуют о ее бОльшей эффективности. Эффективная реклама, в свою очередь, повышает осведомленность потенциальных покупателей о рекламируемом товаре, а это приводит к увеличению его продаж.

Данные:

Финансовые отчеты компаний (за определенный год) с информацией о затратах на рекламу и об объемах продаж йогуртов. Затраты на рекламу должны быть сгруппированны по различным каналам (тв, радио, билборды).

Единица наблюдения – 1 компания

Спецификация модели:

sales – объемы продаж
tv.adv - затраты на рекламу с помощью телевидения
radio.adv - затраты на рекламу с помощью радио
tv.bilboard - затраты на рекламу с помощью билбордов

$$\widehat{\text{sales}} = \hat{\alpha} + \widehat{\beta_1} \times \ln \text{tv.adv} + \widehat{\beta_2} \times \ln \text{radio.adv} + \widehat{\beta_3} \times \ln \text{bilboard.adv}$$

Была выбрана данная спецификация модели, т.к. предполагается, что вложения в рекламу (по любому из каналов) обладают убывающим предельным эффектом. Так, компания, ранее не рекламировшая свою продукцию, получит от увеличенных вложений в рекламу бОльший прирост объемов продаж, чем компания с изначально большими затратами на рекламные компании.

Предельный эффект:

$$\mathrm{ME}_{tv.adv}^{\widehat{sales}} = \frac{\widehat{\beta_1}}{tv.adv}$$

$$\mathrm{ME}_{radio.adv}^{\widehat{sales}} = \frac{\widehat{\beta_2}}{radio.adv}$$

$$\mathrm{ME}_{tv.bilboard}^{\widehat{sales}} = \frac{\widehat{\beta_3}}{tv.bilboard}$$

То есть чем больше затраты на рекламу (по любому каналу), тем меньше предельный эффект на объемы продаж.

Эластичность:

$$E_{tv.adv}^{\widehat{sales}} = \frac{\widehat{\beta_1}}{\widehat{sales}}$$

$$E^{\widehat{sales}}_{radio.adv} = \frac{\widehat{\beta_2}}{\widehat{sales}}$$

$$E^{\widehat{sales}}_{tv.bilboard} = \frac{\widehat{\beta_3}}{\widehat{sales}}$$

Чем больше объем продаж, тем меньше тем меньше процентное изменение затрат на рекламу (пл любому из каналов) влияет на изменение объема продаж.

Факторы, входящие в случайную ошибку:

1. Эффективность рекламной компании определяется не только ее стоимостью, за одни и те же деньги одна компания может сделать гораздо более эффективную рекламу, чем другая компания.
2. Неверные наблюдения: в финансовой отчетности компаний могут присутствовать неточности (например, допущенные с целью уклонения от налогов или с целью мошенничества)
3. Различные каналы рекламы (на тв есть разные каналы, на радио есть разные станции и т.д.). Учесть данную информацию не получится, т.к. если ввести отдельный регрессор для каждого канала рекламы, то модель получится слишком сложной, что усложнит ее интерпретацию. Также в большей части наблюдений большинство регрессоров будет равно нулю, т.к. не каждая компания может позволить себе рекламу на всех ресурсах сразу.

Взаимосвязь между уровнем цен в ресторане и его рейтингом

Исследовательский вопрос: как уровень цен ресторана влияет на его рейтинг?

Гипотеза: рестораны с высоким уровнем цен обладают рейтингом ниже, чем рестораны с низким уровнем цен.

Механизм, лежащий в основе данной взаимосвязи:

Ожидания людей относительно качества сервиса и блюд связаны с уровнем цен ресторана. Люди, посещающие недорогой ресторан, не ожидают от него безупречного обслуживания и поэтому остаются довольны даже относительно невысоким уровнем сервиса. С другой стороны, если человек пришел в дорогой ресторан, он будет ожидать высокого качества обслуживания и будет замечать все недочеты, что приведет к низкому рейтинга ресторана в случае их наличия.

Данные:

Информация о рейтинге, основной кухне и среднем чеке ресторана с сайта restoclub.ru. Все рестораны должны находиться в одном городе.

Единица наблюдения – ресторан

Спецификация модели:

rating – рейтинг ресторана
metro – расстояние от метро
mean.check – средний чек

$$\widehat{\mathrm{rating}} = \widehat{\alpha} + \widehat{\beta_1} \times mean.\mathrm{check} + \widehat{\beta_2} \times metro$$

Была выбрана данная спецификация модели, т.к. предполагается, что рейтинг ресторана зависит линейно от среднего чека и от расстояния до метро.

Предельный эффект:

$$\mathrm{M}E_{mean.check}^{\widehat{rating}} = \widehat{\beta_1}$$

$$\mathrm{M}E_{metro}^{\widehat{rating}} = \widehat{\beta_2}$$

Предельный эффект постоянный

Эластичность:

$$E_{mean.check}^{\widehat{rating}} = \frac{\widehat{\beta_1} \times \mathrm{mean.check}}{\widehat{rating}}$$

$$E_{metro}^{\widehat{rating}} = \frac{\widehat{\beta_2} \times \mathrm{metro}}{\widehat{rating}}$$

Эластичность непостоянна.

Факторы, входящие в ошибку:

1. Необъективность оценок – разные люди оценивают один и тот же ресторан по-разному.
2. Возможная смещенность оценок людей – люди могут поставить рейтинг ресторану только если остались очень недовольны или наоборот остались очень довольны рестораном. Это может привести к смещенности общего рейтинга ресторана.

Задание 2

$$model.df \ = \ k-1 \ = \ 3-1 \ = \ 2$$

$$residual.df \ = n-k \ = \ 539-3 \ = \ 536$$

$$totwork.t = \frac{coef}{std.err} = \frac{-0.1709083}{0.0212731} = -8.034010088$$

$$male.coef = \frac{6.607538 + 159.4909}{2} = 83.049219$$

$$male.std.err = \frac{male.coef}{male.t} = \frac{83.049219}{2.13} = 38.99024368$$

$$cons.t = \frac{cons.coef}{cons.std.err} = \frac{3584.533}{46.08652} = 77.77833952$$

$$\begin{aligned} cons95\%conf.interval&: (cons.coef - t \times cons.std.err; cons.coef + t \times cons.std.err) \\ &= (3584.533 - 1.964 \times 46.08652; 3584.533 + 1.964 \times 46.08652) \\ &= (3494.019114; 3675.046886) \end{aligned}$$

$$RMSE = \sqrt{\frac{RSS}{df_{residuals}}} = \sqrt{\frac{91884830.7}{536}} = \sqrt{171426.9229} = 414.037345$$

$$R^2 = 1 - \frac{RSS}{SS_{total}} = 1 - \frac{91884830.7}{103106773} = 0.1088380712$$

$$R_{adj}^2 = 1 - (1-R^2) \times \frac{n-1}{n-k} = 1 - (1 - 0.1088380712) \times \frac{539-1}{539-3} = 0.1055128401$$

$$F = \frac{R^2}{1-R^2} \times \frac{n-k}{k-1} = \frac{0.1088380712}{1-0.1088380712} \times \frac{539-3}{3-1} = 32.73097979$$

В модели 3 регрессора (включая константу)

При изменении totwrk на 1 единицу переменная sleep уменьшится на 0.17 единиц

При изменении male на 1 единицу переменная sleep увеличится на 83 единиц

На 5% уровне значимости можно сказать, что ни один из коэффициентов не равен нулю.

Вероятность, что все коээффициенты, кроме коэффициента перед констаной, равны нулю, также стремится к нулю (исходя из F статистики).

Данная модель объясняет ~10% дисперсии зависимой переменной.

При этом оценка стандартной ошибки = 414.

Задание 3

1) **Несмещенность оценки**

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{2i}-y_{2i-1}}{x_{2i}-x_{2i-1}}\right)$$

Найдем мат.ожидание оценки

$$E\hat{\beta} = \frac{1}{n}E\sum_{i=1}^{n}\left(\frac{y_{2i}-y_{2i-1}}{x_{2i}-x_{2i-1}}\right)$$

$$E\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{E(y_{2i}-y_{2i-1})}{x_{2i}-x_{2i-1}}\right)$$

$$E\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{E(\alpha+\beta\times x_{2i} - \alpha - \beta\times x_{2i-1} + \varepsilon_{x_{2i}} - \varepsilon_{x_{2i-1}})}{x_{2i}-x_{2i-1}}\right)$$

$$E\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}(\frac{\beta\times x_{2i} - \beta\times x_{2i-1}}{x_{2i}-x_{2i-1}})$$

$$E\hat{\beta} = \frac{n\times\beta}{n} = \beta$$

Мат.ожидание оценки равно реальному коэффициенто, следовательно оценка несмещенная.

2) **Дисперсия оценки**

$$Var\hat{\beta} = E\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{2i}-y_{2i-1}}{x_{2i}-x_{2i-1}}\right)\right)^2\right) - \left(E\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{2i}-y_{2i-1}}{x_{2i}-x_{2i-1}}\right)\right)\right)^2$$

$$\left(E\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{2i}-y_{2i-1}}{x_{2i}-x_{2i-1}}\right)\right)\right)^2 = \widehat{\beta^2}$$

Квадрат суммы раскладывается как сумма квадратов плюс сумма всех попарных перемножий слагаемых (кроме перемножения слагаемого на себя)

Здесь и далее q != p

$$\mathrm{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{2i}-y_{2i-1}}{x_{2i}-x_{2i-1}}\right)\right)^2\right) = \frac{1}{n^2}\times E\left(\left(\sum_{i=1}^{n}\left(\frac{y_{2i}-y_{2i-1}}{x_{2i}-x_{2i-1}}\right)^2\right)+\sum_{q=1}^{n}\sum_{p=1}^{n}\left(\left(\frac{y_{2q}-y_{2q-1}}{x_{2q}-x_{2q-1}}\right)\times\left(\frac{y_{2p}-y_{2p-1}}{x_{2p}-x_{2p-1}}\right)\right)\right)$$

$$\mathrm{E}\left(\sum_{i=1}^{n}\left(\frac{y_{2i}-y_{2i-1}}{x_{2i}-x_{2i-1}}\right)^2\right) = \sum_{i=1}^{n}\left(\frac{E(y_{2i}-y_{2i-1})^2}{(x_{2i}-x_{2i-1})^2}\right)$$
$$= \sum_{i=1}^{n}\left(\frac{E\left(\alpha+\beta\times x_{2i}+\varepsilon_{x_{2i}}-\alpha-\beta\times x_{2i-1}-\varepsilon_{x_{2i-1}}\right)^2}{(x_{2i}-x_{2i-1})^2}\right)$$
$$= \sum_{i=1}^{n}\left(\frac{E\left(\beta\times(x_{2i}-x_{2i-1})+\varepsilon_{x_{2i}}-\varepsilon_{x_{2i-1}}\right)^2}{(x_{2i}-x_{2i-1})^2}\right)$$
$$= \sum_{i=1}^{n}\left(\frac{E\left(\beta^2\times(x_{2i}-x_{2i-1})^2+\left(\varepsilon_{x_{2i}}-\varepsilon_{x_{2i-1}}\right)^2\right)}{(x_{2i}-x_{2i-1})^2}\right)$$
$$= \sum_{i=1}^{n}\left(\frac{\beta^2\times(x_{2i}-x_{2i-1})^2+E\left(\varepsilon_{x_{2i}}-\varepsilon_{x_{2i-1}}\right)^2}{(x_{2i}-x_{2i-1})^2}\right)$$
$$= \sum_{i=1}^{n}\left(\frac{\beta^2\times(x_{2i}-x_{2i-1})^2+E\left(\varepsilon_{x_{2i}}^2-2\times\varepsilon_{x_{2i}}\times\varepsilon_{x_{2i-1}}+\varepsilon_{x_{2i-1}}^2\right)}{(x_{2i}-x_{2i-1})^2}\right)$$

$$\sum_{i=1}^{n}\left(\frac{\beta^2\times(x_{2i}-x_{2i-1})^2+E\left(\varepsilon_{x_{2i}}^2-2\times\varepsilon_{x_{2i}}\times\varepsilon_{x_{2i-1}}+\varepsilon_{x_{2i-1}}^2\right)}{(x_{2i}-x_{2i-1})^2}\right) = \sum_{i=1}^{n}\left(\frac{\beta^2\times(x_{2i}-x_{2i-1})^2+E\left(\varepsilon_{x_{2i}}^2+\varepsilon_{x_{2i-1}}^2\right)-2\times E\times(\varepsilon_{x_{2i}}\times\varepsilon_{x_{2i-1}})}{(x_{2i}-x_{2i-1})^2}\right) =$$
$$\sum_{i=1}^{n}\left(\frac{\beta^2\times(x_{2i}-x_{2i-1})^2+E\left(\varepsilon_{x_{2i}}^2+\varepsilon_{x_{2i-1}}^2\right)}{(x_{2i}-x_{2i-1})^2}\right) = \sum_{i=1}^{n}\left(\frac{\beta^2\times(x_{2i}-x_{2i-1})^2+2\times Var(e))}{(x_{2i}-x_{2i-1})^2}\right) =$$
$$n\times\beta^2+2\times Var(e)\times\sum_{i=1}^{n}\frac{1}{(x_{2i}-x_{2i-1})^2}$$

$$\sum_{q=1}^{n}\sum_{p=1}^{n}\left(\left(\frac{y_{2q}-y_{2q-1}}{x_{2q}-x_{2q-1}}\right)\times\left(\frac{y_{2p}-y_{2p-1}}{x_{2p}-x_{2p-1}}\right)\right)$$
$$= \sum_{q=1}^{n}\sum_{p=1}^{n}\left(\left(\frac{\left(\beta\times\left(x_{2q}-x_{2q-1}\right)+\varepsilon_{x_{2q}}-\varepsilon_{x_{2q-1}}\right)}{x_{2q}-x_{2q-1}}\right)\right.$$
$$\left.\times\left(\frac{\beta\times\left(x_{2p}-x_{2p-1}\right)+\varepsilon_{x_{2p}}-\varepsilon_{x_{2p-1}}}{x_{2p}-x_{2p-1}}\right)\right)$$

$$\sum_{q=1}^{n}\sum_{i=1}^{n}(\frac{(\beta^2 \times (x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1}) + (\varepsilon_{x_{2q}}-\varepsilon_{x_{2q-1}}) \times (\varepsilon_{x_{2p}}-\varepsilon_{x_{2p-1}})) + \beta \times (x_{2q}-x_{2q-1}) \times (\varepsilon_{x_{2p}}-\varepsilon_{x_{2p-1}}) + \beta \times (x_{2p}-x_{2p-1}) \times (\varepsilon_{x_{2q}}-\varepsilon_{x_{2q-1}}))}{(x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1})})$$

$$\sum_{q=1}^{n}\sum_{p=1}^{n}(\frac{(\beta^2 \times (x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1}) + (\varepsilon_{x_{2q}}-\varepsilon_{x_{2q-1}}) \times (\varepsilon_{x_{2p}}-\varepsilon_{x_{2p-1}})) + \beta \times (x_{2q}-x_{2q-1}) \times (\varepsilon_{x_{2p}}-\varepsilon_{x_{2p-1}}) + \beta \times (x_{2p}-x_{2p-1}) \times (\varepsilon_{x_{2q}}-\varepsilon_{x_{2q-1}}))}{(x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1})}) =$$
$$\sum_{q=1}^{n}\sum_{p=1}^{n}(\frac{(\beta^2 \times (x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1}) + (\varepsilon_{x_{2q}} \times \varepsilon_{x_{2p}} + \varepsilon_{x_{2q-1}} \times \varepsilon_{x_{2p-1}} - \varepsilon_{x_{2q}} \times \varepsilon_{x_{2p-1}} - \varepsilon_{x_{2q-1}} \times \varepsilon_{x_{2p}}))}{(x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1})})$$

$$E\sum_{q=1}^{n}\sum_{p=1}^{n}\left(\frac{\left(\beta^2 \times (x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1}) + \left(\varepsilon_{x_{2q}} \times \varepsilon_{x_{2p}} + \varepsilon_{x_{2q-1}} \times \varepsilon_{x_{2p-1}} - \varepsilon_{x_{2q}} \times \varepsilon_{x_{2p-1}} - \varepsilon_{x_{2q-1}} \times \varepsilon_{x_{2p}}\right)\right)}{(x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1})}\right) =$$
$$\sum_{q=1}^{n}\sum_{p=1}^{n}\left(\frac{\left(\beta^2 \times (x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1})\right)}{(x_{2q}-x_{2q-1}) \times (x_{2p}-x_{2p-1})}\right) = \sum_{q=1}^{n}\sum_{p=1}^{n}\beta^2 = n \times (n-1) \times \beta^2$$

$$Var\hat{\beta} = \frac{1}{n^2} \times \left(n \times \beta^2 + 2 \times Var(e) \times \sum_{i=1}^{n}\frac{1}{{(x_{2i}-x_{2i-1})}^2} + n \times (n-1) \times \beta^2\right) -$$
$$\beta^2 = \frac{2}{n^2} \times Var(e) \times \sum_{i=1}^{n}\frac{1}{{(x_{2i}-x_{2i-1})}^2}$$

3) Состоятельность оценки:

$$\lim_{n\to\infty}(\frac{2}{n^2} \times Var(e) \times \sum_{i=1}^{n}\frac{1}{{(x_{2i}-x_{2i-1})}^2}) = 0$$

Задание 4

1) Оценки

$$X'X = \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} \times \begin{bmatrix} x^1 & x^2 & x^3 \end{bmatrix} = \begin{pmatrix} 25 & 0 & 0 \\ 0 & 50 & 50 \\ 0 & 50 & 100 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{50}\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

$$(X'Y)^{-1} = \begin{bmatrix} x^1 \\ x^2 \\ x^3 \end{bmatrix} \times y = \begin{pmatrix} 600 \\ 500 \\ 400 \end{pmatrix}$$

$$\hat{\beta} = (X'X)^{-1}X'Y = \begin{pmatrix} 24 \\ 12 \\ -2 \end{pmatrix}$$

2) Дисперсия оценок

Найдет RSS

$$RSS = e'e = \left((I-P)Y\right)'\left((I-P)Y\right) = Y'Y - (X'Y)'(X'X)^{-1}X'Y$$

$$RSS = 26888 - \begin{pmatrix} 600 & 500 & 400 \end{pmatrix} \frac{1}{50} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 600 \\ 500 \\ 400 \end{pmatrix}$$

$$RSS = 26888 - 19600 = 7288$$

Оценка дисперсии

$$MSE = \frac{RSS}{df} = \frac{7288}{22} = 331{,}2727$$

Дисперсия оцененных коэффициентов

$$Var\widehat{\beta} = \sigma^2 (X'X)^{-1}$$

$$331.2727 \times \frac{1}{50} \times \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 13.250908 & 0 & 0 \\ 0 & 13.250908 & -6.625454 \\ 0 & -6.625454 & 6.625454 \end{pmatrix}$$

Следовательно,

$$var\left(\widehat{beta_1}\right) = 13.250908$$

$$var\left(\widehat{beta_2}\right) = 13.250908$$

$$var\left(\widehat{beta_3}\right) = 6.625454$$

3) 95% доверительный интервал

Рассчитаем дисперсию:

$$\begin{aligned} \mathrm{var}\hat{y} = \mathrm{var}\left(\widehat{\beta_1} + \widehat{\beta_2} x^{(2)} + \widehat{\beta_3} x^{(3)}\right) \\ = \mathrm{var}\left(\widehat{\beta_1}\right) + x^{(2)^2} \mathrm{var}\left(\widehat{\beta_2}\right) + x^{(3)^2} \mathrm{var}\left(\widehat{\beta_3}\right) + 2 \times x^{(2)} \mathrm{cov}\left(\widehat{\beta_1}; \widehat{\beta_2}\right) \\ + 2 \times x^{(3)} \mathrm{cov}\left(\widehat{\beta_1}; \widehat{\beta_3}\right) + 2 \times x^{(2)} x^{(3)} \mathrm{cov}\left(\widehat{\beta_2}; \widehat{\beta_3}\right) \end{aligned}$$

$$\begin{aligned} \mathrm{var}\hat{y} = 13.250908 + 4^2 \times 13.250908 + 3^2 \times 6.625454 + 2 \times 4 \times 0 + 2 \times 3 \times 0 \\ - 2 \times 4 \times 3 \times 6.625454 = 125{,}95 \end{aligned}$$

Рассчитаем среднее ожидаемое значение y при данных регрессорах

$$\hat{y} = 24 + 12 \times 4 - 2 \times 3 = 66$$

Количество степней свободы и t-статистика:

df = 22

t = 2.074

95% доверительный интервал

$$(66 - 2.074 \times \sqrt{}\{125{,}95\}); 66 + 2.074 \times \sqrt{}\{125{,}95\})) = (42{,}72; 89{,}276)$$