

**Резюме по семинару / вебинару по моделям
с переменными взаимодействия. Разбор заданий**

Задание 1. Ниже представлены оценки регрессионной модели. Большему значению зависимой переменной, измеряемой в непрерывной шкале от 0 до 100, соответствует более высокий уровень коррупции. Свобода прессы измерена в непрерывной шкале от 1 до 5, где 5 - более высокий уровень свободы прессы. *Обратите внимание на то, что в модели используется центрированный показатель свободы прессы.* Переменная «бывшая британская колония» бинарная: принимает значение 1, если государство является бывшей британской колонией, 0 - в противном случае.

	коррупция
Свобода прессы (центрированный)	-1.506*** (0.159)
Бывшая британская колония	-11.102*** (1.58)
Переменная взаимодействия	-4.544*** (0.585)
Константа	23.803*** (3.192)

Notes: Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1. Возможно ли исключить из модели только предиктор «свобода прессы»? Объясните свой ответ (к каким последствиям приведет данное исключение). При каком условии (теоретически) исключение показателя, входящего в переменную взаимодействия, было бы возможно?

Ответ: Нет, так как «свобода прессы» входит в переменную взаимодействия. Ее исключение приведет к смещенным оценкам, а также неверно рассчитанному значению предельного эффекта и как следствие неверной интерпретации. Теоретически исключение было бы возможно, если бы корреляция между «свободой прессы» и другими предикторами была бы равна 0, тогда бы это позволило сохранить BLUE-оценки.

2. Проинтерпретируйте все оценки коэффициентов в представленной модели.

Ответ:

Константа: 23.803 – среднее значение коррупции в странах, которые не являются бывшими британскими колониями, и при условии, что свобода прессы принимает среднее значение (или при условии того, что центрированный показатель свободы прессы равен 0).

Свобода прессы (центрированный): при увеличении свободы прессы на единицу измерения в странах, не являющихся бывшими британскими колониями, коррупция в среднем уменьшается на 1.506.

Бывшая британская колония: в странах – бывших британских колониях по сравнению со странами, таковыми не являющимися, при условии того, что свобода прессы принимает среднее значение (или при условии того, что центрированный показатель свободы прессы равен 0), коррупция в среднем ниже на 11.102.

Переменная взаимодействия: в странах – бывших британских колониях – по сравнению со странами, таковыми не являющимися, наблюдается более выраженная отрицательная взаимосвязь свободы прессы и коррупции: в среднем коэффициент взаимосвязи ниже на 4.544 (и составляет -6.05).

3. От чего зависит предельный эффект свободы прессы? Рассчитайте предельный эффект свободы прессы при разных «условиях».

Ответ: Выпишем предельный эффект свободы прессы в общем виде:

$\frac{\partial y_i}{\partial FP_i} = -1.506 + (-4.544)BC_i$, где FP_i – свобода прессы, BC_i – дамми-переменная для бывшей британской колонии.

Таким образом,

для стран, не являющихся бывшими британскими колониями, предельный эффект составляет $-1.506 + (-4.544) \times 0 = -1.506$

для стран, являющихся бывшими британскими колониями, предельный эффект составляет $-1.506 + (-4.544) \times 1 = -6.05$

4. Чему равны оценки констант и коэффициента при предикторе «свобода прессы» в моделях, оцененных на отдельных выборках (такие выборки сформированы на основании значения показателя «бывшая британская колония»)? Зависимой переменной остается уровень коррупции.

Ответ:

Для стран, не являющихся бывшими британскими колониями:

$$\hat{\beta}_0 = 23.803; \hat{\beta}_1 = -1.506$$

Для стран, являющихся бывшими британскими колониями:

$$\hat{\alpha}_0 = 23.803 - 11.102 = 12.701; \hat{\alpha}_1 = -1.506 - 4.544 = -6.05$$

5. *Сделайте самостоятельно после семинара.* Рассчитайте оценки коэффициентов в модели с такой же спецификацией, однако при условии того, что показатель свободы прессы включен без центрирования. Известно, что среднее значение свободы прессы по выборке составляет 2.8. Проинтерпретируйте все полу-

ченные коэффициенты.

Ответ:

Исходная спецификация модели выглядит следующим образом:

$$y_i = \beta_0 + \beta_1(FP_i - \overline{FP_i}) + \beta_2BC_i + \beta_3(FP_i - \overline{FP_i})BC_i + \epsilon_i$$

Распишем эту модель, раскрыв скобки (избавляемся от центрирования):

$$y_i = \beta_0 + \beta_1FP_i - \beta_1\overline{FP_i} + \beta_2BC_i + \beta_3FP_iBC_i - \beta_3\overline{FP_i}BC_i + \epsilon_i$$

Выразим новые константу и коэффициенты при предикторах:

$$(\beta_0 - \beta_1\overline{FP_i}) + \beta_1FP_i + (\beta_2 - \beta_3\overline{FP_i})BC_i + \beta_3FP_iBC_i + \epsilon_i$$

Задание 2. Ниже представлены оценки регрессионной модели. Большему значению зависимой переменной – индекс социального обеспечения **benefit generosity index** (далее – BGI), измеряемой в непрерывной шкале от 0 до 100, соответствует наиболее щедрая социальная политика. Индекс качества государственного управления **QoG** приведен к непрерывной шкале от 0 до 1, где 1 – наиболее высокий уровень качества государственного управления. Показатель **Left** – процент мест «левых» партий в парламенте.

	BGI
QoG	18.9*** (3.39)
Left	-1.57* (-2.16)
QoG×Left	1.92** (3.17)
контрольные переменные	включены
Константа	15.79*** (4.34)

Notes: t statistics are given in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1. Проинтерпретируйте оценки коэффициентов при всех предикторах.

Ответ:

Константа: 15.79 – среднее значение индекса социального обеспечения при условии того, что все предикторы в модели равны 0.

QoG: при увеличении качества государственного управления на единицу измерения индекс социального обеспечения в среднем увеличивается на 18.9 при условии того, что процент мест «левых» партий в парламенте = 0, и при прочих равных контрольных переменных.

Left: при увеличении Left на 1% индекс социального обеспечения в среднем уменьшается на 1.57 при условии того, что качество государственного управления = 0, и при прочих равных контрольных переменных.

QoG × Left: при увеличении Left на 1% при прочих равных контрольных переменных взаимосвязь качества государственного управления и индекса социального обеспечения становится более выраженной положительной: коэффициент взаимосвязи в среднем увеличивается на 1.92.

Возможна также симметричная интерпретация (в исследовании зависит от Ваших гипотез): при увеличении качества государственного управления на единицу измерения при прочих равных контрольных переменных взаимосвязь процента мест «левых» партий в парламенте и индекса социального обеспечения в среднем увеличивается на 1.92.

2. *Сделайте самостоятельно после семинара.* Запишите в общем виде предельный эффект и дисперсию предельного эффекта изменения Left на отклик. От значений какой переменной зависит предельный эффект Left и его дисперсия?

Ответ:

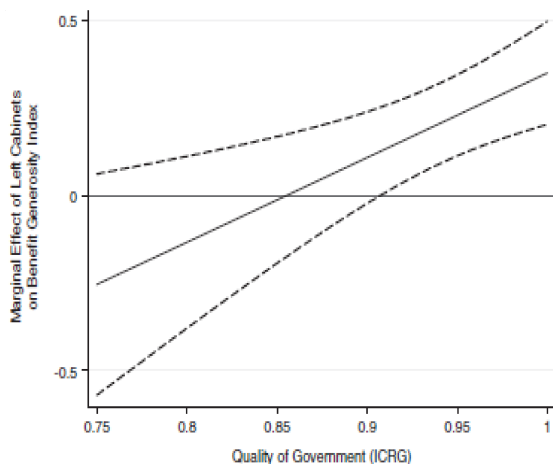
Предельный эффект Left:

$$\frac{\partial BGI_i}{\partial Left_i} = \hat{\beta}_2 + \hat{\beta}_3 QoG_i = -1.57 + 1.92 QoG_i$$

Вариация предельного эффекта Left:

$$\begin{aligned} Var(\hat{\beta}_2 + \hat{\beta}_3 QoG_i) &= Var(\hat{\beta}_2) + Var(\hat{\beta}_3 QoG_i) + 2Cov(\hat{\beta}_2, \hat{\beta}_3 QoG_i) = \\ &= Var(\hat{\beta}_2) + (QoG_i)^2 Var(\hat{\beta}_3) + 2QoG_i Cov(\hat{\beta}_2, \hat{\beta}_3) \end{aligned}$$

3. Проинтерпретируйте график ниже. Определите, в каких случаях предельный эффект Left значим. На графике приведен 95%-ый доверительный интервал.



Ответ:

На графике показана линейная взаимосвязь предельного эффекта переменной Left и качества государственного управления. Взаимосвязь положительная: с ростом качества государственного управления увеличивается и значение предельного эффекта.

Для того, чтобы определить значимость предельного эффекта на 5%-ом уровне, нужно посмотреть, покрывает ли 95%-ый доверительный интервал 0 (границы доверительного интервала обозначены на графике пунктирными линиями): при значениях QoG от 0.75 до примерно 0.91 доверительный интервал содержит 0 (на графике см. горизонтальную линию для обозначения равенства предельного эффекта 0), следовательно, на этом интервале предельный эффект является незначимым. При значениях QoG выше 0.91 обе границы доверительного интервала выше горизонтальной линии (то есть, доверительный интервал не содержит 0), значит, для этой области значений от 0.91 до 1 предельный эффект Left является статистически значимым.

Обратите также внимание на то, что доверительный интервал меняет ширину: можно предположить, что там где доверительный интервал широкий, меньше доступно данных, что дает высокие стандартные ошибки.

Задание 3. Исследователь решил протестировать совместный эффект переменных «респондент – латиноамериканец» (L), «доход респондента» в непрерывной шкале (I), «уровень религиозности» (порядковая шкала от 0 до 5) (R) на уровень доверия Президенту. В качестве предикторов в регрессионную модель исследователь включил все указанные переменные по отдельности, переменную взаимодействия между этими тремя переменными ($L \times I \times R$) и контрольные переменные. Верна ли данная спецификация модели? Если Вы считаете спецификацию ошибочной, предложите необходимые изменения.

Ответ:

Спецификация является неверной, так как пропускает промежуточные переменные взаимодействия (попарные произведения), что грозит неверно рассчитанным предельным эффектом и неверной интерпретацией.

Верная спецификация: $y_i = \beta_0 + \beta_1 L_i + \beta_2 I_i + \beta_3 R_i + \beta_4 L_i I_i + \beta_5 L_i R_i + \beta_6 R_i I_i + \beta_7 L_i I_i R_i + \epsilon_i$

Задание 4. Ниже представлены результаты оценивания взаимосвязи удовлетворенности жизнью и возраста. Удовлетворенность жизни измеряется в категориальной шкале от 0 до 10, где 10 соответствует респондентам, наиболее удовлетворенным своей жизнью. Возраст измеряется в годах.

	Удовлетворенность жизнью
Возраст	-0.359*** (0.021)
Возраст × Возраст	0.003*** (0.0001)
контрольные переменные	включены

Notes: Standard errors are given in parentheses

1. Проинтерпретируйте оценки коэффициентов в модели.

Ответ:

В интерпретации оценок коэффициентов нам поможет предельный эффект возраста на удовлетворенность жизнью:

$$\frac{\partial LS_i}{\partial Age_i} = -0.359 + 2 \times 0.003 Age_i = -0.359 + 0.006 Age_i, \text{ где LS – life satisfaction}$$

Таким образом, оценка коэффициента при возрасте (-0.359) – взаимосвязь возраста и удовлетворенности жизнью при условии, что возраст = 0. Как это понимать? Давайте разбираться.

Для начала введем понятия:

средняя скорость изменения (average velocity, или average rate of change) = $\frac{\Delta y}{\Delta x}$ – отношение того, насколько изменилась зависимая переменная, к изменению в X. Допустим, если мы наблюдаем прирост y на 6 единиц измерения при условии того, что x увеличился на 2 единицы измерения, тогда средняя скорость составляет 3.

Примечание: в английском языке есть разница между понятиями speed и velocity (rate of change). Если velocity может быть как положительной, так и отрицательной, то speed – это модуль значения velocity. К примеру, если при увеличении x на 2 единицы измерения y уменьшился на 6 единиц измерения, тогда velocity (rate of change) составит -3 . А вот speed будет равна 3.

скорость в конкретный момент времени (instantaneous velocity, или instantaneous rate of change) $\frac{\partial y}{\partial x}$, или это то же самое, что $\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$.

Таким образом, возвращаясь к нашей задаче, -0.359 – это скорость (velocity) изменения удовлетворенности жизнью в конкретной точке (когда возраст = 0).

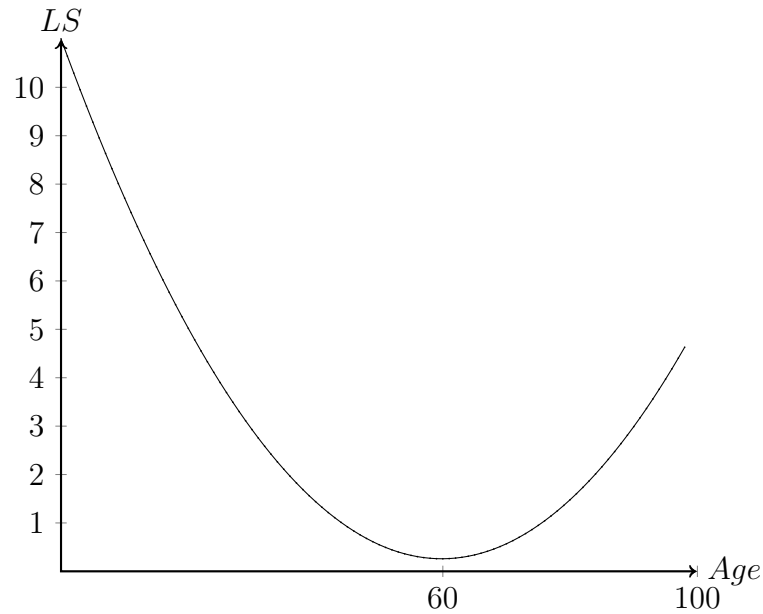
Вторая часть предельного эффекта ($0.006Age$) показывает, что эффект возраста на удовлетворенность жизнью зависит от самого возраста. С каждым последующим годом эффект возраста изменяется. В противном случае при постоянном изменении удовлетворенности жизнью при увеличении возраста на год мы бы наблюдали линейную взаимосвязь (графически – просто прямая). Согласитесь, что наивно было бы полагать, что эффект возраста на удовлетворенность жизнью в любом возрасте остается постоянным, поэтому мы и обратились к одному из видов нелинейной взаимосвязи (включили квадратичный эффект возраста в модель).

Сформулируем интерпретацию: 0.006 – это то, насколько в среднем изменяется взаимосвязь возраста и удовлетворенности жизнью при увеличении возраста на 1 год и при прочих равных контрольных переменных. То есть, к примеру, если скорость изменения при возрасте = 0 составляет -0.359 , то при возрасте = 1 год скорость изменения = $-0.359 + 0.006 = -0.353$. Если хотите проинтерпретировать исходную оценку коэффициента при квадратичном возрасте (0.003), тогда можно сказать, что это половина изменения эффекта возраста на удовлетворенность жизнью при увеличении возраста на год. Но такая трактовка не очень удобна.

2. Изобразите схематично на графике взаимосвязь возраста и удовлетворенности жизнью, заданную в данной модели. Определите «переломную» точку, после которой эффект возраста на удовлетворенность жизни меняет знак.

В оценках модели не была указана константа, я для удобства возьму значение 11 по удовлетворенности жизнью (*константа – это точка пересечения графика и оси ординат OY, не путайте, пожалуйста, с наклоном в точке, где возраст = 0!*). При этом не столь важно, что зависимая переменная измеряется в шкале от 0 до 10, значение 11 – гипотетическое, при условии того, что возраст респондента = 0. Понятно, что никто не измерял удовлетворенность жизнью новорожденного, поэтому мы можем для константы получить значение, выходящее за границы указанной шкалы.

Мы моделировали квадратичный эффект возраста, поэтому графически это будет задаваться параболой. Как мы уже отметили выше, оценка коэффициента при возрасте = -0.359 , а далее с каждым годом отрицательная связь возраста и удовлетворенности жизнью будет ослабляться. Поэтому в первой точке (при возрасте = 0) наклон наиболее выраженный отрицательный, а далее он становится все менее отрицательным, достигает пороговой точки и становится положительным. Отразим это на графике ниже.



Найдем ту самую «переломную» точку, после которой взаимосвязь возраста и удовлетворенности жизнью становится положительной (меняет знак). Для этого приравняем предельный эффект к 0:

$$-0.359 + 0.006Age_i = 0$$

$$Age_i \approx 60$$