ПРОВЕРКА ДОПУЩЕНИЙ В SPSS

ОБЗОР

В основе всех статистических процедур лежат те или иные допущения. В некоторых случаях нарушения этих допущений не будут существенно сказываться на надежности заключений, которые исследователь делает после обработки данных с использованием статистической процедуры. В других случаях нарушения допущений могут существенно сказываться на надежности таких заключений. В содержании файлов с описаниями различных статистических процедур в SPSS всегда присутствует раздел «Допущения», в котором они описываются применительно к конкретной статистической процедуре. Констатация автора, что такие-то данные отвечают допущениям используемой статистической процедуры, является желательным компонентом всех журнальных статей, тезисов и диссертаций, опирающихся на количественные эмпирические исследования.

В этом файле рассматриваются наиболее распространенные виды допущений, использующихся в статистических процедурах.

КЛЮЧЕВЫЕ ПОНЯТИЯ И ТЕРМИНЫ

Допущения лежат в основе любой статистической процедуры. (См. также файл «Уровни данных в SPSS»). Настоящий файл содержит информацию, общую для всех статистических процедур.

Например, **параметрические статистики** являются теми, которые допускают определенное распределение данных (обычно нормальное распределение), допускают интервальный уровень измерения и допускают однородность дисперсий, когда сравниваются две или более выборки. Наиболее распространенные тесты значимости (z-тесты, t-тесты и F-тесты) являются параметрическими. Однако, как это было уже давно установлено, умеренные нарушения параметрических допущений в большинстве случаев оказывают малое или совсем не оказывают влияния на надежность заключений *(напр., Cohen, 1969: 266-267)*.

ДОПУЩЕНИЯ

Качество измерения

- Описательные статистики [descriptive statistics]. Все формы статистического анализа предполагают хорошее качество измерения, относительно свободное от ошибок кодирования. Хорошим способом это проверить является расчет описательных статистик по данным с целью убедиться, что данные в целом отвечают ожиданиям в терминах средних и стандартных отклонений, и что они не содержат значений за пределами ожидаемого диапазона оценок.

- Избегание ослабления [attenuation]. Когда изначальный диапазон оценок какой-либо переменной искусственно сжимается, например, посредством классификации или дихотомизации непрерывной переменной, ее корреляции с другими переменными ослабляются, зачастую приводя к недооценке величины эффекта (влияния) этой переменной.

ОСЛАБЛЕНИЕ

В таблице, показанной ниже, переменные возраста, образования и дохода были продублированы как сжатые переменные с только тремя категориями оценок в каждой

	age	educ	Income	agecat	educat	incomecat
1	60	12	120000.00	3	1	2
2	27	17	100000.00	1	3	1
3	36	12	120000.00	2	1	2
4	21	13	100000.00	1	2	1
5	35	16	120000.00	1	3	2
6	33	16	110000.00	1	3	1
7	43	12	120000.00	2	1	2
8	29	13	120000.00	1	2	2

В корреляционной матрице, приведенной ниже, мы видим, что корреляции сжатых переменных уменьшились по сравнению с исходными непрерывными переменными. Так, корреляции возраста с доходом настолько уменьшилась, что утратила свою прежнюю значимость.

Корреляции

		Age of respondent	Highest year of school completed	Income	Age of respondent (Binned)	Highest year of school completed (Binned)	Income (Binned)
Age of respondent	Pearson Correlation	1	-.175**	-.044*	.907**	-.127**	-.076**
	Sig. (2-tailed)		.000	.028	.000	.000	.000
	N	2828	2816	2501	2828	2816	2501
Highest year of school completed	Pearson Correlation	-.175**	1	.310**	-.126**	.846**	.324**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	2816	2820	2495	2816	2820	2495
Income	Pearson Correlation	-.044*	.310**	1	.000	.219**	.706**
	Sig. (2-tailed)	.028	.000		.998	.000	.000
	N	2501	2495	2503	2501	2495	2503
Age of respondent (Binned)	Pearson Correlation	.907**	-.126**	.000	1	-.096**	-.018
	Sig. (2-tailed)	.000	.000	.998		.000	.373
	N	2828	2816	2501	2828	2816	2501
Highest year of school completed (Binned)	Pearson Correlation	-.127**	.846**	.219**	-.096**	1	.249**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	2816	2820	2495	2816	2820	2495
Income (Binned)	Pearson Correlation	-.076**	.324**	.706**	-.018	.249**	1
	Sig. (2-tailed)	.000	.000	.000	.373	.000	
	N	2501	2495	2503	2501	2495	2503

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Имейте ввиду, что теоретически сжатие может уменьшать корреляции, повышать их или оставлять их неизменными. Однако, наиболее частый (ожидаемый) эффект выражается в ослаблении (уменьшении) корреляций.

- Избегание тавтологической корреляции. Когда оценки латентной переменной А концептуально пересекаются или даже совпадают с оценками латентной переменной Б, это приводит к смешиванию корреляций А и Б с другими латентными переменными. Это становится особенно проблематичным для переменных из независимой и зависимой частей регрессионного уравнения. В более мягком виде концептуальное пересечение может возникать, когда одна и та же переменная используется в роли знаменателя одной переменной из независимой части уравнения и другой переменной в зависимой части уравнения (напр., количество безработных на 100 000 человек населения, использующееся в качестве предиктора количества преступлений на 100 000 человек населения).

Надлежащая спецификация модели

- Ненадлежащая спецификация модели [specification of a model] связана с не включением в модель значимых каузальных переменных или включением коррелирующих, но чуждых, посторонних переменных, а также с некорректным указанием направления стрелок, связывающих переменные в модели. Когда ошибка, связанная с ненадлежащей спецификацией модели корректируется путем изменения модели, все оцениваемые параметры в модели также подвергаются изменениям, не только по величине, но иногда даже по направлению. В уточнении круга переменных, которые надлежит специфицировать, важную роль играет хороший обзор литературы в области изучаемого явления.

  В качестве «правила большого пальца», чем ниже общий эффект (напр., R^2 в множественной регрессии, качество соответствия в логистической регрессии), тем выше вероятность, что какие-то важные переменные были упущены в модели, и что существующие интерпретации модели будут меняться, когда данная модель будет специфицирована корректным образом. Проблема спецификации уменьшается, когда цель исследования состоит в просто сравнении моделей, чтобы увидеть, которая из них лучше соответствует исходным данным, чем когда она заключается в подтверждении правильности некоторой модели и оценке сравнительной важности независимых переменных.

Уровень данных

- Требования к уровню данных [data level] варьируются в зависимости от статистической процедуры, однако большинство процедур требуют интервальных или пропорциональных уровней измерения *(Т.е., чтобы оценки переменных опирались на интервальные шкалы или шкалы отношений.)* В социальных науках, однако, распространено использование дихотомических и порядковых данных, таких как данные в формате лайкертовской шкалы даже в процедурах, которые технически требуют интервального уровня данных. Дихотомические переменные часто включаются в статистические модели (напр., в регрессионные модели), обеспечивающие расщепление двух признаков в соотношении менее чем 90:10 (напр., 65% женщин к 35% мужчин). Порядковые переменные часто включаются в статистические модели, если может быть продемонстрирована нормальность стандартных ошибок, как обсуждается ниже. Нарушение допущений в отношении уровня данных означает, что реальная стандартная ошибка

[standard error] будет выше, чем рассчитанная стандартная ошибка, а значимость переоценена (т.е., шансы Ошибки I типа будут выше, чем это рассчитано).

- *Тест метричности порядковых данных [test of near-metricity]*. Процедура многомерного шкалирования обеспечивает метод тестирования метричности порядковых данных (См. файл «Многомерное шкалирование в SPSS»).

Одномерность

- Тестирование одномерности [testing unidimensionality]. Если исследователь пытается измерить конструкт с множеством переменных-индикаторов *(т.е., многопунктовой методикой)*, он должен продемонстрировать, что его пункты измеряют ту же самую переменную. Существуют несколько методов сделать это, варьирующих по назначению и строгости тестирования на одномерность. Вот некоторые из них:

1. *Альфа Кронбаха*. Вероятно, наиболее часто используемый тест альфа Кронбаха служит измерением интеркорреляции пунктов методики. Если альфа превышает или равен .60, пункты методики считаются одномерными и могут быть объединены при расчете общего показателя по методике. Некоторые исследователи используют более строгую отсечку .70. Альфа Кронбаха можно найти в SPSS, среди других мест, выбрав Анализировать, Шкала, Анализ надежности.

2. *Факторный анализ*. Факторный анализ по методу главных компонент реализуется для всех индикаторов всех конструктов *(т.е., для всех пунктов всех переменных)*, измерявшихся в исследовании. Индикаторы должны обладать более высокими факторными нагрузками на их собственные конструкты, чем на другие конструкты. Некоторые исследователи также требуют, чтобы нагрузки превышали некоторый абсолютный уровень отсечки, такой как .30. Некоторые исследователи также требуют, чтобы индикаторы не имели перекрестных нагрузок на любой другой фактор, кроме своего собственного (напр., чтобы все нагрузки на другие факторы, кроме своего собственного были ниже некоторого абсолютного уровня отсечки, такого как .30). Факторный анализ можно найти в SPSS, выбрав Анализировать, Редукция данных, Фактор.

3. *Конфирматорный факторный анализ в структурном моделировании*. Первый шаг структурного моделирования подразумевает проведение конфирматорного факторного анализа, где модель измерения оценивается отдельно от структурной модели. Если показатели качества соответствия модели измерения исходным данным оказываются приемлемыми, исследователь заключает, что индикаторы адекватно измеряют охватываемые ими конструкты. Можно также оценить одномерность определенного конструкта путем сравнения моделей в структурном моделировании. Для конфирматорного факторного анализа можно использовать AMOS, распространяемый производителями SPSS в виде статистического пакета с процедурами структурного моделирования.

4. *Шкала Гуттмана и другие формы шкалирования*. В шкалах Гуттмана, близости, Моккена и других определенных типах шкал переменные-индикаторы тестируются, чтобы увидеть, формируют ли они определенные связи друг с другом, и исследователь имел основания для включения таких пунктов в ту же саму методику. Такие связи и, соответственно, смыслы одномерности варьируются в зависимости от типа шкалы. Заметьте, что наиболее распространенная форма шкалы, лайкертовская шкала, обычно не подвергается тестированию на одномерность (см. файл «Шкалы и стандартные измерения в SPSS»). Шкалу Гуттмана можно найти в SPSS, выбрав Анализировать, Шкала, Анализ надежности.

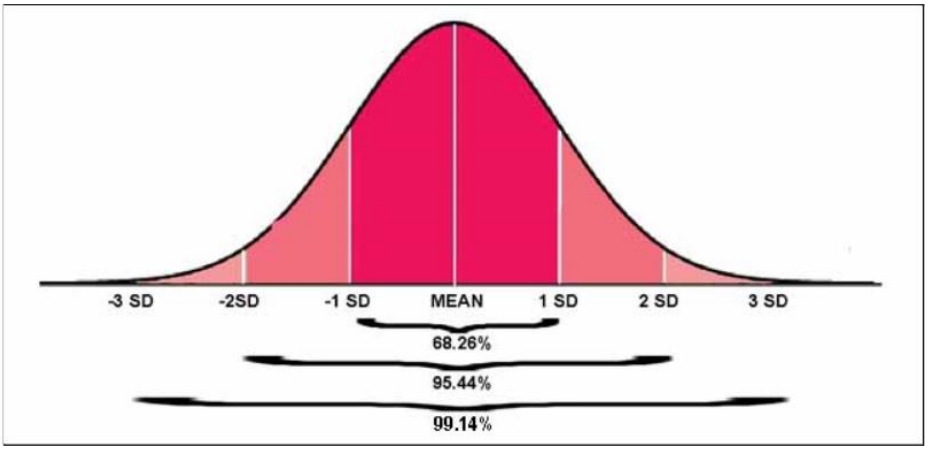
Заметьте, что набор пунктов может рассматриваться одномерным после использования одного из вышеописанных методов даже когда другой метод может не обнаружить статистического оправдания в рассмотрении этих пунктов как измеряющих тот же самый конструкт. Например, будет вполне обычным обнаружить, что пункты в большой шкале Гуттмана не нагружаются на единственный фактор после применения факторного анализа. Обнаружение удовлетворительных показателей качества соответствия в структурном моделировании не гарантирует, что те же самые пункты будут отвечать критериям альфа Кронбаха. Исследователь должен определиться с тем, какая теоретическая основа, какое теоретическое определение и какой критерий для одномерности будут лучше всего отвечать его исследовательской цели. Однако, *некоторый* метод должен всегда использоваться перед тем, как рассчитывать общий показатель по многопунктовой методике.

Нормальность

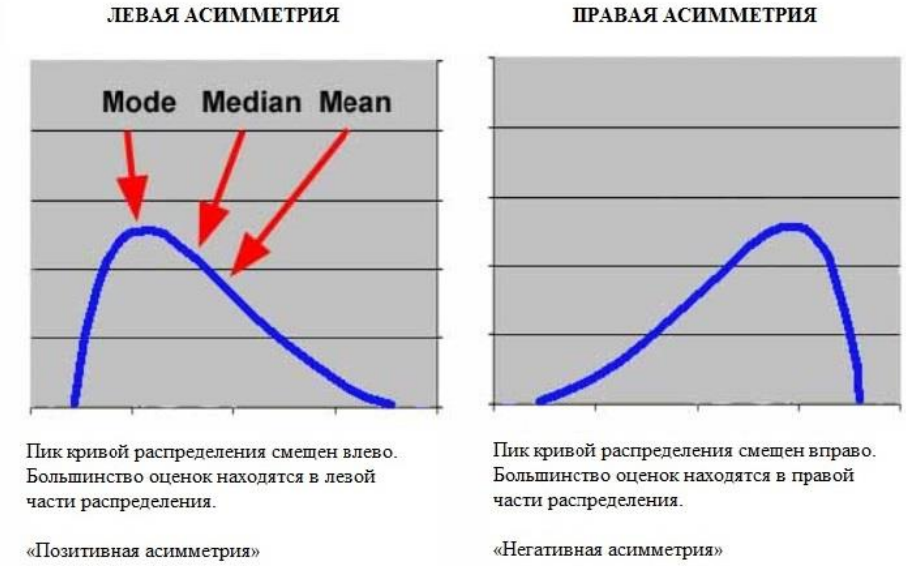
- Нормальное распределение [normal distribution] подразумевается многими статистическими процедурами. Чтобы скорректировать не-нормально распределенные данные, используются разнообразные трансформации. Корреляция, регрессия на основе наименьших квадратов, факторный анализ и связанные линейные процедуры относительно устойчивы [robust] к не очень сильным отклонениям от нормальности при условии, что распределение оценок не является слишком асимметричным *(Vasu, 1979).* Сильная асимметрия может возникать из-за сильных выбросов. Лог-линейный анализ, логистическая регрессия и связанные процедуры, использующие оценку максимального правдоподобия являются даже еще более устойчивыми в отношении умеренных отклонений от нормальности *(см. Steenkamp & van Trijp, 199, p: 285).* Сходным образом, симуляции Монте-Карло показывают, что t-тест оказывается устойчивым в отношении умеренных нарушений нормальности.

  Нормальное распределение имеет форму симметричной колоколообразной кривой. *Стандартное* нормальное распределение – это распределение со средней оценкой 0 и стандартным отклонением 1. *Стандартные оценки*, также называемые z-оценками или стандартизованными данными, этой оценки, которые получаются путем вычитания выборочной средней оценки и затем делятся на выборочное стандартное отклонение, что обеспечивает среднее 0 и стандартное отклонение 1. Нормальность можно оценить визуально путем просмотра гистограммы частот или путем просмотра графика нормальной вероятности, которые обеспечиваются большинством статистических компьютерных программ.

  Область под нормальной кривой репрезентирует вероятность: 68,26% случаев будут лежать в пределах 1 стандартного отклонения от средней, 95,44% в пределах 2 стандартных отклонений и 99,14% в пределах 3 стандартных отклонений. Часто это суждение заменяется с целью сокращения фразой: 1 ст.откл. корреспондирует с 2/3 случаев, 2 ст.откл.. с 95% и 3 ст.откл. с 99%. Другим способом сказать то же самое, что существует меньше .05 шанса, что выбранный случай будет лежать за пределами 2 стандартных отклонений от средней и менее .01 шанса, что он окажется за пределами 3 стандартных отклонений. Такое утверждение аналогично утверждению, относящемуся к уровням значимости .05 и .01 для двусторонних тестов.



- Асимметрия [skew] – это смещение пика кривой распределения в ту или иную сторону от средней оценки. Наиболее распространена левая асимметрия, когда пик кривой смещен в левую сторону от средней. Менее распространена правая асимметрия, где он смещен в правую сторону. «Правило большого пальца» при тестировании нормальности состоит в том, чтобы рассчитать описательные статистики, получить показатели асимметрии и эксцесса, а затем разделить их на стандартные ошибки. Если данные нормально распределены, показатели асимметрии должны лежать в диапазоне от +2 до -2. Некоторые авторы используют диапазон от +1 до -1 в качестве более строгого критерия, когда нормальность имеет решающее значение. В SPSS показатели асимметрии можно получить, выбрав Анализировать, Описательные статистики, Описательные, затем нажав кнопку «Опции» и отметив опцию «Асимметрия».



Вместо терминов «левая асимметрия» и «правая асимметрия» также часто используются термины «позитивная асимметрия» и «негативная асимметрия», в зависимости от знака показателя асимметрии. Показатель левой асимметрии имеет позитивный знак, поэтому ее также называют «позитивной асимметрией», а показатель правой негативный, поэтому ее также называют «негативной асимметрией». При сообщении в отчетах об асимметричных данных для каждого типа асимметрии принято приводить среднее, медиана и мода, т.е. все три измерения центральной тенденции. Асимметрию может нормализовать трансформация Бокса-Кокса (см. ниже эту трансформацию, а также другие способы трансформации). Распределения с левой асимметрией могут соответствовать степенному, логнормальному, гамма, Вейбулла или хи-квадрат распределениям. Распределения с правой асимметрией можно перекодированы в распределения с левой асимметрией.

- Эксцесс [kurtosis] – это уплощение или, наоборот, вытягивание пика кривой распределения. «Правило большого пальца» при тестировании нормальности состоит в том, чтобы рассчитать описательные статистики, получить показатели асимметрии и эксцесса, а затем применить критерий, по которому эксцесс должен находиться в диапазоне от +2 до -2. Некоторые авторы используют более мягкий критерий от +3 до -3, тогда как другие более жесткий от +1 до -1, когда нормальность имеет решающее значение. Негативный эксцесс указывает на слишком большое количество оценок в концах распределения, позитивный эксцесс на слишком малое количество таких оценок. Заметьте, что изначально при расчете эксцесса для нормального распределения используется значение 3 и некоторые статистические пакеты отталкиваются от нормального значения эксцесса как 3, однако настоящее обсуждение предполагает, что 3 вычитается, чтобы его центральное значение составило 0, как это сделано в SPSS и LISREL. Вариант с нормальным распределением, центрирующемся на 0, называется *эксцессом Фишера*, тогда как вариант, центрирующийся на 3, называется *эксцессом Пирсона*. SPSS использует эксцесс Фишера. При островершинном эксцессе [leptocurtosis] пик кривой распределения имеет «острые концы», указывая на эксцесс > 0 (для эксцесса Фишера, или > 3 для эксцесса Пирсона). При плосковершинном эксцессе [platycurtosis] пик кривой имеет «толстые концы», указывая на эксцесс < 0 (для эксцесса Фишера, или < 3 для эксцесса Пирсона).

Для корректировки эксцесса используются различные трансформации: извлечение кубических корней и синусов могут корректировать негативный эксцесс. В SPSS показатели эксцесса можно получить, выбрав Анализировать, Описательные статистики, Описательные, затем нажав кнопку «Опции» и отметив опцию «Эксцесс».

- Дихотомии [dichotomies]. По определению дихотомия не нормально распределена. Многие исследователи будут использовать дихотомии для процедур, требующих нормального распределения, если рас-

щепление обоих признаков находится в соотношении меньше 90:10. Дихотомии не должны использоваться в качестве зависимых переменных в процедурах, подобно регрессии OLS, которые допускают нормально распределенную зависимую переменную.

- W тест Шапиро-Вилка [Shapiro-Wilk`s W test] является формальным тестом нормальности, предлагаемым в модуле SPSS EXAMINE или в процедуре SAS UNIVARIATE. Это стандартный тест нормальности. В SPSS выберите Анализировать, Описательные статистики, Исследовать. Заметьте, меню выбора «Исследовать» вставляет код EXAMINE. Нажмите кнопку «Графики» и отметьте опцию «Графики нормальности с тестами». После нажатия OK будут выведены данные, подобные тем, что представлены ниже:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
StdEduc	.017	932	.200*	.999	932	.960

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Для рассматриваемой переменной W не должно быть значимым, если ее распределение не значимо отличается от нормального, как это видно в случае StdEduc в верхней таблице. Об W можно думать как о корреляции между оценками данной переменой и ее корреспондирующими нормальными оценками, с W = 1, когда оценки данной переменной абсолютно нормально распределены. Когда W значимо меньше 1, допущение о нормальности не выполняется. W тест Шапиро-Вилка рекомендуется для выборок небольших и умеренных размеров до n = 2000. Для более крупных выборок SAS и другими процедурами рекомендуется тест Колмогорова-Смирнова.

- D тест Колмогорова-Смирнова или K-S тест Лиллифорса [Kolmogorov-Smirnov D test or K-S Lilliefors test] является альтернативным тестом нормальности для крупных выборок, доступен в SPSS EXAMINE и SAS UNIVARIATE. Этот тест также можно найти в SPSS, выбрав Анализировать, Описательные статистики, Исследовать, нажав кнопку «Графики» и отметив опцию «Графики нормальности с тестами». Иллюстрация таблицы с результатами приведена выше. D Колмогорова-Смирнова иногда называют *тестом Лиллифорса* из-за коррекции K-S, разработанной Лиллифорсом, которая сейчас обычно применяется. SPSS, например, автоматически применяет коррекцию Лиллифорса к K-S тесту нормальности в модуле EXAMINE (но не в модуле NONPAR). Этот тест, с коррекцией Лиллифорса, является предпочтительным хи-квадрат тестом качества соответствия, когда данные являются интервальными или близкими к интервальным. Применяемый без коррекции Лиллифорса K-S является очень консервативным: то есть, он повышает вероятность обнаружения не-нормальности данных. Заметьте, что K-S тест может проверять качество соответствия применительно к любому теоретическому распределению, а не только к нормальному. Помните, что когда размер выборки велик, даже небольшие отклонения от нормальности могут приобретать значимость в этом и других тестах. По этой причине рекомендуется использовать для суждений о нормальности другие источники, такие как графики распределения частот и ствола и листьев [stem-and-leaf plots].

- Графические методы.

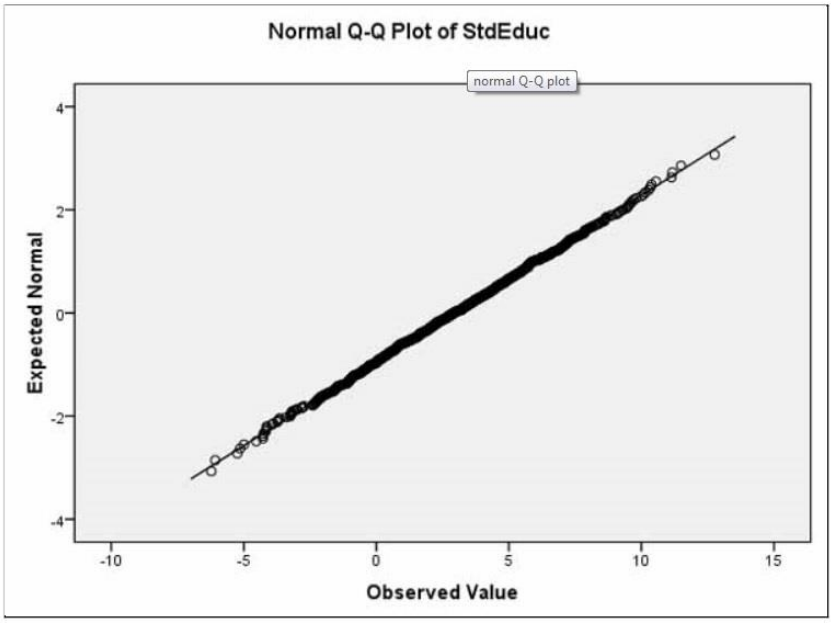
  - *Гистограмма* [histogram] переменной показывает сырую нормальность, а гистограмма остатков, если оценки нормально распределены, часто приводится как свидетельство нормальности переменных.

  - *График кумулятивного распределения эмпирической и теоретической функций* [empirical by theoretical cumulative distribution functions (cdf`s)] просто показывает эмпирическое распределение как пунктирную линию и гипотетическое распределение, или нормальную кривую, как сплошную линию.

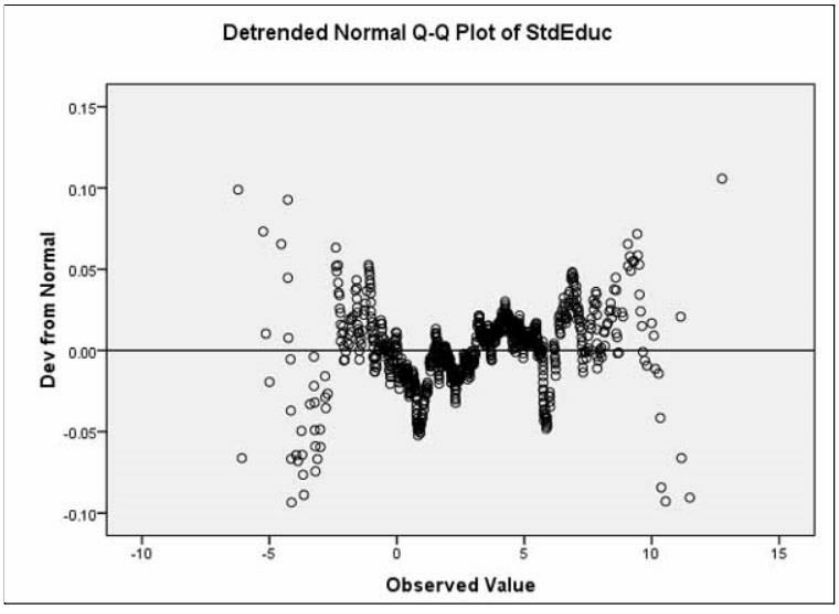
  - *График P-P* [P-P plot] можно найти в SPSS, выбрав Графики, P-P графики. Он позволяет проверить, является ли распределение данной переменной нормальным (или бета, хи-квадрат, экспоненциальным, гамма, полунормальным, Лапласа, логистическим, логнормальным, Парето, t Стьюдента, Вейбулла или однородным) [beta, chi-square, exponential, gamma, half-normal, Laplace, Logistic, Lognormal, Pareto, Student's t, Weibull, or uniform]. График P-P изображает кумулятивные *пропорции* переменной против кумулятивных пропорций проверяемого распределения. Чем прямее линия, формируемая графиком P-P, тем лучше распределение переменной согласуется с

проверяемым распределением (напр., с нормальным). Опции внутри этой процедуры SPSS позволяют предварительно трансформировать данные (натуральный логарифм, стандартизация значений, различие и сезонное различие).

- *Квантиль-квантиль или Q-Q график* [Q-Q plot] формирует линию под углом 45 градусов, когда наблюдаемые оценки согласуются с гипотетическим распределением. Q-Q график изображает *квантили* распределения переменной против квантилей проверяемого распределения. В меню SPSS выберите Графики, Q-Q графики. Диалоговое окно SPSS поддерживает тестирование следующих распределений: бета, хи-квадрат, экспоненциальное, гамма, полунормальное, Лапласа, логистическое, логнормальное, Парето, t Стьюдента, Вейбулла и однородное. Q-Q графики также можно вызвать в SPSS, выбрав Анализировать, Описательные статистики, Исследовать, нажав кнопку «Графики» и отметив опцию «Графики нормальности с тестами».

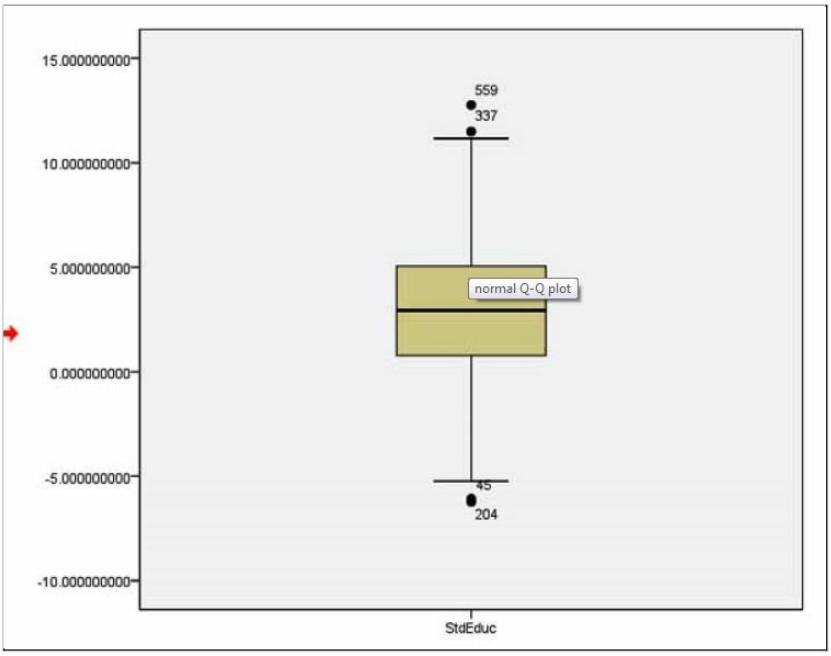


- *Безтрендовый Q-Q график* [detrended Q-Q plot], получаемый в SPSS тем те самым способом, обеспечивает сходную информацию. Если переменная нормально распределена, случаи в таком графике должны сосредотачиваться вокруг горизонтальной линии 0, репрезентирующей нулевые стандартные отклонения от 45-градусной линии, показанной в Q-Q графике сверху. Такой график полезен для поиска выбросов. Однако в иллюстрации ниже нет выбросов в том смысле, что здесь нет случаев, отклоняющихся от линии более чем на +/- .12 стандартных отклонений. Случаи с превышением +/- 1.96 стандартных отклонений являются выбросами с 95%-м доверительным интервалом.



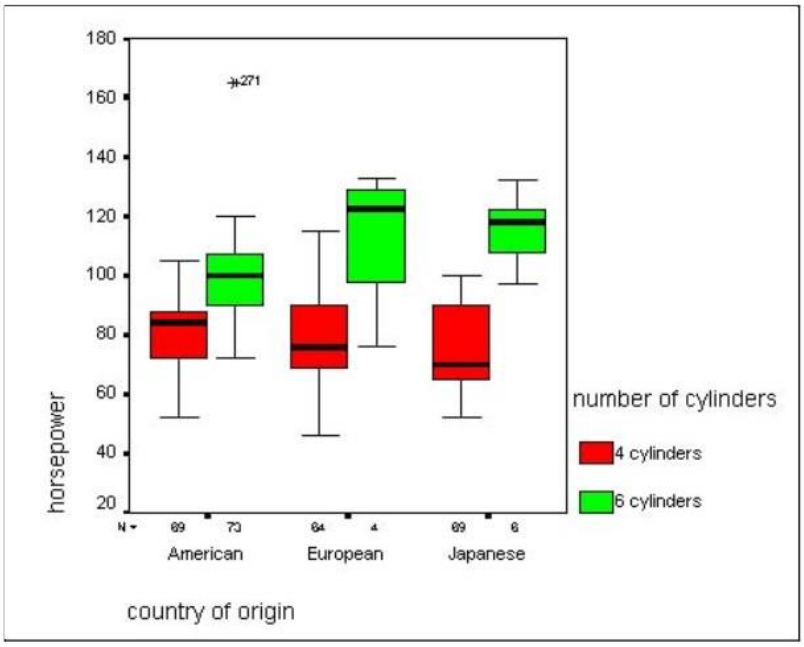
- *Тестирование допущений о нормальности с коробчатыми графиками* [boxplot tests]. Выбросы и асимметрия указывают на не-нормальность, и могут быть проверены при помощи коробчатых графиков. В SPSS эту опцию можно найти выбрав Графики, Коробчатые графики. (А также выбрав Анализировать, Описательные статистики, Исследовать, нажав кнопку «Графики» и отметив «Графики нормальности с тестами».) В случае одной переменной, проверяемой на нормальность, коробчатый график представляет собой диаграмму с этой переменной на оси X и с осью Y, репрезентирующей ее оценки (в иллюстрации ниже, стандартизованные оценки уровня образования, StdEduc). Он состоит из прямоугольника, занимающего пространство от первого до третьего квартиля (то есть, от 25 до 75 процентиля). Линия внутри этого прямоугольника соответствует медиане. В нижней иллюстрации она составляет 2.5. Если большая часть прямоугольника находится по одну или другую сторону от этой линии, это указывает, что переменная асимметрична (не нормальна).

  За пределы прямоугольника выходят «усы», отмечающие минимальное и максимальное значения. Значения, удаленные более чем на полторы длины прямоугольника, помечаются кружками. Значения, удаленные от границ более чем на три длины прямоугольника (экстремальные значения, или выбросы), показываются как номера случаев, выходящих за пределы «усов» и помечаются звездочками. (Вы можете также запросить коробчатый график для двух независимых переменных одновременно, выбрав Графики, Коробчатые графики и нажав кнопку «Кластеризованные».)



  В более сложном случае коробчатый график может также представлять собой диаграмму, в которой категории категориальной независимой переменной или нескольких независимых переменных располагаются на оси X, а оценки интервальной зависимой переменной на оси Y. В иллюстрации ниже имеются две категориальные независимые переменные («страна-производитель автомобилей» и «количество цилиндров»), прогнозирующие непрерывную зависимую переменную «лошадиные силы».

  Внутри графика для каждой категории X приведен прямоугольник, указывающий диапазон оценок (от первого до третьего квартиля) зависимой переменной для данной категории. Если эти прямоугольники расположены на примерно одинаковой высоте Y для всех категорий, это указывает на небольшие различия между группами. Внутри прямоугольников имеется горизонтальная черная линия, показывающая медиану. Если большая часть прямоугольника находится по ту или другую сторону от этой линии, это означает, что зависимая переменная асимметрична (не нормальна) для данной группы (категории). «Усы» и выбросы трактуются, как это было описано выше для случая с одной переменной.



- Повторная выборка [resampling] представляет собой способ протестировать значимость, избегая параметрических допущений, подобных многомерной нормальности. Допущение о многомерной нормальности нарушается, когда используются дихотомические, искусственные и другие дискретные переменные. В таких ситуациях исследователи могут воспользоваться методом повторной выборки.

- Нормализирующие трансформации. Для корректировки асимметрии используют различные трансформации (ниже они рассматриваются применительно к стандартизованным, или z-оценкам переменной)).

    1. Извлечение квадратного корня, логарифма и инверсии (1/x) «втягивают» выбросы и нормализуют левую (позитивную) асимметрию. Инверсивные (реципрокные) трансформации сильнее, чем логарифмические, которые, в свою очередь, сильнее, чем извлечение квадратного корня.

    2. Чтобы скорректировать правую (негативную) асимметрию, сначала вычитают все оценки переменной из наиболее высокой плюс 1, затем применяют извлечение квадратного корня, инверсивную или логарифмическую трансформацию.

    3. Для степенных трансформаций (т.е., связанных с возведением в степень или извлечением квадратного корня) более тонких изменений можно добиться путем добавления константы, C, где C небольшое позитивное число, такое как .50, при трансформации X. Например, $X` = (X + C)^P$. Когда данные исследователя содержать нулевые значения, трансформация с использованием C настоятельно рекомендуется сверх обычной трансформации (напр., квадратный корень из (X+5), а не квадратный корень из (X)), хотя использование C является стандартной практикой во всех случаях. Значения $P < 1$(позволяют извлекать корни) корректируют левую асимметрию (при ее нормализации чаще всего используют $P = 2/3$). Значения $P > 1$ (возведение в степень) корректируют правую асимметрию. Для левой асимметрии уменьшение P уменьшает левую асимметрию. Слишком сильное уменьшение P будет вызывать чрезмерную коррекцию и даже приводить к правой асимметрии. Когда найдено наилучшее значения P, в дальнейшем можно добиться более тонких изменений путем подбора и добавления C. Для левой асимметрии, например, вычитание C будет уменьшать асимметрию.

    4. Логарифмы против извлечения корней: логарифмические трансформации подходят для достижения симметрии в центре распределения, когда симметрия в его концах неважна; извлечение квадратного корня используется, когда важна симметрия в концах распределения; когда важны обе симметрии, может хорошо сработать извлечение корня четвертой степени (такие корни используются для коррекции очень сильной асимметрии).

    5. Логит и пробит трансформации. *Schumacker & Lomax (2004, p.33)* рекомендуют пробит трасформации в качестве средства корректировки асимметрии. Смотри *Lipsey & Wilson (2001, p. 56)* для обсуждения логит и пробит трансформаций как средств трансформации дихотомических данных как части оценки величины эффекта.

6. Проценты можно нормализовать арксинусной трансформацией, которая рекомендуется, когда проценты выходят за пределы диапазона 30% - 70%. Чем больше наблюдений за пределами этого диапазона или чем ближе они к о% или к 100%, тем больше нарушается нормальность и тем настоятельнее необходимость прибегнуть к арксинусной трансформации. Однако арксинусная трансформация неэффективна, когда значительное количество наблюдений составляет 0% или 100%, или когда размер выборки невелик. Обычная арксинусная трансформация выглядит так: p` = arcsin (квадратный корень из p), где p является процентом или пропорцией (долей).

7. Пуассоновские распределения могут быть нормализованы трансформацией с извлечением квадратного корня.

8. Другие стратегии коррекции асимметрии включают объединение отдельных категорий и удаление выбросов.

Предостережение. Трансформации должны иметь теоретический смысл. Например, нормализация такой дихотомии как пол не будет иметь теоретического смысла. Также заметьте, что поскольку логарифм нуля неопределим и ведет к сообщениям об ошибке, исследователи часто добавляют некоторое произвольное малое положительное число (напр., .001) ко всем оценкам, чтобы удалить нули из своих данных. Однако выбор константы может влиять на уровни значимости рассчитываемых коэффициентов (напр., корреляций) таких переменных. Если используется такая стратегия, исследователь должен использовать тонкий анализ с различными константами, чтобы заметить эффекты, которые могут приводить к неверным выводам.

- Трансформации в SPSS: выберите Трансформировать – Рассчитать переменную – Целевая переменная (введите название новой переменной) – Числовое выражение (введите формулу трансформации).

- Трансформации зависимых переменных Бокса-Кокса [Box-Cox transformations pf dependent variables]. Бокс и Кос предложили в 1964 году метод максимального правдоподобия для определения оптимальной степенной трансформации с целью нормализации данных. Степенные трансформации зависимых переменных были предложены для исправления недостатка нормального распределения, недостатка однородности дисперсий и недостатка аддитивности. В контексте регрессии трансформации Бокса-Кокса адресуются к проблеме не-нормальности, на которую указывают асимметричные остатки (в этом случае трансформация выправляет асимметрию) и/или отсутствие гомоскедастичности точек вокруг регрессионной линии. В контексте дисперсионного анализа (ANOVA) трансформации Бокса-Кокса адресуются к проблеме недостатка однородности дисперсий, связанного с корреляцией дисперсий со средними в группах, формируемых независимыми факторами и проявляющегося в асимметричных распределениях внутри групп (в этом случае трансформация уменьшает корреляцию).

  В целом, процедура Бокса-Кокса включает следующие операции: (1) Деление независимой переменной на 10 или подобное число областей; (2) Расчет среднего и стандартного отклонения для каждой области; (3) Построение графика логарифм (стандартного отклонения) против логарифм (среднее) для набора областей; (4) Если график представляет собой прямую линию, отмечается ее наклон, b, затем трансформируется переменная путем возведения зависимой переменной в степень (1-b), и если b = 1, тогда берется логарифм зависимой переменной; и (5) если имеется несколько независимых переменных, повторяются шаги 1-4 для каждой независимой переменной и берется b, которое является рангом тех b, которые вы получили.

  На практике компьютерные пакеты используют итеративный алгоритм максимального правдоподобия для расчета лямбда, параметра Бокса-Кокса, использующегося для определения точной степенной трансформации, которая будет наилучшим образом декоррелировать дисперсии и средине в группах, формируемых независимыми переменными. В качестве «правила большого пальца», если лямбда = 0, в трансформациях нет необходимости. Лямбда +.50 корреспондирует с трансформацией извлечения квадратного корня из зависимой переменной; лямбда 0 корреспондирует с трансформацией на основе натурального логарифма; -.50 корреспондирует с реципрокной трансформацией извлечения квадратного корня; и лямбда -1.0 корреспондирует с реципрокной трансформацией (на основе инверсии). Трансформация Бокса-Кокса пока не поддерживается в SPSS.

  Смотри Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, B, 26, 211-234; see also Maddala, G. S. (1977). *Econometrics*. New York: McGraw-Hill. (page 315-

317); or Mason, R., L., R. F. Gunst, and J. L. Hess (1989). *Statistical design and analysis of experiments with applications to engineering and science.* New York: Wiley.

Многомерная нормальность

- Многомерная нормальность [multivariate normality] присутствует, когда каждая переменная нормально распределена в отношении каждой другой переменной. Множественный дисперсионный анализ (MANOVA), например, требует многомерной нормальности в зависимых переменных. Структурное моделирование и некоторые другие процедуры также допускают многомерную нормальность, которая обсуждается в файле «Структурное моделирование в AMOS».

- Статистика Мардиа [Mardia`s statistic] – это тест многомерной нормальности, Базирующийся на функциях асимметрии и эксцесса, РК Мардиа должен быть менее 3, чтобы было выполнено допущение о многомерной нормальности. PRELIS (входящий в пакет LISREL) выводит показатель РК SPSS пока не поддерживает РК Мардиа. Однако, ДеКарло написал макро SPSS для многомерных тестов асимметрии и эксцесса, доступный на http://www.columbia.edu/~ld208/.

- Одномерный скрининг многомерной нормальности. В качестве «быстрого и грубого» метода некоторые исследователи тестируют нормальность каждой переменной в модели, после чего делают допущение, что раз все они нормально распределены, значит, существует многомерная нормальность. Этот подход не гарантирует корректных заключений.

- Двумерный скрининг многомерной нормальности. Строится двумерная скатерграмма для каждой пары переменных в модели, и если обе переменных линейно связаны и нормально распределены, должна наблюдаться овальная форма множества точек. Хотя эта процедура делает шаг вперед по сравнению с предыдущей, она считается эксплораторным подходом.

- Тест остатков [residuals test]. Этот подход состоит в том, чтобы регрессировать каждую переменную в модели на все другие переменные в модели (т.е., построить ряд регрессионных уравнений), а затем сохранить остатки. Если все переменные остатков нормально распределены (напр., на Q-Q графиках наблюдается приемлемая асимметрия и эксцесс), делается вывод, что данные обладают многомерной нормальностью.

Выбросы [outliers] могут радикально влиять на результаты анализа и также являются нарушениями нормальности. Выбросы возникают по четырем различным причинам, требующим различных способов действий:

1. *Ошибки при вводе данных в таблицу:* просмотрите ваши данные на предмет выхода некоторых оценок за пределы теоретически возможных диапазонов и других ошибок.
2. *Не определенные пропущенные значения:* убедитесь в том, чтобы пропущенные при опросах ответы, или ответы, выходящие за пределы формата выбора ответов (типа «не знаю», «у нас не было такого на работе») не принимались статистической программой за реальные оценки.
3. *Респонденты, которые не должны были войти в выборку:* удалите респондентов, которые ошибочно попали в выборку (напр., гости, пришедшие в домохозяйство, если обследовалось домохозяйство, или не-резиденты города, если обследовались жители города).
4. *Действительное не-нормальное распределение:* для действительно не-нормального распределения с крайними значениями исследователь может трансформировать данные, «втянув» выбросы (см. выше способы трансформации данных) или может предпочесть удалить выбросы и проанализировать их как отдельную выборку.

- Простые выбросы [simple outliers] – это случаи с крайними значениями применительно к единичной переменной. Наиболее распространенной практикой является определять выбросы как случаи (оценки), отличающиеся от выборочной средней оценки более чем на +/- 3 стандартных отклонения.

- Многомерные выбросы [multivariate outliers] – это случаи с крайними значениями применительно к нескольким переменным. Многомерные выбросы операционально определяются как случаи, которые имеют Расстояние Кука [Cook Distance] выше некоторой отсечки (некоторые исследователи используют отсечку 1; другие используют 4/(n – p), где p количество параметров в модели; третьи используют 4/(n – k - 1), где n количество случаев и k количество независимых переменных). Другим связанным способом определения многомерных выбросов является использование Рычага [Leverage], с выбросами, определяемыми как имеющими значение рычага выше некоторой отсечки (некоторые используют .50;, другие используют 2p/n, где p количество параметров, включая интерцепт). Расстояние Махаланобиса [Mahalanobis distance] является третьим и очень распространенным способом измерения много-

мерных выбросов. Случаи с наибольшими значениями D-квадрата Махаланобиса являются наиболее вероятными кандидатами на выбросы, на которые следует обратить внимание.

В SPSS выберите Анализировать, Регрессия, Линейная; нажмите кнопку «Сохранить»; отметьте опции Cook's, Mahalanobis и Leverage values.

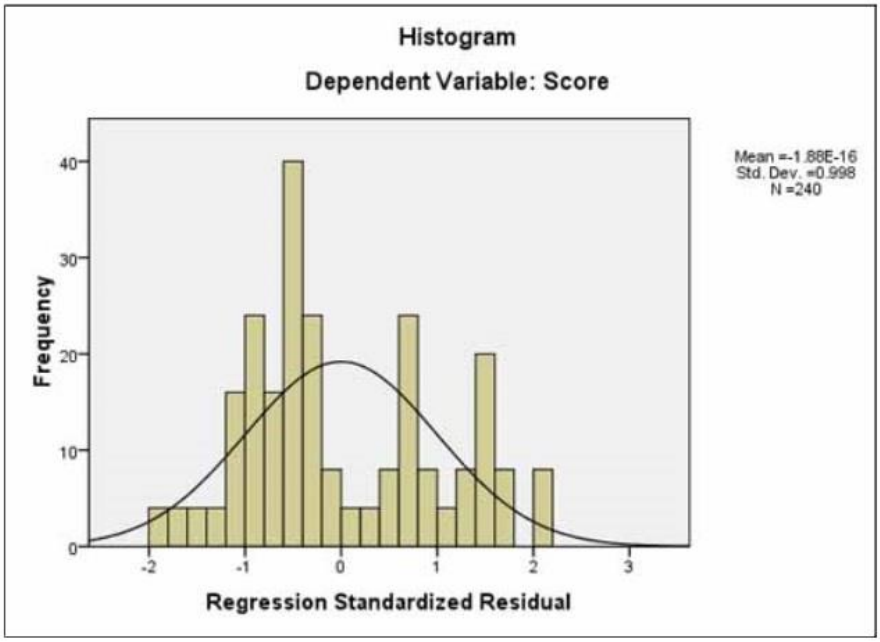
- Винсоризация данных [winsorizing data]. Альтернатива удалению выбросов заключается в том, чтобы их винсоризовать (термин происходит от фамилии автора этой процедуры). В винсоризации исследователь фиксирует наиболее крайние высокие и низкие значения, кажущиеся вполне надежными, затем сбрасывает еще более крайние случаи до этих предельных значений. Альтернативно, предельные значения можно установить в терминах процентилей, таких как 5-й и 95-й процентили значений. Все более крайние значения сводятся к значениям 5-го и 95-го процентилей. Stata обеспечивает команды winsor и wincor для внисоризации данных или корреляции данных на винсоризованной основе. Существует также макрос SAS для винсоризации. В SPSS винсоризацию можно реализовать через процедуру Recode.

  Bollinger & Chandra (2005) и другие обнаружили, что хотя обрезка или винсоризация данных может увеличивать мощность процедуры без значимого увеличения Ошибок I типа в некоторых обстоятельствах, путем «подталкивания» наблюдений в направлении средней также возможно ввести искажения.

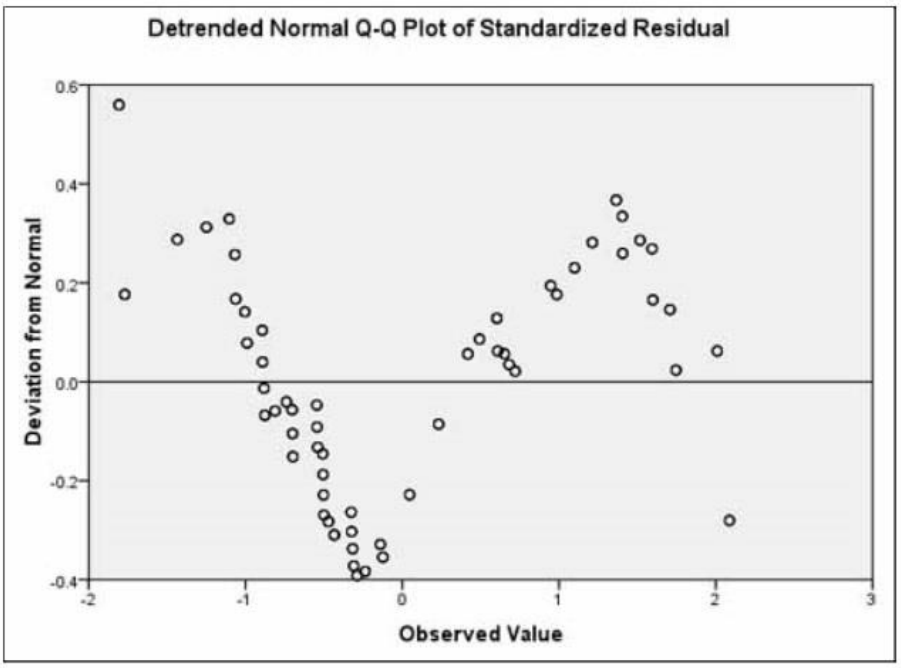
Нормально распределенная ошибка

- Гистограмма стандартизованных остатков [histogram of standardized residuals] должна показать приблизительно нормальную кривую, когда выполняется допущение регрессионного анализа и некоторых других процедур о том, что члены ошибки [error terms] нормально распределены. В SPSS окно результатов выводит нормальные кривые, наложенные на фон таких гистограмм. В SPSS выберите Анализировать, Регрессия, Линейная; нажмите кнопку «Графики»; отметьте опцию «Гистограммы».

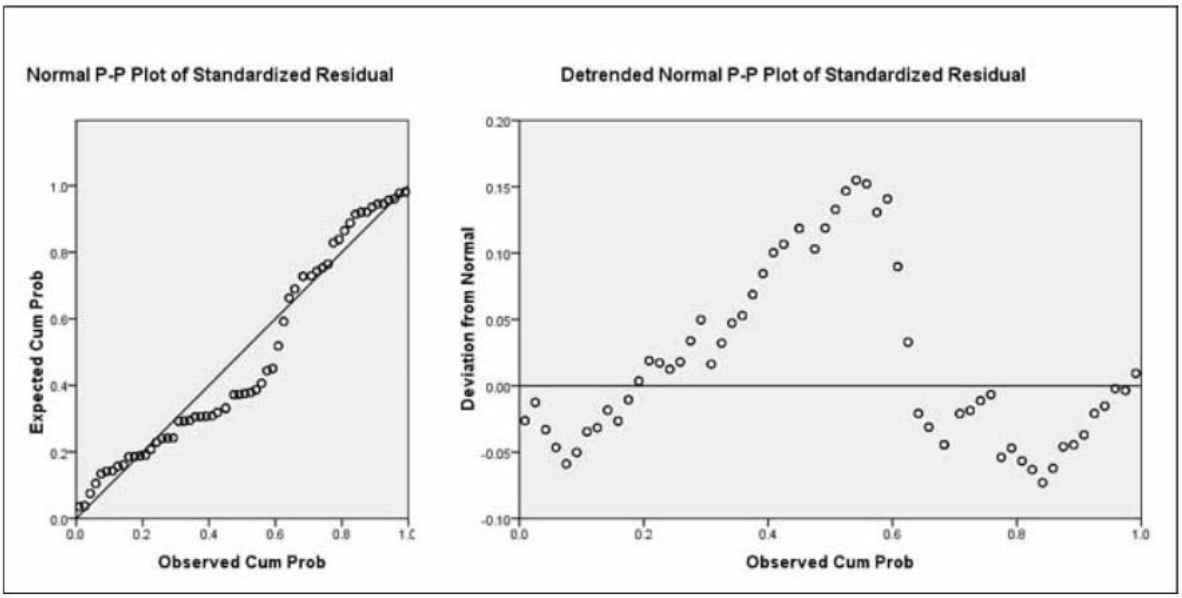
  В примере ниже зависимая переменная Score прогнозируется независимой переменной Seniority. В этом примере осмотр графика вызывает сомнения в том, что ошибка нормально распределена.



  Если сохранить остатки из верхнего примера, выбор в SPSS Анализировать, Описательные статистики, Исследовать сгенерирует, среди других статистик, асимметрию и эксцесс, которые составляют .416 и -.840, соответственно, для этих данных, что в нормальных пределах. Однако при использовании бестрендового Q-Q графика для этих остатков показывает (ниже), что остатки находятся ниже нормальных ожиданий для средней области и выше для левой и правой областей значений. Это указывает скорее на бимодальное, нежели на нормальное распределение ошибки. Это также пример того, когда «правило большого пальца» для асимметрия и эксцесса приводит к ошибочным выводам.



- График нормальной вероятности [normal probability plot], также называемый P-P графиком [P-P plot], является альтернативным методом, изображающим наблюдаемые кумулятивные вероятности возникновения стандартизованных остатков на оси Y и ожидаемые нормальные вероятности на оси X таким образом, что будет возникать 45-градусная линия, когда наблюдаемые значения подчиняются нормально ожидаемым и удовлетворяется допущение о нормально распределенной ошибке. В бестрендовом варианте паттерн точек выше и ниже горизонтальной линии 0 должен быть случайным (не таким, как на иллюстрации ниже). P-P графики можно вызвать в SPSS, выбрав Графики, P-P графики. Или выбрав Анализировать, Описательные статистики, P-P графики. Или выбрав Анализировать, Регрессия, Линейная; нажав кнопку «Графики»; отметив опцию «График нормальной вероятности».



- Тест Колмогорова-Смирнова и другие тесты нормальности [Kolmogorov-Smirnov and other normality tests]. Во множественной регрессии можно воспользоваться кнопкой «Сохранить», чтобы сохранить остатки для переменной, которая будет по умолчанию названа res_1. Другие процедуры позволяют сходным образом сохранять остатки. Затем можно воспользоваться процедурой SPSS «Examine» для применения к этой переменной тестов нормальности, таких как тесты нормальности Колмогорова-Смирнова (Лиллифорса) и Шапиро-Вилкса. Выберите Анализировать, Описательные статистики, Исследовать, нажав кнопку «Графики» и отметив опцию «Графики нормальности с тестами». Эти тесты обсуждались выше в разделе по тестированию нормальности распределения. Оба теста должны быть незначимыми, если остатки нормально распределены.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual	.163	240	.000	.940	240	.000

a. Lilliefors Significance Correction

Однородность дисперсий

- Тест однородности дисперсии Ливена [Levene`s test of homogeneity of variance] проверяет допущение, что каждая группа (категория независимой переменной) обладает той же самой дисперсией в отношении интервальной зависимой переменной. Если статистика Ливена значима на 5%-м уровне или ниже, исследователь отвергает нулевую гипотезу, что группы обладают равными дисперсиями. Тест Ливена устойчив для отклонений от нормальности. Тест Ливена появляется в SPSS в процедурах ONEWAY и T-TEST, а также EXAMINE. Обычно он основывается на отклонениях от средней групповой оценки зависимой переменной. Однако опция «spread-level plot» модуля EXAMINE также выводит в окне результатов версии теста Ливена, базирующиеся на отклонениях от медианы групповой оценки с учетом степеней свободы и 5%-ной обрезанной средней групповой оценке. Тест Ливена более устойчив перед лицом не-нормальности, чем более традиционные тесты, подобные тесту Бартлетта. В SPSS выберите Анализировать, Сравнить средние, Односторонний ANOVA; нажмите кнопку «Опции»; отметьте «Тест однородности дисперсии».

- Тест однородности дисперсий Брауна и Форсайта [Brown & Forsythe`s test of homogeneity of variances] – относительно недавний тест, базирующийся на критике теста Ливена. Он тестирует равенство средних групповых оценок. Тест Брауна-Форсайта более устойчив, чем тест Ливена, когда группы различаются по размеру и абсолютные оценки отклонений (отклонения от средней групповой оценки) обладают сильной асимметрией, вызывая нарушение допущения о нормальности и допущения о равенстве дисперсий. В SPSS выберите Анализировать, Сравнить средние, Односторонний ANOVA; нажмите кнопку «Опции»; отметьте «Браун-Форсайт».

- В примере ниже Zodiac (знак Зодиака) используется для прогнозирования Polviews (либеральных или консервативных политических взглядов). Как и ожидалось, ANOVA является незначимым, показывая, что Zodiac не прогнозирует Polviews. Поскольку статистика Ливена оказалась незначимой, исследователь не отверг нулевую гипотезу, что группы обладают равными дисперсиями. Частоты для Zodiac, не показанные здесь, показывают, что размеры групп не различались заметным образом, так что результаты теста Ливена были приняты. Однако если бы размеры групп заметно различались, потребовалось бы применить тест Брауна и Форсайта. Для этих данных тест Брауна и Форсайта также оказался незначимым и таким образом не отличался в своих выводах от теста Ливена.

ANOVA

Think of Self as Liberal or Conservative

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	22.591	11	2.054	1.120	.341
Within Groups	2601.119	1419	1.833		
Total	2623.711	1430			

Test of Homogeneity of Variances

Think of Self as Liberal or Conservative

Levene Statistic	df1	df2	Sig.
.579	11	1419	.847

Robust Tests of Equality of Means

Think of Self as Liberal or Conservative

	Statistic[a]	df1	df2	Sig.
Welch	1.127	11	547.334	.337
Brown-Forsythe	1.112	11	1358.122	.348

a. Asymptotically F distributed.

- Тест Велча [Welch test] является альтернативой F-теста в ANOVA и используется, когда нельзя предположить равенства средних групповых оценок, как этого не может быть в приведенном примере. Однако в этом примере тест Велча также демонстрирует отсутствие связи Zodiac с Polviews на значимом уровне.

- Тест однородности дисперсии Бартлетта [Bartlett`s test of homogeneity of variance] является более старым альтернативным тестом. Тест Бартлетта является хи-квадрат статистикой с (k – 1) степенями свободы, где k количество категорий в независимой переменной. Тест Бартлетта зависит от выполнения допущения о нормальности и поэтому тест Ливена полностью вытеснил его.

- F-макс тест [F-max test] возникает в контексте ANOVA. Если соотношение размера наибольшей и наименьшей группы не слишком велико (напр., 4:1 или меньше), и если F-макс (соотношение дисперсий в наибольшей и наименьшей группе) 10:1 или меньше, тогда «правило большого пальца» состоит в том, что допущение об однородности дисперсий не оказывается проблемой. Если размеры групп различаются больше, чем 4:1, тогда соотношение дисперсий должно быть не более чем 3:1.

Сферичность

В плане с повторными измерениями таблицы одномерного ANOVA [inivariate ANOVA] не будут интерпретироваться надлежащим образом, если матрица дисперсии/ковариации [variance/covariance matrix] зависимой переменной имеет круговую форму. Сферичность входит в число допущений ряда процедур, например, MANOVA. Чтобы заключить об отсутствии нарушения допущения о сферичности, *тест сферичности Бартлетта* [Bartlett`s test of sphericity] не должен оказываться значимым. В SPSS выберите Анализировать, Общая линейная модель, Многомерная; нажмите кнопку «Опции»; отметьте «Матрица SSCP остатка».

Однородность матриц дисперсии-ковариации

- М тест Бокса [Box`s M test] тестирует многомерную однородность дисперсий и ковариаций, чего требуют MANOVA (см. файл «Многомерный GLM, MANOVA и MANCOVA в SPSS») и некоторые другие процедуры. Когда М оказывается незначимым, исследователь принимает нулевую гипотезу, что группы не различаются. Было показано, что он является консервативным тестом, слишком часто не отвергающим нулевую гипотезу. Он также чрезвычайно чувствителен к нарушениям многомерной нормальности. В SPSS выберите Анализировать, Общая линейная модель, Многомерная; нажмите кнопку «Опции»; отметьте «Тесты однородности».

Однородность регрессий/ тест параллелизма

- В ANCOVA и MANCOVA наклоны регрессионных линий должны быть теми же самыми для каждой группы, формируемой категориальными переменными и рассчитываемыми в отношении зависимых переменных. Чем больше нарушается это допущение, тем более консервативными становятся ANCOVA и MANCOVA (тем выше вероятность совершить Ошибку II типа, ошибочное отвергание нулевых гипотез, приводящую к отрицанию действительно существующих связей). (См. файл «Одномерный GLM, ANOVA и ANCOVA в SPSS»)

- Нарушение допущения об однородности регрессий указывает на эффект взаимодействия между ковариатом (-ами) и фактором (-ами). При запуске модели ANCOVA или MANCOVA в SPSS включите в модель опции взаимодействия между ковариатами и каждым независимым фактором – любые значимые эффекты взаимодействия показывают, что допущение об однородности регрессионных коэффициентов было нарушено.

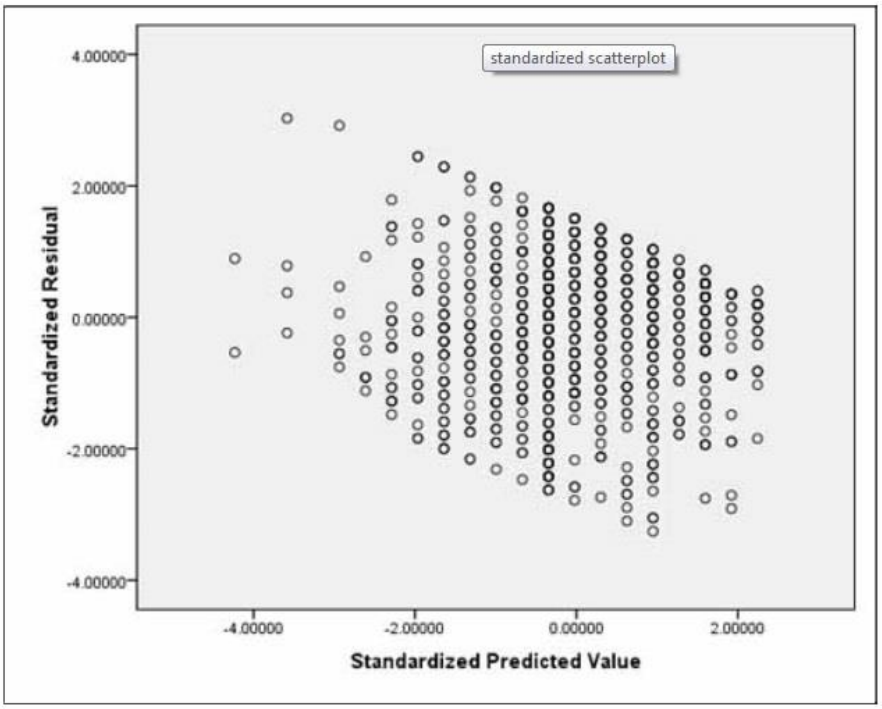
- Однородность регрессии в SPSS можно протестировать, выбрав Анализировать, Общая линейная модель, Одномерная; нажав кнопку «Модель»; выбрав опцию «Вручную»; введя модель со всеми основными эффектами факторов и ковариатов и взаимодействие (-ия) ковариата (-ов) с фактором (-ами). Эти эффекты взаимодействия должны быть незначимыми, если выполняется допущение об однородности регрессии.

- *Тесты параллелизма* сходным образом позволяют проверить, что наклон регрессионных линий является тем же самым для каждого уровня категориальной группирующей переменной. Например, в SPSS тест параллелизма содержится в процедурах «Обычной регрессии» [Ordinal regression] и «Пробит-модели ответа» [Probit response model].

Гомоскедастичность

- Стандартизованная скатерграмма [standardized scatterplot] стандартизованной прогнозируемой зависимой переменной (ZPR_1 в SPSS) и стандартизованных остатков (ZRE_1 в SPSS, или можно использовать стьюдентизованные удаленные остатки, SRE_1; или можно использовать наблюдаемую зависимую переменную против ZRE_1 или SRE_1) должна показывать случайный паттерн на всем диапазоне ZPR_1, когда в регрессии ошибка является гомоскедастичной – то есть, когда регрессионная модель оказывается одинаково точной на всем диапазоне зависимой переменной. Иногда для проверки гомоскедастичности также используется график остатков и актуальной зависимой переменной. В SPSS выберите Анализировать, Регрессия, Линейная; нажмите кнопку «Графики».

- Если ZPR_1, ZRE_1 или другие необходимые переменные были сохранены, вы можете также использовать Графики, Стандартный диалог, Скаттер/Точка. В окне вывода результатов ниже, например, использовалось образование для прогнозирования дохода и были сохранены стандартизованные прогнозируемые и остаточные значения. График преимущественно имеет форму облака (указывая на гомоскедастичность), но в нем есть некоторый паттерн, показывающий, что более высокие прогнозируемые значения имеют более низкие остатки (недостаток гомоскедастичности).



- Тест Голдфелда-Квандта [Goldfeld-Quandt test] является формальным тестом гомоскедастичности, использующимся, когда ошибки принимают форму дыма из трубы (веера). Хотя он и не поддерживается напрямую в SPSS, этот тест подразумевает реализацию отдельных регрессий для более высоких и более низких оценок в терминах значений зависимой переменной, а затем реализацию F-теста отношения суммы квадратов ошибок [error sum of squares] (ESS) для этих двух регрессий. Допущение о гомоскедастичности выполняется в степени, в которой отношение этих двух ESS коэффициентов приближается к 1.0.

- Тест Глейзера [Glejser test] является формальным тестом гомоскедастичности, использующимся, когда ошибки принимают форму галстука-бабочки (бимодальную форму) и когда размер выборки не слишком мал. Остатки регресируются на независимую переменную (или некоторую функцию ее) и если независимая переменная имеет значимый коэффициент b, исследователь заключает об отсутствии гомоскедастичности.

- Тест Парка [Park test] является связанным тестом гомоскедастичности. Квадраты остатков (или их логарифм) регрессируются на независимую переменную (или ее логарифм) и если независимая переменная имеет значимый коэффициент b, исследователь заключает об отсутствии гомоскедастичности.

- Тест Бройша-Пагана-Годфри [Breusch-Pagn-Godfrey test] является другим тестом гомоскедастичности для крупных выборок. Квадраты остатков стандартизуются делением на среднее квадратов остатков (сумма квадратов регрессии (RSS), деленная на N), давая в результате генерализованные остатки. Эти генерализованные остатки затем регрессируются на все зависимые переменные (m переменных), на которые падает подозрение в гетероскедастичности. Сумма квадратов ошибки (ESS), деленная на 2, для этой регрессии подчиняется хи-квадрат распределению с (m – 1) степенями свободы. Обнаружение значимости говорит о том, что нулевая гипотеза отвергается и допущение о гомоскедастичности не выполняется. Тест Бройша-Пагана-Годфри более мощный, чем тесты Голдфелда-Квандта или Вайта.

- Тест Вайта [White`s test] вовлекает регрессирование квадратов остатков на все независимые переменные плюс все квадраты независимых переменных плюс все взаимодействия независимых переменных (в отношении полного набора t переменных-предикторов в целом). Полученный R^2 умножается на N, давая статистику с хи-квадрат распределением с t степенями свободы. Если нулевая гипотеза отвергается, допущение о гомоскедастичности не выполняется. Тест Вайта не требует предварительного знания формы гомоскедастичности. Он менее мощный, чем другие тесты, такие как тест Голдфелда-Квандта.

Линейность

Тестирование нелинейности необходимо, поскольку корреляция, регрессия и другие представители общей линейной модели (GLM) подразумевают линейность. Существует широкое разнообразие методов:

- Графические методы. Простая инспекция скаттерграфиков [scatterplots] является наиболее распространенным не статистическим способом выявления нелинейности в связях переменных. Еще лучше использовать график стандартизованных остатков против стандартизованных оценок (подогнанных значений) зависимой переменной. При отсутствии нелинейности такой график должен показать случайный паттерн. В регрессии, в качестве «правила большого пальца», индикатором возможной нелинейности является превышение стандартного отклонения остатков над стандартным отклонением зависимой переменной. Добавление в модель нелинейных членов, таких как квадраты или кубы независимых переменных с дальнейшим отслеживанием того, наблюдаются ли значимые улучшения R^2 в регрессии или показателей соответствия в структурном моделировании, является другим способом тестирования нелинейности.

- Подгонка кривой тестами различий R-квадратов [curve fitting with R-squared difference tests]. Модуль SPSS «Подгонка кривой» (Анализировать, Регрессия, Подгонка кривой) рассчитывает R-квадрат для линейной модели как для линейной модели, так и для других различных нелинейных моделей. Затем можно использовать F-тест различий R-квадратов между этими моделями, чтобы посмотреть, имеет ли нелинейная модель значимо более высокий R-квадрат, чем линейная модель. Заметьте, что некоторые нелинейные модели не могут быть подогнаны, если в данных есть пропущенные значения, тогда будет необходимо предварительно их вставить, воспользовавшись модулем «Пропущенные значения» (Анализировать, Анализ пропущенных значений), если исследователь желает сравнить эти модели. Предупреждение: этот метод вовлекает множество сравнений, которые делают номинальный уровень значимости более «мягким», чем это должно было бы быть, поэтому Бонферрони или другие корректировки могут потребоваться в зависимости от количества сравниваемых моделей.

- Тест линейности ANOVA. Можно рассчитать таблицу ANOVA для линейных и нелинейных компонентов любой пары переменных. Если F-значение нелинейного компонента оказывается ниже критического уровня (напр., < .05), тогда там присутствует значимая нелинейность. В SPSS выберите Анализировать, Сравнить средние, Средние; нажмите кнопку «Опции»; отметьте «Тест линейности». Сходным образом, опция «Контрасты» в ANOVA может быть использована как тест на существование линейных, квадратических и других полиномиальных связей.

- Эта, отношение корреляции [Eta, correlation ratio]. Эта является коэффициентом нелинейной ассоциации. Для линейных связей eta равен коэффициенту корреляции (r Пирсона). Для нелинейных связей он больше, поэтому разница между eta и r служит измерением степени нелинейности связи. Eta обсуждается в отдельном разделе по ассоциации. В SPSS выберите Анализировать, Сравнить средние, Средние; нажмите кнопку «Опции»; отметьте «Таблица ANOVA и Eta».

- Добавление нелинейных членов в модель. Квадратичный член корреспондирует с нелинейным паттерном единичной кривой, подъемом-опусканием. Кубический член корреспондирует с нелинейным паттерном подъема- опускания или опускания-подъема. В целом, количество кривых оказывается меньше, чем порядок полиномиального члена, например, взведение члена в четвертую степень будет продуци-

ровать три кривых. Более широкий подход состоит в том, чтобы добавить в модель все возможные степенные члены, а затем сбрасывать те, которые оказываются незначимыми.

- RESET-тест Рамсея (тест ошибки спецификации регрессионной модели) [Ramsey`s RESET test (regression specification error test)]. Общий тест Рамсея ошибки спецификации функциональной формы является F-тестом различий R^2 между линейным и нелинейным допущениями. Он обычно используется в виде серии последовательных анализов, чтобы проверить, следует ли добавлять в модель степенные трансформации. Для линейной модели, правильно специфицированной в ее функциональной форме, нелинейные трансформации подобранных значений не должны оказываться полезными в прогнозировании зависимой переменной. Хотя STATA и некоторые статистические пакеты определяют RESET-тест как тест, позволяющий убедиться, что в модели «нет опущенных переменных», это тест линейности, а не общий тест спецификации. Он тестирует, не были ли опущены какие-либо нелинейные трансформации специфических независимых переменных. Но он не тестирует, были ли опущены другие релевантные линейные или нелинейные переменные.
    1. Запустите регрессионный анализ, чтобы получить R_0^2, оригинальную множественную корреляцию.
    2. Сохраните прогнозируемые значения (Y`s).
    3. Перезапустите анализ, используя степенные функции прогнозируемых значений (напр., их квадраты и кубы) как дополнительные независимые переменные для RESET-теста Рамсея функциональной формы, где проверяется, что ни одна из независимых переменных не имеет нелинейной связи с зависимой переменной. Или, альтернативно, перезапустите анализ, используя степенные функции независимых переменных, чтобы протестировать их индивидуально.
    4. Получите $R_н^2$, новую множественную корреляцию.
    5. Примените F-тест, где $F = (R_н^2 - R_0^2)/[(1 - R_н^2)/(n – p)]$, где n размер выборки и p количество параметров в модели,
    6. Интерпретируйте F: для адекватно специфицированной модели F должен быть незначимым.

Мультиколлинеарность

- Мультиколлинеарность в регрессионных моделях – это неприемлемо высокий уровень корреляции среди независимых переменных, так что влияния отдельных независимых переменных на зависимую невозможно отделить друг от друга. При мультиколинеарности регрессионные коэффициенты не искажаются, однако оценки сравнительного влияния независимых переменных и их совместное влияние оказываются ненадежными. (То есть, бета-веса и R-квадраты невозможно надежно интерпретировать, невзирая на то, что прогнозируемые значения по-прежнему остаются наилучшей оценкой при использовании данных независимых переменных.) В качестве «правила большого пальца», корреляции между независимыми переменными выше .80 сигнализируют о возможной проблеме мультиколлинеарности. Сходным образом, о высокой мультиколлинеарности сигнализирует ситуация, в которой высокий R-квадрат и значимые F-тесты модели сочетаются с незначимыми t-тестами коэффициентов.

  Пример мультиколинеарности возник в исследовании Бюро трудовой статистики агентства департамента труда США цен портативных видеокамер. Первоначальная модель включала искусственные переменные Sony и 8mm, обе из которых корреспондировали с высокой ценой. Однако, поскольку Sony являлся в это время единственным производителем 8mm видеокамер, искусственные переменные Sony и 8mm оказались мультиколлинеарными. Схожая проблема возникла в исследовании БТС стиральных машин, где обнаружилось, что «вместимость» и «количество циклов» оказались мультиколлинеарными. В каждом исследовании одну из коллпнеарных переменных потребовалось удалить из модели.

  Если полная мультиколлинеарность приводит к бесконечно большим стандартным ошибкам и неопределенным коэффициентам, более распространенная ситуация высокой мультиколлинеарности приводит к большим стандартным ошибкам, большим доверительным интервалам и снижению мощности (повышению шансов на ошибку II типа – выводу об отсутствии связи, когда она в действительности существует – неверному отверганию нулевой гипотезы о том, что коэффициенты не отличаются от нуля). R-квадрат является высоким. Коэффициенты и их стандартные ошибки становятся чувствительны к изменениям даже при небольшом количестве наблюдений.

  Способы борьбы с высокой мультиколлинеарностью обсуждаются в файле «Множественная регрессия в SPSS».

- Толерантность [tolerance] определяется как 1 - R-квадрат, где R-квадрат является множественным R данной независимой переменной, регрессированной на все другие независимые переменные. Если значение толерантности оказывается ниже некоторого уровня отсечки, обычно .20, данная независимая

переменная должна быть удалена из анализа из-за мультиколлиенарности. Такой подход лучше, чем использование простого r > .80, поскольку толерантность смотрит на независимую переменную в ее связи со всеми другими независимыми переменными и таким образом учитывает эффекты взаимодействия наряду с простыми корреляциями. В SPSS выберите Анализировать, Регрессия, Линейная; нажмите кнопку «Статистики»; отметьте опцию «Диагностики коллинеарности».

- Фактор инфляции дисперсии, VIF [variance inflation factor]. Заметьте, что фактор инфляции дисперсии, VIF, может использоваться вместо толерантности, поскольку VIF является просто обратным показателем толерантности. «Правило большого пальца» состоит в том, что VIF > 4.0, когда мультиколлинеарность оказывается проблемой. Некоторые авторы используют более мягкий уровень отсечки VIF = .5.0. В SPSS выберите Анализировать, Регрессия, Линейная; нажмите кнопку «Статистики»; отметьте опцию «Диагностики коллинеарности».

- Показатели состояния [condition indices]. Обсуждаются более подробно в файле «Множественная регрессия в SPSS», показатели состояния свыше 15 указывают на возможные проблемы мультиколлинеарности, свыше 30 указывают на серьезные проблемы мультиколлинеарности. В SPSS выберите Анализировать, Регрессия, Линейная; нажмите кнопку «Статистики»; отметьте опцию «Диагностики коллинеарности».

- Мультиколлинеарность в структурном моделировании (SEM):

    - *Стандартизованные регрессионные веса*. Поскольку всем латентным переменным в модели SEM назначается метрика 1, все стандартизованные регрессионные веса должны находиться внутри диапазона от -1 до +1. Когда существует проблема мультиколлинеарности, вес, близкий к 1 указывает, что две данные переменные близки друг другу настолько, что являются идентичными. Когда эти две практически идентичные латентные переменные затем используются как причины третьей латентной переменной, процедура SEM будет испытывать трудности с расчетом отдельных регрессионных весов для двух путей от этих практически идентичных переменных к третьей переменной. В результате она может прийти к одному стандартизованному регрессионному весу, превышающему +1 и одному весу меньше -1 для этих двух путей.

    - *Стандартные ошибки нестандартизованных регрессионных весов.* Сходным образом, когда существуют две практически идентичные латентные переменные, и они используются как причины третьей латентной переменной, трудности с расчетом отдельных регрессионных весов могут отразиться в гораздо более высоких стандартных ошибках для этих путей, чем для других путей в модели, что является признаком высокой мультиколлинеарности этих двух практически идентичных переменных.

    - *Ковариации оценок параметров.* Аналогичным образом, те же самые трудности с расчетом отдельных регрессионных весов могут отразиться в высоких ковариациях оценок параметров для этих путей, более высоких, чем ковариации оценок параметров для других путей в модели.

    - *Оценки дисперсий.* Другим эффектом синдрома мультиколлинеарности могут быть негативные оценки дисперсий. В верхнем примере с двумя практически идентичными латентными переменными, влияющими на третью латентную переменную, оценка дисперсии этой третьей переменной может оказаться негативной.

Независимость

- Коэффициент Дурбина-Ватсона [Durbin-Watson coefficient]. Независимые наблюдения допускаются многими статистическими процедурами, включая множественную и логистическую регрессию. Примерами не-независимых наблюдений являются исследования «до-после», панельные исследования и данные с парными сравнениями. Независимость тестируется коэффициентом Дурбина-Ватсона, использующем стьюдентизов анные остатки. Статистика Дурбина-Ватсона должна быть в пределах между 1.5 и 2.5 для независимых наблюдений. В окне вывод результатов SPSS она находится в таблице «Результаты модели». В SPSS выберите Анализировать, Регрессия, Линейная; нажмите кнопку «Статистики»; отметьте опцию «Дурбин-Ватсон».

- Графический метод. Остатки можно изобразить против номеров случаев, отражающих порядок внесения данных в таблицу. В графике наблюдений не должно наблюдаться каких-либо систематических паттернов, если наблюдения являются независимыми. В SPSS выберите Графики, Скаттерплот.

Случайность

- Тест непрерывных последовательностей [runs test]. Используется для тестирования случайности в некоторой выборке. Заметьте, что это необходимый, но не достаточный тест на случайность выборки. Неслучайная доступная выборка, скажем, студентов определенного курса, может быть очень искаженной репрезентацией всех студентов в университете, несмотря на то, что внутри этого курса выборка может быть случайной и такая выборка может проходить тест случайных последовательностей. С другой стороны, если целенаправленно подобранная случайная выборки не проходит такого теста, это указывает, что в ней присутствуют необычные, неслучайные последовательности порядке выборки, не согласующиеся со случайной выборкой. В SPSS выберите Анализировать, Непараметрические тесты, Последовательности (Runs).

Аддитивность

- Тест аддитивности Тьюки [Tukey`s Test for additivity]. Оценивает степень, в которой количество пунктов в некотором наборе необходимо увеличить для того, чтобы достичь аддитивности. В SPSS выберите Анализировать, Шкала, Анализ надежности; нажмите кнопку «Статистики»; отметьте опцию «Тест аддитивности Тьюки».

- Трансформации для аддитивности. Если данные не проходят теста аддитивности, иногда это можно скорректировать путем добавления логарифма у (зависимой переменной) как способа трансформации.

Равенство средних

- Т-квадрат Хотеллинга [Hotelling`s T-square] – это многомерный тест равенства средних внутри некоторого набора пунктов. В SPSS выберите Анализировать, Шкала, Анализ надежности; нажмите кнопку «Статистики»; отметьте опцию «Т-квадрат Хотеллинга».

БИБЛИОГРАФИЯ

- Bollinger, Christopher R. & Chandra , Amitabh (2005). Iatrogenic specification error: A cautionary tale of cleaning data. *Journal of Labor Economics*, 23(2), 235-257.
- Boneau, C. A. (1960). The effect of violation of assumptions underlying the t-test. *Psychological Bulletin*, 57: 49-64.
- Cohen, Jacob (1969). *Statistical power analysis for the behavioral sciences*. NY: Academic Press.
- Hutcheson, Graeme and Nick Sofroniou (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Thousand Oaks, CA: Sage Publications. Chapter two covers data screening for assumptions under GLM.
- Lipsey, Mark W. & Wilson, David B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Schumacker, Randall E. and Richard G. Lomax (2004). *A beginner's guide to structural equation modeling, Second edition.* Mahwah, NJ: Lawrence Erlbaum Associates.
- Shapiro, S. S. & Wilk. M. B. (1965). An analysis of variance test for normality. *Biometrika*, 52(3), 591-599.
- Steenkamp, J.E.M., & van Trijp, H.C.M. (1991). The use of LISREL in validating marketing constructs. International Journal of Research in Marketing, 8: 283-299.
- Vasu, E. S. (1979). Non-normality in regression analysis: Monte Carlo investigation under the condition of multicollinearity. Working papers in methodology. Number 9. Institute for Research in Social Science, University of North Carolina. Chapel Hill, North Carolina, USA.