Г. В. Суходольский

СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ В ПСИХОЛОГИИ[1]

Научные гипотезы — это хотя бы в принципе проверяемые предположения. Соответственно предмету, языку формулировок и методам проверок следует различать физические, социологические и т. д. — психологические и статистические гипотезы.

Психологические гипотезы формулируются о свойствах психики, условиях, на нее влияющих, и т.д. Язык формулировок — язык научной и практической психологии, а методы — эмпирические: наблюдение, психологический эксперимент, тестирование и т. д.

Статистические гипотезы предметом своим имеют теоретико-вероятностные объекты — распределения вероятностей и их параметры. Язык формулировок предельно лаконичен: ♦ сходство объектов сравнения не случайно, следовательно, различия случайны» или «сходство объектов сравнения случайно, следовательно, различия не случайны». Методы статистической проверки стандартны. О них пойдет речь в дальнейшем.

Так как все психологические объекты суть случайные, то проверять психологические гипотезы приходится через посредство статистических гипотез. При этом психологическую гипотезу необходимо преобразовать в статистическую.

Вот как, например, это делается. В автотранспортной психологии предполагается, что время реакции (ВР) водителя увеличивается по мере увеличения длительности пребывания его за рулем. Но время реакции — это с математико-психологической точки зрения случайная величина, которая полностью определяется своим распределением вероятностей. Условие — длительность пребывания за рулём — приводит к необходимости рассматривать распределение времени реакции в зависимости (или нет) от времени. Иначе говоря, перед нами случайный процесс ВР(t). И речь, следовательно, идет о том, является этот процесс нестационарным или нет. Проверить это можно, взяв хотя бы две различные длительности «за рулём» и эмпирически определив для них условные распределения времени реакции. Сравнивая эти распределения, устанавливают статистическую степень сходства-различия, и в зависимости от вывода отклоняют или принимают психологическую гипотезу. Разумеется, длительность интервала t между распределениями должна быть достаточной для того, чтобы проявился эффект утомления и увеличения времени реакции. Иначе, при недостаточном сдвиге $\tau - t_1 - t_0$, можно получить ошибочный вывод.

---

[1] *Суходольский Г.В.* Математические методы в психологии. 3-е изд., испр. Харьков: Изд-во «Гуманитарный центр», 2008.

Так же как, изучая отдельное событие, мы должны изучать и его альтернативу, образующую вместе с ним полную группу, проверяя гипотезу, мы всегда должны иметь в виду ее альтернативу. В связи с этим принято использовать альтернативные гипотезы: $\{H_0 \vee H, \}$, где H_0 — обычно гипотеза о сходстве сравниваемых объектов, а H_1 — гипотеза об их различии, принимаемая, если будет отклонена H_0.

С формальной точки зрения надо различать простые и сложные гипотезы: сложные гипотезы есть обычно конъюнктивно дизъюнктивные функции от простых. Психологу полезно уметь записывать формулы гипотез — прежде всего статистических. Рассмотрим эти формулы:

$$H_0 = \{F_1 \sim F_2\} \text{ или } \{P_1 \sim P_2\},$$

когда сравниваются два распределения частот либо частостей; при этом мерность распределений может быть любой;

$H_0 = \{F_1 \sim F_2 \sim ... \sim F_k\}$ или $\{P_1 \sim P_2 \sim ... \sim P_k\}$, когда сравнивается несколько эмпирических распределений. Альтернативными при этом будут тоже простые гипотезы:

$$H_1 = \{F_1 \neq F_2\} \text{ или } \{P_1 \neq P_2\}, \text{ или } \{\neq F_i\}, \text{ или } \{\neq P_i\};$$

дальнейшее попарное сравнение поможет установить, есть ли в неэквивалентных группах распределений частью однородные.

Гипотезы об отдельных параметрах тоже являются простыми. Но ведь главный объект — это распределение, а распределения — в одномерном варианте — описываются двумя параметрами. Поэтому приходится рассматривать сложные гипотезы.

Для одномерного случая:

$H_0 = (M_1 \sim M_2) \wedge (D_1 \sim D_2)$, но $H_1 = [(M_1 \sim M_2) \wedge (D_1 \neq D_2)] \vee [(M_1 \neq M_2) \wedge (D_1 \sim D_2)] \vee [(M_1 \neq M_2) \wedge (D_1 \neq D_2)]$.

Какой из этих трёх вариантов будет иметь место, не известно априори. Правда, для многих «человеческих» работ свойствен первый вариант: дисперсия первая «страдает» от утомления. А для статистических методов принят как характерный второй вариант, позволяющий легко определять погрешности измерений. С третьим вариантом — зависимости дисперсий и среднего от аргумента — мы уже сталкивались в одном из числовых примеров случайного процесса.

Параметрическая формулировка статистических гипотез усложняется уже для двумерного случая: ведь здесь кроме средних дисперсий приходится сравнивать меры корреляции. В общем должно быть ясно, что выгодней исследовать распределения и проверять гипотезы о распределениях.

Рассмотрим теперь ситуацию проверки гипотезы исследователем. Формально эта ситуация описывается следующим множеством событий.

Гипотеза имеет два состояния — она либо истинна, либо ложна: τ = (И, Л); априорное распределение вероятностей этих состояний:

$$\mathcal{P}(\mathcal{T}) = (p_{\text{и}}, p_{\text{л}}).$$

Проверяя гипотезу, исследователь может совершить тоже два действия — принять либо отклонить гипотезу: D = (П, О); но при этом, учитывая состояния гипотезы как условия, он совершает правильные действия: П/Л — принять гипотезу, если она истинна, и О/Л — отклонить ее, если гипотеза ложна; он совершает также два ошибочных действия: П/Л — принять гипотезу, хотя она ложна, и О/И — отклонить гипотезу, хотя она истинна. Условные вероятности

этих правильных и ошибочных действий имеют стандартные обозначения, собственные названия и записываются в виде ассоциированной матрицы условных распределений:

$$\mathcal{P}(\mathcal{D}/\mathcal{T}) = \begin{matrix} & \text{И} & \text{Л} \\ \text{П} & 1-\alpha & \beta \\ \text{О} & \alpha & 1-\beta \end{matrix},$$

где $1 - \alpha$ = р(П / И) — доверительная вероятность, $\alpha = p$(О / И) — вероятность ошибки I рода, $\beta = p$(П / Л) — вероятность ошибки II рода, $1 - \beta$ = р(О / Л) — мощность критерия.

Еще раз обратим внимание на ошибки исследователя. Отклонение гипотезы при условии, что она истинна, т.е. О/И, называется ошибкой I рода, или «пропуском цели», или «риском заказчика» (пропускающего некондиционный товар). Напротив, принятие гипотезы при условии, что она ложна, т.е. П/Л, называется ошибкой II рода, или «ложной тревогой», или «ошибкой поставщика» (у которого заказчику чудится некондиционный товар).

Обратим внимание, что перед нами двумерная система случайных событий, заданная двумя частными распределениями, по которым можно восстановить полное совместное распределение. Сделаем это:

$$\mathcal{P}(\mathcal{DT}) = \mathcal{P}(\mathcal{D}/\mathcal{T}) * \mathcal{P}(\mathcal{T}) = \begin{matrix} & \text{И} & \text{Л} \\ \text{П} & 1-\alpha & \beta \\ \text{О} & \alpha & 1-\beta \end{matrix} * \begin{matrix} \text{И} & \text{Л} \\ p_{\text{и}} & p_{\text{л}} \end{matrix} = \begin{matrix} & \text{И} & \text{Л} \\ \text{П} & (1-\alpha)p_{\text{и}} & \beta p_{\text{л}} \\ \text{О} & \alpha p_{\text{и}} & (1-\beta)p_{\text{л}} \end{matrix}.$$

Заметим, что поведение исследователя в ситуации проверки гипотезы складывается из верных и ошибочных поступков: верные поступки заключаются в том, чтобы принять истинную гипотезу или отклонить ложную:

ВП = ПИ v ОЛ,

а ошибочные поступки складываются из принятий ложных гипотез и отклонений гипотез истинных:

ОП = ОИ v ПЛ

В полученном выше совместном распределении вероятностей поступков исследователя и состояний гипотезы по главной диагонали матрицы расположены вероятности верного поведения:

p(ВП) = p(ПИ) + p(ОЛ) = $(1-\alpha)p_{И}+(1-\beta)p_{Л}$.

По контрдиагонали расположены вероятности ошибочного поведения:

р(ОП) = р(ПЛ) + р(ОИ) = $\alpha p_{И} + \beta p_{Л}$.

Анализируя эти равенства (удобнее последнее), можно видеть следующее. Если вероятности состояний гипотезы примерно равны: $p_{И} \sim p_{Л}$, то для минимизации вероятности ошибочных поступков р(ОП) необходимо, чтобы вероятности ошибок I и II рода были одинаково малы $\alpha \approx \beta$. Иначе говоря, исследователю следует быть бдительным, но не сверхбдительным.

Если гипотеза скорее ложна, чем истинна, т.е. $p_{\text{И}} << p_{Л}$, то для минимизации Р(ОП) необходимо, чтобы $\alpha >> \beta$, т.е. чтобы вероятнее были пропуски цели, а не ложные тревоги. Заметим, что это ситуация неожиданности. В такой ситуации внимание притуплено и любой человек скорее пропустит цель, нежели совершит ложную тревогу. Есть, однако, люди невнимательные, рассеянные, у которых вообще вероятность ошибки I рода (α) является повышенной. О таких людях в соответствующих ситуациях говорят: «проморгал», «прохлопал» и т.п.

Наконец, рассмотрим ситуации, в которых гипотеза скорее истинна, чем ложна, т.е. $p_И >> p_Л$. Для минимизации функционала вероятности ошибочных поступков теперь требуется, чтобы $\alpha << \beta$. Психологически это хорошо известная ситуация напряженного ожидания, когда чудится, мерещится, кажется и т. п. Многие люди обладают свойством повышенной тревожности, что как раз и означает повышенный уровень вероятности ошибок II рода.

Восстановив полное двумерное распределение событий в нашей ситуации, мы получили возможность рассчитать и другие распределения, возможные в данной системе.

Так, суммируя полное распределение по столбцам (по состояниям гипотезы), получаем безусловное распределение поступков

$$P(\mathcal{D}) = \sum_{\mathcal{T}} P(\mathcal{D}\mathcal{T}) \Leftrightarrow \begin{matrix} \\ \text{П} \\ \text{О} \end{matrix} \begin{pmatrix} (1-\alpha)p_И + \beta p_Л \\ \alpha p_И + (1-\beta)p_Л \end{pmatrix}.$$

Наконец, деля смешанным делением полное распределение на полученную матрицу поступков исследователя, получаем ассоциированную матрицу условных распределений состояний гипотезы при условии принятия или отклонения гипотезы. Это распределение так называемых апостериорных вероятностей гипотезы — частный случай формулы Байеса:

$$P(\mathcal{D}\mathcal{T}) \div P(\mathcal{D}) =$$

$$= \begin{matrix} \\ \text{П} \\ \text{О} \end{matrix} \begin{matrix} \text{И} & \text{Л} \\ \end{matrix} \begin{pmatrix} \dfrac{(1-\alpha)p_И}{(1-\alpha)p_И + \beta p_Л} & \dfrac{\beta p_Л}{(1-\alpha)p_И + \beta p_Л} \\ \dfrac{\alpha p_И}{\alpha p_И + (1-\beta)p_Л} & \dfrac{(1-\beta)p_Л}{\alpha p_И + (1-\beta)p_Л} \end{pmatrix} \Leftrightarrow$$

$$\Leftrightarrow \begin{matrix} \\ \text{П} \\ \text{О} \end{matrix} \begin{matrix} \text{И} & \text{Л} \\ \end{matrix} \begin{pmatrix} P(\text{И}/\text{П}) & P(\text{Л}/\text{П}) \\ P(\text{И}/\text{О}) & P(\text{Л}/\text{О}) \end{pmatrix}.$$

ПРОВЕРКА ГИПОТЕЗ ПО КРИТЕРИЯМ

Статистические критерии представляют собой особые случайные величины, полученные как неслучайные функции от сравниваемых теоретико-вероятностных объектов — рядов и функций распределений вероятностей, отдельных параметров. Кроме того, в распределениях значений критериев учитывается количество наблюдений, из которого получены сравниваемые объекты, и доверительные вероятности $1 - \alpha$ или α — так называемые уровни значимости, при которых принимаются решения об истинности либо ложности проверяемой гипотезы.

Принято различать критерии по мощности, различать непараметрические и параметрические, а также парные и множественные критерии.

Вспомним, что мощность критерия — это вероятность с его помощью отбрасывать гипотезу, если она ложна. Среди критериев наиболее мощным считается χ^2 — К. Пирсона. По нему поверяются все другие статистические критерии, которые, как правило, менее мощны. В течение XX века статистиками выдумано достаточно много разнообразных критериев, с которыми можно познакомиться по таблицам математической статистики[1].

Для инструментария психолога необходимы и достаточны пять критериев: x^2 —

[1] *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. 3-е изд. М., 1988.

К. Пирсона, λ — Колмогорова—Смирнова, *t* — Стьюдента, *F* — Снедекора—Фишера и *G* — Кохрана. С ними мы познакомимся на примерах.

χ^2 — непараметрический критерий. Это означает, как следует из названия, что он «работает» с распределениями, а не с параметрами. Теоретически x^2 представляет собой сумму квадратов стандартных нормально-распределенных величин, и он зависит только от числа слагаемых, которые называются «числом степеней свободы».

Надо сразу же разобраться с этим числом. Пусть суммируется вектор длиной *k;* *k* – 1 слагаемое может быть каким угодно, но *k-е* должно дополнять сумму компонент до заданной константы. Например, по условию нормировки, сумма вероятностей в *k* интервалах группировки должна быть равна единице. При этом *k –1* вероятность, конечно в пределах своего определения, может быть любой, а вот *k-я* должна дополнять сумму *k* –1 слагаемого до единицы. И так при суммировании компонентов любого вектора связывается одна степень свободы, а если суммирований s, то всего связывается (*k* – l) (s –1) степеней свободы.

К. Пирсон первоначально использовал (и ввел) χ^2 для сравнения эмпирического и теоретического распределений частот или частостей:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - f_i^*)^2}{f_i^*} = n \sum_{i=1}^{k} \frac{(p_i - p_i^*)^2}{p_i^*},$$

где f_i, p_i — эмпирические, f_i^*, p_i^* — теоретические частоты либо вероятности, *n* — объём выборки, *k* — число интервалов группировки, одинаковое для эмпирического и теоретического распределений.

Позже стало ясно, что формулу χ^2 можно модифицировать для сравнения ряда эмпирических распределений, для оценки значимости корреляции по Чупрову, для проверки однородности корреляционных матриц, значения которых имеют размерность вероятностей. Рассмотрим в качестве примера использования χ^2 простейшую задачу: в двух выборках получены оценки вероятности события А. Спрашивается, можно ли их считать статистически различными?

Вспомним, что говоря о вероятности А, мы должны иметь в виду и вероятность «не-А», — иначе говоря, здесь речь идет о сравнении двух выборочных распределений. При этом расчетная формула эмпирического значения χ^2 будет такой:

$$\chi^2_э = \sum_{i=1}^{s} n_i \sum_{j=1}^{k} \frac{(p_{ij} - p_i^*)^2}{p_i^*},$$

где *s* — число сравниваемых выборок, n_i — их объёмы, i = $\overline{1,S}$, k — число интервалов группировки. Заметим, что в качестве p_i * — теоретической вероятности должна выступать средняя взвешенная из эмпирических вероятностей выборок: разумно считать, что рассеивание выборок симметрично:

$$p_i^* = \frac{1}{\sum_{i=1}^{s} n_i} \sum_{j=1}^{k} p_{ij} n_i.$$

Запишем числовые данные задачи в виде матрицы, в последнем столбце которой вычислены теоретические вероятности:

n	30	60	90
pA	0,6	0,4	0,47
pA	0,4	0,6	0,53

Теперь по формуле x^2 вычисляем:

$$\chi^2_Э = 50\left(\frac{0,13^2}{0,47}+\frac{0,13^2}{0,53}\right)=$$

$$=100\left(\frac{0,07^2}{0,47}+\frac{0,07^2}{0,53}\right)=3,18.$$

Проверка гипотезы сводится к сопоставлению эмпирического значения критерия с теоретическим квантилем, выбираемым для эмпирического числа степеней свободы из специальных таблиц. В табл. 1 показан фрагмент такой таблицы для χ^2-критерия.

Таблица 1. Фрагмент таблицы квантилей X^2-критерия
(1 -α — доверительные вероятности, *v* — число степеней свободы)

V	0,900	0,950	0,975	0,990	0,995
1	2,71	3,84	5,02	6,64	7,88
2	4,60	5,99	7,38	9,21	10,60
4	7,78	9,49	11,19	13,28	14,86
6	10,64	12,58	14,45	16,81	18,55
9	14,68	16,98	19,32	21,67	23,59

Существует два способа принятия решения при проверке гипотезы. Один, наиболее распространенный, так как не требует обстоятельных таблиц, состоит в том, что заранее выбирают доверительную вероятность, или уровень значимости, из таблиц для него находят при эмпирическом числе степеней свободы теоретический квантиль и сравнивают: если эмпирический квантиль не превышает теоретический, принимается гипотеза о сходстве; если превышает, то сходство считают случайным и принимают гипотезу о различиях. В нашем примере, действуя по этому способу для числа степеней свободы $(s-1)(k-1) = (2-1)(2-1) = 1$ и доверительной вероятности, скажем, 0,95, находим «критический» квантиль 3,84. Сравнивая с ним эмпирическое значение $\chi^2_Э$ = 3,18, видим, что оно меньше, следовательно, нулевую гипотезу о сходстве этих так, на первый взгляд, различающихся вероятностей (0,6 против 0,4) отклонить нельзя.

Второй способ более гибкий и верный. Ну почему надо ограничиваться уровнем значимости 5%? А если 6% или 7%. Велика ли разница? — Нет. Второй способ оставляет решение пользователю данных: лишь утверждается, что нулевая гипотеза может быть отклонена при определенном уровне значимости. Для этого в табл. 12 надо найти значение, равное или близкое эмпирическому, и интерполяцией определить соответствующую доверительную вероятность. В нашем случае на 0,01 доверительной вероятности приходится (3,84-2,71):0,05 = 0,22: следовательно, $\chi^2_Э$ = 3,18 соответствует доверительная вероятность 0,92. Таким образом, по второму способу мы можем отклонить H_0 о сходстве вероятностей на уровне значимости 8%. Это, конечно, более правильно, нежели по общепринятому способу критического значения.

Итак, вывод по примеру: выбранные вероятности нельзя признать одинаковыми. Они различны с доверительной вероятностью примерно 0,92.

В конце главы 19[2] рассматривался пример оценки, по Чупрову, корреляции педагогических оценок у десяти студентов.

Было получено значение K^2 = 0,294. Заметим, что в формуле коэффициента записа-

[2] *Суходольский Г.В.* Математические методы в психологии. 3-е изд., испр. Харьков: Изд-во «Гуманитарный центр», 2008.

на уменьшенная в n раз величина $\chi^2_Э$. Поэтому для проверки значимости корреляции, по Чупрову, надо эмпирическое значение определять как $n * K^2 = \chi^2_Э$. В примере это дало бы 2,94. По табл. 1, опять так же интерполируя, находим приращение 0,22 на 0,01 доверительной вероятности. Становится ясно, что отклонить нулевую гипотезу об отсутствии корреляции в том примере можем с доверительной вероятностью 0,914÷0,915. Иначе говоря, значение $K^2 = 0{,}294$, при п = = 10, отличается от нулевого с доверительной вероятностью 0,91 ÷ 0,915, или на уровне значимости 9% ÷8,5%.

Критерий λ — Колмогорова—Смирнова — это парный непараметрический критерий, предназначенный для сравнения эмпирической и теоретической интегральных функций. Теоретически λ — это параметр распределения Колмогорова. Он табулирован и представлен в табл. 2.

Таблица 2. Фрагмент таблицы квантилей λ-критерия

λ	0,40	0,45	0,50	0,55	0,60	0,65	0,70
1-α	0,997	0,987	0,964	0,923	0,864	0,792	0,711

Практически $\lambda = d_{max}\sqrt{n}$, где d_{max} — максимальная разность значений эмпирической и теоретической функций распределения, п — объем выборки. Например, сравним эмпирическое распределение педагогических оценок с теоретическим равномерным распределением:

x	$F(x)$	$Fэ(x)$	d
2	0,125	0,150	0,025
3	0,375	0,355	0,020
4	0,625	0,635	0,010
5	0,875	0,825	0,050

$n = 100$

Можно видеть, что $d_{max} = 0{,}05$. Так что $\lambda = 0{,}05\sqrt{100} = 0{,}5$. В табл. 13, где сопоставлены значения λ и соответствующие значения доверительных вероятностей, ищем подходящее значение и находим для $\lambda = 0{,}5$ при $1 - \alpha = 0{,}964$; следовательно, на 3,6%-ном уровне значимости нулевая гипотеза должна быть принята.

Перейдем к рассмотрению параметрических критериев. Среди них на первом месте стоит t-критерий Стьюдента — как исторически, так и по распространённости.

Создателем t-критерия был молодой английский статистик Уильям Госсет. Он работал в пивной компании, занимался контролем качества пива и, возможно, по соображениям коммерческой тайны публиковал свои результаты под псевдонимом «студент». Так и вошел в историю.

Значение t представляет собой разность двух средних арифметических, определенных на малых выборках (а как контролировать пиво большими выборками?). Теоретически $t = \Delta M/\sqrt{x^2} : v$. Распределение Стьюдента при увеличении объема выборки более 20 практически сходится к нормальному распределению. В табл. 3 приведен фрагмент большой таблицы квантилей t-критерия.

Таблица 3. Фрагмент таблицы квантилей t-критерия Стьюдента

v\1-α	0,900	0,950	0,975	0,990	0,995
2	2 35	3,18	4,54	5,84	7,45
5	2,02	2,57	3,36	4,03	4,78
7	1,89	2,06	3,00	3,45	4,03
10	1,84	2,13	2,76	3,17	3,58
20	1,72	2,19	2,53	2,84	3,15
30	1,70	2,04	2,46	2,75	3,03

На практике эмпирический квантиль *t*-распределения рассчитывается по формуле:

$$t_э = \frac{|M_1 - M_2|}{\sqrt{\frac{D_1}{n_1} + \frac{D_2}{n_2}}},$$

где *M,D* и *n* — это средние арифметические значения, дисперсии и объемы 1-й и 2-й выборок. Число степеней свободы определяется суммой объемов без двух единиц: $v = n_1 + n_2 - 2$; причем это «–2» не всегда имеет значение. Гораздо важней независимость сравниваемых выборок (потому что для зависимых линейно выборок в подкоренное выражение пришлось бы добавлять среднюю ковариацию, деленную на среднее из объемов выборок).

Рассмотрим пример. Пусть в двух группах по 16 студентов получены средние баллы за сессию: $M_1 = 3{,}5$ и $M_2 = 4{,}5$ при $D_1 = D_2 = 0{,}8$. Спрашивается, значимы ли различия средних? Вычисляем:

$$t_э = \frac{|4{,}5 - 3{,}5|}{\sqrt{\frac{0{,}8}{16} \times 2}} \cong 3{,}2, \quad v = 16 \times 2 - 2 = 30.$$

В табл. 14 находим, что при наибольшей доверительной вероятности 0,995 при $v = 30$ теоретический квантиль *t*-распределения равен 3,03; это меньше эмпирического значения 3,2; следовательно, с большой значимостью отклоняется гипотеза о статистическом сходстве оценок за сессию у двух групп студентов и принимается гипотеза о различии этих оценок.

Критерий Снедекора-Фишера представляет собой отношение двух дисперсий — большей к меньшей (некоторые об этом забывают). Есть сведения, что F-критерий придумал ученик Рональда Фишера — Джордж Снедекор и обозначил первой буквой фамилии учителя. Так это или нет, но в ряде руководств двойное название критерия указано.

Теоретически *F* — это отношение двух $\boldsymbol{x}^2$, деленных на свои степени свободы. Практически *F*-критерий используется в двух амплуа: он парный для дисперсий и множественный для выборочных средних арифметических. В табл. 4 приведён фрагмент таблицы квантилей *F*-критерия. Заметим, что используются два показателя степеней свободы: v_1 — для числителя и v_2 — для знаменателя. В ряде случаев оказывается, что нужно брать одинаковое число степеней свободы и для числителя, и для знаменателя (обоснование дадим в другой главе).

Таблица 4. Фрагмент таблицы квантилей *F*-критерия

v_2 \ v_1	2	5	10	1-α
2	9,00	9,24	3,39	0,900
	19,00	19,30	19,40	0,950
	99,00	99,30	99,40	0,990
5	3,78	3,45	3,30	0,900
	5,79	5,05	4,74	0,950
	13,27	10,97	10,05	0,990
10	2,92	2,52	2,32	0,900
	4,10	3,33	2,98	0,950
	7,56	5,64	4,85	0,990

Рассмотрим пример. Выше для случайного процесса научения студентов за четыре года были получены «годовые» средние арифметические значения на выборках с n = 5:

1 2 3 4

Mx/t = (3,0 3,6 4,2 4,6).

На первый взгляд эти данные образуют возрастающую прямую, но визуальные впечатления могут быть обманчивы. Поэтому проверим по F-критерию, значимы ли различия этих выборочных средних. Эмпирический квантиль — $F_Э$ вычисляется по формуле:

$$F_Э = \frac{n \cdot D[M_i] + M[D_i]}{M[D_i]},$$

где i = 1, 2, 3, 4; n = 5. Вычисляем по полученным раньше данным:

$D[M_i] = (3^2 + 3,\ 6^2 + 4{,}2^2 + 4,\ 6^2) : 4 - 3{,}85^2 = 0{,}37;$

$$M[D_i] = (0{,}4 + 1{,}04 + 0{,}56 + 0{,}24) : 4 = 0{,}376;$$

$$F_Э = \frac{5 \cdot 0{,}37 + 0{,}376}{0{,}376} = 5{,}94.$$

Число степеней свободы будет одинаковым: 5 и 5.

По табл. 15 при ν_1 / ν_2 = 5/5 находим теоретический квантиль $F_Т$ = 5,05. Это меньше эмпирического значения 5,94, следовательно, с доверительной вероятностью 0,950, т.е. на 5%-ном уровне значимости, гипотезу о сходстве отклоняем и принимаем гипотезу о различии выборочных средних.

Последний из необходимых и достаточных практически статистических критериев — это критерий G — Кохрана.

Этот критерий необходим для сравнения ряда выборочных дисперсий, больше двух. Теоретические квантили этого критерия приведены в табл. 16. Эмпирический квантиль вычисляется просто, это отношение максимальной дисперсии к сумме всех дисперсий:

$$G = D_{\max} : \sum_{i=1}^{m} D_i,$$

где m — число выборок.

Заметим, что $G < 1$, поэтому в табл. 16 приведены лишь значимые цифры, нули опущены.

В качестве числового примера воспользуемся четырьмя дисперсиями, полученными выше и, на первый взгляд, тоже разными: $G_Э$ = 1,04 : (0,4 + 1,04 + 0,56 + 0,24) = 0,553.

В табл. 16 при ν = 6, что близко к n = 5, и m = 4 находим для доверительной вероятности 0,95, что теоретический квантиль равен 0,560 против эмпирического 0,553; ясно, что можем принять гипотезу о сходстве при уровне значимости 5%. Дисперсии признаются равными.

Таблица 16. Фрагмент таблицы квантилей G-критерия Кохрана

" ... m V	3	**4**	**5**	i - a
2	976	906	841	0,95
	993	968	928	0,99
6	677	560	478	0,95
	761	641	553	0,99
12	580	466	392	0,95
	674	527	445	0,99
20	526	416	338	0,95
	577	461	345	0,99
40	469	364	298	0,95
	505	397	327	0,99

Заканчивая, хочу еще раз подчеркнуть, что не следует пользоваться многочисленными непараметрическими критериями, мощность которых мала либо не определена вовсе. Рассмотренный набор из пяти статистических критериев является достаточным для психологической практики.