Блок заданий III. Статистическая проверка гипотез

Результаты эксперимента $x^{(n)} = (x_1, \dots, x_n)$ можно трактовать как независимые реализации случайной величины (вектора) ξ, распределение которой неизвестно. Необходимо проверить ряд гипотез относительно этого распределения.

III.1. Проверить гипотезу об изменении математического ожидания после воздействия по одновыборочному (разностному) критерию Стьюдента

– данные: столбцы A(Z1_A) – измерения до, B(Z1_B) – измерения после воздействия
– условия: уровень значимости α и направление ожиданий исследователя – строка 2.

III.2. Проверить гипотезу о точности измерений двух приборов (о дисперсиях наблюдений) по критерию Фишера

– данные: столбцы C(Z2_A) – 1-й прибор, D(Z2_B) – 2-й прибор
– условия: уровень значимости α и направление ожиданий исследователя – строка 3.

III.3. Проверить гипотезу о вероятности осуществления события A

– данные: столбец E(Z3)
– условия: уровень значимости α и направление ожиданий исследователя – строка 4.

III.4. Проверить гипотезу о различиях в распределении наблюдений в двух группах по критерию Вилкоксона

– данные: столбцы F(Z4_A) – 1-я группа, G(Z4_B) – 2-я группа
– условия: уровень значимости α и направление ожиданий исследователя – строка 5.

III.5. Проверить гипотезу однородности наблюдений в двух группах по критерию хи-квадрат

– данные: столбцы H(Z5_A) – 1-я группа, I(Z5_B) – 2-я группа
– условия: уровень значимости α, начальная точка $X0$ и шаг разбиения Delta – строка 6.

Теоретические основы

HT – Help_Teorija – «Теоретические аспекты заданий курсового проекта по математической статистике» – стр. 38-40, 40-41, 43-46, 46-47, 48-50

HEx – Help_Excel – «Как выполнить курсовой проект по математической статистике в Excel» – стр. 29-32, 33-36, 40-43, 44-46

Предположим, что имеется некий «физический» объект, относительно которого высказано некое предположение (гипотеза). Например, объект – новое лекарство, гипотеза – понижает давление (препятствует возникновению ОРЗ и т.п.), или объект – две специфические группы людей, гипотеза – на выборах в первой группе чаще, чем во второй, голосуют за партию «любителей математической статистики» (или – распределение голосов в этих группах различное).

Для проверки этого предположения (гипотезы) берётся выборка $x^{(n)} = (x_1, \dots, x_n)$ представителей объекта, например, измеряется показатель прочности ряда образцов, производится частичный опрос перед выборами, лекарство скармливается группе подопытных пациентов.

Математическая статистика начинается тогда, когда результаты наблюдений изменяются от представителя к представителю случайным образом. Анализируемая гипотеза может быть выражена в терминах характеристик распределения этой случайной величины, например, математическое ожидание разности измерений давления до лечения и после лечения меньше нуля. Часто распределение наблюдений F_θ зависит от нескольких неизвестных параметров $\theta = (\theta_1, \dots, \theta_k) \in \Theta$ (в ситуациях, не попадающих под такое описание, можно считать, что параметр θ имеет бесконечную размерность и просто совпадает с истинной функцией распределения F).

Высказанной гипотезе относительно объекта соответствует подмножество $\widetilde{\Theta} \subset \Theta$ возможных значений параметра θ. По результатам выборочных данных необходимо сделать вывод о справедливости или несправедливости гипотезы. Здесь можно совершить ошибку двух типов – принять решение в пользу выдвинутой гипотезы, когда на самом деле она ложна, или принять решение, отвергающее гипотезу, когда в действительности она верна. Ввиду случайности наблюдений мы можем контролировать лишь относительные частоты этих ошибок – вероятности ошибок. Более того, в большинстве ситуаций контролю подвластна только вероятность ошибочного отвержения в пользу так называемой нулевой гипотезы, которая чаще всего противоположна ожиданиям исследователя и соответствует ситуации, когда отсутствует эффект воздействия, т.е. воздействие имеет нулевой эффект.

Статистическая задача состоит в том, чтобы по результатам наблюдений $x^{(n)} = (x_1, \dots, x_n)$ принять решение d_0 в пользу нулевой гипотезы $\mathbf{H}_0\colon \theta \in \Theta_0$ или решение d_1 в пользу альтернативной гипотезы $\mathbf{H}_1\colon \theta \in \widetilde{\Theta}$ о значении некоторого параметра θ, характеризующего распределение наблюдаемой случайной величины.

Решение принимается с помощью решающей функции.

Определение. Измеримая функция $\delta\colon x^{(n)} \mapsto \{d_0, d_1\}$ называется решающей функцией (критерием, тестом).

Почти всегда решающая функция может быть описана с помощью некоторой статистики T в виде

$$\delta\left(x^{(n)}\right)=\begin{cases}d_1, \text{если } T \in O_{crit},\\ d_0, \text{если } T \notin O_{crit},\end{cases}$$

где так называемая критическая область O_{crit} соответствует тем значениям тестовой статистики T, при которых нулевая гипотеза отвергается (принимается альтернатива, т.е. предположения исследователя считаются подтверждёнными).

Вид критической области зависит от выбранной статистики и обычно записывается как $T < C$ или $T > C$, или $|T| > C$. Выбор критической константы C зависит от заданного исследователем (заранее до проведения эксперимента) уровня значимости – уровня допустимой вероятности ошибки первого рода.

Определения. *Вероятность ошибки* I-*ого рода* критерия δ есть вероятность отвержения нулевой гипотезы, если в действительности нулевая гипотеза верна:

$$\mathbf{P}_\theta\left\{\delta\left(X^{(n)}\right)=d_1\right\}, \qquad \theta \in \Theta_0.$$

Критерий δ называется *критерием уровня* α, если его вероятность ошибки I-ого рода при всех $\theta \in \Theta_0$ не превосходит заданного уровня значимости:

$$\mathbf{P}_\theta\left\{\delta\left(X^{(n)}\right)=d_1\right\} \leq \alpha, \qquad \forall\, \theta \in \Theta_0.$$

Все известные критерии устроены так, что для выполнения последнего условия при всех $\theta \in \Theta_0$, достаточно, чтобы оно выполнялось только в граничной точке θ_0 (поэтому, кстати, часто нулевую гипотезу описывают в виде простой гипотезы $\mathbf{H}_0\colon \theta=\theta_0$). Если критическая область имеет вид $T > C$, то в этом случае критическая константа C находится из условия

$$\mathbf{P}_{\theta_0}\{T > C\} \leq \alpha \text{ или (лучше) } \mathbf{P}_{\theta_0}\{T > C\} = \alpha, \tag{*}$$

если последнее равенство, в принципе, возможно. В ситуации, когда альтернативная гипотеза соответствует «мечтам» исследователя, уровень значимости представляет собой степень опасений заказчика исследования (спонсора) за свои зазря потраченные средства.

Для нахождения критической константы и построения критерия необходимо знать распределение тестовой статистики. Здесь и кроется причина, по которой контролируют только вероятность ошибки I-ого рода – распределение тестовой статистики (точное или асимптотическое при объёме выборки $n \to \infty$) чаще всего доступно для вычислений только при $\theta=\theta_0$. Таким образом, если F_0 – функция распределения тестовой статистики при $\theta=\theta_0$, то:

для критической области $T > C$ уравнение (*) принимает вид

$$1-F_0(C)=\alpha,$$

т.е критическая константа C совпадает с верхней квантилью порядка α для F_0,

для критической области $T < C$ уравнение (*) принимает вид

$$F_0(C)=\alpha,$$

т.е. критическая константа C совпадает с (нижней) квантилью порядка α для F_0,

для критической области $|T| > C$ уравнение (*) принимает вид

$$1-F_0(C)+F_0(-C)=\alpha,$$

что в случае симметричного распределения F_0 (когда $F_0(-C)=1-F_0(C)$) даёт уравнение

$$2\left(1-F_0(C)\right)=\alpha,$$

т.е критическая константа C совпадает с верхней квантилью порядка $\frac{\alpha}{2}$ для F_0.

Если используемый для обработки пакет программ умеет вычислять квантиль только одного вида, то следует помнить, что верхняя квантиль порядка α совпадает с (нижней) квантилью порядка $1-\alpha$ и наоборот. Кроме того, часто (нижнюю) квантиль называют обратной функцией F_0. И ещё. Эпитеты «верхняя» и «нижняя» относятся к значениям функции распределения (т.е. к оси OY). Если же их перенести на аргументы этой функции (т.е. на ось OX), то употребляют эпитеты «правая» или «правый хвост» (ПХ) и соответственно «левая» или «левый хвост» (ЛХ).

Прежде чем переходить к обработке выборочных данных следует записать вид критической области с найденной (заранее) критической константой $C = C_{crit}$:

$$T < C_{crit} \quad \text{или} \quad T > C_{crit}, \quad \text{или} \quad |T| > C_{crit}.$$

Теперь можно обратиться к выборочным экспериментальным данным $x^{(n)}$ и вычислить значение тестовой статистики $t_{\text{эксп}} = T\left(x^{(n)}\right)$.

Если $t_{\text{эксп}}$ попадает в критическую область, то нулевая гипотеза отвергается и делается вывод о справедливости притязаний исследователя. При этом употребляют обороты типа «значимо» и «статистически значимо», например: уменьшение давления после лечения статистически значимо на 5%-ом уровне, данные частичного опроса значимо свидетельствуют в пользу предположения о различиях в распределении голосов в изучаемых группах и т.п.

Если $t_{\text{эксп}}$ не попадает в критическую область, то нулевая гипотеза не отвергается. В ситуации, когда значение $t_{\text{эксп}}$ свидетельствует в пользу предположений альтернативы (например, разность выборочных средних, как и ожидалось, больше нуля), но это свидетельство слишком мало, чтобы отвергнуть нулевую гипотезу, говорят, что отклонение от нулевой гипотезы «не значимо».

Чуть более информативный способ проверки гипотез предполагает вычисление (вместо критической константы) так называемого критического уровня значимости (ещё его называют достигнутым, фактическим, наблюдённым и много ещё разно как; более современное название p-value – p-значение). Критический уровень значимости (p-val) связан с применяемым критерием и равен минимальному возможному уровню значимости, при котором этот критерий будет отвергать нулевую гипотезу по имеющимся данным (т.е. при полученном значении $t_{\text{эксп}}$).

Способ вычисления p-значения зависит от типа критической области и функции распределения F_0 тестовой статистики T:

для критической области $T > C$

$$p\text{-val} = 1 - F_0(t_{\text{эксп}}),$$

т.е p-val совпадает с функцией надёжности распределения F_0 (правый хвост F_0 – ПХ),

для критической области $T < C$

$$p\text{-val} = F_0(t_{\text{эксп}}),$$

т.е p-val совпадает с функцией распределения F_0 (левый хвост F_0 – ЛХ),

для критической области $|T| > C$

$$p\text{-val} = 1 - F_0(|t_{\text{эксп}}|) + F_0(-|t_{\text{эксп}}|) = 2(1 - F_0(|t_{\text{эксп}}|)),$$

где второе равенство справедливо для симметричных распределений F_0 (в Excel называется два хвоста – 2Х).

В соответствии с определением p-val, нулевая гипотеза отвергается, если заданный уровень значимости $\alpha > p$-val (т.е. при малых значениях p-val). Вывод о справедливости той или иной гипотезы оформляется с указанием полученного значения p-val. Например: «предположение об эффективности лекарства отвергается при уровне значимости $p = 0{,}47$» или «различие между двумя группами статистически значимо ($p < 0{,}001$)». Сокращённая запись вида $p < 0{,}01, p < 0{,}001, p < 0{,}00001$ используется в ситуациях, когда p-val слишком мало, чтобы указывать его точное значение, например $p = 1{,}23456e - 35$ хуже, чем $p < 0{,}00001$.

Замечание. Выводы на основе критической области и на основе p-значения **всегда совпадают**. Если произошло невозможное, следует искать ошибку в вычислениях. Один из вариантов возникновения ошибки – неправильно понятые возможности встроенных функций применяемого пакета. Например, иногда вместо функции распределения вычисляется функция надёжности (правый хвост). Наиболее простой способ избежать ошибки – обратиться к Help'у используемого пакета.

Описание статистических критериев

III.1. (**НТ** стр. 38-40) Одновыборочный (разностный) критерий Стьюдента

применяется в ситуациях, когда требуется проанализировать эффект воздействия способа обработки, например лекарства. Для этого воздействие применяется к ряду объектов (субъектов) и измерения соответствующей характеристики производятся как до, так и после исследуемого воздействия. Например: показатель прочности металла измеряется до и после обработки, среднее давление пациентов (за несколько замеров в течение дня) измеряется до и после приёма лекарства и т.п. Иногда, исходя из природы измерений (– как для показателя прочности) или из теоретико-вероятностных соображений (– по центральной предельной теореме для усреднённых измерений давления), можно предположить, что пара измерений (ξ, η) до и после воздействия есть реализация двумерного нормального случайного вектора со средним μ_1 до проведения воздействия и μ_2 – после воздействия. Таким образом, изменение $\xi - \eta$ имеет нормальное распределение со средним $\mu_1 - \mu_2$; ожидания исследователя формализуются в виде утверждения относительно разности $\mu_1 - \mu_2$ (< 0, > 0 или $\neq 0$). Пусть, для примера, исследователь ожидает, что после воздействия характеристика увеличится, т.е. $\mu_1 - \mu_2 < 0$.

Статистическая задача: по наблюдениям $(x_1, y_1), \ldots, (x_n, y_n)$, представляющим собой реализации двумерного нормального вектора с вектором средних (μ_1, μ_2), проверить нулевую гипотезу $H_0\colon \mu_1 - \mu_2 \geq 0$.

Для вычисления статистики Стьюдента необходимо найти среднее арифметическое $\overline{u}$ и дисперсию (смещённую) S_u^2 разностей $u_i = x_i - y_i, i = 1, \ldots, n$. Статистика Стьюдента равна

$$T = \frac{\overline{u}}{\sqrt{S_u^2}}\sqrt{n-1}.$$

Легко понять, что функция распределения этой статистики монотонно убывает при возрастании параметра $\delta = \mu_1 - \mu_2$:

$$\mathbf{P}_{\delta''}\{T < t\} \;=\; \mathbf{P}_{\delta'}\left\{T + \frac{\delta'' - \delta'}{\sqrt{S_u^2}}\sqrt{n-1} < t\right\} \;\leq\; \mathbf{P}_{\delta'}\{T < t\}$$

при $\delta' < \delta''$. Поэтому при нахождении критической константы (и при вычислении p-значения) достаточно считать, что $\delta = 0$ (т.е. рассматривать только пограничную ситуацию между нулевой гипотезой и альтернативой).

При $\mu_1 - \mu_2 = 0$ статистика Стьюдента имеет распределение Стьюдента $\mathbb{St}_{(n-1)}$ с $(n-1)$ степенями свободы. Функция распределения Стьюдента симметрична около нуля, т.е.

$$\mathbb{St}_{(n-1)}(-C) = 1 - \mathbb{St}_{(n-1)}(C).$$

Поэтому при $\alpha < \frac{1}{2}$ верхняя α-квантиль этого распределения всегда положительна, а (нижняя) α-квантиль отличается от верхней только знаком: $t_{(\alpha)} = -t^{(\alpha)}$.

Замечание. Критерий называется одновыборочным, поскольку его можно применять при сравнении распределения наблюдений $(x_1, \ldots, x_n)$ в одной выборке с некоторой нормой μ_0. В этом случае статистика Стьюдента вычисляется на основе разностей $u_i = x_i - \mu_0$.

III.2. (**НТ** стр. 46-47) Критерий Фишера

предназначен для сравнения дисперсий двух совокупностей (не совокупностей данных, а дисперсий генеральных совокупностей, из которых эти данные извлечены). Напомним, что дисперсия есть показатель разброса (изменчивости, вариабельности) случайной величины около её среднего значения. Поэтому наиболее естественно этот критерий применять в случае сравнения точности измерений двух приборов, не имеющих систематической ошибки. Часто в таких случаях можно считать, что измерения есть реализации нормальной (или приближённо нормальной) случайной величины с дисперсиями σ_1^2 и σ_2^2.

Для вычисления статистики Фишера проводятся n_1 измерений $\left(x_1, \dots, x_{n_1}\right)$ с помощью 1-го прибора (в 1-й группе) и n_2 измерений $\left(y_1, \dots, y_{n_2}\right)$ с помощью 2-го прибора (во 2-й группе). Статистика Фишера равна

$$F = \frac{\tilde{S}_1^2}{\tilde{S}_2^2},$$

где $\tilde{S}_j^2$ – несмещённая оценка дисперсии в j-й группе (j-го прибора), $j = 1{,}2$. (В пособии **НТ**, стр. 47 при определении статистики F прокралась опечатка).

Легко понять, что распределение статистики Фишера монотонно зависит от отношения дисперсий $\delta = \frac{\sigma_1^2}{\sigma_2^2}$:

$$P_{\delta''}\{F < x\} = P_{\delta'}\left\{F\frac{\delta''}{\delta'} < x\right\} \le P_{\delta'}\{F < x\},$$

если $\delta' < \delta''$. Поэтому достаточно знать точное распределение этой статистики только при $\delta = 1$. Известно, что если выборки в двух группах независимы между собой и имеют нормальное распределение с одинаковыми дисперсиями $\sigma_1^2 = \sigma_2^2$ статистика Фишера имеет распределение Фишера $\mathbb{F}_{(n_1-1, n_2-1)}$ с $(n_1 - 1, n_2 - 1)$ степенями свободы.

Замечание (которое лучше не игнорировать). Иногда с вычислительной точки зрения в числителе статистики Фишера удобнее поставить оценку дисперсии той совокупности, которая в соответствии с альтернативой (по предположению исследователя) должна быть больше. В этом случае критическая область будет всегда иметь вид $F > C$ и для нахождения критической константы нужно уметь вычислять только верхнюю квантиль распределения Фишера, а для нахождения p-значения – правый хвост этого распределения.

III.3. (**НТ** стр. 40-41) Проверка гипотезы о вероятности «успеха» (критерий знаков).

В эксперименте наблюдается бинарная случайная величина, принимающая два значения 1 и 0 (A и B). Если значение 1 (A) трактовать как успех воздействия (– лекарство привело к понижению давления у пациента, термическая обработка повысила прочность металлического прутка, вакцина позволила избежать ОРЗ и т.п.), то статистическую задачу можно связать с вероятностью этого успеха. Например: вероятность подхватить ОРЗ в группе принимавших новую вакцину меньше нормы, принятой для данного региона, или – вероятность того, что после приёма лекарства у пациента понижается давление, больше ½ . Классический пример: эксперименты Менделя по скрещиванию двух чистых линий гороха. В соответствии с генной теорией вероятность появления во втором поколении гороха с рецессивным признаком равна ¼ .

Проводится n испытаний и подсчитывается количество T успешных исходов. Если действительная (неизвестная) вероятность успеха равна θ, то статистика T имеет биномиальное распределение с параметрами (n, θ):

$$\mathbf{P}_\theta\{T \le k\} = \mathbb{Bin}(k \mid n, \theta) = \sum_{j=0}^{k} \mathrm{C}_n^j \theta^j (1-\theta)^{n-j}, \ \ k = 0,1, \dots, n.$$

Замечание 1. При построении критической области можно воспользоваться тем, что относительная частота успехов $\frac{T}{n}$ представляет собой оценку неизвестной вероятности θ.

Замечание 2. Для дискретных распределений необходимо следить за типом используемого неравенства ($<$ или $\le$, $>$ или $\ge$) и соотносить его с тем, которое используется в применяемом пакете программ. Например, в Excel функция БИНОМ.РАСП вычисляет вероятность $\mathbf{P}_\theta\{T \le k\}$, поэтому $\mathbf{P}_\theta\{T \ge k\} = 1 -$БИНОМ.РАСП$(k-1; n; \theta; 1)$.

Замечание 3. Критерий называют критерием знаков, т.к. его часто применяют при анализе эффекта воздействия – знаком «+» отмечают наличие эффекта (успех), а знаком «–» отсутствие эффекта. При этом ожидается, что вероятность обнаружения эффекта больше ½, т.е. нулевая гипотеза $H_0\colon \theta = \frac{1}{2}$ при альтернативе $H_1\colon \theta > \frac{1}{2}$.

III.4. (**НТ** стр. 43-45) Проверка гипотезы о распределениях по критерию Вилкоксона.

Некоторая характеристика объекта сравнивается в двух группах. Например: вес арбузов, выращенных с применением и без применения специального удобрения, или время безотказной работы электроламп, изготовленных на разных заводах. Ожидания формулируются в виде соотношений типа «в первой группе больше, чем во второй». Это соотношение нельзя воспринимать буквально, т.к. между парами случайных реализаций двух величин возможны отношения обоих типов – и меньше, и больше. Для формализации сказанного предположим, что случайная величина ξ_1 из 1-ой группы имеет функцию распределения F_1, а из второй сл.в. ξ_2 – ф.р. F_2, тогда говорят, что ξ_1 стохастически меньше ξ_2, если

$$F_1(x) \geq F_2(x), \quad \forall\, x \in \mathcal{R},$$

т.е. вероятность «малых» значений ($< x$) для ξ_1 больше, чем для ξ_2. Обозначим это как $\xi_1 \underset{d}{\prec} \xi_2$, т.е. меньше on distribution.

Статистическая задача: по наблюдениям в двух независимых группах $(x_1, \dots, x_{n_1})$ (с функцией распределения F_1) и $(y_1, \dots, y_{n_2})$ (с функцией распределения F_2) требуется проверить гипотезу $\mathbf{H}_0: F_1 \equiv F_2$ при альтернативе $\mathbf{H}_1: \xi_1 \underset{d}{\prec} \xi_2$ или альтернативе $\mathbf{H}_1: \xi_1 \underset{d}{\succ} \xi_2$.

Критерий Вилкоксона (Wilcoxon) основан на рангах (занятых местах) данных из одной группы в общем наборе данных. Пусть r_j – место (в порядке возрастания), которое занимает значение x_j, из 1-й группы среди общей совокупности данных $(x_1, \dots, x_{n_1}, y_1, \dots, y_{n_2})$. Статистика Вилкоксона

$$W = \sum_{j=1}^{n_1} r_j\,.$$

Интуитивно понятно, что если $\xi_1 \underset{d}{\prec} \xi_2$, то ожидаются малые значения W, т.е. при альтернативе $\mathbf{H}_1: \xi_1 \underset{d}{\prec} \xi_2$ критическая область имеет вид $W < C$. Аналогично, при альтернативе $\mathbf{H}_1: \xi_1 \underset{d}{\succ} \xi_2$ критическая область имеет вид $W > C$.

Если справедлива гипотеза $\mathbf{H}_0: F_1 \equiv F_2$ и функция распределения F_1 всюду непрерывна, то W имеет распределение Вилкоксона с параметрами (n_1, n_2). Это распределение при небольших значениях (n_1, n_2) доступно только в табличном виде. Искусство применения этих таблиц требует некоторого навыка. Ситуацию сильно облегчает тот факт, что уже при небольших значениях (n_1, n_2) это распределение хорошо приближается нормальным законом:

$$\mathbf{P}_{\mathbf{H}_0}\{W < C\} \approx \Phi\left(\frac{C - \mu_W}{\sigma_W}\right),$$

где Φ – функция распределения стандартного нормального $(0,1)$ закона, $\mu_W = \frac{n_1(n_1+n_2+1)}{2}$ (здесь n_1 – объём выборки той группы, для которой подсчитываются ранги), $\sigma_W = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$.

Замечание. Непрерывность F_1 гарантирует отсутствие совпадающих значений в выборках. На практике по разным причинам такие значения всегда присутствуют. В этом случае рекомендуется одинаковым значениям присваивать один и тот же ранг, равный среднему значению занимаемых ими мест. Например, выборке из 7 элементов $(0, 0, 0, 1, 1, 3, 4)$ будет соответствовать вектор рангов (2, 2, 2, 4.5, 4.5, 6, 7). Обратим внимание на то, что в этом случае сумма всех рангов вектора из m элементов, как и положено по формуле арифметической прогрессии, равна $m(m+1)/2$. Пакет Excel последних версий умеет вычислять такие ранги. К сожалению, при этом приходится пересчитывать станд.отклонение σ_W. Мы не будем этого делать из-за двигателя прогресса.

III.5. (**НТ** стр. 48-50) Проверка однородности двух групп (критерий согласия хи-квадрат).

Этот критерий носит название критерия согласия, поскольку он применяется в ситуациях, когда нулевая гипотеза совпадает с ожиданиями исследователя и альтернативное утверждение включает в себя «всё остальное». Например: разработчик языка программирования использует алгоритм генерации псевдослучайных чисел, от которых он ожидает, что они будут распределены по равномерному закону, или – по Менделю доля гороха с рецессивным признаком во втором поколении равна ¼. В подобных ситуациях вопрос ставится о согласии выборочных данных с выдвинутой (нулевой) гипотезой.

Критерий хи-квадрат однородности (т.е. одинаковой распределённости) наблюдений в двух независимых группах основан на таблицах частот, используемых при построении гистограммы. Пусть $I_1, \dots, I_r$ – интервалы разбиения всей числовой прямой: $\mathcal{R} = I_1 + \dots + I_r$, $(x_1, \dots, x_{n_1})$ – выборка из 1-й группы, $(y_1, \dots, y_{n_2})$ – выборка из 2-й группы. Для каждой из этих выборок $(k = 1{,}2)$ подсчитывается количество ν_{jk} попаданий в j-й интервал $(j = 1, \dots, r)$. Пусть $\nu_{j*} = \nu_{j1} + \nu_{j2}$ – общее число наблюдений (в обеих группах), попавших в j-й интервал. Статистика критерия хи-квадрат вычисляется по формуле

$$X^2 = n_1 n_2 \sum_{j=1}^{r} \frac{1}{\nu_{j*}} \left(\frac{\nu_{j1}}{n_1} - \frac{\nu_{j2}}{n_2} \right)^2.$$

Идея построения этой статистики абсолютно прозрачна. Относительная частота ν_{jk}/n_k представляет собой оценку вероятности попадания случайной величины из k-й группы в j-й интервал. Поэтому если распределения сл. в., наблюдаемых в группах, одинаковы, то разность этих частот не должна быть слишком большой. Таким образом, гипотезу однородности следует отвергнуть, если статистика X^2 примет достаточно большое значение: $X^2 > C$.

Распределение статистики X^2 при конечных объёмах выборок n_1, n_2 зависит как от выбранных интервалов, так и от истинного распределения наблюдений. Однако известно, что при справедливости гипотезы однородности $(F_1 = F_2)$ распределение X^2 можно аппроксимировать распределением хи-квадрат с $(r-1)$-й степенью свободы:

$$\mathbf{P}_{F_1=F_2}\{X^2 < x\} \to \mathbb{K}_{r-1}(x) \qquad (n_1, n_2 \to \infty).$$

Следовательно, критическая константа C_{crit} приблизительно равна верхней α-квантили распределения хи-квадрат с $(r-1)$-й степенью свободы, а критический уровень значимости (p-value) p-val $\approx 1 - \mathbb{K}_{r-1}(x^2)$, где x^2 – значение статистики X^2, полученное на основе представленных выборочных данных (здесь x^2, X^2 – просто обозначения; дополнительно возводить в квадрат значения статистики запрещается).

Замечание. Поскольку распределение хи-квадрат используется в основном в контексте критериев согласия, многие разработчики статистических программ под функцией распределения хи-квадрат понимают функцию надёжности (1 минус ф.р.), а под квантилью – верхнюю квантиль. Дабы не запутаться, лучше всегда обращаться к Help'у. Другой способ – воспользоваться свойствами распределений: функция распределения возрастает при возрастании аргумента (x), а верхняя квантиль убывает при возрастании уровня значимости α.

Общая схема построения статистических критериев,
которой следует **обязательно** придерживаться при представлении отчёта.

1. Описать «реальную» исследовательскую задачу и ожидания исследователя.
2. Описать наблюдаемую случайную выборку с указанием объёма выборок.
3. Описать вероятностную модель распределения наблюдений $f(x|\theta), \theta \in \Theta$.
4. Связать параметры вероятностной модели с ожиданиями исследователя – гипотеза $H_1: \theta \in \Theta_1$. Описать нулевую гипотезу $H_0: \theta \in \Theta_0$.
5. Указать выбранный уровень значимости.
6. Указать применяемый критерий и описать соответствующую ему тестовую статистику T. Выбрать вид критической области в зависимости от значений T.
7. Описать распределение тестовой статистики (для всех θ или только в граничной точке θ_0.
8. Выбрать способ вычисления критической константы. Найти её значение по своим входным параметрам (объём выборки, α).
9. Записать окончательный вид критической области с найденной критической константой.
   a. По своим данным найти значение тестовой статистики $T = t_{ex}$.
   b. Сделать вывод о справедливости (или несправедливости) той или иной гипотезы.
10. Вычислить p-value. Сделать вывод о справедливости (или несправедливости) той или иной гипотезы.
11. Представить отчёт о проделанной работе.

В своих отчётах – заполнить самостоятельно применительно к своему заданию и самостоятельно предложить вариант реальной задачи в пункте 1 (интернет, книги и т.д. и т.п.).

Пользуясь моими таблицами отчётов, не оставляйте мои пояснения – будет считаться грубой ошибкой.

Пример отчёта.

Имярек Джон Карлович гр. 09-0101 (ZadanMS278)

Задание III.1.

1. В соответствии с молекулярной теорией при термической обработке металла должно происходить изменение его прочности (увеличение или уменьшение).
2. Для проверки предположения были произведены измерения прочности до и после обработки $n = 74$ металлических прутков, изготовленных из одной плавки металла.
3. По соображениям физики процесса можно предположить, что измерения суть реализации нормальной случайной величины $N(\mu, \sigma^2)$, где μ – математическое ожидание показателя прочности в образцах, σ^2 – дисперсия, характеризующая степень изменчивости этого показателя от образца к образцу, а также точность метода измерения. До проведения обработки $\mu = \mu_1$, $\sigma^2 = \sigma_1^2$; после проведения обработки $\mu = \mu_2, \sigma^2 = \sigma_2^2$.
4. Ожидается, что $\theta = \mu_1 - \mu_2 \neq 0$, т.е. в среднем прочность металла изменяется. Нулевая гипотеза $H_0: \theta = 0$ при альтернативе $H_1: \theta \neq 0$.
5. Уровень значимости $\alpha = 0.06$.
6. Ввиду предположения нормальности наблюдений следует применить одновыборочный (разностный) критерий Стьюдента, основанный на разности $\overline{x}_1 - \overline{x}_2$ выборочных средних значений до и после обработки. Для вычисления статистики Стьюдента необходимо найти среднее арифметическое $\overline{u}$ и дисперсию (смещённую) S_u^2 разностей $u_i = x_i - y_i, i = 1, \dots, n$. Статистика Стьюдента равна
$$T = \frac{\overline{u}}{\sqrt{S_u^2}}\sqrt{n-1}.$$
В соответствии с теоретическими предпосылками ожидается, что абсолютная величина $|\overline{x}_1 - \overline{x}_2|$ будет принимать большие положительные значения. Нулевая гипотеза должна отвергаться, когда абсолютное значение статистики Стьюдента $|T| \geq C$.
7. Функция распределение тестовой статистики в граничной точке $\theta_0 = 0$ совпадает с функцией распределения Стьюдента $\mathbb{St}_{(n-1)}$ с $n - 1 = 73$ степенями свободы.
8. Критическая константа C_α находится из уравнения
$$P\{|T| > C\} = 2\big(1 - \mathbb{St}_{73}(C)\big) = 0.06,$$

т.е. – равна верхней 0.03-квантили распределения Стьюдента с 73 степенями свободы. Воспользовавшись таблицей (пакетом Excel, Phyton, Wolfram, Auchan), нашли, что $C_\alpha = 1{,}910$.

9. Окончательный вид критической области $|T| \geq 1{,}910$.

a. По представленным данным найдено

		До	После	По разностям
Объём выборки	n	74	74	74
Среднее	$\overline{x}$	163.66	163.23	–0.429
Станд.отклонение	s	9.516	7.996	6.473
Станд.ошибка среднего	m	1.113	0.936	0.758
Статистика Стьюдента			$T = -0.567 = t_{\text{эксп}}$	
6%-ая критическая область			$\|T\| \geq 1.910$	
Гипотеза отсутствия эффекта			принимается	
с критическим уровнем значимости			p-val = 0.57	
<u>Вывод.</u> Отклонение от нулевой гипотезы статистически не значимо.				

10. Критический уровень значимости p-value вычислялся по формуле
$$\text{p-val} = 2\big(1 - \mathbb{St}_{73}(|T|)\big) = 2\big(1 - \mathbb{St}_{73}(0.567)\big) = 0.57.$$

Т.к. p-val > 0.06, следует считать наблюдения не противоречащими гипотезе отсутствия эффекта обработки.

Задание III.2.

1. Требуется сравнить точность измерений, производимых двумя хорошо откалиброванными (т.е. без систематической ошибки) приборами. Второй прибор был изготовлен по новой технологии, которая, как заявляют изобретатели, повышает точность.
2. Первым прибором было произведено 66 измерений эталонных образцов, вторым прибором – 46 измерений.
3. Можно предположить, что ошибка измерения каждым из приборов носит случайный характер и имеет нормальное распределение со средним ноль и дисперсиями σ_1^2 и σ_2^2.
4. Ожидается, что $\sigma_1^2 > \sigma_2^2$. Т.е. в терминах параметра $\theta = \frac{\sigma_1^2}{\sigma_2^2}$ нулевая гипотеза $\mathbf{H}_0: \theta \leq 1$ при альтернативе $\mathbf{H}_1: \theta > 1$.
5. Уровень значимости $\alpha = 0.05$.
6. В силу нормальности распределения наблюдений, можно применить критерий Фишера. Тестовая статистика Фишера
$$F = \frac{\tilde{S}_1^2}{\tilde{S}_2^2},$$
где $\tilde{S}_j^2$ – несмещённая оценка дисперсии в j-й группе (j-го прибора), $j = 1,2$. Ожидания разработчиков нового прибора будут подтверждены, если F примет достаточно большие значения, т.е. критическая область имеет вид $\{F > C\}$.
7. В граничной точке θ_0=1 распределение статистики Фишера совпадает с распределением Фишера $\mathbb{F}_{(n_1-1, n_2-1)}$ с $n_1 = 65$ и $n_2 = 45$ степенями свободы.
8. Критическая константа C_α находится как решение уравнения
$$P\{F > C_\alpha\} = 1 - \mathbb{F}_{(65,45)}(C) = 0.05,$$
т.е. равна верхней 0.05-квантили распределения Фишера. По таблице распределения Фишера (процедуре Excel, Phyton, Wolfram), находим $C_\alpha = 1{,}594$.
9. Окончательный вид критической области $\{F > 1{,}594\}$.

a. По представленным данным:

	1-й прибор	2-й прибор
n	66	46
$\overline{x}$	158,24	160,93
$\tilde{S}^2$	211,487	100,153
Статистика Фишера $F = s_1^2/s_2^2$		2,112
5%-ая критическая область		$F > 1{,}594$
Гипотеза $\mathbf{H}_0: \sigma_1^2 \leq \sigma_2^2$		отвергается
Вывод: предположение о повышенной точности 2-го прибора статистически подтверждается		
10. p-значение		0,0045

10. p-value вычисляется по формуле
$$\text{p-val} = 1 - \mathbb{F}_{(65,45)}(2{,}112) = 0.0045.$$

Поскольку p-val значительно меньше 5%-го уровня значимости, можно сделать вывод о высокой значимости согласия данных с претензиями изобретателей прибора.

III.3.

1. В регионе ¾ населения ежегодно страдало от ОРЗ. Фармацевтическая компания разработала средство профилактики ОРЗ и обещала, что это средство приведёт к понижению доли заболевших.
2. Для проверки этого заявления предполагается разработанное средство применить к группе $n = 60$ случайно отобранных пациентов.
3. Таким образом, в эксперименте наблюдаются бернуллиевские случайные величины с вероятностью успеха (не подхватить ОРЗ) θ.
4. Ожидается, что $\theta > \frac{1}{4}$. Нулевая гипотеза $\mathbf{H}_0 : \theta \le \frac{1}{4}$ при альтернативе $\mathbf{H}_1 : \theta > \frac{1}{4}$.
5. Уровень значимости $\alpha = 0.01$.
6. Применим критерий знаков, основанный на числе избежавших заболевания T после вакцинации. Ясно, что ожидания фармкомпании будут подтверждены, если T примет достаточно большое значение, т.е. критическая область имеет вид $\{T > C\}$.
7. В граничной точке $\theta_0 = \frac{1}{4}$ функция распределения статистики T есть функция биномиального распределения $\mathbb{Bin}(k \mid n, \theta_0) = \mathbf{P}_{\theta_0}\{T \le k)$ с $n = 60, k = 0,1, \dots, n$.
8. Критическая константа C_α находится как решение неравенства
$$\mathbf{P}\{T > C_\alpha\} = 1 - \mathbb{Bin}(C_\alpha \mid n, \theta_0) \le 0.01,$$
причём, из всех таких констант нужно выбрать наименьшую, т.е. C_α равна квантили порядка 0.99 биномиального распределения. По таблице биномиального распределения (с помощью процедуры Excel, Phyton, Wolfram), находим $C_\alpha = 23$.
9. Окончательный вид критической области: нулевая гипотеза отвергается, если $\{T > 23\}$.

a. По представленным данным:

Частота появления А (не заболевших)	0.55
	27 из 60
1%-ая критическая область	> 23
Гипотеза $\mathbf{H}_0 : p < 0.25$	отвергается
Вывод. Отклонение от нулевой гипотезы статистически значимо. Имеются все основания одобрить применение профилактического средства.	
Критический уровень значимости	$\alpha_{crit} = 0.00022$

10. p-value вычисляется по формуле

$$\text{p-val} = 1 - \mathbb{Bin}(27 | \, 60; \; 0{,}25\,) = 0.000224.$$

Поскольку p-val значительно меньше 1%-го уровня значимости, можно сделать вывод о высокой значимости согласия данных с ожиданиями фармкомпании.

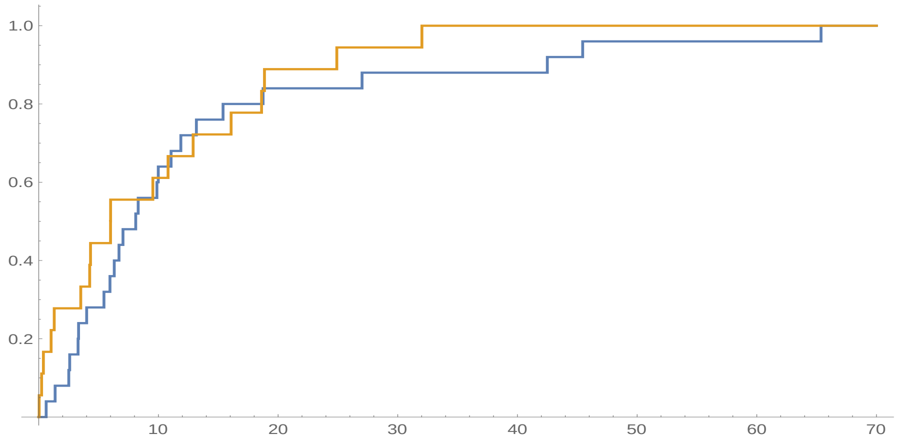
III.4.

1. Для увеличения срока службы электроламп был разработан новый дизайн цоколя.
2. Чтобы проверить действенность этой модификации предлагается провести испытания на долговечность в одинаковых условиях партии $m = 18$ ламп, изготовленных со старым цоколем, и, независимо, партии $n = 25$ ламп с новым цоколем.
3. Время службы каждой лампы есть случайная величина с функцией распределения F_1 (для старых образцов – 1-я выборка) или F_2 (для новых образцов – 2-я выборка).
4. Ожидается, что $F_1(x) > F_2(x)$ для $\forall x > 0$ (т.е. для ламп старого образца более вероятен выход из строя до момента x или, другими словами, $\xi_2 \underset{d}{>} \xi_1$ – время службы новых ламп стохастически больше). Нулевая гипотеза $\mathbf{H}_0 : F_1(x) \equiv F_2(x)$ при альтернативе $\mathbf{H}_1 : F_1(x) > F_2(x), \forall x$.
5. Уровень значимости $\alpha = 0.05$.
6. Применим критерий Вилкоксона, основанный на сумме W рангов 1-й выборки в общем ряду данных. Если справедлива альтернативная гипотеза (наблюдения в 1-й выборке стохастически меньше наблюдений во 2-й), то ожидаются небольшие значения W. Другими словами, критическая область имеет вид $\{W \le C\}$.
7. При справедливости нулевой гипотезы распределение статистики W есть распределение Уилкоксона с параметрами $(18, 25)$. Можно применить нормальную аппроксимацию с матем.ожиданием $\mu_W = 396$ и стандартным отклонением $\sigma_W = 40.62$.
8. Т.о., критическая константа C_α находится как целая часть решения уравнения
$$\mathbf{P}\{W \le C\} \approx \Phi\left(\frac{C - 396}{40.62}\right) = 0.05,$$
т.е. C_α равна квантили порядка 0.05 нормального закона. По таблице (с помощью процедуры Excel, Phyton, Wolfram, пакета Auchan), находим $C_\alpha = 329$.
9. Окончательный вид критической области: нулевая гипотеза отвергается, если $\{W \le 329\}$.

a. По представленным данным:

Объемы выборок		m=18	n=25
Сумма рангов 1-й выборки W		363	
Математическое ожидание μ_W		396	
Стандартное отклонение σ_W		40,6	
6. 5%-я критическая область		$W < 329$	
8. Вывод	Нулевая гипотеза о совпадении распределений	не отвергается	
9.	с критическим уровнем значимости	$\alpha_{crit} = 0.21$	
Заключение. Новый дизайн цоколя лампы не приводит к увеличению срока службы.			

Сравнение эмпирических функций распределения
(синяя линия – новые лампы)



10. p-value вычисляется по формуле

$$\text{p-val} \approx \Phi\left(\frac{363 - 396}{40.62}\right) = 0.21.$$

Поскольку p-val значительно больше 5%-го уровня значимости, то нет оснований считать новый цоколь более надёжным.

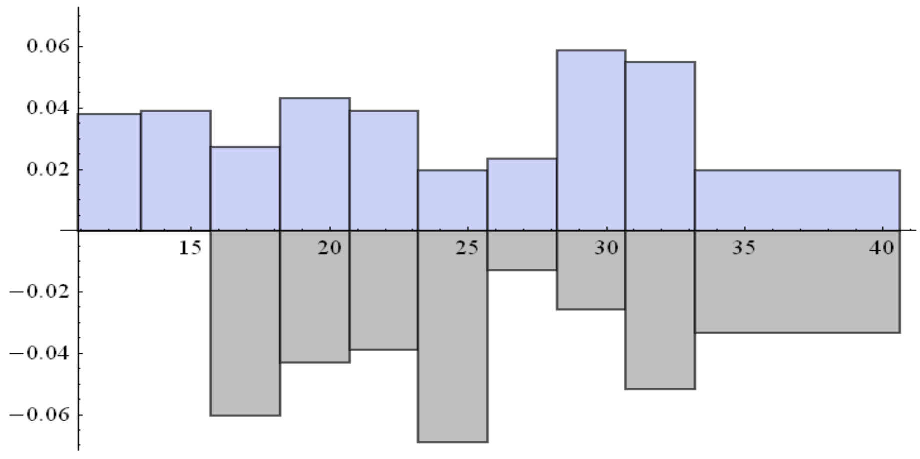
Замечание для особо интересующихся. Полученное в эксперименте значение статистики Вилкоксона не слишком сильно, но всё-таки меньше среднего ожидаемого значения $\mu_W = 396$. Это говорит о том, что, может быть, есть некоторое повышение срока службы, но оно слишком мало, чтобы его можно было обнаружить при столь небольшом объёме выборки. Графики эмпирических функций распределения также говорят в пользу этой версии.

III.5.

1. Производители кока-колы уверяют, что содержание витамина С в банках идентично содержанию этого витамина в стеклянных бутылках.
2. Измерено содержание витаминов в $n_1 = 103$ банках (группа А) и в $n_2 = 93$ бутылках (группа В).
3. Содержание витаминов в продукте есть случайная величина с функцией распределения F_1 (для группы А) или F_2 (для группы В).
4. Ожидания производителей можно формализовать в виде $F_1(x) = F_2(x)$ для $\forall x > 0$ (т.е. содержание витаминов стохастически одинаково). Т.о., нулевая гипотеза $\mathbf{H}_0\colon F_1 \equiv F_2$ – гипотеза однородности совокупностей (без альтернативы).
5. Уровень значимости $\alpha = 0.10$.
6. Применим критерий однородности хи-квадрат, основанный статистике X^2, равной сумме квадратов разностей частот попадания данных в $r = 10$ интервалов группировки. Ожидания компании будут подтверждены, если X^2 примет маленькое значение, т.е. критическая область имеет вид $\{X^2 > C\}$.
7. При справедливости нулевой гипотезы функцию распределения статистики X^2 можно приблизить функцией хи-квадрат распределения $\mathbb{Khi}(x \mid r-1) = \mathbf{P}_{\mathrm{H}_0}\{X^2 < x)$ с $r - 1 = 9$ степенями свободы.
8. Критическая константа C_α находится как решение неравенства
$$\mathbf{P}_{\mathrm{H}_0}\{X^2 > C_\alpha) = 1 - \mathbb{Khi}(C_\alpha \mid 9) = 0.10,$$
т.е. C_α равна квантили порядка 0.975 хи-квадрат распределения с 9 ст.св. По таблице хи-квадрат распределения (процедурой Excel, Phyton, Wolfram), находим $C_\alpha = 14{,}68$.
9. Окончательный вид критерия: гипотеза однородности отвергается, если $\{X^2 > 14{,}68\}$.

a. По представленным данным:

	частоты				
Границы	Группа А		Группа В		χ^2
13.2	9	0.0882	0	0	8.206
15.7	10	0.0980	0	0	9.118
18.2	7	0.0686	14	0.1510	3.031
20.7	11	0.1080	10	0.1080	0.000045
23.2	10	0.0980	9	0.0968	0.00080
25.7	5	0.0490	16	0.1720	6.837
28.2	6	0.0588	3	0.0323	0.744
30.7	15	0.1470	6	0.0645	3.078
33.2	14	0.1370	12	0.1290	0.025
∞	15	0.1370	23	0.2470	2.509
Σ	102	1	93	1	**33.55**
10%-я критическая область					$X^2 > 14{,}68$
Вывод	Гипотеза однородности групп				отвергается
p-value	с критическим уровнем значимости				0.00011
Вывод. Содержание витамина С высоко значимо различаются по способу разлива.					



Сравнение гистограмм распределений (группа А – голубая)

10. p-value вычисляется по формуле $\text{p-val} = 1 - \mathbb{Khi}(33.55 \mid 9) = 0.00011.$
Поскольку p-val значительно меньше 10%-го уровня значимости, есть все основания считать не идентичными способы разлива продукта.