Проверка статистических гипотез

Лекция №5
для студентов 2 курса,
обучающихся по специальности 060609 – Медицинская кибернетика

доц. Шапиро Л.А.

Красноярск, 2015 г.

План лекции:

1. Актуальность темы. Общие понятия в теории проверки гипотез. Ошибки 1 и 2 рода.
2. Подходы к сравнению критериев. Минимаксный подход. Байесовский подход.
3. Построение оптимальных критериев. Критерий отношения правдоподобия
4. Сравнение эмпирических и теоретических распределений.
5. Заключение

Актуальность темы.

Одним из наиболее важных разделов математической статистики является проверка статистических гипотез.

Проверка статистических гипотез

- *Статистической называется гипотеза* о виде неизвестного распределения или о параметрах известных распределений, формулируемая на основе выборки.

Примеры статистических гипотез:

- генеральная совокупность распределена по нормальному закону;
- математические ожидания двух выборок из генеральной совокупности равны.

Гипотезы формулируются только для параметров генеральной совокупности

- Гипотезы, в основе которых нет никаких допущений о конкретном виде закона распределения, называют *непараметрическими*, в противном случае – *параметрическими*.
- ***Гипотезу***, утверждающую, что различие между сравниваемыми характеристиками отсутствует, а наблюдаемые отклонения объясняются лишь случайными колебаниями в выборках, на основании которых производится сравнение, ***называют нулевой*** (основной) гипотезой (H_0).
- Наряду с основной гипотезой рассматривают и ***альтернативную*** (конкурирующую, противоречащую) ей гипотезу H_1. И если нулевая гипотеза будет отвергнута, то будет иметь место альтернативная гипотеза.

- ***Гипотезу называют простой***, если она однозначно характеризует параметр распределения случайной величины. Например, если λ является параметром экспоненциального распределения, то гипотеза H_0 о равенстве λ =10 – простая гипотеза.
- ***Сложной называют гипотезу***, которая состоит из конечного или бесконечного множества простых гипотез. Гипотеза H_0 о том, что математическое ожидание нормального распределения равно двум при неизвестной дисперсии, является сложной.

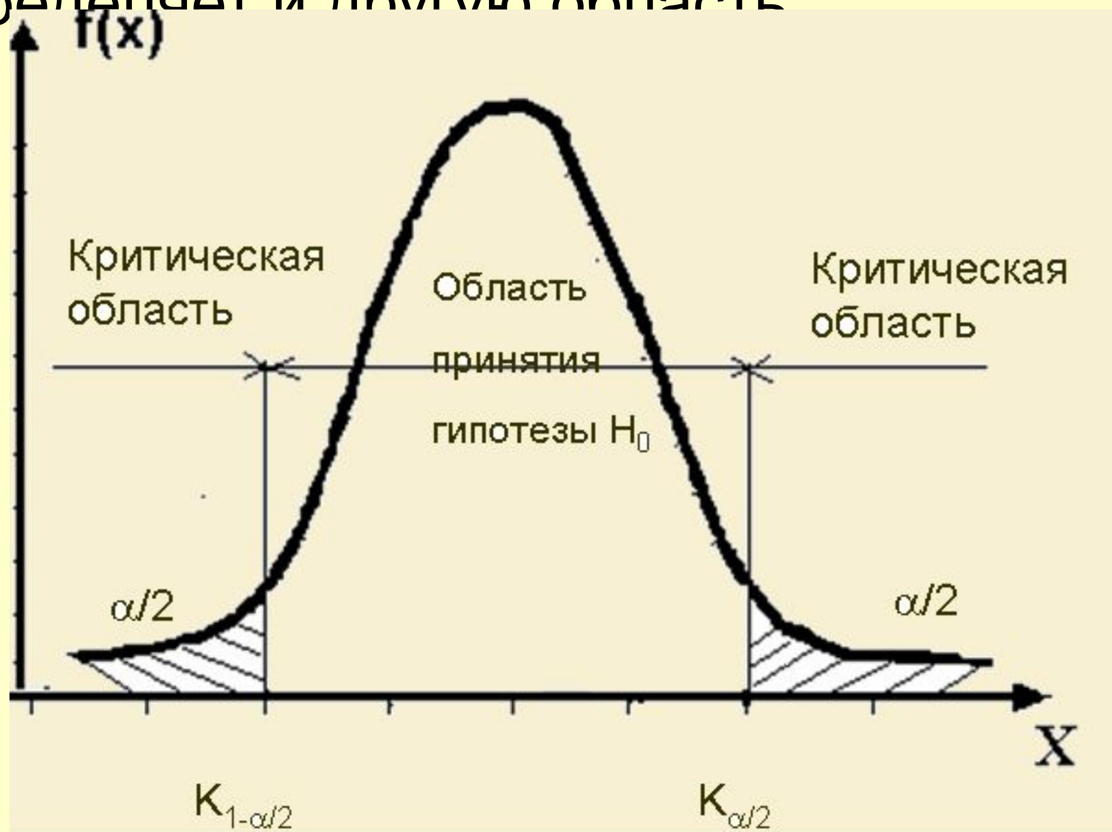
Проверка гипотезы основывается на вычислении некоторой случайной величины – критерия, точное или приближенное распределение которого известно. Обозначим эту величину через K, ее значение является функцией от элементов выборки

$$K=K(x_1, x_2, \ldots, x_n).$$

Процедура проверки гипотезы предписывает каждому значению критерия одно из двух решений – принять или отвергнуть гипотезу. Тем самым все выборочное пространство и соответственно множество значений критерия делятся на два непересекающихся подмножества S_0 и S_1. Если значение критерия K попадает в область S_0, то гипотеза принимается, а если в область S_1, – гипотеза отклоняется.

Двусторонняя критическая область

- Множество S_0 называется *областью принятия гипотезы или областью допустимых значений*, а множество S_1 – *областью отклонения гипотезы или критической областью*. Выбор одной области однозначно определяет и другую область.



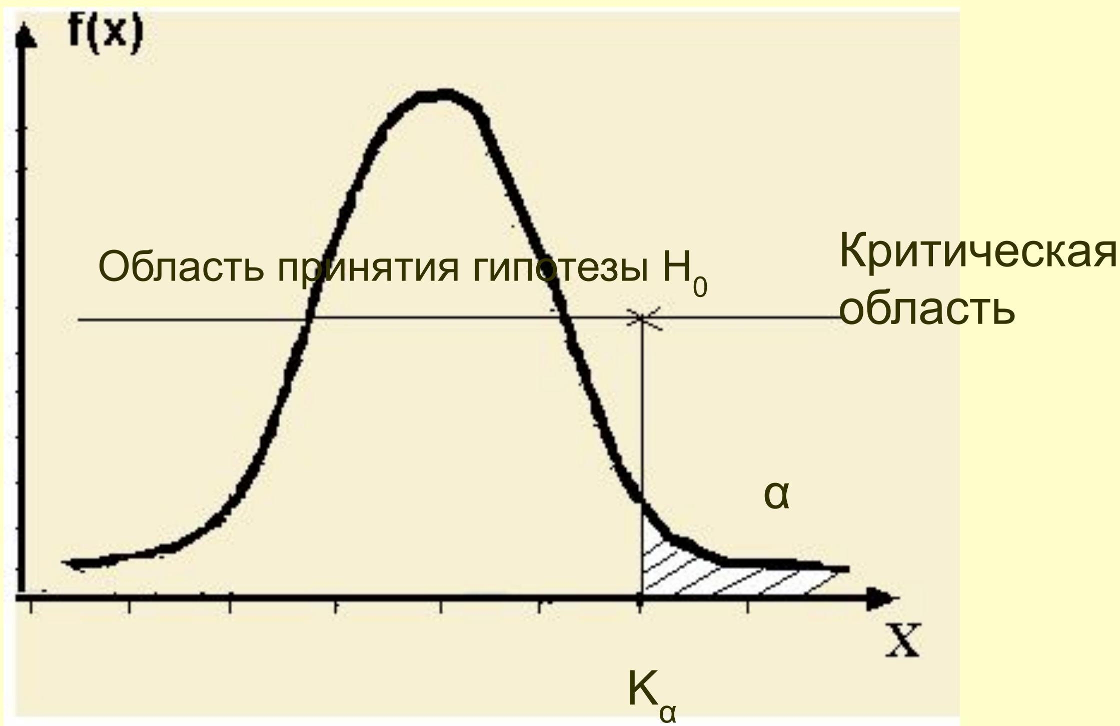
- **Двусторонней** называют критическую область, определяемую неравенством:

$$K<k_1,\ K>k_2, \qquad k_2>k_1$$

- **Правосторонней** называют критическую область, определяемую неравенством $K>k_{кр}$,

$k_{кр}>0$

- **Левосторонней** называют критическую область, определяемую неравенством $K<k_{кр}$, $k_{кр}<0$



- Принятие или отклонение гипотезы H_0 по случайной выборке соответствует истине с некоторой вероятностью и, соответственно, возможны два рода ошибок.
- ***Ошибка первого рода*** возникает с вероятностью α тогда, когда отвергается верная гипотеза H_0 и принимается конкурирующая гипотеза H_1.
- ***Ошибка второго рода*** возникает с вероятностью β в том случае, когда принимается неверная гипотеза H_0, в то время как справедлива конкурирующая гипотеза H_1.

Доверительная вероятность – это вероятность не совершить ошибку первого рода и принять верную гипотезу H_0. Вероятность отвергнуть ложную гипотезу H_0 называется *мощностью критерия*.

Гипотеза H_0	Решение	Вероят-ность	Примечание
Верна	Принимается	$1-\alpha$	Доверительная вероятность
	Отвергается	α	Вероятность ошибки первого рода
Не верна	Принимается	β	Вероятность ошибки второго рода
	Отвергается	$1-\beta$	Мощность критерия

Алгоритм проверки статистических гипотез

- Располагая выборочными данными ($x_1, x_2, \ldots, x_n$) формулируют нулевую гипотезу (H_0) и конкурирующую гипотезу (H_1).
- Задают уровень значимости (ошибка 1 рода).
- Рассматривается выборочная статистика наблюдений (критерий К)
- На основании выборки ($x_1, x_2, \ldots, x_n$) определяют эмпирическое значение критерия ($К_{эмп}$)
- В зависимости от вида альтернативной гипотезы по соответствующей таблице выбирают квантили критерия для двусторонней или односторонней критической области ($К_{кр}$)
- Если значения критерия попадают в критическую область ($К_{эмп} > К_{кр}$), то нулевая гипотеза отвергается и принимается альтернативная гипотеза H_1.
- Если значения критерия $К_{эмп} < К_{кр}$, нулевая гипотеза не отвергается.

Подходы к сравнению критериев

Пример. Имеется выборка объёма $n = 1$ из нормального распределения $N_{a,1}$ и две простые гипотезы $H_1=\{a=0\}$ и $H_2=\{a=1\}$.

Рассмотрим при некотором $b \in R$ следующий критерий:

$$K(X_1) = \begin{cases} H_1, & \text{если } X_1 \le b, \\ H_2, & \text{если } X_1 > b. \end{cases}$$

Изобразим на графике соответствующие гипотезам плотности распределений и вероятности ошибок первого и второго рода критерия K

$$\alpha=P_{H1}(X_1>b), \quad \beta=P_{H2}(X_1 \le b)$$

- С ростом числа b вероятность ошибки первого рода α уменьшается, но вероятность ошибки второго рода β растёт. Общая тенденция: при попытке уменьшить одну из вероятностей ошибок другая, как правило, увеличивается.

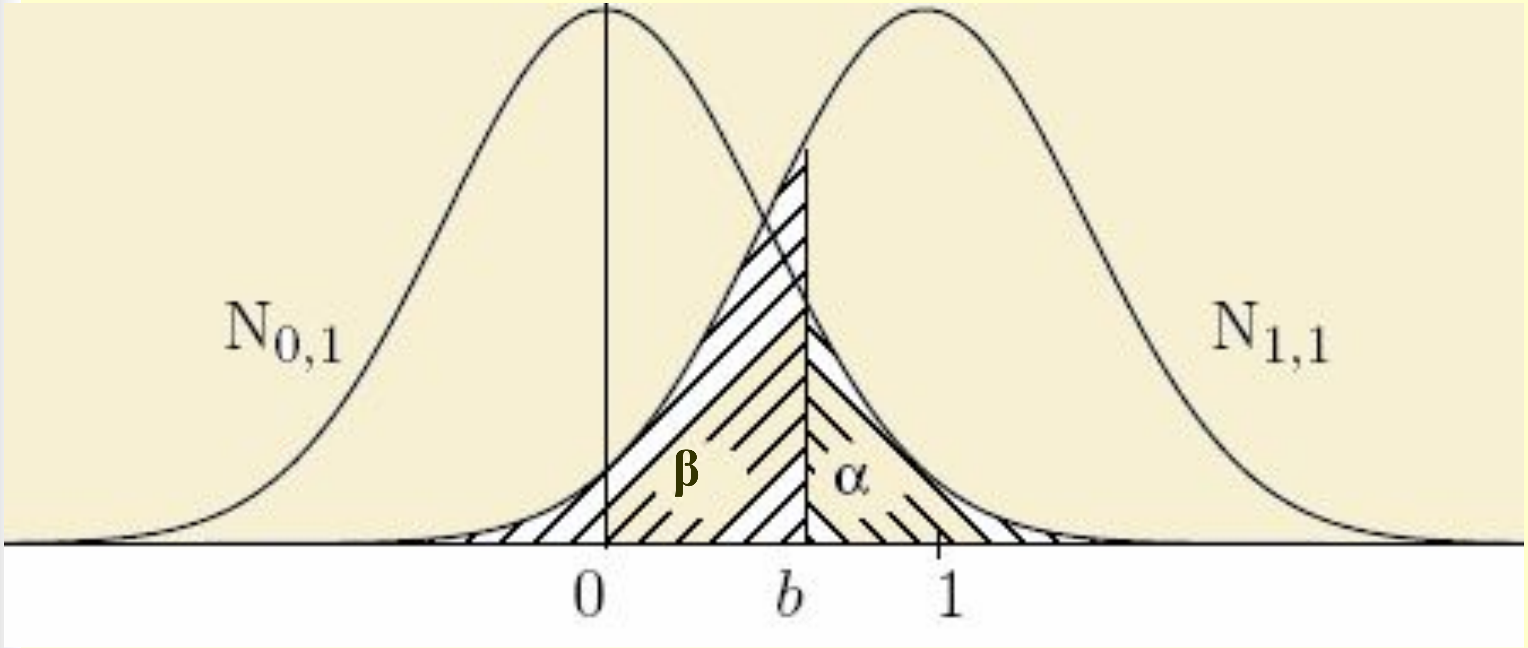


Рис1. Две простые гипотезы.

П р и м е р. Пусть любое изделие некоторого производства оказывается браком с вероятностью р. Контроль продукции допускает ошибки: годное изделие бракует с вероятностью α, а бракованное пропускает (признаёт годным) с вероятностью β.

Если ввести для проверяемого изделия гипотезы H_1 = {изделие годное} и H_2 = {изделие бракованное}, а критерием выбора одной из них считать контроль продукции, то α -вероятность ошибки первого рода этого критерия, а β -второго рода:

$\alpha = P_{H1}(K = H_2)$ = P(контроль забраковал годное изделие);

$\beta = P_{H2}(K = H_1)$ = P(контроль пропустил бракованное изделие);

Упражнение. Вычислить вероятности ошибок первого и второго рода того же критерия, если гипотезы занумеровать иначе:

H_1 = изделие бракованное, H_2 = изделие годное.

Подходы к сравнению критериев

Пусть сравниваем две простые гипотезы. Имеется два критерия с вероятностями ошибок 1 и 2 рода: $\alpha(К_1)$, $\beta(К_1)$ и $\alpha(К_2)$, $\beta(К_2)$.

1. **Минимаксный подход.**

Говорят, что критерий $К_1$ не хуже $К_2$ в смысле минимаксного подхода, если:

$$\max\{\alpha(К_1), \beta(К_1)\} \le \max\{\alpha(К_2), \beta(К_2)\}$$

Минимаксный критерий имеет самую маленькую «наибольшую ошибку» $\max\{\alpha(К_1), \beta(К_1)\}$ среди всех критериев.

2. Байесовский подход.

Этот подход применяют в двух случаях:

а) если известно априори, что с вероятностью r справедлива гипотеза H_1, а с вероятностью $s=1-r$ - гипотеза H_2,

б) если задана «линейная функция потерь»: потери от ошибочного решения равны r если происходит ошибка 1 рода и s - если второго. $r + s$ не обязательно равно 1, но потери можно свести к 1 нормировкой $r'=r/(r+s)$ $s'=s/(r+s)$.

Пусть априорные вероятности или потери заданы. Говорят, что критерий $К_1$ не хуже $К_2$ в смысле байесовского подхода, если:

$$r\alpha(К_1)+s\beta(К_1)\leq r\alpha(К_2)+s\beta(К_2)$$

Байесовский критерий имеет самую маленькую «средневзвешенную ошибку» $r\alpha(К_1)+s\beta(К_1)$ среди всех прочих критериев.

По формуле полной вероятности это есть вероятность ошибки критерия в случае а) или математическое ожидание потерь в случае б).

Выбор наиболее мощного критерия (НМК).

Ошибки 1 и 2 рода не равноправны. Если зафиксировать ошибку 1 рода, то лучшим будет считаться критерий с наименьшей ошибкой 2 рода.

Построение оптимальных критериев.

Оптимальные во всех трех смыслах (минимаксные, байесовские и наиболее мощные) критерии могут быть построены в самом общем случае простым выбором различных констант в одном и том же критерии - **критерии отношения правдоподобия**.

Критерий отношения правдоподобия

Пусть имеется выборка $X=X_1, X_2, \dots, X_n$ (набор независимых, одинаково распределенных величин), и имеются две гипотезы о распределении X_i:

H_1- X_i имеют распределение F_1

H_2- X_i имеют распределение F_2.

Оба распределения либо дискретны, либо непрерывны

$$f_1(X) = f_1(X_1,\dots,X_n) = \prod_{i=1}^{n} f_i(X_i)$$

$$f_2(X) = f_2(X_1,\dots,X_n) = \prod_{i=1}^{n} f_2(X_i)$$

Соответствующие функции правдоподобия (ф. п.=плотность распределения выборки)

Отношением правдоподобия называется частное:

$$T(x) = T(x_1,\ldots,x_n) = \frac{f_2(x_1,\ldots,x_n)}{f_1(x_1,\ldots,x_n)}$$

Если нужно получить критерий с заданной ошибкой 1 рода α или иметь возможность варьировать и размер и мощность критерия, следует рассмотреть класс похожих критериев, введя свободный параметр (1 или какая-то константа С).

Если вторая плотность в С раз превышает первую, принимается вторая гипотеза, в противном случае принимается гипотеза H_1. $K_c(X)$-критерий отношения правдоподобия.

$$K_c(X) = \begin{cases} H_1 \;\; \textit{если}\, T(X) < C \\ H_2 \;\; \textit{если}\, T(X) \geq C \end{cases}$$

Закон распределения-нормальный?

Параметрическая статистика

Да

M±σ, M±m,
M (95% ДИ)

Сравнение 2-х выборок по критерию Стьюдента

Корреляция по Пирсону

Непараметрическая статистика

Нет

Me [25%-75%],
Mo, Min-Max

Сравнение 2-х выборок по критериям Манна-Уитни, Вилкоксона

Корреляция по Спирмену

Сравнение эмпирических и теоретических распределений

- Нулевая гипотеза (H_0)- эмпирическое распределение студентов по росту не отличается от нормального распределения
- Выбираем уровень значимости для проверки гипотезы $\alpha=0,05$ (5% ошибки)
- Гипотезу проверяем по критерию $t_{эмп}$ для коэффициентов асимметрии и эксцесса
- По таблице вероятностей для нормального распределения выбираем квантиль критерия для двусторонней критической области ($t_{кр}=1,96$)
- Если значения критерия попадают в критическую область ($t_{эмп} > 1,96$), то нулевая гипотеза отвергается и принимается альтернативная гипотеза H_1.
- Если значения критерия $t_{эмп} < 1,96$, нулевая гипотеза не отвергается. Распределение студентов по росту не отличается от нормального распределения

Коэффициенты асимметрии и эксцесса:

m_i	$(< x_i > - \bar{x})^3$	$(< x_i > - \bar{x})^3 \cdot m_i$	$(< x_i > - \bar{x})^4$	$(< x_i > - \bar{x})^4 \cdot m_i$
4	-6434,86	-25739,4	119688,3	478753,3
14	-636,06	-8904,78	5470,08	76581,14
20	2,744	13333,9	3,84	76,832
9	1481,5	29401,03	16889,6	152006,4
3	9800,3	555	209727,4	629182,1
Σ=50		**8145,6**		**1336600**

$$A = \frac{8145{,}6}{50 \cdot 9{,}6^3} = 0{,}184 \qquad E = \frac{1336600}{50 \cdot 9{,}6^4} - 3 = 0{,}147$$

Вычислим по выборке коэффициенты асимметрии и эксцесса

$$A = \frac{\sum (x_i - \overline{x})^3 m_i}{n \cdot s^3} = 0{,}184 \qquad E = \frac{\sum (x_i - \overline{x})^4 m_i}{n \cdot s^4} - 3 = 0{,}147$$

Критерий нормированного отклонения t показывает отклонение параметра по выборке от значения в генеральной совокупности, выраженное в единицах среднеквадратической ошибки.

$$t_{эмп} = \frac{|A - m_3|}{s_A} \quad t_{эмп} = \frac{|E - m_4|}{s_E}$$

Так как в генеральной совокупности для нормального распределения коэффициенты асимметрии и эксцесса равны 0, то

$$t_{эмп} = \frac{|A - 0|}{s_A} \quad t_{эмп} = \frac{|E - 0|}{s_E}$$

Ошибки коэффициентов асимметрии и эксцесса (n=50):

$$s_A = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \qquad s_E = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}$$

$$s_A = \sqrt{\frac{6 \cdot 49}{51 \cdot 53}} \approx 0{,}33 \qquad s_E = \sqrt{\frac{24 \cdot 50 \cdot 48 \cdot 47}{49^2 \cdot 53 \cdot 55}} \approx 0{,}62$$

$$t_{эмп} = \frac{0{,}184}{0{,}33} = 0{,}56 \qquad t_{эмп} = \frac{0{,}147}{0{,}62} = 0{,}24$$

Вывод: т.к. 0,56<1,96   и   0,24<1,96

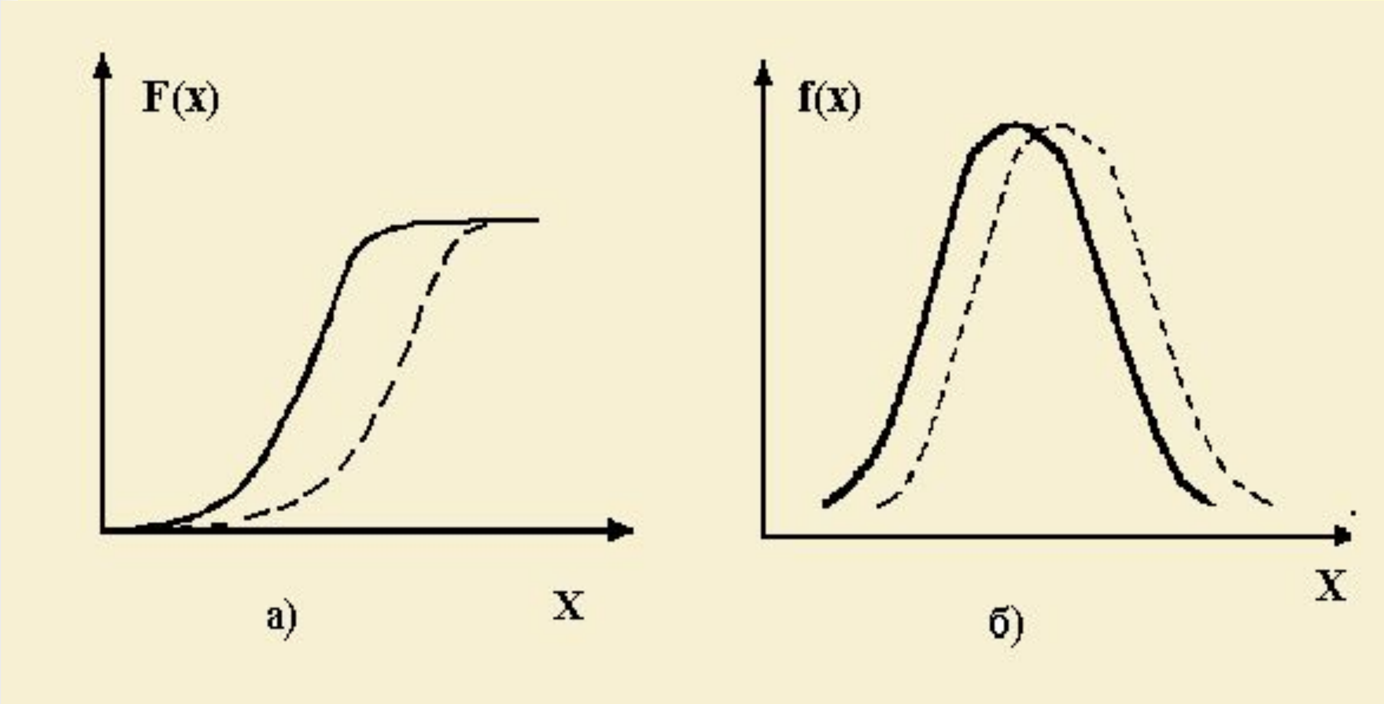
Нулевая гипотеза не отвергается – распределение студентов по росту соответствует нормальному закону с уровнем ошибки 0,05 (5%).

Критерии согласия

Пусть задана некоторая случайная величина, измеряющая отклонение эмпирического распределения от теоретического.

Критерии согласия принимают или отвергают основную гипотезу исходя из величины этой функции отклонения.

Критерии χ^2 Пирсона, Колмогорова, Смирнова, Мизеса, Шапиро-Уилки и т.д.



Сравнение теоретических и экспериментальных распределений по:

а) критерию Колмогорова – Смирнова,

б) критерию Пирсона.

Пунктирная линия – эмпирическое распределение, сплошная – теоретическое распределение.

Проверка гипотезы о нормальном распределении генеральной совокупности по критерию χ^2 .

Для проверки гипотезы будем сравнивать эмпирические m_i (наблюдаемые) и теоретические np_i (вычисленные в предположении нормального распределения) частоты.

В качестве критерия проверки нулевой гипотезы принимаем случайную величину χ^2 :

Затем по таблице критических точек распределения χ^2 (таблица Приложения) по заданному уровню значимости α и числу степеней свободы найдем критическое значение $\chi^2_{кр.}(\alpha, df)$.

Если $\chi^2_{эмп.} \le \chi^2_{кр.}(\alpha, df)$, то считаем, что данный критерий оснований для отклонения гипотезы не дает, а в противном случае считаем, что гипотеза не согласуется с экспериментальными данными и ее надо отвергнуть.

Критерий Пирсона

$$\chi^2_{\text{эмп.}} = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}$$

где $\mathbf{m_i}$ – экспериментальные частоты попадания значения случайной величины в интервал,

$\mathbf{np_i}$ – теоретические частоты.

Вероятность попадания значения случайной величины в интервал от a до b:

$$P(a < x < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$=\Phi(t_2)-\Phi(t_1)$$

причем $\Phi(-t) = 1 - \Phi(t)$

Критерий χ^2

m Эксперимент. частоты	np_i Теоретические частоты	$\frac{(m_i - np_i)^2}{np_i}$
4	3	1/3
14	11	9/11
20	22	4/22
9	10	1/10
3	4	1/4

$\Sigma=1,68$

- **Число степеней свободы** – это общее число величин, по которым вычисляются соответствующие статистические показатели, минус число тех условий, которые связывают эти величины, то есть уменьшают возможности вариации между ними. Число степеней свободы определяется по следующей **формуле:**

  df=k–r–1, где k – число интервалов, r – число параметров предполагаемого распределения. Для нашего случая **r=2**, следовательно, **df=k–3.**
- По заданному уровню значимости **(α)** и числу степеней свободы **df,** находим критическое значение $\chi^2_{кр}$ **(α,df).**

$\chi^2_{кр}$ **(α=0,05,df=2)=5,99**

1,68<5,99

Нулевая гипотеза не отвергается. Распределение студентов по росту подчиняется нормальному закону.

Допущения:

1. Число интервалов =5 и более
2. $m_i \geq n \cdot \alpha$, в противном случае интервалы можно объединить с соседними

Заключение

Нами рассмотрены:

1. Основные положения проверки статистических гипотез.
2. Принципы построения оптимальных критериев.
3. Примеры проверки гипотезы о законе распределения по критерию асимметрии и эксцесса и критерию χ^2.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА:

Основная литература:

- Попов А.М. Теория вероятней и математическая статистика /А.М. Попов, В.Н. Сотников. – М.: ЮРАЙТ, 2011. – 440 с.
- Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие / В.Е. Гмурман. – М. : Высш. шк., 2011. – 479 с.
- Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман. – М. : Высш. шк., 2011. – 404 с.
- Балдин К. В. Основы теории вероятностей и математической статистики : учебник / К. В. Балдин. – М. : Флинта, 2010. – 488с.

Учебно–методические пособия:

- Шапиро Л.А., Шилина Н.Г. Руководство к практическим занятиям по медицинской и биологической статистике Красноярск: ООО «Поликом». – 2003.

БЛАГОДАРЮ ЗА ВНИМАНИЕ