ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

ОСНОВНЫЕ ПОНЯТИЯ

Определение: *Статистической гипотезой* называется любое предположение о виде или параметрах неизвестного закона распределения генеральной совокупности

Определение: Подлежащую проверке гипотезу будем называть основной или *нулевой* и обозначать H_0.

Определение: *Конкурирующей* (альтернативной) гипотезой (H_1) будем называть гипотезу противоположную нулевой.

Определение: Гипотеза называется *простой*, если она однозначно характеризует параметр или свойство генеральной совокупности, например H_0: $\theta = \theta_0$; или H_0: $\xi \in N(m_\xi, \sigma_\xi^2)$- генеральная совокупность распределена по нормальному закону). В противном случае гипотеза называется *сложной*. Например H_0: $\theta_0 \le \theta \le \theta_1$ или H_0: $\xi \notin N(m_\xi, \sigma_\xi^2)$ закон распределения СВ отличен от нормального

Определение: Статистические гипотезы относительно неизвестного истинного значения параметра θ распределения называют *параметрическими гипотезами.*

Определение: Статистические гипотезы о характере распределения генеральной совокупности называются *непараметрическими гипотезами*.

Определение: *Критерием*, или *статистическим критерием*, проверки гипотез называют однозначно определенное правило, по которому по данным апостериорной выборки принимается решение о справедливости H_0 или H_1 гипотез (т.е. проверяемую гипотезу H_0 следует либо принять, либо отклонить)

ВЫБОРОЧНОЕ ПРОСТРАНСТВО

Определение: Выборочное пространство – это множество всех возможных значений вектора $\xi_{1,n}$.

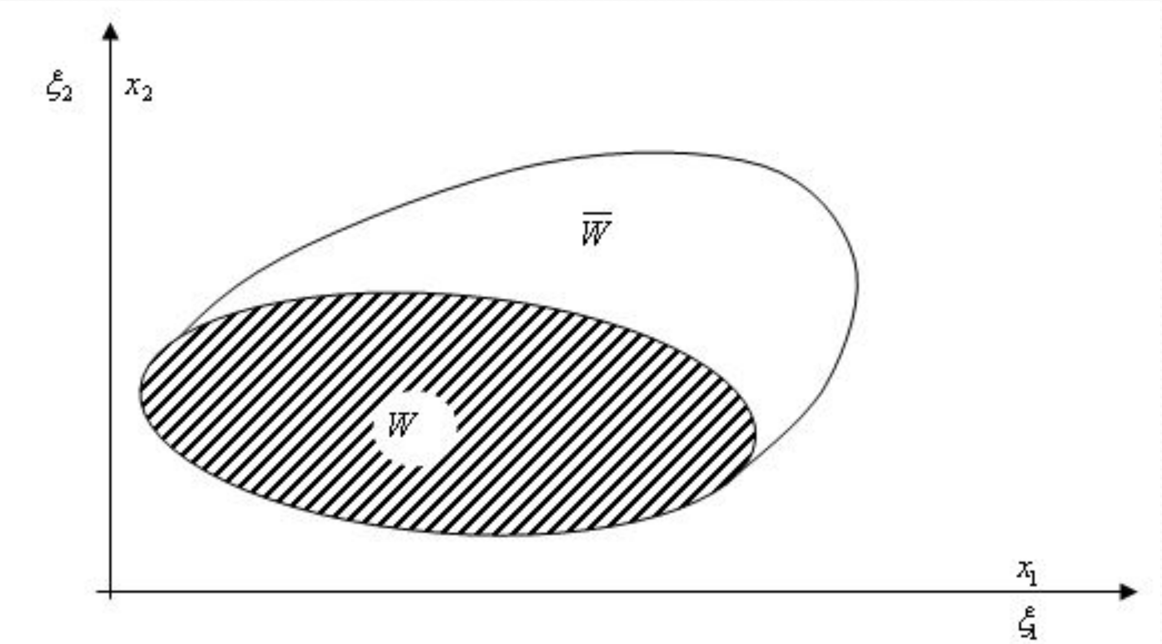
$W \cup \overline{W}$ - выборочное пространство;

W - критическая область (область отклонения гипотезы H_0);

$\overline{W}$ - область принятия гипотезы H_0.

Если выборка $x_{1,n} \in W$, то - гипотеза H_0 отвергается и принимается гипотеза H_1.

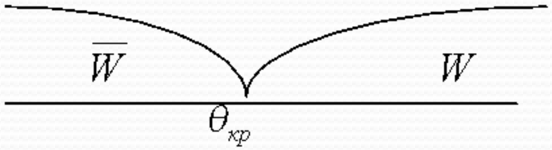
Если выборка $x_{1,n} \in \overline{W}$, то- гипотеза H_1 отвергается и принимается гипотеза H_0.



КРИТИЧЕКИЕ ТОЧКИ

Пусть из генеральной совокупности значений СВ ξ извлечена выборка объемом n и вычислена статистика $\hat{\theta}(\xi_{1,n})$, точное или приближенное распределение которой известно, как правило определяет меру расхождения выборочных данных с высказанной гипотезой H_0.

Затем по этому выборочному распределению определяется критическое значение $\theta_{кр}$, которое все множество возможных значений статистики θ_n^* будем разбивать на два непересекающихся подмножества (области): критическую область W (область отклонения гипотезы) и область принятия решений $\overline{W}$ (гипотезы).



Точки, разделяющие критическую область и область принятия решения, будем называть **критическими.**

Наша задача, научиться находить критические точки.

КРИТИЧЕСКАЯ ОБЛАСТЬ

Зададимся достаточно малой величиной "α", называемой **уровнем значимости** критерия и определим **критическую область** (обозначим W), как множество таких значений $\hat{\theta}(\xi_{1,n})$, вероятность попадания которых в W равнялась бы "α", то есть

$$P(\hat{\theta}(\xi_{1,n}) \in W) = \alpha$$

- если, полученное по данным выборки, значения статистики критерия попадают в критическую область $\hat{\theta}(\xi_{1,n}) \in W$, то гипотезу H_0 отвергают
- если наблюденные значения статистики критерия не попадают в критическую область $\hat{\theta}(\xi_{1,n}) \notin W$, т.е. $\hat{\theta}(\xi_{1,n}) \in \overline{W}$ то гипотезу H_0 не отвергают (принимают).

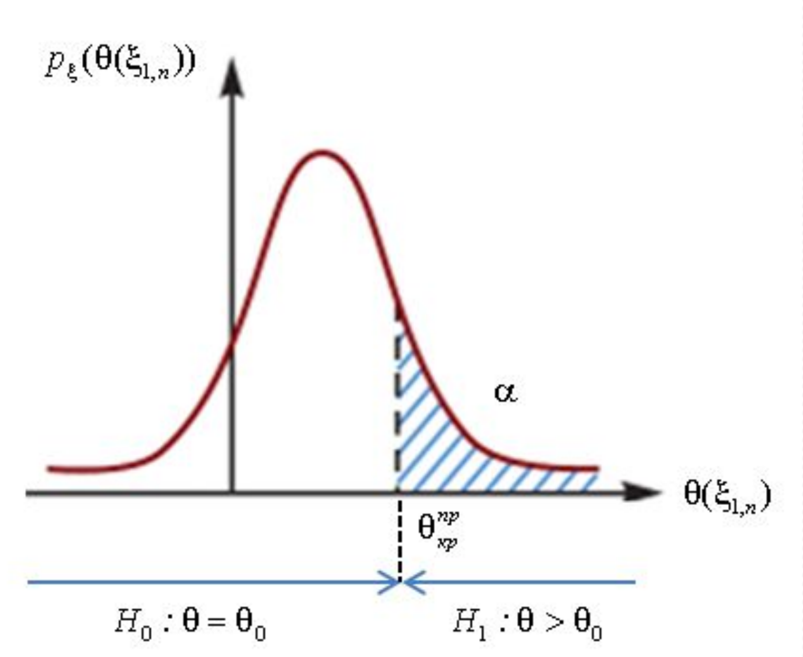
ОШИБКИ ПЕРВОГО И ВТОРОГО РОДА

Реальная ситуация	Статистическое решение	
	H_0 принимается	H_0 отклоняется
H_0 верна	Правильное решение	Ошибка I рода
H_0 не верна	Ошибка II рода	Правильное решение

Определение: **Уровнем значимости** $\alpha = 1 - \gamma$ называют вероятность совершить ошибку первого рода $P(\hat{\theta}(\xi_{1,n} \in W / H_0) = \alpha \to \min$

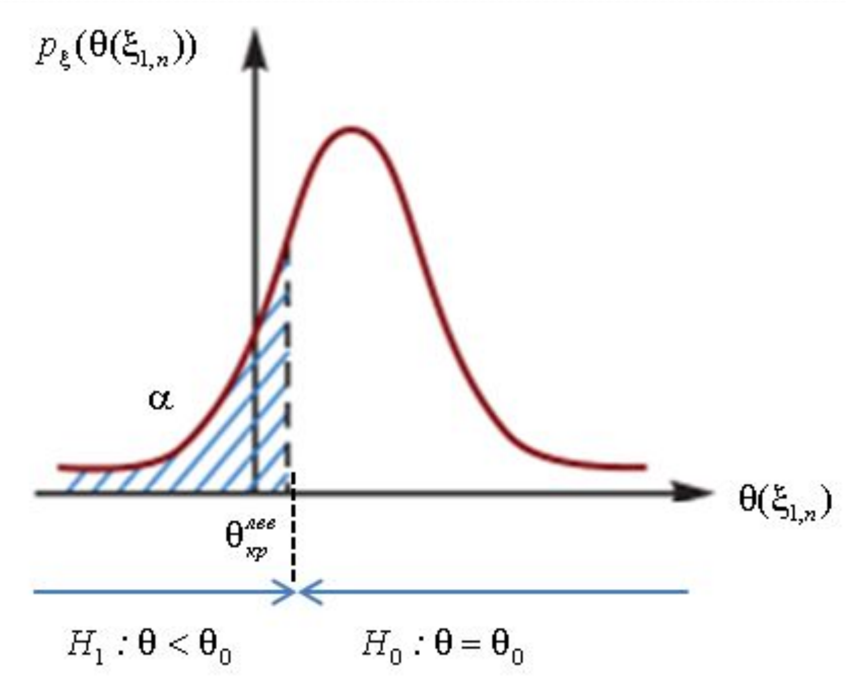
Определение: **Мощностью статистического критерия** называют величину $1-\beta$, равную вероятности отклонить основную гипотезу H_0, когда она неверна (т.е. вероятность не совершить ошибку второго рода) $P(\hat{\theta}(\xi_{1,n} \in W / H_1) = 1 - \beta \to \max$

ПРАВОСТОРОННЯЯ КРИТИЧЕСКАЯ ОБЛАСТЬ



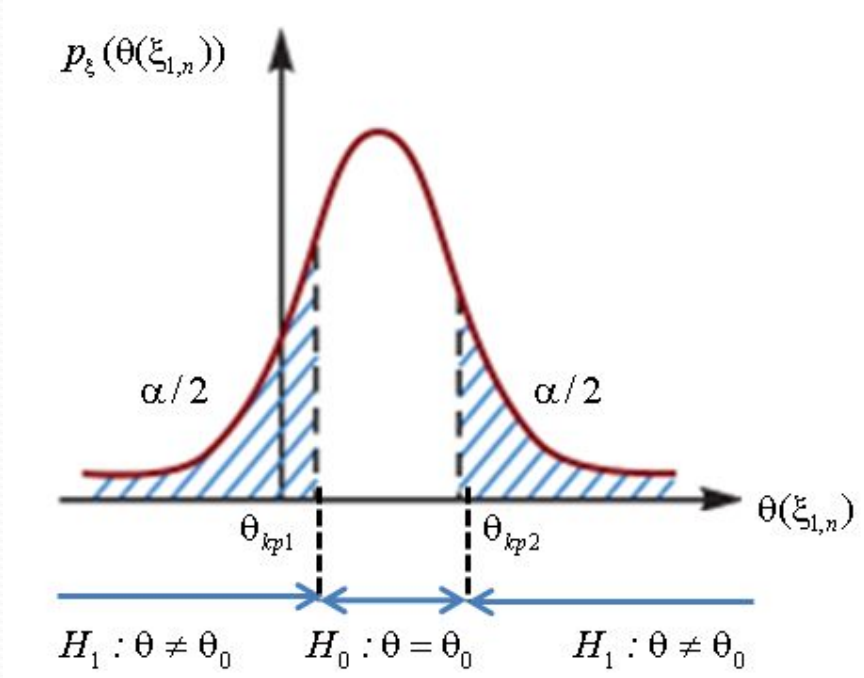
$$P(\hat{\theta}(\xi_{1,n}) > \theta_{кр}^{пр}) = \alpha$$

ЛЕВОСТОРОННЯЯ КРИТИЧЕСКАЯ ОБЛАСТЬ



$$P(\hat{\theta}(\xi_{1,n}) < \theta_{кр}^{лев}) = \alpha$$

ДВУСТОРОННЯЯ КРИТИЧЕСКАЯ ОБЛАСТЬ



$$P(\hat{\theta}(\xi_{1,n}) < \theta_{kp1}) = \frac{\alpha}{2} \qquad P(\hat{\theta}(\xi_{1,n}) > \theta_{kp2}) = \frac{\alpha}{2}$$

ПРОВЕРКА НЕПАРАМЕТРИЧЕСКИХ ГИПОТЕЗ

Постановка задачи

Пусть ξ - СВ (дискретная или непрерывная) с неизвестным законом распределения в форме функции распределения $F_\xi(x)$ или плотности распределения $p_\xi(x)$;

$\xi_{1,n} = (\xi_1, \xi_2, \ldots \xi_n)$ - априорная случайная повторная выборка с взаимно независимыми и одинаково распределенными компонентами;

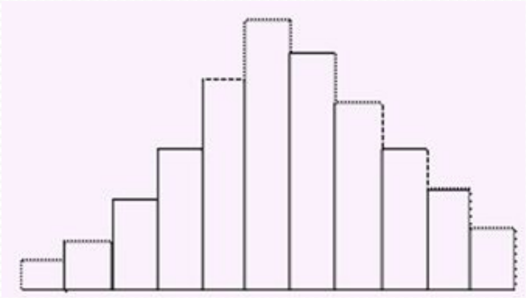
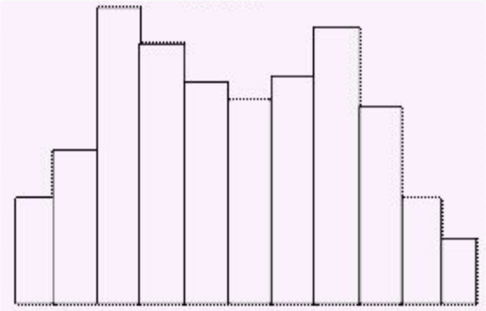
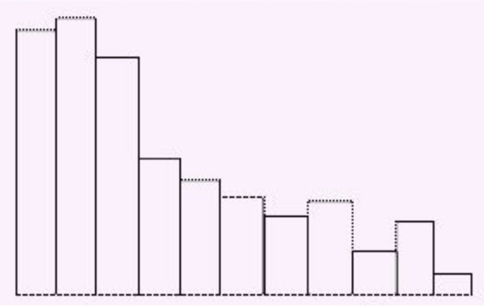
$x_{1,n} = (x_1, x_2, \ldots x_n)$ - реализация случайной выборки $\xi_{1,n}$ на основе, которой получены эмпирические $\hat{F}_n(x)$ и/или $\hat{p}_n(x)$;

$F_0(x)$ - предполагаемая функция распределения генеральной совокупности;

$p_0(x)$ - предполагаемая плотность распределения генеральной совокупности.

Требуется выяснить, согласуется ли гипотеза $H_0 : \xi \in F_0(x)$ или $H_0 : \xi \in p_0(x)$ с полученными экспериментальными данными $x_{1,n}$ или не согласуется.

НЕПАРАМЕТРИЧЕСКИЕ ГИПОТЕЗЫ: ОБЩИЙ СЛУЧАЙ

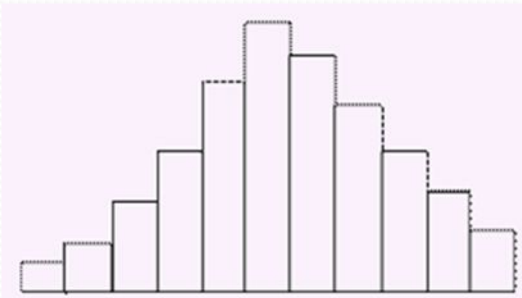
  

На основе гистограммы или полигона выдвигается гипотеза о характере распределения генеральной совокупности:

$$H_0:\ \xi \in F_0(x) \quad \text{или} \quad H_0:\ \xi \in p_0(x)$$

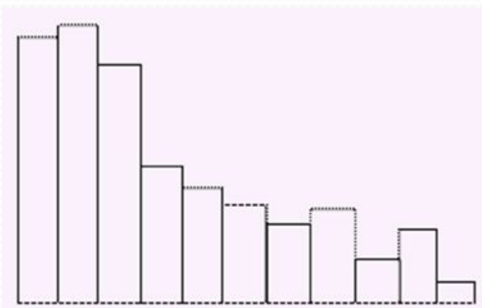
$$H_1:\ \xi \notin F_0(x) \qquad\qquad H_1:\ \xi \notin p_0(x)$$

НЕПАРАМЕТРИЧЕСКИЕ ГИПОТЕЗЫ: ЧАСТНЫЕ СЛУЧАИ



H_0: Распределение СВ не отличается от нормального $H_0: \xi \in N(m, \sigma^2)$

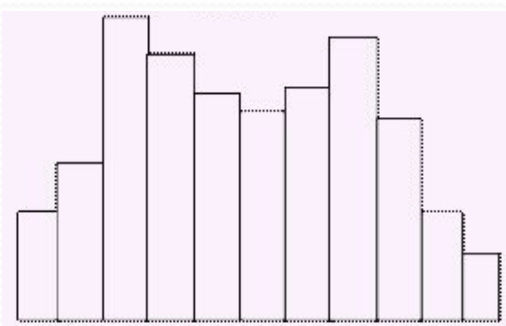
H_1: Распределение СВ отличается от нормального $H_1: \xi \,\overline{\in}\, N(m, \sigma^2)$



H_0: Закон распределения не отличается от экспоненциального $\xi \in F_\xi(x)$ или $\xi \in p_\xi(x)$,

H_1: Закон распределения отличается от экспоненциального $\xi \,\overline{\in}\, F_\xi(x)$ или $\xi \,\overline{\in}\, p_\xi(x)$,

где $F_\xi(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ $\quad p_\xi(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$



H_0: Закон распределения не отличается от равномерного $\xi \in F_\xi(x)$ или $\xi \in p_\xi(x)$,

H_1: Закон распределения отличается от равномерного $\xi \,\overline{\in}\, F_\xi(x)$ или $\xi \,\overline{\in}\, p_\xi(x)$,

где $F_\xi(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a < x \leq b \\ 1, & x > b \end{cases}$ $\quad p_\xi(x) = \begin{cases} \dfrac{1}{b-a}, x \in [a,b] \\ 0, x \notin [a,b] \end{cases}$

КРИТЕРИЙ СОГЛАСИЯ

Для статистической проверки гипотез о теоретическом (модельном) виде закона распределения вероятностей исследуемой случайной величины используют **критерии согласия.**

Критериями согласия называют статистические критерии, предназначенные для обнаружения расхождения между гипотетической статистической моделью (теоретическим распределением) и реальными данными, которые эта модель призвана описать.

Пусть статистика $T(\xi_{1,n})$, являющаяся случайной величиной, характеризующей меру отклонения эмпирического закона распределения от гипотетического (статистический критерий);

t_α - критическое значение, отделяющее область принятия гипотезы H_0 от критической области, и определяемое из условия:

$$P(T(\xi_{1,n}) \geq t_\alpha) = \alpha$$

где α - уровень значимости нулевой гипотезы ($\alpha = 0{,}01; 0{,}05; 0{,}001; 0{,}005$)

Если значение статистики критерия, найденной по результатам проведенных испытаний $T(x_{1,n}) > t_\alpha$, то гипотеза H_0 отвергается;

если значение статистики критерия, найденной по результатам проведенных испытаний $T(x_{1,n}) < t_\alpha$, то гипотеза H_0 принимается.

КРИТЕРИЙ СОГЛАСИЯ ПИРСОНА

Строится статистика: $\chi^2(\xi_{1,n}) = \sum_{i=1}^{k} \frac{(n_i^Э - n_i^T)^2}{n_i^T} = \sum_{i=1}^{k} \frac{(n_i^Э - np_{0i})^2}{np_{0i}} \in \chi^2 \ (k-r-1)$

где $n_i^Э$ – эмпирическая частота - частота варианты x_i дискретного вариационного ряда или число вариант, которые находятся в $i-ом$ интервале интервального вариационного ряда;

k - число возможных вариант дискретного вариационного ряда или число интервалов интервального вариационного ряда;

n_i^T или np_{0i} – теоретическая частота.

p_{0i} - вероятность наблюдения варианты x_i или вероятность попадания исследуемой случайной величины в $i-ый$ интервал интервального вариационного ряда $\sum_{i=1}^{k} p_i = 1$

В случае проверки гипотезы о **нормальном характере распределения** вычисляются по интегральной теореме Лапласа

$$p_{0i} = P(V_{i-1} < \xi < V_i) = \Phi\left(\frac{V_{i-1} - \bar{x}_B}{S}\right) - \Phi\left(\frac{V_i - \bar{x}_B}{S}\right)$$

В случае **равномерного закона распределения**

$p_{0i} = P(V_{i-1} < x < V_i) = \frac{V_i - V_{i-1}}{b-a}$, где $a = V_0$, $b = V_k$, все вероятности должны иметь одинаковые значения

В случае **экспоненциального закона распределения**

$$p_{0i} = P(V_{i-1} < x < V_i) = e^{-\lambda V_{i-1}} - e^{-\lambda V_i}$$

КРИТЕРИЙ СОГЛАСИЯ ПИРСОНА: ПРОДОЛЖЕНИЕ

r - число параметров предполагаемого теоретического закона

В случае нормального закона распределения $r = 2\ (m_\xi, \sigma_\xi^2)$

В случае равномерного закона распределения $r = 2\ (a, b)$

В случае показательного закона распределения $r = 1\ (\lambda)$

Для проверки гипотезы H_0 строим двухстороннюю критическую область, критические точки определяются из уравнений:

$$P(\chi^2 > \chi^2_{kp2}) = \frac{\alpha}{2}, \quad \text{откуда } \chi^2_{kp2} = P_i^{-1}(\frac{\alpha}{2})$$

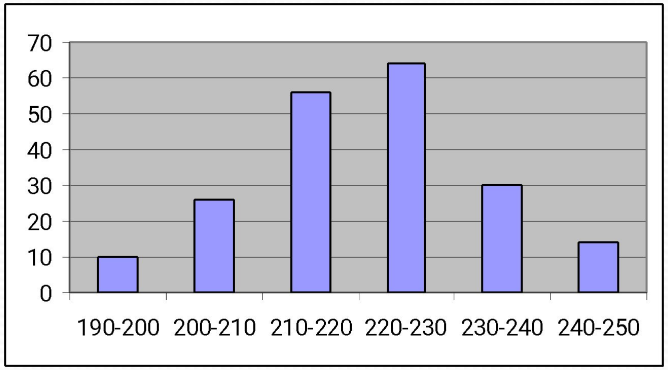
$$P(\chi^2 < \chi^2_{kp1}) = \frac{\alpha}{2}, \text{ откуда } \quad \chi^2_{kp1} = P_i^{-1}(1 - \frac{\alpha}{2})$$

Если $\chi x^2_{kp1}(1 - \frac{\alpha}{2}, k - r - 1) < \chi^2_{набл} < \chi^2_{kp2}(\frac{\alpha}{2}, k - r - 1)$, то гипотеза H_0 принимается и делается вывод в пользу проверяемого закона распределения, в противном случае гипотезу H_0 следует отклонить.

ПРОВЕРКА НЕПАРАМЕТРИЧЕСКОЙ ГИПОТЕЗЫ: ПРИМЕР

Имеются сгруппированные данные о дневной выручке в магазине электротоваров.

Сумма продаж	Число продаж
190-200	10
200-210	26
210-220	56
220-230	64
230-240	30
240-250	14



Требуется проверить гипотезу о характере распределения СВ ξ - сумма продаж $(\alpha = 0{,}05)$

По виду гистограммы можно предположить, что СВ ξ распределена по нормальному закону. Проверим данное предположение.

H_0: Распределение СВ не отличается от нормального

H_1: Распределение СВ отличается от нормального

ПРОВЕРКА НЕПАРАМЕТРИЧЕСКОЙ ГИПОТЕЗЫ: ПРИМЕР
продолжение

Для проверки гипотезы воспользуемся статистикой Пирсона :

$$\chi^2(\xi_{1,n}) = \sum_{i=1}^{k} \frac{(n_i^Э - n_i^T)^2}{n_i^T} = \sum_{i=1}^{k} \frac{(n_i^Э - np_{0i})^2}{np_{0i}}$$

распределена по закону χ^2-распределение с числом степеней свободы $\nu = k - r - 1$

где $p_{0i} = P(V_{i-1} < \xi < V_i) = \Phi\left(\frac{V_{i-1} - \bar{x}_B}{S}\right) - \Phi\left(\frac{V_i - \bar{x}_B}{S}\right)$

Предварительно были определены выборочная средняя и выборочная дисперсия $\bar{x} = 221$, $S^2 = 152$, $S = 12{,}3$

ПРОВЕРКА НЕПАРАМЕТРИЧЕСКОЙ ГИПОТЕЗЫ: ПРИМЕР

продолжение

Для расчета $\chi^2(\xi_{1,n}) = \sum_{i=1}^{k} \frac{(n_i^Э - n_i^T)^2}{n_i^T}$ составим таблицу

интервалы	$n_i^Э$	p_{0i}	$n_i^T = n \cdot p_{0i}$	$\frac{(n_i^Э - n_i^T)^2}{n_i^T}$
190-200	10	0,0376	7,52	0,817872
200-210	26	0,1432	28,64	0,243352
210-220	56	0,2814	56,28	0,001393
220-230	64	0,2992	59,84	0,289198
230-240	30	0,1709	34,18	0,511188
240-250	14	0,0527	10,54	1,135825
итого	200	≈1	≈ 200	$\chi_н^2 = 2{,}999$

$$\chi_н^2 = 2{,}999$$

$$x_{кр1}^2\left(1-\frac{\alpha}{2}, k-r-1\right) = P_i^{-1}\left(1-\frac{\alpha}{2}, k-r-1\right) = P_i^{-1}(0{,}975{,}3) = 0{,}22$$

$$x_{кр2}^2\left(\frac{\alpha}{2}, k-r-1\right) = P_i\left(\frac{\alpha}{2}, k-r-1\right) = P_i(0{,}025{,}3) = 9{,}35$$

$$p_{01} = P(190 < x < 200) = \Phi\left(\frac{200-221}{12.3}\right) - \Phi\left(\frac{190-221}{12.3}\right) =$$
$$= \Phi(-1.71) - \Phi(-2.52) = -0.4565 + 0.4941 = 0.0376$$

$$p_{02} = P(200 < x < 210) = \Phi\left(\frac{210-221}{12.3}\right) - \Phi\left(\frac{200-221}{12.3}\right) =$$
$$= \Phi(-0.89) - \Phi(-1.71) = -0.3133 + 0.4565 = 0.1432$$

$$p_{03} = P(210 < x < 220) = \Phi\left(\frac{220-221}{12.3}\right) - \Phi\left(\frac{210-221}{12.3}\right) =$$
$$= \Phi(-0{,}08) - \Phi(-0{,}89) = -0{,}0319 + 0{,}3133 = 0{,}2814$$

$$p_{04} = P(220 < x < 230) = \Phi\left(\frac{230-221}{12.3}\right) - \Phi\left(\frac{220-221}{12.3}\right) =$$
$$= \Phi(0{,}73) - \Phi(-0{,}08) = 0{,}2673 + 0{,}0319 = 0{,}2992$$

$$p_{05} = P(230 < x < 240) = \Phi\left(\frac{240-221}{12.3}\right) - \Phi\left(\frac{230-221}{12.3}\right) =$$
$$= \Phi(1{,}54) - \Phi(0{,}73) = 0{,}4382 - 0{,}2673 = 0{,}1709$$

$$p_{06} = P(240 < x < 250) = \Phi\left(\frac{250-221}{12.3}\right) - \Phi\left(\frac{240-221}{12.3}\right) =$$
$$= \Phi(2{,}36) - \Phi(1{,}54) = 0{,}4909 - 0{,}4382 = 0{,}0527$$

Т.к. $x_{кр1}^2 < x_{набл}^2 < x_{кр2}^2$, гипотеза о нормальном характере распределения принимается.

ПРОВЕРКА ПАРАМЕТРИЧЕСКИХ ГИПОТЕЗ

ПРОВЕРКА ГИПОТЕЗ О ПАРАМЕТРАХ НОРМАЛЬНО РАСПРЕДЕЛЕННЫХ ГЕНЕРАЛЬНЫХ СОВОКУПНОСТЕЙ

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ: ПОСТАНОВКА ЗАДАЧИ

Пусть $\xi_{1,n}$ - случайная выборка из генеральной совокупности $\xi \in N(m;\sigma^2)$, где математическое ожидание m неизвестно.

Точечной оценкой m является выборочное среднее $\bar{x}(\xi_{1,n}) = \frac{1}{n}\sum_{i=1}^{n}\xi_i$, известно значение выборочного среднего $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$.

Ставится задача на уровне значимости α проверить гипотезу $H_0 : m = m_0$.

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ: СЛУЧАЙ 1

Дисперсия σ^2 известна

Тогда при условии справедливости гипотезы H_0 статистика $u(\xi_{1,n}) = \dfrac{\bar{x}(\xi_{1,n}) - m_0}{\sigma}\sqrt{n} \in N(0;1)$.

Пусть $H_1 : m > m_0$.

Строим **правостороннюю критическую область**: $P\{u(\xi_{1,n}) > u_{\text{кр. прав}}\} = \alpha$, $u_{\text{кр.прав}} = \Phi^{-1}(\frac{1}{2} - \alpha)$ - определяется по таблице Лапласа. Правосторонняя критическая область имеет вид: $(u_{\text{кр. прав}};+\infty)$.

Пусть $H_1 : m < m_0$.

Строим **левостороннюю критическую область**: $P\{u(\xi_{1,n}) < u_{\text{кр.лев}}\} = \alpha$, $u_{\text{кр.лев}} = -\Phi^{-1}(\frac{1}{2} - \alpha)$ - определяется по таблице Лапласа. Левосторонняя критическая область имеет вид: $(-\infty; u_{\text{кр.лев}})$.

Пусть $H_1 : m \neq m_0$.

Строим двустороннюю критическую область:

$P\{u(\xi_{1,n}) < u_{\text{кр}1}\} = \alpha/2$, $P\{u(\xi_{1,n}) > u_{\text{кр}2}\} = \alpha/2$;

$u_{\text{кр}1} = -\Phi(\frac{1}{2} - \frac{\alpha}{2})$, $u_{\text{кр}2} = \Phi(\frac{1}{2} - \frac{\alpha}{2})$ - определяются по таблице Лапласа.

Двусторонняя критическая область имеет вид: $(-\infty; u_{\text{кр}1}) \cup (u_{\text{кр}2};+\infty)$.

Если $u_{\text{набл}} = \dfrac{\bar{x} - m_0}{\sigma}\sqrt{n}$ принадлежит критической области, то гипотеза H_0 отвергается с вероятностью совершить ошибку первого рода α.

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ: СЛУЧАЙ 2

Дисперсия σ^2 неизвестна

Тогда при условии справедливости гипотезы H_0 статистика $t(\xi_{1,n}) = \dfrac{\overline{x}(\xi_{1,n}) - m_0}{S(\xi_{1,n})}\sqrt{n-1}$, где $S(\xi_{1,n}) = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{x}(\xi_{1,n}))^2}$ имеет распределение Стьюдента с $(n-1)$ степенями свободы.

Пусть $H_1 : m > m_0$. Строим **правостороннюю критическую область**:
$P\{t(\xi_{1,n}) > t_{\text{кр.прав}}\} = \alpha$; $t_{\text{кр.прав}} = St^{-1}(2\alpha, n-1)$ - определяется по таблице Стьюдента с числом степеней свободы $\nu = n-1$.
Критическая область имеет вид: $(t_{\text{кр.прав}}; +\infty)$.

Пусть $H_1 : m < m_0$. Строим **левостороннюю критическую область**:
$P\{t(\xi_{1,n}) < t_{\text{кр.лев}}\} = \alpha$; $t_{\text{кр.лев}} = -St^{-1}(2\alpha, n-1)$ - определяется по таблице Стьюдента с числом степеней свободы $\nu = n-1$.
Левосторонняя критическая область имеет вид: $(-\infty; t_{\text{кр.лев}})$.

Пусть $H_1 : m \neq m_0$. Строим **двустороннюю критическую область**:
$P\{t(\xi_{1,n}) < t_{\text{кр}1}\} = \alpha/2$, $P\{t(\xi_{1,n}) > t_{\text{кр}2}\} = \alpha/2$;
$t_{\text{кр}1} = -St^{-1}(\alpha, n-1)$, $t_{\text{кр}2} = St^{-1}(\alpha, n-1)$ - определяются по таблице Стьюдента с числом степеней свободы $\nu = n-1$.
Двусторонняя критическая область имеет вид: $(-\infty; t_{\text{кр}1}) \cup (t_{\text{кр}2}; +\infty)$.

Если $t_{\text{набл}} = \dfrac{\overline{x} - m_0}{S}\sqrt{n-1}$, где $S = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$, принадлежит критической области, то гипотеза H_0 отвергается с вероятностью совершить ошибку первого рода α.

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ: ПРИМЕР 1

Автоматическая линия фасует пакеты с мукой весом в 1 кг. (1000 грамм). Производитель утверждает, что точность наладки линии достаточно высока и средний вес наполненных пакетов в точности равен 1 кг. Покупатель же партии фасованной муки сомневается в точности веса и выдвигает предположение, что вес пакета меньше одного кг. Проведена случайная выборка 100 пакетов с мукой. Повторное их взвешивание показало, что средний выборочный вес пакета $\bar{x} = 995$ грамм. Предполагается, что σ известна и равна 10 граммам. На 5%-ом уровне значимости, проверить, прав производитель или покупатель.

Решение:

Выдвигаем гипотезу $H_0 : m = 1000$

и альтернативную ей $H_1 : m < 1000$

$$u_{набл} = \frac{\bar{x} - m_0}{\sigma}\sqrt{n} = \frac{995 - 1000}{10}\sqrt{100} = -5$$

Так как $m < m_0$, строится левосторонняя критическая область

$$u_{кр.лев} = -\Phi^{-1}\left(\frac{1}{2} - \alpha\right) = -1{,}645$$

Так как $u_{набл} < u_{кр.лев}$ H_0 отвергается $m < 1000$ (прав покупатель, вес пакета с мукой меньше 1 кг)

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ: ПРИМЕР 2

На основании сделанного прогноза средняя дебиторская задолженность однотипных предприятий региона должна составить $m_0 = 120$ ден. ед. Выборочная проверка 10 предприятий установила, что средняя задолженность $\bar{x}$ =135 ден. ед., $S = 20$ ден. ед. На уровне значимости α =0,05 выяснить, можно ли принять данный прогноз?

Решение:

выдвигается нулевая гипотеза $H_0 : m = 120$,

альтернативная гипотеза $H_1 : m > 120$

Наблюдаемое значение статистики $t = \dfrac{\bar{x} - m_0}{S}\sqrt{n-1} = \dfrac{135-120}{20}\sqrt{10-1} = 2{,}25$,

Так как $m > m_0$ строим правостороннюю критическую область

$t_{кр.прав} = St^{-1}(2\alpha, n-1)$

$t_{кр.прав} = St^{-1}(0{,}1;9) = 1{,}83$ - по таблице Стьюдента.

$t_{набл} > t_{кр.прав}$, следовательно, гипотеза H_0 отвергается, т.е. на 5% уровне значимости сделанный прогноз должен быть отвергнут.

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ ГЕНЕРАЛЬНОЙ ДИСПЕРСИИ: ПОСТАНОВКА ЗАДАЧИ

Пусть $\xi_{1,n}$ - случайная выборка из генеральной совокупности $\xi \in N(m;\sigma^2)$, с неизвестными параметрами m и σ^2.

Точечной оценкой m является выборочное среднее $\bar{x}(\xi_{1,n}) = \frac{1}{n}\sum_{i=1}^{n}\xi_i$, значение выборочного среднего $\bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$.

Точечной оценкой σ^2 является выборочная дисперсия $S^2(\xi_{1,n}) = \frac{1}{n}\sum_{i=1}^{n}\left(\xi_i - \bar{x}(\xi_{1,n})\right)^2$, значение выборочной дисперсии $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Ставится задача на уровне значимости α проверить гипотезу $H_0 : \sigma^2 = \sigma_0^2$.

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ ГЕНЕРАЛЬНОЙ ДИСПЕРСИИ

Для проверки нулевой гипотезы H_0 используется статистика $\chi^2(\xi_{1,n}) = \dfrac{nS^2(\xi_{1,n})}{\sigma_0^2}$, где $S^2(\xi_{1,n}) = \dfrac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{x}(\xi_{1,n}))^2$, которая имеет распределение χ^2 - Пирсона с числом степеней свободы $(n-1)$.

Пусть $H_1 : \sigma^2 > \sigma_0^2$. Строим **правостороннюю критическую область**:
$P\{\chi^2(\xi_{1,n}) > \chi^2_{кр.прав}\} = \alpha$; $\chi^2_{кр.прав} = P_i^{-1}(\alpha, n-1)$ - определяется по таблице Пирсона с числом степеней свободы $\nu = n-1$.
Критическая область имеет вид: $(\chi^2_{кр.прав}; +\infty)$.

Пусть $H_1 : \sigma^2 < \sigma_0^2$. Строим **левостороннюю критическую область**:
$P\{\chi^2(\xi_{1,n}) < \chi^2_{кр.лев}\} = \alpha$; $\chi^2_{кр.лев} = P_i^{-1}(1-\alpha, n-1)$ - определяется по таблице Пирсона с числом степеней свободы $\nu = n-1$.
Левосторонняя критическая область имеет вид: $(0; \chi^2_{кр.лев})$.

Пусть $H_1 : \sigma^2 \neq \sigma_0^2$. Строим **двустороннюю критическую область**:
$P\{\chi^2(\xi_{1,n}) < \chi^2_{кр1}\} = \alpha/2$, $P\{\chi^2(\xi_{1,n}) > \chi^2_{кр2}\} = \alpha/2$;
$\chi^2_{кр1} = P_i^{-1}(1-\dfrac{\alpha}{2}, n-1)$, $\chi^2_{кр2} = P_i^{-1}(\dfrac{\alpha}{2}, n-1)$- определяются по таблице Пирсона с числом степеней свободы $\nu = n-1$.
Двусторонняя критическая область имеет вид: $(0; \chi^2_{кр1}) \cup (\chi^2_{кр2}; +\infty)$.

Если $\chi^2_{набл} = \dfrac{nS^2}{\sigma_0^2}$, где $S^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, принадлежит критической области, то гипотеза H_0 отвергается с вероятностью совершить ошибку первого рода α.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДВУХ МАТЕМАТИЧЕСКМХ ОЖИДАНИЙ: ПОСТАНОВКА ЗАДАЧИ

Пусть $\xi \in N(m_\xi, \sigma_\xi^2), \eta \in N(m_\eta, \sigma_\eta^2)$ – нормально распределенные генеральные совокупности с известными дисперсиями σ_ξ^2 и σ_η^2 и неизвестными математическими ожиданиями m_ξ и m_η.

$(\xi_1, \xi_2, ... \xi_n)$ и $(\eta_1, \eta_2, ... \eta_n)$ - две случайные выборки из генеральных совокупностей ξ $и$ η.

$x_{1,n} = (x_1, x_2, .. x_n)$ – реализация случайной выборки $\xi_{1,n}$.

$y_{1,n} = (y_1, y_2, .. y_n)$ – реализация случайной выборки $\eta_{1,n}$.

По выборкам (объема n_ξ и n_η) найдены выборочные оценки: $\overline{X}(\xi_{1,n})\, u\, \overline{Y}(\eta_{1,n})$

Требуется проверить гипотезу: H$_0$: $m_\xi = m_\eta$

при одной из альтернативных гипотез

H$_1$: $m_\xi > m_\eta$

H$_1$: $m_\xi < m_\eta$

H$_1$: $m_\xi \neq m_\eta$.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДВУХ МАТЕМАТИЧЕСКМХ ОЖИДАНИЙ: СЛУЧАЙ 1

Дисперсии σ_ξ^2 и σ_η^2 известны

Тогда при условии справедливости гипотезы H_0 статистика $u(\xi_{1,n},\eta_{1,n}) = \dfrac{\overline{X}(\xi_{1,n}) - \overline{Y}(\eta_{1,n})}{\sqrt{\dfrac{\sigma_\xi^2}{n_\xi} + \dfrac{\sigma_\eta^2}{n_\eta}}} \in N(0,1)$.

Пусть $H_1 : m_\xi > m_\eta$. Строим **правостороннюю критическую область**: $P\{u(\xi_{1,n}) > u_{\text{кр.прав}}\} = \alpha$, $u_{\text{кр.прав}} = \Phi^{-1}(\frac{1}{2} - \alpha)$ - определяется по таблице Лапласа. Правосторонняя критическая область имеет вид: $(u_{\text{кр.прав}};+\infty)$.

Пусть $H_1 : m_\xi < m_\eta$. Строим **левостороннюю критическую область**: $P\{u(\xi_{1,n}) < u_{\text{кр.лев}}\} = \alpha$, $u_{\text{кр.лев}} = -\Phi^{-1}(\frac{1}{2} - \alpha)$ - определяется по таблице Лапласа. Левосторонняя критическая область имеет вид: $(-\infty; u_{\text{кр.лев}})$.

Пусть $H_1 : m_\xi \neq m_\eta$. **Строим двустороннюю критическую область**:

$P\{u(\xi_{1,n}) < u_{\text{кр}1}\} = \alpha/2$, $P\{u(\xi_{1,n}) > u_{\text{кр}2}\} = \alpha/2$;

$u_{\text{кр}1} = -\Phi^{-1}(\frac{1}{2} - \frac{\alpha}{2})$, $u_{\text{кр}2} = \Phi^{-1}(\frac{1}{2} - \frac{\alpha}{2})$ - определяются по таблице Лапласа.

Двусторонняя критическая область имеет вид: $(-\infty; u_{\text{кр}1}) \cup (u_{\text{кр}2};+\infty)$.

Если $u_{\text{набл}} = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_\xi^2}{n_\xi} + \dfrac{\sigma_\eta^2}{n_\eta}}}$ принадлежит критической области, то гипотеза H_0 отвергается с вероятностью совершить ошибку первого рода α.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДВУХ МАТЕМАТИЧЕСКМХ ОЖИДАНИЙ: СЛУЧАЙ 2

Дисперсии σ_ξ^2 и σ_η^2 неизвестны.

Тогда при условии справедливости гипотезы H_0 статистика $T = \dfrac{\overline{X}(\xi_{1,n}) - \overline{Y}(\eta_{1,n})}{\sqrt{\dfrac{n_\xi S^2(\xi_{1,n}) + n_\eta S^2(\eta_{1,n})}{n_\xi + n_\eta - 2}}} \cdot \sqrt{\dfrac{n_\xi n_\eta}{n_\xi + n_\eta}}$, где $S^2(\xi_{1,n}) = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{x}(\xi_{1,n}))^2$, $S^2(\eta_{1,n}) = \frac{1}{n}\sum_{i=1}^{n}(\eta_i - \overline{y}(\eta_{1,n}))^2$, имеет распределение Стьюдента с $(n_\xi + n_\eta - 2)$ степенями свободы.

Пусть $H_1 : m_\xi > m_\eta$. Строим **правостороннюю критическую область**:
$P\{t(\xi_{1,n}) > t_{\text{кр.прав}}\} = \alpha$; $t_{\text{кр.прав}} = St^{-1}(2\alpha, n_\xi + n_\eta - 2)$ - определяется по таблице Стьюдента с числом степеней свободы $\nu = n_\xi + n_\eta - 2$.
Критическая область имеет вид: $(t_{\text{кр.прав}}; +\infty)$.

Пусть $H_1 : m_\xi < m_\eta$. Строим **левостороннюю критическую область**:
$P\{t(\xi_{1,n}) < t_{\text{кр.лев}}\} = \alpha$; $t_{\text{кр.лев}} = -St^{-1}(2\alpha, n_\xi + n_\eta - 2)$ - определяется по таблице Стьюдента с числом степеней свободы $\nu = n_\xi + n_\eta - 2$.
Левосторонняя критическая область имеет вид: $(-\infty; t_{\text{кр.лев}})$.

Пусть $H_1 : m_\xi \neq m_\eta$. Строим **двустороннюю критическую область**:
$P\{t(\xi_{1,n}) < t_{\text{кр}1}\} = \alpha/2$, $P\{t(\xi_{1,n}) > t_{\text{кр}2}\} = \alpha/2$;
$t_{\text{кр}1} = -St^{-1}(\alpha, n_\xi + n_\eta - 2)$, $t_{\text{кр}2} = St^{-1}(\alpha, n_\xi + n_\eta - 2)$ - определяются по таблице Стьюдента с числом степеней свободы $\nu = n_\xi + n_\eta - 2$.
Двусторонняя критическая область имеет вид: $(-\infty; t_{\text{кр}1}) \cup (t_{\text{кр}2}; +\infty)$.

Если $t_{\text{набл}} = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{n_\xi S_X^2 + n_\eta S_Y^2}{n_\xi + n_\eta - 2}}} \cdot \sqrt{\dfrac{n_\xi n_\eta}{n_\xi + n_\eta}}$,, где $S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$, $S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$ принадлежит критической области, то гипотеза H_0 отвергается с вероятностью совершить ошибку первого рода α.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДВУХ ДИСПЕРСИЙ: ПОСТАНОВКА ЗАДАЧИ

Пусть $\xi \in N(m_\xi, \sigma_\xi), \eta \in N(m_\eta, \sigma_\eta)$ - нормально распределенные генеральные совокупности.

$(\xi_1, \xi_2, ... \xi_n)$ и $(\eta_1, \eta_2, ... \eta_n)$ - две случайные выборки из генеральных совокупностей ξ *и* η

$x_{1,n} = (x_1, x_2, ... x_n)$ – реализация случайной выборки $\xi_{1,n}$.

$y_{1,n} = (y_1, y_2, ... y_n)$ – реализация случайной выборки $\eta_{1,n}$.

По выборкам (объема n_ξ и n_η) найдены выборочные оценки: $\hat{S}^2(\xi_{1,n}), \hat{S}^2(\eta_{1,n}), \quad (\hat{S}^2(\xi_{1,n}) > \hat{S}^2(\eta_{1,n}))$

Требуется проверить гипотезу H_0: $\sigma_\xi^2 = \sigma_\eta^2$

H_1: $\sigma_\xi^2 > \sigma_\eta^2$.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДВУХ ДИСПЕРСИЙ

Строится статистика: $F(\xi_{1,n},\eta_{1,n})=\dfrac{\hat{S}^2(\xi_{1,n})}{\hat{S}^2(\eta_{1,n}),}$, которая при справедливости нулевой гипотезы распределена по закону Фишера-Снедекора (F-распределение) с $\nu_1=n_\xi-1$ и $\nu_2=n_\eta-1$ степенями свободы.

Для проверки гипотезы строится правосторонняя критическая область из уравнения: $P(F(x_{1,n})>F_{kp.прав})=\alpha$, $F_{kp.прав}=F^{-1}(\alpha,n_x-1,n_y-1)$, определяется по таблице Фишера-Снедекора

Если $F_{набл}<F_{kp.прав}$, то H_0 принимается, если $F_{набл}>F_{kp.прав}$ H_0 отвергается.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ: ПРИМЕР

Произведены две выборки урожая пшеницы: при своевременной уборке урожая и уборке с некоторым опозданием. В первом случае при наблюдении 8 участков выборочная средняя урожайность составила 16,2 ц/га, а среднее квадратическое отклонение - 3,2 ц/га; во втором случае при наблюдении 9 участков те же характеристики равнялись соответственно 13,9 ц/га и 2,1 ц/га. На уровне значимости 0,05 выяснить влияние своевременности уборки урожая на среднее значение урожайности.

Решение.

H_0: $m_\xi = m_\eta$, т.е. средние значения урожайности при своевременной уборке урожая и с некоторым опозданием равны.

H_1: $m_\xi > m_\eta$, принятие которой означает существенное влияние на урожайность сроков уборки.

Значение статистики составило:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{n_x s_x^2 + n_y s_y^2}{n_x + n_y - 2}}} \cdot \sqrt{\frac{n_x n_y}{n_x + n_y}} = \frac{16{,}2 - 13{,}9}{\sqrt{\frac{9 \cdot 3{,}2^2 + 8 \cdot 2{,}1^2}{8 + 9 - 2}}} \cdot \sqrt{\frac{9 \cdot 8}{9 + 8}} = 1{,}62$$

Так как H_1: $m_\xi > m_\eta$, то строим правостороннюю критическую область:

$P(t > t_{kp}) = \alpha;$ $\quad t_{kp} = St^{-1}(2\alpha; n_x + n_y - 2) = St^{-1}(2 \cdot 0{,}05; 15) = St^{-1}(0{,}1; 15) = 1{,}75$

Так как $t = 1{,}62 < t_{kp} = 1{,}75$, то гипотеза H_0 принимается.

Имеющиеся выборочные данные на 5% уровне значимости не позволяют считать, что некоторое запаздывание в сроках уборки оказывает существенное влияние на величину урожая.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ГЕНЕРАЛЬНЫХ ДИСПЕРСИЙ: ПРИМЕР

До наладки станка была проверена точность изготовления 10 изделий и найдена оценка дисперсии контролируемого признака $s_1^2 = 9{,}6$. После наладки измерено еще 15 изделий и получена оценка дисперсии $s_1^2 = 5{,}7$. Можно ли считать, что точность изготовления изделий после наладки повысилась? ($\alpha = 0.05$)

РЕШЕНИЕ:

$H_0: \ \sigma_1^2 = \sigma_2^2$

$H_1: \ \sigma_1^2 > \sigma_2^2$

Значение статистики $F_{набл} = 9{,}6 / 5{,}7 = 1{,}68$

Для проверки гипотезы строится правосторонняя критическая область:

$F_{кр.прав}(0{,}05; 9; 14) = 2{,}65$

Так как $F_{набл} = 1{,}68 < F_{кр.прав} = 2{,}65$, то гипотеза H_0 принимается ($\sigma_1^2 = \sigma_2^2$), точность изготовления изделий до и после наладки одинаковая