Выполните задания в Excel   (в одном файле, но каждое на своём листе)

Задача 1

Используя процедуру ***Генерация случайных чисел*** из *Пакета анализа*, сделайте три класса нормальных выборок:

	мат.ожидание	дисперсия	объём	столбец
первая выборка	$a_1 = 1$	$\sigma^2 = 1$	$n_1 = 50$	A
вторая выборка	$a_2 = 2$	$\sigma^2 = 1$	$n_2 = 40$	B
третья выборка	$a_3 = 1$	$\sigma^2 = 4$	$n_3 = 60$	C

Выполните задания:

для каждой пары выборок проверить гипотезу H0 : a1 = a2 при альтернативе H1 : a1 $\neq$ a2. при уровне значимости α =0,05; Рассмотреть два случая: дисперсия известна и неизвестна.

для каждой пары выборок проверить гипотезу H0 : $\sigma^2_1 = \sigma^2_2$ при альтернативе H1: $\sigma^2_1 \neq \sigma^2_2$ при уровне значимости α =0,05;

для пары выборок A и C проверить гипотезу H0 : a1 = a3 при альтернативе H1 : a1 $\neq$ a3 при уровне значимости α = 0,01.

Решение

Для проверки каждой из гипотез необходимо рассчитать соответствующую статистику критерия и сравнить ее с соответствующим критическим значением.
Результаты решения задачи оформить в виде таблицы со следующими заголовками:

критерий	гипотеза	альтернатива	значение статистики критерия	критическое значение	вывод о принятии основной гипотезы

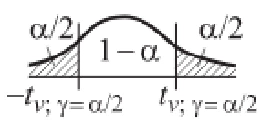
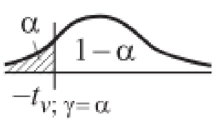
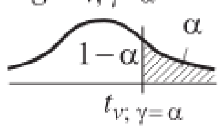
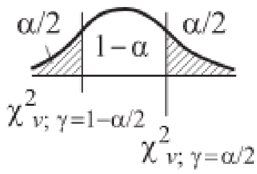
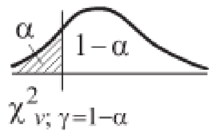
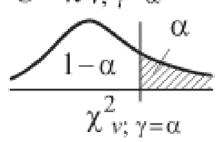
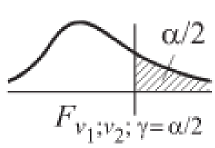
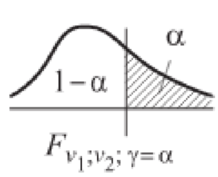
Для проверки гипотез согласия необходимо использовать указанные ниже варианты критериев, реализованные в Excel:

1. Используя инструмент анализа ***Двухвыборочный t-тест с одинаковыми дисперсиями*** из надстройки *Пакет анализа*, проверить гипотезу H0 : a1 = a2 при альтернативе H0 : a1 $\neq$ a2
и уровне значимости α = 0,05.

2. Используя инструмент анализа ***Двухвыборочный t-тест с разными дисперсиями*** из надстройки *Пакет анализа,* проверить гипотезу H0 : a1 = a3 при альтернативе H1 : a1 $\neq$ a3
и уровне значимости α = 0,01.

3. Используя инструмент анализа ***Двухвыборочный f-тест*** из надстройки *Пакет анализа*, рассмотреть гипотезу H0 : $\sigma^2_1 = \sigma^2_2$ при альтернативе H1: $\sigma^2_1 \neq \sigma^2_2$ и уровне значимости α = 0,05.

В качестве комментария ниже представлены некоторые стандартные критерии, позволяющие проверять гипотезы о значениях математических ожиданий и дисперсий нормальных генеральных совокупностей при независимых наблюдениях в выборке:

Проверяемая гипотеза H_0 и альтернативная гипотеза H_1	Информация о параметрах распределения	Статистика g, ее обозначение для каждого критерия	Распределение статистики g при справедливой гипотезе H_0	Критическая область
$H_0: m_x = m_0$ $H_1: m_x \neq m_0$	σ_x^2 неизвестно	$t = \dfrac{\bar{x} - m_0}{s_x/\sqrt{N}}$	t-распределение с $\nu = N-1$ степенями свободы	$\lvert g \rvert > t_{\nu;\,\gamma=\alpha/2}$ 
$H_0: m_x = m_0$ $H_1: m_x < m_0$				$g < -t_{\nu;\,\gamma=\alpha}$ 
$H_0: m_x = m_0$ $H_1: m_x > m_0$				$g > t_{\nu;\,\gamma=\alpha}$ 
$H_0: \sigma_x^2 = \sigma_0^2$ $H_1: \sigma_x^2 \neq \sigma_0^2$	m_x неизвестно	$t = \dfrac{(N-1)s_x^2}{\sigma_0^2}$	χ^2-распределение с $\nu = N-1$ степенями свободы	$g < \chi^2_{\nu;\,\gamma=1-\alpha/2}$ $g > \chi^2_{\nu;\,\gamma=\alpha/2}$ 
$H_0: \sigma_x^2 = \sigma_0^2$ $H_1: \sigma_x^2 < \sigma_0^2$				$g < \chi^2_{\nu;\,\gamma=1-\alpha}$ 
$H_0: \sigma_x^2 = \sigma_0^2$ $H_1: \sigma_x^2 > \sigma_0^2$				$g > \chi^2_{\nu;\,\gamma=\alpha}$ 
$H_0: \sigma_{x_1}^2 = \sigma_{x_2}^2$ $H_1: \sigma_{x_1}^2 \neq \sigma_{x_2}^2$	m_{x_1}, m_{x_2} неизвестны	$F = s_1^2/s_2^2$, где $s_1^2 = \max\{s_{x_1}^2,\ s_{x_2}^2\}$	F-распределение с ν_1, ν_2 степенями свободы, где ν_1 и ν_2 — число степеней свободы числителя и знаменателя	$g > F_{\nu_1;\nu_2;\,\gamma=\alpha/2}$ 
$H_0: \sigma_{x_1}^2 = \sigma_{x_2}^2$ $H_1: \sigma_{x_1}^2 > \sigma_{x_2}^2$		$F = s_{x_1}^2/s_{x_2}^2$		$g > F_{\nu_1;\nu_2;\,\gamma=\alpha}$ 
$H_0: \sigma_{x_1}^2 = \sigma_{x_2}^2$ $H_1: \sigma_{x_1}^2 < \sigma_{x_2}^2$		$F = s_{x_2}^2/s_{x_1}^2$		

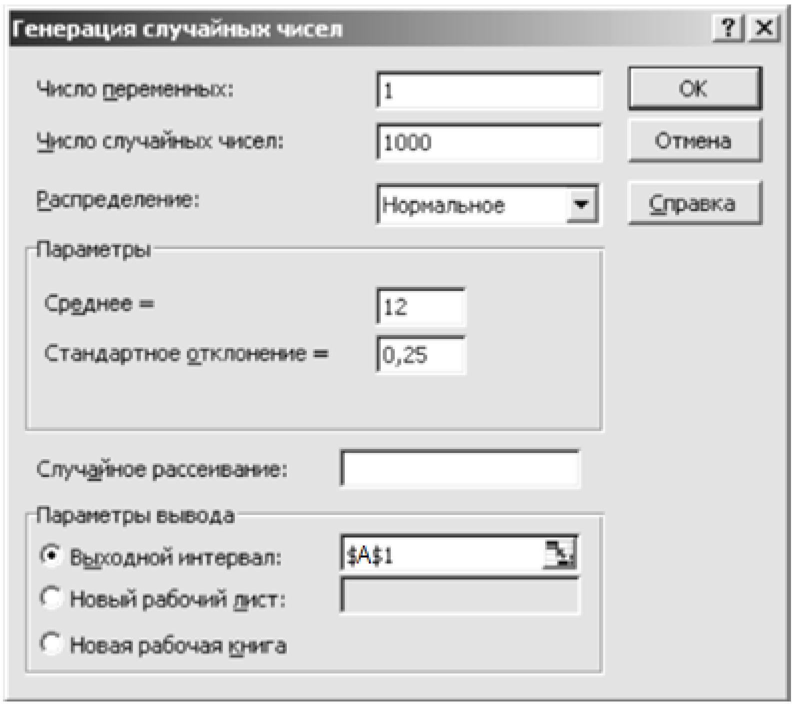
Задача 2

Проверка гипотезы о нормальном распределении по критерию Пирсона χ^2

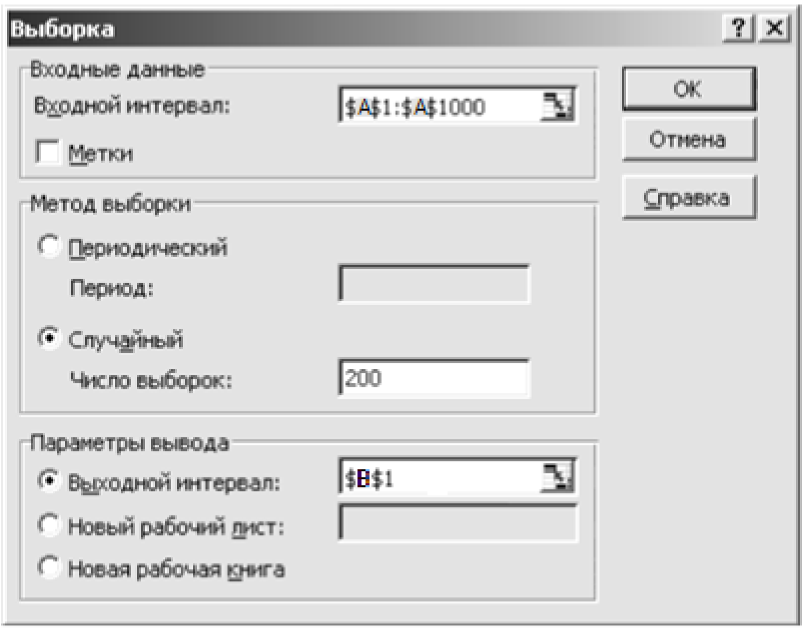
В типовой постановке задачи требуется, используя критерий Пирсона χ^2, проверить гипотезу о том, что генеральная совокупность X распределена нормально (*The chi-square goodness of fit test).*

В случае проверки сложных гипотез мы задаем только форму распределения, и параметры распределения, в отличие от простой гипотезы, неизвестны, т.е. из выборки сначала нужно оценить эти неизвестные параметры, затем вычислить статистику χ^2 (как и для простых гипотез).

На новом листе в Excel смоделируйте нормально распределенную совокупность из 1000 элементов с помощью инструмента «Генерация случайных чисел» из пакета «Анализ данных», разместите эти данные в столбце А.



Сформируйте случайную выборку из 200 элементов для этой совокупности с помощью инструмента «Выборка» из пакета «Анализ данных». Разместите эту выборку в столбце В.



Используя критерий χ^2 (хи-квадрат) проверим, действительно ли выборка сделана из нормально распределенной генеральной совокупности.

В качестве точечных оценок математического ожидания и дисперсии примите соответствующие выборочные характеристики. Найдите их, используя «Описательную статистику» из пакета «Анализ данных». Разместите результаты расчета в столбцах C и D.

С помощью инструмента «Гистограмма» найдите опытные частоты n_i . Включать саму диаграмму не нужно, достаточно оставить только расчет разбиения в столбцах F,G.

Расчётные частоты p_i вычисляются через вероятности попадания нормально распределенной величины в соответствующий интервал:

$$p_i = \Phi\left(\frac{x_{i+1} - m}{\sigma}\right) - \Phi\left(\frac{x_i - m}{\sigma}\right),$$

где функция стандартного нормального распределения Φ(·) вычисляется с помощью встроенной статистической функции НОРМ.РАСП(x, среднее значение m, стандартное отклонение σ, интегральный/весовой).

Аргументы функции НОРМ.РАСП: x - граница интервала, вводится адрес соответствующей ячейки; m и σ - вводятся абсолютные адреса характеристик, полученных с помощью «Описательной статистики»; значение Интегральный = (Истина) для функции распределения, в противном случае (Ложь) вычисляется плотность распределения.

В нашем случае таблицу вычисленных значений функции НОРМ.РАСП рассчитаем в колонке I . Вероятности p_i вычисляются как разности между значениями НОРМ.РАСП в текущей и предыдущей строках. В колонке J подсчитаем расчётные частоты $n*p_i$ для ***n*** = 200 [например, =(*I3* - *I2*)*200]

Для вычисления статистики χ^2 (хи-квадрат) в Excel есть функция ХИ2.ТЕСТ (интервал фактических данных, интервал ожидаемых данных). В качестве фактического интервала вводятся опытные частоты, в качестве ожидаемого – расчетные. Функция ХИ2.ТЕСТ возвращает ***р-вероятность*** того, что (при условии независимости) может быть получено такое значение статистики *χ2*, которое будет по крайней мере не ниже значения, рассчитанного по формуле

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(A_{ij} - E_{ij}\right)^2}{E_{ij}}$$

Для вычисления этой вероятности, функцией ХИ2.ТЕСТ используется распределение χ^2 с соответствующим числом степеней свободы (df).

Сделайте вывод для данного примера о принятии гипотезы H_0 "Выборка взята из нормально распределенного набора данных" при уровне значимости 0,05 [=ЕСЛИ(0,05>J17;"Отклонить";"Нет оснований для отклонения")].

На новом листе Excel, используя критерий Пирсона χ^2 при уровне значимости 0,05, проверьте, согласуется ли гипотеза о нормальном распределении совокупности X с эмпирическим распределением выборки объема N = 100, полученной в модели семинара 2 характеристики «длительность пребывания» (колонка W).

Комментарий

При использовании критерия χ^2 количество опытных значений в каждом интервале должно быть не менее 5. Если в каком-то интервале их меньше, то интервалы объединяют. Например, если в промежутке от 4 до 6 оказалось три значения, а в промежутке от 6 до 8 – четыре, то вводится новый интервал от 4 до 8 с семью значениями. С учетом этого перестройте таблицу частот вручную. Например, в колонках Карман – Частота (F:G) будут данные, полученные автоматически, то в колонках Границы – Опытные частоты (I:J) данные надо пересчитать частично вручную.

При применении критерия χ^2 необходимо следить за тем, чтобы объем выборки ***n*** был достаточно большой, иначе будет неправомочна аппроксимация χ^2 -распределением распределения статистики χ^2 . Обычно считается, что для этого достаточно, чтобы наблюдаемые частоты (Observed) были больше 5. Если это не так, то малые частоты объединяются в одно или присоединяются к другим частотам, причем объединенному значению приписывается суммарная вероятность и, соответственно, уменьшается число степеней свободы χ^2 -распределения. Для того чтобы улучшить качество применения критерия χ^2 (увеличить его мощность), необходимо уменьшать интервалы разбиения (увеличивать количество степеней свободы), однако этому препятствует ограничение на количество попавших в каждый интервал наблюдений (нужно >5).

В случае сложной гипотезы, *p*-значение, которое мы сравниваем с уровнем значимости, рассчитывается с использованием χ^2-распределения с L-k-1 степеней свободы, где k – количество оцениваемых параметров.

Если вероятность, того что случайная величина имеющая χ^2-распределение с L-k-1 степенями свободы примет значение больше вычисленной статистики χ^2 , т.е. $\chi^2_{L-k-1} > \chi^2_0$ меньше уровня значимости, то нулевая гипотеза отклоняется.

Граница критической области – это квантиль распределения хи-квадрат, которая также может быть найдена с помощью встроенной функции ХИ2.ОБР.ПХ (вероятность, степени свободы). Аргумент «вероятность» – это уровень значимости (α = 0,05), а «степени свободы» df$=k-l-1$ определяются как количество интервалов (здесь k = 11) за вычетом количества оцениваемых параметров (здесь два – m и σ) и минус 1.