

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №8
по дисциплине «Медицинская информатика»
для зарубежных студентов
по направлению подготовки «Лечебное дело»

Тема: «Компьютерный анализ медицинских данных»

Законы распределения случайных величин. Методы математической статистики: проверка статистических гипотез

ЗАДАНИЕ 1. В рабочей тетради составить опорный конспект по теме занятия, используя следующий материал:

Статистическая гипотеза – это любое предположение, которое относится к виду распределения или отдельным параметрам случайных величин.

Практически – это любое высказывание о генеральной совокупности, проверяемое по выборочным данным.

Статистическая гипотеза может относиться к одной выборке или к нескольким. Например:

- 1) данная выборка извлечена из генеральной совокупности с нормальным законом распределения;
- 2) две выборки относятся к одной генеральной совокупности, т.е. имеют одинаковые генеральные средние и дисперсии.

Гипотезы о типах распределения и о значениях параметров распределения называют параметрическими.

Гипотезы, в которых о типе распределения никаких предположений не делается, называются непараметрическими.

Первая из приведенных выше гипотез – параметрическая, а вторая – непараметрическая.

Задача проверки гипотез. Ошибки первого и второго рода. Уровень значимости.

Вернемся к примеру, представленному в таблице. В процессе проверки гипотез всегда имеются две стороны: «скептик» и «оптимист». «Скептик» считает, что первая методика не лучше второй, а лучший показатель числа необходимых процедур получился случайно. Из-за малого объема выборки.

Он говорит $\bar{X}_Г = \bar{Y}_Г$ (гипотеза скептика). «Оптимист» считает, что $\bar{X}_Г < \bar{Y}_Г$ (гипотеза оптимиста). Арбитром между «скептиком» и «оптимистом» выступает статистика.

При проверке гипотез принято формулировать сразу две гипотезы:

H_0 – *нулевая гипотеза* – гипотеза об отсутствии различий между сравниваемыми выборками.

H_1 – *альтернативная (конкурирующая) гипотеза* – гипотеза о наличии различий между сравниваемыми выборками.

Задача проверки гипотез заключается в том, чтобы на основании анализа выборочных данных принять решение о справедливости одной из них.

Однако, при проверке гипотез могут быть допущены ошибки двух видов:

1. Принимается H_1 , когда верна H_0 – *ошибка первого рода*.
2. Принимается H_0 , когда верна H_1 – *ошибка второго рода*.

Вероятность совершить ошибку первого рода должна быть достаточно мала, т.к. нужны веские основания для признания того, что один метод лечения лучше другого, т.е. $\overline{X}_Г \neq \overline{Y}_Г$. Эта вероятность называется *уровнем значимости* и обозначается α .

При проверке статистических гипотез уровнем значимости (α) называется максимально приемлемая вероятность, отвергающая нулевую гипотезу, если она верна. И тогда можно сказать, что результаты значимы на уровне α .

В медицинских исследованиях обычно используют $\alpha=0.05$ или $\alpha=0.01$.

Как правило, для большинства случаев 5%-ный уровень значимости является вполне достаточным. И только при рассмотрении вопросов, в которых ошибочное отклонение нулевой гипотезы может повлечь тяжелые последствия (например, увеличение частоты летальных исходов), используют меньший уровень значимости.

Процедура (алгоритм) проверки гипотез. Критерий.

Процедура – это правило (алгоритм действий), руководствуясь которым принимается статистически обоснованное решение о справедливости одной из двух гипотез.

Она состоит из следующих этапов:

1. Формулируют (выдвигают) нулевую гипотезу H_0 об отсутствии различий между группами, об отсутствии существенного отличия фактического распределения от некоторого заданного, например, нормального, экспоненциального и др.

Сущность нулевой гипотезы H_0 : разница между сравниваемыми генеральными параметрами равна нулю, и разница, наблюдаемая между выборочными характеристиками, носит случайный характер, т.е. эти выборки принадлежат одной генеральной совокупности.

2. Формулируют противоположную нулевой альтернативную гипотезу H_1 .

3. Задают уровень значимости α . Уровень значимости – это вероятность отвергнуть нулевую гипотезу, если на самом деле она верна. При $\alpha \leq 0.05$ ошибка возможна в 5% случаев.

4. Для проверки выдвинутой гипотезы используют критерии.

Критерий – это случайная величина K , которая служит для проверки нулевой гипотезы и описывается определенной функцией. Эти функции распределения известны и табулированы. Критерий зависит от двух параметров: от числа степеней свободы f и от уровня значимости α . Выбирают критерий, который соответствует данной задаче и вычисляют по

соответствующим формулам фактическую величину критерия по данным наблюдения $K_{набл}$.

5. По таблице определяют критическое значение, превышение которого при справедливости гипотезы маловероятно $K_{крит}(\alpha, f)$.

6. Сравнивают $K_{набл}$ и $K_{крит}(\alpha, f)$.

Если $K_{набл} > K_{крит}(\alpha, f)$, то H_0 отвергают и принимают H_1 .

Если $K_{набл} \leq K_{крит}(\alpha, f)$, то принимают H_0 .

Это правило для параметрических критериев.

Если использованы непараметрические критерии, то наоборот: если $K_{набл} \geq K_{крит}(\alpha, f)$, то принимается H_0 .

7. Делают вывод: различие статистически значимо при $\alpha \leq 0.05$ или незначимо (см. пункт 1).

Проверка гипотез относительно средних больших выборок. z – критерий

Предположим, что надо сравнить состояние больного до и после лечения. Для этого сравнивают друг с другом две независимые выборки объемом n_1 и n_2 , взятых из нормально распределенных совокупностей с параметрами $M(X_1)$ и $M(X_2)$

Для выборочных данных найдены выборочные средние \bar{x}_1 и \bar{x}_2 и исправленные дисперсии S_1^2 и S_2^2 . Уровень значимости задан.

1. Нулевая гипотеза $H_0: M(X_1) = M(X_2)$
2. Конкурирующая гипотеза $H_1: M(X_1) \neq M(X_2)$
3. Для проверки нулевой гипотезы в случае больших выборок можно использовать Z-критерий сравнения средних.

$$z_{набл} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1 - 1} + \frac{\sigma_2^2}{n_2 - 1}}}$$

Найденное значение $z_{набл}$ надо сравнить с $\Phi(z_{кр})$ – функция Лапласа

$$\Phi(z_{кр}) = \frac{1 - \alpha}{2} \quad \text{если конкурирующая гипотеза } M(X_1) \neq M(X_2)$$

$$\text{Или } \Phi(z_{кр}) = \frac{1 - 2\alpha}{2} \quad \text{если конкурирующая гипотеза } M(X_1) > M(X_2)$$

4. Сравниваем $z_{крит}$ и $z_{набл}$

Если $z_{крит} \geq z_{набл} \Rightarrow$ принимается H_0 .

Если $z_{крит} < z_{набл} \Rightarrow H_0$ отвергается и принимается H_1 о том, что различия достоверны при данном уровне значимости α .

Проверка гипотез относительно средних. t – критерий Стьюдента

Предположим, что надо сравнить состояние больного до и после лечения. Для этого сравнивают друг с другом две независимые выборки объемом n_1 и n_2 , взятых из **нормально** распределенных совокупностей с параметрами $M(X_1)$ и $M(X_2)$. Дополнительно предполагаем, что неизвестные генеральные дисперсии **равны** между собой, т.е. $D(X_1)=D(X_2)$.

Для выборочных данных найдены выборочные средние \bar{x}_1 и \bar{x}_2 и исправленные дисперсии S_1^2 и S_2^2 . Уровень значимости задан.

1. Нулевая гипотеза $H_0: M(X_1) = M(X_2)$
2. Конкурирующая гипотеза $H_1: M(X_1) \neq M(X_2)$
3. Для проверки нулевой гипотезы в этом случае можно использовать t -критерий Стьюдента сравнения средних.

Величину критерия находим по формуле:

$$t_{\text{набл}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} \cdot \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

В статистике доказано, что величина $t_{\text{набл}}$ при справедливости нулевой гипотезы имеет t -распределение Стьюдента с $f=n_1+n_2-2$ степенями свободы.

4. Задаем уровень значимости и по таблице находим $t_{\text{крит}}(\alpha, f)$
5. Сравниваем $t_{\text{крит}}(\alpha, f)$ и $t_{\text{набл}}$

Если $t_{\text{крит}}(\alpha, f) \geq t_{\text{набл}} \Rightarrow$ принимается H_0 .

Если $t_{\text{крит}}(\alpha, f) < t_{\text{набл}} \Rightarrow H_0$ отвергается и принимается H_1 о том, что различия достоверны при данном уровне значимости α .

Обратите внимание, что данная формула используется только для случая выборок из нормально распределенных генеральных совокупностей с одинаковыми дисперсиями. **Поэтому предварительно** проводится процедура проверки гипотез о равенстве генеральных дисперсий.

В случае, если генеральные дисперсии неравны, применяется модифицированный критерий.

Критерий для проверки гипотезы о сравнении двух дисперсий.

Пусть имеются две нормально распределенные генеральные совокупности X и Y . Из них извлечены независимые выборки объемов соответственно n_1 и n_2 , по которым вычислены исправленные выборочные дисперсии s_x^2 и s_y^2 . Требуется при заданном уровне значимости α проверить нулевую гипотезу $H_0: D(X) = D(Y)$ о равенстве дисперсий рассматриваемых генеральных совокупностей. Учитывая несмещенность исправленных выборочных дисперсий, можно записать нулевую гипотезу так:

$$H_0: M\left(\frac{s_x^2}{x}\right) = M\left(\frac{s_y^2}{y}\right).$$

Замечание. Конечно, исправленные дисперсии, вычисленные по выборкам, обычно оказываются различными. При проверке гипотезы выясняется, является ли это различие незначимым и обусловленным случайными

причинами (в случае принятия нулевой гипотезы) или оно является следствием того, что сами генеральные дисперсии различны.

В качестве критерия примем случайную величину

$$F = \frac{S^2}{S_M^2}$$

- отношение большей выборочной дисперсии к меньшей. Она имеет распределение Фишера-Снедекора со степенями свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$, где n_1 – объем выборки, по которой вычислена большая исправленная дисперсия, а n_2 – объем второй выборки. Рассмотрим два вида конкурирующих гипотез:

- пусть $H_1: D(X) > D(Y)$. Наблюдаемым значением критерия будет отношение большей из исправленных дисперсий к меньшей: $F_{набл} = \frac{s^2}{S_M^2}$. По таблице

критических точек распределения Фишера-Снедекора можно найти критическую точку $F_{набл}(\alpha; k_1; k_2)$. При

$F_{набл} < F_{кр}$ нулевая гипотеза принимается, при $F_{набл} > F_{кр}$ отвергается.

- если $H_1: D(X) \neq D(Y)$, то критическая область является двусторонней и определяется неравенствами $F < F_1, F > F_2$, где $p(F < F_1) = p(F > F_2) = \alpha/2$.

При этом достаточно найти правую критическую точку $F_2 = F_{кр}(\frac{\alpha}{2}, k_1, k_2)$.

Тогда при $F_{набл} < F_{кр}$ нулевая гипотеза принимается, при $F_{набл} > F_{кр}$ отвергается.

ЗАДАНИЕ 2. Используя программу MS Excel выполнить следующие задачи, результаты и выводы оформить в тетради:

1) Анализ двух выборок

Критерий Стьюдента (t) наиболее часто используется для проверки гипотезы: «Средние двух выборок относятся к одной и той же совокупности». Критерий позволяет найти вероятность того, что оба средних относятся к одной и той же совокупности. Если эта вероятность p ниже уровня значимости ($p < 0,05$), то принято считать, что выборки относятся к двум разным совокупностям.

Задача 1. Рассматриваются ежемесячные количества реализованных фирмой путевок за периоды до и после начала активной рекламной кампании.

С рекламой	Без рекламы
162	135
156	126
144	115
137	140
125	121
145	112
151	130

Выявить, достоверны ли отличия при сравнении данных реализации путевок за периоды до и после начала активной рекламной кампании.

1. Введите данные в рабочую таблицу.
2. Для выявления достоверности отличий табличный курсор установите в ячейку A11. Откройте диалоговое окно Мастер функций и выберите в категории Статистические функцию ТТЕСТ, после чего нажмите кнопку Ок. В диалоговом окне ТТЕСТ указателем мыши введите диапазон данных контрольной группы в поле Массив 1 (A2:A8). В поле Массив 2 введите диапазон данных исследуемой группы (B2:B8). В поле Хвосты всегда вводится с клавиатуры цифра 2, а в поле Тип введите – цифру 3. Нажмите кнопку Ок. В ячейке A11 появится значение вероятности – 0,006295.
3. Т.к. величина вероятности случайного появления анализируемых выборок меньше уровня значимости, то нулевая гипотеза отвергается. Следовательно, различия между выборками не случайные и средние выборок считаются достоверно отличающимися друг от друга. Поэтому на основании критерия Стьюдента можно сделать вывод о большей эффективности реализации путевок после начала рекламной кампании ($p < 0,05$).

При использовании t -критерия выделяют два основных случая. В первом случае его применяют для проверки гипотезы о равенстве генеральных средних двух независимых, несвязанных выборок (так называемый двухвыборочный t -критерий). В этом случае есть две различных выборки, количество элементов в которых может быть также различно. При заполнении диалогового окна ТТЕСТ при этом указывается Тип 3.

Во втором случае, когда одна и та же группа объектов порождает числовой материал для проверки гипотез о средних, используется так называемый парный t -критерий. Выборки при этом называют зависимыми, связанными (при заполнении диалогового окна ТТЕСТ указывается Тип 1).

Задача 2. Сравнивается количество наличных денег у двух групп студентов:

30	10
60	20
40	30
50	40
60	50

Необходимо определить достоверность различия между группами при двух вариантах постановки задачи:

- группы состоят из различных студентов (тип 3);
- группы состоят из одних и тех же студентов, но первая – до посещения буфета, а вторая – после (тип 1).

1. В ячейки C1:C5 введите количество денег у студентов первой группы. В ячейки D1:D5 введите количество денег у студентов второй группы.
2. Табличный курсор установите в ячейку C6. Откройте диалоговое окно Мастер функций, в категории Статистические выберите функцию ТТЕСТ, нажмите кнопку Ok. В диалоговом окне ТТЕСТ указателем мыши введите диапазон данных первой группы в поле Массив 1 (C1:C5). В поле Массив 2 введите диапазон данных второй группы (D1:D5). В поле Хвосты всегда вводится цифра 2, а в поле Тип введите цифру 3. Нажмите кнопку Ok. В ячейке C6 появится значение вероятности – 0,228053.

Т.к. величина вероятности случайного появления анализируемых выборок больше уровня значимости, то нулевая гипотеза не может быть отвергнута (принимается). Следовательно, различия между выборками могут быть случайными и средние выборок не считаются достоверно отличающимися друг от друга. Поэтому на основании критерия Стьюдента нельзя сделать вывод о достоверности отличий двух групп студентов по количеству карманных денег, имеющихся у них ($p < 0,05$).

3. Табличный курсор установите в ячейку D6. Откройте диалоговое окно ТТЕСТ. Указателем мыши введите диапазон данных первой группы в поле Массив 1 (C1:C5). В поле Массив 2 введите диапазон данных второй группы (D1:D5). В поле Хвосты всегда вводится цифра 2, а в поле Тип введите цифру 1. Нажмите кнопку Ok. В ячейке D6 появится значение вероятности – 0,003883.

Т.к. величина вероятности случайного появления анализируемых выборок меньше уровня значимости, то нулевая гипотеза отвергается. Следовательно, различия между выборками не могут быть случайными и

средние выборок считаются достоверно отличающимися друг от друга. Поэтому на основании критерия Стьюдента можно сделать вывод о том, что в двух группах студентов выявлены достоверные отличия по количеству карманных денег ($p < 0,05$), что явилось результатом посещения буфета.

Критерий согласия χ^2 является непараметрическим критерием. Используется в тех случаях, когда закон распределения данных отличается от нормального или неизвестен. Принимается нулевая гипотеза о том, что выборки принадлежат к одной генеральной совокупности. Кроме того, определяется ожидаемое значение результата. Обычно это среднее значение между выборками рассматриваемого показателя. Затем оценивается вероятность того, что ожидаемые значения и наблюдаемые принадлежат к одной генеральной совокупности.

Задача 3. Пусть после окончания двух медицинских институтов трудоустроилось по специальности из первого института 90 человек, а из второго – 60 (обе группы молодых специалистов включали по 100 человек).

1. Принимается нулевая гипотеза, что выборки принадлежат к одной генеральной совокупности.
2. Определяется ожидаемое значение результата (среднее значение между выборками: $(60+90)/2=75$), т.е. мы ожидали, что разницы между группами нет, и в обоих случаях должно было трудоустроиться по 75 человек.
3. Затем вычисляется значение вероятности того, что изучаемые события произошли случайным образом. Для этого введите данные в рабочую таблицу: 90 – в ячейку E1, 60 – в F1, 75 – в E2, F2. Табличный курсор установите в ячейку E3. Откройте диалоговое окно Мастер функций, в категории Статистические выберите функцию ХИ2ТЕСТ, после чего нажмите кнопку Ok. В диалоговом окне ХИ2ТЕСТ указателем мыши введите диапазон данных наблюдавшегося количества трудоустроившихся в поле Фактический интервал (E1:F1). В поле Ожидаемый интервал введите диапазон данных предполагаемого количества трудоустроившихся (E2:F2). Нажмите кнопку Ok. В ячейке E3 появится значение вероятности – 0,014306.

Т.к. величина вероятности случайного появления анализируемых выборок меньше уровня значимости, то нулевая гипотеза отвергается. Следовательно, различия между выборками не могут быть случайными и выборки считаются достоверно отличающимися друг от друга. Поэтому на основании критерия хи-квадрат можно сделать вывод о том, что в двух группах выпускников выявлены достоверные отличия по успешности

трудоустройства ($p < 0,05$), что, по-видимому, явилось результатом более высокой репутации выпускников первого института.

Задача 4. В ходе социологического опроса на вопрос о перенесенном в детстве заболевании ответы распределились следующим образом:

	Да	Нет	Не помню
Мужчины	58	11	10
Женщины	35	25	23

Есть ли достоверные отличия в ответах женщин и мужчин?

Задача 5. Определите, имеют ли выборки $\{6, 7, 9, 15, 21\}$ и $\{20, 28, 31, 38, 40\}$ различные уровни разнородности (отличаются ли дисперсии)?

Задача 6. Даны результаты бега на дистанции 100 м в секундах в двух группах студентов. Студенты первой группы в течение года посещали факультативные занятия по физкультуре. Определите, достоверны ли отличия по результатам бега в этих группах.

Посещавшие факультатив	Не посещавшие
12,6	12,8
12,3	13,2
11,9	13,0
12,2	12,9
13,0	13,5
12,4	13,1

2) Проверка соответствия теоретическому распределению

Наиболее часто проверяется предположение о нормальном распределении генеральной совокупности, т.к. большинство статистических процедур ориентировано на выборки, полученные из нормально распределенной генеральной совокупности.

Для оценки соответствия имеющихся экспериментальных данных нормальному закону распределения обычно используют графический метод, выборочные параметры формы распределения и критерии согласия.

Наиболее убедительные результаты дает использование критериев согласия. Здесь нулевая гипотеза представляет собой утверждение о том, что распределение генеральной совокупности, из которой получена выборка, не отличается от нормального. Среди критериев согласия большое

распространение получил непараметрический критерий χ^2 (хи-квадрат). Он основан на сравнении эмпирических частот интервалов группировки с теоретическими (ожидаемыми) частотами, рассчитанными по формулам нормального распределения.

В MS Excel критерий χ^2 реализован функцией ХИ2ТЕСТ, которая вычисляет вероятность совпадения наблюдаемых (фактических) значений и теоретических значений. Если вычисленная вероятность ниже уровня значимости (0,05), то нулевая гипотеза отвергается и утверждается, что наблюдаемые значения не соответствуют нормальному закону распределения. Если вычисленная вероятность близка к 1, то можно говорить о высокой степени соответствия экспериментальных данных нормальному закону распределения.

Задача 7. Проверить соответствие выборочных данных: 64, 57, 63, 62, 58, 61, 63, 60, 60, 61, 65, 62, 62, 60, 64, 61, 59, 59, 63, 61, 62, 58, 58, 63, 61, 59, 62, 60, 60, 58, 61, 60, 63, 63, 58, 60, 59, 60, 59, 61, 62, 62, 63, 57, 61, 58, 60, 64, 60, 59, 61, 64, 62, 59, 65 нормальному закону распределения.

Решение

1. В диапазон A1:E12 введите данные как показано на рисунке.
2. Выберите ширину интервала 1 кг. Тогда при крайних значениях веса 57 кг и 65 кг получится 9 интервалов. В диапазон G1:G12 введите названия и граничные значения интервалов.
3. Введите заголовки создаваемой таблицы (диапазон H1:L3).
4. Заполните столбец абсолютных частот. Для этого выделите блок ячеек H4:H12, вызовите функцию ЧАСТОТА, в появившемся диалоговом окне в поле Массив_данных введите диапазон наблюдений (A2:E12), в рабочее поле Массив_интервалов введите диапазон интервалов (G4:G12). Нажмите комбинацию клавиш Ctrl+Shift+Enter. В столбце H4:H12 появится массив абсолютных частот.
5. В ячейке H13 найдите общее количество наблюдений.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Наблюдения						Вес	Абсолютные	Относи-	Накопленные	Теоретические	Теоретические
2	64	62	58	63	61		кг	частоты	тельные	частоты	частоты	частоты
3	57	62	63	58	58				частоты			
4	63	60	61	60	60		57	2	0,036	0,036	0,033205828	1,82632055
5	62	64	59	59	64		58	6	0,109	0,145	0,073795567	4,058756212
6	58	61	62	60	60		59	7	0,127	0,273	0,129258576	7,109221655
7	61	59	60	59	59		60	10	0,182	0,455	0,178443849	9,814411704
8	63	59	60	61	61		61	9	0,164	0,618	0,194158732	10,67873029
9	60	63	58	62	64		62	8	0,145	0,764	0,16650428	9,157735407
10	60	61	61	62	62		63	7	0,127	0,891	0,112540024	6,189701326
11	61	62	60	63	59		64	4	0,073	0,964	0,059951732	3,297345259
12	65	58	63	57	65		65	2	0,036	1,000	0,025171529	1,384434082
13								55	60,855	2,050		0,984
14									ср.знач.	станд.отклон.		p
15												
16												
17	<p>Диаграмма относительных и накопленных частот</p>											
18												
19												
20												
21												
22												
23												
24												
25												
26												
27												
28												
29												
30												
31												
32												
33												

- Заполните столбец относительных частот. В ячейку I4 введите формулу для вычисления относительной частоты: =H4/H\$13. Нажмите клавишу Enter. Протягиванием скопируйте введенную формулу в диапазон I5:I12.
- Заполните столбец накопленных частот. В ячейку J4 скопируйте значение относительной частоты из ячейки I4. В ячейку J5 введите формулу: =J4+I5. Нажмите Enter. Протягиванием скопируйте введенную формулу в диапазон J6:J12.
- Найдите теоретические частоты нормального распределения. Для этого предварительно необходимо найти среднее значение и стандартное отклонение выборки. В ячейке I13 найдите среднее значение для данных из диапазона A2:E12. В ячейке J13 найдите стандартное отклонение для этих же данных. В ячейки K1:L3 введите названия столбцов. Затем с помощью функции НОРМРАСП найдите теоретические частоты. Установите курсор в ячейку K4, вызовите указанную функцию и заполните ее рабочие поля: x – G4, Среднее –

σ_{I13} , Стандартное_откл – σ_{J13} , Интегральный – 0. Далее протягиванием скопируйте содержимое ячейки K4 в диапазон ячеек K5:K12. Затем установите курсор в ячейку L4 и введите формулу: $=N_{I13} * K4$. Протягиванием скопируйте содержимое ячейки L4 в диапазон ячеек L5:L12.

9. С помощью функции ХИ2ТЕСТ определите соответствие данных нормальному закону распределения. Для этого установите табличный курсор в ячейку L13. Выберите функцию ХИ2ТЕСТ и в появившемся диалоговом окне в рабочие поля введите фактический $N4:N12$ и ожидаемый $L4:L12$ диапазоны частот. Нажмите Ок. В ячейке L13 появится значение вероятности того, что выборочные данные соответствуют нормальному закону распределения.

Вывод: Т.к. полученная вероятность соответствия экспериментальных данных $p=0,98$ много больше, чем уровень значимости $\alpha=0,05$, от можно утверждать, что нулевая гипотеза не может быть отвергнута и, следовательно, данные не противоречат нормальному закону распределения. Более того, поскольку полученная вероятность близка к 1, можно говорить о высокой степени вероятности того, что экспериментальные данные соответствуют нормальному закону.

Задача 8. Найдите соответствие экспериментальных данных нормальному закону распределения для следующей выборки весов детей (кг): 21 (2), 22 (9), 23 (6), 24 (11), 25 (17), 26 (14), 27 (2).