Глава 6

Критерии однородности

Мы будем рассматривать следующую задачу: проверку гипотезы однородности — по нескольким выборкам $X_{1,1}, ..., X_{1,n_1}, X_{2,1}, ..., X_{2,n_2}, ..., X_{k,1}, ..., X_{k,n_k}$, $n_1 + ... + n_k = n$, проверить гипотезу о том, что выборки из одного распределения. В англоязычной литературе такого рода критерии называют k-sample test.

6.1 Задача о двух выборках

При этом будет изучать две альтернативы — общую H_1 : выборки из разных распределений или доминирования $H_2 : F_1(x) \geq F_2(x)... \geq F_k(x)$ при всех x, т.е. каждое следующее распределение в целом выдает более большие наблюдения.

Вопрос 1. Показать, что $F_X(x) \geq F_Y(x)$ при всех x равносильно тому, что на некотором вероятностном пространстве можно задать $X_1 \stackrel{d}{=} X$, $Y_1 \stackrel{d}{=} Y$, т.ч. $Y_1 \geq X_1$ п.н.
Начнем с модификации известных нам критериев.

6.1.1 Старые знакомые

1) Критерий Смирнова.
Для $k = 2$ и непрерывных распределений F_X, F_Y можно модернизировать критерий Колмогорова.
Пусть $\hat{F}_{i,n_i}$ — ЭФР, распределения F_i непрерывны, $i = 1, 2$. Тогда положим

$$D_n = \sup |\hat{F}_{1,n_1} - \hat{F}_{2,n_2}|.$$

Мы понимаем, что если гипотеза верна и функции распределения выборок одинаковы, то их эмпирические функции распределения близки. Более точно, оказывается, что при выполнении гипотезы $\sqrt{\frac{n_1 n_2}{n_1+n_2}} D_n \to K \sim K(x)$, где $K(x)$ — ф.р. Колмогорова, $n_1, n_2 \to \infty$, откуда получаем критическое множество

$$\sqrt{\frac{n_1 n_2}{n_1 + n_2}} D_n > k_{1-\alpha}.$$

В случае альтернативы доминирования H_2 рассматривают статистику

$$D_n^+ = \sup(\hat{F}_{2,n_2} - \hat{F}_{1,n_1}).$$

При выполнении основной гипотезы $\sqrt{\frac{n_1 n_2}{n_1+n_2}} D_n^+ \to 1 - e^{-2x^2}$ при $n_1, n_2 \to \infty$ откуда получаем критическое множество

$$\frac{\sqrt{n_1 n_2}}{\sqrt{n_1 + n_2}} D_n^+ > \sqrt{-\ln \sqrt{\alpha}}.$$

В R критерий Колмогорова доступен в той же функции ks.test, параметр alternative может быть выбран "less" или "greater" для одностороннего критерия. В Python для него есть функция ks_2samp все в том же scipy.stats.

2) Критерий Розенблатта (two sample Lehman test).

Этот критерий является модификацией критерия Крамера-Мизеса. Тестовая статистика критерия

$$\int\limits_{\mathbb{R}} (\hat{F}_{1,n_1}(x) - \hat{F}_{2,n_2}(x))^2 d\hat{H}_2(x),$$

где $\hat{H}_2(x) = (n_1\hat{F}_{1,n_1}(x) + n_2\hat{F}_{2,n_2}(x))/(n_1+n_2)$ — совместная эмпирическая фукция распределения двух выборок. При выполнении гипотезы однородности она будет сходиться к тому же распределению, что и статистика Крамера-Мизеса.

Другой формой той же статистики является

$$\frac{1}{n_1 n_2}\left(\frac{1}{6} + \frac{1}{n_2}\sum_{i=1}^{n_1}(R_i - i)^2 + \frac{1}{n_1}\sum_{i=1}^{n_2}(S_i - i)^2\right) - \frac{2}{3},$$

где R_i — ранги элементов $X_{1,i}$, S_i — ранги элементов $X_{2,i}$ в общем вариационном ряду.

Этот критерий является модификацией критерий Крамера-Мизеса.

Задача 1. Проверить критерии на выборках из $\mathcal{N}(0,1)$ и Лапласа размера а) 30 б) 50 в) 100.

3) Критерий Андерсона-Дарлинга.

Для нескольких выборок можно модернизировать и критерий Андерсона-Дарлинга. Такой критерий был предложен F.W. Scholz, M.A. Stephens в 1987 году. Они предложили рассматривать статистику

$$A_k^2 = \sum_{i=1}^{k} n_i \int_B \frac{(\hat{F}_{i,n_i}(x) - H_k(x))^2}{H_k(x)(1-H_k(x))} dH_k(x),$$

где H_k — совместная ЭФР всех выборок, $B = \{x \in \mathbb{R} : H_k(x) < 1\}$. Оказывается, что для неперывных величин A_k^2 с ростом длин выборок сходится по распределению к $\sum_{j=1}^{\infty} \frac{1}{j(j+1)} Y_j$, $Y_j \sim \chi^2_{j-1}$ независимы. Критерий можно применять и в не непрерывном случае. В R этот критерий представлен функцией AndersonDarling пакета MissMech со слегка неудобным интерфейсом — для ее использования выборки нужно склеить в одну и второй переменной указать вектор из длин выборок.

В scipy.stats в Python этот же метод представлен функцией anderson_ksamp (для нее также нужно соединить выборки в набор массивов через []).

4) Критерии хи-квадрат и отношения правдоподобия.

Однородность нескольких выборок можно проверять с помощью критерия хи-квадрат. Для этого положим Δ_1,...,Δ_m — разбиение области значений наших величин на множества, найдем $\nu_{i,j}$ — количество $X_{i,l}$, попавших в Δ_j, $\nu_{\cdot,j} = \sum_i \nu_{i,j}$, $\nu_{i,\cdot} = \sum_j \nu_{i,j}$. Тогда при выполнении основной гипотезы

$$\sum_{i=1}^{k}\sum_{j=1}^{m} \frac{\left(\nu_{i,j} - \frac{\nu_{i,\cdot}\nu_{\cdot,j}}{n}\right)^2}{\nu_{\cdot,j}\nu_{i,\cdot}/n} \xrightarrow{d} Y \sim \chi^2_{mk-k-m+1}.$$

Аналогично

$$2\sum_{i=1}^{k}\sum_{j=1}^{m} \ln\left(\frac{\nu_{i,j} n}{\nu_{i,\cdot}\nu_{\cdot,j}}\right) \xrightarrow{d} Y \sim \chi^2_{mk-k-m+1}.$$

Вопрос 2. Как вывести этот результат из общего критерия отношения правдоподобий?

Критерий хи-квадрат для проверки однородности двух выборок реализован в chisq.test. Первым аргументом ему задается так называемая таблица сопряженности, то есть матрица из $\nu_{i,j}$. Для дискретной выборки получить ее строки можно с помощью команды table(). В Python это же делает chi2_contingency.

В случае двух бернуллиевских выборок может быть также использован явный критерий Фишера fisher.test, которые не апеллирует к асимптотическим свойствам.

Замечание 1. Приведенные критерии асимптотические, при небольших выборках p-value в указанных функциях рассчитывается не из предельного распределения, а исходя из метода Монте-Карло.

Задача 2. Перед вами результаты медицинских исследований. Из 1500 мужчин, испытывающих лекарство, выздоровели 700, из 210 не принимавших выздоровели 80. Из 220 принимавших женщин — 150, из 680 не принимавших — 400. Проверить, влияет ли лекарство на мужчин? На женщин? На людей обоих полов?

6.1.2 Новые подходы

К непараметрическим критериям 1) и 2) из прошлого раздела добавим еще несколько:

1) Критерий перестановок (Permutation test).

Этот критерий олицетворяет общую идею, которая, в частности, может использоваться для проверки однородности. Рассмотрим какую-то асимметричную статистику $T(x_1, ..., x_n)$, которая, как мы считаем, должна быть большой при альтернативе и небольшой при гипотезе. Тогда мы можем рассмотреть критерий вида $\{T > c\}$, где c — какая-то константа, но не знаем, как найти фактический уровень значимости такого критерий.

Критерий предлагает нам представить, что среди возможных выборок есть только перестановки нашей выборки. Они равновероятны, поэтому вероятность $T > c$ есть

$$\frac{1}{n!}\sum_{\sigma\in S_n} I_{T(X_{\sigma(1)},...,X_{\sigma(n)})>c},$$

где S_n — множество всех перестановок. Значит, фактический уровень значимости на реализации $x_1, ..., x_n$ есть

$$\frac{1}{n!}\sum_{\sigma\in S_n} I_{T(x_{\sigma(1)},...,x_{\sigma(n)})>T(x_1,...,x_n)}.$$

Поскольку рассматривать все перестановки слишком трудоемко, достаточно выбрать некоторое количество N случайных перестановок и оценить фактический уровень значимости величиной

$$\frac{1}{N}\sum_{i=1}^{N} I_{T(x_{\sigma_i(1)},...,x_{\sigma_i(n)})>T(x_1,...,x_n)},$$

где σ_i — выбранные перестановки.

Для случая проверки однородности $X_1, ..., X_n$ и $Y_1, ..., Y_m$ в качестве тестовой статистики можно использовать, например, $|\overline{X} - \overline{Y}|$ или как в t-критерии

$$\frac{\overline{X}-\overline{Y}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\frac{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2}{n_1+n_2-2}}}.$$

Вопрос 3. Что можно сказать про однородность выборок $(X_1, X_2) = (1, 9)$,$Y_1 = 3$?

Метод достаточно прост и может быть легко запрограммирован, а может быть взят, например, permTS из пакета perm в R.

Задача 3. Массив Salary из пакета wPerm содержит информацию о зарплатах сотрудников частных или публичных университетов. Проверить гипотезу о том, что эти зарплаты одинаковы.

Рангом x_1 в наборе $x_1, ..., x_n$ назовем номером x_1 по возрастанию среди x_i.

2) Ранговый критерий Манна-Уитни-Уилкоксона.

Этот критерий используется для проверки гипотезы однородности против альтернативы доминирования H_2.

Для величин $X_1, ..., X_n$, $Y_1, ..., Y_m$ найдем ранги $R_1, ..., R_m$ величин $Y_{(1)}$,...,$Y_{(m)}$ в общем вариационном ряду. На их основе подсчитаем $V = R_1 + ... + R_m$. При выполнении гипотезы эта статистика распределена также как $Z_1 + ... + Z_m$, где Z_i — выборка без возвращения из чисел $\{1, ..., n + m\}$. При больших n, m

верна сходимость

$$\frac{V-(n+m+1)m/2}{\sqrt{nm(n+m+1)/12}} \xrightarrow{d} Z \sim \mathcal{N}(0,1),$$

при малых параметрах оценить фактический уровень значимости можно методом Монте-Карло. В R этот критерий задается wilcox.test, в Python mannwhitneyu из scipy.stats

Вопрос 4. Убедиться, что при выполнении гипотезы $\mathbf{E}V=(n+m+1)m/2$, $\mathbf{D}V=nm(n+m+1)/12$. Стоит отметить, что критерий Манна-Уитни по факту проверяет гипотезу $\mathbf{P}(X>Y)=1/2$. Вообще говоря, не принципиально, чтобы $X_1,...,X_n$ были из одного распределения, достаточно чтобы $\mathbf{P}(X_i>Y_j)$ все были смещены от $1/2$ в одну сторону.

3) Критерий Баумгартнера-Вейсса-Шиндлера.

Критерий используется для проверки гипотезы однородности с общей альтернативой и базируется на статистике

$$B=\frac{1}{2}(B_X+B_Y),\ B_X=\frac{1}{n}\sum_{i=1}^{n}\frac{\left(S_i-\frac{n+m}{n}i\right)^2}{\frac{i}{n+1}\left(1-\frac{i}{n+1}\right)\frac{m(n+m)}{n}},\ B_Y=\frac{1}{m}\sum_{i=1}^{m}\frac{\left(R_i-\frac{n+m}{m}i\right)^2}{\frac{i}{m+1}\left(1-\frac{i}{m+1}\right)\frac{n(n+m)}{m}},$$

где S — ранги X в общем вариационном ряду, а R — ранги Y, упорядоченные по возрастанию.

При гипотезе числители достаточно малы и статистика имеет предельное распределение. В случае неоднородности ранги одной из выборок будут сгруппированы не столь однородно и статистика будет принимать большие значения. При этом засчет знаменателя эта статистика представляет близка к статистике Андерсона-Дарлинга и внимательно реагирует на отличия "на хвостах".

Предельное распределение для теста достаточно громоздко. В R метод реализован в пакете BWStest в функции bws_test. Чтобы не писать тест руками, вы можете и в Python импортировать его через R. Например, в Python есть пакет r_bwstest. Можно, впрочем, просто запустить код через rpy2.

В отличие от критерия Манна-Уитни, критерий Баумгартнера способен отслеживать отличия не только сдвига, но и масштаба. Этот критерий достаточно новый и по-видимому является удачной альтернативой популярным критериям Стьюдента и Манна-Уитни.

Задача 4. Сравнить критерии Баумгартнера и Манна-Уитни на выборках размера а) 15 б) 100 в) 1000, где $X_i \sim \mathcal{N}(0,1)$ 1) $Y_i \sim \mathcal{N}(0,A)$; 2) $Y_i \sim \mathcal{N}(A,2)$; 3) $X_i \sim Laplace(0,\sqrt{2})$.

6.2 Многократные повторные выборки

В ряде случаев мы сталкиваемся с зависимостью следующего характера — в выборках $X_{1,1},...,X_{1,n}$,...,$X_{k,1},...,X_{k,n}$ величины $X_{i,j}$ зависимы между собой при разных i. Типичным примером этого может быть исследования каких-либо объектов в изменяющихся условиях. Например, мы изучали параметры цветов на грядке до применения удобрений и после. При этом наблюдения до и после зависимы, поскольку растения одни и те же.

Итак, многократными повторными выборками мы будем называть $X_{i,j}$, $i\le k$, $j\le n$, т.ч. векторы $(X_{1,j},...,X_{k,j})$ независимы и одинаково распределены.

Рассмотрим случай парных повторных выборок ($k=2$). Положим $Z_j=X_{2,j}-X_{1,j}$ и предположим, что Z_j имеют непрерывные распределения, причем $Z_j=\theta+\varepsilon_j$, где $P(\varepsilon_j<0)=P(\varepsilon_j>0)=1/2$. Гипотеза H_0 будет заключаться в том, что $\theta=0$, альтернатива $H_1:\theta\neq 0$. Тогда используют следующие критерии:

1) Критерий знаков.

Рассмотрим S — число $Z_i>0$. Тогда при гипотезе S будет иметь биномиальное n, 0.5 распределение, а при альтернативе биномиальное с другим параметром p. Эту гипотезу можно проверить асимптотически при $n\to\infty$, а можно в явном виде.

2) Критерий знаковых рангов Уилкоксона.

При дополнительном предположении симметричности распределений Z_j, справедлив критерий Уилкоксона, основывающийся на статистике

$$T = R_1U_1 + ...R_nU_n,$$

где $U_i = I_{Z_i>0}$, R_i — ранг $|Z|_i$ в ряду $|Z|_1,...,|Z|_n$. При выполнении гипотезы $(T-ET)/\sqrt{DT}$ сходится к нормальному распределению.
В R критерий Уилкоксона задан все той же функцией wilcox.test, но с указание paired = TRUE. В Python он задается wilcoxon().

Задача 5. Проверить гипотезу для Z_i а) $Z_i \sim \mathcal{N}(0,1)$, б) $Z_i \sim \mathcal{N}(0.5,3)$, в) $Z_i+1 \sim \exp(1)$.
3) Критерий Фридмана.
В случае, если выборки многократные повторные (то есть при тех же условиях мы рассматриваем k выборок), используется критерий Фридмана. Мы предполагаем, что $X_{i,j} = \mu + a_i + b_j + \varepsilon_{i,j}$, где a_i эффект строки, b_j — столбца. При этом мы предполагаем, что распределения ошибок $\varepsilon_{i,j}$ одинаковы и непрерывны. Основная гипотеза заключается в том, что все b_j одинаковы. Для проверки гипотезы строится $Q_{i,j}$ — ранг $X_{i,j}$ в своей строке, $T_j = \sum_{i=1}^n Q_{i,j}/n$. При верной гипотезе все статистики T_j близки, а значит $T_j - T_{.,.}$) близко к нулю, где $T_{.,.} = (k+1)/2$ — средний ранг. Критерий предлагает использовать статистику

$$F = \frac{12}{k(k+1)} \sum_{j=1}^{k} (T_j - T_{.,.})^2$$

и сравнивать ее с квантилью χ^2_{k-1} уровня $1-\alpha$.
В R этот критерий задан friedman.test, в Python критерий Фридмана задается friedmanchisquare().

Если мы обнаружили, что выборки неоднородны, то нам нужен дополнительный анализ, чтобы разделить их друг от друга. Можно провести последующий анализ попарных сравнений выборок (такой анализ называют post hoc). В R это делают с помощью posthoc.friedman.nemenyi.test пакета PMCMR. В Python большой набор post hoc критериев предложен в scikit-posthocs.

6.3 Нормальные выборки

1) Критерии Фишера и Стьюдента.
Классический подход предлагал сравнивать однородность нормальных выборок путем применения
а) критерия Фишера (F-критерия) для проверки равенства дисперсий, основанном на том, что статистика

$$\frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2/(n_1-1)}{\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2/(n_2-1)}$$

при равенстве дисперсий имеет распределение Фишера;
б) критерия Стьюдента или t-критерия, основывающемся на статистике

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{i=1}^{n_2}(Y_i - \overline{Y})^2}{n_1+n_2-2}}},$$

которая при равенстве средних и дисперсий имеет распределение Стьюдента. Существенным недостатком классического подхода заключается в том, что критерий Стьюдента значимо опирается на равенство дисперсий, в котором мы не можем быть уверенным наверняка.
В R критерии Стьюдента и Фишера заданы функциями t.test и var.test. В Python ttest_ind
Можно убрать шаг проверки равенства дисперсий, используя более сложные подходы, в частности, подход Уэлча (Welch). Этот критерий рассматривает статистику

$$\frac{\overline{X} - \overline{Y}}{\sqrt{\sum_{i=1}^{n_1}(X_i - \overline{X})^2/n_1 + \sum_{i=1}^{n_2}(Y_i - \overline{Y})^2/n_2}}$$

и приближенно находит ее распределение. По умолчанию t.test использует именно критерий Уэлча для случая различных дисперсий.
В целом проблема сравнения средних нормальных выборок без предположения равных дисперсий (она называется проблемой Беренса-Фишера) достаточно сложна, и точный критерий для нее построить не удается.

Задача 6. Испытать t.test для сравнения средних а) $\mathcal{N}(0,1)$ и $\mathcal{N}(0,3)$, б) $\mathcal{N}(0,1)$ и $T(0,1,2,3,0.01)$, в) $\mathcal{N}(0,1)$ и $T(0,1,2,3,0.05)$, г) $T(0,1,2,3,0.05)$ и $T(0,1,2,3,0.05)$ д) t_7 и t_7, где t — критерий Стьюдента, на выборках размера 100, 500, 1000.
Здесь $T(a_1,\sigma_1,a_2,\sigma_2,p)$ — смесь $\mathcal{N}(a_1,\sigma_1^2)$ и $\mathcal{N}(a_2,\sigma_2^2)$, где первое распределение берется с вероятностью $1-p$, а второе — p.

2) Дисперсионный анализ ANOVA.

Этот метод обобщает пункт 1) на случай k выборок. В предположении равенства дисперсий, положим $\overline{X}_{\cdot,j}$ — среднее по выборке j, $\overline{X}_{\cdot,\cdot}$ — среднее по всем выборкам

$$SSTR=\sum_{j=1}^{k} n_j(\overline{X}_{\cdot,j}-\overline{X}_{\cdot,\cdot})^2,\ SSE=\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{i,j}-\overline{X}_{\cdot,j})^2,.$$

При этом $SSE\sim\chi^2_{n_1+\ldots+n_k-k}$ из леммы Фишера, при выполнении гипотезы $SSTR\sim\chi^2_{k-1}$ и эти величины независимы. Значит если $MSTR=SSTR/(k-1)$, $MSE=SSE/(n-k)$, то $F=MSTR/MSE$ будет иметь распределение Фишера с $k-1,n-k$ степенями свободы при выполнении гипотезы и будет принимать большие значения иначе. Так называемые ANOVA table, иллюстрирующие данные анализа ANOVA, имеют вид

$$\begin{array}{cccc} SSTR & MSTR & F & p-value \\ SSE & MSE & & \end{array}.$$

В R получить такую таблицу можно следующий образом. Запишем все выборки в один вектор X подряд. В другой вектор I на место i запишем номер выборки, к которой принадлежит X_i. Тогда $anova(lm(X\sim I))$ построит таблицу ANOVA.
В Python используется f_oneway.

Задача 7. Сгенерировать 5 выборок размера 100 из $\mathcal{N}(a_i,1)$ или $R[a_i-1/2,a_i+1/2]$ при а) $a_1=...=a_5$, б) $a_1=a_2=a_3=a_4=0$, $a_5=0.5$ в) $a_i=i/10$ и исследовать их с помощью ANOVA. Повторить эксперимент в случае, если у последней выборки дисперсия 2.
Полезно также использовать метод Тьюки, позволяющий оценить, какие из переменных значимо отличаются. Интерфейс этого метода в R $TukeyHSD(aov(X\sim I))$. Здесь I как и прежде вектор, идентифицирующий принадлежность элементов вектора X к различным выборкам. Для метода Тьюки он должен быть в формате factor, это можно сделать функцией factor(I).

Задача 8. Применить метод Тьюки к массиву данных warpbreaks для того, чтобы понять, какие уровни натяжения (tension) значимо отличаются с точки зрения количества разрывов волокна (breaks).
В однофакторной модели, рассмотренной нами в части 2) предыдущего раздела, можно использовать и ранговые критерии. Если $X_{i,j}=\mu_i+\varepsilon_{i,j}$, $i\le k$, $j\le n_i$, где $\varepsilon_{i,j}$ имеют одинаковое непрерывное (но уже необязательно нормальное) распределение, то для проверки гипотезы $\mu_1=...\mu_k$ можно использовать критерий Краскелла-Уоллеса:

1) Критерий Краскелла-Уоллеса. Положим $R_{i,j}$ — ранг $X_{i,j}$ в общем вариационном ряду. Тогда пусть $S_i=\sum_{j=1}^{n_i}R_{i,j}$, $R_{i,\cdot}=S_i/n_i$, $R_{\cdot,\cdot}=\sum R_{i,j}/N$, $N=n_1+...+n_k$. Критерий Краскелла-Уоллеса основан на статистике

$$W=\frac{12}{N(N+1)}\sum_{i=1}^{k} n_i(R_{i,\cdot}-R_{\cdot,\cdot})^2,$$

сходящейся к χ^2_{k-1} при $n_i\to\infty$.
В R он задается kruskall.test, в Python kruskal.

2) Критерий Джонкхиера-Терпстры.

Для альтернативы $\mu_1 \leq \mu_2 .. \leq \mu_k$ используют критерий, задаваемый статистикой

$$\sum_{r<s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} I_{X_{i,r}<X_{j,s}}.$$

В R этот критерий задан JonckheereTerpstraTest из пакета DescTools.

6.4 Ответы на вопросы

Ответ 1. Если $F_X(x) \geq F_Y(x)$, то $F_X^{-1}(x) \leq F_Y^{-1}(x)$. Тогда $X_1 = F_X^{-1}(R) \leq F_Y^{-1}(R) = Y_1$, где $R \sim R[0,1]$, что в силу метода обратной функции и дает требуемое.

Ответ 2. Правдоподобие для нашей задачи имеет вид

$$L = \prod_{i=1}^{k} \prod_{j=1}^{m} p_{i,j}^{\nu_{i,j}},$$

где $\nu_{i,j}$ — количество наблюдений вида j в i-й выборке, $p_{i,j}$ — соответствующая вероятность, $\sum_j p_{i,j} = 1$. Наше параметрическое семейство в общем случае имеет размерность $k(m-1)$, поскольку последняя вероятность $p_{i,m}$ определяется остальными при каждом i. ОМП, как мы убедились во втором ссеминаре, имеет вид

$$\hat{p}_{i,j} = \nu_{i,j}/\nu_{i,\cdot},\ \nu_{i,\cdot} = \sum_{j=1}^{m} \nu_{i,j}.$$

При гипотезе параметров $m-1$, правдоподобие имеет вид

$$L = \prod_{j=1}^{m} p_{i,j}^{\nu_{\cdot,j}},\ \nu_{\cdot,j} = \sum_{i=1}^{k} \nu_{i,j},$$

а ОМП $\hat{p}_{i,j} = \nu_{i,\cdot}/n$. Остается поделить полученные правдоподобия и применить критерий отношения правдоподобий.

Ответ 3. Величина $|\overline{X} - \overline{Y}|$ при данной выборке (и перестановке, меняющей местами 1 и 9) принимает значение 1.5, а при остальных перестановках 5, 5, 7 и 7. Следовательно, фактический уровень значимости 1/3.

Ответ 4. В силу представления $V = Z_1 + ... + Z_m$, имеем $\mathbf{E}V = \mathbf{E}Z_1 + ... + \mathbf{E}Z_m = m(n+m+1)/2$. Математическое ожидание каждой из Z_i очевидно равно $(n+m+1)/2$, поскольку Z_i равновероятно принимает значения от 1 до $n+m$.
С дисперсией ситуация обстоит чуть сложнее, поскольку $\mathbf{D}V = \mathbf{D}Z_1 + ... + \mathbf{D}Z_m + 2\sum_{i>j} cov(Z_i, Z_j) = m\mathbf{D}Z_1 + m(m-1)cov(Z_1, Z_2)$. При этом

$$\mathbf{E}Z_1^2 = \frac{1}{n+m} \sum_{i=1}^{n+m} i^2 = \frac{(n+m)(n+m+1)(2n+2m+1)}{6(n+m)} = \frac{(n+m+1)(2n+2m+1)}{6},$$

$$\mathbf{E}Z_1 Z_2 = \frac{1}{n+m} \frac{1}{n+m-1} \sum_{i \neq j} ij = \frac{1}{(n+m)(n+m-1)} \left(\sum_{i=1}^{n+m} i \sum_{j=1}^{n+m} j - \sum_{i=1}^{n+m} i^2 \right) = \frac{((n+m)(n+m+1))^2/4 - (n+m)(n+m+1)(2n+2m+1)/6}{(n+m)(n+m-1)}.$$

Итого

$$\frac{m(n+m+1)(2n+2m+1)}{6}+\frac{m(m-1)(n+m+1)(3(n+m)(n+m+1)-2(2n+2m+1))}{12(n+m-1)}-\frac{m^2(n+m+1)^2}{4}=\frac{mn(n+m+1)(2n+2m+1)}{6(n+m-1)}+\frac{m(n+m+1)^2((n+m)(m-1)-(n+m-1)m)}{4(n+m-1)}=\frac{nm(n+m+1)}{12}.$$