第八章 假设检验

§1 假设检验

§2 正态总体均值和方差的假设检验

§3 假设检验中的p值法

§4 分布假设检验

§5 多重假设检验

回顾：

参数的假设检验的一般步骤(临界值法):

1. 根据实际问题建立原假设H_0和备择假设H_1；
2. 确定检验统计量Z，并明确其分布；
3. 给定显著性水平$\alpha > 0$，根据统计量的分布，由$P\{|Z| \geq z_{\alpha/2}\} \leq \alpha$确定拒绝域（临界值$z_{\alpha/2}$）；
4. 由样本值具体计算统计量Z的观察值z，并做出判断，若$|Z| \geq z_{\alpha/2}$，则拒绝H_0，接受H_1；若$|Z| < z_{\alpha/2}$，则接受H_0。

§3 假设检验中的p值法

假设检验方法$\begin{cases} \text{临界值法} \\ p\text{值法} \end{cases}$

例 1 设总体$X \sim N(\mu, \sigma^2)$， μ未知，σ^2=100, 现有样本$x_1, x_2, \dots, x_{52}$，算得$\bar{x} = 62.75$。现在检验假设

$$H_0: \mu \le \mu_0 = 60, \qquad H_1: \mu > 60$$

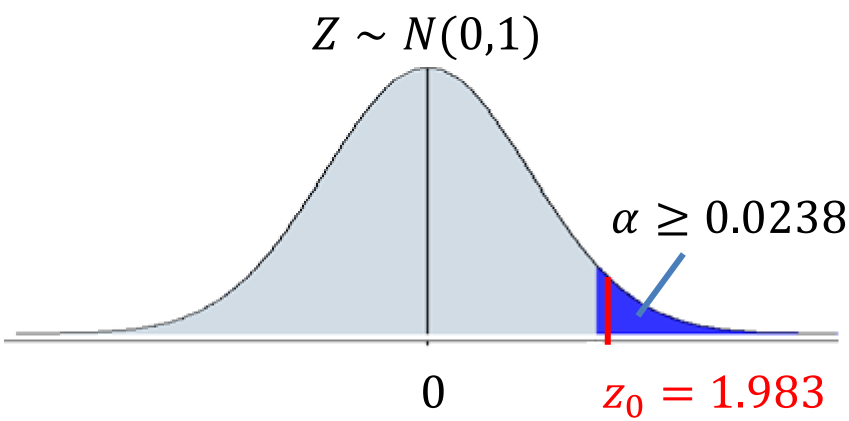
采用Z检验法，检验统计量为$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$。以样本值代入，得到$Z$的观察值为$z_0 = 1.983$。

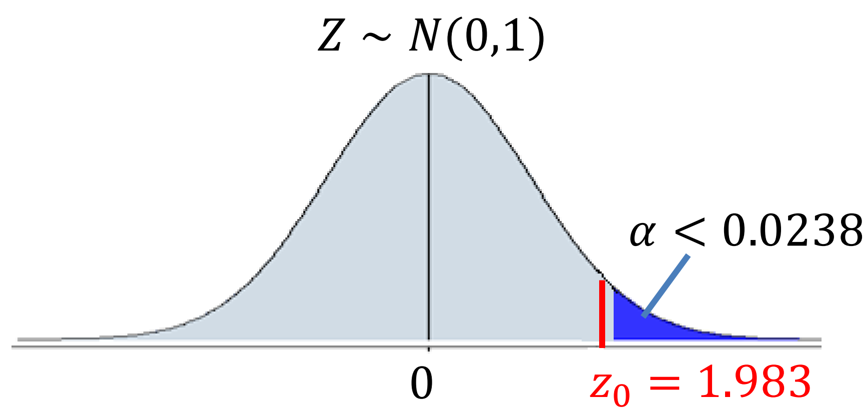
例 1（续）

$P\{Z \geq z_0\} = P\{Z \geq 1.983\} = 1 - \Phi(1.983) = 0.0238$

此概率为Z检验法的右边检验的p值，记为

$$P\{Z \geq z_0\} = p\text{值}\ (=0.0238)$$





若$\alpha \geq p$，对应临界值$z_\alpha \leq 1.983$，观察值z_0落在拒绝域内，因而拒绝H_0；若$\alpha < p$，则$z_\alpha > 1.983$，观察值z_0落在拒绝域外，因而接受H_0

$p = 0.0238$是H_0被拒绝的最小显著性水平

定义 假设检验问题的p值(probability value)是由检验统计量的样本观察值得出的原假设可被拒绝的最小显著性水平。

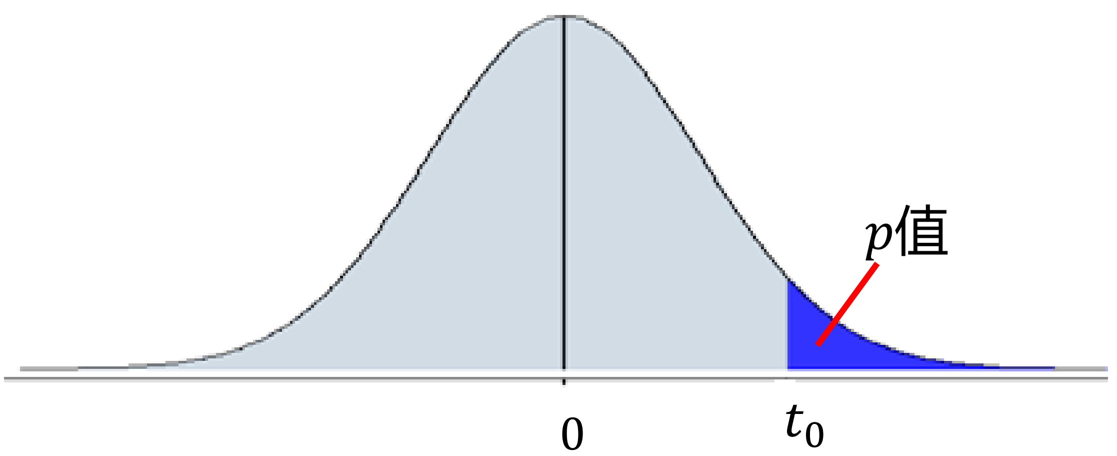
任何检验问题的p值可以根据检验统计量的样本观察值以及检验统计量在H_0下的一个特定参数值对应的分布求出。

例如 正态总体均值检验中，当σ^2未知时，可采用检验统计量$t=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$，$t\sim t(n-1)$。

如果由样本求得统计量t的观察值为t_0，那么以下检验问题的p值分析如下：

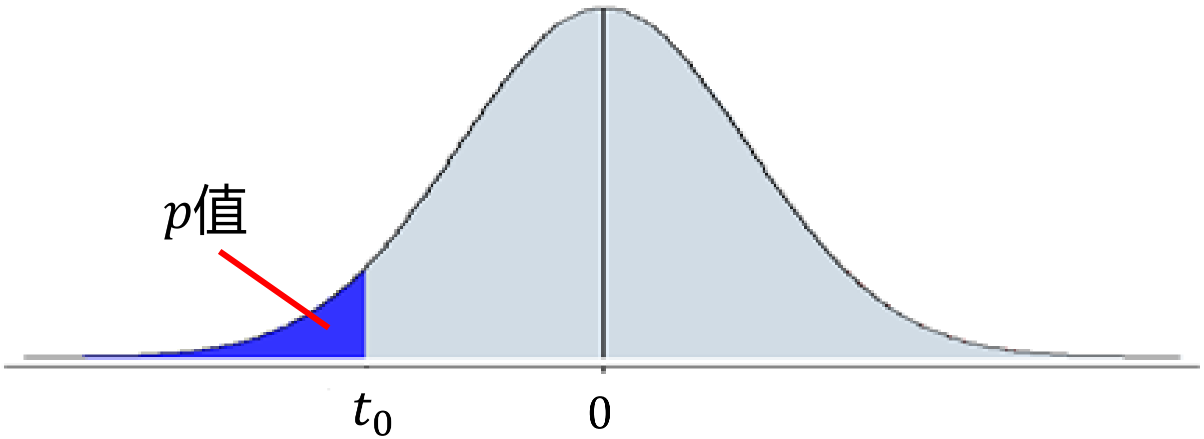
(1) $H_0:\mu\le\mu_0,\quad H_1:\mu>\mu_0$

$$p\text{值}=P_{\mu_0}\{t\ge t_0\}=t_0\text{右侧尾部面积}$$



(2) $H_0: \mu \geq \mu_0$, $H_1: \mu < \mu_0$

p值 $= P_{\mu_0}\{t \leq t_0\} =$ t_0左侧尾部面积

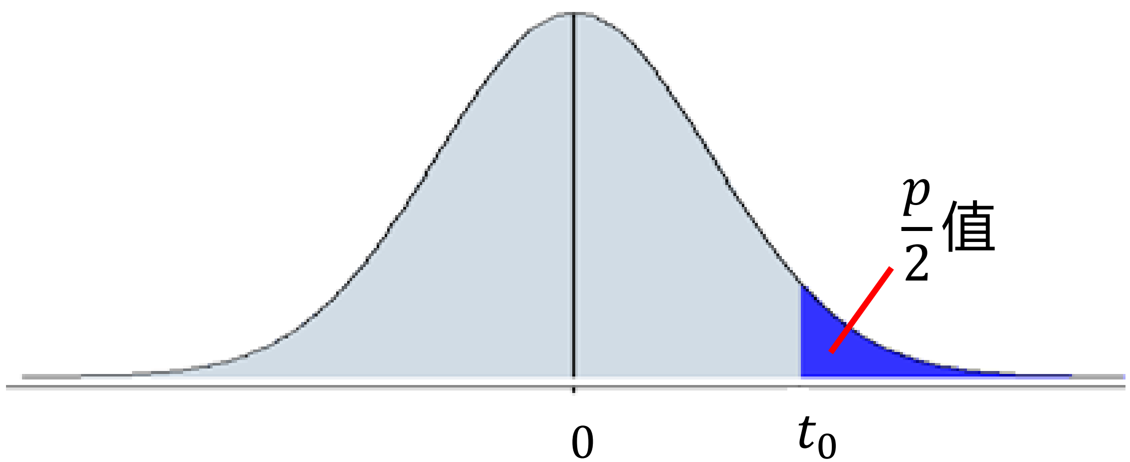


(3) $H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0$

(a) 当$t_0 > 0$时

p值 $= P_{\mu_0}\{|t| \geq t_0\} = P_{\mu_0}\{(t \leq -t_0) \cup (t \geq t_0)\}$

$= 2 \times$ (t_0右侧尾部面积)

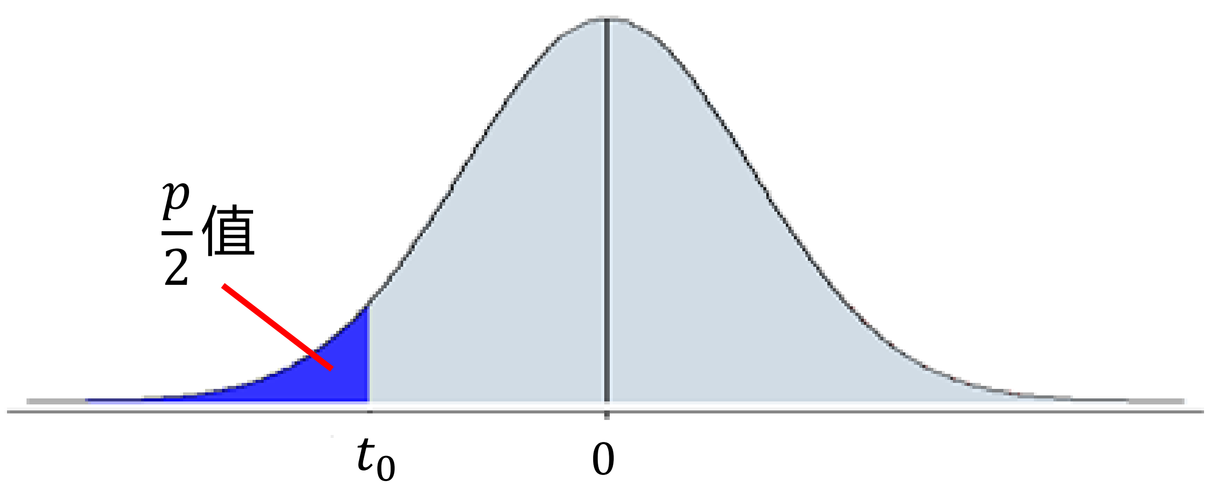


(3) $H_0: \mu = \mu_0, \quad H_1: \mu \neq \mu_0$

(b) 当$t_0 < 0$时

p值 $= P_{\mu_0}\{|t| \geq -t_0\} = P_{\mu_0}\{(t \leq t_0) \cup (t \geq -t_0)\}$

$= 2 \times (t_0$左侧尾部面积$)$



按照p值的定义，对于任意指定的显著性水平α，有

(1) 若p值 $\leq \alpha$, 则在显著性水平α下拒绝H_0

(2) 若p值 $> \alpha$, 则在显著性水平α下接受H_0

这种利用p值来确定是否拒绝H_0的方法称为p值法。

注意：

(1) 与临界值法相比，p值法给出了有关拒绝域的更多的信息。

(2) p值表示反对原假设H_0的依据的强度，p值越小，反对H_0的依据越强、越充分。

例 2 天然牛奶的冰点温度近似服从正态分布，均值$\mu_0 = -0.545$℃，标准差$\sigma = 0.008$℃。牛奶掺水可以使冰点温度升高至水的冰点温度（ 0℃ ）。测得生产商提交的 5 批牛奶的冰点温度，其均值为$\bar{x} = -0.535$℃，问是否可以认为生产商在牛奶中掺了水？

用p值法检验：

$$H_0: \mu \le \mu_0(\text{未掺水})，H_1: \mu > \mu_0(\text{已掺水})$$

检验统计量$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$的观察值

$$z_0 = \frac{-0.535-(-0.545)}{0.008/\sqrt{5}} = 2.795$$

p值 $= P_{\mu_0}\{Z \ge 2.795\} = 1 - \Phi(2.795) = 0.0026$

p值$<\alpha = 0.05$，故拒绝H_0。

§4 分布假设检验

参数检验：总体分布类型已知，对其中的未知参数进行检验

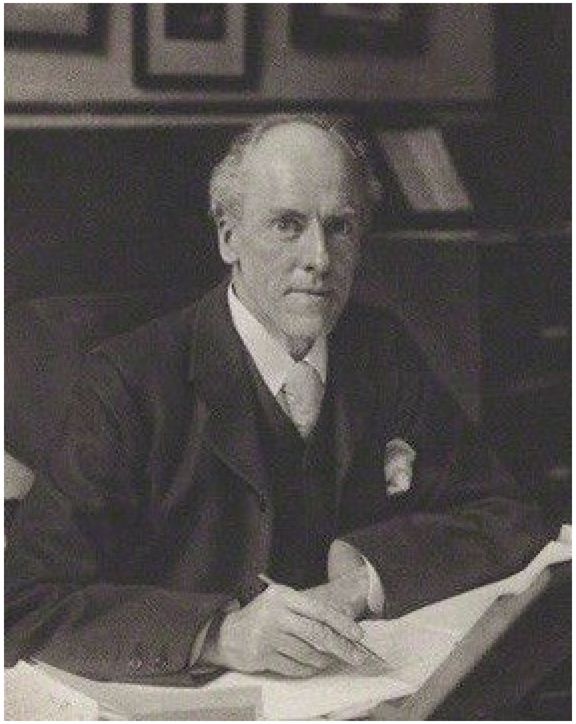
非参数检验：根据总体的样本对其分布进行推断和检验。

☆单个分布的χ^2拟合检验

☆分布族的χ^2拟合检验

☆置换检验

Karl Pearson (1857-1936)



引例

(1) 1500年到1931年的432年间，每年爆发战争的次数可以看作一个随机变量X，据统计，这432年间共爆发299次战争。根据经验，可以假设X服从泊松分布。

战争次数X	0	1	2	3	4
发生X次战争的年数	223	142	48	15	4

(2) 检验某工厂制造的色子是否均匀。

分布拟合检验—— 检验总体是否具有某一个指定的分布或属于某一个分布族。

(1) 单个分布的χ^2拟合检验

设总体X的分布未知，$x_1, x_2, \ldots, x_n$是来自X的样本值。检验假设：

$$H_0: \text{总体}X\text{的分布函数为}F(x)$$

$$H_1: \text{总体}X\text{的分布函数不是}F(x)$$

其中$F(x)$设不含未知参数。

定义检验统计量：

$$\sum_{i=1}^{k} C_i \left(\frac{f_i}{n} - p_i \right)^2$$

- 设总体 X 的取值范围分成互不相交的子集：$A_1, A_2, \ldots, A_k$
- 样本的观察值为: $x_1, x_2, \ldots, x_n$
- f_i 为样本落在子集A_i中的个数

$$p_i = P(A_i)$$

- C_i 为给定的常数

上述统计量度量了样本与H_0中假设分布的吻合程度。

定理 若n充分大（$n \geq 50$），则当H_0为真时统计量

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{f_i^2}{np_i} - n$$

近似服从$\chi^2(k-1)$分布。（证略）

如果样本观察值使得

$$\chi^2 \geq \chi_\alpha^2(k-1)$$

则在显著性水平α下拒绝H_0，否则接受H_0。这就是所谓的χ^2拟合检验法。

注意：

(1) $n \geq 50$; $np_i \geq 5$，否则适当合并A_i；

(2) 待检验的可以是分布、分布率或分布密度

- 如果总体分布为离散型，原假设的形式 为 H_0: 总体X的分布律为$P\{X = x_i\} = p_i, i = 1, 2, \ldots$
- 如果总体分布为连续型，原假设的形式为 H_0: 总体X的概率密度函数为$f(x)$

例 3 下表列出了某一地区在夏季的一个月中由100个气象站报告的雷暴雨的次数

i	0	1	2	3	4	5	≥6
f_i	22	37	20	13	6	2	0
A_i	A_0	A_1	A_2	A_3	A_4	A_5	A_6

其中f_i是报告雷暴雨次数为i的气象站数。试用χ^2拟合检验法检验雷暴雨的次数X是否服从均值$\lambda = 1$的泊松分布（取$\alpha = 0.05$）。

解：$H_0\colon P\{X=i\}=\dfrac{\lambda^i e^{-\lambda}}{i!}=\dfrac{e^{-1}}{i!}, i=0,1,\ldots$

X所有可能的取值为$\Omega=\{0,1,2,\ldots\}$，将Ω分成如表所示的两两不相交的子集$A_0,A_1,\ldots,A_6$，则有

<table>
<tr><th>A_i</th><th>f_i</th><th>p_i</th><th>np_i</th><th>$f_i^2/(np_i)$</th></tr>
<tr><td>$A_0\colon\{X=0\}$</td><td>22</td><td>e^{-1}</td><td>36.788</td><td>13.16</td></tr>
<tr><td>$A_1\colon\{X=1\}$</td><td>37</td><td>e^{-1}</td><td>36.788</td><td>37.21</td></tr>
<tr><td>$A_2\colon\{X=2\}$</td><td>20</td><td>$e^{-1}/2$</td><td>18.394</td><td>21.75</td></tr>
<tr><td>$A_3\colon\{X=3\}$</td><td>13</td><td>$e^{-1}/6$</td><td rowspan="4">$\left.\begin{matrix}6.131\\1.533\\0.307\\0.059\end{matrix}\right\}8.03$</td><td rowspan="4">54.92</td></tr>
<tr><td>$A_4\colon\{X=4\}$</td><td>6</td><td>$e^{-1}/24$</td></tr>
<tr><td>$A_5\colon\{X=5\}$</td><td>2</td><td>$e^{-1}/120$</td></tr>
<tr><td>$A_6\colon\{X\geq 6\}$</td><td>0</td><td>$1-\sum_{i=0}^{5}p_i$</td></tr>
</table>

$\sum=127.04$

解（续）：

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{f_i^2}{np_i} - n \sim \chi^2(3)$$

$k = 4, n = 100$

$\chi^2_{0.05}(k-1) = \chi^2_{0.05}(3) = 7.815$

由统计量的观察值得：

$$\chi^2 = \sum_{i=1}^{k} \frac{f_i^2}{np_i} - n = 127.04 - 100 = 27.04 > 7.815$$

所以，在显著性水平$\alpha = 0.05$下拒绝H_0，认为样本不是来自均值$\lambda = 1$的泊松分布。

例 4（孟德尔豌豆杂交试验）用圆润黄色豌豆与皱皮绿色豌豆杂交，得到后代的可能有4种类型：圆润黄色、皱皮黄色、皱皮绿色、圆润绿色。四种类型服从多项式分布，由孟德尔遗传理论预测的参数为：

$$p_0=\left(\frac{9}{16},\frac{3}{16},\frac{3}{16},\frac{1}{16}\right)$$

现有556个豌豆样本，四种类型的数量分别为

$$X=(315,101,108,32)$$

试分析该数据是否与孟德尔的理论相符。

解： $H_0: p = p_0 = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$

检验统计量：

$$\chi^2 = \sum_{i=1}^{4} \frac{(f_i - np_i)^2}{np_i} = \frac{(315-312.75)^2}{312.75} + \frac{(101-104.25)^2}{104.25} + \frac{(108-104.25)^2}{104.25} + \frac{(32-34.75)^2}{34.75} = 0.47$$

查表得$\chi^2_{0.05}(3) = 7.815$，$\chi^2 < \chi^2_\alpha(k-1)$，所以在显著性水平$\alpha = 0.05$下接受$H_0$，认为该数据与孟德尔的理论相符。

小结：

1. χ^2拟合检验法的基本步骤：
   1) 将总体X的取值范围分成k个互不重叠的小区间，记为$A_1, A_2, \dots, A_k$
   2) 落在第i个小区间的样本值个数记作n_i (实测频数，$f_1 + f_2 + \cdots + f_k = n$)
   3) 根据所假设的分布，算出总体X的值落在每个A_i的概率p_i（则np_i为样本的理论频数）
   4) 引进检验统计量$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - np_i)^2}{np_i}$，当充分大时 ($n \geq 50$)，$\chi^2 \sim \chi^2(k-1)$
   5) 根据样本值计算统计量是否落入拒绝域 $\chi^2 > \chi^2_\alpha(k-1)$

小结：

1. χ^2拟合检验法的基本步骤：
2. χ^2拟合检验法是在总体X的分布未知时，根据来自总体的样本检验关于总体分布的假设的一种试验方法
3. 一般地 ，我们可以根据样本观察值用直方图和经验分布函数推断出总体可能服从的分布，然后做检验

(2) 分布族的χ^2拟合检验

H_0: 总体X的分布函数是$F(x;\theta_1,\theta_2,\dots,\theta_r)$

其中， F的形式已知，参数$\Theta=(\theta_1,\theta_2,\dots,\theta_r)$未知，$F(x;\theta_1,\theta_2,\dots,\theta_r)$ 代表一族分布。

定义检验统计量：

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-n\widehat{p_i})^2}{n\widehat{p_i}}$$

总体 X 的取值范围分成k（ $k\geq r+1$ ）个互不相交的子集：$A_1,A_2,\dots,A_k$

$p_i=p_i(\theta_1,\theta_2,\dots,\theta_r)=P(A_i)$, $\widehat{p_i}$是由样本求出的估计值。

可以证明，在某些条件下，当H_0为真时近似地有

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-n\widehat{p_i})^2}{n\widehat{p_i}}\sim\chi^2(k-r-1)$$

类似地，分布族的假设检验问题的拒绝域为

$$\chi^2\geq\chi_\alpha^2(k-r-1)$$

这就是检验分布族的χ^2拟合检验法。

例 5 自1965年1月1日至1971年2月9日共2231天中，全世界里氏4级和4级以上地震总计162次，统计如下：

间隔天数	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	≥40
频数	50	31	26	17	10	8	6	6	8

试检验相继两次地震间隔的天数X服从指数分布（$\alpha = 0.05$）

解：需检验假设H_0: X的概率密度为

$$f(x)=\begin{cases}\dfrac{e^{-\frac{x}{\theta}}}{\theta}, x>0\\ 0, \quad\quad x\leq 0\end{cases}$$

其中参数θ未知。

由最大似然估计法求得θ的估计值为：

$$\hat{\theta}=\bar{x}=\frac{2231}{162}=13.77$$

当H_0成立时，X可能取值的全体为$[0,+\infty)$。将此区间分成9个互不重叠的小区间：

$A_1=[0,4.5], A_2=(4.5,9.5], \ldots, A_9=(39.5,+\infty)$

解(续)：若H_0为真，则X的分布函数的估计为：

$$\hat{F}(x)=\begin{cases}1-e^{-\frac{x}{13.77}}, & x>0\\ 0, & x\le 0\end{cases}$$

由此可得p_i的估计：

$$\widehat{p_i}=\hat{P}(A_i)=\widehat{p_i}\{a_i<X\le a_{i+1}\}=\hat{F}(a_{i+1})\text{-}\hat{F}(a_i)$$

A_i	f_i	$\hat{p}_i$	$n\hat{p}_i$	$f_i^2/(n\hat{p}_i)$
$A_0: 0 \le x \le 4.5$	50	0.2788	45.1656	55.3519
$A_1: 4.5 < x \le 9.5$	31	0.2196	35.5752	27.0132
$A_2: 9.5 < x \le 14.5$	26	0.1527	24.7374	27.3270
$A_3: 14.5 < x \le 19.5$	17	0.1062	17.2044	16.7980
$A_4: 19.5 < x \le 24.5$	10	0.0739	11.9718	8.3530
$A_5: 24.5 < x \le 29.5$	8	0.0514	8.3268	7.6860
$A_6: 29.5 < x \le 34.5$	6	0.0358	5.7996	6.2073
$A_7: 34.5 < x \le 39.5$	6	0.0248	$\left.\begin{matrix} 4.0176 \\ 9.2016 \end{matrix}\right\} 13.2192$	14.8269
$A_8: 39.5 < x < \infty$	8	0.0568		

解(续)：

将计算结果列成表格，合并第8和第9区间，并计算：

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-n\hat{p}_i)^2}{n\hat{p}_i}=163.5633-162=1.5633$$

$$\chi^2_{0.05}(k-r-1)=\chi^2_{0.05}(6)=12.592>1.5633$$

所以，在显著性水平0.05下接受H_0，认为X服从指数分布。

注意：

本题答案是"接受H_0，认为X服从指数分布"，即认为$X\sim E(\theta)$。不能说"X服从参数$\theta=13.77$的指数分布"

(3) 置换检验

由R. A. Fisher在1930年提出

问题：$X_1, \dots, X_m \sim F_X$和$Y_1, \dots, Y_n \sim F_Y$为独立样本，检验假设$H_0: F_X = F_Y$，$H_1: F_X \neq F_Y$

置换检验：一种重采样技术，用于上述假设检验问题。

置换样本：对原始样本$Z = (X_1, \dots, X_m, Y_1, \dots, Y_n)$进行$N = m + n$次无放回采样，得到$N$个样本，称为一个置换样本。

- 共有$N!$个置换样本

令$T(X_1,\ldots,X_m,Y_1,\ldots,Y_n)$为检验统计量，每个置换样本对应一个统计量的值

$$T_i=T(X_1^i,\ldots,X_m^i,Y_1^i,\ldots,Y_n^i)$$

共有$N!$个检验统计量：$T_1,\ldots,T_{N!}$

令t_o为检验统计量T的观察值。假设当T很大时拒绝原假设，则假设检验问题的p值为：

$$p\text{值}=P\{T>t_o\}=\frac{1}{N!}\sum_{i=1}^{N!}I(T_i>t_o)$$

当H_0成立时，$F_X=F_Y$，X与Y的地位相等，所以置换样本的分布为$Uniform(1,N!)$

例 6：假设数据为$(X_1, X_2, Y_1) = (1,9,3)$,

令$T(X_1, X_2, Y_1) = |\bar{X} - \bar{Y}| = 2$，则置换为：

置换	T值	概率
(1, 9, 3)	2	1/6
(9, 1, 3)	2	1/6
(1, 3, 9)	7	1/6
(3, 1, 9)	7	1/6
(3, 9, 1)	5	1/6
(9, 3, 1)	5	1/6

则p值 $= P\{T > 2\} = \frac{4}{6}$

当N较大时，通常不用所有的$N!$个置换样本，而是从置换样本集合中随机选择一部分样本，来近似计算p值。

置换检验的一般步骤：

1. 计算检验统计量T的观测值
$$t_o = T(X_1, \dots, X_m, Y_1, \dots, Y_n)$$
2. 随机置换数据，并计算置换样本的统计量，重复B次，得到$T_1, \dots, T_B$
3. p值可近似计算为：
$$\frac{1}{B}\sum_{j=1}^{B} I(T_i > t_o)$$

例 7：DNA阵列可用来度量基因的表达水平。数据包括每个基因的 mRNA的水平（mRNA可以衡量该基因产生蛋白质的数量）。粗略地讲，该数值越大，基因越活跃。下表[来自Efron et al. JASA, 2001]显示10位携带两类肝癌细胞的病人的基因的表达水平，试验中共涉及2638个基因。

	类型1					类型2				
病人	1	2	3	4	5	6	7	8	9	10
基因1	230	-1350	-1580	-400	-760	970	110	-50	-100	-200
基因2	470	-850	-0.8	-280	120	390	-1730	-1360	-1	-330
…	…	…	…	…	…	…	…	…	…	……

检验：两组基因中基因1的中值是否有明显差异

解：

v_1：类型1中基因1的中值

v_2：类型2中基因1的中值

检验统计量：

$$T = |\hat{v}_1 - \hat{v}_2| = 710$$

通过仿真估计置换分布，计算得到：

$$p\text{值} = 0.045$$

如果取$\alpha = 0.05 > p$，则拒绝H_0，即认为两个中值有显著差异。

§5 多重假设检验

即便每个检验的显著性水平控制为α，多重检验中犯一次弃真错误的概率依然很高。

考虑m重假设检验：

$$H_{0i} \quad \text{vs.} \quad H_{1i}, \qquad i = 1, \dots, m$$

令$P_1, \dots, P_m$表示以上检验的m个p-值。

Bonferroni方法：

已知m重假设检验的p-值$P_1, \dots, P_m$，如果$P_i < \alpha/m$，则拒绝原假设H_{0i}。

定理：使用Bonferroni方法做多重假设检验，犯弃真错误的概率小于等于α。

证明：令$R=\{$至少犯一个弃真错误$\}$，

$R_i=\{$第i个假设检验犯了弃真错误$\}$

由于$P\left(\bigcup_{i=1}^{k} A_i\right) \le \sum_{i=1}^{k} P(A_i)$，所以

$$P(R)=P\left(\bigcup_{i=1}^{m} R_i\right) \le \sum_{i}^{m} P(R_i)=\sum_{i}^{m} \frac{\alpha}{m}=\alpha$$

Bonferroni方法是一种比较保守的方法。

更合理的想法是控制错误发现率（false discovery rate, FDR）

FDR: 是错误拒绝原假设的个数占所有被拒绝的原假设个数的比例的期望值。

- 1995年Benjamini和Hochberg首次提出了FDR的概念，并给出了在多重检验中对它的控制方法(BH方法)

在m重检验中，如果有m_0个正确的原假设，m_1个错误的原假设，那么假设检验可以按下表划分：

	接受H_0	拒绝H_0	总计
H_0正确	U	V	m_0
H_0错误	T	S	m_1
总计	$m-R$	R	m

错误发现比例（false discovery proportion, FDP）

$$FDP = \begin{cases} \frac{V}{R}, R > 0 \\ 0, R = 0 \end{cases}$$

$$FDR = E[FDP]$$

BH方法：

1. 令$P_1 < \cdots < P_m$表示排序后的p-值
2. 定义

$$l_i = \frac{i\alpha}{c_m m},\quad R = \max\{i: P_i < l_i\}$$

当p-值独立时c_m取值1，否则$c_m = \sum_{i=1}^{m} \frac{1}{i}$。

3. 令$t = P_R$，t称为BH拒绝阈值
4. 对于所有$P_i < t$，拒绝原假设H_{0i}

可以证明，BH方法能够将FDR控制在显著性水平范围内，即：$FDR = E[FDP] \leq \alpha$

假设检验小结

- 参数检验
  - 正态总体均值的检验
    - Z 检验
    - t 检验
  - 正态总体方差的检验
    - χ^2检验
    - F 检验
- 非参数检验
  - 单个分布的χ^2拟合检验
  - 分布族的χ^2拟合检验
  - 置换检验

- 临界值法
- p值法

假设检验的类比

- 原假设：被告无罪；备择假设：被告是有罪的
- 原则：无罪推论（保护原假设）
- 最后的判决：当有明显犯罪证据时，判决被告有罪
  - ✓ 证据是否明显由临界值决定（法律条文的规定）
  - ✓ 法律条文的严厉程度，决定了对被告的保护程度
  - ✓ 如法律条文严厉，保护被告的程度弱
  - ✓ 不管怎样制定条文，总有好人被冤枉，总有坏人漏网，即总会犯错误，但这样的事件是小概率事件。

作 业

《概率论与数理统计》：p.222-223, #25， #27，#31