概率论与数理统计 B　第十三周作业　5 月 9 日 周六

PB18151866 龚小航

8.13 令 $X_1,\dots,X_{10}$ 是从 $N(\mu,16)$ 中抽取的随机样本，假设样本均值 $\overline{X}=4.8$. 在 5% 显著水平下,检验:
(1) $H0: \mu=7 \leftrightarrow H1: \mu\neq 7$;　(2) $H0: \mu\geq 7 \leftrightarrow H1: \mu<7$.　(3) $H0: \mu\leq 2 \leftrightarrow H1: \mu>2$;

解：$n=10$，$X_1,\dots,X_{10}\sim N(\mu,4^2)$. 取标准化的检验统计量：其中$\mu$指总体的$\mu$.

$$\mathrm{U}=\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\sim N(0,1)$$

(1) $\alpha=0.05$时，有$P_{H_0}(|\mathrm{U}|>\alpha/2)=\alpha$，即检验的拒绝域为 $\{|\mathrm{U}|>\alpha/2\}$

即当观测值满足

$$\frac{\sqrt{n}|\overline{X}-\mu|}{\sigma}=\frac{\sqrt{10}|4.8-\mu|}{4}>u_{0.05/2}=1.9600$$

就会拒绝H_0. 对三题直接代入数据计算即可:

计算 $|\mathrm{U}|$ 再与1.9600比较即可:

$$|\mathrm{U}|=\frac{\sqrt{n}|\overline{X}-\mu|}{\sigma}=\frac{\sqrt{10}|4.8-7|}{4}=1.7393<1.9600$$

因此这时无法拒绝H_0

(2) $\alpha=0.05$时，有$P_{H_0}(\mathrm{U}<-u_\alpha)=\alpha$，即检验的拒绝域为 $\{\mathrm{U}<-u_\alpha\}$

即当观测值满足

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}=\frac{\sqrt{10}(4.8-\mu)}{4}<-u_{0.05}=-1.6449$$

就会拒绝H_0. 对三题直接代入数据计算即可:

计算 U 再与−1.6449比较即可:

$$\mathrm{U}=\frac{\sqrt{10}(\overline{X}-\mu)}{\sigma}=\frac{\sqrt{10}(4.8-7)}{4}=-1.7393<-1.6449$$

因此这时拒绝H_0

(3) $\alpha=0.05$时，检验的拒绝域为 $\{\mathrm{U}>u_\alpha\}$

即当观测值满足

$$\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}=\frac{\sqrt{10}(4.8-\mu)}{4}>u_{0.05}=1.6449$$

就会拒绝H_0. 对三题直接代入数据计算即可:

计算 U 再与1.6449比较即可:

$$\mathrm{U}=\frac{\sqrt{10}(\overline{X}-\mu)}{\sigma}=\frac{\sqrt{10}(4.8-2)}{4}=2.2136>1.6449$$

因此这时拒绝H_0

8.20. 2015 年全国人口调查中男女性别比例为 $\mu=105.02$ (女=100)，为检验这一比例，随机抽取了8个省份，其男女性别比例如下表，假设各省的性别比服从正态分布，

北京	内蒙古	辽宁	安徽	河南	海南	重庆	宁夏
109.45	104.32	100.45	104.90	103.99	110.47	100.60	106.16

在 5% 显著水平下，检验 $H0: \mu=105.02 \leftrightarrow H1: \mu\neq 105.02$

解：本例属于正态分布均值检验，且σ^2未知。计算样本的均值与方差:

$$\overline{X}=\frac{1}{8}\sum_{i=1}^{8}x_i=105.0425;\quad S=\sqrt{\frac{1}{8-1}\sum_{i=1}^{8}(x_i-105.0425)^2}=3.63730$$

σ^2未知时检验正态总体的均值，统计量为：其中μ就是待检验的参数，此时需要利用t检验

$$T=\frac{\sqrt{n}(\overline{X}-\mu)}{S}$$

由教材 206 页的结论，检验H_0的拒绝域为:

$$\left|\frac{\sqrt{n}(\overline{X}-\mu)}{S}\right|>t_{n-1}(\alpha/2)=2.365$$

带入本题数据，计算$|T|$再与$t_{n-1}(\alpha/2)$比较即可。

$$|T|=\left|\frac{\sqrt{8}(105.0425-105.02)}{3.63730}\right|=0.01750<2.365$$

因此此时不能拒绝H_0

8.30. 随机从一批钉子中抽取 9 枚，测得其长度（cm）为：

2.15,　2.13,　2.10,　2.14,　2.15,　2.16,　2.12,　2.11,　2.13,

假设钉子长度服从正态分布，分别在下面两种情况，5%显著水平下，

检验 $H0:\sigma \le 0.01 \leftrightarrow H1:\sigma > 0.01$ ：

(1) $\mu = 2.12$　　(2) μ 未知.

解：本例属于正态分布方差检验。　要检验的方差为 $\sigma_0 = 0.01$

(1) 当总体均值已知时，直接利用讲义上的结论，

由题知 σ^2 的极大似然估计：$\hat{\sigma}^2 = \frac{1}{9}\sum_{i=1}^{9}(x_i-\mu)^2 = 0.0005$ (仅用于统计量的计算)

构造的检验统计量为：

$$\chi^2 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{n\hat{\sigma}^2}{\sigma_0^2}$$

H_0拒绝域为：

$$\chi^2 > \chi_n^2(\alpha) = \chi_9^2(0.05) = 16.919$$

带入本题数据，可知：

$$\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 45 \quad ; \quad \sigma_0\chi_{n-1}^2(\alpha) = 0.01 * \chi_9^2(0.05) = 0.16919$$

显然有$\chi^2 > \chi_n^2(\alpha)$成立，因此这时$H_0$被拒绝。

(2) 当总体均值未知时，直接利用讲义上的结论，

构造的检验统计量为：

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

其中S为样本方差，计算之：

$$\overline{X} = \frac{1}{9}\sum_{i=1}^{n}x_i = 2.13222 \qquad S = \sqrt{\frac{1}{9-1}\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2} = 0.019861$$

H_0拒绝域为：

$$\chi^2 > \chi_{n-1}^2(\alpha) = \chi_8^2(0.05) = 15.507$$

带入本题数据，可知：

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{8 * 3.94444 * 10^{-4}}{0.01^2} = 31.5556$$

显然处于拒绝域中，因此这时H_0被拒绝。

8.35. 一个以减肥为主要目的的健美俱乐部声称，其训练班至少可以使肥胖者平均减少体重 8 kg 以上. 为检验该宣传是否可信，调查人员随机调查了 9 名参加者，得到他们训练前后的体重数据如下(单位：kg)：

训练前	104.5	94.0	104.7	96.4	91.6	90.9	92.0	99.9	109.7
训练后	94.2	86.6	97.5	91.7	82.6	83.8	81.3	92.2	101.0

现假设训练前后人的体重均服从正态分布. 问在 0.05 的显著水平下，是否可以认为该俱乐部的宣传是可信的?

解：本题是两个正态总体均值差的检验。 $\alpha = 0.05$

记训练前的总体为X，训练后的总体为Y，两者的差值 $Z = X - Y$

X	104.5	94.0	104.7	96.4	91.6	90.9	92.0	99.9	109.7
Y	94.2	86.6	97.5	91.7	82.6	83.8	81.3	92.2	101.0
Z	10.3	7.4	7.2	4.7	9.0	7.1	10.7	7.7	8.7

和 8.20 类似，接下来就是一个方差未知时，正态总体的均值检验。

记检验问题： $H_0:\mu \ge 8 \leftrightarrow H_1:\mu < 8$

计算样本的均值与方差：

$$\overline{Z} = \frac{1}{9}\sum_{i=1}^{8}x_i = 8.0888889; \quad S = \sqrt{\frac{1}{8-1}\sum_{i=1}^{8}(x_i - 8.0888889)^2} = 1.82992$$

σ^2未知时检验正态总体的均值，统计量为：

$$T = \frac{\sqrt{n}(\overline{X}-\mu)}{S}$$

其中μ就是待检验的参数，此时需要利用t检验. 由教材 206 页的结论，检验H_0的拒绝域为：

$$\left|\frac{\sqrt{n}(\overline{X}-\mu)}{S}\right| > t_{n-1}(\alpha/2) = 2.365$$

带入本题数据，计算$|T|$再与$t_{n-1}(\alpha/2)$比较即可。

$$|T| = \left|\frac{\sqrt{9}(8.088889-8)}{1.82992}\right| = 0.14573 < 2.365$$

因此此时不能拒绝H_0，该俱乐部的宣传是可信的。

8.40. 为了解甲、乙两企业职工工资水平，分别从两企业各随机抽取若干名职工调查，得如下数据(单位：元)：

甲公司	3750	5300	3750	9100	5700	5250	5000	
乙公司	5000	9500	4500	9000	6000	8500	9750	6000

设两企业职工工资分别服从正态分布，而总体独立且均值方差均未知. 试根据以上数据判断：两企业职工工资的方差是否相等？甲企业职工平均工资是否低于乙企业职工平均工资(α=0.05)?

解：本题是两个正态总体的情况。直接根据教材 212，213 页的结论，记检验问题：

$$H_0\colon \frac{\sigma_1^2}{\sigma_2^2}=1 \quad \leftrightarrow \quad H_1\colon \frac{\sigma_1^2}{\sigma_2^2}\neq 1$$

检验统计量为：

$$F=\frac{S_1^2}{S_2^2}\sim F_{6,7}$$

检验的拒绝域为：

$$\left\{F>F_{6,7}\left(\frac{\alpha}{2}\right)\ \ or\ \ F<F_{6,7}\left(1-\frac{\alpha}{2}\right)\right\}$$

再算出观察到的样本均值与方差：

$$S_{甲}=\sqrt{\frac{1}{7-1}\sum_{i=1}^{7}(x_{甲i}-\overline{X}_{甲})}=1798.941\,;\quad S_{乙}=\sqrt{\frac{1}{8-1}\sum_{i=1}^{8}(x_{乙i}-\overline{X}_{甲})}=2127.362$$

需要查表，当 $m=6, n=7, \alpha'=0.025$时，查得$F_{6,7}(0.025)=5.12$，$F_{6,7}(0.975)=\frac{1}{F_{7,6}(0.025)}=0.18$

则$F=\frac{S_1^2}{S_2^2}=0.715$，不在拒绝域中，因此可以认为两企业职工工资方差相等。

再判断甲企业的平均工资是否低于乙企业的平均工资：此时认为 $\sigma_1=\sigma_2$. 均值方差都未知

记检验问题：

$$H_0\colon \mu_{甲}-\mu_{乙}\geq 0 \quad \leftrightarrow \quad H_1\colon \mu_{甲}-\mu_{乙}<0$$

即为均值方差都未知的情况下两个正态总体均值差的检验。【教材 208 页】

引进统计量：

$$\mathrm{T}=\sqrt{\frac{mn}{m+n}}\frac{\overline{X}-\overline{Y}}{S}\sim t_{m+n-2}=t_{13}$$

其中，$\overline{X}=5407$; $\overline{Y}=7281$; $S_{甲}=1798.941$; $S_{乙}=2127.362$

$$S=\sqrt{\frac{1}{m+n-2}\left(\sum_{i=1}^{m}\left(X_i-\overline{X}\right)^2+\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2\right)}=1983$$

检验拒绝域为：$\{T<-t_{13}(\alpha)=1.771\}$

再将各数据带入计算T，可得，$T=-1.8$ 显然落在拒绝域内。

因此此时拒绝H_0，即甲企业的平均薪资低于乙企业。

8.44. 从甲、乙两处煤矿各随机抽取矿石 5 个和 4 个，分析其含灰率(%)得到如下结果.假设各煤矿含灰率都服从正态分布且方差相等，问甲乙两矿的含灰率有无显著差异(α=0.05)?

甲矿	24.3	20.8	23.7	21.3	17.4
乙矿	18.2	16.9	20.2	16.7	

解：记甲矿的样本总体为X，乙矿的样本总体为Y。方差相等，即有$X\sim N(\mu_1,\sigma^2)$，$Y\sim N(\mu_2,\sigma^2)$

这是总体均值方差都未知的检验问题。

记检验问题：

$$H_0\colon \mu_1=\mu_2 \quad \leftrightarrow \quad H_1\colon \mu_1\neq\mu_2$$

构造检验统计量：

$$T=\sqrt{\frac{mn}{m+n}}\frac{\overline{X}-\overline{Y}}{S}\sim t_{m+n-2}=t_7$$

其中，$\overline{X}=21.5$; $\overline{Y}=18$; $S_{甲}=2.7395$; $S_{乙}=1.6104$

$$S=\sqrt{\frac{1}{m+n-2}\left(\sum_{i=1}^{m}\left(X_i-\overline{X}\right)^2+\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2\right)}=2.3$$

检验拒绝域为$\{|T|>t_7(0.05/2)\}=2.365$

再带入所有数据，计算出由样本得出的$T=2.25$ ，不在拒绝域内。

因此无法拒绝H_0，即可以认为甲乙两矿含灰率无显著差异。

概率论与数理统计 B　第十三周作业　5 月 12 日 周二

PB18151866 龚小航

8.54. 2000 年的全国人口普查表明某城市的 65 岁以上老年人所占的比例为 13.55%, 现在为了调查人口的变动情况, 随机抽取 400 名居民, 发现其中有 57 人年龄在 65 岁以上. 试问该市现在老年人所占的比例较 2000 年普查时是否有变化(α=0.05)?

解: 记总体为X，显然X为二项分布，由于样本容量比较大，可以利用中心极限定理:

每抽取一名居民都是一次独立重复事件，记为 $X_1, X_2, \cdots\cdots, X_{400}$，它们服从独立同分布

若抽取出老年人，则 $X_i = 1$，否则$X_i = 0$。显然 $P(X_i = 1) = p$，$P(X_i = 0) = 1 - p$.

对题目给出的问题，可以记检验问题:

$$H_0: p = 13.55\% \quad \leftrightarrow \quad H_1: p \neq 13.55\%$$

再对这个问题进行检验:

$$\overline{X} = \frac{57}{400} = 0.1425 = 14.25\%$$

则检验量的观测值为:

$$T = \sqrt{n}\frac{\overline{X} - p}{\sqrt{p(1-p)}} = \sqrt{400}\frac{0.1425 - 0.1355}{\sqrt{0.1355(1 - 0.1355)}} = 0.4090$$

检验的拒绝域为:

$$\left\{|T| > u\left(\frac{\alpha}{2}\right)\right\} = \{|T| > u(0.025)\} = \{|T| > 1.96\}$$

显然样本的观测值在拒绝域外，因此不可以拒绝H_0

即说明该市老年人所占比例自 2000 年以来没有变化是可接受的。

8.59. 袋中装有 8 个球, 其中红球数未知. 在其中任取 3 个, 记录红球的个数 X, 然后放回; 再任取 3 个, 记录红球的个数, 然后放回. 如此重复进行 112 次, 得到结果如下:

x	0	1	2	3
次数	1	31	55	25

试在 α=0.05 水平下检验假设 $H0$: 红球的个数为 5.

解: 利用拟合优度检验，直接利用教材 234 页结论：此时$n = 112$

记一次实验（取三个球）得到的红球个数为随机变量X.

$$P(X=0) = p_0 = \frac{C_3^3}{C_8^3} = \frac{1}{56}\,; \qquad P(X=1) = p_1 = \frac{C_5^1 * C_3^2}{C_8^3} = \frac{15}{56}$$

$$P(X=2) = p_2 = \frac{C_5^2 * C_3^1}{C_8^3} = \frac{15}{28}\,; \qquad P(X=3) = p_3 = \frac{C_5^3}{C_8^3} = \frac{5}{28}$$

计算其理论值: $np_0 = 2$; $np_1 = 30$; $np_2 = 60$; $np_3 = 20$; $k = 4$

$$Z = \sum_{i=1}^{k}\frac{(np_i - v_i)^2}{np_i} = \frac{(2-1)^2}{2} + \frac{(30-31)^2}{30} + \frac{(60-55)^2}{60} + \frac{(20-25)^2}{20} = \frac{11}{5} = 2.2$$

检验的拒绝域为:

$$\{Z > \chi_{k-1}^2(\alpha)\} = \{Z > \chi_3^2(0.05)\} = \{Z > 7.815\}$$

此时显然观测到的Z落在拒绝域之外，因此不能拒绝假设H_0.

因此，在$\alpha = 0.05$的条件下，“红球的个数为 5”是可以接受的。

8.61. 为了研究蜗牛的种类是否与其生活的珊瑚礁种类有关, 选取了 3 种珊瑚礁作为检验样本, 记为 I, II, III , 记录下 A 和 B 两种蜗牛分别在 3 种珊瑚礁中生存的数目, 得到如下数据. 试问 A 和 B 两种蜗牛的分布是否在 3 种珊瑚礁中都是一样的 (α=0.05) ?

	I	II	III	合计
A	6	8	14	28
B	7	21	5	33
合计	13	29	19	61

解: 这是对列联表的应用，齐一性检验【教材 242 页 例 3.6】

$$Z = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{\left(nn_{ij} - n_i n_j\right)^2}{nn_i n_j}$$

其中 n_{ij}表示表格中第i行第j列格子中的数据，n_i表示第i行的合计，n_j表示第j列的合计。$n = 61$

只需计算六个格子内的值，再将它们相加即可:

$$Z_0 = \frac{(61*6 - 13*28)^2}{61*13*28} + \cdots\cdots + \frac{(61*5 - 19*33)^2}{61*19*33} = 9.82383$$

$$p(Z_0) = 1 - K_2(9.82383) < 0.01$$

因此结果高度显著,即有明显证据说明蜗牛的种类与其生活的珊瑚礁种类有关

8.63. 为了解男性和女性对三种类型的啤酒：淡啤酒、普通啤酒和黑啤酒的偏好有没有差异，分别调查了 180 位男士和 120 位女士的喜好，得如下数据

	淡啤酒	普通啤酒	黑啤酒
男性	49	31	100
女性	51	20	49

请问男性和女性对这三种类型的啤酒的偏好有显著差异吗? (α=0.05)

解：同上题，对列联表的应用，齐一性检验：

	淡啤酒	普通啤酒	黑啤酒	合计
男性	49	31	100	150
女性	51	20	49	120
合计	100	51	149	270

$$Z=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{\left(nn_{ij}-n_in_j\right)^2}{nn_in_j}$$

其中 n_{ij}表示表格中第i行第j列格子中的数据，n_i表示第i行的合计，n_j表示第j列的合计。$n=61$

只需计算六个格子内的值，再将它们相加即可：

$$Z_0=\frac{(270*6-100*150)^2}{270*100*150}+\cdots\cdots+\frac{(270*5-149*120)^2}{270*149*120}=8.197$$

$$p(Z_0)=1-K_2(8.197)<0.1$$

因此结果显著,即有证据说明男性和女性对三种类型的啤酒偏好有差异。

8.64. 检查一本书的 150 页，记录各页中印刷错误的个数，其结果为

错误的个数f_i	0	1	2	3	4	5	6	7
含f_i个错误的页数	86	40	19	2	0	2	1	0

试在显著性水平 0.05 下检验假设 H0：每页上的印刷错误个数服从泊松分布

解：由题意，写出检验问题：【教材 245 页 例 3.7】

$$H_0\text{:每页上错误个数服从泊松分布}$$

则有：$\hat{\lambda}=\overline{X}=\dfrac{0*86+1*40+2*19+3*2+4*0+5*2+6*1+7*0}{150}=\dfrac{2}{3}$

由于错误较多的页面较少，将$f_i\geq 3$的错误页面合成一组。先算理论值：

$$np_0=150*\frac{\left(\frac{2}{3}\right)^0e^{-\frac{2}{3}}}{0!}=77.01\,;\qquad np_1=150*\frac{\left(\frac{2}{3}\right)^1e^{-\frac{2}{3}}}{1!}=51.34$$

$$np_2=150*\frac{\left(\frac{2}{3}\right)^2e^{-\frac{2}{3}}}{2!}=17.11\,;\qquad np_3=150-n(p_0+p_1+p_2)=4.54$$

$$Z_0=\sum_{i=1}^{k}\frac{(np_i-v_i)^2}{np_i}=\frac{(77.01-86)^2}{77.01}+\frac{(51.34-40)^2}{51.34}+\frac{(17.11-19)^2}{17.11}+\frac{(4.54-5)^2}{4.54}=3.80$$

检验的拒绝域为：

$$\{Z>\ \chi^2_{4-1-1}(\alpha)\}=\{Z>\ \chi^2_2(0.05)\}=\{Z>5.991\}$$

此时显然观测到的Z落在拒绝域之外，因此不能拒绝假设H_0.

再计算拟合优度，$p(Z_0)=1-K_2(3.80)\mathrm{K}\in(10\%,25\%)$

即有上述机会产生比比例数据更大的偏离。

可以认为每页上的错误服从泊松分布。