第六章 基本气候状态的统计检验

章节概述

通过某一气候变量序列的均值和方差可以了解其变化的平均状态和变化的幅度。但气候系统常常发生突变（均值突变、方差突变、趋势突变、关系突变），但不清楚这种状态是否稳定、变化是否显著。因此，需要进行统计检验。

气候稳定性检验涉及两种情形：① 检验某一地区的气候是否具有稳定性：比较不同时段气候变量的均值或方差是否发生显著变化。② 检验两个地区的气候是否存在显著差异：比较两个地区气候变量的均值或方差来判断。

6.1 均值的检验

检验方法 u检验和t检验

合成分析 在气候变化研究中，常要研究某些**特殊年份**有何**显著特点**。经常使用的方法就是将**这些特殊年份**的气象要素与**其它年份的平均值**进行比较。

6.1.1 u检验

适用情况 ① **总体方差已知**（总体理论上是无限的，但如果有很长的资料，可以用它的方差代表总体方差）

② 对遵从**正态分布**的观测对象样本量**大或小**均适用。

③ 若样本量足够大，即使观测对象不遵从正态分布也适用。

因为总体方差往往不知道，u检验在实际应用中不经常使用。

统计量 ① **单个变量**总体均值的检验，检验一地气候是否稳定：(其中μ_0和σ为**总体**均值和标准差)

$$u = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n}$$

② **两个总体均值**的检验，检验两地气候变化是否存在显著差异：

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

实际研究

实际研究中，研究对象的定义往往十分重要。比如“台风季”的定义，往往在 6~10 月，但具体时间随年份有很大变化。既然样本覆盖范围不同，那么得到的结论明显不同。

检验判断 随后，对于给定的**显著性水平α (0.05)**，可以查表得到相应的u_α，若$|u| \geq u_\alpha$ 则拒绝假设；若$|u| < u_\alpha$ 接受假设。

应用案例 1

经正态检验，中国 1910-1989 年年平均气温等级遵从正态分布，其均值为 2.94，标准差为 0.30。又观测得到 1990-1994 年中国年平均气温等级分别为 2.60，3.30，3.70，3.10 和 2.40，样本均值$x = 3.02$。检验在显著性水平$\alpha = 0.05$下，中国年平均气温等级的总体均值与样本均值有无显著差异，即总体均值有无改变。这里样本量$n = 5$。检验有：

① 提出原假设H_0: $\mu = \mu_0$。即总体均值与样本均值之间没有显著差异。

② 计算统计量。将特征值代入方程：$u = \frac{3.02-2.94}{0.30} \approx 0.604$。

③ 当$\alpha = 0.05$时，查正态分布函数表$u_\alpha = 1.96$，那么$|u| < u_\alpha$，接受原假设。至此，可以得出结论：在$\alpha = 0.05$显著性水平上，可以认为 1990-1994 年样本均值与年平均气温总体均值无显著差异，即年平均气温变化是稳定的。这里应强调，这一结论是在 $\alpha = 0.05$ 的显著性水平上得出的，如果以更低的显著性水平进行检验，有可能得出不同的结论。

应用案例 2

赤道东太平洋地区(0°~10°S，180°~90°W)春季(3~5 月)平均海表温度 39 年平均值为 27.5℃，方差为 2.07℃。西风漂流区(40°~20°N，180°~145°W)春季平均海表温度 39 年平均值为 17.3℃，方差为 2.08℃。检验两地区海温平均值有无显著差异。这里样本量$n_1 = n_2 = 39$。检验步骤为：

① 提出原假设$H_0: \mu_1 = \mu_2$。两总体均值之间没有显著差异。

② 计算统计量，$u = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} = \frac{27.5-17.3}{\sqrt{\frac{2.07^2}{39}+\frac{2.08^2}{39}}} = 21.4$。

③ 当显著性水平$\alpha = 0.05, u_\alpha = 1.96$。那么，$u > u_\alpha$，拒绝原假设，在$\alpha = 0.05$的显著性水平上，认为赤道东太平洋春季海温均值与西风漂流区春季海温的均值之间存在显著性差异。

6.1.2 t检验

适用情况 ① **总体方差未知**

② 遵从正态分布的均值检验，小样本也适用。

t统计量 ① 构造检验**一个总体均值稳定性**的统计量：

$$\boldsymbol{t = \frac{\bar{x}-\mu_0}{s}\sqrt{n}}$$

$\bar{x}$ 和 s 分别代表**样本**均值和标准差，μ_0为**总体均值**。对于给定的显著性水平 α，根据**自由度$\boldsymbol{n-1}$**查t分布表得到相应的 t_α，作判别。

② 构造检验**两个总体均值稳定性**的统计量：

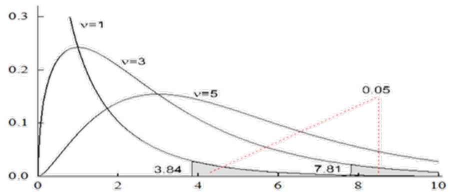
$$t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \xRightarrow{\text{足够大时}} t = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$\bar{x}, \bar{y}, s_1^2, s_2^2, n_1, n_2$ 分别为两个样本的均值、方差和样本个数。对于给定的显著性水平 α，根据**自由度$\boldsymbol{(n_1+n_2-2)}$**查t分布表得到相应的t_α，作判别。

6.2 方差的检验

6.2.1 χ^2检验

适用范围 检验**某一地区**气候是否稳定



$\chi^2_{0.05(1)} = 3.84 = (1.96)^2 = Z^2_{0.05/2}$

$\chi^2_{0.01(1)} = 6.63 = (2.5758)^2 = Z^2_{0.01/2}$

统计量 $\boxed{\boldsymbol{\chi^2 = \frac{(n-1)s^2}{\sigma^2}}}$ σ 是总体方差，s 是样本方差，**自由度为$\boldsymbol{\nu = n-1}$**

检验判断 若 $\chi^2 > \chi^2_{\alpha/2}$ 或 $\chi^2 < \chi^2_{1-\alpha/2}$ 则认为总体方差有显著变化。

均值已知 若**总体均值已知**，可以用下面的统计量进行检验：

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

其中，n为样本数，x_i为观测样本，上式遵从**自由度为$\boldsymbol{n}$**的χ^2分布。

卡方分布 若 $Z \sim N(0,1)$，则 Z^2 的分布称为**自由度为 1** 的卡方分布(χ^2,*chi-square distribution*)，记为 $\chi^2_{(1)}$ 图形如上图，从纵轴某个点开始单调下降，先凸后凹。

应用案例

已知上海 10 月逐日相对湿度(单位:%)近似遵从正态分布，且$\sigma^2 = 102.9$，又测得 5 天相对湿度，算出 $s^2 = 46.4$。

① 提出原假设$H_0: \sigma = \sigma_0$。可表述为总体方差与样本方差无显著差异。

② 计算统计量。将特征量代入，算出$\chi^2 = \frac{(5-1)\times 46.4}{102.9} = 1.80$。

③ 确定显著性水平$\alpha = 0.10$，自由度$v = 5-1 = 4$，查分布表$\chi^2_{\alpha/2} = 9.49$，$\chi^2_{1-\alpha/2} = 0.71$。$\chi^2_{1-\alpha/2} < \chi^2 < \chi^2_{\alpha/2}$，所以接受原假设，认为总体方差与样本方差之间无显著差异。

6.2.2 F检验

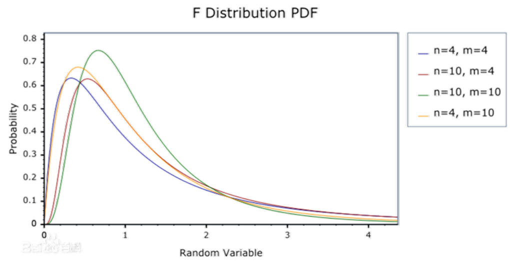
适用范围　在总体方差未知的情况下，假定是分别来自两个相互独立的正态总体的样本方差

统计量　$\boxed{F=\frac{n_1 s_1^2}{n_1-1}\Big/\frac{n_2 s_2^2}{n_2-1}}$　遵从**自由度$v_1=n_1-1, v_2=n_2-1$**的F 分布

检验判断　若$F>F_{\alpha/2}$ 或 $F<F_{1-\alpha/2}$，则认为两地样本方差有显著差异，或者说气候有显著差异。

注意　由于F分布有两个自由度，通常只对几个显著性水平α制表，用这种表只能查出临界值F_α，而查不出临界值$F_{1-\alpha}$。要得到$F_{1-\alpha}$的值，**可交换分子、分母自由度**再求所查得**临界值的倒数**。

$$F_{df_1,df_2,1-\alpha}=\frac{1}{F_{df_2,df_1,\alpha}}$$



应用案例 1

用F检验两个总体的样本方差有无显著差异。赤道东太平洋春季海温样本方差$s_1^2=3.1$℃，西风漂流区春季海温样本方差$s_2^2=2.3$℃。这里$n_1=9$，$n_2=11$。

① 原假设 H_0: $\sigma_1=\sigma_2$

② 计算统计量，算得$F=1.38$。

③ 给定显著性水平$\alpha=0.10$，自由度$v_1=9-1=8$，$v_2=11-1=10$，查F分布表，$F_{\alpha/2}=3.07$，$F<F_{\alpha/2}$，接受原假设，认为赤道东太平洋春季海温与西风漂流区春季海温的样本方差无显著差异。

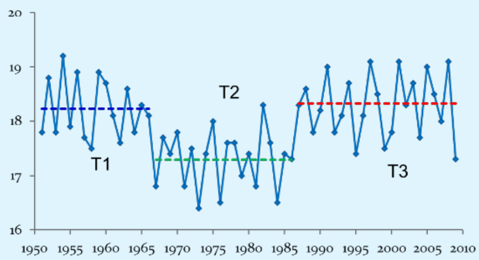
应用案例 2

	均值	标准差	样本量	t_α临界值	t统计值	$F_{\alpha/2}$临界值	F统计值
T1	18.2	0.54	16	2.03	14.2(T12)	2.23	1.1(T12)
T2	17.3	0.52	20	2.02	-15.4(T23)	2.09	0.8(T23)
T3	18.3	0.57	23	2.02	-1.5 (T13)	2.15	0.9(T13)

① T1,T2 两者均值显著差异，因为t统计值大于t_α临界值。

② T1,T2 两者方差存在不存在显著差异，因为$F<F_{\alpha/2}$。

③ 其他内容也可以类似分析



1951-2009年黑龙江呼玛市8月气温逐年演变曲线