

定量分析：数据思维与 商业统计

陈文波

cwb@whu.edu.cn

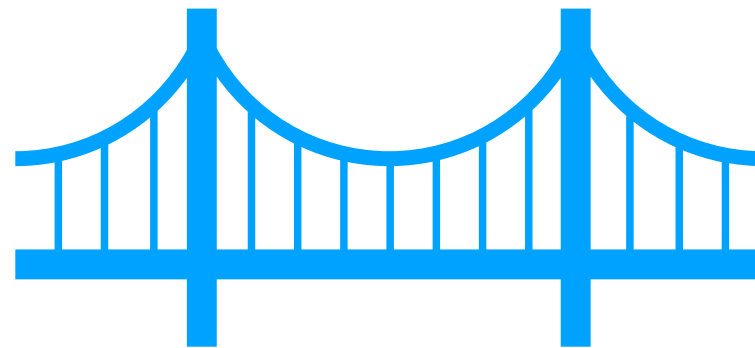
2021年10月

课程梗概

描述性分析与EDA

- 数值方法：三类指标
- 图形方法：五种图形

R



- 抽样分布、CLT
- 正态分布
- T分布
- 卡方分布
- F分布

推断性分析

- 区间估计
- 假设检验
- 应用：
 - 分类 vs. 分类
 - 分类 vs. 数值
 - 数值 vs. 数值

课程梗概

- Session 1 : 内容简介、基本概念、描述性统计1
- Session 2: 描述性统计2
- Session 3: 抽样、中心极限定理、正态分布
- Session 4: 其他三个分布, 区间估计
- Session 5: 假设检验, 两个总体均值和比例的推断
- Session 6: 总体方差和方差比的推断、分类数据 Vs. 数值型数据: 卡方检验与列联表分析
- Session 7: 分类数据 Vs. 数值型数据: 方差分析; 数值型数据: 回归分析
- Session 8: 课程总结

关于方差的推断

- 最关键的是？

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

例子

- 武汉公交公司正在推行一项质量改进运动，公司希望减少到站时间的方差，从而塑造公司准点到达的形象。
- 附件数据是公司对某站点24次到站时间的抽查记录。公司希望知道这一站点是否符合公司提出的方差小于4这一标准。

假设检验过程

- 五步法

判断标准

	Lower Tail Test	Upper Tail Test	Two-Tailed Test
Hypotheses	$H_0: \sigma^2 \geq \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 \leq \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$
Test Statistic	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
Rejection Rule: p-value Approach	Reject H_0 if $p\text{-value} \leq \alpha$	Reject H_0 if $p\text{-value} \leq \alpha$	Reject H_0 if $p\text{-value} \leq \alpha$
Rejection Rule: Critical Value Approach	Reject H_0 if $\chi^2 \leq \chi_{(1-\alpha)}^2$	Reject H_0 if $\chi^2 \geq \chi_{\alpha}^2$	Reject H_0 if $\chi^2 \leq \chi_{(1-\alpha/2)}^2$ or if $\chi^2 \geq \chi_{\alpha/2}^2$

两个总体方差的推断

- 最关键的是？

$$\frac{s_1^2}{s_2^2} \sim F(n_1 - 1, n_2 - 1)$$

例子

- 磁疗可以缓解疼痛?
 - 为了测试磁疗是否可以缓解疼痛，通过对实验组和对照组分别给与磁疗和安慰剂，并询问受试疼痛的缓解程度，得到如下数据：
 - 对照组： $n = 20, \bar{x} = 0.44, s = 1.4$
 - 实验组： $n = 20, \bar{x} = 0.49, s = 0.96$
 - 可否认为给予安慰剂的对照组具有更大的方差？

练习

- 数据: [WE.xlsx](#)
- 报告:
 - 检验流失组与非流失组在10个指标均值上是否有显著差异;
 - 计算各指标的方差, 检验流失组和非流失组在上述指标的方差上是否有显著差异。

churn	mean_chi_m0	mean_chi_m01	mean_support_m0	mean_support_m01	mean_sp_m0	mean_sp_m01	mean_login_m0	mean_blog_m01	mean_view_m01	mean_days_last_login
0	88.60591	5.530213	0.7242696	-0.0092961	0.8295759	0.0326818	16.13894	0.1711487	106.6096	3.511454
1	63.27245	-3.736842	0.3715170	0.0371517	0.4995577	-0.0166962	8.06192	-0.1021672	-95.7678	8.486068

推断总结

- 推断性统计关注的也是总体的三类指标
- 要实现从样本到总体的推断，需要知道：
 - 统计量的抽样分布
- 从样本出发，构造一张网（区间）来罩住真实值，或者
- 看样本统计值在假定的总体参数下，是大概率还是小概率事件

统计推断的应用

x

y

分类

分类

数值

数值

χ^2

F ANOVA

logit

lm

分类数据 vs. 分类数据

- 列联表分析
- 拟合优度检验

列联表分析

泰坦尼克悲剧

Source:
Wiki

时间：1912.4.15

地点：

西经49度56分54.23秒

北纬41度43分38.01秒



项目	男性	女性	儿童	总计
幸存	338	316	56	710
死亡	1352	109	53	1514
总计	1690	425	109	2224

★ Ottawa

圣彼得岛 ★ St Pierre



W49 56'54.23 N41 43'38.01

北大西洋

百慕大

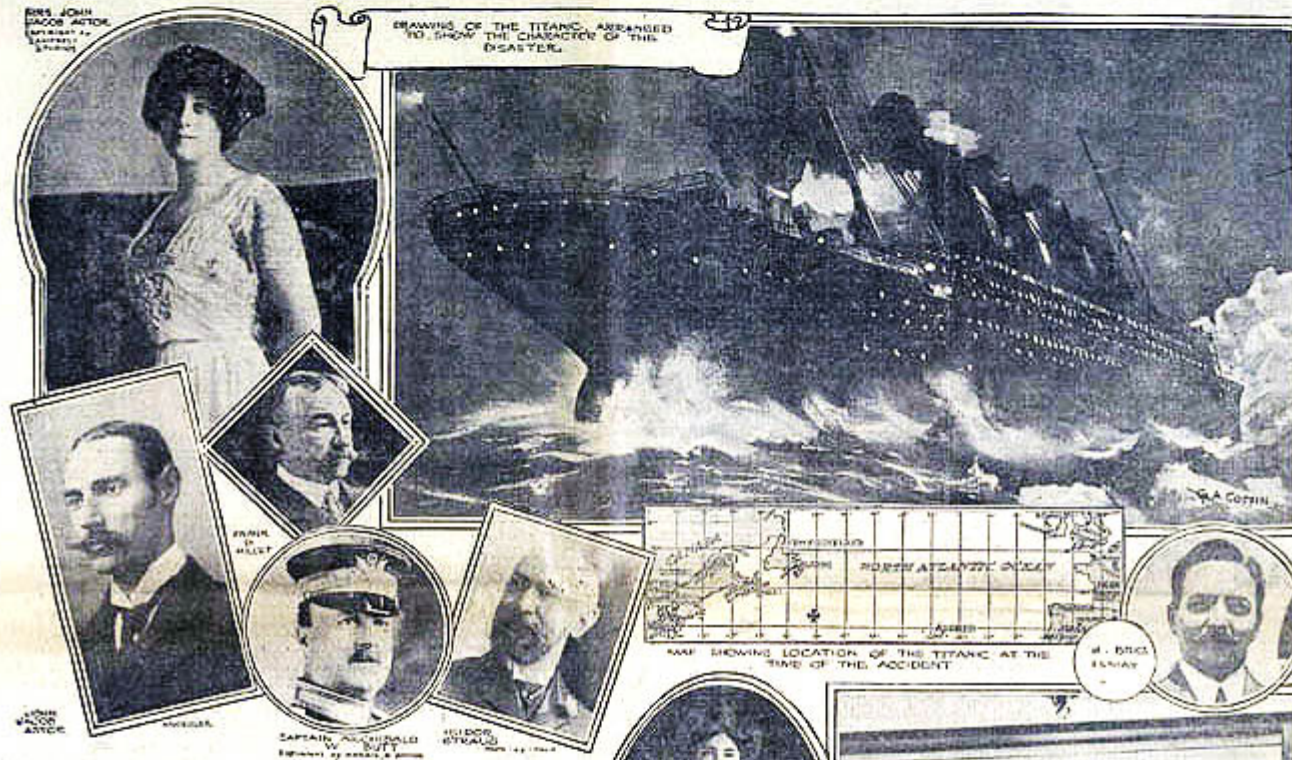
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
US Dept of State Geographer
© 2011 Europa Technologies
© 2011 Google

©2010 Google™

41°43'38.01" 北 49°56'54.23" 西 海拔 -3575 米

视角海拔高度 3185.98 公里

THE TITANIC SINKS WITH 1,800 ON BOARD; ONLY 675, MOSTLY WOMEN AND CHILDREN, SAVED



MOST APPALLING DISASTER IN MARINE HISTORY OCCURS WHEN WORLD'S LARGEST STEAMSHIP STRIKES GIGANTIC ICEBERG AT NIGHT

Scores of World's Most Widely Known Persons, Including Colonel John Jacob Astor and His Wife, William T. Stead and Isidor Straus Among Those Whose Fate Is in Doubt.

J. Bruce Ismay, Head of the White Star Line; Mr. and Mrs. George D. Widener, the Countess of Rothes and Mrs. Henry B. Harris Are Some of Those Saved.

In the darkness of night and in water two miles deep the Titanic, swiftest of the White Star fleet and greatest of all ocean steamships, sank to the bottom of the sea at twenty minutes past two o'clock yesterday morning. Dispatches received late last night from the Cape Race Wireless station in Newfoundland and elsewhere, made at the same time by the New York office of the White Star Company, warrant the fear that of the 2,200 persons who were aboard the great vessel when she received her mortal wound in collision with an iceberg more than 1,500 have gone to their death in her shattered hull, while 675, most of whom are women and children, have been saved.

Should these given figures be verified, the loss of the Titanic—entirest, most powerful, greatest of all the ocean fleet—while speeding westward on her maiden voyage will take rank in maritime history as the most terrible of all recorded disasters of the sea.

A dispatch from St. John, N. F., received at fifteen minutes past two o'clock this morning, gives me the hope that the steamship Virginia may have some of the Titanic's survivors on board. The message said she would bring to St. John such survivors as she "may rescue." Hope arises from the fact that the steamer is putting in here at all, which she would scarcely need to do were there not some humane humanity for that cause. She was outward bound for Liverpool.

One point is known from which may be deduced a sad satisfaction. In a desperate situation when the salvation of all was not possible the women and children were sent first. There was sent away in the first of the boats launched from the sinking ship, the only boat apparently which did not state the fate of the man who remained.

America and Britain are spared the horrors that attended the sinking of the French ship La Bourgogne, in 1898, when the women and children were trampled under feet and cut down with thrusts.

The Carpathia's Survivors List

A partial list of the survivors who are aboard the Carpathia, leaving for this port, will be found on page 2 of this morning's edition.



It was learned late last night that the full text of the message New York aboard the Carpathia, of the Cunard line, the loss of life from the steamship Olympic reporting the sinking of the Titanic, only indicated would be 1,491 persons.

The full text of the dispatch received from the Olympic is as follows:—"Carpathia reached Titanic's position at daybreak. Titanic sank about twenty minutes past two o'clock in the morning, in 41.16 north latitude, 70.14 west longitude. All her boats accounted for, containing about 675 souls saved, crew and passengers included. Nearly all saved women and children. Layland Roger Californian remained and searching exact position of disaster. Lost 2,201 souls."

By midnight Bostoning Gones in front of the White Star offices

检验1： 性别与生存

项目	男性	女性	总计
幸存	338	316	654
死亡	1352	109	1461
总计	1690	425	2115

检验1： 性别与生存

项目	男性	女性	总计
幸存	338	316	654
期望	522.58	131.42	
死亡	1352	109	1461
期望	1167.42	293.58	
总计	1690	425	2115

检验1： 性别与生存

检验2： 仓位等级与生存

项目	头等舱	二等舱	三等舱	船员舱	总计
幸存	202	118	178	212	710
死亡	123	167	528	696	1514
总计	325	285	706	908	2224

Is the nurse a serial killer?

	Shifts with a death	Shifts without a death
Gilbert was working	40	217
Gilbert was not working	34	1350



拟合优度检验

- Goodness of fit

Last Digit	Frequency
0	7
1	14
2	6
3	10
4	8
5	4
6	5
7	6
8	12
9	8

Wrap-up

- 关于方差的推断
- 独立性检验（列联表分析）
- 拟合优度检验