

# 定量分析：数据思维与 商业统计

陈文波

[cwb@whu.edu.cn](mailto:cwb@whu.edu.cn)

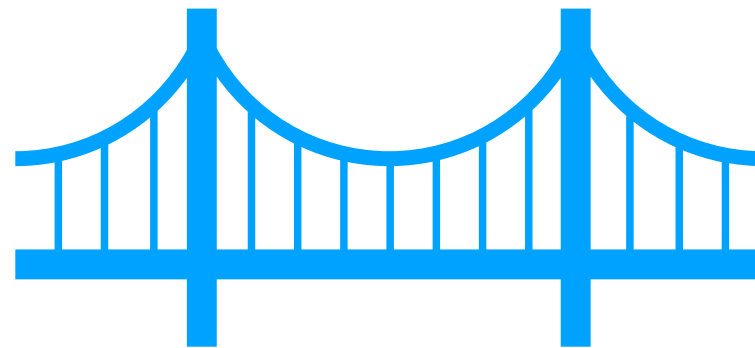
2021年10月

# 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

# Review

---

- 统计量的抽样分布

# S5-假设检验、两个总体

---

- 假设检验基本思想与过程
- 两类错误
- 均值、比例的假设检验
- 两个总体的均值和比例

# 基本思想

---

- 小概率事件反证法

# 过程： 五步

---

1. 提假设

2. 给定显著性水平

3. 计算检验统计量

4. 计算p-value

5. 做出判断

**OR**

1. 提假设

2. 给定显著性水平

3. 计算检验统计量

4. 计算临界值

5. 做出判断

# 例子：急救中心服务水平评价

---

- 为了衡量急救中心的反应速度，随机抽取了40辆救护车的作为测试对象。发现40辆救护车平均到现场的时间是13.25分钟。从历史资料可知，总体标准差是3.2分钟。
- 医疗主管部门想知道，在0.05的显著性水平下，医疗主管部门要求的最多12分钟到现场这一服务水平承诺是否达到。





# 两个要点

---

- 原假设的提法
- 单侧与双侧检验中的判断标准

# 总结

---

- 需要知道对应统计量的抽样分布

1

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

2

$$\bar{X} \sim t(n - 1)$$

3

$$\bar{P} \sim N\left(p, \frac{\bar{p} \times (1 - \bar{p})}{n}\right)$$

# 两个总体的均值和比例

---

- 两个总体：方差已知和未知
- 两个总体：匹配样本
- 两个总体：比例之差

# 两个总体，方差已知

---

$$\overline{X}_1 - \overline{X}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

# 两个总体，方差未知

---

$$\overline{X}_1 - \overline{X}_2 \sim t(df)$$

# 两个总体，匹配样本

---

# 两个总体：比例的推断

---

# Wrap-up

---

- 小概率事件反证法
- 假设检验五步法
- 两个总体：统计量的抽样分布！