Agile Development

# A++

敏捷开发小手册

宋世傑（Edmond S.K. SUNG） 彭国明

# A++ 敏捷开发小手册

– 从个人提升到以数据说话

*impossible* publisher 2022
printed blindfolded
design: LaTeX

# Contents

# Part I

# 个人提升

# Part II

# 自我管理开发过程

软件开发工程师自我改进过程

# Part III

# 利用评估开始过程改进

团队必须先获得管理层的关注与支持

# Part IV

# 从做好迭代回顾开始

做好迭代回顾让团队向量化管理迈进一大步

# Part V

# 基础：代码质量

这部分参照 Jeffery 先生总结极限编程 (XP) 的三层图（在第一章里），利用实例，先从内层的测试驱动开发 (TDD)，重构 (Refactoring), 与结对编程 (PairProgramming) 等核心最佳实践开始，然后到中层的持续集成 (Continuous-Integration)，最后是外层的团队各角色协作/互补 (Whole Team) 和如何做好需求分析等。

# Part VI

# 度量与分析

当团队开始养成每天收集数据的习惯，在迭代回顾时分析数据，公司便可以开始准备度量与分析。

(注: 要做好度量与分析，跟过程改进一样，要针对目标，制定度量计划。) 本部分开头会以问卷调查为例，介绍度量计划的主要元素，

然后会介绍统计分析，控制图，假设检验等数据分析的基本功。

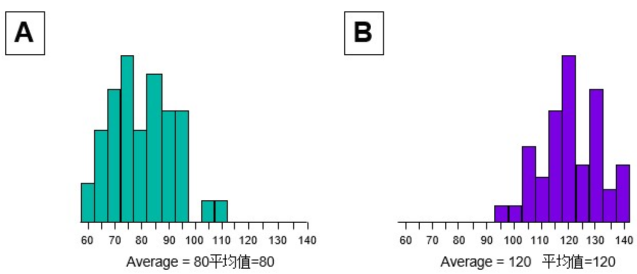
最后利用某银行软件维护数据分析案例，简单介绍数据收集后，做分析的主要步骤。

注：请不要误以为度量与分析必须由管理层驱动。若要团队能从定性升到定量管理，并能持续，而且有效果，必须从团队开始，而不是靠管理者总体策划

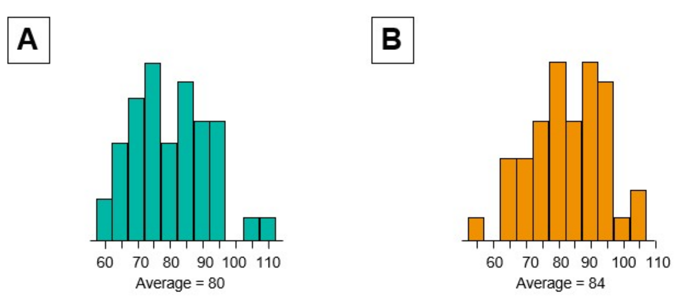
# Chapter 23

# 假设检验

上章介绍了双样本 T 检验，是假设检验的一种；下面介绍假设检验，及它帮我们解决什么问题。




想比较 A B 两组实验结果是否有显著差异。如果像上图应该看到 B 很明显比 A 高。
但如果像下图，我就很难看得出来：



假设检验利用统计分析方法帮我们区分 - 两组数据是否有显著的差异。

## 假设检验的步骤

零假设: A 与 B 没有显著差异

备选假设: A 与 B 有显著差异

挑选对应的检验方法后，

可以利用工具计算出的 P 值来判断：

如果 P 值大于 0.05，就不能拒绝零假设
如果 P 值小于 0.05，就可以拒绝

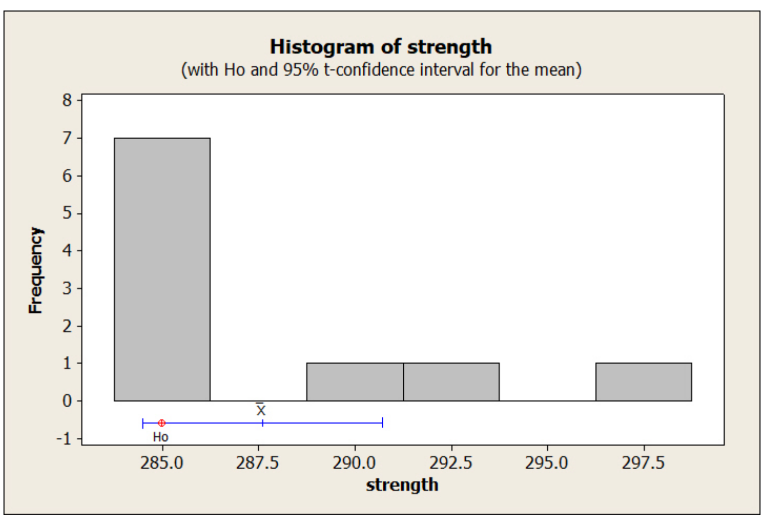
因 A 与 B 都包含非常多样本数据，我们只可以利用 A 的随机抽样，和 B 的随机抽样数值，来判断 A 与 B 是否有显著差异，这便是双样本 T 检验。
要深入了解便要与数据，先举个比较简单的假设检验实例：

## 检验单个正态总体的均值

按历史数据，某厂生产的钢丝折断力是正态分布（平均值 =285 ，标准差 =4），更换了新的钢材供货商，随机抽取以下 10 样本，想判断更换供货商后折断力和原先是否有显著提升？

289 , 286 , 285 , 284 , 286 , 285 , 285 , 286 , 298 , 292

从下面样本钢丝折断力柱状图看到 10 个样本中有七个接近 285, 有 3 个是明显大于 285，应如何判断是否有显著差异呢？



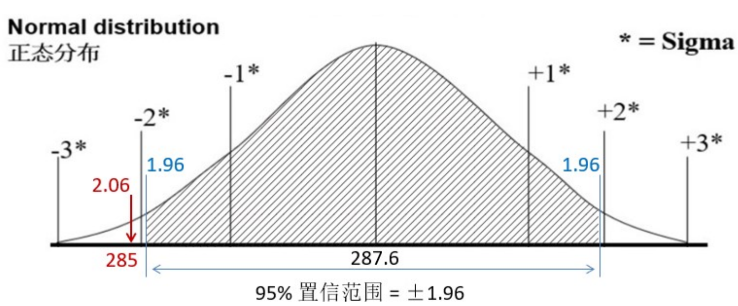

按假设检验步骤：

1. 零假设 *：新钢材的折断力小于或等于 285 （$\mathbf{H}_0 : \mu \leq 285$）

2. 备选假设：新钢材的折断力大于 285 （$\mathbf{H}_1 : \mu > 285$）

(* 通常设零假设为原本情况，没有变化，备选假设为希望的情况，有提升）

这例子，可以使用正态分布相关方程式，估计概率，如概率低于 5% ，拒绝零假设，（折断力 > 285) （详见附件）
如果不想研究相关方程式，可利用概率分布图 “看” 到：

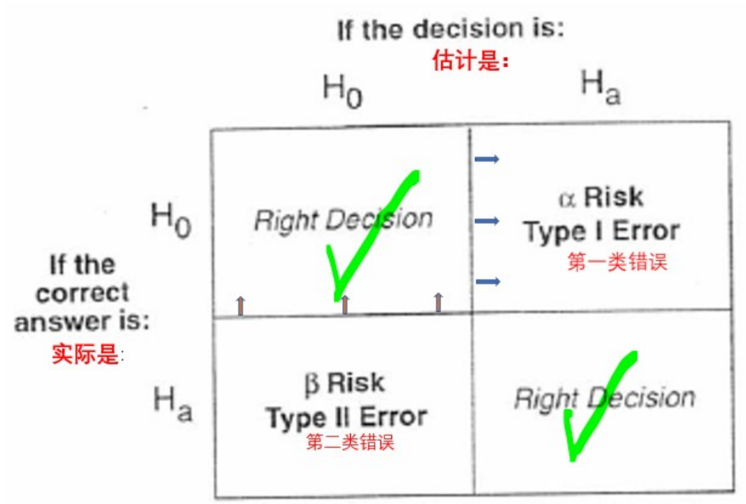



利用正态概率分布图，抽样的平均值是 287.6
历史的平均值是 285，之间的差异超出了 95% 置信区间, 所以我们拒绝零假设，样本与 285 有显著差异（提升）。

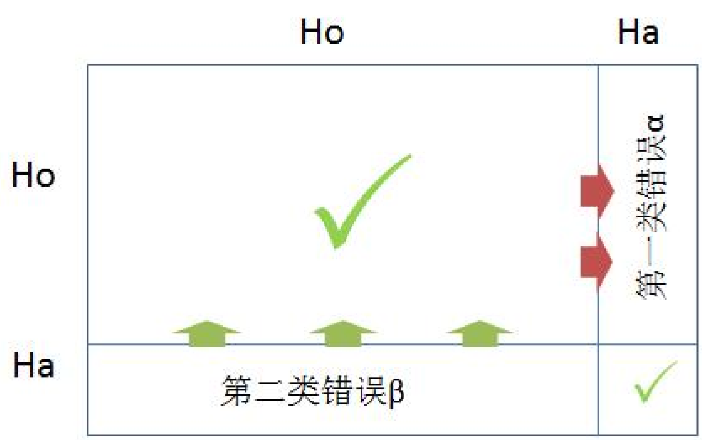
也可以利用工具，计算单样本 Z 检验 (one-sample Z test) 的 P 值，因 P 值 =0.04, 少于 0.05，判断拒绝零假设，新供货商的钢丝折断力比原本 285 高。

要了解 P 值是什么？为什么低于 0.05 可以拒绝？首先要理解的会发生两类错误
检验可能犯错误, 所谓犯错误就是检验的结论与实际情况不符, 有两种情况:

1. 实际情况是 $\mathbf{H}_0$ 成立, 而检验的结果表明 $\mathbf{H}_0$ 不成立, 即拒绝了 $\mathbf{H}_0$，这时称该检验犯了第一类错误 (Type I error) 或 “弃真” 错误
2. 实际情况是 $\mathbf{H}_0$ 不成立,$\mathbf{H}_1$ 成立, 而检验的结果表明 $\mathbf{H}_0$ 成立, 即接受了 $\mathbf{H}_0$, 这时称该检验犯了第二类错误 (Type II error), 或称 “取伪” 错误




这两类错误 $\alpha$ 与 $\beta$ 之间是有关系：
如果 $\alpha$ 风险（概率从本来 0.05 降低）下降，便会引起 $\beta$ 风险提高，所以不能同时降低 $\alpha$ 与 $\beta$ 。$\alpha$ 不能无限降低，一般选 $\alpha$ =0.05 。




我们可以简单理解 P 值为发生第一类错误的概率，如果低于 5% 我们就有信心拒绝零假设。

假设检验用于比较两组数据有没有显著差异的例子；

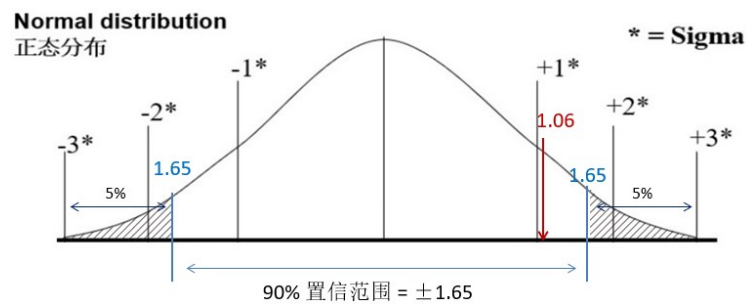
## 比较两个正态总体的均值

一名研究人员假设，大学为男生提供的运动项目的平均数量大于大学为女生提供的运动项目的平均数量。对大学为男生和女生提供的体育项目的随机抽样显示，设定临界概率为 5%（Alpha $\alpha$=0.05），是否有足够的证据支持这种说法?（假设方差都一样，$\sigma = 3.3$）。

| 男生 Males | | | | | 女生 Females | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 11 | 11 | 8 | 15 | 6 | 8 | 11 | 13 | 8 |
| 6 | 14 | 8 | 12 | 18 | 7 | 5 | 13 | 14 | 6 |
| 6 | 9 | 5 | 6 | 9 | 6 | 5 | 5 | 7 | 6 |
| 6 | 9 | 18 | 7 | 6 | 16 | 10 | 7 | 8 | 5 |
| 15 | 6 | 11 | 5 | 5 | 16 | 10 | 7 | 8 | 5 |
| 9 | 9 | 5 | 5 | 8 | 7 | 5 | 5 | 6 | 5 |
| 8 | 9 | 6 | 11 | 6 | 9 | 18 | 13 | 7 | 10 |
| 9 | 5 | 11 | 5 | 8 | 7 | 8 | 5 | 7 | 6 |
| 7 | 7 | 5 | 10 | 7 | 11 | 4 | 6 | 8 | 7 |
| 10 | 7 | 10 | 8 | 11 | 14 | 12 | 5 | 8 | 5 |

1. 零假设：男女运动项目平均数量没有区别（$\mathbf{H}_0 : \mu_1 \leq \mu_2$）
2. 备选假设：男生运动项目平均数量比女生多（$\mathbf{H}_1 : \mu_1 > \mu_2$）

男生的均值是 8.6 ; 女生是 7.9 ; 相差是 0.7
假定使用 90% 置信区间:
正态分布 $\pm 1.65\sigma$ (标准差) 内的面积是总面积的 90%，从下图看到 1.06 落在 90% 置信区间之内，所以不能拒绝零假设。




（有关计算 1.06 的方程式，详见附件）
使用工具跑出来的 P 值是 0.146 ，大于 0.05 ，所以不能拒绝零假设

以上两个例子都是假定：分布的方差已知

但很多时，我们不一定知道分布的方差或标准差是多少？我们便需要从样本数估计分布的方差（或标准差）。下面是一个结对 T 检验 (Paired T test) 例子

## 检验单个正态总体的均值, 方差未知

营养师正在观察如果每天饮食都添加某类矿物质，是否会改变人的胆固醇水平。对随机选取 6 位对象在实验之前进行了水平测试。然后对他们进行为期 6 周的测试，在它们大食物里添加矿物质补充，实验之后再测试，结果显示在下表 (胆固醇水平以每分升毫克为单位)。能否得出结论，在 $=0.1$ 时，胆固醇水平是否发生了变化。

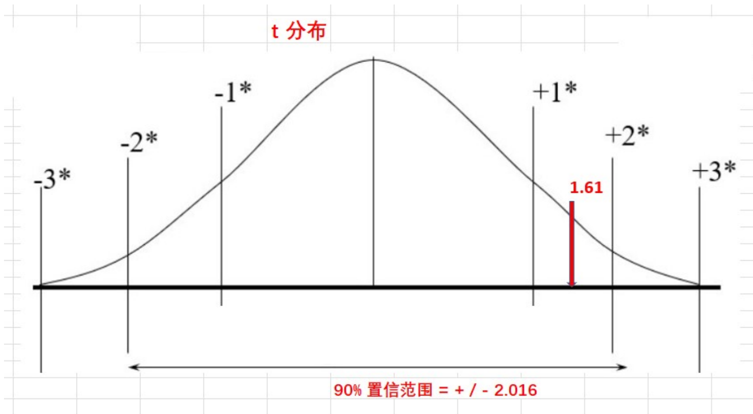
| 对象 Subject | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 前($X_1$) | 210 | 235 | 208 | 190 | 172 | 244 |
| 后($X_2$) | 190 | 170 | 210 | 188 | 173 | 228 |

1. 零假设：没有变化（$\mathbf{H}_0 : \mu_D = 0$）
2. 备选假设：有显著变化（$\mathbf{H}_1 : \mu_D \neq 0$）

注意：与上面大学男女生的各 50 个数据成一组，两组 50 个数之间没有对应关系

这例子不一样，因前后是一一对应，我们需要用结对 T 检验 (Paired T test)。因标准差不是已知，是从样本估算，所以要用 T 检验，非 Z 检验。

可以使用相关方程式，计算出 T 值 =1.61 ,




从上图看到 1.61 落在 90% 置信区间 (2.016) 之内，所以不能拒绝零假设。

也可以利用工具，计算双样本 T 检验 (Paired T test) 的 P 值，因 P 值 =0.165, 大于 0.05，判断不能拒绝零假设

如果想比较比较超过两组数据便要用方差分析 (ANOVA)

## 方差分析 (ANOVA test)

例子一

比较三种车型（小轿车，舒适型，豪华车）的省油系数有没有显著差异？

Worksheet 5 ***

| ↓ | C1 | C2 | C3 |
|---|---|---|---|
| | small | sedans | luxury |
| 1 | 36 | 43 | 29 |
| 2 | 44 | 35 | 25 |
| 3 | 34 | 30 | 24 |
| 4 | 35 | 29 | |
| 5 | | 40 | |
| 6 | | | |

用统计工具跑单因子方差分析得出 P 值是 0.0385，低于 0.05，所以我们便可以拒绝零假设

**One-way ANOVA: small, sedans, luxury**

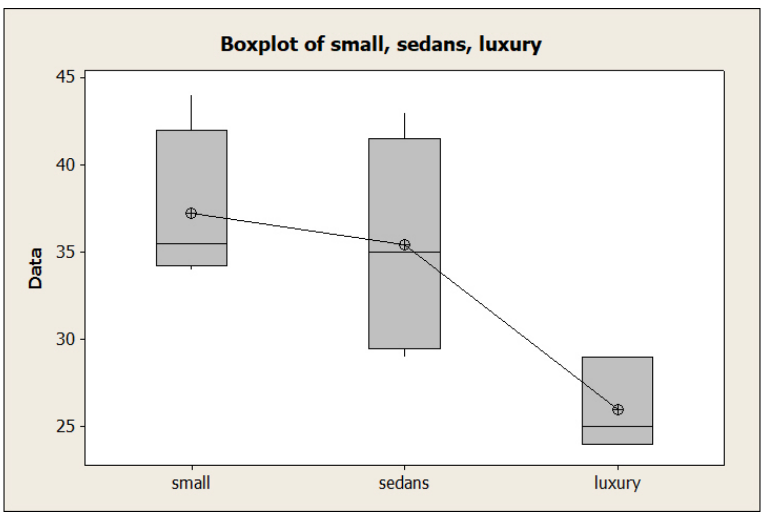
```
Source  DF     SS     MS     F      P
Factor   2  242.7  121.4  4.83  0.038
Error    9  226.0   25.1
Total   11  468.7

S = 5.011   R-Sq = 51.79%   R-Sq(adj) = 41.08%


                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N    Mean  StDev  --------+---------+---------+---------+-
small   4  37.250  4.573                      (---------*---------)
sedans  5  35.400  6.107                    (-------*-------)
luxury  3  26.000  2.646  (----------*----------)
                          --------+---------+---------+---------+-
                               24.0      30.0      36.0      42.0

Pooled StDev = 5.011
```

从下面工具的简单画图的也看到豪华型的系数范围比其他两种有显著差异：




方差分析只能识别有没有显著差异，但若要要明确哪些地方有差异，便要用 Tukey-Kramer 做多组之间比较

方差分析 **2 (ANOVA test)**

某降落伞工厂想比较 4 家布料供货商的布料拉力（因拉力不足会影响使用者的性命安全）
下面是它们随机抽样的结果：

| 布料供货商(Supplier) | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 随机样本拉力 | 18.5 | 26.3 | 20.6 | 25.4 |
| | 24.0 | 25.3 | 25.2 | 19.9 |
| | 17.2 | 24.0 | 20.8 | 22.6 |
| | 19.9 | 21.2 | 24.7 | 17.5 |
| | 18.0 | 24.5 | 22.9 | 20.4 |
| 样本均值 Sample Mean | 19.52 | 24.26 | 22.84 | 21.16 |
| 样本标准差 Sample Standard Deviation | 2.69 | 1.92 | 2.13 | 2.98 |

用工具跑方差分析 P 值是 0.042，少于 0.05，判断拒绝零假设

- Minitab 出来方差分析结果表，与上面的总结表能一一对应；

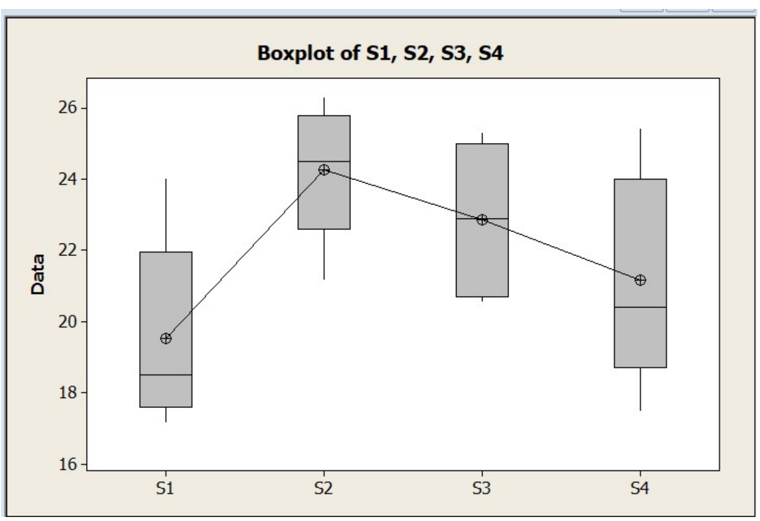
**One-way ANOVA: S1, S2, S3, S4**

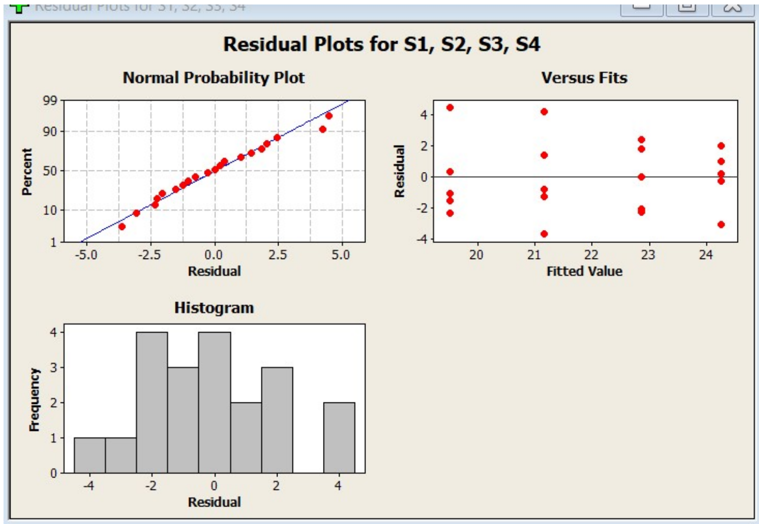
```
Source  DF      SS     MS     F      P
Factor   3   63.47  21.16  3.45  0.042
Error   16   97.98   6.12
Total   19  161.45

S = 2.475   R-Sq = 39.31%   R-Sq(adj) = 27.93%


                         Individual 95% CIs For Mean Based on Pooled StDev
Level  N    Mean  StDev   -+---------+---------+---------+--------
S1     5  19.520  2.690   (--------*--------)
S2     5  24.260  1.919                     (--------*--------)
S3     5  22.860  2.162               (--------*---------)
S4     5  21.160  2.984         (---------*--------)
                          -+---------+---------+---------+--------
                         17.5      20.0      22.5      25.0

Pooled StDev = 2.475
```






- 从 4 组的分布，你可能看出第一组与第二组的差异很明显，其他差不多。但当差异没有这么明显，便需要用统计分析判断。

**Tukey-Kramer** 多组比较

实例（比较 4 供货商布料的拉力）:
#$|\overline{X_1} - \overline{X_2} = |19.52 - 24.26| = 4.74$

1. $|\overline{X_1} - \overline{X_3} = |19.52 - 22.84| = 3.32$
2. $|\overline{X_1} - \overline{X_4} = |19.52 - 21.16| = 1.64$
3. $|\overline{X_2} - \overline{X_3} = |24.26 - 22.84| = 1.42$
4. $|\overline{X_2} - \overline{X_4} = |24.26 - 21.16| = 3.10$
5. $|\overline{X_3} - \overline{X_4} = |22.84 - 21.16| = 1.68$

用 Tukey-Kramer 相关方程式（详见附件），得出临界范围为 4.47，所以判断唯一供货商 1 与 2 的布料拉力是有显著差异。

## 分组（分类）数据

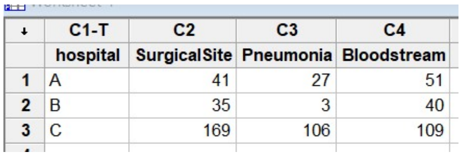
上面都是用于连续数据的例子

假设检验也可以用于分组（分类）数据

**卡方检验 (Chi Square test)**

想研究三家医院针在三类病人感染的主因:（外科手术 Surgical Site ，肺炎 Pneumonia，血液 Bloodstream）最终导致病人死亡

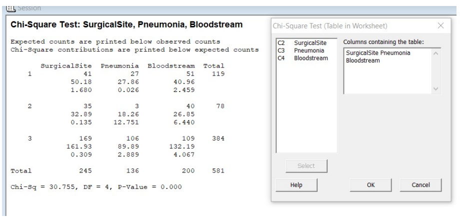
是否有显著差异?



| ↓ | C1-T | C2 | C3 | C4 |
|---|---|---|---|---|
| | hospital | SurgicalSite | Pneumonia | Bloodstream |
| 1 | A | 41 | 27 | 51 |
| 2 | B | 35 | 3 | 40 |
| 3 | C | 169 | 106 | 109 |

1. 零假设：那家医院与感染主因没有关系
2. 备选假设：有显著关系

卡方检验得出 P 值等于零，所以可以拒绝零假设，医院与感染主因有关系



Session

**Chi-Square Test: SurgicalSite, Pneumonia, Bloodstream**

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

      SurgicalSite  Pneumonia  Bloodstream  Total
    1           41         27           51    119
             50.18      27.86        40.96
             1.680      0.026        2.459

    2           35          3           40     78
             32.89      18.26        26.85
             0.135     12.751        6.440

    3          169        106          109    384
            161.93      89.89       132.19
             0.309      2.889        4.067

Total          245        136          200    581

Chi-Sq = 30.755, DF = 4, P-Value = 0.000
```



1: 零假设: 医院 (Hospitals) 与感染种类 (Infections) 之间没有相关

2: 为了计算卡方值 (Chi-square)，首先使用以下公式计算每个列联表的预计值 E(Expected Value)，

$E = \frac{(row\ \ sum)*(column\ \ sum)}{grand\ \ total}$

再用下面公式从预计值与实际值 O(Observed value), 计算卡方值：

$\chi^2 = \sum \frac{(O-E)^2}{E}$

3: 计算得出卡方值 = 30.696 （注），比关键值 9.5 高

（对应 DF=4 (=(3-1)*(3-1)) ，$\alpha = 0.05$，从 $\chi^2$ 统计表得出 9.5)，

所以拒绝零假设，医院 (Hospitals) 与感染种类 (Infections) 之间相关。

（注：对应 Minitab 工具自动算出的 30.755）

# 附件

## 检验单个正态整体均值相关方程式

平均值分布的方 = 样本方差 / n ，并可以用以下方程式估算（如果样本数 n 越大，样本均值的方差就越少)：

$$Z_n = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma}$$

$\overline{X} = 287.6$

$$\frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma}=\frac{\sqrt{10}(287.6-285)}{4}=2.06$$

## 比较两个正态总体的均值

男生：$\overline{X_1}=8.6$ ；$\overline{\sigma_1}=3.3$
女生：$\overline{X_2}=7.9$ ；$\overline{\sigma_2}=3.3$

$$\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{(8.6-7.9)-0}{\sqrt{\frac{(3.3)^2}{50}+\frac{(3.3)^2}{50}}}=1.06$$

## 检验单个正态总体的均值, 方差未知

$\overline{D}=\sum D/n=100/6=16.7$
$(D=X_1-X_2\ )$
$\mathbf{H}_0$ 的拒绝域为：
$\frac{\sqrt{n}(\overline{D}-\mu_0)}{S_D}>t_\alpha(n-1)$

- 查 student-t 分布表，$t_{0.05}(5)=2.016$
- 从样本估计方差

$$s_D^2=\frac{\sum(D_i-\overline{D})^2}{(n-1)}=\frac{\sum D_i^2-n\times\overline{D}^2}{(n-1)}=\frac{4890-6\times(16.7)^2}{(6-1)}=645$$
$s_D=25.4$

- 计算得：

$t=\frac{\sqrt{n}(\overline{D}-\mu_0)}{S_D}=\frac{\sqrt{6}(16.7-0)}{25.4}$
$=1.61$

## **Tukey-Kramer** 相关方程式

- 计算显著差异的临界值：

Critical range= $Q_u\sqrt{\frac{MSW}{2}(\frac{1}{n_j}+\frac{1}{n_j'})}$

对应布料拉力例子：
$\alpha=0.05$，c = 4 , n - c= 20-4 = 16
从统计表得出对应 $Q_u=4.05$ ;
MSW =6.1 (如何计算 MSW, 详见下面” 如何计算 MSW”)

代入上面方程式：

Critical range= $4.05\sqrt{(\frac{6.1}{2})(\frac{1}{5}+\frac{1}{5})}$

$= 4.47$

如何计算 **MSW**，参考 **ANOVA** 方程式

- 总离差平方和 Total Variation = Sum of Squares for Total (SST)：
- SST 由组内离差平方和 (SSW)，和组间离差平方和 (SSA)，双加出来：
- 如果组间离差远远大于组内离差，表示组间差异大，应拒绝零假设；反之，如差异不大，便难以拒绝零假设，没有显著差异。
- 单因素方差分析计算 F = MSA / MSW，如果 F 值大于临界值，拒绝零假设

统计工具，如 Minitab，能直接算出下面总结报告：

| 原因 | 自由度 | 离差平方和 (Sum of Squares) | (Mean Square) | F |
|---|---|---|---|---|
| 组间 | c - 1 | SSA | MSA = SSA/(c-1) | MSA/MSW |
| 组内 | n - c | SSW | MSW = SSW /(n-c) | |
| 总 | n - 1 | SST | | |

相关方程式

* 总离差平方和

SST $= \sum_{j=1}^{c}\sum_{i=1}^{n_j}(X_{ij}-\overline{\overline{X}})^2$

SST = $\sum_{j=1}^{c}\sum_{i=1}^{n_j}(X_{ij}-\overline{\overline{X}})^2$

注：总体平均数 $\overline{\overline{X}} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n_j}X_{ij}}{n}$

- 组间离差平方和

SSA = $\sum_{j=1}^{c} n_j(\overline{X_j}-\overline{\overline{X}})^2$

注：组 j 的平均值 $\overline{X_j}$

- 组内离差平方和

SSW = $\sum_{j=1}^{c}\sum_{i=1}^{n_j}(X_{ij}-\overline{X_j})^2$

注：组 j 的平均值 $\overline{X_j}$

- 单因素方差分析均方和

MSA = $\frac{SSA}{c-1}$

MSW = $\frac{SSW}{n-c}$

MST = $\frac{SST}{n-1}$

实例（比较 4 供货商布料的拉力）：

- 组间自由度 = c - 1 = 4 - 1 =3
- 组内自由度 = n - c = 20 - 4 =16
- 总自由度 = n - 1 = 20-1 = 19

总体平均数 $\overline{\overline{X}} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n_j} X_{ij}}{n} = \frac{438.9}{20} = 21.945$

SSA $= \sum_{j=1}^{c} n_j(\overline{X_j} - \overline{\overline{X}})^2$

$$= (5)(19.52-21.945)^2 + (5)(24.26-21.945)^2 + (5)(22.84-21.945)^2 + (50(21.16-21.945)^2$$

$$= 63.285$$

SSW $= \sum_{j=1}^{c}\sum_{i=1}^{n_j}(X_{ij} - \overline{X_j})^2$

$$= (18.5 - 19.52)^2 + .. + (18.0 - 19.52)^2 + (26.3 - 24.26)^2 + .. + (24.5 - 24.26)^2$$

$$+(25.4 - 21.16)^2 + .. + (20.4 - 21.16)^2$$

$$= 97.5$$

SST $= \sum_{j=1}^{c}\sum_{i=1}^{n_j}(X_{ij} - \overline{\overline{X}})^2$

$$= (18.5 - 21.945)^2 + (24 - 21.945)^2 + .. + (20.4 - 21.945)^2$$

$$= 160.7895$$

MSA $= \frac{SSA}{c-1} = \frac{63.285}{4-1} = 21.1$
MSW $= \frac{SSW}{n-c} = \frac{97.5}{16} = 6.1$
F = MSA / MSW = 21.1 / 6.1 = 3.46
能对应文中 Minitab 出来的结果：
MSA = 21.16, MSW = 6.12, F = 3.45

# Part VII

# 总结