



# 基本统计分析知识点回顾

## ➤ 基础知识点回顾：

在数据被组织成合适的形式后，开始使用图形探索数据，而下一步通常就是使用数值描述每个变量的分布，接下来则是两两探索所选择变量之间的关系。

回顾接下来用到了，今天要实践的知识点：

描述性统计分析

频数表（列联表）

独立性检验

相关性检验

t检验

方差分析

组间差异的非参数检验

# 实训项目

**数据1介绍:**  
数据是关于一些医院的基础信息，数据包含13个字段，包含的字段分别为： 供应商编号、医院名、地址1、地址2、地址3、所处城市、所处州、邮编、所处乡镇、电话、医院类型、医院所有权和是否提供急救服务。数据显示如下：

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Provider Number	Hospital Name	Address 1	Address 2	Address 3	City	State	ZIP Code	County	Phone Number	Hospital Type	Hospital Ownership	Emergency Services	
10001	SOUTHEAST ALABAMA	1108 ROSS CLARK CIRCLE			DOTHAN	AL	36301	HOUSTON	3347938701	Acute Care Hospitals	Government - Hospital	Yes	
10005	MARSHALL MEDICAL	2505 U S HIGHWAY 431 NORTH			BOAZ	AL	35957	MARSHALL	2565938310	Acute Care Hospitals	Government - Hospital	Yes	
10006	ELIZA COFFEE MEMORIAL	205 MARENGO STREET			FLORENCE	AL	35631	LAUDERDALE	2567688400	Acute Care Hospitals	Government - Hospital	Yes	
10007	MIZELL MEMORIAL	702 N MAIN ST			OPP	AL	36467	COVINGTON	3344933541	Acute Care Hospitals	Voluntary non-profit	Yes	
10008	CRENSHAW COMMUNITY	101 HOSPITAL CIRCLE			LUVERNE	AL	36049	CRENSHAW	3343353374	Acute Care Hospitals	Proprietary	Yes	
10009	HARTSELLE MEDICAL	201 PINE STREET NORTHWEST			HARTSELL	AL	35640	MORGAN	2567736511	Acute Care Hospitals	Proprietary	Not Available	
10010	MARSHALL MEDICAL	8000 ALABAMA HIGHWAY 69			GUNTERSVILLE	AL	35976	MARSHALL	2565718000	Acute Care Hospitals	Government - Hospital	Yes	
10011	ST VINCENT'S EAST	50 MEDICAL PARK EAST DRIVE			BIRMINGHAM	AL	35235	JEFFERSON	2058383122	Acute Care Hospitals	Voluntary non-profit	Yes	
10012	DEKALB REGIONAL	200 MED CENTER DRIVE			FORT PAYNE	AL	35968	DE KALB	2568453150	Acute Care Hospitals	Proprietary	Yes	
10015	SOUTHWEST ALABAMA	33700 HIGHWAY 43			THOMASVILLE	AL	36784	CLARKE	3346366221	Acute Care Hospitals	Proprietary	Not Available	
10016	SHELBY BAPTIST MEDICAL	1000 FIRST STREET NORTH			ALABASTER	AL	35007	SHELBY	2056208100	Acute Care Hospitals	Voluntary non-profit	Yes	
10018	CALLAHAN EYE FOUNDATION	1720 UNIVERSITY BLVD			BIRMINGHAM	AL	35233	JEFFERSON	2053258100	Acute Care Hospitals	Voluntary non-profit	Yes	
10019	HELEN KELLER MEMORIAL	1300 SOUTH MONTGOMERY AVENUE			SHEFFIELD	AL	35660	COLBERT	2563864556	Acute Care Hospitals	Government - Hospital	Yes	
10021	DALE MEDICAL CENTER	126 HOSPITAL AVE			OZARK	AL	36360	DALE	3347742601	Acute Care Hospitals	Government - Hospital	Yes	
10022	CHEROKEE MEDICAL	400 NORTHWOOD DR			CENTRE	AL	35960	CHEROKEE	2569275531	Acute Care Hospitals	Voluntary non-profit	Yes	
10023	BAPTIST MEDICAL	2105 EAST SOUTH BOULEVARD			MONTGOMERY	AL	36116	MONTGOMERY	3342882100	Acute Care Hospitals	Voluntary non-profit	Yes	
10024	JACKSON HOSPITAL	1725 PINE STREET			MONTGOMERY	AL	36106	MONTGOMERY	3342938000	Acute Care Hospitals	Voluntary non-profit	Yes	
10025	GEORGE H. LANIER	4800 48TH ST			VALLEY	AL	36854	CHAMBERS	3347561400	Acute Care Hospitals	Voluntary non-profit	Yes	
10027	ELBA GENERAL HOSPITAL	987 DRAYTON STREET			ELBA	AL	36323	COFFEE	3348972257	Acute Care Hospitals	Government - Hospital	Yes	
10029	EAST ALABAMA MEDICAL	2000 PEPPERELL PARKWAY			OPELIKA	AL	36801	LEE	3347493411	Acute Care Hospitals	Government - Hospital	Yes	
10032	WEDOWEE HOSPITAL	209 NORTH MAIN STREET			WEDOWEE	AL	36278	RANDOLPH	2563572111	Acute Care Hospitals	Government - Local	Yes	
10033	UNIVERSITY OF ALABAMA	619 SOUTH 19TH STREET			BIRMINGHAM	AL	35233	JEFFERSON	2059344011	Acute Care Hospitals	Government - State	Yes	
10034	COMMUNITY HOSPITAL	805 FRIENDSHIP ROAD			TALLASSEE	AL	36078	ELMORE	3342836541	Acute Care Hospitals	Voluntary non-profit	Yes	
10035	CULLMAN REGIONAL	1912 ALABAMA HIGHWAY 157			CULLMAN	AL	35058	CULLMAN	2567372000	Acute Care Hospitals	Government - Hospital	Yes	
10036	ANDALUSIA REGIONAL	849 SOUTH THREE NOTCH STREET			ANDALUSIA	AL	36420	COVINGTON	3342228466	Acute Care Hospitals	Proprietary	Yes	
10038	STRINGFELLOW MEMORIAL	301 EAST 18TH ST			ANNISTON	AL	36201	CALHOUN	2562358900	Acute Care Hospitals	Proprietary	Yes	
10039	HUNTSVILLE HOSPITAL	101 SIVLEY RD			HUNTSVILLE	AL	35801	MADISON	2562651000	Acute Care Hospitals	Government - Hospital	Yes	

# 实训项目

## 数据2介绍:

选取人类死亡率数据库（HMD，2007），提供了1951-2005年瑞典人口信息。这是个多变量的数据集，变量描述如下表：

变量名称	描述
Year	年份：1951,1952，...,2005
Age	年龄：0,1,...,109
Female_Exp	女性生存人口数
Male_Exp	男性生存人口数
q_female	女性死亡率
q_male	男性死亡率
Female_death	女性死亡人数
Male_death	男性死亡人数
L_female_exp	对数女性生存人口数
L_male_exp	对数男性生存人口数

# 实训项目

## 数据3介绍:

该数据是记录超过4000家医院关于心脏病、心力衰竭、肺炎的30天内的死亡率和重入院率。（这份数据变量解释，可参照Hospital Revised Flatfiles.pdf中的Outcome of Care Measures.csv的变量请参考第17页编号19），数据显示如下：

Provider	Hospital	Address	Address	Address	City	State	ZIP Code	County	Phone	Nur	Hospital	Comparis	Lower	Moi	Upper	Moi	Number	of	Footnote	Hospital	Comparis	Lower	Moi	Upper	Moi	Number	of	Footnote	Hospital	C	
10001	SOUTHEAS	1108 ROSS CLARK CIRCLE			DOTHAN	AL	36301	HOUSTON	3.35E+09		14.3	No Differ	12.1		17		666				11.4	No Differ	9.5		13.7		741			10.9	N
10005	MARSHALL	2505 U S HIGHWAY 431 NORTH			BOAZ	AL	35957	MARSHALL	2.57E+09		18.5	No Differ	14.7		23		44				15.2	Worse th	12.2		18.8		234			13.9	N
10006	ELIZA CO	205 MARENGO STREET			FLORENCE	AL	35631	LAUDERDAI	2.57E+09		18.1	No Differ	14.8		21.8		329				11.3	No Differ	9.1		13.9		523			13.4	N
10007	MIZELL MI	702 N MAIN ST			OPP	AL	36467	COVINGTON	3.34E+09		Not Avail	Number of	Not Avail		Not Avail		14	number of			13.6	No Differ	10		18.2		113			14.9	N
10008	CRENSHAW	101 HOSPITAL CIRCLE			LUVERNE	AL	36049	CRENSHAW	3.34E+09		Not Avail	Number of	Not Avail		Not Avail		9	number of			13.8	No Differ	9.9		18.7		53			15.8	N
10010	MARSHALL	8000 ALABAMA HIGHWAY 69			GUNTERSVI	AL	35976	MARSHALL	2.57E+09		Not Avail	Number of	Not Avail		Not Avail		22	number of			12.5	No Differ	9.9		15.6		163			8.7	B
10011	ST VINCE	50 MEDICAL PARK EAST DRIVE			BIRMINGH	AL	35235	JEFFERSON	2.06E+09		17.7	No Differ	14.2		22.1		220				10.9	No Differ	8.5		13.8		348			16.2	W
10012	DEKALB RI	200 MED CENTER DRIVE			FORT PAY	AL	35968	DE KALB	2.57E+09		18	No Differ	14.1		22.8		99				16.6	Worse th	12.4		22		137			15.8	W

# 实训项目

## 题目1:

- 1.通过读取“hospital-data.csv”将数据保存到df中，获取该数据的条数；查看数据中的前5条数据。
- 2.查看数据概况；求得邮编的范围。
- 3.我们默认电话号码是个数值，没有实际意义；应用sapply函数，通过调用用户自定义函数，返回电话号码的最大值，最小值，均值，中位数，标准差和方差。
- 4.利用aggregate求取各州的电话号码的中位数。
- 5.利用by求取各城市的电话号码的最大值和最小值；显示结果的前3条数据。
- 6.对所属州生成简单的频数统计表；并将这个频数表转化为比例值。
- 7.建立所属州和医院类型的二维列联表，命名为mycontable；按列生成边际和。
- 8.利用CrossTable建立所属乡镇和是否提供急救服务字段的二维列联表，命名为mycrosstable。（注：安装gmodels包）

# 实训项目

## 题目2:

- 1.通过读取文件death rate.csv获取数据保存到death中, 通过卡方检验来检验年龄与男性生存人口数（二级列联表）之间是否独立。
- 2.通过assocstats函数来度量年龄与男性的死亡率（二级列联表）之间的相关性。
- 3.计算年龄与男性的死亡率之间的Pearson和Spearman相关系数;和death中所有变量的协方差。
- 4.检验女性生存人口数和男性生存人口数之间的相关性的显著性。



# 实训项目

## 题目3:

1.通过读取文件outcome-of-care-measures.csv获取数据保存到care\_df中,通过写一个名为 best的函数, 找到一个州中最好的医院, 函数有两个参数, 一个是有两个字母缩写的州名称, 另一个是结果名称, 包括 (heart attack, heart failure, pneumonia) , 函数的返回结果为30天死亡率最低的医院名称。在处理排名的时候, 如果出现死亡率相等的情况, 将医院按照字母表顺序排序, 取第一家医院。

**注意检查读入数据的类型, 以免排序的时候出现问题!!**

函数模板如下

```
best <- function(state, outcome) {  
##code  
}
```

函数要检查输入变量的有效性, 如果输入一个无效的州名, 函数应该停止程序, 并提示" invalid state", 同样当输入的outcome参数有误时, 应提示" invalid outcome" 。

函数检验:

```
best("TX", "heart failure")  
best( "MD" , "heart attack" )  
best("MD", "pneumonia")  
best("BB", "heart attack")  
best("NY", "hert attack")
```

**要求提交的文件包含代码和输出结果或是word文档**



# THANKS

