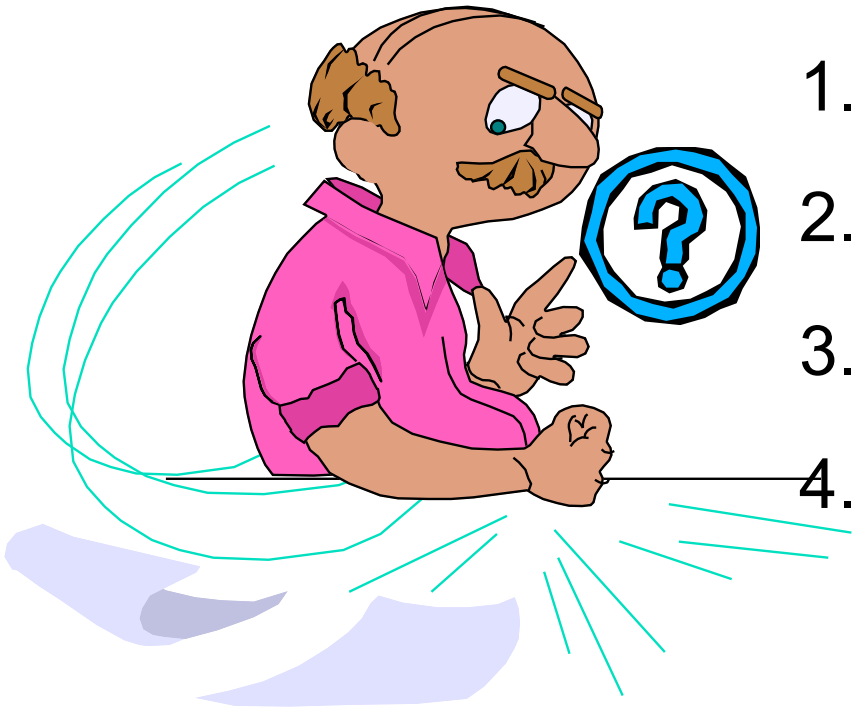
统计学概念

收集、分析、表述和解释数据的科学



1. 数据收集：取得数据
2. 数据分析：分析数据
3. 数据描述：图表展示数据
4. 数据解释：结果的说明

计数
资料

1 总体与样本

总体：(population) 根据研究目的所确定的同质观察单位的全体

例1　调查2004年某市7岁正常男童的身高

总体：该地2004年全体7岁男童的身高

同质：　(homogeneity)：事物的性质、影响条件或背景相同或非常相近
变异：　(variation)：同质个体间的差异

例1　调查2004年某市7岁正常男童的身高

同质：2004年、某市、7岁男童

变异：身高相同

1 总体与样本

样本：(sample)从总体中随机抽取的部分观察单位

2变量和资料

变量： **(variable) 每个观察单位的某项特征**

资料： **(data) 由变量值构成**

婴儿生长发育对照表

身高（cm）

		出生	一周	二周	三周	四周	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月	十五月	十八月	二岁	二岁三月	二岁六月	二岁九月	三岁	三岁六月
实测																									
应增值	男		0.5	1.1	0.9	1.4	3.4	3.1	2.7	2.2	2.0	1.4	1.4	1.2	1.3	1.0	1.4	2.8	2.8	2.0	2.4	2.2	2.1	2.0	2.2
	女		0.9	1.2	1.0	1.4	3.5	3.0	2.8	2.4	1.9	1.6	1.2	1.0	1.2	1.0	1.4	2.4	2.6	2.4	2.4	2.6	2.2	2.2	1.8
实际值	男	50.8	51.0	51.8	53.0	54.4	57.9	61.6	64.6	66.9	69.0	70.0	72.3	72.8	74.3	75.2	76.2	79.4	82.5	84.6	87.2	90.0	92.2	93.9	96.3
	女	49.8	50.1	52.2	52.3	53.8	57.6	59.9	62.6	64.6	67.2	68.6	69.8	71.0	72.0	73.3	74.6	77.8	81.4	84.0	86.0	89.4	91.7	93.6	95.7

体重（kg）

		出生	一周	二周	三周	四周	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月	十五月	十八月	二岁	二岁三月	二岁六月	二岁九月	三岁	三岁六月
实测																									
应增值	男		0.04	0.21	0.30	0.45	1.22	1.03	0.77	0.58	0.55	0.40	0.44	0.40	0.43	0.34	0.31	0.55	0.45	0.75	0.83	0.66	0.47	0.52	0.80
	女		-0.16	0.24	0.34	0.37	1.24	0.83	0.75	0.60	0.56	0.44	0.45	0.34	0.44	0.35	0.38	0.57	0.49	0.52	0.66	0.58	0.46	0.56	0.78
实际值	男	3.39	3.39	3.55	3.87	4.22	5.54	6.65	7.43	8.00	8.52	8.91	9.33	9.69	10.09	10.35	10.69	11.23	11.56	12.34	13.09	13.63	14.28	14.28	15.0
	女	3.28	3.18	3.40	3.71	3.96	5.17	6.62	6.91	7.58	8.06	8.39	8.83	9.13	9.82	9.82	10.29	10.78	11.02	11.88	12.65	13.43	13.96	14.55	14.8

婴儿的正常生理指标体重：6个月内：出生体重（kg）+月龄X0.7
7-12个月：6个月体重+(月龄-6)X0.44
2岁体重:4倍于出生体重
2-12岁体重:年龄X2+8kg

婴儿的正常生理指标身高：6个月：65cm
1岁：75cm
2岁：85cm
2-12岁：年龄X7+70cm

昵图网 www.nipic.com

By:你一定要相信 No.20121227191547073394

2变量和资料

资料：**(data)** 由变量值构成

1. 计量资料
2. 计数资料
3. 等级资料

计量资料

计量资料： *(measurement data)*

又称定量资料，用仪器、工具等定量方法对观察单位测量(measure)某指标值所得到的资料。

特点：一般有计量单位 -连续型

如：患者身高(cm)

体重(kg)

血压(mmHg)

脉搏(次/分)

红细胞计数(10^{12}/L)等

计数资料

计数资料：*(counting data)*

又称定性资料（**Qualitative data**），分二分类和多分类，按观察单位品质标志分组，再清点各组的例数所得的资料。

特点：一般无固有计量单位- 离散型

如：心电图检查结果（正常、异常）

性别（男、女）

血型（A、B、O、AB）

职业（工、农、兵）

性别	男	女
例数	90	45

等级资料

等级资料：*(ordinal data)*

又称半定量半定性资料，根据观察单位某指标量的大小，深浅或严重程度分组，得到的各等级组观察单位数。

特点：有大小顺序，故又称有序分类资料-有序型（ordinal category data）。

如 ① 癌症分期：早、中、晚；
② 药物疗效：治愈、好转、无效、死亡；
③ 尿蛋白： －（阴性），±, +, ++, +++及以上；

资料类型例1

住院号	年龄	身高	体重	住院天数	职业	文化程度	分娩方式	妊娠结局
2025655	27	165	71.5	5	无	中学	顺产	足月
2025653	22	160	74.0	5	无	小学	助产	足月
2025830	25	158	68.0	6	管理员	大学	顺产	足月
2022543	23	161	69.0	5	无	中学	剖宫产	足月
2022466	25	159	62.0	11	商业	中学	剖宫产	足月
2024535	27	157	68.0	2	无	小学	顺产	早产
2025834	20	158	66.0	4	无	中学	助产	早产
2019464	24	158	70.5	3	无	中学	助产	足月
2025783	29	154	57.0	7	干部	中学	剖宫产	足月

计量资料

计数资料

资料类型例2

100例高血压患者治疗后的临床记录

NO	group	age	sex	收缩压（kPa）	舒张压（kPa）	心电图	疗效
1	A	37	M	18.67	11.47	正常	治疗
2	C	45	F	20.00	12.53	正常	有效
3	B	43	M	17.33	10.93	异常	有效
4	C	59	F	22.67	14.67	异常	无效
……	……	……	……	……	……	……	……
100	B	54	F	16.80	11.73	正常	有效
	计数资料	计量资料	计数资料	计量资料		计数资料	等级资料

资料类型例3

1. 例：一组20~40岁成年人的血压

计量资料

等级资料

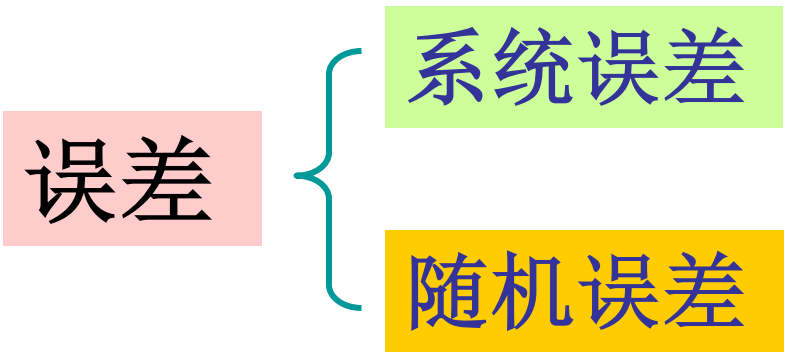
<8	低血压
8~	正常血压
12~	轻度高血压
15~	中度高血压
17~	重度高血压

计数资料

以12kPa为界分为正常与异常两组，统计每组例数

3 误差

误差：（error）指实测值与真值之差



系统误差(systematic error)

a. 仪器标准试剂未经校正；

b. 测量者掌握尺度不同；

c. 测量者的某种感官障碍等原因所导致测量结果呈倾向性偏大或偏小。

特点：有倾向性；可避免

随机误差(random error)

重复测量误差（error of replication)

抽样误差(sampling error):由于抽样所导致样本指标与总体指标的差异（主要由变异引起）

特点：无倾向性；不可避免

4 概率和频率

概率： (probability) 度量随机事件发生可能性大小的一个数值，用大写的P表示；取值[0，1]。

必然事件： P=1

不可能事件： P=0

重要结论：小概率事件很难发生！

4 概率和频率

频率： **(frequency)** 事件实际发生次数与可能发生次数的比率，设在相同条件下，独立重复进行n次试验，事件A出现f 次，则事件A出现的频率为f / n。

概率和频率关系

实验者	投掷次数	出现"正面"次数	频率	概率
1	4040	2048	0.506931	0.5
2	12000	6019	0.501583	0.5
3	24000	12012	0.5005	0.5

频率与概率的关系

- 样本含量*n*越大，波动幅度越小，频率越接近概率；
-