

【Part - 1】

数据分析的基础概念



1: 描述性统计

描述性统计

观察样本基本情况的一种方法。通常考察样本的平均数、中位数（以及四分位数）、众数、几何平均数、调和平均数、方差、标准差、分布等。

目的是从表象和整体观察样本的特征。

描述性统计是数据分析的入门甜点，也是后续高阶分析的基础，需要认真掌握。

描述性统计是这样描述数据，并快速找到切入点的

描述性统计的工具和方法		业务含义	典型问题
中位数		按照从小到大排序，处于中间位置的数	中位数不是平均值
平均值		算数平均值、几何平均值、加权平均值	平均值一般都有偏差
分位数		按照从小到大排序，第x位置的数	-
方差		描述数据离散程度的指标	评估变化是否在正常范围
异常值		Q1-1.5*IQR Q3+1.5*IQR	数学上给出的异常值定义

2: 分析变化

观察变化

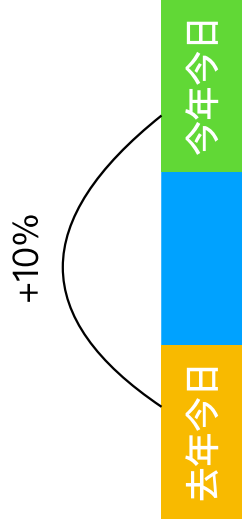
通常情况下，发生变化的指标一般就是指标关联的业务环境发生了某种变化。
目的是通过观察变化量，来寻找可能的业务问题点。

变化的异常情况包括了：下跌，不变，以及上升。

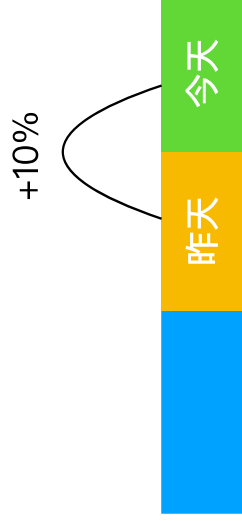
寻找变化，深入观察

评估变化的方法和工具	业务含义	典型问题
同比	对比同期的变化	和上周五比 和去年过年比
环比	对比连续周期	今天和昨天比 本月和上月比 今年和去年比
增长率	评估累计型指标的有力工具 评估业务增长的潜力和趋势	收入不断在涨，增长幅度如何 APP用户不断增长，增长趋势如何

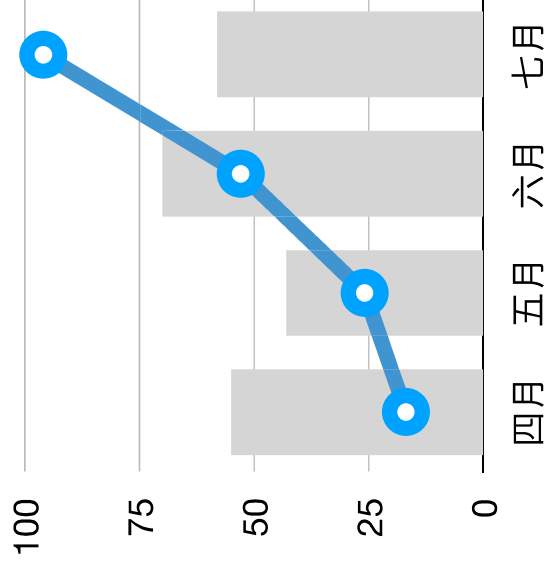
看图说话：评估变化的三个工具



同比 +10%



环比 +10%



增长率

3: 多维分析

多维分析

还记得之前磊叔分享指标体系的时候，除了指标外，还提到了什么吗？

没错，就是维度。

维度是描述指标的不同角度，通过多维度分析，来寻求指标变化的可能原因。

广义的多维分析，不仅包括从指标的不同维度进行分析，也包括拆分为多个子指标进行分析。

常见的维度

分类	常见维度
基础/通用	年龄、性别、学历、地域、手机型号、操作系统
产品	产品类型、归属业务
运营	归属渠道、投放周期、活动类型
营销	市场推广、营销方式、营销目的

3: 常见的分析偏差

维度偏差

维度偏差

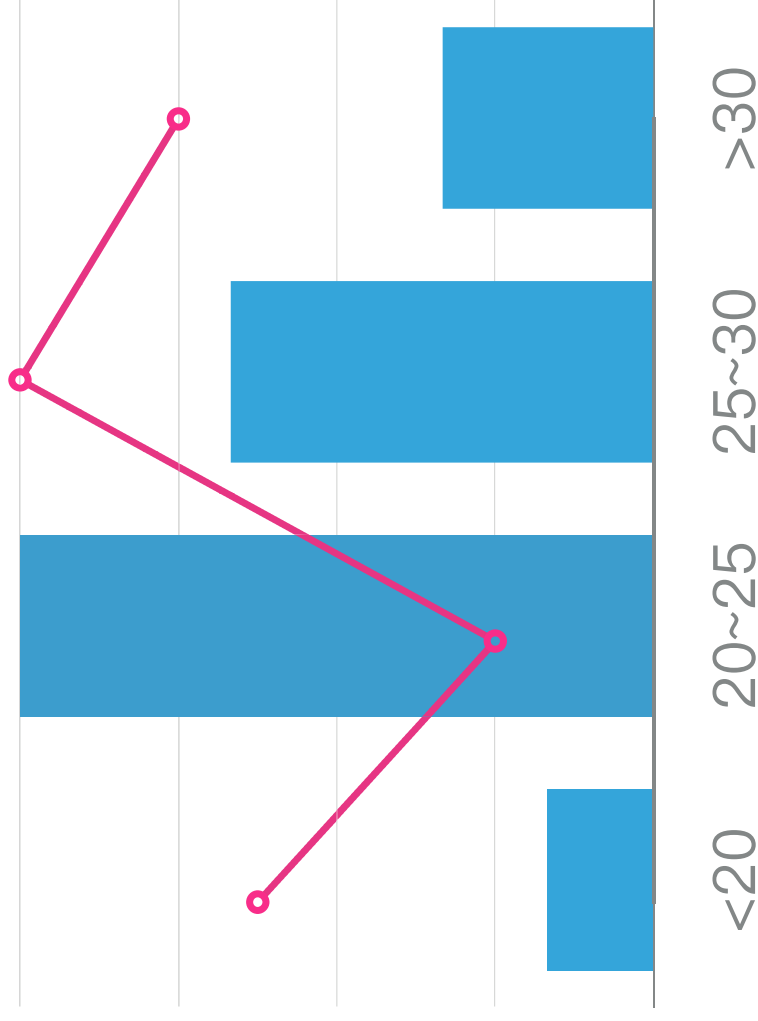
入门级小坑

大数据涉及的维度很多

单一维度分析会出现偏差

多个维度组合起来的时候

可能会得出相反的结论



幸存者偏差

幸存者偏差

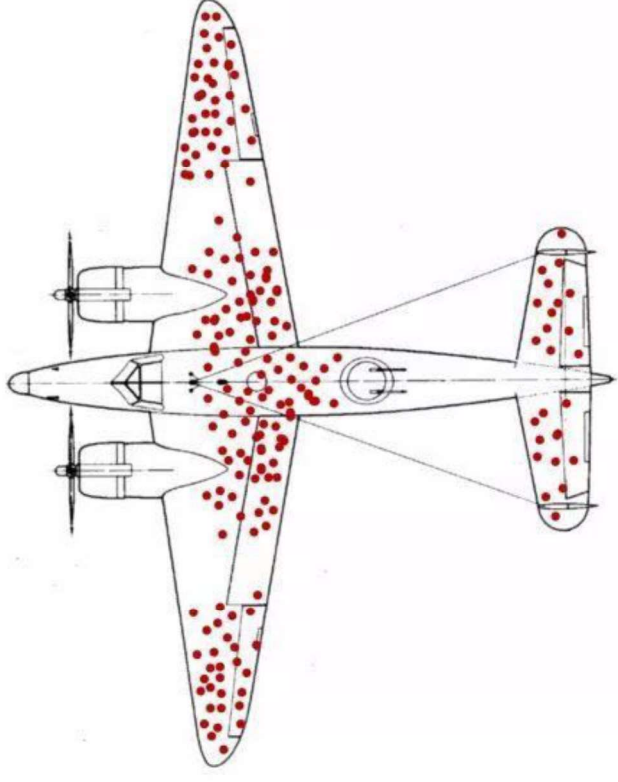
神级大坑

多听失败者的经验

增大观测样本量

一人成功经验是特例，多人就有共性

单次成功有运气，多次成功真功夫



5: 相关性分析

相关性分析

指对两个或多个具备相关性的变量元素进行分析，从而衡量两个变量因素的相关密切程度。相关性的元素之间需要存在一定的联系或者概率才可以进行相关性分析。

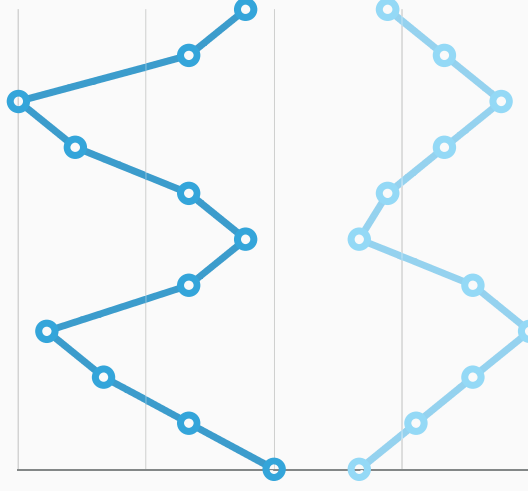
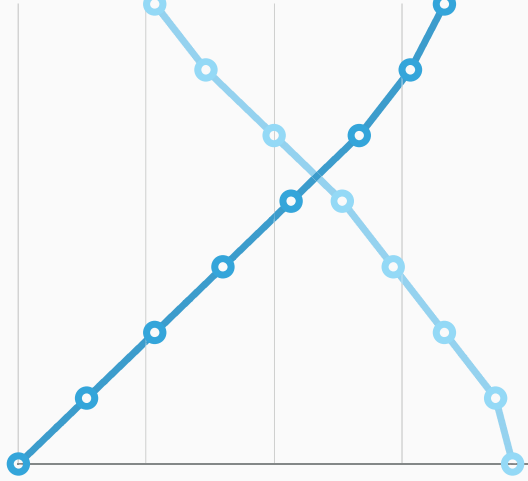
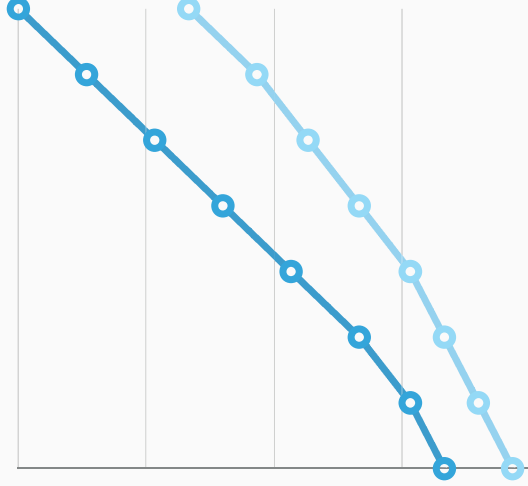
在业务中，通常是为了量化评估各种因素对于核心指标的影响程度，寻找出对业务影响最大的原因。

特别的，相关性分析也可以察觉肉眼难以分辨的变化和关系。

相关性分析

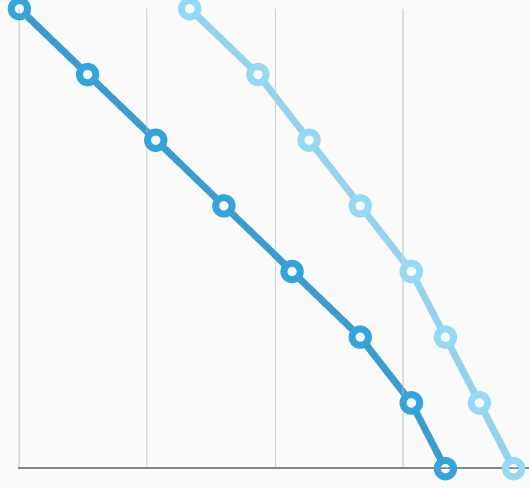
定量确定变量间相互影响的分析方法

用来评估变量的影响大小

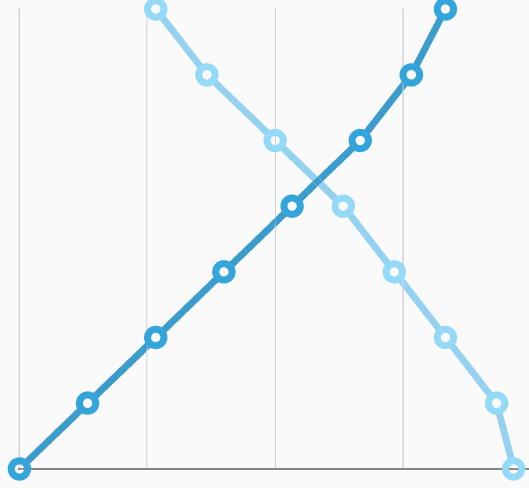


相关性分析

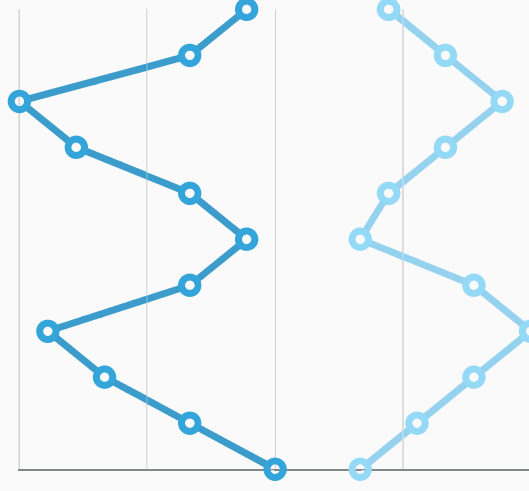
正相关



负相关



还是负相关



6：预测分析

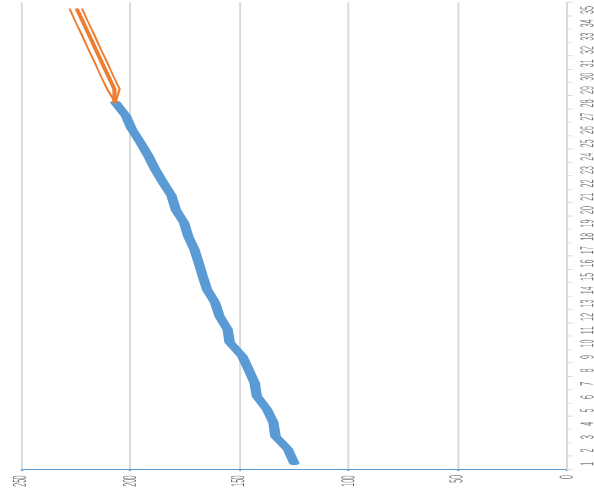
趋势预测

预测分析是一种统计或数据挖掘解决方案，包含可在结构化和非结构化数据中使用以确定未来结果的算法和技术。

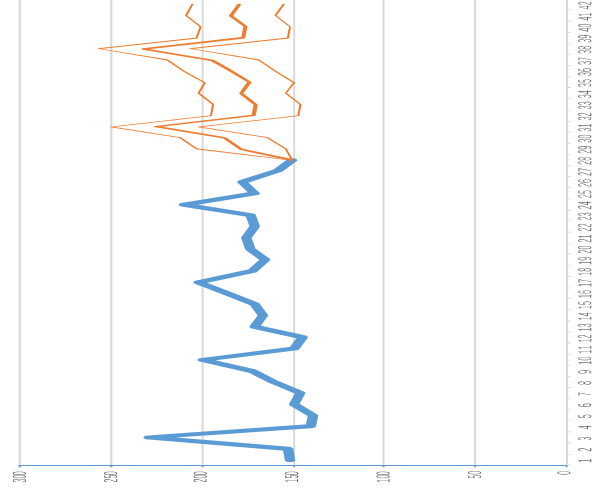
在业务中，通常是为了定性判断核心指标发展趋势。

趋势预测，大概是数据分析领域最迷人的部分了。

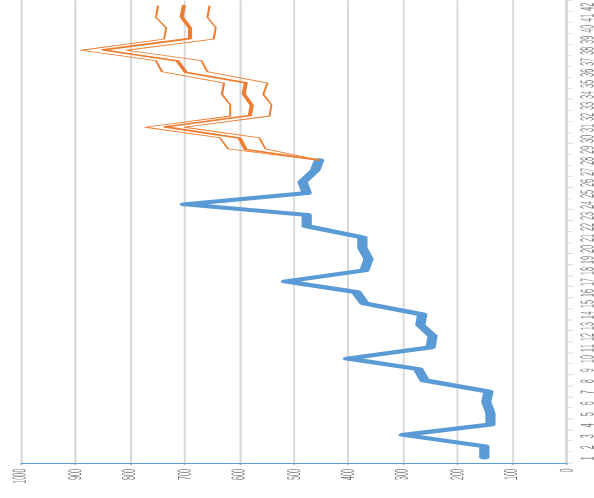
回到未来、趋势预测



一般时间序列预测



季节性时间序列预测



复合型时间序列预测

【Part - 2】

数据分析的常用产品



1：指标和指标体系

行为类			业务类		交易类	
新增访客	活跃访客	活跃访问时段	注册用户数 登录用户数 新增用户数	购买人数 订单人数 购买转化率 复购率 客单价 笔单价		
每日平均访问时长	每日启动次数	留存率				
停留时间	访问页面数					
页面点击渗透率	页面跳出率	页面转化率				
活动参与人数	页面流失率	活动转化率				

2: A/B TEST

A/B Test

A



B

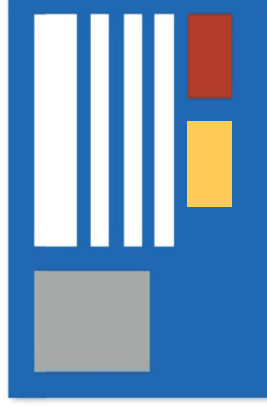


A/B Test

A



B



3: 漏斗分析

—系列 / 连续步骤的 / 转化和流失

两种转化率计算方式

