

54课

随机变量之差的方差

相互独立的随机变量 X 、 Y ，令随机变量 Z 为两者之差，即 $Z=X-Y$ ，那么 Z 的方差就等于 X 和 Y 的方差之和，即 $\text{Var}(Z)=\text{Var}(X)+\text{Var}(Y)$ 。这一节重点讲解了这一性质

X Y independent random variables 網易公开课

$$E(X) = \mu_X \quad E(Y) = \mu_Y$$
$$\text{Var}(X) = E((X - \mu_X)^2) = \sigma_X^2 \quad \text{Var}(Y) = E((Y - \mu_Y)^2) = \sigma_Y^2$$
$$Z = X + Y \quad A = X - Y$$
$$E(Z) = E(X + Y) = E(X) + E(Y) \quad E(A) = E(X - Y) = E(X) - E(Y)$$
$$\mu_Z = \mu_X + \mu_Y \quad \mu_A = \mu_X - \mu_Y$$

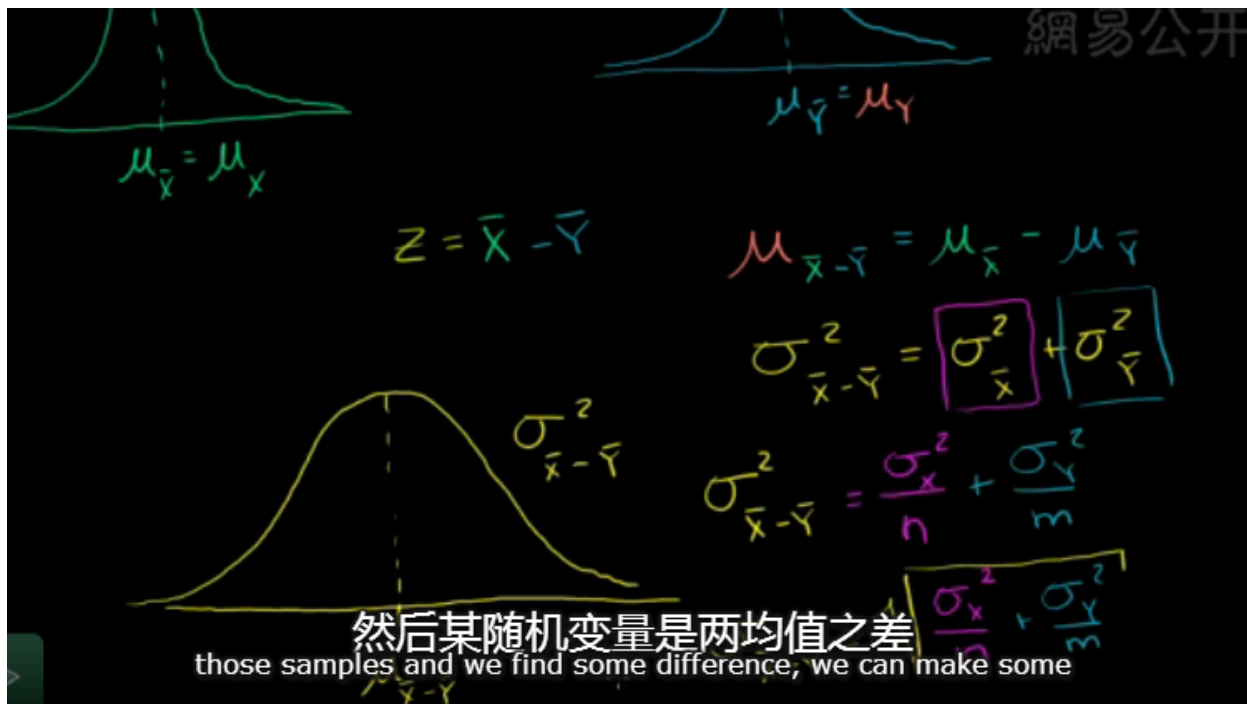
下面再来看看随机变量 Z 和 A 的方差

$$\mu_Z = \mu_X + \mu_Y \quad \mu_A = \mu_X - \mu_Y$$
$$\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y) \quad \sigma_A^2 = \sigma_{X-Y}^2 = \sigma_{X+(-Y)}^2 = \sigma_X^2 + \sigma_{-Y}^2$$
$$\sigma_Z^2 = \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 \quad \sigma_{-Y}^2 = \text{Var}(\underline{-Y}) = E((-Y - E(-Y))^2)$$

55课

样本均值之差的分布

一个随机变量 X 和一个随机变量 Y ，分别抽取样本计算均值得到 \bar{X} 和 \bar{Y} ，令 $Z = \bar{X} - \bar{Y}$ ，于是可以得到统计量 Z 的抽样分布，当样本量足够大时，根据中心极限定理， Z 的抽样分布也近似是正态分布



56课

均值之差的置信区间

这一节紧接着上一节，一个随机变量X和一个随机变量Y，其均值分别为 μ_X 和 μ_Y ，那么 $\mu_X - \mu_Y$ 也可以求出一个置信区间。

58课

均值之差的假设检验

对于某减肥新方法的实验组，分别求其样本均值和方差，另外对普通减肥方法的对照组求样本均值和方差，如何通过假设检验知道这种新方法是否有效呢？这一节讲解这一问题。

We're trying to test whether a new, low-fat diet actually helps obese people lose weight. 100 randomly assigned obese people are assigned to group 1 and put on the low fat diet. Another 100 randomly assigned obese people are assigned to group 2 and put on a diet of approximately the same amount of food, but not as low in fat. After 4 months, the mean weight loss was 9.31 lbs. for group 1 (s=4.67) and 7.40 lbs. (s=4.04) for group 2.

null hypothesis 零假设：低脂节食没有作用，也就是低脂组的总体均值-对照组的总体均值=0

显著性水平为希腊字母alpha，或则零假设threshold（之前某一节课中语）取5%？

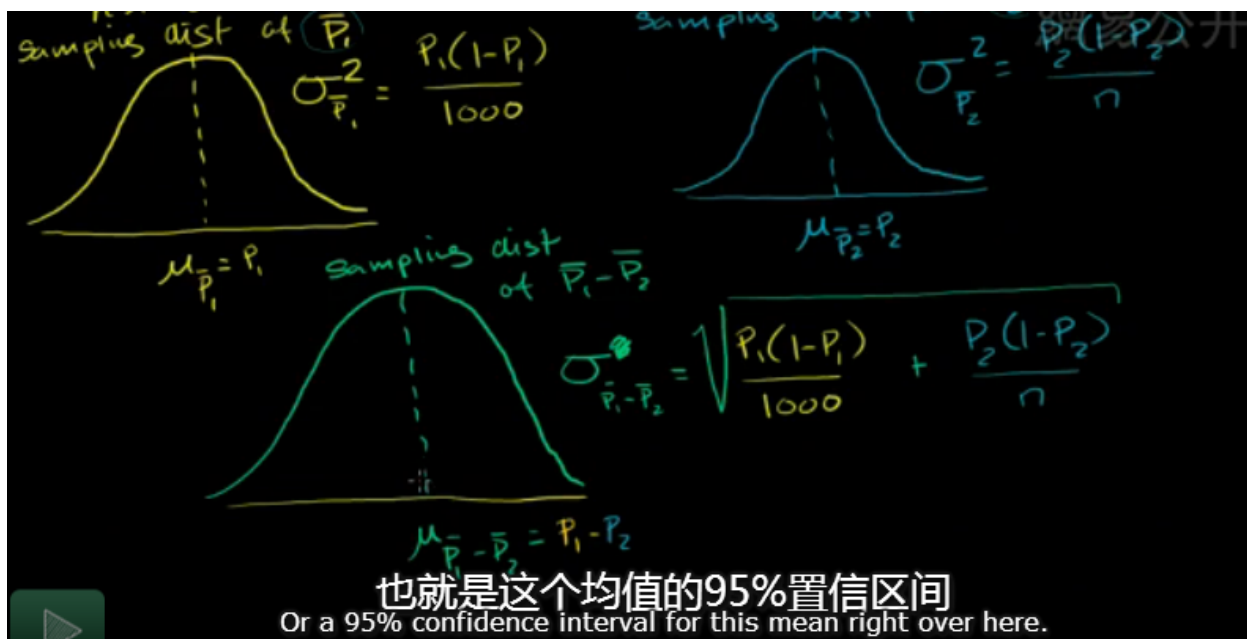
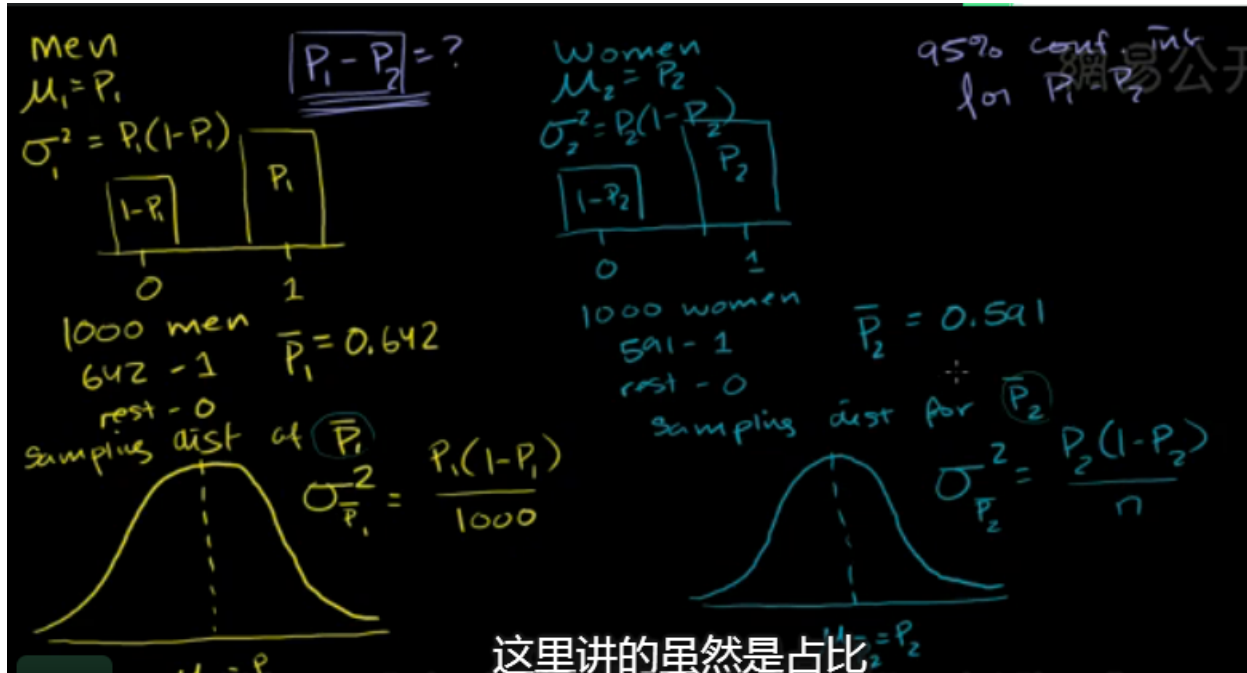
59课

总体占比的比较1

选举时，一部分男性中有 p_1 人投给某候选人，其它人没投给此候选人，女性中有 p_2 人投给此候选人，其它人没投给此候选人。那么如何使用抽样的方法得知男性和女性投给此候选人的占比 p_1 和 p_2 之间有没有差值

呢？这就是这一节所处理的实际问题。这一节列出了式子。

求男女投票是否有显著差异 (meaningful difference)



这里算样本方差为什么不是除以999而是1000呢？

顶[2] 踩[0] 回复 分享

对于泊松分布，方差是可以通过公式 $p*(1-p)$ 计算出来的，所以就不用样本方差去近似了。只有在总体方差不知道的情况下，才用样本方差/($n-1$)来近似得到总体方差。

网易重庆市手机网友 (223.104.*.*)

1

这里算样本方差为什么不是除以999而是1000呢？

[顶\[2\]](#) [踩\[0\]](#) [回复](#) [分享](#)

这里除以1000算的是样本均值抽样分布的方差，根据中心极限定理，这个方差等于总体方差/样本容量

所以关键是总体方差是什么，视频中直接用样本方差近似为总体方差。

61课

总体占比比较的假设检验

选举时，一部分男性中有 p_1 人投给某候选人，其它人没投给此候选人，女性中有 p_2 人投给此候选人，其它人没投给此候选人。那么如何使用抽样的方法得知男性和女性投给此候选人的占比 p_1 和 p_2 之间有没有差值呢？之前两节计算了置信区间，这一节将直接从假设检验角度审视这一问题。

网易陕西省西安手机网友 (117.22.*.*)

1

请问，检验什么情况用单侧？什么情况用双侧？

huoyuan13 [网易浙江省杭州手机网友]

2

我也想问这个

假设检验：

（当我们想说服自己或者别人，我们的数据为某个事实提供了证据的时候，需要用到假设检验）

证明逻辑：——“概率反证法”思想，即：

(1)为了检验一个零假设(即虚拟假设)是否成立，先假定它是成立的，然后看接受这个假设之后，是否会导致不合理结果。如果结果是合理的，就接受它；如不合理，则否定原假设。

(2)所谓导致不合理结果，就是看是否在一次观察中，出现小概率事件。通常把出现小概率事件的概率记为 α ，即显著性水平。它在次数函数图形中是曲线两端或一端的面积。因此，从统计检验来说，就涉及到双侧检验和单侧检验问题。在实践中采用何类检验是由实际问题的性质来决定的。一般可以这样考虑：

①双侧检验。如果检验的目的是检验抽样的样本统计量与假设参数的差数是否过大(无论是正方向还是负方向)，就把风险平分在右侧和左侧。比如显著性水平为0.05，即概率曲线左右两侧各占，即0.025。(95%置信区间)

②单侧检验。这种检验只注意估计值是否偏高或偏低。只注意偏低，则临界值在左侧，称左侧检验；如只注意偏高，则临界值在右侧，称右侧检验。

这里算标准差 σ_{p1-p2} 的时候不是应该根据假设, $p1=p2=p$, p (平均) $=1/2$ 然后 $\sigma_{p1-p2}=\sqrt{2*1/2(1-1/2)/1000}=0.0223$ 么?

换个与视频不一样的理解方式, 有体现除以2的直觉:

第一直觉, p 应该就是 $p1+p2$ 的一半, 那么, 就用样本值 $0.5*(p1+p2)$ 去近似估计; (注意它们的区别: $p1, p2$: 来自单个的样本, 可用于近似估计其他值; p : 来自均值之差 样本均值抽样分布, 并且是在假设检验中假设出来的, 是需要被估计得值。)

1、零假设中, 假设均值 $p1=p2$, 并令 $p=p1=p2$, 那么可得到 $p=0.5*(p1+p2)$;

2、根据大数定律, 可用样本均值去近似估计样本均值抽样分布的均值, 在这里表示:

$p1=642/1000=0.642$;

$p2=591/1000=0.591$,

因此 $p=0.5*(p1+p2)=(0.642+0.591)*0.5=0.6165$ 。