1. 分析下面两个因果论断错误的可能原因：

   (a) 美国大萧条时期，文化程度较高的群体失业率较低。所以教育程度越高，失业机会越小。

   Ans：文化程度高的群体从事的职业可能大部分比较稳定（比如医生、律师、政府职员等），这些职业的失业率不太受经济的影响。如果对特定的职业（比如，银行职员）考察失业率，教育程度高的人反而失业率可能更高, 这是因为他们的工资待遇较高（待遇与学历挂钩）。

   (b) 化石分析发现：地域分布广的物种，其生存年代更为久远。所以一个物种如果在地域上分布广泛，则难以绝迹。

   Ans：地域分布广的物种被发现的机会也大，更容易发现历史久远的化石。另一方面，也可能是因果颠倒，即难以绝迹的物种在时间、空间上跨度大，在地域上分布更广。

2. Ans: Cross-over 设计中，同一个人作为其本身的对照能很好地控制与时间无关的干扰因素，但同一个人在两个阶段处于不同的时间，其状态会有差异。比如经过第一阶段药物治疗之后，如果药物有效，会改变病人的疾病状态，而且药物残留的效应会显现在第二阶段。

3. 为期半年的减肥训练班结束的时候，体重超标最严重的几人的减重效果都较明显，这是否意味着训练有效？

   Ans: 未必，即使训练没有任何效果也可能出现这种现象（回归效应），实际上另外一方面，体重较轻的那些人可能平均体重会增加。

   假设只有 3/4 的学员坚持到训练班活动结束，为了衡量训练效果，假设我们比较这些坚持到底的学员训练开始和结束时候的体重，并以成对 t 检验考察训练效果的显著性。你认为这种检验方法是否恰当？如果不恰当，应该如何检验训练效果？

   因为每个个体测量两次，可看作是成对设计，故应该应用成对 t 检验（假设正态）。但只比较坚持完成训练的人的前后两次体重很有可能是不恰当的，因为退出者可能是因为感觉训练没效果才退出的。正确的做法是测量所有人的两次体重进行比较：如果训练无效，那么包括进退出者的数据不影响检验统计量的原假设下的分布。

   具体讨论如下：

   假设 $x_1,...,x_n$ 是 n 个人前后两次体重之差，假设 $x_1,...,x_n$ iid $\sim N(\delta,\sigma^2)$, 原假设是 $H_0:\delta=0$ (训练无效)。成对 t 检验统计量

   $$t=\sqrt{n}\bar{r}/s \sim_{H_0} t_{n-1}$$

   其中 s^2 是 $x_1,....,x_n$ 的样本方差，$\bar{x}$ 是样本均值。

   (1) 如果部分学员退出不是随机的，比如是因为感觉体重变化不大退出的，那么剩余的样本，比如为 $x_1,...,x_m, m<n$, 是所有样本中那些值比较大的，从而，即使 H_0 成立，我们也不能认为 $x_1,...,x_m \sim N(0,\sigma^2)$，从而基于 $x_1,...,x_m$ 构建的 t 统计量在原假设下不服从 t_{m-1} 分布, 即基于没退出学员的数据计算得到的成对 t 检验在原假设下不服从 t_{m-1}。

   (2) 如果存在非随机退出，正确的做法是使用所有体重差 $x_1,...,x_m, x_{m+1},...,x_n$, 假设其中后面 $n-m$ 个 x's 是退出的那些人的体重差。在 H_0 成立的条件下，前 m 个 x 和后 $n-m$ 个 x 分布没有差别，都服从 $N(0,\sigma^2)$ 分布，从而基于所有 $n=m+(n-m)$ 的样本的 t 检验在原假设下依然服从 t_{n-1} 分布。如果原假设不成立，那么 $x_1,...,x_m$ iid $\sim N(\delta>0,\sigma^2)$ (实际测得的体重差比较大), 而退出的那些

$x_{m+1},...,x_n$ iid $\sim N(0,\sigma^2)$（实际测得的体重差在 0 附近，均值为 0），从而基于所有样本的 $\bar{x}$ 的期望

$$E(\bar{x}) = \sum_{i=1} Ex_i/n = (m\delta + (n-m)*0)/n = m\delta/n < \delta$$

因此基于所有样本的 t 检验的功效相对于没人退出的情况会偏小。

4. 霍乱传播案例数据提供了除 Southwark & Vauxhall 和 Lambeth 之外其它供水公司客户的霍乱死亡率 (下表 "Rest of London")，试检验三家公司 (SV, Lambeth, Rest of London) 客户死亡率是否相同（3×2 列联表的 Pearson 卡方）。

Table 2. Death rate from cholera by source of water. Rate per 10,000 houses. London. Epidemic of 1854. Snow's table IX.

	No. of Houses	Cholera Deaths	Rate per 10,000
Southwark & Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

Ans: 3x2 列联表，计算 $\sum(O-E)^2/E$.

```
3x2表：
1263 40046-1263
98   26107-98
1422 256423-1422
```

5. Ans: 对比 Screened, Refused 组的其它疾病死亡率，结果显著。因为其它疾病可以认为与 X 光筛查无关，所以两组人除了接受与不接受筛查之外还存在其它系统性的差异。

6. 验证 $X^2 = z^2$ (略)

7. (未布置) 为什么案例 1 的 intention-to-treat analysis 是正确的方法？在 HIP 腺癌临床试验案例中，为了检验 Mammagraph 筛查是否有效，正确的做法是比较处理组（含接受和拒绝筛查的所有人）和对照组的死亡率。试讨论为什么（提示：一个检验如果能把 I 型错误率控制在给定的水平就是正确的）。

   Ans: 因为随机化分配，在原假设（X 光筛查无效）下，处理组（含 Refused）和对照组没有系统性差异，因而检验的 I 型错误率能控制在给定的水平，所以检验是正确的。而如果不包括进 Refused 组，只比较 Screened 组和对照组，那么即使在 X 光筛查无效的情况下，两组也因为个体的自我选择存在系统性的差异（比如职业、教育背景等），那么 I 型错误率会偏大，不能控制在给定的水平。

   当然，处理组包含 Refused 组（被邀请但未接受筛查的研究对象）会降低检验的功效，即如果筛查确实有效的情形下，将 Refused 组合并到 Screened 组会稀释筛查的效果。