**檔案:**

鐵達尼號電影曾經風靡全世界，讀者可於維基百科查詢到鐵達尼沈船的詳細情形。

鐵達尼號是一艘奧林匹克級郵輪，是當時最大的客運輪船，但因為人為錯誤，於1912年4月14日23點40分撞上冰山，事發2小時40分鐘後，即4月15日凌晨02點20分，船裂成兩半後沉入大西洋，死亡人數超越1500人，堪稱20世紀最大的海難事件，同時也是最廣為人知的海難之一。

**Story**

[**www.encyclopedia-titanica.org**](http://www.encyclopedia-titanica.org)

**Titanic Survival Predictor:**

[**http://demos.datasciencedojo.com/demo/titanic/**](http://demos.datasciencedojo.com/demo/titanic/)



**來源:**D. P. Byar and S. B. Green. The choice of treatment for cancer patients based on covariate information: Application to prostate cancer. *Bulletin Cancer, Paris*, 67:477-488, 1980.   
**問題:**

1. **該資料有哪些變數? 分別代表什麼意義? 請利用表整理出這些變數的意義與敘述統計。**
2. **男子(male)中，哪些姓氏的人死亡人數出現最多? 平均年齡多少?**
3. **您覺得票價(fare)大概是呈現何種分布? 票價和什麼變數相關? 票價合理嗎?**
4. **如果是你們組員不幸搭上這艘船，請你用迴歸分析方法預測他們的存活率。**
5. **同上，請你用分類樹方法預測他們的存活率! 並比較與上面方法的結論有何不同。**
6. **你們成員針對這次分析的心得?**

**Due date: 10/14(三) 6pm, 頁數限制12頁**