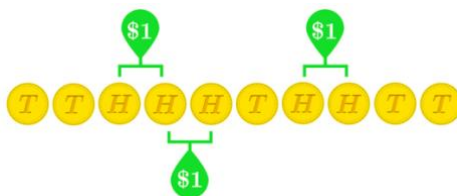


# STA101: 魅力统计 Fascinating Statistics

## 作业 2

1. (10 分)有这样一个游戏: 你抛一枚均匀的硬币 10 次, 按次序记下正反面的情况, 每连续出现一对正面, 你就能获得 1 枚金币。例如, 如果一次游戏的结果是 T<sup>\*</sup>TTHHHTHHT<sup>\*</sup>T(H 表示正面, T 表示反面), 你将获得 3 枚金币, 如下图所示。请问这个游戏的期望奖励是多少? 请给出具体计算细节。



2. (10 分)随机变量 $X$ 服从泊松分布 $P(\lambda)$ , 请给出其期望和方差的具体计算步骤。
3. (10 分)第一章中提到的 Francis Galton 做过一个实验叫高尔顿钉板实验, 请你自行上网搜索学习, 解释这是一个什么实验, 它和中心极限定理有什么关系?
4. (10 分)在第四章数据的收据中, 我们讲述了几个由于样本不具有代表性而得出错误结论的例子, 包括美国大选民意调查、长相与智商、战斗机加固。除了这几个例子, 请你再讲述一个类似的例子, 介绍故事的背景、样本有什么问题、得出了什么错误结论等。
5. (10 分)除课堂上讲过的数据可视化例子外, 请你讲述一个令你印象深刻的数据可视化例子, 介绍数据的背景、让你印象深刻的原因、数据可视化有什么特色等。
6. (15 分)在第四章数据的可视化中, 我们用 R 的 ggplot2 软件包画了包括柱状图、直方图、箱线图、相关热图在内的四种常见统计图形。请你自行上网搜索, 挑选另外两种统计图形, 并选择一个 ggplot2 软件包中自带的数据集, 用你喜欢的颜色画出这两种统计图形。请在答题区给出对这两种统计图形的介绍、所画图形的截图、作图使用的 R 语言代码。

7. (15 分)  $X_1, X_2, \dots, X_n$  和  $Y_1, Y_2, \dots, Y_m$  为来自同一个总体的两个独立的简单随机样本, 样本容量分别为  $n, m$ 。令总体的均值为  $\theta_1$ , 方差为  $\theta_2$ ,  $\bar{X}, \bar{Y}$  分别为两个样本的样本均值。
- (1) 证明对于任意满足  $a + b = 1$  的常数  $a, b$ ,  $\hat{\theta}_1 = a\bar{X} + b\bar{Y}$  都是  $\theta_1$  的无偏估计量。(5 分)
- (2) 确定使得(1)中定义的  $\hat{\theta}_1$  的方差达到最小的常数  $a, b$ 。(10 分)
8. (15 分) 美国劳工统计局发现, 一个容量  $n = 6000$  人的样本中有 516 个人失业。
- (1) 求美国总体失业率的 95% 置信区间。(5 分)
- (2) 在显著性水平  $\alpha = 0.05$  下检验美国总体失业率是否高于 8%。(10 分)
9. (5 分) 期末报告大家已经分好小组, 老师设置了一个腾讯文档供大家选择期末报告的时间, 最后一节课(2024 年 6 月 3 日)会有一次集中报告, 可容纳 8 个小组, 其他小组会分布在 10-15 周的上课时间, 每次课程最多两个小组。请各个小组商量好报告时间, 点击[这里](#)填写选好的时间, 并将报名截图贴至答题区, 如果你想选择的时间已经被其他小组优先占用, 请不要修改其他小组的选择。期末报告指引见课程 Blackboard/学习评价/期末报告。