**统计数据 201 数据分析**

**任务 1， NEFU， 2021**

**有关此任务的说明：**

此作业的主要目的是简化您进入分配程序和使用统计包 **R。**我们将通过使用**R工作室**和使用**R标记来达到这个目标**。

我们为您提供了一个名为**统计 20**1**\_2021\_NEFU\_A1.rmd**的 R Markdown文档，其中将已经填写了一些答案。您将需要填写并完成文档的其余部分。您将用于分配的数据文件在 questi 上进行描述，并可在线获得。请确保将这些数据文件放在将 R M方舟文件放在同一位置，因为它将在那里查找它们。您需要对 markdown 文档进行的第一个更改是将您的姓名和 ID 号放在顶部。

问题1。 •15 标记|

我们希望探讨世界各国的用电量与国内生产总值（GDP）之间的关系。国内生产总值是一国经济表现的指标，经购买力平价调整后，考虑到国与国之间的价格水平差异。为世界上28个人口最多的国家选取了资料。数据存储在"电.csv"中，并包含变量s

以 数十亿千瓦时的电力消耗。

国内生产总值 （GDP）以数十亿美元计算。

国家 名称。

这个问题的很多分析已经为你填写了。有几个部分和附加al 问题供您解决：

* 评论数据的初始图。
* 我们安装了最初的线性模型，但GDP最高的两个国家存在问题。识别这两个国家， 重新绘制消除这两个国家的数据， 对此图进行评论，然后重新安装简单的线性回归模型（包括假设检查），而没有这两个国家。
  + 创建一个散射图，将新安装模型的安装线叠加在它上。
* 写一份适当的 **执行摘要**。
* 使用最终模型来 预测一个国内生产总值为1000亿美元的国家的用电量。解释此间隔。基于此间隔，预测模型有多大用处？

问题2。 [8 标记]

一位研究人员推测，世界各国的平均预期寿命约为68岁。他从55个国家的随机样本中收集了数据，并希望利用这些数据来检验他的假设。 数据存储在"国家.csv"中，并包含变量

每个国家的预期寿命（以年为数）。

* 对图/探索性数据分析的评论。
* 手动计算将平均预期寿命与 68 岁和相应的 95% 置信区间进行比较的 *t*统计。
* 为什么*t*测试输出和空线性模型的 P 值不同？
* 写一份适当的 **执行摘要**。

**问题3。 [16** **标记]**

美国俄勒冈州尤金市的一位研究人员对房屋的销售价格如何受房屋年龄的影响很感兴趣。2005年，她随机抽取了76套单户住宅。数据存储在"家庭.csv"中，并包含可变s

以 千美元的价格出售房屋。

房子的年龄，定义为2005年减去房子建造的年份）。

* + 评论分析的兴趣/目标问题。
  + 评论数据的初始图。
  + 适合适当的线性模型，包括模型检查。
  + 将数据与叠加在它上的估计模型进行绘图。
  + 编写适当的 **方法和假设检查** 和 **执行摘要。**

分配说明

对于很多任务问题，我们将简单地给你描述如何和为什么收集数据和最小的指导。您将看到如何分析课堂案例研究中的data，但以下是回答开放数据问题的一般方法：

* 评论感兴趣的问题或分析的目标。
* 查看数据（绘制、获取汇总统计数据）并对此发表评论。
* 将模型适合数据
* 检查模型 假设。
* 根据需要更改模型并重复检查。您可能需要不止一次地这样做。

• 从最终模型生成推理输出。

• 编写 **M伦理 和假设检查** 部分。

* 这将详细说明您采取的步骤以及您为什么在构建模型时采取这些步骤。
* 它将包括模型假设检查的简要描述。
* 它将包括您安装的最终模型的数学语句。

• 撰写 **执行摘要**。

**确保您阅读下一页上的笔记！**

**一些非常重要的说明：**

* 在以案例研究为指导时，不要盲目地遵循一个案例研究。所有数据集都有自己的属性，不太可能完全匹配您找到的案例研究。相反，智能地使用案例研究来指导你。
* 在评论情节时，请保持评论简短和相关。在评论假设时，您不需要详细阐述。如果情节显示没有问题，可以这样说。如果情节显示问题， brie飞描述问题， 然后说可以做些什么。
* 撰写**执行摘要**时：
* We 希望从最初提出的问题中得出主要结论。
* 如果数据存在关键问题或目标，请确保直接回答，然后根据需要详细了解。例如，如果一项研究询问片剂是否影响血压，你可以有一个句子，如"我们证明片剂会增加血压"，然后用适当的量化来支持这一点（例如，给出服用片剂后平均血压折痕的一系列数字）。不要让我们推断量化的结果。
* 指出对模型做出的任何异常步骤或更改，这些步骤或更改都用易于理解的术语表示。
* 你应该使用易于理解的自然语言。您应该避免使用可变名称、大量小数位和不必要的细节。
* 国家单位在已知的情况下。
* 这应该是一个简短的，容易阅读的分析摘要，一个对统计学知之甚少的人可以理解，而无需通读分析。
* 如果你想检查你的 E x 执行 Summary， 找一个没有做统计的朋友来阅读它， 并告诉你他们认为这意味着什么。

获取 R和 **R 工作室**副本，供在家使用

首先，下载并安装最新版本的**R** 与最新的 s20x 库：

前往： [www.stat.auckland.ac.nz](http://www.stat.auckland.ac.nz)

单击屏幕右上角**的主机**标题下方的**CRAN**

在框标签d**下载和安装 R，**c舔 **Linux** 或 **（Mac） OS X** 或 **Windows** 取决于您的计算机系统

**什么是 R 标记？**

**R Markdown**是一种相对较新的标记语言内置到 **R 工作室**。它是一种创作格式，能够轻松创建动态文档和 R 报告。

这意味着您处理主文档（包括 R 命令）并"编织"在一起，R Studio 在后台运行 R，并创建最终文档，将所有图形和计算机输出插入正确位置。如果你想改变你的文件，你只是"编织"的文件再次： 您不必重做 分析，因为它 会自动重新运行。 您也不需要重复复制和粘贴输出。

**开始：**

在行动中看到这一点的最佳方式是尝试这一点。

运行 R 工作室。在文件菜单下，去打开文件，并导航找到和打开 统计201\_2021\_NEFU\_A1.rmd.

接下来，find针织按钮（旁边的一些羊毛和针织针的图标），并选择**针织字**（除非你在家里，没有安装Word在这种情况下选择**针织HTML）。**这将将模拟泰利将文档"编织"到 Word（或 HTML） 文件中，运行R代码并创建输出文件。

恭喜你-你创造了你的第一个文件！

**三重针织：**

您可以将文档编织为 3 种格式：HTML、Word 或 PDF。我们将接受使用三种格式中的任何一种交上来的任务，尽管我们更喜欢 Word，因为它更清洁，而且如果您愿意，您可以轻松调整绘图大小以整理并减少作业中的页面数。

要编织到 Word，您必须在计算机上安装 Microsoft Word。

要编织到 PDF，您必须在计算机上安装乳胶。乳胶是一种自由、极其强大的数学排版语言。不幸的是，我们不能支持您安装这个，但欢迎您这样做。

**编辑文档：**

现在，您应该尝试编辑文档，看看会发生什么。

* 更改标题以包含您的姓名和 ID 号。
* 编辑一些评论。
* 看看当你尝试方程的公式时会发生什么。
* 添加一些 R 代码。

确保您保存了原始文档，并记住经常备份。您通过尝试新软件工具来了解它们。你总是可以消除错误。

方程需要多一点练习。您会注意到，我们已经为您提供了 A寄售 1 的方程，使您的生活更轻松。

在下一页，我们提供了一些信息，说明如何更改文档的格式，例如添加头条、列表、粗体、意花。一点点练习就相当容易了。

**一些格式化选项：**

要开始新段落，请结束两个空格的行。

\*意花\* 生成 *意花* \*\*大胆\*\* 生成 **粗体**

# 头 产生 **头**

## 头 产生 **头**

## 头生产 **头**

...

####头产生 **头**

要制作未排序列表：

\* 项目 1 = 项目 1

\* 项目 2 要点 = 项目 2

• 子项 1 = 子项目 1

• 子项 2 = 子项目 2

要制定订单列表：

1. 项目11。 项目 1

2. 项目2gives2。 项目2

• 子项 1 = 子项目 1

• 子项 2 = 子项目 2

要编写方程，在 $符号之间封装它们。

$beta\_0=55美元给予

要获得 Greek 字母： • 西格玛产生σ等。

s上记 +2 给超级脚本2  subscripts\_1给子脚本1

\ne 给≠ ] 时间给×

想在符号上戴一顶帽子吗？ $帽子 [贝塔] $ 给 

想 拿 一个方根？

想赚一小部分吗？

想做一个愚蠢的复杂公式吗？

[帽子] [\ s ][ 弗拉克] beta\_0 [ 2] [西格玛] 时间 [ 酒吧]

给