Homework 5

1. Following the academic example discussed in the class, you are expected to deliver an empirical analysis and answer one of the two questions below as you like:

   (you may download the dataset from the Canvas system)

   根据课堂上讨论的学术论文案例，请进行实证分析并在以下两个问题中任选一个作答： (您可以从 Canvas 系统下载数据集)

   1) Suppose we focus on the high-intensity counties (indexed "1" in the column "high_county_T1"). Does the video treatment have the same impact on counties with different economic conditions (urban (indexed "1" in column "urban") or rural county)?

   You can also analyze the above question for the low-intensity counties and explain your results.

   2) Suppose we focus on the high-intensity counties (indexed "1" in the column "high_county_T1"). Does the video treatment have the same impact on counties with different political statuses (Republican   (indexed "1" in column "majority_gop") or Democratic)?

   You can also analyze the above question for the low-intensity counties and explain your results.

   1) 假设我们关注高强度干预的县(在"high_county_T1"列中索引为"1")。 视频推送对不同经济状况的县(城市("urban"一栏索引为"1")或农村)是否有相同的影响?

   您也可以对低强度干预的县做以上分析，并简析您的结果。

   2) 假设我们关注高强度干预的县(在"high_county_T1"列中索引为"1")。 视频推送对不同政治偏向的县(共和党（"majority_gop"一栏索引为"1")或民主党)是否有相同的影响?

   您也可以对低强度干预的县做以上分析，并简析您的结果。

2. **Background:** As the COVID-19 continues spreading globally in 2021, universities have adopted two forms of final exams: **(1) online; (2) onsite**, where students can choose one of them. Based on the knowledge you have learned in business statistics, you are expected to judge whether the different forms of exams (online and onsite) **significantly influence** the scores of various courses.

   背景：随着 2021 年全球新冠疫情的持续流行，全国各地大学均采用线上和线下相结合的考核方案。基于商务统计知识，本题拟分析线上线下教学和考核形式对不同学科的考试成绩是否会有显著性影响.

   1) Please design an experiment to collect students' data

Hint: Consider the five decisions to finalize your experiment: a) What is the Overall Evaluation Criterion (OEC)?   b) What is the randomization unit?   c) What population of randomization units do we want to target?   d) How large (size) does our experiment need to be?   e) How long do we run the experiment?

请设计实验方案收集学生数据

提示：思考实验设计的五个决策：a）综合评估标准（OEC）是什么？b）随机化单元是什么？c）目标群体是什么？d）实验需要多大的样本量？e）实验需要运行多久？

2) Now we have collected the exam scores of 600 students, which include both forms of online and onsite exams and three different subjects (business statistics, financial engineering, and stochastic processes). The dataset is shown in the table below. Please briefly answer how to design a hypothesis testing scheme based on the above data:

   a) Will different forms (online or onsite exams) significantly impact students' performance?

   b) Are there existing significant differences in average performance among different courses?

   c) Do different forms of exams have the same impact on different courses?

已收集 600 个学生样本的考试成绩，其中包含参加线上线下两种考试形式，三种不同科目（商务统计学，金融工程，随机过程），如下表所示。请简要回答如何通过上述数据设计假设检验方案:

a) 线上线下不同考试形式是否会对学生成绩产生显著影响？

b) 不同科目之间平均成绩是否有显著差异？

c) 线上线下考试形式对不同科目的影响是一致的吗？

		Factor B:	
		1(online, 线上)	2（onsite, 线下）
Factor A	1 (business statistics 商务统计学)	$x_{111}, x_{112}, \ldots, x_{11r}$	$x_{121}, x_{122}, \ldots, x_{12r}$
	2 (financial engineering 金融工程)	$x_{211}, x_{212}, \ldots, x_{21r}$	$x_{221}, x_{222}, \ldots, x_{22r}$
	3 (stochastic processes 随机过程)	$x_{311}, x_{312}, \ldots, x_{31r}$	$x_{321}, x_{322}, \ldots, x_{32r}$

r=100.

3) Bonus question： What are the potential problems in the data above? How can we improve it?

附加题：上述数据可能的问题是什么？我们如何进行改进？