

公共选修课“可计量的社会”·课后作业 题目

命题人：【国际教育学院】何濯羽

2024 年 11 月 18 日

答题须知

1. 本作业为开卷作业，学生可查阅任何线上资料与网络信息。
2. 本作业满分为 100 分，占本门课程总分的 30%，包含 5 道填空题（4 分/题）、5 道填空题（4 分/题）、12 道简答题（5 分/题）。

题号	1	2	3	4	5	总分
分值	20	20	20	20	20	100

3. 本作业考察知识范围为本门课程的第一至第五单元，包括：概率论与描述性统计、推断性统计与回归模型、随机控制试验、工具变量、二元选择模型。
4. 学生需要在答题纸（请在钉钉群共享文件中下载打印）上手写作答，并在 **2024 年 12 月 2 日（周一）19:00 前**将答题纸上交给授课教师，逾期一律计为 0 分。

第 1 大题 (20 分) 单选题

1-1 以下哪个选项不属于频率学派所提出的试验概率 (frequentist probability) 的缺陷?

- A. 难以确保每次试验的程序完全相同
- B. 无法确认具体需要重复试验多少次
- C. 某些试验在道德上不被允许
- D. 无法应用于已知的非等概率的情况

1-2 以下哪个函数无法完美地描述一个随机变量的概率分布?

- A. 矩量母函数
- B. 累积分布函数
- C. 分位数函数
- D. 概率体积函数

1-3 当一个估计量的_____等于被估量的真实值时, 我们称该估计量具有无偏性。

- A. 数学期望
- B. 方差
- C. 标准差
- D. 偏度

1-4 综合健康计划 iThrive 是由哪个美国大学设计并在校园内推行的?

- A. 芝加哥大学
- B. 伊利诺伊大学芝加哥分校
- C. 伊利诺伊大学厄巴纳-香槟分校
- D. 约翰霍普金斯大学

1-5 当我们在构建 Probit 模型 $Y = \Phi(\beta X) + e$ 时, 我们实际上是在假设误差项 e 服从_____概率分布。

- A. $t(n-1)$
- B. $\chi^2(n)$
- C. $N(0,1)$
- D. $U[0,1]$

第 2 大题 (20 分) 填空题

2-1 数学期望衡量的是一个随机变量取值的_____, 但它并非唯一可以衡量随机变量这一特征的统计工具 (中位数和众数也可以)。方差衡量的是一个随机变量取值的_____, 但为了解决单位的问题, 统计学家们通常选择使用标准差 (即方差的开方) 来衡量该特征。

2-2 X 与 Y 代表两个不同的随机变量。当 $Pr(X = x, Y = y) = Pr(X = x) \cdot Pr(Y = y)$ 时, 我们称 X 与 Y 相互_____独立; 当 $Cov(X, Y) =$ _____时, 我们称 X 与 Y 线性不相关。

2-3 条件期望函数是统计学回归分析中最核心的概念。

2-3-1 条件期望函数 $E(Y|X)$ 是一个以_____为自变量的函数。

- 2-3-2** 任何一个随机变量 Y 都可以被表达为如下的分解形式： $Y = E(Y|X) + e$ ，其中_____表示可以被 X 解释的部分，_____表示不可以被 X 解释的部分。
- 2-3-3** 出于计算与解释简便的考虑，研究者通常会假设条件期望函数是一个线性函数。这样，回归模型可以被写为 $Y = \alpha + \beta X + e$ ，其中 Y 被称作_____， X 被称作_____。

第 3 大题 (20 分) 简答题：数字平台与毕业生就业

- 根据 Kelly, Ksoll & Magruder (2024) 及课堂内容回答以下问题。
- 3-1** 在新冠疫情之后，青年失业 (youth unemployment) 问题成为几乎全世界所有国家需要面对的难题。请问什么是青年失业问题？
- 3-2** Kelly, Ksoll & Magruder (2024) 的研究问题是什么？
- 3-3** 为了解决他们的研究问题，Kelly, Ksoll & Magruder (2024) 采用了随机对照试验 (RCT) 方法。在他们的 RCT 中，共有两个实验组。请问这两个实验组的区别是什么？
- 3-4** Kelly, Ksoll & Magruder (2024) 发现：在 Job Shikari 注册账号并收到短信后，“普通”实验组毕业生的就业率下降了 9.2 个百分点。请问他们是如何解释这一发现背后的原理的？

第 4 大题 (20 分) 简答题：教育与收入

- 根据 Angrist & Krueger (1991) 及课堂内容回答以下问题。
- 4-1** Angrist & Krueger (1991) 使用了工具变量 (IV) 去估计教育对收入的影响。请问他们的工具变量是什么？
- 4-2** 下表来自 Angrist & Krueger (1991) 的 Table IV，请阅读该表并回答问题。

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0802 (0.0004)	0.0769 (0.0150)	0.0802 (0.0004)	0.1310 (0.0334)	0.0701 (0.0004)	0.0669 (0.0151)	0.0701 (0.0004)	0.1007 (0.0334)
Race (1 = black)	—	—	—	—	0.2980 (0.0043)	-0.3055 (0.0353)	-0.2980 (0.0043)	-0.2271 (0.0776)
SMSA (1 = center city)	—	—	—	—	0.1343 (0.0026)	0.1362 (0.0092)	0.1343 (0.0026)	0.1163 (0.0198)
Married (1 = married)	—	—	—	—	0.2928 (0.0037)	0.2941 (0.0072)	0.2928 (0.0037)	0.2804 (0.0141)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region of residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	0.1446 (0.0676)	0.1409 (0.0704)	—	—	0.1162 (0.0652)	0.1170 (0.0662)
Age-squared	—	—	-0.0015 (0.0007)	-0.0014 (0.0008)	—	—	-0.0013 (0.0007)	-0.0012 (0.0007)
χ^2 [dof]	—	36.0 [29]	—	25.6 [27]	—	34.2 [29]	—	28.8 [27]

- 4-2-1** (1) 列和 (2) 列对应的回归模型分别使用了什么估计方法？
- 4-2-2** 与 (2) 的回归模型相比，(4) 列的回归模型加入了哪些额外的控制变量？
- 4-2-3** 根据 (8) 列的估计结果，受教育时间增加 1 年，会使得工作薪资增长多少百分比？（计算结果保留两位小数）

第 5 大题 (20 分) 简答题：经济萧条时的饮酒行为

Outcome	(a)	(b)	(c)	(d)	(e)	(f)
Drinker	-4.5E-4 (7.8E-4)	-6.1E-4 (8.2E-4)	-0.0018 (0.0013)	-0.0019 (0.0013)	-0.0021 (0.0014)	-3.3E-4 (0.0031)
Log of number of drinks	-0.0151 (0.0024)	-0.0184 (0.0027)	-0.0313 (0.0048)	-0.0099 (0.0045)		-0.0408 (0.0141)
Alcohol-involved driving	-4.1E-4 (1.8E-4)	-1.8E-4 (2.0E-4)	-8.9E-4 (3.4E-4)	-4.8E-4 (3.7E-4)	-5.1E-4 (2.0E-4)	-6.7E-4 (4.7E-4)
Type of data	Individual	Individual	Individual	Individual	Individual	Aggregate
State-specific time-trends	No	Yes	Yes	Yes	Yes	Yes
Year effects	No	No	No	Yes	No	No
Sampling weights	No	No	Yes	Yes	Yes	Yes
Estimation technique	OLS	OLS	WLS	WLS	Probit	WLS

上表来自 Ruhm & Black (2002) 的 Table 2, 请阅读该表, 并根据 Ruhm & Black (2002) 中该表的附注内容回答以下问题。

- 5-1** (a) 列和 (c) 列对应的回归模型分别使用了什么估计方法?
- 5-2** 根据 (c) 列展示的结果, 当州级失业率上升 1 个百分点时, 我们预计当地居民参与饮酒的概率升高或降低多少?
- 5-3** 根据 (c) 列展示的结果, 当州级失业率上升 1 个百分点时, 我们预计当地居民饮酒杯数的增加或减少了百分之几? (计算结果保留两位小数)
- 5-4** 在 (e) 列, 研究者们使用了 Probit 模型, 请问第二行为什么没有结果 (即空白)?