Short Questions

Bookmark this page

Comprehensive Review due Jul 31, 2024 07:30 CST Completed
Question 1

1.0/1.0 point (graded)

You realize your wallet is missing. You have either left in the office or at a friends' place. It seems very likely, say 80% likely, that your friend would have texted you by now if you had left it at her place. You know it is unlikely, (say 5% probability), that someone in the office would have tracked you down by now if you had lost it there. You have not been contacted.

What other information do you need in order compute the probability that you left it in the office? Select all that apply and select none of the above if none of the choices can be used to compute the proabbility that you left it in the office.

- [ ] the conditional probability that you left it at your friends.
- [ ] the conditional probability that you left it in the office.
- [x] the prior probability that you left your wallet in the office.
- [ ] None of the above.

Explanation
We want to find the probability that the wallet was lost at the office conditional on not having received a phone call, $P(O|Call^C)$. By Bayes rule:

$$P(O|Call^C) = \frac{P(Call^C|O)P(O)}{P(Call^C|O)P(O)+P(Call^C|F)P(F)}$$

where we know $P(Call^C|O) = 1 - P(Call|O) = 0.95$ and $P(Call^C|F) = 1 - P(Call|F) = 0.80$
Hence, we would need $P(O)$ and $P(F)$, which are the prior probabilities that you left the wallet in the office, and at your friends' place, respectively. Given that $P(O) = 1 - P(F)$, you could also find the probability that the wallet was lost at the office conditional on not having received a phone call with either $P(O)$ or $P(F)$. Since $P(F)$ is not one of the options, the prior probability that you left your wallet in the office is the correct choice.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 2

1.0/1.0 point (graded)

True or False? The Poisson distribution is a special case of the exponential distribution.

- ( ) True
- (•) False

Explanation
The exponential distribution models the waiting time between two events in a Poisson process. One is NOT a special case of the other. In particular, the Poisson is a discrete distribution and the exponential is a continuous distribution.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 3

1.0/1.0 point (graded)

Consider a $\mathcal{B}(n,p)$ and a $\mathcal{H}(A,B,n)$ where $p = A/(A+B)$. Which of the following statements is true? Select all that apply.

- [ ] The variance of the binomial is always smaller than the variance of the hypergeometric.
- [x] The variance of the hypergeometric is always smaller than or equal to the variance of the binomial.
- [ ] The relative sizes of the variances depend on the choice of both n and p.
- [ ] The variance of the binomial is always smaller than or equal to the variance of the hypergeometric.

Explanation
Let $X \sim B(n,p)$. Then, $V(X) = np(1-p)$. When $p = \frac{A}{A+B}$, $V(X) = n\frac{AB}{(A+B)^2}$.
Let $Y \sim H(A,B,n)$. Then, $V(Y) = n\frac{AB}{(A+B)^2}\frac{A+B-n}{A+B-1}$. Hence $V(X) > V(Y) \iff 1 > \frac{A+B-n}{A+B-1}$. But this is always true. Hence the variance of the hypergeometric is always smaller than the variance of the binomial, unless $n = 1$, in which case they are equal.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 4

1.0/1.0 point (graded)

True or False? Knowing the distribution of the test statistic under the null allows you to calculate α (the probability of a type I error) and β (the probability of a type II error).

- ( ) True
- (•) False

Explanation
Recall, knowing the distribution of the test statistic under the null allows you to calculate α (the probability of a type I error), but does not allow you to compute β.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 5

1.0/1.0 point (graded)

Suppose that judges can decide to whom they give a GPS bracelet to when individuals who have been arrested for a crime are released on bail (pending judgement).

Separately, a researcher runs an RCT for the impact of bracelets: people who are released on bail are randomly assigned to recieve a bracelet or not. She finds that the bracelet reduces the probability of committing a crime while on bail by 5 percentage points (and this difference is significant).

Another researcher obtains a separate data set on the regular program (where the judge can decide whether or not to give the bracelet), that has many variables about the people who were arrested (including whether or not they were given a bracelet when released), and runs a machine learning algorithm to find out what predicts whether someone will commit a crime while on bail.They find that the bracelet tends to predict **greater** recidivism (recidivism means relapse into criminal behavior)

This implies that one of the two studies must be incorrect.

- ( ) True
- (•) False

Explanation
Machine learning is concerned with prediction, whereas the RCT is concerned with causal estimation. Since judges probably give out the bracelets to individuals who are more likely to be arrested for crime, the machine learning algorithm could be picking that up.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Long Question 1 - Jail Sentences and Recidivism

Bookmark this page

Comprehensive Review due Jul 31, 2024 07:30 CST Completed

You are interested in doing a project on jail sentences and recidivism. You find publicly available data listing plea deals and court decisions resulting in jail sentences for Middlesex County, Massachusetts. The list contains de-identified information on offense, sex, age, and prison of incarceration. You scrape the data and perform some preliminary analysis. You wish to contact the prisoners and ask each to participate in a survey. You then survey those who are willing and analyze the resulting data.

Question 6

1.0/1.0 point (graded)

You must obtain approval from your IRB before _____.

- You administer your survey.
- You do any preliminary analysis.
- You contact the prisoners.
- You scrape the data. ✔

✔

Explanation

You should obtain approval from your IRB prior to contacting the prisoners and asking them to participate in the survey, as this would meet the criterion of "data obtained through intervention or interaction with the individual."

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 7

1.0/1.0 point (graded)

Suppose you conclude, later in continuing research, that recidivism (whether released prisoners commit additional crimes) is a function of length of original sentence, S, distance of prison from hometown, D, and sex of offender, M ($=1$ if male). In particular, the probability of recidivism, $P(R)$, is $S/32 + D/40 + 0.1M$. Suppose also that S follows a uniform distribution $U(1, 15)$; D follows an exponential distribution $\exp(0.05)$; and 80% of offenders are male.

What is the expected recidivism rate?

Please express your answer to the nearest whole percentage, i.e if your answer is 23.2, please enter 23 and DO NOT enter in a $\%$ sign

83 ✔ Answer: 83 **or** 0.83

83

Explanation

$[R] = \left[\frac{S}{32} + \frac{D}{40} + 0.1M\right] = \frac{[S]}{32} + \frac{[D]}{40} + 0.1\,[M]\,[S] = \frac{1+15}{2} = 8,\ [D] = \frac{1}{0.05} = 20,$ and $[M] = 0.8.$ So, $[R] = \frac{8}{32} + \frac{20}{40} + 0.1 * 0.8 = \frac{83}{100}$ The recidivism rate is thus $83\%.$

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Long Question 2 - Prof. Ellison's Commute

Bookmark this page

Comprehensive Review due Jul 31, 2024 07:30 CST Completed

Sara Elison needs to commute to MIT everyday. She currently lives near campus, but is considering moving to a place near Fenway stadium (Boston's baseball stadium). Her colleagues who live near there tell her that the commute from the office to the Fenway is independent across days and follows a $\mathcal{N}(20, 9)$ (where 20 is the average commute and 9 is the variance) on days where there is a home game at Fenway and $\mathcal{N}(12, 4)$ (where 12 is the average commute and 4 is the variance) on other days.

Question 8

1.0/1.0 point (graded)

What is the probability that the commute on a particular game day exceeds 22 minutes?

Please round your answer to 2 decimal points, e.g. if your answer is 0.987, please round to .99 and if it is 0.981, round to 0.98)

0.25 ✔ Answer: [0.245, 0.255]

$\mathbf{0.25}$

Explanation

Let $X_G \sim N(20, 9)$ be the commute distribution on a game day (i.e., there is a home game at Fenway). We want to find $P(X_G > 22)$.

First, we standardize X_G by subtracting its mean and dividing by the standard deviation on both sides of the inequality:

$$P(X_G > 22) = P(Z > \tfrac{22-20}{3}) = P(Z > \tfrac{2}{3})$$

where $Z \sim N(0, 1)$

Then, by the symmetry of the normal distribution: $P(X_G > 22) = 1 - P(Z \leq \tfrac{2}{3})$. So, we can look up in the table that $P(Z \leq \tfrac{2}{3}) = 0.7486$. Hence, $P(X_G > 22) = 0.2514$. We also gave full credit if you rounded $\tfrac{2}{3}$ as 0.66.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 9

1.0/1.0 point (graded)

What is the probability that all commutes on a particular 3-game homestand (3 games played at Fenway) exceed 22 minutes?

Please round your answer to 2 decimal points)

0.02 ✔ Answer: [0.015, 0.025]

$\mathbf{0.02}$

Explanation

We want the probability that the commute is greater than 22 minutes on three days. But these are independent events, hence the answer is given by:

$$P(X_1 > 22\ AND\ X_2 > 22\ AND\ X_3 > 22) = P(X_G > 22)^3 = 0.2514^3 = 0.01588897274 \approx 0.02$$

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 10

1.0/1.0 point (graded)

What is the probability that the commute on a particular game day exceeds the commute on a particular non-game day?

Please round your answer to 2 decimal points.

0.99 ✔ Answer: 0.99

$\mathbf{0.99}$

Explanation

Let $X_N \sim N(12, 4)$ be the commute distribution on a non-game day. We want to find $P(X_G > X_N) = P(X_G - X_N > 0) = 1 - P(X_G - X_N \leq 0)$. But since X_G and X_N are independent normal distributions, $X_G - X_N \sim N(20 - 12, 9 + 4)$. Hence,

$$1 - P(X_G - X_N \leq 0) = 1 - P(Z \leq \tfrac{0-8}{\sqrt{13}}) = P(Z \leq \tfrac{8}{\sqrt{13}}) = 0.9868$$

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Long Question 3 - Scrabble Pizza Party

Bookmark this page

Comprehensive Review due Jul 31, 2024 07:30 CST Completed

You and your friend decide throw a scrabble pizza party. You don't know how many pizzas to order, so you want to obtain an estimate of slices each person would eat. You perform a poll of ten randomly selected patrons to see how many slices they each ate. Their responses are denoted $X_i, i = 1, \ldots, 10,$ and

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i = 4.3$$

$$s^2 = \frac{1}{9}\sum_{i=1}^{10} (X_i - \bar{X})^2 = 2.7$$

Question 11

1.0/1.0 point (graded)

Assuming that the X_is are *i.i.d.* normal, construct a 95 confidence interval for the mean of the X_is. Enter the lower and upper bounds on the interval $[a, b]$.

Please round your answer to 2 decimal points

Lower bound **a**:

3.12 ✔ Answer: [3.115, 3.125]

3.12

Upper bound **b**:

5.48 ✔ Answer: [5.475, 5.485]

5.48

Explanation

$$\bar{X} \pm t_{0.025,n-1}\frac{s}{\sqrt{n}}$$

Plugging in the relevant numbers from the table:

$$4.3 \pm 2.26\frac{\sqrt{2.7}}{\sqrt{10}}$$

$$[3.12, 5.48]$$

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 12

1.0/1.0 point (graded)

How many patrons would you have had to poll to limit the width of the confidence interval to one slice?

*Please round your answer **up** to the nearest whole number.*

42 ✔ Answer: 42

42

Explanation

We want to find the n for which the width of the confidence interval will be limited to one slice. Using our notation, we want to find n such that:

$$\frac{1.96\sqrt{2.7}}{\sqrt{n}} < 0.5$$

Note that we can also rewrite this as:

$$2\frac{1.96\sqrt{2.7}}{\sqrt{n}} < 1$$

This is because since the total width of the confidence interval must be less than 1 slice, each side of the confidence interval must be less than 0.5 slice. Solving for n, we have:

$$n > (2 * 1.96 * \sqrt{2.7})^2$$

$$n > 41.49$$

Rounding up, we would need to have $n > 42$ to ensure the width of the confidence interval is less than 1 slice. If we used critical values from the t-distribution, we would compute slightly different answers (i.e. 44).

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 13

1.0/1.0 point (graded)

True or False? Suppose your friend gives you the additional information that the X_is have a uniform distribution on $[0, \theta]$ instead. Unfortunately, he has already thrown away the original data and kept the sample mean and the sample variance he computed. You can still construct a 95% confidence interval for the mean of the X_is.

- (•) True
- ( ) False

✔

Explanation

You can still construct a 95% confidence interval for the mean of the X_i, you can just use $2\bar{X}$ as an estimate for θ.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Previous Next

Long Question 4 - Neyman Analysis and Fisher Exact Test

Bookmark this page

Comprehensive Review due Jul 31, 2024 07:30 CST Completed

Suppose that the 8 large regions of Vietnam were randomized into one of the following groups: in some regions, the local health care centers continued to be run by the government (control group: 4 regions) or in others, the health centers were subcontracted out to a NGO (treatment group: 4 regions).

You have access to information from a child health survey, which covers 1,000 children per region and gives you information on whether or not the children have been fully immunized.

Question 14

1.0/1.0 point (graded)

A collaborator proposes to run a standard Neyman analysis, on the sample of $\mathbf{4,000}$ treatment and $\mathbf{4,000}$ control children, ignoring the region altogether.

Denote $\overline{Y_T} = \mathbf{0.80}$ the sample average immunization rate in the treatment group, $\overline{Y_C} = \mathbf{0.58}$ the sample average immunization rate in the control group, $\sigma_T^2 = \mathbf{1.2^2}$ the estimated variance in the treatment group and $\sigma_C^2 = \mathbf{2.3^2}$ the estimated variance in the control group.

For this question, please round your answer to 2 decimal points

What is the collaborator's estimate of the average treatment effect?

0.22 ✔ Answer: [0.2178, 0.2222]

$\mathbf{0.22}$

For each of the following questions, please round your answer to 3 decimal points

What is the collaborator's estimate of the associated variance?

0.002 ✔ Answer: [0.0016825, 0.0021)

$\mathbf{0.002}$

Explanation

$$\hat{\tau} = \bar{Y}_t - \bar{Y}_c = 0.80 - 0.58 = 0.22$$

$$\hat{V}_{\text{Neyman}} = \frac{\sigma_T}{N_T} + \frac{\sigma_C}{N_C} = \frac{1.2^2}{4000} + \frac{2.3^2}{4000} = \frac{6.73}{4000} = 0.0016825$$

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 15

1.0/1.0 point (graded)

You object to the collaborator's approach, and instead propose to use the fact that the randomization was done at the region level very seriously and aggregate the data at the region level. Since the sample is small, you propose to run a Fisher exact test.

True or False: The test will test the hypothesis H_0 that the average treatment effect is significantly different from $\mathbf{0.}$

- True
- False ✔

Explanation

The Fisher exact test tests the hypothesis that the treatment effect is identically $\mathbf{0}$ for all treatment units.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 16

1.0/1.0 point (graded)

At the regional level, the rates of fully immunized children in treatment regions are as follows:

- Treatment regions: $\mathbf{85\%, 99\%, 100\%, 76\%}$
- Control regions: $\mathbf{26\%, 45\%, 97\%, 72\%}$

(Round your answer to 2 decimal points)

I. Using a Permutation Table or R code, construct your Fisher exact test. Please enter the p-value you obtained from the test you constructed.

0.11 ✔ Answer: [0.109, 0.115]

$\mathbf{0.11}$

II. True or False? You can you reject H_0 at the $\mathbf{5\%}$ level.

- True
- False ✔

Explanation

Refer back to Homework 8. There are ${}_8C_4$ possible permutations of the treatment assignments, so $\mathbf{70}$ combinations. In $\mathbf{8}$ of them the test statistic is greater than $\mathbf{30,}$ which implies that the p-value is $\mathbf{0.11.}$ Hence, we fail to reject H_0 at the $\mathbf{5\%}$ level.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Previous Next

Long Question 5 - Flowers in China

Bookmark this page

Comprehensive Review due Jul 31, 2024 07:30 CST Completed

China suffers from enormous gender imbalance: there are many more boys than girls. Part of this is due to selective abortion, and part is due to worse treatment of girls.

Nancy Qian was interested in finding out whether parents consider the future possible wages of a girl when deciding how much to feed them and take care of them. To this end, she exploits the reform that brought household responsibility system reform in China. After Den Xiao Ping replaced Mao in 1979, households were given the choice about what crop to grow (before, they essentially had to grow cereals), and suitable regions started producing tea and orchards.

Women are particularly useful for tea growing, which requires nimble hands. Therefore she proposed that parents would start taking better care of girls in regions that produce tea. Girls would be more likely to survive, and this would translate into a relatively lower share of males in those region after the reform, thus justifying a difference in difference approach.

Let $POST$ be a dummy for post reform, and TEA be a dummy for whether the region produces tea. Let Y_{it} be the fraction of boys in region i at time t

She runs the following regression:

$$Y_{it} = \beta_0 + \beta_1 TEA_i + \beta_2 POST_t + \beta_3 POST_t * TEA_i + \epsilon_{it}$$

Question 17

1.0/1.0 point (graded)

This question has 3 parts:

Given the regression she runs, which of the following denotes the the average fraction of males in tea-regions, pre-reform?

- β_0
- β_1
- β_2
- β_3
- (•) $\beta_0 + \beta_1$
- $\beta_1 + \beta_2$
- $\beta_2 + \beta_3$

Given the regression she runs, which of the following denotes the the average fraction of males in non-tea regions, pre-reform?

- (•) β_0
- β_1
- β_2
- β_3
- $\beta_0 + \beta_1$
- $\beta_1 + \beta_2$
- $\beta_2 + \beta_3$

In this strategy, which coefficient gives her the causal effect of growing tea on the average fraction of males?

- β_0
- β_1
- β_2
- (•) β_3
- $\beta_0 + \beta_1$
- $\beta_1 + \beta_2$
- $\beta_2 + \beta_3$

Explanation

Part I. $\mathbb{E}[Y_{it}|TEA_i = 1, POST = 0] = \beta_0 + \beta_1$

Part II. $\mathbb{E}[Y_{it}|TEA_i = 0, POST = 0] = \beta_0$

Part III. $\text{causal effect} = [\mathbb{E}[Y_{it}|TEA_i = 1, POST = 1] - \mathbb{E}[Y_{it}|TEA_i = 1, POST = 0] -$
$[\mathbb{E}[Y_{it}|TEA_i = 0, POST = 1] - \mathbb{E}[Y_{it}|TEA_i = 0, POST = 0]]$
$= [\beta_0 + \beta_1 + \beta_2 + \beta_3 - \beta_0 - \beta_1] - [\beta_0 + \beta_2 - \beta_0]$
$= \beta_2 + \beta_3 - \beta_2$
$= \beta_3$

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 18

1.0/1.0 point (graded)

True or False? Instead of including the TEA dummy, she could include one dummy for each of the regions (excluding one) to account for inherent differences between regions.

- (•) True
- False

Explanation

Including the TEA dummy controls for inherent differences between regions that grow tea and regions that don't grow tea. However, including region fixed effects controls for inherent differences between regions.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 19

1.0/1.0 point (graded)

This question has 2 parts.

Table 4 – Differences-in-Differences Estimates of the Effect of Planting Tea and Orchards on Sex Ratios:
Coefficients of the Interactions between Dummies Indicating Whether a Cohort was Born Post Reform and Dummies Indicating Whether Any Tea Was Planted in the County of Birth

	Dependent Variable : Fraction of Male			
	(1)	(2)	(3)	(4)
Tea * Post	-0.0081	-0.0086	-0.0074	-0.0074
	(0.0024)	(0.0026)	(0.0026)	(0.0026)
Orchard * Post			0.0096	0.0093
			(0.0033)	(0.0033)
Cashcrop * Post		0.0007	-0.0016	-0.0016
		(0.0007)	(0.0011)	(0.0011)
Han	N	N	N	Y
Observations	49082	49082	49082	49082
R-squared	0.09	0.09	0.09	0.09

All regressions include county fixed effect and controls for post and cash crops *post.
Orchard and cashcrop are dummy variables for the amount of orchards and cashcrop planted in each county.
Post = 1 for cohorts born 1979-1990.
Standard errors clustered at county level.

I. Look at column 1 in the table above, what is the t-statistic for the hypothesis H_0 that the coefficient on tea*post is zero? *Please round your answer to two decimal points.*

-3.38 ✔ Answer: -3.38

-3.38

II. What is the 90 confidence interval for the coefficient tea*post. Enter the lower and upper bounds on the interval $[a, b]$.

Please round your answer to 3 decimal points

a :

-0.012 ✔ Answer: -0.012048

-0.012

b :

-0.004 ✔ Answer: -0.004152

-0.004

Explanation

The t-statistic is the point estimate divided by its standard error. In this case, the standard error is in shown in parenthesis.

$$t - statistic = -0.0081/0.0024$$

The CI is: $a = -0.0081 - 1.645 * 0.0024 = -0.012048$ $b = -0.0081 + 1.645 * 0.0024 = -.004152$

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 20

1.0/1.0 point (graded)

We only observe tea production in the regions that have *chosen* to produce tea. Your friend who is an anthropology major argues that in some regions, people are more likely to prefer girls, for historical reasons. Could these regions then decide to grow more tea?

If yes, what assumption underlying this design strategy would this violate?

- the independence assumption
- (•) the parallel trends assumption
- the exclusion restriction
- None of the above ✔

Explanation

If regions that grew tea and regions that did not grow tea, would follow different trends at the time of the policy change, then even they maintained parallel trends before the reform. The effect of tea growth would be conflated with this inherent difference.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 21

1.0/1.0 point (graded)

To solve this problem, Qian uses the fact that some regions are more suitable to tea production than others: in particular, a certain amount of rain, elevation and slopes are needed to produce tea. She decides to propose an instrumental variables strategy.

What is the first stage equation?

- (•) An OLS regression of TEA on rain, elevation, and slope.
- An OLS regression of the fraction of boys on rain, elevation, and slope.
- An OLS regression of TEA on rain, elevation, and slope with fixed effects.
- An OLS regression of the fraction of boys on rain, elevation, and slope including region fixed effects and a dummy for whether or not the country grows Orchards.

Explanation

Qian's proposed IV strategy is to use geographic features (rain, elevation, and slope) as an instrument for whether or not a region grows tea. So the first stage is an OLS regression of TEA (the variable that is instrumented for) on the instruments.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 22

1.0/1.0 point (graded)

What assumptions are needed for the instrument to be a good instrument?

- Rain, elevation, and slope don't affect the fraction of boys except through tea growth.
- Rain, elevation, and slope affect whether or not a region grows tea.
- Rain, elevation, and slope vary randomly across regions that grow tea and regions that don't.
- (•) All of the above.

Explanation

Refer back to the lecture questions for the past 2 lectures- all of the above are necessary conditions for the instrument's validity.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Long Question 6 - More on Flowers in China

Bookmark this page

Comprehensive Review due Jul 31, 2024 07:30 CST Completed
Download the data set used in Qian's paper (qian.csv). The data contains the following variables:

- **admin**: an id for each region in China.
- **birthyear**: a variable that corresponds to year.
- **sex**: the sex ratio $\left(\frac{male}{female}\right)$ that were born in that region in that year.
- **teasown**: whether tea is produced in region *j*.

Load the data in R and now answer the following questions:

Question 23
1.0/1.0 point (graded)

Explore the data and input the following variables:

Number of observations:

51766 ✔ Answer: 51766

$\mathbf{51766}$

Mean of $\boldsymbol{birthyear}$:

1976 ✔ Answer: [1975.9, 1976.1]

$\mathbf{1976}$

75$^{\boldsymbol{th}}$ percentile of $\boldsymbol{sex}$:

Please round your answer to the second decimal place, i.e. if your answer is 0.1287, round to 0.13, if it is 0.1223, round to 0.12

0.56 ✔ Answer: [0.555, 0.565]

$\mathbf{0.56}$

Maximum value of $\boldsymbol{teasown}$ **:**

1 ✔ Answer: 1

$\mathbf{1}$

Explanation
You can use the command $\boldsymbol{\mathtt{summary}}$ to calculate these variables. A complete R code will be posted once the final is complete.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 24
1.0/1.0 point (graded)

Create a variable $\boldsymbol{post = 1}$ if $\boldsymbol{birthyear >= 1979}$. Similarly, create the interaction between $\boldsymbol{teasown}$ and this variable.

In how many observations is the dummy post switched on?

In part 2, please round your answer to the third decimal place, i.e. if your answer is 0.1245, round to 0.125 and if it is 0.1243, round to 0.124

Observations:

21309 ✔ Answer: 21309

$\mathbf{21309}$

What is the mean of the interaction?

0.081 ✔ Answer: [0.075, 0.0815]

Explanation
You can use the command $\boldsymbol{\mathtt{summary}}$ to find these values. A complete R code will be posted once the final is complete.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 25
1.0/1.0 point (graded)

Estimate the following model in R:

$$sex_{jt} = \beta_0 + \beta_1 teasown_j + \beta_2 post_t + \beta_3 teasown_j \times post_t + \varepsilon_{jt}$$

Based on your estimation input the following values:

Please round your answer to the third decimal place, i.e. if your answer is 0.1245, round to 0.125 and if it is 0.1243, round to 0.124

$\hat{\beta}_0$:

0.503 ✔ Answer: [0.5025, 0.5035]

$\mathbf{0.503}$

$\hat{\beta}_3$:

-0.009 ✔ Answer: [-0.0095, -0.0085]

$\mathbf{-0.009}$

p-value: $\boldsymbol{H_0 : \beta_3 = 0}$:

0.004 ✔ Answer: [0.0035, 0.0045]

$\mathbf{0.004}$

R^2

0.005 ✔ Answer: [0.0045, 0.0055]

$\mathbf{0.005}$

Explanation
You can use the command $\boldsymbol{\mathtt{lm}}$ to estimate this model. A complete R code will be posted once the final is complete.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 26
1.0/1.0 point (graded)

True or False: If you estimate this model instead (where $\boldsymbol{\gamma}$ is a set of region fixed effects) instead:

$$sex_{jt} = \alpha_0 + \alpha_2 post_t + \alpha_3 teasown_j \times post_t + \gamma_j + \varepsilon_{jt}$$

you would have $\hat{\beta}_3 = \hat{\alpha}_3$?

- True
- ✔ False

Explanation
The first model is imposing the same estimated average in the pre period to all the regions that produce tea. This model allows to have a different intercept for each region in the pre period.

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Question 27
1.0/1.0 point (graded)

Go through the R documentation and estimate this fixed effects model. Which of the following statements are true? (Select all that apply)

Do not use the absolute values of the point estimates

- ✔ Our point estimates show that $\hat{\alpha}_3 \geq \hat{\beta}_3$.
- Our point estimates show that $\hat{\alpha}_3 \leq \hat{\beta}_3$.
- ✔ The p-value associated to $\boldsymbol{H_0 : \beta_3 = 0}$ is larger than the p-value associated to $\boldsymbol{H_0 : \alpha_3 = 0}$
- The p-value associated to $\boldsymbol{H_0 : \beta_3 = 0}$ is smaller than the p-value associated to $\boldsymbol{H_0 : \alpha_3 = 0}$

✔

Explanation
For $\hat{\beta}_3$ we have that:

$$teapost \quad -0.0086573 \; 0.0029746 \quad -2.910 \; 0.00361 \; **$$

For $\hat{\alpha}_3$ we have that:

$$teapost \quad -0.008369 \; 0.002853 \quad -2.934 \; 0.00335 \; **$$

Show answer

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Previous Next