**STATS PROJECT REPORT**

**Group No.2**

**Group Members:**

**i)Aditya Anand ii) Prasoon Bajpai iii) Rohit iv) Sagar Mazumdar**

**Section A**

**Q. What are the attitude towards carbon pricing?**

**Method:** We are going to fix a variable which reflects the respondents attitude towards carbon pricing say “rich\_pay” and will try to analyse if there is any change in repondents attitude as their age/gender/income/education level change. We are going to make use of two sample z test to measure the change in attitude.

**I)Does gender affect attitude towards carbon pricing?**

Results: Zstat= 1.36 (**Refer R code for calculation of Zstat**)

**Setting up the hypothesis:**

We want to measure the change in respondents attitude as their gender change. So we are going to setup the hypothesis as:

H0: P1 = P2

Ha: P1 != P2

P1: proportion of population(male) agreeing to maximum burden of carbon tax on rich

P2: proportion of population(female) agreeing to maximum burden of carbon tax on rich

**Zcritical(alpha=.05) = 1.96**

Since, |Zstat | < |Zcritical|,we fail to reject the null hypothesis. Therefore, we conclude that in our sample respondents voting for a higher tax on the rich is independent of their gender.

**proportion test power calculation**:

n1 = 80, n2=38, d = 0.09080225, sig.level = 0.05 , **power = 0.0742581**

**calculating number of observations needed for power of the test to be 0.8:**

**n = 1904.862**, delta = 0.09080225, sig.level = 0.05, **power = 0.8**

**II)Does level of education affect attitude towards carbon pricing?**

Result: Zstat= -1.91 (**Refer R code for calculation of Zstat**)

**Setting up the hypothesis:**

We want to measure the change in respondents attitude as their maximum education level change. So we are going to setup the hypothesis as:

H0: P1 = P2

Ha: P1 != P2

P1: proportion of population with senior secondary and undergraduate level of education agreeing to maximum burden of carbon tax on rich

P2: proportion of population with PG and Phd level of education agreeing to maximum burden of carbon tax on rich

**Zcritical(alpha=.05) = 1.96**

Since |Zstat | < |Zcritical|, we fail to reject the null hypothesis. Therefore, we conclude that in our sample respondents voting for a higher tax on the rich and their level of education has no correlation.

**proportion test power calculation**:

n1 = 73, n2=45, d = 0.08694618, sig.level = 0.05 , **power = 0.07403125**

**calculating number of observations needed for power of the test to be 0.8:**

**n = 2077.484**, delta = 0.08694618, sig.level = 0.05, **power = 0.8**

**III)Does income level affect attitude towards carbon pricing?**

Results: X-squared = 3.2724, df = 2, p-value = 0.1947 (**Refer R code for calculation of** X-squared)

**Setting up the hypothesis:**

We want to measure if the respondents attitude towards carbon pricing is independent of their income levels. So we are going to setup the hypothesis as:

H0: income and attitude towards carbon pricing are independent

Ha: income and attitude towards carbon pricing are dependent

**X2critical(alpha=.05,df=2) = 5.991**

Since, X-squared < **X2critical(alpha=.05,df=2),**we fail to reject the null hypothesis. Therefore, we conclude that in our sample, respondents voting for a higher tax on the rich has no correlation with their level of income .

**chisq test power calculation**:

w = 0.1177544, N = 118, df = 2, sig.level = 0.05 , **power = 0.1910785**

**calculating number of observations needed for power of the test to be 0.8:**

w = 0.1177544, **N = 694.8376**, sig.level = 0.05, **power = 0.8**

**IV)Does age affect attitude towards carbon pricing?**

Results: Zstat= -0.051(**Refer R code for calculation of Zstat**)

**Setting up the hypothesis:**

We want to measure the change in respondents attitude as there is a change in their age. So we are going to setup the hypothesis as:

H0: P1 = P2

Ha: P1 != P2

P1: proportion of population with age b/w 18 and 24 agreeing to maximum burden of carbon tax on rich

P2: proportion of population with age not b/w 18 and 24 agreeing to maximum burden of carbon tax on rich

**Zcritical(alpha=.05) = 1.96**

Since |Zstat | < |Zcritical|, we fail to reject the null hypothesis. Therefore, we conclude that in our sample age of respondents and them voting for a higher tax on the rich are independent of each other .

**proportion test power calculation**:

n1 = 86, n2=32, d = 0.09974626, sig.level = 0.05 , **power = 0.07652935**

**calculating number of observations needed for power of the test to be 0.8:**

**n = 1578.734**, delta = 0.09974626, sig.level = 0.05, **power = 0.8**

**Section B**

Now, in the following section we will change the attitude variable to “maximum % tax that respondents are willing to pay as carbon tax” will analyse if their attitude change as independent variables such as gender/education/age/income change.

**Method:** The method of analysis is going to remain the same as used in section A. We are going to divide the responses into 2 categories based on max tax % of income that they are willing to pay i)below 5% ii) above 5%

**I)Does gender affect attitude towards carbon tax rate?**

Results: Zstat= 0.112(**Refer R code for calculation of Zstat**)

**Setting up the hypothesis:**

We want to measure the change in respondents attitude as their gender change. So we are going to setup the hypothesis as:

H0: P1 = P2

Ha: P1 != P2

P1: proportion of population(male) who want tax rate b/w 0-5%

P2: proportion of population(female) who want tax rate b/w 0-5%

**|Zcritical(alpha=.05)| = 1.96**

Since |Zstat | < |Zcritical|, we fail to reject the null hypothesis. Therefore, we conclude that in our sample respondents attitude towards carbon pricing is independent of their gender.

**proportion test power calculation**:

n1 = 80, n2=37, d = 0.05429366, sig.level = 0.05 , **power =** **0.05844329**

**calculating number of observations needed for power of the test to be 0.8:**

**n = 5326.203**, delta = 0.05429366, sig.level = 0.05, **power = 0.8**

Therefore, we need about 5326 observation for us to conclude correctly ,80% of the time, if there is correlation b/w respondents gender and their willingness to pay a carbon tax.

**II)Does living in city affect the willingness to pay a carbon tax?**

Results: Zstat= -3.11(**Refer R code for calculation of Zstat**)

**Setting up the hypothesis:**

We want to measure the change in respondents willingness to pay caron tax change as their place of living change. So we are going to setup the hypothesis as:

H0: P1 = P2

Ha: P1 != P2

P1: proportion of population (living in city) who want tax rate b/w 0-5%

P2: proportion of population (living in village) who want tax rate b/w 0-5%

**Zcritical(alpha=.05) = 1.96**

Since, |Zstat|> |Zcritical|, we reject the null hypothesis. Therefore, we conclude that in our sample respondents willingness to pay carbon tax and their location are dependent on each other.

**proportion test power calculation**:

n1 = 113, n2=4, d = 0.05429366, sig.level = 0.05 , **power =** **0.05028488**

**calculating number of observations needed for power of the test to be 0.8:**

Number of observation = 23985.8. Therefore, we need about 23986 observations for us to conclude correctly ,80% of the time, if there is correlation b/w respondents location and their willingness to pay a carbon tax.

**III)Does age affect attitude towards carbon tax rate?**

Results: Zstat= -1.39(**Refer R code for calculation of Zstat**)

**Setting up the hypothesis:**

We want to measure the change in repondents attitude as there is a change in their age. So we are going to setup the hypothesis as:

H0: P1 = P2

Ha: P1 != P2

P1: proportion of population with age b/w 18 and 24 who want tax rate b/w 0-5%

P2: proportion of population with age not b/w 18 and 24 who want tax rate b/w 0-5%

**Zcritical(alpha=.05) = 1.96**

Since |Zstat | < |Zcritical|, we fail to reject the null hypothesis. Therefore, we conclude that in our sample respondents maximum willingness to pay a carbon tax and their age are independent of each other.

**proportion test power calculation**:

n1 = 86, n2=32, d = 0.0443751, sig.level = 0.05 , **power=** **0.05519034**

**calculating number of observations needed for power of the test to be 0.8:**

Number of observation = 7972. Therefore, we need about 7972 observations for us to conclude correctly ,80% of the time, if there is correlation b/w respondents age and their willingness to pay a carbon tax

**IV)Does income level affect attitude towards carbon tax rate?**

For analysing the affect of income on the maximum willingness to pay carbon tax we are going to make use of chi-square of independence

**Sample R code for calculating CHIstat:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Tax rate** | **0-5lpa** | **5-20lpa** | **20lpa+** |
| **0-5%** | **31** | **53** | **24** |
| **>5%** | **3** | **5** | **1** |

Results: X-squared = 0.61168, df = 2, p-value = 0.7365

**Setting up the hypothesis:**

We want to measure the change in respondents attitude as their income levels change. So we are going to setup the hypothesis as:

H0: respondents income and their willingness to pay tax are independent

Ha: respondents income and their willingness to pay tax are not independent

**Xcritical(alpha=0.05, df=2) = 5.991**

Since, X-squared < **Xcritical**, we fail to reject the null hypothesis. Therefore, we conclude that respondents income and their willingness to pay carbon tax are independent.

**chisq test power calculation**:

w = 0.05088205, N = 118, df = 2, sig.level = 0.05 , **power = 0.07370586**

**calculating number of observations needed for power of the test to be 0.8:**

Number of observation = 3721. Therefore, we need about 3721 observations for us to conclude correctly ,80% of the time, if there is correlation b/w respondents income and their willingness to pay a carbon tax

**Q. Are the attitudes randomly distributed?**

From the results we got from our tests on attitude variable (such as amount and rich\_pay) and variables such as age/gender/income/education we observe that the attitudes of the respondent towards carbon pricing are independent of their age/gender/income/education. However, as we saw in Section B.II respondents location and willing to pay a carbon tax are dependent. As the number of respondents who live in village this dependence cannot be representative of the Indian population at large.

**Section C**

**Q. Is the sample representative of the average Indian?**

In section of the report, we will try to analyse whether the survey response is representative of the average Indian.

**Method:** We are going to make use of simple Z teststo check if the sample data is representative of the average Indian. To perform the Z test we are going to use the following variables:

i)age ii) gender iii) income iv) education

**I)Is the mean age of the sample representative of mean age of Indian population?**

**Result: Zstat= 6.73**

**Setting up the hypothesis:**

We want to analyse whether average age of the sample representative of mean age of Indian population. So we are going to setup the hypothesis as:

H0: P = .4302

Ha: P != .4302

P: census data on proportion of population b/w the age 18-24 -> .4302

**Zcritical(alpha=.05) = 1.96**

Since |Zstat | > |Zcritical|, we reject the null hypothesis. Therefore, we conclude that in our mean age respondents is not representative of mean age of population.

**Power of test calculation:**

**Result: Power = 0.088**

The above result means that we have an approximately 8.8% chance of making a correct decision of rejecting the null hypothesis when the null hypothesis is false when the level of significance is 5%.

**Result for number of observations with power of test set to 0.8:**

Number of observations = 2409.367. This means that we need approx. 2410 observations for correctly rejecting null hypothesis when the null hypothesis is false 80% of the time.

**II)Is the proportion of gender in the sample representative of Indian population?**

**Result: Zstat= 3.524**

**Setting up the hypothesis:**

We want to analyse whether gender proportions representative of Indian population. So we are going to setup the hypothesis as:

H0: P = .51511

Ha: P != .51511

P: census data on proportion of male population -> .51511

**Zcritical(alpha=.05) = 1.96**

Since |Zstat | > |Zcritical|, we reject the null hypothesis. Therefore, we conclude that in our sample is not representative of Indian population in terms of gender proportions.

**Power of test calculation:**

**Result: Power = 0.0508**

The above result means that we have an approximately 5.08% chance of making a correct decision of rejecting the null hypothesis when the null hypothesis is false when the level of significance is 5%.

**Result for number of observations with power of test set to 0.8:**

Number of observations = 2434.185. This means that we need approx. 2434 observations for correctly concluding, that proportion of male population is not 0.51511 ,80% of the time.

**III)Is the maximum education level in the sample representative of Indian population?**

**Result: Zstat= 24.31**

**Setting up the hypothesis:**

We want to analyse whether maximum education level of the respondents representative of Indian population. So we are going to setup the hypothesis as:

H0: P = .1526

Ha: P != .1526

P: census data on proportion of population with minimum undergraduate level of education

-> .51511

**Zcritical(alpha=.05) = 1.96**

Since |Zstat | > |Zcritical| we reject the null hypothesis. Therefore, we conclude that our sample is not representative of average education level of Indian population.

**Power of test calculation:**

**Result: Power = 0.6063**

The above result means that we have an approximately 60.63% chance of making a correct decision of rejecting the null hypothesis when the null hypothesis is false when the level of significance is 5%.

**Result for number of observations with power of test set to 0.8:**

Number of observations = 1261.168. This means that we need approx. 1261 observations for correctly concluding, that proportion of Indian population with maximum UG level of education is not .1526 ,80% of the time.

**IV)Is the income level of the respondents representative of Indian population?**

**Result: X-squared = 73.28, df = 2**

**Setting up the hypothesis:**

We want to analyse whether income level of the respondents representative of Indian population. So, we are going to setup the hypothesis as:

H0: proportional representation of different income groups in sample is same as proportional representation of different income groups in population

Ha: proportional representation of different income groups in sample is not same as proportional representation of different income groups in population

**Xcritical(alpha=0.05, df=2) = 5.991**

Since, **X-squared > Xcritical(alpha=0.05, df=2),**we reject the null hypothesis. Therefore, we conclude that proportional representation of different income groups in sample is not same as proportional representation of different income groups in population

**Power of test calculation:**

**Result: Power = 0.99**

The above result means that we have an approximately 99% chance of making a correct decision of rejecting the null hypothesis when the null hypothesis is false when the level of significance is 5%.

**Result for number of observations with power of test set to 0.8:**

Number of observation = 31.This means that we need about 31 observations for correctly concluding proportional representation in sample is not that same as proportional representation in population w.r.t different income group 80% of the time.