Econ B2000

NAME: Jason Seda

**Homework 1**

**Group:**

Michael Stewart

Kerra Sinanan

Jason Seda

**Dice results:**

**Hypothesis:**

**H0- Result will return a 6**

**H1- Result will return any number other than a 6**

**Results with control Die:**

**4/20 returned result confirming a 6 outcome**

**0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1,1,0,0**

**Results with altered Die:**

**5/20 returned result of a 6**

**0,1,0,0,1,1,1,0,0,0,1,0,0,0,0,0,0,0,0,0**

| > load("d\_HHP2020\_24.RData")  > #glimpse(d\_HHP2020\_24) try this later  > d\_HHP2020\_24[1:10,1:6]  Age Gender Education Mar\_Stat income\_midpoint Race  1 34 female college grad Married 62500 white  2 65 male some college divorced 30000 white  3 44 female college grad Married 225000 other  4 56 male some college divorced 12500 white  5 57 female adv degree never 62500 white  6 44 female adv degree Married 125000 white  7 37 female adv degree Married 62500 Black  8 59 male college grad Married 82500 white  9 51 female lt hs never 12500 Black  10 29 female assoc deg Married 40000 white  > attach(d\_HHP2020\_24)  > summary(d\_HHP2020\_24)  Age Gender Education  Min. :17.00 male :410536 lt hs : 6787  1st Qu.:39.00 female:566464 some hs : 14934  Median :52.00 trans : 1989 high school :122541  Mean :52.25 other : 5801 some college:210698  3rd Qu.:65.00 assoc deg :103575  Max. :88.00 college grad:279400  adv degree :246855  Mar\_Stat income\_midpoint Race  Married :556611 Min. : 12500 white:806002  widowed : 54162 1st Qu.: 40000 Black: 80846  divorced :152705 Median : 82500 Asian: 48885  separated: 17850 Mean : 95461 other: 49057  never :195037 3rd Qu.:125000  NA's : 8425 Max. :225000  NA's :187771  Hispanic Number\_people\_HH Number\_kids\_HH  not Hispanic:895979 Min. : 1.000 Min. :0.000  Hispanic : 88811 1st Qu.: 2.000 1st Qu.:0.000  Median : 2.000 Median :0.000  Mean : 2.715 Mean :0.623  3rd Qu.: 4.000 3rd Qu.:1.000  Max. :10.000 Max. :5.000    Number\_adults\_HH private\_health\_ins  Min. : 1.000 0 : 74413  1st Qu.: 2.000 has private health insurance:607599  Median : 2.000 no private health insurance :149384  Mean : 2.092 NA's :153394  3rd Qu.: 2.000  Max. :10.000    public\_health\_ins  0 : 74413  has public health insurance:302958  no public health insurance :425600  NA's :181819        work\_kind  employed by govt : 96450  employed by private co :320047  employed by nonprofit or charity: 74364  self employed : 68547  work for family business : 11698  NA's :413684    workloss DOWN  yes recent household loss of work:171404 Min. :1.000  no :794667 1st Qu.:1.000  NA's : 18719 Median :1.000  Mean :1.634  3rd Qu.:2.000  Max. :4.000  NA's :108234  ANXIOUS WORRY INTEREST  Min. :1.000 Min. :1.000 Min. :1.000  1st Qu.:1.000 1st Qu.:1.000 1st Qu.:1.000  Median :2.000 Median :1.000 Median :1.000  Mean :1.906 Mean :1.718 Mean :1.649  3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000  Max. :4.000 Max. :4.000 Max. :4.000  NA's :106951 NA's :108419 NA's :108683  YEAR Begin\_Date K4SUM  Min. :20.00 Min. :2020-04-23 Min. : 4.000  1st Qu.:20.00 1st Qu.:2020-12-09 1st Qu.: 4.000  Median :22.00 Median :2022-04-27 Median : 6.000  Mean :21.73 Mean :2022-05-03 Mean : 6.908  3rd Qu.:23.00 3rd Qu.:2023-08-23 3rd Qu.: 8.000  Max. :24.00 Max. :2024-07-23 Max. :16.000  NA's :111831  income\_midpoint\_factor  125000 :145006  62500 :134183  82500 :112727  225000 : 92900  40000 : 85421  (Other):226782  NA's :187771  > summary(Age[Gender == "female"])  Min. 1st Qu. Median Mean 3rd Qu. Max.  17.00 39.00 52.00 51.62 64.00 88.00  > summary(Age[Gender == "male"])  Min. 1st Qu. Median Mean 3rd Qu. Max.  17.00 40.00 54.00 53.29 67.00 88.00  > summary(Age[Gender == "trans"])  Min. 1st Qu. Median Mean 3rd Qu. Max.  17.00 26.00 31.00 36.02 41.00 88.00  > summary(Age[Gender == "other"])  Min. 1st Qu. Median Mean 3rd Qu. Max.  17.00 31.00 43.00 45.88 59.00 88.00  > mean(Age[Gender == "female"])  [1] 51.61668  > sd(Age[Gender == "female"])  [1] 15.59165  > mean(Age[Gender == "male"])  [1] 53.28593  > sd(Age[Gender == "male"])  [1] 16.28551 |
| --- |
|  |
|  |

4. After reviewing the data, it is difficult to fully trust the accuracy of this dataset. While statistical correlations can certainly be performed, how reliable are the results when a significant portion of the data contains missing values (NAs)? Would the removal of these NAs and potentially tainted entries strengthen the validity of the correlations by leaving only complete cases, thereby offering a clearer indication of the sample size? Or, conversely, would excluding the NAs reduce the dataset to the point where meaningful conclusions become limited or even unattainable?