assignment-1

Raul Rodriguez

2023-05-23

assignment question 1

A farmer wants to determine the proportion of pumpkin seeds planted in her field that successfully grow into pumpkins. It would take too much time to count the total number of seeds planted in the field and the total yield of pumpkins. Thus, she decides that she needs to take a sample to estimate this proportion.

a. State the population and the variable of interest to this farmer. The population in this case are the pumpkins in the farmer's field. The variable of interest is the proportion of seeds that successfully turn into pumpkins.

b. Give an example of a way the farmer could perform a convenience sample. A convenience sample for the farmer is to count an area that is convenient for her. A good example of this scenario is the farmer counting the pumpkins and seeds along the base and height of her field. She would then calculate: seeds_base * seeds_height = total_number_of_seeds in field pumpkins_base * pumpkins_height = total_number_of_pumpkins in field total_number_of seeds in field / total_number_of_pumpkins = seed to pumpkin ratio.

c. Give an example of a way the farmer could perform a simple random sample. For a simple random sample, the farmer would choose 4 random patches of land (assuming that each patch is 25 squared cm.) This would give the farmer a total of 1 square meter of land that is formed from random samples across her field. She can then use the seed to pumpkin ratio in that patch to calculate the rest of her field's ratio.

d. Give an example of a way the farmer could perform a stratified random sample. A simple random sample would be for the farmer to count the seeds and pumpkins in a square meter of her field (strata). This sample represents a part of the population that She can use to calculate the **seed to pumpkin ratio in a square meter of the field**. This result can be used to calculate the total seed to pumpkin ratio by doing the following: base_of_field * height_of_field * square_meter_of_seeds = total_seeds base_of_field * height_of_field * square_meter_of_pumpkins = total_pumpkins total_seeds / total_pumpkins = seed to pumpkin ratio

e. What is the population parameter of interest? What would be a good statistic to use to estimate this? The population of interest are the pumpkins because the farmer wants to know what is the pumpkin yield out of a determined number of seeds, specifically, the number of seeds that turn into pumpkins.

assignment question 2

a. Describe what information is contained in the data set. What command in R did you use to find this information?

The Lynx data set contains numeric data. I used the is.numeric() method to find the information.

```
## [1] TRUE
```

b. Create a character vector called years which contains the years of the trappings.

according to (stat.ethz)[https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/lynx.html (https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/lynx.html)] "Annual numbers of lynx trappings for 1821–1934 in Canada.".

```
##   [1] 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835
##  [16] 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850
##  [31] 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865
##  [46] 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880
##  [61] 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895
##  [76] 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910
##  [91] 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925
## [106] 1926 1927 1928 1929 1930 1931 1932 1933 1934
```

c. Set the names of the lynx vector equal to years.

```
## Time Series:
## Start = 1821
## End = 1934
## Frequency = 1
## 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836
##  269  321  585  871 1475 2821 3928 5943 4950 2577  523   98  184  279  409 2285
## 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852
## 2685 3409 1824  409  151   45   68  213  546 1033 2129 2536  957  361  377  225
## 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868
##  360  731 1638 2725 2871 2119  684  299  236  245  552 1623 3311 6721 4254  687
## 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884
##  255  473  358  784 1594 1676 2251 1426  756  299  201  229  469  736 2042 2811
## 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900
## 4431 2511  389   73   39   49   59  188  377 1292 4031 3495  587  105  153  387
## 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916
##  758 1307 3465 6991 6313 3794 1836  345  382  808 1388 2713 3800 3091 2985 3790
## 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932
##  674   81   80  108  229  399 1132 2432 3574 2935 1537  529  485  662 1000 1590
## 1933 1934
## 2657 3396
```

d. How many lynx were trapped in 1901? Use years as your index.

```
## 1901
##  758
```

e. What is the average number of annual lynx trappings from 1821 to 1920, inclusive? Hint: You want to compute the average trappings for the first 100 data points.

```
## [1] 188.5
```

assignment question 3

a. Read in the data file and name the data frame casino.

```
##         Name BlackJack  Poker   Slots Roulette   Craps
## 1      Betty     50.46  41.68  262.88  -114.46  106.59
## 2       John      6.80   4.00  212.70    48.46  890.84
## 3     Dwayne    -98.29 -54.82  252.58   -66.82   38.65
## 4     Sophia    183.73  59.49   95.19  -115.82   15.20
## 5      Luisa     43.12  38.79  -10.95  -230.82   29.88
## 6     Carlos     49.40  68.40 -289.88    53.92 -275.19
## 7     Andrew    343.51 -59.21   74.33    85.42 -577.07
## 8  Charlotte    284.26  24.08  -67.36    72.55 -370.38
## 9      Calum   -233.37   7.29  -76.96    84.13 -519.10
## 10     Layla    -81.39  42.39   -6.81   -59.80 -110.55
```

b. Use the head() function to determine the games these friends played in the casino.

```
##     Name BlackJack  Poker   Slots Roulette   Craps
## 1  Betty     50.46  41.68  262.88  -114.46  106.59
## 2   John      6.80   4.00  212.70    48.46  890.84
## 3 Dwayne    -98.29 -54.82  252.58   -66.82   38.65
## 4 Sophia    183.73  59.49   95.19  -115.82   15.20
## 5  Luisa     43.12  38.79  -10.95  -230.82   29.88
## 6 Carlos     49.40  68.40 -289.88    53.92 -275.19
```

c. Create a character vector called friends which contains the values from the first column of the data set.

```
##  [1] "Betty"     "John"      "Dwayne"    "Sophia"    "Luisa"     "Carlos"
##  [7] "Andrew"    "Charlotte" "Calum"     "Layla"
```

d. Using the R command as.matrix(), create a matrix called winnings which contains all the columns except the first one from the casino data set.

```
##       BlackJack  Poker   Slots Roulette   Craps
##  [1,]     50.46  41.68  262.88  -114.46  106.59
##  [2,]      6.80   4.00  212.70    48.46  890.84
##  [3,]    -98.29 -54.82  252.58   -66.82   38.65
##  [4,]    183.73  59.49   95.19  -115.82   15.20
##  [5,]     43.12  38.79  -10.95  -230.82   29.88
##  [6,]     49.40  68.40 -289.88    53.92 -275.19
##  [7,]    343.51 -59.21   74.33    85.42 -577.07
##  [8,]    284.26  24.08  -67.36    72.55 -370.38
##  [9,]   -233.37   7.29  -76.96    84.13 -519.10
## [10,]    -81.39  42.39   -6.81   -59.80 -110.55
```

e. Create a vector called totals which contains the row sums of the matrix winnings. What do the values in this vector represent?

```
##  [1]  347.15 1162.80   71.30  237.79 -129.98 -393.35 -133.02  -56.85 -738.01
## [10] -216.16
```

f. Set the names of the vector totals equal to friends.

```
##     Betty      John    Dwayne    Sophia     Luisa    Carlos    Andrew Charlotte
##    347.15   1162.80     71.30    237.79   -129.98   -393.35   -133.02    -56.85
##     Calum     Layla
##   -738.01   -216.16
```

g. Use the R functions min(), max(), which.max() and which.min() to determine which friend won the most money and which friend lost the most money in the casino.

```
## [1] "Andrew"
```

```
## [1] "Sophia"
```

h. What was the average amount of money won or lost by the group of friends on the trip? Page 2

```
## [1] 15.167
```