Statistics HW9 b03705002 林軒逸

5.3

a.

Step1: My study purpose is “The relationship between smoking and lung cancer”.

Step2: I’ll do questionnaire to people, asking them whether they smoke and have lung cancer.

Step3: I’ll present my result with a pie chart of proportion of having lung cancer in the people that smoke.

Step4: My questionnaire will be very simple. Whether you smoke or not, and whether you have had lung cancer.

Step5: Although many people might not know they have lung cancer or not, I’ll provide an option is “Don’t know” for it, and that questionnaire will be abandon.

Step6: Collect the data.

Step7: Error Controlling.

Step8: Making out an document of “Relationship between smoking and lung cancer”.

b.

Observational. Survey by questionnaire is a kind of observational study.

5.9

Since the sample is the readers of the newspaper, but the sample mentioned here is ”The people that are interested in the topic and willing to send a email of yes/no”. Thus, the sample is not right, that is, a self-selected sample, that has bias on choosing the samples. As a result, this statistic should be ignored.

5.13

First, we’ll randomly choose students in three groups, school of business, faculty of art and sciences, and the graduate school. In this example, the president wants to know the information of these groups, thus we choose samples from this groups. In determining the response of the whole school, we can use the samples we have chosen, and this is called a stratification random sampling, however, in the result, we will get bigger sampling errors. If the fund is enough, the president can do another random sampling to present the response of the students, and mention that the samples are different. Then, this result will have a smaller sampling errors from the stratification random sampling one.

5.16

No. Since a census is doing a study by the population. And the results we do is to use random samples to estimate the condition of population, thus there will be sampling errors. Thus, I don't think the results we do can be better than census.

5.EX

(1)

Average: 3.644421

Standard Deviation(Population): 0.846394

(2)

Average:3.60375

Standard Deviation(Sample): 0.808601

(3):

Proportion of Female: 60%, 144 samples

Proportion of Male: 40%, 96 samples

Average: 3.625833

Standard Deviation(Sample): 0.767596

(4):

Proportion of 1: 25%, 60 samples

Proportion of 2: 45%, 108 samples

Proportion of 3: 30%, 72 samples

Average: 3.645

Standard Deviation: 0.825468

(PS: There are some recurrences in Age 1 and 3's random numbers, and I've eliminated them and replace with new ones.)

(5):

Sample Errors of Average:

(2) |3.60375-3.644421|= 0.040671

(3) |3.625833-3.644421|= 0.01859

(4) |3.645 -3.644421|= 0.000579

The sampling error of average in (2) is bigger than (3) & (4).

Sample Errors of Standard Deviation:

(2) |0.808601-0.846394|= 0.037793

(3) |0.767596-0.846394|= 0.0788

(4) |0.825468-0.846394|= 0.02093

The sampling error of standard deviation in (3) is bigger than (2), and the error in (2) is bigger than (4).

(6):

First, let's look at the standard deviation of Age and Gender.

Gender:

M: 0.6560084

F: 0.770712

Combined: 0.767596

As we can see, the combined standard deviation is not bigger than both M & F, it implies that there is no special relationship in group M or F. That is, I will conclude it as a cluster random sampling.

Let's calculate the standard deviation of the cluster sampling to calculate the difference between groups.

×Standard Deviation(Cluster): 0.047028

As we can see, the difference between clusters are small. That is, there is little gender difference in the diet. Also, this follows the trait that clusters usually have little difference between each other.

Age:

1: 0.650291113

2: 0.70910722

3: 0.887302845

Combined: 0.825468

As we can see, the combined standard deviation is not bigger than 1, 2, and 3. It implies that there is no special relationship in group 1, 2, or 3. That is, I will conclude it as a cluster random sampling.

Let's calculate the standard deviation of the cluster sampling to calculate the difference between groups.

×Standard deviation(Cluster): 0.048758436

As we can see, the difference between clusters are small. That is, there is little age difference in the diet. Also, this follows the trait that clusters usually have little difference between each other.