Exercises 5

Sample size determination:

Error of estimation

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}} S}{d}\right)^2 \quad or \quad n = \left(\frac{Z_{1-\frac{\alpha}{2}} \sigma}{d}\right)^2 \quad \text{, in case (WR).}$$

$$n = \frac{N\sigma^2}{(N-1)D^2+\sigma^2}, \quad D = \frac{d}{z} \quad \text{, in case (WOR).}$$

Coefficient of variation

$$n = \frac{1}{CV^2}\frac{s^2}{\bar{y}^2}$$

Relative error (Relative Differences)

$$n = \left(\frac{Z_{1-\frac{\alpha}{2}} S}{a\, p}\right)^2 \; or \; n = \frac{Z^2_{1-\frac{\alpha}{2}} \; S^2}{(a\, \bar{y}\,)^2} \quad a\text{: is error should not exceed}$$

[***Note***: We always round up; the sample size formulas always generate the minimum number of subjects needed to ensure the specified precision.]

Example 1: (Sample size estimation for proportion in survey)

Researcher interested to know the sample size for conducting a survey for measuring the prevalence of obesity in certain community. Previous literature gives the estimate of an obesity at 20% in the population to be surveyed, and assuming 95% confidence interval (or 5% level of significance) and 10% Relative error. Find the sample sizes required (WR)?

The sample size can be calculated as follow as

$$p = 0.2\,, 1-\alpha = 0.95\,, \alpha = 0.05$$

$$n = \frac{Z^2_{1-\frac{\alpha}{2}} \; pq}{(a\,p)^2} = \frac{1.96^2 * 0.2 * (1-0.2)}{(0.1 * 0.2)^2} = 1536.64 \approx 1537 \text{ for a simple random sampling design.}$$

Exercises 2: (Sample size estimation for average in survey)

Consider a camera club with 1800 members, where it is required to estimate the average number of rolls of film used during a year. Consider also the information from the past that the average and standard deviation of the number of rolls of film have been around 6 and 4, respectively.

Find the sample sizes required (WR)to estimate the average number of rolls of film with following information:

a) With an error not exceeding 1, and 5% level of significance, with the normal approximation.

$$Z_{0.975} = 1.96 \;, n = \left(\frac{Z_{1-\frac{\alpha}{2}}\, S}{d}\right)^2 = \left(\frac{1.96 * 4}{1}\right)^2 = 61.46 \approx 62$$

b) With C.V. should not exceed 8%.

With the information on the mean and standard deviation, $c_y = c.v = \frac{s}{\bar{y}}$

$$n = \frac{1}{CV^2}\frac{s^2}{\bar{y}^2} = \frac{c.v^2}{CV^2} = \frac{1}{0.08^2}\left(\frac{4}{6}\right)^2 = 69.44 \approx 70$$

c) If the relative error should not exceed (0.1), except for α=0.05.

$$n = \frac{Z^2_{1-\frac{\alpha}{2}}\; S^2}{(a\,\bar{y}\,)^2} = \frac{(1.96 * 4)^2}{(0.1 * 6)^2} = 170.73 \approx 171$$

Exercises 3: (Sample size estimation for average in survey)

The metropolitan area and the suburbs together in a region consist of 5, 10, and 10 thousand families with one, two, and three or more children. For these three types of families, preliminary estimates of the averages and standard deviations of the number of hours of television watching in a week are (10, 15, 20) and (6, 10, 15), respectively.

Find the sample sizes required (WR) to estimate the above average for each group if the error of estimation should not exceed 2 hours in each case, except for α=0.05.

$$Z_{0.975} = 1.96 \;, n = \left(\frac{Z_{1-\frac{\alpha}{2}}\, S}{d}\right)^2$$

$$n_1 = \left(\frac{1.96 * 6}{2}\right)^2 = 34.56 \approx 35$$

$$n_2 = \left(\frac{1.96 * 10}{2}\right)^2 = 96.04 \approx 97$$

$$n_3 = \left(\frac{1.96 * 15}{2}\right)^2 = 216.09 \approx 217$$

We note the sample size are differences, because the standard errors are different from each group.

H.W

Exercises 4. (Sample size estimation for proportion in survey). (H.W)

One university has 1000 students in each of the four classes. The percentages of the Freshman–Sophomore, Junior, and Senior classes expressing interest in professional training after graduation were guessed to be 20, 50, and 80%, respectively.

(a) For each of these three groups, find the sample sizes required to estimate the percentage if the estimate should not differ from the actual value by not more than 20% of the actual value except for α=0.05, and present the reason for the differences in the sample sizes.

(b) Find the sample sizes needed for each of the three groups for estimating the above percentage if the **error of estimation** should not exceed 10% except for α=0.05 and present the reason for the differences in the sample sizes.

Exercises 5 : A survey is to be conducted to estimate the average monthly income of a locality with 5000 households. It is known that the squared variability in the income is 250000 Rayals. How many households should we select (WOR) so that the marginal error of estimated income is no more than 200 Rayals? (use α=0.05)

Solution: we have N= 5000, σ^2 =250000 and d=200

$$n = \frac{N\sigma^2}{(N-1)D^2 + \sigma^2}, \quad D = \frac{d}{z}$$

$D = \frac{200}{z_{0.975}} = \frac{200}{1.96} = 102.0408 \quad , n = \frac{5000*250000}{4999*102.0408^2+250000} = 23.9 \approx 24$

So, a sample of 24 householder is required to conduct the survey.

Exercises 6: It is known that the proportion of smokers in a society of 4000 individuals is 0.2. How many individuals we need to select (WOR) if we want to estimate the proportion of smokers with error in estimated "marginal error" equal to 4%. (use α=0.05)

Solution: we have N= 4000, d=0.04 , P =0.2 so Q=1-0.2=0.8

The sample size for estimation of population mean is given as

$$n = \frac{N\sigma^2}{(N-1)D^2 + \sigma^2}, \quad D = \frac{d}{z}$$

Further, the sample size for estimation of population proportion is given as

$$n = \frac{NPQ}{(N-1)D^2 + PQ}$$

$$D^2 = \left(\frac{d}{Z_{0.975}}\right)^2 = \left(\frac{0.04}{1.96}\right)^2 = 0.0004$$

$$n = \frac{NPQ}{(N-1)D^2 + PQ} = \frac{4000 * 0.2 * 0.8}{(3999 * 0.0004) + (0.2 * 0.8)} = 363.72 \approx 364$$

Hence, we need a sample of 364 individuals to estimate the proportion of smokers.

Exercises 6

#Cost -sample size

Example 1: student was asked to take up the problem about estimate the average time per week devoted to study in University library by the students of this university. He was provided with $150, including overhead cost of $ 24. The cost of contacting the students, and collecting information is $ 3 per student. How many students would he select in the sample, for collecting the desired information.

C(n) = cost of taking n samples =150
c_o = fixed cost= 24
c_1 = cost for each sample interview

$$C(n) = c_o + n * c_1$$

$$n = \frac{C(n) - c_o}{c_1} = \frac{150 - 24}{3} = 42$$

#Power in sample size

Power $(\mathbf{1 - \beta})$	$\boldsymbol{Z_{1-\beta}}$
0.70	**0.524**
0.75	**0.674**
0.80	**0.841**
0.85	**1.036**
0.90	**1.282**
0.95	**1.645**
0.99	**2.326**

Alpha	$\boldsymbol{Z_{1-\frac{\alpha}{2}}}$
0.2	**1.281**
0.15	**1.44**
0.10	**1.645**
0.05	**1.960**
0.01	**2.576**
0.001	**3.29**

NOTE:
» The higher power will require a larger sample size.
» Sample Size for One Sample, Continuous Outcome $H_0: \mu = \mu_1 vs\ H_1: \mu \neq \mu_1$

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 S^2}{(d)^2}, \qquad d = \mu_1 - \mu_0$$

»Sample Sizes for Two Independent Samples, Continuous Outcome $H_0: \mu_1 = \mu_2 and\ H_1: \mu_1 \neq \mu_2$ under the assumption of common variance (i.e equal variability in the two populations)

$$n = \frac{2\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 S^2}{(d)^2}, \qquad d = \mu_1 - \mu_2$$

»Sample Sizes for Two Independent Samples, Dichotomous Outcomes $H_0: p_1 = \mu_2 vs\ H_1: \mu_1 \neq \mu_2$

$$n = \frac{2\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 \bar{p}\bar{q}}{(p_1 - p_2)^2}$$

Where $\bar{p}$ is mean of the proportions in the two comparison groups, assuming that the groups will be of approximately equal size.

Example 2: An investigator wishes to compare two treatments for nausea, one being placebo and the other being a new experimental drug. The absolute risk of nausea on placebo is predicted to be 50% and it is thought that the new treatment would be worth using if it reduced the absolute risk of nausea to 30%, meaning that the treatment effect would have an absolute risk reduction of 20%. The trial will have 90% power and a two-sided significance level of 5%.

How many students would he select in the sample? (Two-sided test)

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 2\bar{p}\bar{q}}{(p_1 - p_2)^2}; \quad \bar{p} = \frac{p_1 + p_2}{2}$$

$$\bar{p} = \frac{p_1 + p_2}{2} = \frac{0.5 + 0.3}{2} = 0.4\,, \quad \bar{q} = 1 - 0.4 = 0.6$$

$$Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96\,; \quad Z_{1-\beta} = Z_{0.90} = 1.282$$

$$n = \frac{(1.96 + 1.282)^2 2 * 0.4 * 0.6}{(0.2)^2} = 126.126 \approx 127\ per\ group$$

$$n_1 = 127, n_2 = 127$$

Example 3: It is believed that the proportion of patients who develop complications after undergoing one type of surgery is 5% while the proportion of patients who develop complications after a second type of surgery is 15%. How large should the sample be in each of the 2 groups of patients if an investigator wishes to detect, with a power of 90%, and 5% level of significance? (Two-sided test)

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 2\bar{p}\bar{q}}{(p_1 - p_2)^2}; \quad \bar{p} = \frac{p_1 + p_2}{2}$$

$$\bar{p} = \frac{p_1 + p_2}{2} = \frac{0.05 + 0.15}{2} = 0.10\,, \quad \bar{q} = 1 - 0.1 = 0.90$$

$$Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96\,; \quad Z_{1-\beta} = Z_{0.90} = 1.282$$

$$n = \frac{(1.282 + 1.96)^2 * 2 * (0.10) * (0.90)}{(0.05 - 0.15)^2} = 189.19 = 190;$$

Example 4: (H.W)

An investigator wishes to compare two treatments for patients who've suffered from a heart attack. It is known that 20% of people in placebo group die within one year, while 5% of people in treatment group die within one year.

How large should the sample be in each of the 2 groups, with a power of 80%, and 5% level of significance? (Two-sided test)

Example 5: (Sample Sizes for One Samples)

In a study for estimating the weight of population and wants the error of estimation to be less than 2 kg of true mean (that is expected difference of weight to be 2 kg), the sample standard deviation was 5 kg. Find the sample size required for this study, with statistical power of 90% at 5% level significance.

the sample size estimated as

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 S^2}{(d)^2}$$

$$Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96\,; \quad Z_{1-\beta} = Z_{0.90} = 1.282$$

$$n = \frac{(1.96 + 1.282)^2 \ 5^2}{2^2} = 65.69 \ \approx 66$$

Example 6: (Sample Sizes for Two Independent Samples, Continuous Outcome)

An investigator is planning a clinical trial to evaluate the efficacy of a new drug designed to reduce blood pressure. The plan is to enroll participants and to randomly assign them to receive either the new drug or a placebo. If the new drug shows a 5 unit reduction in mean blood pressure, this would represent a clinically meaningful reduction.

How many patients should be enrolled in the trial to ensure that the power of the test is 80% to detect this difference? A two-sided test will be used with a 5% level of significance. In addition to, the standard deviation of blood pressure from previous Heart Study was 19 unit.

$$n = \frac{\mathbf{2}\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2 S^2}{(d)^2}$$

$$Z_{1-\frac{\alpha}{2}} = Z_{0.975} = 1.96\,; \quad Z_{1-\beta} = Z_{0.80} = 0.84$$

$$n = \frac{(1.96 + 0.84)^2 * 2 * (19)^2}{(5)^2} = 226.4 \approx 227$$

#Sampling weighting

Determine weighting for the following sampling designs:

1) If N=20000 , n= 400 . $>>> w = \frac{N}{n} = \frac{20000}{400} = 50$

2) If N=1559 , n= 60 . $>>> w = \frac{N}{n} = \frac{1559}{60} = 25.983$

3) If N=335 , n= 230 . $>>> w = \frac{N}{n} = \frac{335}{230} = 1.456$