Editing file

Feedback

How to calculate sample size

10th practical lecture

Color Index:

- Boys' Slides
- Girls' Slides
- Doctors Notes
- Golden Notes
- Important
- Extra

Sample Size Estimation

438's notes

To calculate the sample size you need to identify **two things**:

1. **Type of outcome variable**: quantitative/continuous OR qualitative/categorical
   a. Categorical variables: They are those that have the responses falling into fixed categories. Different types of categorical variables exit, which depends on the potential responses.
      i. Dichotomous variable: It is where the answer could be one of two possibilities, such as "hypertensive" or "non-hypertensive".
      ii. Nominal variable: It is a categorical variable with more than two responses, where there is no specified order in the responses, such as blood type A, B, AB, and O.
      iii. Ordinal variable: It is where a categorical variable could have more than two responses with a predefined order, such as severity of disease (mild, moderate, and severe).
   b. Continuous variables: They are those that consist of any value within the normal defined limits. An example of a continuous variable is age, blood pressure, weight, height, etc.
2. **Type of study**: descriptive OR analytical
   a. Descriptive: no comparisons or associations only description (single mean/proportion)
   b. Analytical: comparisons or associations (two means/proportions)

Outcome Variable

Quantitative/continuous

outcome measure is **mean**

Example: Height, weight, BMI, HB, BP etc.,

a) For a **single mean**:
Sample size: $n = Z_\alpha^2 S^2 / d^2$
Where, S (= sd, get from the literature review or from the pilot study)

b) For **two means**:
Sample size: $n = 2S^2 (Z_\alpha+Z_\beta)^2 / d^2$, per armWhere, S (= sd, get from the literature review or from the pilot study)

The relationship between margin of error and sample size is an inverse relationship because the two move in opposite directions. As the sample size increases, the margin of error decreases and vice versa

Qualitative/categorical

outcome measure is **proportion**

Example: proportion of smokers, diabetes, anemia etc.,

a) For a **single proportion**:
Sample size: $n = Z_\alpha^2 P(1-P) / d^2$
Where, P (= proportion/prevalence, get from the literature review or from the pilot study)

b) For **two proportions**:
Sample size: $n = (Z_\alpha+Z_\beta)^2((p_1q_1)+(p_2q_2)) / (p_1-p_2)^2$, per arm, where $q_1 = (1-p_1)$, $q_2 = (1-p_2)$ Where, P1 and P2 (are proportions for group1 and group2 we are studying, for example, obese-non obese, smokers-non smokers etc., get from the literature review or from the pilot study)

- **d** = precision (the researcher has to decide hence it has no cut off)
- Z_α = 1.96 for 95% confidence level, usually
- Z_β = 1.282 for 90% power, usually

Z = The standard normal distribution

Sample Size Estimation

1

Question

We want to estimate the mean hemoglobin of Saudi females. The standard deviation is around 5 grams/deciliter and we wish to estimate the true mean to within 2 grams/deciliter with 95% confidence. What is the required sample size?

Answer

1. Outcome variable = mean hemoglobin (continuous)
2. Type of study = descriptive

According to the outcome variable and study type we will use single mean formula

Findings: $Z_\alpha = 1.96$ for 95% confidence interval, S = 5, d = 2
$n = Z_\alpha^2 S^2 / d^2$
$n = 1.96^2 \times 5^2 / 2^2 = 24.01 \sim 24$
n = 24 + 20% non-response rate = 24 + 4.8 = 28.8 ~ 29

2

Question

A researcher wanted to estimate average/mean number of cigarettes smoked per week by undergraduate students studying in a certain city. How many students are to be selected in to the sample such that the estimate of mean number of cigarettes smoked is to be within 2 of the true average with 95% confidence? (Based on a pilot study, it was found that the Sd. of number of cigarettes smoked is 30

Answer

1. Outcome variable = mean number of cigarettes (continuous)
2. Type of study = descriptive

According to the outcome variable and study type we will use single mean formula

Findings: $Z_\alpha = 1.96$ for 95% confidence interval, S = 30, d = 2
$n = Z_\alpha^2 S^2 / d^2$
$n = 1.96^2 \times 30^2 / 2^2 = 864.36 \sim 864$
n= 864 + 20% non-response rate = 864 + 172.8 = 1036.8 ~ 1037

The difference between Q1 & Q2 is Sd which refers to the variations in the population as you see the smaller the variations the smaller the sample size and vice versa.

Sample Size Estimation

3

Question

We wish to estimate the proportion of Saudi males who smoke. What sample size do we require to achieve a 95% confidence interval of width ± 5% (that is to be within 5% of the true value)? A study some years ago found approximately 30% were smokers?

Answer

1. Outcome variable = proportion of Saudi males who smoke (categorical)
2. Type of study = descriptive

According to the outcome variable and study type we will use single proportion formula

Findings: $Z_\alpha = 1.96$ for 95% confidence interval, P = 0.3, d = 0.05
$n = Z_\alpha^2 P(1-P) / d^2$
$n = 1.96^2 \times 0.3 \times (1-0.3) / 0.05^2 = 322.6944 \sim 323$
n = 323 + 20% non-response rate = 323 + 64.6 = 387.6 ~ 388

4

Question

An epidemiologist was asked to estimate the Knowledge level (%) towards Covid-19 in a particular community. How many subjects he should select, if the resulting estimate is to fall within 10% (width of confidence interval) of the true proportion with 95% confidence? What will happen to sample size if width of confidence interval is 5%. (As no literature is available researcher assumes that only 30% of subjects had good knowledge level)

Answer

1. Outcome variable = knowledge level (categorical)
2. Type of study = descriptive

According to the outcome variable and study type we will use single proportion formula

Findings: $Z_\alpha = 1.96$ for 95% confidence interval, P = 0.3, d = 0.1
$n = Z_\alpha^2 P(1-P) / d^2$
$n = 1.96^2 \times 0.3 \times (1-0.3) / 0.1^2 = 80.6736 \sim 81$
n = 81 + 20% non-response rate = 81 + 16.2 = 97.2 ~ 97

What will happen to sample size if width of confidence interval is 5%?

Findings: $Z_\alpha = 1.96$ for 95% confidence interval, P = 0.3, d = 0.05
$n = Z_\alpha^2 P(1-P) / d^2$
$n = 1.96^2 \times 0.3 \times (1-0.3) / 0.05^2 = 322.6944 \sim 323$
n = 323 + 20% non-response rate = 323 + 64.6 = 387.6 ~ 388

The smaller the precision (d) the larger the sample size.

Sample Size Estimation

5

Question

An epidemiologist wants to test whether an iron supplement for pregnant women will increase their Hb level. One group of women will receive new supplement and the other group the usual supplement. From a pilot study the sd of Hb is 4 g/dl and is assumed to be same for both groups. what is the sample size required to test the hypothesis of no difference in mean Hb level at 99% level of confidence and 90% power of detecting an increase of 2 g/dl.

Answer

1. Outcome variable = hemoglobin level (continuous)
2. Type of study = analytical

According to the outcome variable and study type we will use two means formula

Findings: $Z_\alpha = 2.58$ for 99% confidence interval, $Z_\beta = 1.282$ for 99% power, S = 4, d = 2
$n = 2S^2 (Z_\alpha+Z_\beta)^2 / d^2$, per arm
$n = 2 \times 4^2 \times (2.58+1.282)^2 / 2^2 = 119.320 \sim 119$
n = 119 + 20% non-response rate = 119 + 23.8 = 142.8 ~ 143, per group
Total sample size = 143 x 2 = 286

6

Question

Suppose it has been estimated that the rate of caries is 800 per 1000 school children in one district and 600 per 1000 in another district. What is the sample size required from each district to determine whether the difference is significant at the 95% level if we wish to have an 90% of chance of detecting the difference if it is real?

Answer

1. Outcome variable = rate of caries (categorical)
2. Type of study = analytical

According to the outcome variable and study type we will use two proportions formula

Findings: $Z_\alpha = 1.96$ for 95% confidence interval, $Z_\beta = 1.282$ for 99% power, $p_1 = 800/1000 = 0.8$, $p_2 = 600/1000 = 0.6$, $q_1 = 1-0.8 = 0.2$, $q_2 = 1-0.6 = 0.4$, difference $= p_1-p_2 = 0.8 - 0.6 = 0.2$
$n = (Z_\alpha+Z_\beta)^2((p_1q_1)+(p_2q_2)) / (p_1-p_2)^2$, per arm, where $q_1 = (1-p_1)$, $q_2 = (1-q_2)$
$n = (1.96+1.282)^2 \times ((0.8\times0.2) + (0.6\times0.4)) / 0.2^2 = 105.106 \sim 105$
n = 105 + 20% non-response rate = 105 + 21 = 126, per group
Total sample size = 126 x 2 = 252

Sample Size Estimation

TABLE 1A	SAMPLE SIZES FOR A SINGLE MEAN FOR VARIOUS d and sd for 95% level, Za=1.96																			
	d																			
sd	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	16	4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	35	9	4	3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	62	16	7	4	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
5	97	24	11	7	4	3	2	2	2	1	1	1	1	1	1	1	1	1	1	1
6	139	35	16	9	6	4	3	3	2	2	2	1	1	1	1	1	1	1	1	1
7	189	48	21	12	8	6	4	3	3	2	2	2	2	1	1	1	1	1	1	1
8	246	62	28	16	10	7	6	4	4	3	3	2	2	2	2	1	1	1	1	1
9	312	78	35	20	13	9	7	5	4	4	3	3	2	2	2	2	2	1	1	1
10	385	97	43	25	16	11	8	7	5	4	4	3	3	2	2	2	2	2	2	1
11	465	117	52	30	19	13	10	8	6	5	4	4	3	3	3	2	2	2	2	2
12	554	139	62	35	23	16	12	9	7	6	5	4	4	3	3	3	2	2	2	2
13	650	163	73	41	26	19	14	11	9	7	6	5	4	4	3	3	3	3	2	2
14	753	189	84	48	31	21	16	12	10	8	7	6	5	4	4	3	3	3	3	2
15	865	217	97	55	35	25	18	14	11	9	8	7	6	5	4	4	3	3	3	3
16	984	246	110	62	40	28	21	16	13	10	9	7	6	6	5	4	4	4	3	3
17	1111	278	124	70	45	31	23	18	14	12	10	8	7	6	5	5	4	4	4	3
18	1245	312	139	78	50	35	26	20	16	13	11	9	8	7	6	5	5	4	4	4
19	1387	347	155	87	56	39	29	22	18	14	12	10	9	8	7	6	5	5	4	4
20	1537	385	171	97	62	43	32	25	19	16	13	11	10	8	7	7	6	5	5	4
21	1695	424	189	106	68	48	35	27	21	17	15	12	11	9	8	7	6	6	5	5
22	1860	465	207	117	75	52	38	30	23	19	16	13	12	10	9	8	7	6	6	5
23	2033	509	226	128	82	57	42	32	26	21	17	15	13	11	10	8	8	7	6	6
24	2213	554	246	139	89	62	46	35	28	23	19	16	14	12	10	9	8	7	7	6
25	2401	601	267	151	97	67	49	38	30	25	20	17	15	13	11	10	9	8	7	7
26	2597	650	289	163	104	73	53	41	33	26	22	19	16	14	12	11	9	9	8	7
27	2801	701	312	176	113	78	58	44	35	29	24	20	17	15	13	11	10	9	8	8
28	3012	753	335	189	121	84	62	48	38	31	25	21	18	16	14	12	11	10	9	8
29	3231	808	359	202	130	90	66	51	40	33	27	23	20	17	15	13	12	10	9	9
30	3458	864	385	217	139	97	71	55	43	35	29	25	21	18	16	14	12	11	10	9
31	3692	923	411	231	148	103	76	58	46	37	31	26	22	19	17	15	13	12	11	10
32	3934	984	438	246	158	110	81	62	49	40	33	28	24	21	18	16	14	13	11	10
33	4184	1046	465	262	168	117	86	66	52	42	35	30	25	22	19	17	15	13	12	11
34	4441	1111	494	278	178	124	91	70	55	45	37	31	27	23	20	18	16	14	13	12
35	4706	1177	523	295	189	131	97	74	59	48	39	33	28	25	21	19	17	15	14	12
36	4979	1245	554	312	200	139	102	78	62	50	42	35	30	26	23	20	18	16	14	13
37	5260	1315	585	329	211	147	108	83	65	53	44	37	32	27	24	21	19	17	15	14
38	5548	1387	617	347	222	155	114	87	69	56	46	39	33	29	25	22	20	18	16	14
39	5844	1461	650	366	234	163	120	92	73	59	49	41	35	30	26	23	21	19	17	15
40	6147	1537	683	385	246	171	126	97	76	62	51	43	37	32	28	25	22	19	18	16

This is another way to calculate the sample size for a single mean. All you have to do is choose a row (d) and a column (sd) for a given a confidence interval of 95% and a Za of 1.96.
E.g., Question 1

Sample Size Estimation

TABLE 1B	SAMPLE SIZES FOR A SINGLE PROPORTION FOR VARIOUS P and d for 95% level, Za=1.96																			
	d																			
P	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
0.01	16	4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0.02	31	8	4	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0.03	45	12	5	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
0.04	60	15	7	4	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
0.05	73	19	9	5	3	3	2	2	1	1	1	1	1	1	1	1	1	1	1	1
0.06	87	22	10	6	4	3	2	2	2	1	1	1	1	1	1	1	1	1	1	1
0.07	101	26	12	7	5	3	3	2	2	2	1	1	1	1	1	1	1	1	1	1
0.08	114	29	13	8	5	4	3	2	2	2	1	1	1	1	1	1	1	1	1	1
0.09	126	32	14	8	6	4	3	2	2	2	2	1	1	1	1	1	1	1	1	1
0.1	139	35	16	9	6	4	3	3	2	2	2	1	1	1	1	1	1	1	1	1
0.11	151	38	17	10	7	5	4	3	2	2	2	2	1	1	1	1	1	1	1	1
0.12	163	41	19	11	7	5	4	3	3	2	2	2	1	1	1	1	1	1	1	1
0.13	174	44	20	11	7	5	4	3	3	2	2	2	2	1	1	1	1	1	1	1
0.14	186	47	21	12	8	6	4	3	3	2	2	2	2	1	1	1	1	1	1	1
0.15	196	49	22	13	8	6	4	4	3	2	2	2	2	1	1	1	1	1	1	1
0.16	207	52	23	13	9	6	5	4	3	3	2	2	2	2	1	1	1	1	1	1
0.17	217	55	25	14	9	7	5	4	3	3	2	2	2	2	1	1	1	1	1	1
0.18	227	57	26	15	10	7	5	4	3	3	2	2	2	2	2	1	1	1	1	1
0.19	237	60	27	15	10	7	5	4	3	3	2	2	2	2	2	1	1	1	1	1
0.2	246	62	28	16	10	7	6	4	4	3	3	2	2	2	2	1	1	1	1	1
0.21	255	64	29	16	11	8	6	4	4	3	3	2	2	2	2	1	1	1	1	1
0.22	264	66	30	17	11	8	6	5	4	3	3	2	2	2	2	2	1	1	1	1
0.23	273	69	31	18	11	8	6	5	4	3	3	2	2	2	2	2	1	1	1	1
0.24	281	71	32	18	12	8	6	5	4	3	3	2	2	2	2	2	1	1	1	1
0.25	289	73	33	19	12	9	6	5	4	3	3	3	2	2	2	2	1	1	1	1
0.26	296	74	33	19	12	9	7	5	4	3	3	3	2	2	2	2	2	1	1	1
0.27	303	76	34	19	13	9	7	5	4	4	3	3	2	2	2	2	2	1	1	1
0.28	310	78	35	20	13	9	7	5	4	4	3	3	2	2	2	2	2	1	1	1
0.29	317	80	36	20	13	9	7	5	4	4	3	3	2	2	2	2	2	1	1	1
0.3	323	81	36	21	13	9	7	6	4	4	3	3	2	2	2	2	2	1	1	1
0.31	329	83	37	21	14	10	7	6	5	4	3	3	2	2	2	2	2	2	1	1
0.32	335	84	38	21	14	10	7	6	5	4	3	3	2	2	2	2	2	2	1	1
0.33	340	85	38	22	14	10	7	6	5	4	3	3	3	2	2	2	2	2	1	1
0.34	345	87	39	22	14	10	8	6	5	4	3	3	3	2	2	2	2	2	1	1
0.35	350	88	39	22	14	10	8	6	5	4	3	3	3	2	2	2	2	2	1	1
0.36	355	89	40	23	15	10	8	6	5	4	3	3	3	2	2	2	2	2	1	1
0.37	359	90	40	23	15	10	8	6	5	4	3	3	3	2	2	2	2	2	1	1
0.38	363	91	41	23	15	11	8	6	5	4	3	3	3	2	2	2	2	2	2	1
0.39	366	92	41	23	15	11	8	6	5	4	4	3	3	2	2	2	2	2	2	1
0.4	369	93	41	24	15	11	8	6	5	4	4	3	3	2	2	2	2	2	2	1

You can also calculate the sample size for a single proportion by using this table. Depending on the variables you are given, choose a row (d) and a column (P) for a given a confidence interval of 95% and a Za of 1.96.
e.g.,Question 3

Sample Size Estimation

	Table 2A Za for 99% level=2.58 Zb for 90% power = 1.28				SAMPLE SZES for two means for various values of d and sd d															
Sd	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	30	7	3	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	119	30	13	7	5	3	2	2	1	1	1	1	1	1	1	0	0	0	0	0
3	268	67	30	17	11	7	5	4	3	3	2	2	2	1	1	1	1	1	1	1
4	477	119	53	30	19	13	10	7	6	5	4	3	3	2	2	2	2	1	1	1
5	745	186	83	47	30	21	15	12	9	7	6	5	4	4	3	3	3	2	2	2
6	1073	268	119	67	43	30	22	17	13	11	9	7	6	5	5	4	4	3	3	3
7	1460	365	162	91	58	41	30	23	18	15	12	10	9	7	6	6	5	5	4	4
8	1907	477	212	119	76	53	39	30	24	19	16	13	11	10	8	7	7	6	5	5
9	2414	603	268	151	97	67	49	38	30	24	20	17	14	12	11	9	8	7	7	6
10	2980	745	331	186	119	83	61	47	37	30	25	21	18	15	13	12	10	9	8	7
11	3606	901	401	225	144	100	74	56	45	36	30	25	21	18	16	14	12	11	10	9
12	4291	1073	477	268	172	119	88	67	53	43	35	30	25	22	19	17	15	13	12	11
13	5036	1259	560	315	201	140	103	79	62	50	42	35	30	26	22	20	17	16	14	13
14	5841	1460	649	365	234	162	119	91	72	58	48	41	35	30	26	23	20	18	16	15
15	6705	1676	745	419	268	186	137	105	83	67	55	47	40	34	30	26	23	21	19	17
16	7629	1907	848	477	305	212	156	119	94	76	63	53	45	39	34	30	26	24	21	19
17	8612	2153	957	538	344	239	176	135	106	86	71	60	51	44	38	34	30	27	24	22
18	9655	2414	1073	603	386	268	197	151	119	97	80	67	57	49	43	38	33	30	27	24
19	10758	2689	1195	672	430	299	220	168	133	108	89	75	64	55	48	42	37	33	30	27
20	11920	2980	1324	745	477	331	243	186	147	119	99	83	71	61	53	47	41	37	33	30
21	13141	3285	1460	821	526	365	268	205	162	131	109	91	78	67	58	51	45	41	36	33
22	14423	3606	1603	901	577	401	294	225	178	144	119	100	85	74	64	56	50	45	40	36
23	15764	3941	1752	985	631	438	322	246	195	158	130	109	93	80	70	62	55	49	44	39
24	17164	4291	1907	1073	687	477	350	268	212	172	142	119	102	88	76	67	59	53	48	43
25	18625	4656	2069	1164	745	517	380	291	230	186	154	129	110	95	83	73	64	57	52	47

Table 2B Sample sizes for 95% Confidence level, (Zα=1.96) and for 90% Power(Zβ= 1.282) Here P1 is Larger proportion and P2 is Smaller proportion																			
	P1																		
P2	***10%***	***15%***	***20%***	***25%***	***30%***	***35%***	***40%***	***45%***	***50%***	***55%***	***60%***	***65%***	***70%***	***75%***	***80%***	***85%***	***90%***	***95%***	***100%***
10%		**914**	**263**	**130**	**79**	**53**	**39**	**29**	**22**	**18**	**14**	**11**	**9**	**7**	**5**	**4**	**3**	**2**	**1**
15%			**1209**	**331**	**158**	**93**	**62**	**44**	**32**	**25**	**19**	**15**	**12**	**9**	**7**	**5**	**4**	**3**	**2**
20%				**1461**	**389**	**181**	**105**	**69**	**48**	**35**	**26**	**20**	**16**	**12**	**9**	**7**	**5**	**4**	**3**
25%					**1671**	**436**	**200**	**114**	**74**	**51**	**37**	**27**	**21**	**16**	**12**	**9**	**7**	**5**	**4**
30%						**1839**	**473**	**214**	**121**	**77**	**53**	**38**	**28**	**21**	**16**	**12**	**9**	**6**	**5**
35%							**1965**	**499**	**223**	**125**	**79**	**53**	**38**	**27**	**20**	**15**	**11**	**8**	**6**
40%								**2050**	**515**	**228**	**126**	**79**	**53**	**37**	**26**	**19**	**14**	**10**	**7**
45%									**2092**	**520**	**228**	**125**	**77**	**51**	**35**	**25**	**18**	**12**	**9**
50%										**2092**	**515**	**223**	**121**	**74**	**48**	**32**	**22**	**15**	**11**
55%											**2050**	**499**	**214**	**114**	**69**	**44**	**29**	**19**	**13**
60%												**1965**	**473**	**200**	**105**	**62**	**39**	**25**	**16**
65%													**1839**	**436**	**181**	**93**	**53**	**32**	**20**
70%														**1671**	**389**	**158**	**79**	**43**	**25**
75%															**1461**	**331**	**130**	**62**	**32**
80%																**1209**	**263**	**97**	**42**
85%																	**914**	**184**	**60**
90%																		**578**	**95**
95%																			**200**
100%																			

Research Team

RESEARCH TEAM
438
Keep it simple, Keep it focused

Thank you for checking our work!

Leaders:

Shuaa Khdary Sarah AlQuwayz

Abdulrhman Alsuhaibany

Member:

Muneerah Alsadhan

Note Taker:

Rakan Aldohan

Contact us:

Research4390@gmail.com