BIST P8130: Biostatistics Methods I

Recitation 03 – Sample Size Determination and T-test in SAS

Zilan Chai

Fall 2017

This recitation's big ideas:

- Compute sample size for a given power and effect size
- Use PROC TTEST in SAS 9.4 to perform:
  1. One-sample t-test
  2. Two-sample t-test (Paired and independent samples)
- Use PROC POWER in SAS to perform power and sample size calculation

Ophthalmology Example: A new drug is proposed to prevent the development of glaucoma in people with high intraocular pressure (IOP). A pilot study is conducted with the drug among 10 patients. After 1 month of using the drug, their mean IOP decreases by 5 mm Hg with a standard deviation of 10 mm Hg.

What sample size do we need in order to achieve 80% power for detecting a minimum clinically meaningful difference of 30mmHg with a standard deviation of 10mmHg?

In general, if $X_1, X_2, ...X_n$ are iid sample from a normal distribution with unknown mean μ and known variance σ^2.

For hypothesis $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, the z-test statistic $z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ under H_0. For Type I error α, the H_0 is rejected if $|z| > z_{1-\frac{\alpha}{2}}$ where $z_{1-\frac{\alpha}{2}}$ satisfies $2 \times (1 - \Phi(z_{1-\frac{\alpha}{2}})) = 2 \times [1 - P(N(0,1) \leq z_{1-\frac{\alpha}{2}})] = \alpha$

$$\text{Power} = P(\text{reject } H_0 | H_1 \text{ is true}) = P(|\text{test statistic}| > \text{critical value } |H_1 \text{ is true})$$

From the calculation above, and recall that our goal is to find the value of n such that the power is $1-\beta$, we have,

$$\text{Power} = \Phi\left[-z_{1-\frac{\alpha}{2}} + \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}\right] = 1-\beta$$

$$z_{1-\beta} = -z_{1-\frac{\alpha}{2}} + \frac{|\mu_0 - \mu_1|}{\sigma/\sqrt{n}}$$

$$\sqrt{n} = \frac{(z_{1-\beta} + z_{1-\frac{\alpha}{2}})\sigma}{|\mu_0 - \mu_1|}$$

$$n = \frac{(z_{1-\beta} + z_{1-\frac{\alpha}{2}})^2\sigma^2}{(\mu_0 - \mu_1)^2}$$

Note, always round up the number.

Factors Affecting the Sample Size

(1) The sample size increases as σ^2 increases.

(2) The sample size increases as the significance lever is made smaller (α decreases).

(3) The sample size increases as the required power increases.

(4) The sample size increases as the effect size increases.

Sample-Size Estimation When Testing for the Mean of a Normal Distribution (Two-Sided)

$$n = \frac{(z_{1-\beta} + z_{1-\frac{\alpha}{2}})^2\sigma^2}{(\mu_0 - \mu_1)^2}$$

Sample Size Needed for Comparing the Means of Two Normally Distributed Samples of Equal Size Using a Two-Sided Test

$$n = \frac{(z_{1-\beta} + z_{1-\frac{\alpha}{2}})^2(\sigma_1^2 + \sigma_2^2)}{(\mu_1 - \mu_2)^2} = \text{sample size for each group}$$

Sample Size Needed for Comparing the Means of Two Normally Distributed Samples of Unequal Size Using a Two-Sided Test

$$n_1 = \frac{(z_{1-\beta} + z_{1-\frac{\alpha}{2}})^2(\sigma_1^2 + \sigma_2^2/k)}{(\mu_1 - \mu_2)^2} = \text{sample size of first group}$$

$$n_2 = \frac{(z_{1-\beta} + z_{1-\frac{\alpha}{2}})^2(k\sigma_1^2 + \sigma_2^2)}{(\mu_1 - \mu_2)^2} = \text{sample size of first group}$$

where $k = n_2/n_1 =$ the projected ratio of the two sample sizes.

Use PROC TTEST in SAS 9.4 to perform a t-test:

Example: To assess the impact of oral contraceptive use on bone mineral density (BMD), researchers in Canada carried out a study comparing BMD for women who had used oral contraceptives for at least 3 months to BMD for women who had never used oral contraceptives. Data on BMD (in grams per centimeter) as well as contraceptive use status are recorded in the file "Bonemineral". Let's assume BMD is normally distributed.

Obs	Status	BMD
1	never	0.82
2	never	0.89
	…	…
19	used	1.29
20	used	1.31

One Sample Hypothesis Testing and Confidence Intervals:

First, researchers wanted to test whether BMD (among all participants) **was different from** 1.06, the mean BMD in healthy Americans.

TTEST procedure can be used to compare the BMD with a specific value:

```
proc ttest data=bonemineral h0=1.06; * alpha=0.01, and sides=L,U;
var BMD;
run;
```

ALPHA=*p* specifies that confidence intervals to be 100(1-p)% confidence intervals, where 0<p<1.
H0=*m* requests tests against a null value of *m*,
SIDES=L specifies lower one-sided tests, in which the alternative hypothesis indicates a mean less than the null value
SIDES=U specifies upper one-sided tests, in which the alternative hypothesis indicates a mean greater than the null value

N	Mean	Std Dev	Std Err	Minimum	Maximum
20	1.0770	0.1370	0.0306	0.8200	1.3100

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1.0770	1.0129	1.1411	0.1370	0.1042	0.2002

DF	t Value	Pr > |t|
19	0.55	0.5855

Exercise: researchers wanted to test whether BMD (among all participants) is lower than 1.06 at 0.10 level of significance.

Two Sample Hypothesis Testing and Confidence Intervals (independent samples):

Next, researchers wanted to test whether the mean BMD for non-OC users is different than the mean BMD for OC users. This situation calls for a two sample t-test because we have independent samples.

```
proc ttest data=bonemineral alpha=0.05;
  class Status;
  var BMD;
run;
```

`CLASS` statement: giving the name of the classification (or grouping) variable. Equality of variance test is part of the output when you see a CLASS statement.

Status	N	Mean	Std Dev	Std Err	Minimum	Maximum
never	10	0.9700	0.0726	0.0229	0.8200	1.0500
used	10	1.1840	0.0945	0.0299	1.0600	1.3100
Diff (1-2)		-0.2140	0.0843	0.0377		

Status	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
never		0.9700	0.9181	1.0219	0.0726	0.0499	0.1325
used		1.1840	1.1164	1.2516	0.0945	0.0650	0.1726
Diff (1-2)	Pooled	-0.2140	-0.2932	-0.1348	0.0843	0.0637	0.1246
Diff (1-2)	Satterthwaite	-0.2140	-0.2936	-0.1344			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	18	-5.68	<.0001
Satterthwaite	Unequal	16.873	-5.68	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	1.70	0.4429

Two Sample Hypothesis Testing and Confidence Intervals (paired samples):

Suppose that researchers are investigating the effect of a new diet on weight. The dataset looks like the following:

Obs	wbefore	wafter
1	200	185
2	175	154
3	188	176
4	198	193
5	197	198
6	310	275
7	245	224
8	202	188

```
proc ttest data=weight;
 paired wbefore*wafter;
run;
```

N	Mean	Std Dev	Std Err	Minimum	Maximum
8	15.2500	10.9381	3.8672	-1.0000	35.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
15.2500	6.1055	24.3945	10.9381	7.2320	22.2621

DF	t Value	Pr > |t|
7	3.94	0.0056

Power in SAS:

SAS can handle power and sample size calculations for different tests; to find power or sample size, use one of the options with a PROC POWER statement.

```
PROC POWER < options > ;
  MULTREG < options > ;
  ONECORR < options > ;
  ONESAMPLEFREQ < options > ;
  ONESAMPLEMEANS < options > ;
  ONEWAYANOVA < options > ;
  PAIREDFREQ < options > ;
  PAIREDMEANS < options > ;
  TWOSAMPLEFREQ < options > ;
  TWOSAMPLEMEANS < options > ;
  TWOSAMPLESURVIVAL < options > ;
```

(1) One-Sample t Test, Confidence Interval Precision, or Equivalence Test
    Statement: ONESAMPLEMEANS
    Test=T

(2) Two Sample t Test, Confidence Interval Precision, or Equivalence Test
    Statement: TWOSAMPLEMEANS
    Test=DIFF/EQUIV

(3) Paired t Test, Confidence Interval Precision, or Equivalence Test
    Statement: PAIREDMEANS
    Test=DIFF/EQUIV_DIFF

(4) One-Way ANOVA Including Single-Degree-of-Freedom Contrasts
    Statement: ONEWAYANOVA
    Test=OVERALL/CONTRAST

(5) Chi-Square, Likelihood Ratio, and Fisher's Exact Tests for Two Independent Proportions
    Statement: TWOSAMPLEFREQ
    Test=PCHI/LRCHI/FISHER

Example: A sample size of 36 patients per group was determined to detect a difference in means for the change in modified Rodnan skin score of 4.7 from baseline to 24 weeks, based on an estimated common SD of 6.99 using a two-group t test with a 5% two-sided significance level. What is the power of this test?

```
proc power;
twosamplemeans test=diff
meandiff=4.7     /* Mean diff. */
stddev=6.99       /* Std.dev. */
npergroup=.
power=0.8
;
run;
```

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean Difference	4.7
Standard Deviation	6.99
Sample Size Per Group	36
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed Power
Power
0.803

For this example, what should the sample size be for 90% power?

```
proc power;
twosamplemeans test=diff
meandiff=4.7
stddev=6.99
npergroup=.
power=0.9
;
run;
```

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean Difference	4.7
Standard Deviation	6.99
Nominal Power	0.9
Number of Sides	2
Null Difference	0
Alpha	0.05

Computed N Per Group	
Actual Power	N Per Group
0.903	48