AMS 572 Data Analysis I
Part I: Power and sample size calculation for one population mean μ
Part II: Inference on one population variance σ^2

Pei-Fen Kuan

Applied Math and Stats, Stony Brook University

Power calculation for one population mean μ

Type I error, Type II error and Power

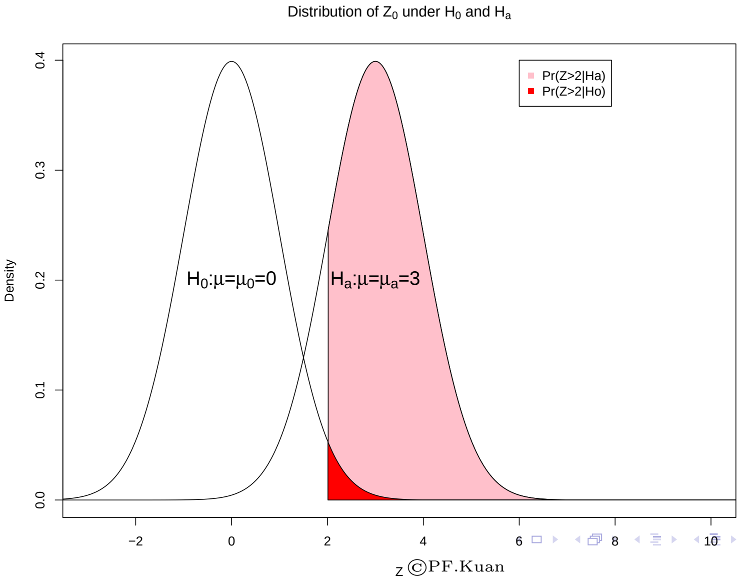
		Truth	
		H_0	H_a
Decision	Do not reject		Type II error
	Reject	Type I error	Power

Case 1: Power calculation for normal population, σ^2 is known

- Suppose we are testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_a > \mu_0$
- Test statistic : $Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$ under H_0.
- At the significance level α, we reject H_0 if $Z_0 \geq z_\alpha$
- Power =
- What is the distribution of Z_0 when $\mu = \mu_a$?

Case 1: Power calculation for normal population, σ^2 is known

Case 1: Power calculation for normal population, σ^2 is known



Case 1: Power calculation for normal population, σ^2 is known

- For hypothesis test $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_a < \mu_0$
- Test statistic : $Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$ under H_0.
- At the significance level α, we reject H_0 if $Z_0 \leq -z_\alpha$
- Power =

Case 1: Power calculation for normal population, σ^2 is known

- For hypothesis test $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_a \neq \mu_0$
- Test statistic : $Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1)$ under H_0.
- At the significance level α, we reject H_0 if $|Z_0| \geq z_{\alpha/2}$
- Power =

Case 2: Power calculation for normal population, σ^2 is unknown, small sample

- Suppose we are testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_a > \mu_0$
- Test statistic : $T_0 = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$ under H_0.
- At the significance level α, we reject H_0 if $T_0 \geq t_{n-1,\alpha}$
- Power $= 1 - \beta = \mathrm{P}(\text{reject } H_0 | H_a) = \mathrm{P}(T_0 \geq t_{n-1,\alpha} | \mu = \mu_a)$
- What is the distribution of T_0 when $\mu = \mu_a$?

Case 2: Power calculation for normal population, σ^2 is unknown, small sample

Note that if $U \sim N(\mu, 1)$ and $V \sim \chi^2_k$ and U independent of V, then $\frac{U}{\sqrt{V/k}}$ is a non-central t distribution with k degrees of freedom and non-centrality parameter μ

Power $=$ 1-$\beta = \mathrm{P}(\text{reject } H_0|H_a) = \mathrm{P}(T_0 \geq t_{n-1,\alpha}|\mu = \mu_a)$

- Reasonably good approximation to the power of the t-test can be obtained from the z-test. If z-test is used, there is a slight over estimation of the power.

Case 2: Power calculation for normal population, σ^2 is unknown, small sample

- Note: Shapiro-Wilk test can be used to determine whether the population is normal
- Power derivation for testing
  - $H_0 : \mu = \mu_0$ vs $H_a : \mu = \mu_a < \mu_0$
  - $H_0 : \mu = \mu_0$ vs $H_a : \mu = \mu_a \neq \mu_0$

  is similar to Case 1 and thus omitted.

Case 3: Power calculation for any population, large sample

- Suppose we are testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_a > \mu_0$
- Test statistic : $Z_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \mathrel{\dot\sim} N(0, 1)$ under H_0.
- At the significance level α, we reject H_0 if $Z_0 \geq z_\alpha$
- Power of the test = P(reject $H_0 | H_a$) $= 1 - \beta$

Case 3: Power calculation for any population, large sample (similar to case 1)

$\boxed{\mu_a - \mu_0}$ → effect size

$$\text{Power} = 1-\beta = P(Z_0 > z_\alpha \mid H_a: \mu = \mu_a)$$

$$= P\Big(\underbrace{\frac{\bar{x}-\mu_a}{s/\sqrt{n}}}_{N(0,1)} + \frac{\mu_a-\mu_0}{s/\sqrt{n}} > z_\alpha \,\Big|\, H_a\Big)$$

$$= P\Big(Z > z_\alpha - \frac{\mu_a-\mu_0}{s/\sqrt{n}}\Big) \quad \text{where } Z \overset{\text{approx}}{\sim} N(0,1)$$

Case 3: Power calculation for any population, large sample

- Power derivation for testing
  - $H_0 : \mu = \mu_0$ vs $H_a : \mu = \mu_a < \mu_0$
  - $H_0 : \mu = \mu_0$ vs $H_a : \mu = \mu_a \neq \mu_0$

  is similar to Case 1 and thus omitted.

Sample size determination for one population mean μ

Introduction

- Choosing an appropriate sample size is not just a study design issue, it is an ethical issue
- For a (new) study to be ethical, it must be designed to have sufficient power to detect meaningful differences
- There are ethical issues even if using already-collected data – wasting resources if the sample size is too small
- Power and sample size are mathematically related
- In some situations we can calculate sample sizes explicitly
- In complicated situations, one may need to use simulation to determine the sample size

Introduction

To estimate sample size we need to specify:

- The study design
- The significance level 5%
- The test statistic and its distribution
- The null and alternative hypothesis
- The value μ_a that we want to be able to detect
- The desired power to detect this μ_a 80%
- More complex models may require specifying other parameters, such as covariances (for measures taken at multiple time-points or if adjusting for confounders)

These need to be specified when designing the study

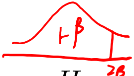
Introduction

How do we decide which values to use?

- Some are reasonably standard, e.g. $\alpha = 0.05$
- Obtain estimates from pilot studies or studies done elsewhere, e.g. μ_a, variances, covariances
- For example, in clinical studies, μ_1 may be what is regarded as the smallest clinically meaningful effect
- We often calculate sample size for a few representative values of what the underlying parameters might be
- We often calculate the sample size for two or more choices of the study power (typically 0.8 and 0.9)
- We often choose a few sample sizes and calculate the associated power

Case 1a: Sample size determination for normal population, σ^2 is known, for a given power $(1-\beta)$

$H_a = H_1$ $\quad \Phi(Z_\beta) = 1-\beta$



- Suppose we are testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_a > \mu_0$

$$1-\beta = P(Z_0 > z_\alpha \mid H_a : \mu = \mu_0)$$

$$Z = Z_0 - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}}$$

$$= P\left(Z > z_\alpha - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} \,\Big|\, H_a\right) \qquad Z \sim N(0,1)$$

$$= P\left(Z \le -z_\alpha + \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} \,\Big|\, H_a\right)$$

$$Z_\beta = -z_\alpha + \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} \Rightarrow n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2}$$

Case 1a: Sample size determination for normal population, σ^2 is known, for a given power

- Suppose we are testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_a < \mu_0$

Case 1a: Sample size determination for normal population, σ^2 is known, for a given power $(1-\beta)$

$1-\beta = P(|Z_0| > z_{\alpha/2} \mid H_a: \mu = \mu_a)$.

- Suppose we are testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu = \mu_a \neq \mu_0$
-
$$1 - \beta = P(Z \geq z_{\alpha/2} - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} | \mu = \mu_a)$$
$$+P(Z \leq -z_{\alpha/2} - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} | \mu = \mu_a)$$

  Assume $\mu_a > \mu_0$. Then,
  $P(Z \leq -z_{\alpha/2} - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} | \mu = \mu_a) \to 0$. So, we can neglect it.

Case 1a: Sample size determination for normal population, σ^2 is known, for a given power

$$1-\beta \overset{\approx}{\geq} P\left(Z > z_{\frac{\alpha}{2}} - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}} \,\middle|\, \mu = \mu_a\right)$$

$$-z_\beta \overset{\approx}{\leq} z_{\frac{\alpha}{2}} - \frac{\mu_a - \mu_0}{\sigma/\sqrt{n}}$$

$$n \overset{\approx}{\leq} \frac{(z_{\frac{\alpha}{2}} + z_\beta)^2 \sigma^2}{(\mu_a - \mu_0)^2}$$

Sample size calculation for the two sided test differs from that for the one-sided tests in two aspects:

1. α is replaced by $\alpha/2$
2. It is an approximate formula (a more conservative estimate)

Case 1b: Sample size determination for normal population, σ^2 is known, for a given CI length

P.Q. $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$100(1-\alpha)\%$ CI for μ : $\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$$L = (\bar{X} + Z_{\frac{\alpha}{2}}\sigma/\sqrt{n}) - (\bar{X} - Z_{\frac{\alpha}{2}}\sigma/\sqrt{n})$$

$$= 2Z_{\frac{\alpha}{2}}\sigma/\sqrt{n}$$

$$n = \left(\frac{2Z_{\alpha/2}\sigma}{L}\right)^2$$

Case 1b: Sample size determination for normal population, σ^2 is known, for a given margin of error E

$$P(|\overline{X} - \mu| \leq E) = 1 - \alpha$$

$$\Leftrightarrow P(-E \leq \bar{X} - \mu \leq E) = 1-\alpha$$

$$\Leftrightarrow P\left(\frac{-E}{\sigma/\sqrt{n}} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{E}{\sigma/\sqrt{n}}\right) = 1-\alpha$$

$$\frac{E}{\sigma/\sqrt{n}} = z_{\frac{\alpha}{2}}$$

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{E}\right)^2$$

Compare the above formula to:

$$n = (\frac{2z_{\alpha/2}\sigma}{L})^2$$

For a given α, $L = 2E$

Case 2: Sample size determination for normal population, σ^2 is unknown

- Exact sample size calculation is based on non central t distribution.
- Reasonably good approximation to the required sample size can be obtained from the z-test. If z-test is used, there is a slight under estimation of the sample size.

Using SAS

For the inference on one population mean, three procedures are most relevant:

- `Proc means`
- `Proc univariate`
- `Proc ttest`

Data entry in SAS

character variable.

```
data one;
input ID $ weight ;
X = weight - 100 ;
datalines;
P1 100
P2 93
P3 88
p4 106
p5 90
p6 95
p7 97
p8 102
;
run ;
```

ID

weight

Data entry in SAS

```
/*-----Alternative way-----*/
data two;
input ID $ weight @@ ;
X = weight - 100 ;
datalines;
P1 100 P2 93
P3 88 p4 106 p5 90
p6 95
p7 97
p8 102
;
run ;
```

- One may also use `infile` to read data stored in other files already, e.g. excel files.

Inference on one population mean in SAS

single variable.

```
proc univariate data = one normal;
var X;
run;
```

- Generate different tests for normality
- Output t-test, non-parametric tests (sign and signed-rank tests)

Inference on one population mean in SAS

```
/*----One may also use proc means----*/
proc means data=one t prt ;
var X ;
run ;

/*----Yet another way using proc ttest----*/
proc ttest data = one;
var X;
run;
```

prob.

- `prt` : p-value of $\begin{cases} H_0 : \mu = 0 \\ H_a : \mu \neq 0 \end{cases}$
- `proc ttest` : 1 population t-test / 2 populations t-test (paired and independent)

$H_0: \mu_0 = 525$ vs $H_a: \mu_0 > 525$

Example 1: Jerry is planning to purchase a sports goods store. He calculated that in order to make profit, the average daily sales must be > \$525. He randomly sampled 36 days and found $\overline{X} = \$565$ and $S = \$150$

$n > 30$

(a) In order to estimate the average daily sales to within \$20 with 95% reliability, how many days should Jerry sample?

(b) If the true average daily sales is \$530, what is the power of Jerry's test at the significance level of 0.05?

(c) Suppose $\mu = \$530$. In order to guarantee $\alpha = 0.05$ and $\beta = 0.2$, how many days should Jerry sample?

Example 2: John Pauzke, president of Cereal's Unlimited Inc, wants to be very certain that the mean weight μ of packages satisfies the package label weight of 16 ounces. The packages are filled by a machine that is set to fill each package to a specified weight. However, the machine has random variability measured by σ^2. John would like to have strong evidence that the mean package weight is greater than 16 oz. George Williams, quality control manager, advises him to examine a random sample of 25 packages of cereal. From his past experience, George knew that the weight of the packages follows a normal distribution with standard deviation 0.4 oz. At the significance level $\alpha = 0.05$,

Z-type

(a) What is the decision rule (rejection region) in terms of the sample mean $\overline{X}$?

(b) What is the power of the test when $\mu = 16.13$ oz?

(c) How many packages of cereal should be sampled if we wish to achieve a power of 80% when $\mu = 16.13$ oz?

Example 3: The seven scores listed below are axial loads (in pounds) for a random sample of 7 12-oz aluminum cans manufactured by ALUMCO. An axial load of a can is the maximum weight supported by its sides, and it must be greater than 165 pounds, because that is the maximum pressure applied when the top lid is pressed into place.
270, 273, 258, 204, 254, 228, 282

(a) As the quality control manager, please test the claim of the engineering supervisor that the average axial load is greater than 165 pounds. Use $\alpha = 0.05$. What assumptions are needed for your test?

(b) Write a SAS and R program to do part (a).

SAS Code

```
data cans ;
input pressure @@ ;
newvar = pressure-165;
datalines;
270 273 258 204 254 228 282
;
run;

proc univariate data=cans normal;
var newvar;
run;

/*----Alternatively, we can use
 the proc ttest procedure as follows:*/

proc ttest data=cans h0=165 sides=u alpha = 0.05;
var pressure;
run;
```

x

var newvar;

u/L/2

SAS Output

```
                    The SAS System    22:58 Saturday, September 12, 2015   1

                  The UNIVARIATE Procedure
                      Variable:  newvar

                           Moments

N                          7     Sum Weights                 7
Mean              87.7142857     Sum Observations          614
Std Deviation     27.6327962     Variance           763.571429
Skewness          -1.0164993     Kurtosis           0.20245235
Uncorrected SS         58438     Corrected SS       4581.42857
Coeff Variation   31.5031879     Std Error Mean     10.4442153


                 Basic Statistical Measures

        Location                    Variability

    Mean     87.71429     Std Deviation            27.63280
    Median   93.00000     Variance                763.57143
    Mode       .          Range                    78.00000
                          Interquartile Range      45.00000


                 Tests for Location: Mu0=0

       Test           -Statistic-    -----p Value------

       Student's t    t  8.398361    Pr > |t|    0.0002
       Sign           M       3.5    Pr >= |M|   0.0156
       Signed Rank    S        14    Pr >= |S|   0.0156
```

SAS Output

```
                     Tests for Normality

 Test                  --Statistic---    -----p Value------

 Shapiro-Wilk          W     0.907963    Pr < W      0.3820
 Kolmogorov-Smirnov    D     0.232841    Pr > D     >0.1500
 Cramer-von Mises      W-Sq  0.059584    Pr > W-Sq  >0.2500
 Anderson-Darling      A-Sq  0.351151    Pr > A-Sq  >0.2500

                      The SAS System    22:58 Saturday, September 12, 2015   3

                     The TTEST Procedure

                     Variable:  pressure

N         Mean     Std Dev     Std Err     Minimum     Maximum

7        252.7     27.6328     10.4442       204.0       282.0

   Mean        95% CL Mean        Std Dev      95% CL Std Dev

  252.7       232.4  Infty        27.6328     17.8064  60.8492

                     DF     t Value    Pr > t

                      6        8.40    <.0001
```

R Code and Output

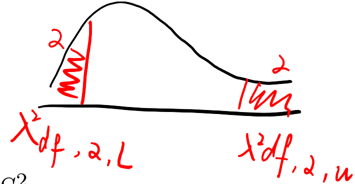
```
> x <- c(270,273,258,204,254,228,282)
> t.test(x,mu=165,alternative='greater')

One Sample t-test

data:  x
t = 8.3984, df = 6, p-value = 7.761e-05
alternative hypothesis: true mean is greater than 165
95 percent confidence interval:
 232.4193      Inf
sample estimates:
mean of x
 252.7143
```

Inference for one population variance σ^2

Inference for one population variance σ^2 for normal distribution

- Let $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$
- Let $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ be the pivotal quantity for the inference on σ^2

Inference for one population variance σ^2 for normal distribution



<u>Confidence interval for σ^2</u>

$$P(\chi^2_{n-1,\alpha/2,L} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1,\alpha/2,U}) = 1 - \alpha$$

$$\Longleftrightarrow P\left(\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2},U}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2},L}}\right) = 1-\alpha$$

$100(1-\alpha)\%$ CI for σ^2 is

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2},U}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2,L}}\right]$$

Inference for one population variance σ^2 for normal distribution

$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}.$

Hypothesis test on σ^2(one-tailed)

$$\begin{cases} H_0 : \sigma^2 \leq \sigma_0^2 \\ H_a : \sigma^2 > \sigma_0^2 \end{cases}$$

=

$P(W_0 < \chi^2_{n-1,\alpha,L})$

$= \alpha.$

- $$E(S^2) = \sigma^2$$
- Test statistic: $W_0 = \frac{(n-1)S^2}{\sigma_0^2} \overset{H_0}{\sim} \chi^2_{n-1}$
- At the significance level α, we reject H_0 if $W_0 \geq \chi^2_{n-1,\alpha,U}$
- p-value $= p_U = P(W_0 \geq w)$

observed test statistic compute from data

Inference for one population variance σ^2 for normal distribution

Hypothesis test on σ^2 (one-tailed)

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_a : \sigma^2 < \sigma_0^2 \end{cases}$$

- At the significance level α, we reject H_0 if $W_0 \leq \chi^2_{n-1,\alpha,L}$ or
- p-value $= p_L = P(W_0 \leq w)$

Inference for one population variance σ^2 for normal distribution

$P(W_0 \le \chi^2_{n-1,\frac{\alpha}{2},L})$

$+ P(W_0 \ge \chi^2_{n-1,\frac{\alpha}{2},U})$

$= \alpha$

Hypothesis test on σ^2 (two-tailed)

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_a : \sigma^2 \neq \sigma_0^2 \end{cases}$$

- $$E(S^2) = \sigma^2$$
- Test statistic: $W_0 = \frac{(n-1)S^2}{\sigma_0^2} \overset{H_0}{\sim} \chi^2_{n-1}$
- At the significance level α, we reject H_0 if $W_0 \le \chi^2_{n-1,\alpha/2,L}$ or $W_0 \ge \chi^2_{n-1,\alpha/2,U}$
- p-value $= 2\min(p_U, p_L = 1 - p_U)$

What if the population is NOT normal? & sample size is small

- If the population distribution is known, one can carry out the **LR test** (likelihood ratio test)
- If the population distribution is unknown, one can try Box-Cox normal transformation, or apply non-parametric procedures or resampling method (e.g. Bootstrap resampling).