MIE 1807
Principles of Measurement

2017-02-27

Neil Montgomery

Lecture notes:
https://github.com/mie1807-winter-2017

Point "Estimator"

An *estimator* is a statistic (in particular, a random variable) used to guess the value of a parameter.

Desirable properties:

- *accurate* (correct on average) $\longleftarrow E(\cdot)$
- *precise* (not too variable)
- *consistent* (improves as $n \nearrow$)

$\left.\begin{array}{l}\textit{precise}\\ \textit{consistent}\end{array}\right\}\text{---}\ \mathrm{Var}(\cdot)$

Example point estimators

Population: $N(\mu, 25)$ Sample $X_1, \ldots, X_n$

Usual estimator: $\overline{X}$ Stupid estimator: $\frac{X_1+X_2}{2}$

Both are "accurate":

$$E\left(\overline{X}\right) = \mu \qquad E\left(\frac{X_1 + X_2}{2}\right) = \mu$$

Example estimators

The usual estimator is more precise:

$$\mathrm{Var}\left(\overline{X}\right) = \frac{25}{n} \qquad \mathrm{Var}\left(\frac{X_1 + X_2}{2}\right) = \frac{25}{2}$$

Also from this, the usual estimator gets better with $n \nearrow$, but stupid one just sits there being stupid.

Most situations have a "usual" estimator.

Another example point estimator

One factor experiment with only two levels, 1 and 2. Population divided into two groups $N(\mu_1, 100)$ and $N(\mu_2, 100)$.

Usual question: what is the difference?

Estimate $\mu_1 - \mu_2$

Sample: $Y_{11}, \ldots, Y_{1n}, Y_{21}, \ldots Y_{2n}$

Usual estimator: $\overline{Y}_{1.} - \overline{Y}_{2.}$

Another example estimator

$$E\left(\overline{Y}_{1.} - \overline{Y}_{2.}\right) = \mu_1 - \mu_2$$

$$\begin{aligned}\mathrm{Var}\left(\overline{Y}_{1.} - \overline{Y}_{2.}\right) &= \mathrm{Var}\left(\overline{Y}_{1.}\right) + \mathrm{Var}\left(\overline{Y}_{2.}\right) \\ &= \frac{100 + 100}{n}\end{aligned}$$

Interval Estimation

Point estimators do not provide any assessment of precision.

One could instead report a range of plausible values based on two statistics L and U that satisfy (e.g. in the case of estimating a mean μ):

$$P(L \leq \mu \leq U) = 1 - \alpha$$

"Confidence Interval"

$$P(L \leq \mu \leq U) = 1 - \alpha$$

(L, U) is a $100(1-\alpha)\%$ *confidence interval*

L and U are the *confidence limits*.

$100(1-\alpha)$ is the *confidence level* and is typically close to 100, such as 90, 95, 99 etc.

Example of such an L and U

Population: $N(\mu, 25)$ Sample $X_1, \ldots, X_n$

$$\alpha = 0.05$$

$$P\left(-1.96 \leq \frac{\overline{X} - \mu}{5/\sqrt{n}} \leq 1.96\right) = 0.95$$

$$P\left(\overline{X} - 1.96\frac{5}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{5}{\sqrt{n}}\right) = 0.95$$

Informally: "$\overline{X} \pm 1.96\frac{5}{\sqrt{n}}$"

Numerical Example

Calcium concentration in an oil additive is $N(\mu, 400)$. A sample of size 25 is taken. The observed sample average is 491ppm.

A 95% confidence interval for μ is:

$$491 \pm 1.96\frac{20}{5} = (483.16, 498.84)$$

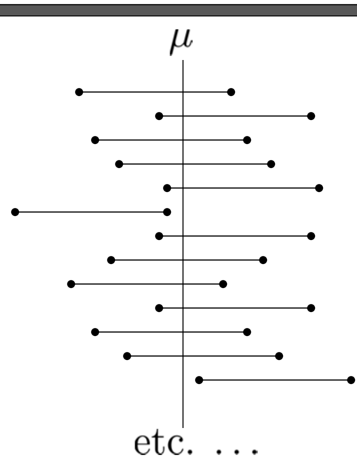
Use/Myth/Meaning

USE: To give a range of plausible values for μ that accounts for the sampling procedure undertaken.

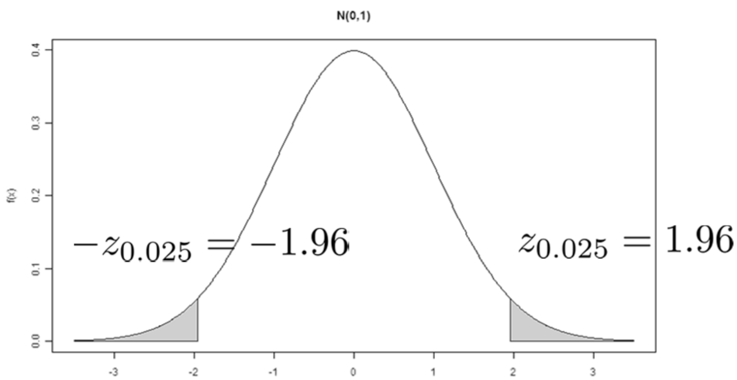
MYTH: $P(483.16 \leq \mu \leq 498.84) = 0.95$

LITERAL MEANING: If many (hypothetical) samples were observed, and a C.I. computed each time, then $100(1-\alpha)\%$ of the intervals will contain μ.

Meaning



What influences the width of a confidence interval?



Generally: $\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Sample size for a given "margin of error"

The difference between estimator and parameter (e.g. $\overline{X}$ and μ) can be called *margin of error* denoted by m

Problem: determine n so that $\left|\overline{X} - \mu\right| < m$ with probability $1 - \alpha$.

Solution: since $|m| = \left|z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right|$, just solve for n in:

$$m = z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \Longrightarrow \quad n = \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2$$

Sample size numerical example

Calcium concentration in an oil additive is $N(\mu, 400)$. What sample size is required to obtain an estimate for μ within 5ppm with probability 0.95?

$$n = \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2 = \left(\frac{1.96 \cdot 20}{5}\right)^2$$
$$= 61.4656$$

Sample size required is 62 (conservative).

A more realistic situation

It is unlikely to have $N(\mu, 400)$, say, as a population model, i.e. with variance known.

More realistic: $N(\mu, \sigma^2)$ both parameters unknown.

Confidence interval based on:

$$P\left(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

What to do about the unknown variance?

What is the obvious thing to try?

Estimate σ^2 with something—obviously (?) with the sample variance s^2

Proposed new formula: $\overline{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

Is it valid? No - the "coverage" probability is too low for small to moderate n.

A new (family of) distribution(s)

Sample $X_1, X_2, \ldots, X_n$ from a $N(\mu, \sigma^2)$

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

"t distribution" with parameter $n-1$, where n is the sample size.

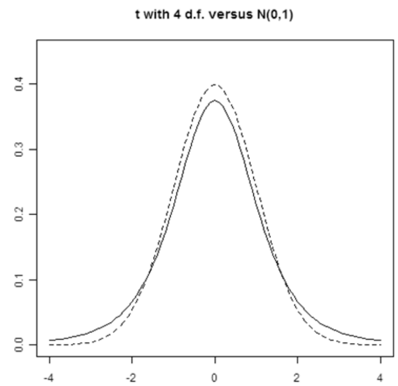
Notation: t_{n-1}

Fundamental Difference between *t* and *Z*

Densities:

$$f_Z(z) \propto e^{-z^2/2}$$

$$f_t(t) \propto \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$



"One Sample *t* confidence interval for the mean"

$$\overline{X} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}$$

Calcium concentration in an oil additive is $N(\mu, \sigma^2)$. A sample of size 25 is taken. The observed sample average is 491ppm and the observed sample variance is 426..

A 95% confidence interval for μ is:

$$491 \pm 2.064\frac{\sqrt{426}}{5} = (482.47, 499.56)$$

More on Estimating the Variance

Why $n-1$ and not n in:

$$S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

The formula with n is *biased*.

Practical Note on Sample Size Calculation

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

But σ is usually unknown. Some practical options:

- use a prior study or gather a pilot sample to estimate σ

- alternative: use the upper limit from a C.I. for σ

- find a software/online implementation of the calculation.

CI when the population distribution is not known

Similar to the CLT we also have:

$$\frac{\overline{X} - \mu}{s/\sqrt{n}} \sim_{approx} t_{n-1}$$

...for n large enough - same criteria.

Procedure:
- plan and gather data (incl. sample size)
- look at plots to assess (non-)normality as soon as possible.
- perform computation

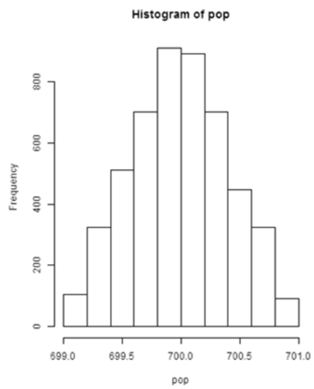
Example

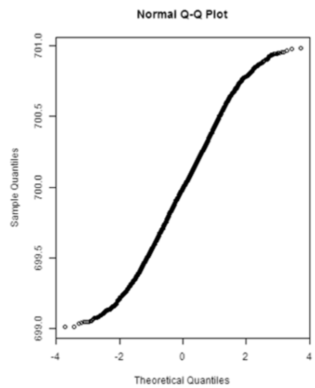
A pre-fabricated furniture maker wants to estimate the mean width of wooden component it gets from a supplier. The nominal width is 700mm.

Goal: a 95% CI with a margin of error of 0.05mm.

Sample size problem: will gather pilot sample of 50 parts to estimate population variance, and to assess (non-)normality.

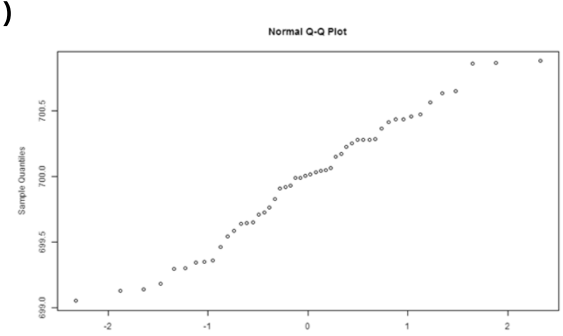
Example – *Some population info – not realistic!*





Example – Information from pilot sample

```
> summary(pilot.widths)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  699.2   699.7   700.1   700.0   700.3   700.8
> sd(pilot.widths)
[1] 0.3899095
```



Example – Sample size calculation

$$n = \left(\frac{z_{\alpha/2}\sigma}{m}\right)^2 \approx \left(\frac{1.96 \cdot 0.39}{0.05}\right)^2$$

$$= 233.7$$

So we need to collect 233 or so (234, whatever) samples.

Could we use the 50 already collected?

Example – The interval

We collect another 234 samples.

```
> mean(sample.widths);
sd(sample.widths)
[1] 699.9934
[1] 0.431008
```

$t_{233,\alpha/2} \approx -1.96$ since the sample size is quite big (but is actualy -1.9702 by computer)

$$699.99 \pm 1.96\frac{0.431}{\sqrt{234}} \Longrightarrow (699.94, 700.05)$$

Example – Using Software (R) Directly

```
> t.sample.widths <- t.test(sample.widths)
> t.sample.widths
      One Sample t-test

data:  sample.widths
t = 24843.71, df = 233, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 699.9379 700.0490
sample estimates:
mean of x
 699.9934
```


Terminology: *Standard Error*

If $X_1, \ldots, X_n$ is a sample from a $N(\mu, \sigma^2)$ distribution...

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

$\frac{s}{\sqrt{n}}$ is the *estimated standard deviation* for the estimator $\overline{X}$.

Can be shortened to: *standard error* or just $\mathrm{se}(\overline{X})$

Universal Confidence Interval Formula

$\boldsymbol{X} = \{X_1, \dots, X_n\}$ is a sample and $\hat{\theta}(\boldsymbol{X})$ is an estimator for the parameter θ.

A $100 \cdot (1-\alpha)\%$ confidence interval will be:

$$\hat{\theta}(\boldsymbol{X}) \pm q_{\alpha/2} \cdot se(\hat{\theta}(\boldsymbol{X}))$$

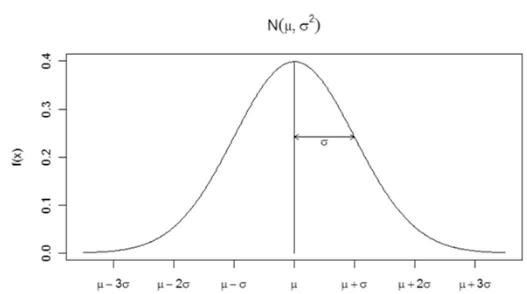
where $q_{\alpha/2}$ is a number computed from the appropriate distribution. (For us usually a t distribution or the $N(0,1)$)

Prediction – The General Problem

Given a population $N(\mu_0, \sigma_0^2)$, with both parameters known, consider the problem of predicting a new obsvervation X^*.

Best predictor? μ_0

“95% Interval?”



Prediction with unknown parameters

Suppose the population were $N(\mu, \sigma^2)$ parameters unknown, but you still want to predict X^*.

Obtain a sample $X_1, X_2, \ldots, X_n$

Best predictor? $\overline{X}$

"Interval"?

Prediction Interval Background

We need to consider the nature of $\overline{X} - X^*$.

Distribution shape: Normal.

$$E\left(\overline{X} - X^*\right) = \mu - \mu = 0$$

$$\begin{aligned}\text{Var}\left(\overline{X} - X^*\right) &= \text{Var}\overline{X} + \text{Var}X^* \\ &= \frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(\frac{1}{n} + 1\right)\end{aligned}$$

Estimate σ^2 as usual.

Prediction Interval

A $100\cdot(1-\alpha)\%$ *prediction interval* for X^* is given by:

$$\overline{X} \pm t_{n-1,\alpha/2}\, s \sqrt{1+\frac{1}{n}}$$

BUT!

The central limit theorem (and friends) don't help here:

$$\frac{\overline{X} - X^*}{\sigma\sqrt{1+1/n}} \sim N(0,1)$$

Prediction Interval "Example"

From those 234 samples:

```
> mean(sample.widths);
sd(sample.widths)
[1] 699.9934
[1] 0.431008
```

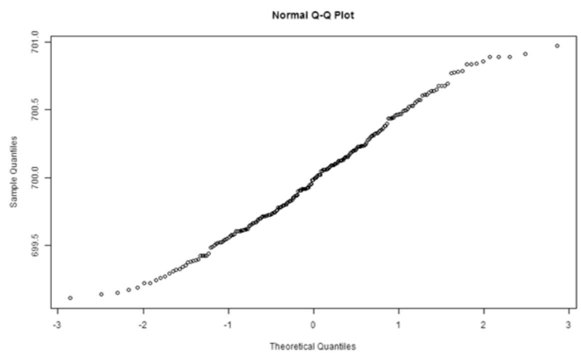
The interval would be

$$699.9954 \pm (1.9702)(0.431008)\sqrt{1+1/234}$$

$$(699.1424, 700.8444)$$

Prediction Interval "Example"

However, from the plot the population isn't normal, so the interval is actually too ????????????



Prediction Interval Procedure

- Make plan to collect data, including possible pilot samples or an assessment of previous work, and sample size requirements.
- Look at relevant plots as soon as possible – normality essential for the formulae given in this course.
- Do calculations
- If non-normal, PIs can still be done, but a specific population model must first be determined.

Hypothesis Testing

Sometimes specific statements about parameters have practical meaning:

Model:

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

with $i \in \{1, 2\}$, i.e. "two samples".

Statements: "$\mu_1 = \mu_2$" "$\mu_1 \neq \mu_2$"

Null Hypothesis/ Alternative Hypothesis

Hypothesis testing involves an evaluation of two hypotheses "null" and "alternative".

The hypotheses define disjoint parts of the parameter range.

The null hypothesis, denoted H_0, tends to embody "no effect".

Example from previous page: $H_0 : \mu_1 = \mu_2$

Other typical "nulls"

Model: $Y_j = \mu + \varepsilon_j$, i.e. "1 sample".

$$H_0 : \mu = \mu_0 \qquad \mu_0 \text{ is some constant.}$$

Model: $Y_{ij} = \mu_i + \varepsilon_{ij}$ with $i \in \{1, \ldots, I\}$, i.e. "I samples".

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$$

Model: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$H_0 : \beta_1 = 0$$

Model: $Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i$

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

Alternative Hypotheses

In almost all cases, aside from special exceptions, the alternative hypothesis H_a will be the "complement" of the null.

Two samples: $H_0 : \mu_1 = \mu_2$

$$H_a : \mu_1 \neq \mu_2$$

Aside: The Myth of the "One-sided Alternative"

Textbooks go on about "choosing" the "appropriate" alternative hypothesis, based on little more than the hopes and dreams of the experimenter.

I believe this to be nonsense, usually.

Most common exception: H_0 is at the "edge" of the parameter range.

Classical Hypothesis Testing

Goal: decide to *reject* or *not reject* H_0.

Method: assume H_0 is true, collect a sample, and see if the sample contradicts H_0.

Motivating example: one sample model $Y_j = \mu + \varepsilon_j$ with $\varepsilon_j \sim N(0, \sigma^2)$.

Equivalent: $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$

Classical Hypothesis Testing Motivating Example

Story: does the supplier send doors with a mean of 700mm, or not?

$$H_0 : \mu = 700$$

$$H_a : \mu \neq 700$$

Sample: $X_1, \ldots, X_{10}$

What *statistic* to use? $\overline{X}$ *test statistic*

What values of $\overline{X}$ should surprise us (assuming H_0)?

Classical Hypothesis Testing Motivating Example

Not quite enough information. Suppose $\sigma = 0.5$ is known, somehow.

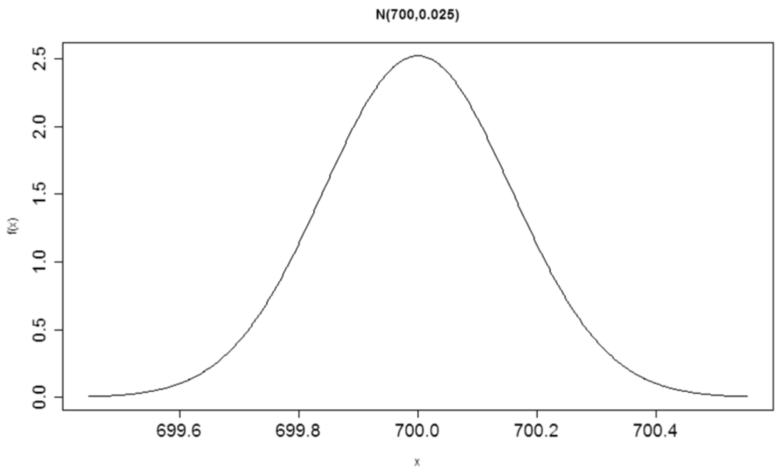
Then, assuming H_0, $\overline{X} \sim N(700, 0.025)$

$\Downarrow$

null distribution

What values of $\overline{X}$ should surprise us (assuming H_0)?

Classical Hypothesis Testing Motivating Example



Classical Hypothesis Testing - Details

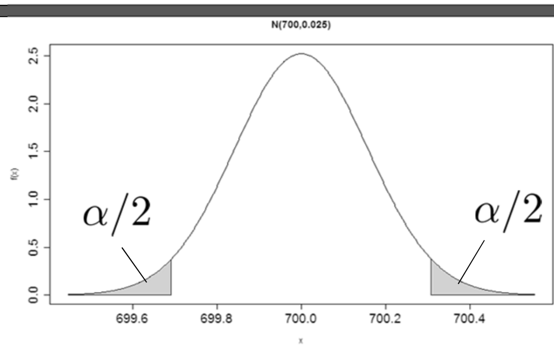
"The values that would surprise us" are defined in advanced according to a pre-set probability α

α is called *level of significance* or *size*.

α is usually small: 0.1, 0.05, 0.01, 0.001, etc.

α is the probability of rejecting H_0 when it is in fact true.

Classical Hypothesis Testing - Details



Suppose $\alpha = 0.05$

The "area of suprise" (*rejection region* or *critical region*) is: $\overline{X} \leq 699.6901$ and $\overline{X} \geq 700.3099$

Types of Error

	Reject	Not Reject
H_0 True	Type I Error	
H_0 False		Type II Error

"Power"

$\alpha = P(\text{Type I Error})$ $\qquad \beta = P(\text{Type II Error})$

$1 - \beta$ is called the *power* of the test.

(computing power requires a specific alternative)

Example

(From text #2, p.156) A new LED light needs to meet a standard of 75,000 hours average life.

Assume population is $N(\mu, 6250^2)$.

$$H_0 : \mu = 75000 \qquad H_a : \mu \neq 75000$$

Set $\alpha = 0.01$

Example

Null distribution: $\overline{X} \sim N(75000, 6250^2/100)$

Find the regions on either side of H_0 that have 0.01 probability (total).

$$\overline{X} \leq 73390.11 \qquad \overline{X} \geq 76609.89$$

A sample of size 100 is collected: $\overline{x} = 76500$

Do not reject H_0.

Example Power Calculation

What if in fact $\mu = 76000$?

Then really $\overline{X} \sim N(76000, 6250^2/100)$

And the power of the test procedure is:

$$P(\overline{X} \leq 73390.11) + P(\overline{X} \geq 76609.89)$$

$$= 0.0000148 + 0.1646$$