MIE 1807
Principles of Measurement

2017-03-13

Neil Montgomery

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

Universal Confidence Interval Formula

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

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

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

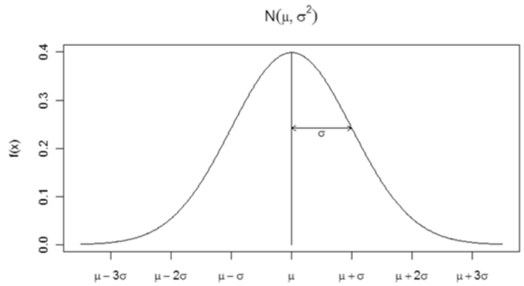
where $q_{\alpha/2}$ is a number computed from the appropriate distribution. (For us usually a t distribution or the $N(0, 1)$)

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
```

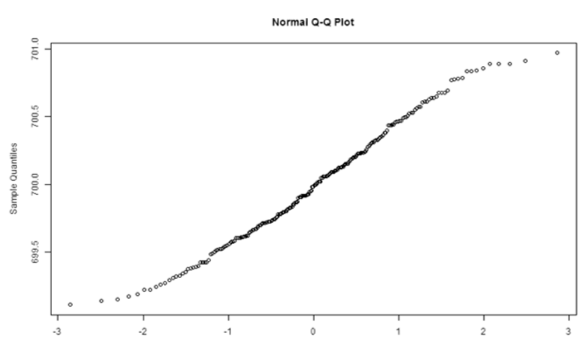
The interval would be

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

$$(699.1424, 700.8444)$$

Prediction Interval "Example"

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



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 $i \in \{1, \dots, I\}$, 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: $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$

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

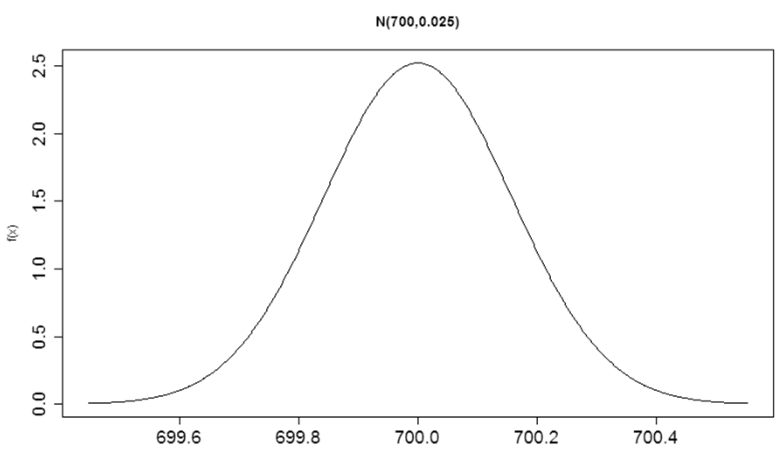
Not quite enough information. Suppose $\sigma = 0.5$ is known, somehow.

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

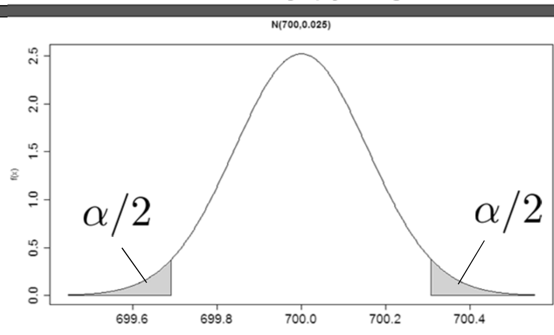
"The values that would surprise us" are defined in advanced according to a pre-set probability α

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:

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

$$= 0.0000148 + 0.1646$$

Some Issues with (Classical) Hypothesis Testing

The null hypothesis is not ever actually true.

"Statistical significance" versus "practical significance".

The "file drawer" problem.

Solution: *think!*

Reject/not reject framework is arbitrary.

Alternative: *p-values*

Sample Size Requirement Example

Consider the LED example. Say the old ads claimed 75000 hours.

The Marketing Dept. wants to be able to say "New and Improved!" without getting sued.

They decide the minimum practical difference is +1000h. They would like to be able to detect this difference using a 0.05-sized test with power $1 - \beta = 0.8$.

Sample Size Requirement

Population $N(\mu, \sigma^2)$; sample $X_1, \ldots, X_n$

$$H_0 : \mu = \mu_0 \text{ versus } H_a : \mu \neq \mu_0$$

Rejection region will be:

$$\left\{\overline{X} \le \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} \cup \left\{\overline{X} \ge \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

The sample size will come from computing the probability of this region, given $\mu = \mu_1$ (defined as minimally interesting)

Sample Size Requirement

$$1-\beta = P_{\mu_1}\left(\overline{X} \le \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) + P_{\mu_1}\left(\overline{X} \ge \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

In practice, one probability is going to be very small!

$$1 - \beta = P_{\mu_1}\left(\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} \le \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}\right)$$

$$z_\beta = \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \quad \Longrightarrow \quad n = \sigma^2 \frac{(z_\beta + z_{\alpha/2})^2}{(\mu_0 - \mu_1)^2}$$

Sample Size Requirement Example Solution

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

$\sigma^2 = 6250^2$ $\qquad |\mu_0 - \mu_1| = 1000$

$z_{0.025} = 1.96$ $\qquad z_{0.2} = 0.842$

$$n = 306.7$$

Equivalence of Classical Hypothesis Testing and Confidence Intervals

$N(\mu, \sigma^2)$ population, $H_0 : \mu = \mu_0$ vs. $H_a : \mu \neq \mu_0$ gives R.R.:

$$\left\{\overline{X} \le \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} \cup \left\{\overline{X} \ge \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

The $100(1-\alpha)\%$ confidence interval is:

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

So an α-sized test can be performed using a C.I. simply by seeing if it contains μ_0

Example

Calcium concentration in an oil additive is $N(\mu, 400)$. It is desired to test $H_0 : \mu = 500$ versus $H_a : \mu \neq 500$ with $\alpha = 0.05$

A sample of size 25 is taken. The observed sample average is 491ppm. The 95% confidence interval is $(483.16, 498.84)$

Conclusion: reject H_0.

P-values

Classical hypothesis testing is good for learning, theory, and some computations (e.g. sample size requirements), but less good for decision-making.

p-value: the probability, assuming H_0, of observing an even more extreme value of the test statistic.

Assesses strength of evidence against H_0 (smaller means more evidence.)

P-value example

Two competing companies A and B sell LEDs whose lives follow $N(75000, 6250^2)$ and $N(70000, 5800^2)$ respectively.
Both test a new production process, and test to see if the new process changes the life length, using samples of size 300.

$H_0^A : \mu_A = 75000$ $\quad$ $H_0^A : \mu_B = 70000$

$H_a^A : \mu_A \neq 75000$ $\quad$ $H_a^A : \mu_B \neq 70000$

$\overline{X}_A = 75914$ $\quad$ $\overline{X}_B = 71021$

Using Classical approach with $\alpha = 0.05$, say, both would conclude "reject H_0"

P-value example

Company A:

$$p = P(\overline{X} \geq 75914) + P(\overline{X} \leq 74086)$$

$$= 0.0113 + 0.0113 = 0.0226$$

Company B:

$$p = P(\overline{X} \geq 71021) + P(\overline{X} \leq 68979)$$

$$= 0.0023 + 0.0023 = 0.0046$$

p-value is more informative.

Experimental Designs

- We've covered the basics of inference.
- We'll no longer assume population variance is known.
- We'll now go over a variety of practical experimental settings, considering:
  - Sample size requirements
  - How to make the inferences
  - How to check the model assumptions
- Recall terms "factor" and "level"

One "factor" with one "level"

Model: Population $N(\mu, \sigma^2)$

Or, Model: $Y_i = \mu + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2)$

Sample: $Y_1, \ldots, Y_n$ Key fact: $\frac{\overline{Y}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

Either exactly or approximately… (plots)

Confidence interval: $\overline{Y} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$

Hypothesis testing: $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$

One factor/one level – Hypothesis Testing

Reject region or p-value comes from t distributions.

Example: LED lighting, $H_0 : \mu = 75000$ etc.

Sample $Y_1, \dots, Y_n$ is gathered; $n = 30$; observed sample average and sample s.d. are 75782 and 6821.

Classical method (with $\alpha = 0.05$) results in R.R. of:

$$\left\{\overline{Y} \leq \mu_0 - t_{\alpha/2}\frac{s}{\sqrt{n}}\right\} \cup \left\{\overline{Y} \geq \mu_0 + t_{\alpha/2}\frac{s}{\sqrt{n}}\right\}$$

One factor/one level – Hypothesis Testing

Classical method results in R.R. of:

$$\left\{\overline{Y} \leq 72453\right\} \cup \left\{\overline{Y} \geq 77547\right\}$$

"Do not reject." (Plus, check plot of data.)

–OR–

P-value approach:

$$\begin{aligned} & P(\overline{Y} \leq 74218) + P(\overline{Y} \geq 75782) \\ &= P(t_{29} \leq -0.628) + P(t_{29} \geq 0.628) \\ &= 0.535 \end{aligned}$$

(Plus check plot of data.)