MIE 1807
Principles of Measurement

2017-03-27

Neil Montgomery

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

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:

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

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.:

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

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.

$H_0^A : \mu_A = 75000$	$H_0^A : \mu_B = 70000$
$H_a^A : \mu_A \neq 75000$	$H_a^A : \mu_B \neq 70000$
$\overline{X}_A = 75914$	$\overline{X}_B = 71021$

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
- Terminology: "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 $Y_1, \ldots, Y_n$ 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.)

Scenario: One Factor, Two Levels

Model: Populationz $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$

Or, Model: $Y_{ij} = \mu_i + \varepsilon_{ij}$ with $\varepsilon_{ij} \sim N(0, \sigma^2)$ and $i \in \{1, 2\}$

Samples: $Y_{11}, \ldots, Y_{1n_1}$ and $Y_{21}, \ldots, Y_{2n_2}$

$$\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

Scenario: One Factor, Two Levels

Model: Populationz $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$

Or, Model: $Y_{ij} = \mu_i + \varepsilon_{ij}$ with $\varepsilon_{ij} \sim N(0, \sigma^2)$ and $i \in \{1, 2\}$

Samples: $Y_{11}, \ldots, Y_{1n_1}$ and $Y_{21}, \ldots, Y_{2n_2}$

$$\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot} \sim N\left(\mu_1 - \mu_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

One Factor, Two Levels

$$\frac{\left(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}\right) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Key fact:

$$\frac{\left(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}\right) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

One Factor, Two Levels

The Key Fact is approximately true as long as *both* sample sizes are big enough (check normal plots of both samples), and the true population variances are equal (I'll show you a formal test later.)

One Factor, Two Levels: Example

Two competing companies A and B sell LEDs whose lives follow unknown distributions. Are the life lengths the same?

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_a : \mu_1 \neq \mu_2$$

Samples taken:

$$\begin{array}{ll} \bar{y}_{1\cdot} = 75914 & \bar{y}_{2\cdot} = 71201 \\ s_1 = 6125 & s_2 = 5962 \\ n_1 = 25 & n_2 = 27 \end{array}$$

One Factor, Two Levels: Example

$$s_p = \sqrt{\frac{24(6125^2) + 26(5962^2)}{50}} = 6040.8$$

(Pretend the normal plots are good. The sample SDs are close.)

P-value:

$$P\left(t_{50} \geq \frac{75914 - 71201}{6040.8\sqrt{\frac{1}{25} + \frac{1}{27}}}\right) \cdot 2 = 0.0070$$

One Factor, Two Levels: Example

What if I had wanted to estimate the difference, with 95% confidence, say?

$$\left(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot}\right) \pm t_{n_1+n_2-2,0.025} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Easy!

One Factor, Two Levels: Variations

You might encounter these situations in practice:

Situation	Approach
Truly Unequal Variances	Look up "unequal variances" in any textbook.
Small non-normal samples.	Don't give up - hire statistician.