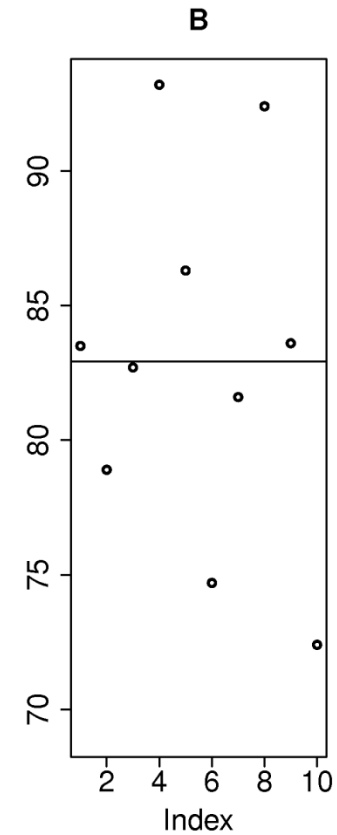
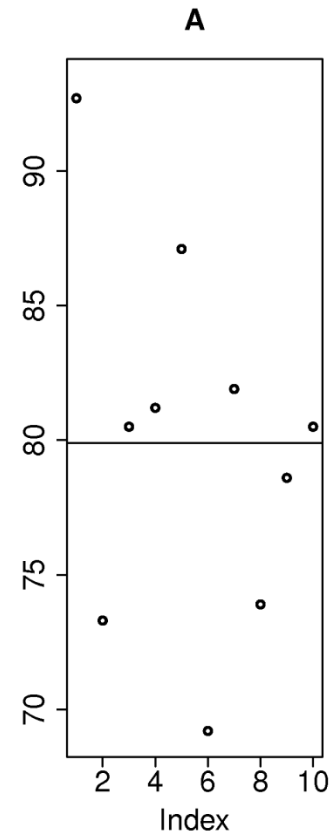
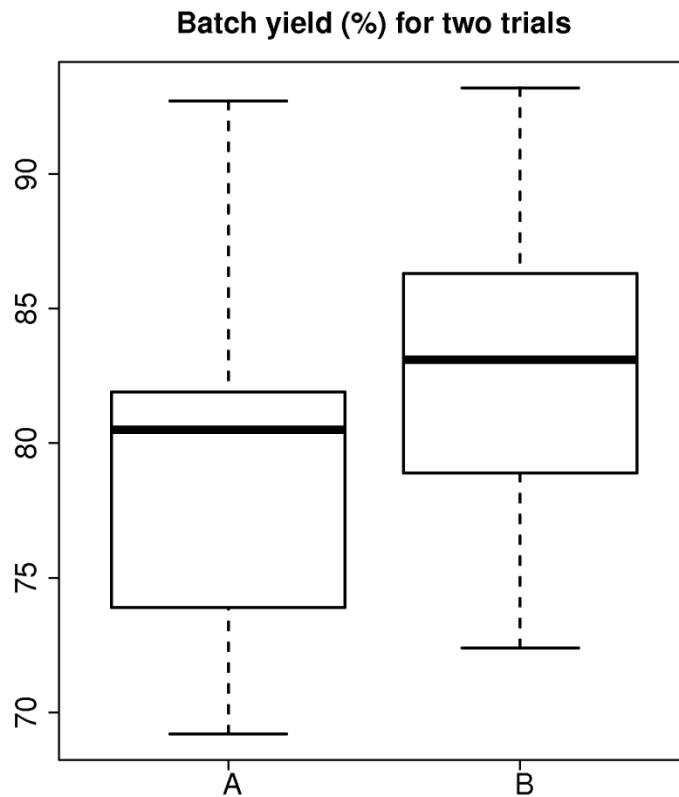


# Test for differences/similarity

- We want to change to a cheaper material, B. Does it work as well as A?
- We want to introduce a new catalyst B. Does it improve our product properties over the current catalyst A?
- Example
  - An engineer needs to verify that new feedback controller (B) on batch reactor leads to improved yields. Compare with yields from feedback controller A.
    - There are 10 sequential runs with system A, then 10 runs with system B.
    - System B will cost us \$100,000 to install, and \$20,000 in annual software license fees.
    - A significant difference means long-run implementation of B will lead to an improved yield (not due to chance)

# Test for differences/similarity

➤ What to do first?



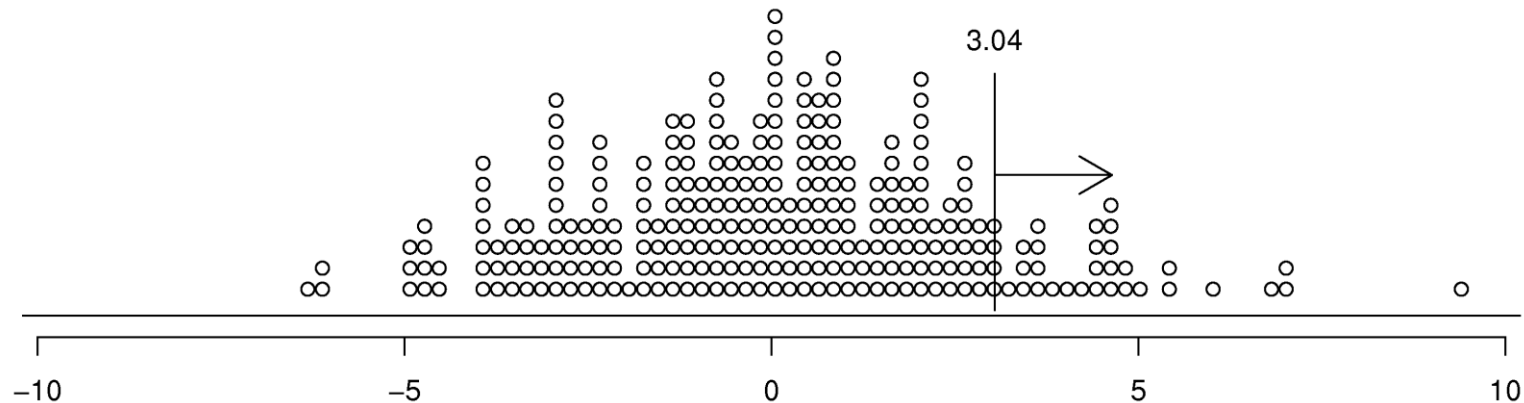
# Test for differences/similarity

➤ Data acquired

Experiment number	Feedback system	Yield	Experiment number	Feedback system	Yield
1	A	92.7	11	B	83.5
2	A	73.3	12	B	78.9
3	A	80.5	13	B	82.7
4	A	81.2	14	B	93.2
5	A	87.1	15	B	86.3
6	A	69.2	16	B	74.7
7	A	81.9	17	B	81.6
8	A	73.9	18	B	92.4
9	A	78.6	19	B	83.6
10	A	80.5	20	B	72.4
Mean		79.89	Mean		82.93
Standard deviation		6.81	Standard deviation		6.70

# Test for differences/similarity

- What would be the simplest yet not a bad method?
  - Compare with long-term reference set.
    - Ex. 300 batches last 10 years.



- 31 historical differences out of 281 had a difference value higher than 3.04.
  - 89 % of previous batches had lower yield.
- What if no reference set?

# Test for differences/similarity

- Don't have a suitable reference, e.g. just the 20 runs.
- But still want to know  $\mu_A > \mu_B$  or the other.
  - Comparing  $\mu_A$  &  $\mu_B$  requires C.Is of  $\mu_A$  &  $\mu_B$ . Why?
- There are several cases to consider
  - Unpaired, variance known
  - Comparison of variances
  - Unpaired, variances unknown but equal
  - Unpaired, variances unknown and unequal
  - Paired

## Case1: Unpaired data, $\sigma^2_1$ and $\sigma^2_2$ are known

- No pairing of data: there is no relationship between an individual value in one sample and an individual value in the other sample (i.e., independent)
- Compute:

$$\bar{X} - \bar{Y}$$

We ask the question, “**is 0 a plausible value for  $\bar{X} - \bar{Y}$  ?**” i.e. is there no difference between the means?

## Case1: Unpaired data, $\sigma^2_1$ and $\sigma^2_2$ are known

We can use confidence intervals to answer this question.

To do this we need to know the distribution of  $\bar{X} - \bar{Y}$ .

Since  $\bar{x}$  and  $\bar{y}$  are both normally distributed with variances  $\frac{\sigma_x^2}{n_x}$  and  $\frac{\sigma_y^2}{n_y}$  respectively.

$$(\bar{X} - \bar{Y}) \sim N\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}\right)$$

i.e. the variance of  $\bar{X} - \bar{Y}$  is  $\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$  assuming that X and Y are independent

# Case1: Unpaired data, $\sigma^2_1$ and $\sigma^2_2$ are known

$$1 - \alpha = \Pr ob \left\{ -z_{\alpha/2} \leq \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} \leq z_{\alpha/2} \right\}$$

and the  $100(1-\alpha)$  % confidence interval for the difference in population means is

$$\left[ (\bar{X} - \bar{Y}) - z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}, (\bar{X} - \bar{Y}) + z_{\alpha/2} \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right]$$

**If the confidence interval contains zero, then there is no statistically significant evidence (at the  $1-\alpha$  level) of a difference in means.**



## Case 2a: variances are unknown, but the same

➔ Use pooled variances

$$s_p^2 = \frac{\nu_x s_x^2 + \nu_y s_y^2}{\nu_x + \nu_y}, \quad \nu_x \text{ \& \ } \nu_y : \text{D.O.F of each samples}$$

$$\frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}} \sim t_\nu$$

$$\left[ (\bar{X} - \bar{Y}) - t_{\nu, \alpha/2} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}, \quad (\bar{X} - \bar{Y}) + t_{\nu, \alpha/2} \sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}} \right]$$

## Case 2a: variances are unknown, but the same

➔ Comparison of sample variances

$$\frac{s_x^2 \sigma_y^2}{s_y^2 \sigma_x^2} \sim F_{\nu_x, \nu_y}$$

$$1 - \alpha = \Pr \text{ob} \left\{ \frac{s_x^2}{s_y^2 F_{\nu_x, \nu_y, \alpha/2}} < \frac{\sigma_x^2}{\sigma_y^2} < \frac{s_x^2}{s_y^2 F_{\nu_x, \nu_y, 1-\alpha/2}} \right\}$$

$$\left[ \frac{s_x^2}{s_y^2 F_{\nu_x, \nu_y, \alpha/2}}, \frac{s_x^2}{s_y^2 F_{\nu_x, \nu_y, 1-\alpha/2}} \right] \quad \text{If this contains 1, then.....}$$

## Case 2b: variances are unknown and different

- There is no exact procedure for this case but there are several approximate procedures.
- One such procedure is known as Cochran's procedure
- A crude approach is to construct a confidence interval for  $\mu_x$  based on the estimated variance of the x sample, and then construct a confidence interval for  $\mu_y$  using the estimated variance for the y sample. If the confidence intervals overlap, then there is no evidence at the  $100(1-\alpha)$  % confidence level of a difference.
- Minitab can handle this.

## Case 3: Paired comparisons between two groups

- The idea is to remove extraneous sources of variation by taking measurements in pairs, each pair comprising one measurement from each of the two groups. Each pair should have as much in common as possible except the difference that is trying to be detected.
- Each pair is called a block. The idea of blocking will be discussed in more detail when we talk about Design of Experiments.

# Paired comparisons between two groups

$$w_i = x_i - y_i$$

$$\bar{w} = \frac{\sum_{i=1}^n w_i}{n}$$

$$s_w^2 = \frac{\sum_{i=1}^n (w_i - \bar{w})^2}{n-1} \quad \text{with } v_w = n-1 \text{ degrees of freedom}$$

$$\left[ \bar{w} - t_{v, \alpha/2} s_w / \sqrt{n}, \bar{w} + t_{v, \alpha/2} s_w / \sqrt{n} \right]$$

# Example

- In an oxide coating operation on silicon chips, a series of test were conducted to determine whether position in the furnace had any appreciable effect on the resulting thickness of the oxide coating. Within each furnace run, chips were randomly assigned to two positions, denoted by A and B, within the furnace. 10 independent furnace runs were carried out and a single thickness measurement (in Angstroms) was made at the centre of each chip.

• <b>Run</b>	<b>A</b>	<b>B</b>	<b>w</b>
• 1	920	923	-3
• 2	914	924	-10
• 3	927	913	14
• 4	891	881	10
• 5	943	923	20
• 6	902	884	18
• 7	910	887	23
• 8	856	858	-2
• 9	937	916	21
• 10	857	857	0

# Solution

$$[ \bar{w} - t_{\nu, \alpha/2} s_w / \sqrt{n} , \bar{w} + t_{\nu, \alpha/2} s_w / \sqrt{n} ]$$

$$[ 9.1 - t_{13, 0.025} s_w / \sqrt{n} , 9.1 + t_{13, 0.025} s_w / \sqrt{n} ]$$

$$[ 9.1 - 2.26 \frac{11.9}{\sqrt{10}} , 9.1 + 2.26 \frac{11.9}{\sqrt{10}} ]$$

$$[ 0.6, 17.6 ]$$

# Advantages of Pairing

- Extraneous sources of variation are removed or minimized.
- The test of significance is more sensitive because the total variation is minimized.
- Measurements are paired based on proximity in time or space, or in some cases, blocking (pairing) is done by piece, person or animal
- Within each pair, the order of the measurements should be randomized.
- Pairing helps to remove one or more sources of variation.
  - Randomization helps to spread the effects of extraneous sources of variation over all measurements so as to avoid accidentally confounding the effect of interest with an extraneous source of variation.



# Examples

## ➤ 1-sample Z / 1-sample t

- Previously, process lead time (L/T) was 30 days. Recently, process L/T is surveyed (see the table below) after successful completion of a six sigma project. From this data, can you say that the six sigma project was successful to reduce process L/T?

Data: 23, 28, 35, 26, 22, 25, 28, 29, 24, 30, 18, 23, 24, 29, 28

(1) When  $\sigma^2$  is known to be  $4.09^2$ .

Stat > Basic statistics > 1-sample Z

(2) When  $\sigma^2$  is unknown.

Stat > Basic statistics > 1-sample t

# Examples

## ➤ 2-sample t

- The components library was standardized in order to reduce set-up loss of SMD process. Following data were measured to know whether actual set-up loss time was reduced. What do you think?

Before	12.50	12.63	13.25	13.00	12.86	12.75	13.22	12.90	12.96	13.15
after	12.35	12.81	12.00	11.95	12.12	12.64	12.56	12.75	12.19	12.17

Stat > basic statistics > 2-sample t

# Hypothesis test (가설의 검증)

- When a factor appears to have an effect, it is important to be able to state **with confidence** that **the effect was really due to the factor and not just due to chance**.

Ex. We want to introduce a new catalyst B. Does it improve our product properties over the current catalyst A?

Ex. Drug test



- How can one be sure that the drug treatment rather than chance occurrences were responsible for the difference between the groups?

# [FYI] Vocabulary

- Confidence level,  $1 - \alpha$  (or  $100(1 - \alpha)\%$ )
  - The confidence level tells you how sure you can be. It is expressed as a percentage and represents how often the true percentage of the population who would pick an answer lies within the confidence interval.
  - The 95% confidence level means you can be 95% certain.
- Significance level,  $\alpha$  (or  $100\alpha\%$ )
  - the criterion used for rejecting the null hypothesis.
- Null hypothesis (歸無/虛無 가설)
  - Ex. The increased yield of catalyst B is the result of pure chance.
- Alternative hypothesis
  - Ex. The null hypothesis is false. The increased yield of catalyst B is not the result of pure chance, but is .

# [FYI] meaning of a p-value

- A measure of how much evidence we have **against** the null hypothesis.
- The general rule is that a **small p-value is evidence against the null hypothesis** while a **large p-value means little or no evidence against the null hypothesis**.
- The p-value is **not** the probability that the null hypothesis is true.
- In a court room,
  - $H_0$ : the defendant is innocent.     $H_1$ : the defendant is guilty.

Judge's decision \ Fact	He is innocent	He is guilty
He is innocent (Accept $H_0$ )	right	Wrong Type II error (2종 오류)
He is guilty (Reject $H_0$ )	Wrong Type I error (1종 오류)	right

# Examples

- Redo previous examples using (formal) hypothesis test.
  - What would be appropriate null & alternative hypotheses for each example?

# 가설검정 로드맵

