Section 4.1-4.2 Pairwise Comparison & Contrasts
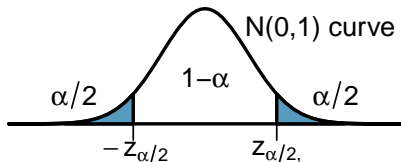
Yibi Huang

Analysis of Factor Level Means
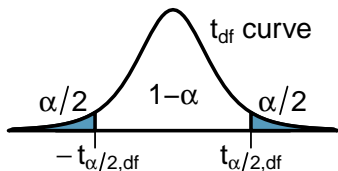
- When *F*-test is significant; there exist differences among the means. Now what?
- Want to determine which means are different and identify treatments statistically of the same effect

Notations for Normal and t-Critical Values

In the remainder of the course, we use $z_{\alpha/2}$ to denote the critical value of $N(0,1)$ such that $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$ for $Z \sim N(0,1)$.

We use $t_{\alpha/2,df}$ to denote the critical value such that $P(-t_{\alpha/2,df} < T < t_{\alpha/2,df}) = 1 - \alpha$ where T has a t-distribution w/ df degrees of freedom

Confidence Interval (CI) for One-Sample Mean (Review)

Recall that for a one-sample problem, $X_1, X_2, \ldots, X_n$ are i.i.d. from a population with **unknown mean** μ and SD σ.

If σ is known, a $(1 - \alpha)100\%$ CI for μ is given by

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

If σ is unknown and is estimated with the sample SD s, a $(1 - \alpha)100\%$ CI for μ is given by

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

Confidence Interval for Treatment Means

For one way ANOVA model,

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

a naive $100(1-\alpha)\%$ CI for μ_i is

$$\overline{y}_{i\bullet} \pm t_{\alpha/2, n_i - 1} \times \frac{s_i}{\sqrt{n_i}}, \quad \text{where } s_i = \text{sample SD of the } i\text{th group}$$

- using only data in the ith group, ignoring the rest,
- not optimal

A better $100(1-\alpha)\%$ CI for μ_i is

$$\overline{y}_{i\bullet} \pm t_{\alpha/2, N-g} \frac{\sqrt{\text{MSE}}}{\sqrt{n_i}}$$

- using observations in all groups to estimate the unknown σ
- with a higher df $= N - g$, not $n_i - 1$

Case Study: Grass/Weed Competition

Treatment	1N	1Y	2N	3N	4N	4Y
Mean $\overline{y}_{i\bullet}$	95	82.25	81.5	68.25	50.5	52

$\text{MSE} = 17.97$

As $N = 24$, $g = 6$, for $\alpha = 0.05$, the t-critical value is $t_{0.05/2, 24-6} \approx 2.101$.

```
> qt(0.05/2, df = 24-6, lower.tail=F)
[1] 2.100922
```

The 95% CI for μ_{4Y} is

$$\overline{y}_{i\bullet} \pm t_{\alpha/2, N-g} \frac{\sqrt{\text{MSE}}}{\sqrt{n_i}} = 52 \pm 2.101 \times \frac{\sqrt{17.97}}{\sqrt{4}} \approx 52 \pm 4.45$$

Interpretation: For plots received 800 mg N/kg soil and 1 cm of irrigation per week, we estimate that 52.0% of living material is blue stem on average one year after the plot is seeded with quack grass, with a margin of error of 4.45% at 95% confidence.

If Analyzing Like One-Sample Data...

Treatment	1N	1Y	2N	3N	4N	4Y
Mean $\bar{y}_{i\bullet}$	95	82.25	81.5	68.25	50.5	52
SD s_i	2.160	3.775	3.512	5.560	5.000	4.546

, MSE = 17.97

If using only the data in Group 4Y, the critical value would be $t_{\alpha/2, n_{4Y}-1} = t_{0.05/2, 4-1} \approx 3.182$.

```
> qt(0.05/2, df = 4-1, lower.tail=F)
[1] 3.182446
```

The naive 95% CI for μ_{4Y} would be

$$\bar{y}_{4Y\bullet} \pm t_{\alpha/2, n_{4Y}-1} \frac{s_{4Y}}{\sqrt{n_{4Y}}} \approx 52 \pm 3.182 \times \frac{4.546}{\sqrt{4}} \approx 52 \pm 7.23.$$

Observe that the margin of error 7.23% for the naive CI is bigger than the margin of error 4.45% for the CI in the previous page.

Confidence Intervals for Pairwise Differences

The $100(1-\alpha)\%$ confidence interval (C.I.) for $\mu_i - \mu_k$ is

$$\overline{y}_{i\bullet} - \overline{y}_{k\bullet} \pm t_{\alpha/2, N-g}\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}.$$

Note this is NOT a two-sample CI assuming equal SDs

$$\overline{y}_{i\bullet} - \overline{y}_{k\bullet} \pm t_{\alpha/2, n_i+n_k-2}\sqrt{s_p^2\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}, \text{ where } s_p^2 = \frac{(n_i-1)s_i^2 + (n_k-1)s_k^2}{n_i + n_k - 2},$$

NOR the two-sample CI not assuming equal SDs

$$\overline{y}_{i\bullet} - \overline{y}_{k\bullet} \pm t_{\alpha/2, df}\sqrt{\frac{s_i^2}{n_i} + \frac{s_k^2}{n_k}}$$

- MSE calculated using the entire dataset is a more accurate estimator of σ^2 than s_p^2 or s_i^2, s_k^2 calculated using only data in the two groups compared

- The critical value for the two-sample C.I. is larger
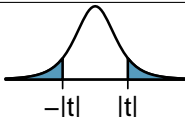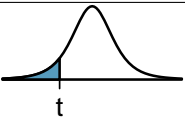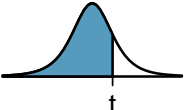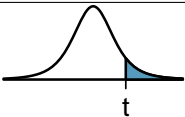
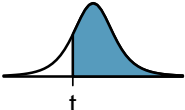$$t_{\alpha/2, n_i+n_k-2} > t_{\alpha/2, N-g}$$

Hypothesis Testing for Difference

For testing the hypothesis H_0: $\mu_i - \mu_k = 0$, the test statistic is

$$t_0 = \frac{\overline{y}_{i\bullet} - \overline{y}_{k\bullet}}{\text{SE}(\overline{y}_{i\bullet} - \overline{y}_{k\bullet})} = \frac{\overline{y}_{i\bullet} - \overline{y}_{k\bullet}}{\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}}$$

which has a $t-$distribution with d.f. $= N - g$.

The calculation of the p-value depends on H_a as follows

H_a	$\mu_i - \mu_k \neq 0$	$\mu_i - \mu_k < 0$	$\mu_i - \mu_k > 0$
p-value			

The bell curve above is the t-curve with df $= N - g$.

Case Study: Grass/Weed Competition

Treatment	1N	1Y	2N	3N	4N	4Y
Mean $\overline{y}_{i\bullet}$	95	82.25	81.5	68.25	50.5	52
SD s_i	2.160	3.775	3.512	5.560	5.000	4.546

, MSE $= 17.97$

A 95% confidence interval for $\mu_{1N} - \mu_{1Y}$ is

$$\overline{y}_{1N\bullet} - \overline{y}_{1Y\bullet} \pm t_{0.025,18} \times \sqrt{\mathsf{MSE}\left(\frac{1}{n_{1N}} + \frac{1}{n_{1Y}}\right)}$$

$$= 95 - 82.25 \pm 2.101 \times \sqrt{17.97\left(\frac{1}{4} + \frac{1}{4}\right)} = 12.75 \pm 6.65$$

in which $t_{0.025,18} = 2.101$ is found using the R command

```
> qt(0.05/2, df = 18, lower.tail=F)
[1] 2.100922
```

With irrigation, we see the percentage of bluestem is reduced by 12.75% on average, with a margin of error of 6.65% at 95% level.

Case Study: Grass/Weed Competition

If one wants to test whether treatment 1N and treatment 1Y have the same effect

$$H_0 : \mu_{1N} - \mu_{1Y} = 0 \quad \text{v.s.} \quad H_a : \mu_{1N} - \mu_{1Y} \neq 0$$

the test statistic is

$$t = \frac{\overline{y}_{1N\bullet} - \overline{y}_{1Y\bullet}}{\sqrt{\text{MSE}\left(\frac{1}{n_{1N}} + \frac{1}{n_{1Y}}\right)}} = \frac{95 - 82.25}{\sqrt{17.97\left(\frac{1}{4} + \frac{1}{4}\right)}} \approx \frac{12.75}{2.9975} \approx 4.253$$

with df $= N - g = 24 - 6 = 18$. The two-sided P-value is

```
> 2*pt(4.235, df = 18, lower.tail=F)
[1] 0.0004979698
```

As the P-value < 0.05, we again confirm that irrigation makes bluestem less competitive.

Pairwise *t*-Tests in R

The R command `pairwise.t.test` can perform pairwise comparisons between all pairs of treatments, but it shows the *P*-values only.

```
> pairwise.t.test(grass$percent, grass$trt, p.adjust="none")

Pairwise comparisons using t tests with pooled SD
data:  grass$percent and grass$trt
    1N      1Y      2N      3N      4N
1Y 0.00048 -       -       -       -
2N 0.00027 0.80527 -       -       -
3N 5.0e-08 0.00019 0.00033 -       -
4N 1.5e-11 3.7e-09 5.3e-09 1.3e-05 -
4Y 2.7e-11 7.8e-09 1.1e-08 3.8e-05 0.62287

P value adjustment method: none
```

Note that we must include `p.adjust="none"` in the command. Otherwise the *P*-value is not calculated using *t*-tests.

Underline Diagrams (p.88, Section 5.4.1)

a concise way to summarize pairwise comparisons

```
        1N        1Y        2N        3N        4N
1Y  0.00048  -         -         -         -
2N  0.00027  0.80527   -         -         -
3N  5.0e-08  0.00019   0.00033   -         -
4N  1.5e-11  3.7e-09   5.3e-09   1.3e-05   -
4Y  2.7e-11  7.8e-09   1.1e-08   3.8e-05   0.62287
```

How to make a **underline diagram**?

1. Write out treatment labels horizontally in **increasing order sorted by treatment means**
2. (Write the group mean $\overline{y}_{i\bullet}$ under each corresponding treatment label) (may skip)
3. Draw a line segment under a group of treatments if no pair of treatments in that group is significantly different.

4Y	4N	3N	2N	1Y	1N
50.5	52	68.25	81.5	82.25	95

Underline Diagrams

In an experiment with 5 treatments A, B, C, D and E, the underline diagram for all pairwise comparisons of the 5 treatments is as follows.

Answer the following questions:

▶ Order the means of the 5 groups from low to high.

$$C < B < A < D < E$$

▶ Check all the pairs that are significantly different from each other.

Least Significant Difference (LSD)

- It's lots of work to to compare all pairs of treatments. One needs to compute the SE, the t-statistic, and P-value for each pair of treatments. When there g treatments, there are $\binom{g}{2} = g(g-1)/2$ pairs to compare with.

- When all groups are of the same size n, an easier way to do pairwise comparisons of all treatments is to compute the **least significant difference** (LSD), which is the minimum amount by which two means must differ in order to be considered statistically different.

Least Significant Difference (LSD)

- When all groups are of the same size n, the SEs of pairwise comparisons all equal to

$$\text{SE} = \sqrt{\text{MSE}\left(\frac{1}{n} + \frac{1}{n}\right)}$$

- To be significant at level α, the t-statistic for pairwise comparison

$$t = \frac{\overline{y}_{j\bullet} - \overline{y}_{i\bullet}}{\text{SE}}$$

  must be at least $t_{\alpha/2, N-g}$ in absolute value

- So treatment i and j are significantly different at level α if and only if their difference in mean $\overline{y}_{j\bullet} - \overline{y}_{i\bullet}$ is at least $t_{\alpha/2, N-g}\sqrt{\text{MSE}\left(\frac{1}{n} + \frac{1}{n}\right)}$ in absolute value. So

$$\text{LSD} = t_{\alpha/2, N-g}\sqrt{\text{MSE}\left(\frac{1}{n} + \frac{1}{n}\right)}.$$

Least Significant Difference (LSD)

For the Grass/Weed experiment, the critical value for t% significance is $t_{\alpha/2, N-g} = t_{0.025, 24-6} \approx 2.101$, the LSD at 5% level is

$$\text{LSD} = t_{\alpha/2, N-g} \sqrt{\text{MSE}\left(\frac{1}{n} + \frac{1}{n}\right)} = 2.101 \sqrt{17.97\left(\frac{1}{4} + \frac{1}{4}\right)} \approx 6.30$$

Two treatments are significantly different at 5% level if and only if their mean differ by 6.30 or more.

The only two pairs with no significantly difference are (4Y, 4N) and (2N, 1Y), as they are the only pairs differ less than 6.30 in mean.

4Y	4N	3N	2N	1Y	1N
50.5	52	68.25	81.5	82.25	95

Questions of Interest Other Than Pairwise Comparisons

For the Grass/Weed Competition experiment, we are interested in

- Irrigation effect: $\mu_{1N} - \mu_{1Y}$ or $\mu_{4N} - \mu_{4Y}$ or the combination

$$\frac{\mu_{1Y} + \mu_{4Y}}{2} - \frac{\mu_{1N} + \mu_{4N}}{2}$$

- Nitrogen effect: $\mu_{1N} - \mu_{2N}$, $\mu_{2N} - \mu_{3N}$, etc.

- Does the irrigation effect change with nitrogen levels?

$$\underbrace{(\mu_{1N} - \mu_{1Y})}_{\substack{\text{irrigation effect when} \\ \text{nitrogen level} = 1}} \quad - \quad \underbrace{(\mu_{4N} - \mu_{4Y})}_{\substack{\text{irrigation effect when} \\ \text{nitrogen level} = 4}}$$

- Is the nitrogen effect **linear**?

$$\underbrace{(\mu_{1N} - \mu_{2N})}_{\substack{\text{slope when nitrogen level} \\ \text{changes from 1 to 2}}} \quad - \quad \underbrace{(\mu_{2N} - \mu_{3N})}_{\substack{\text{slope when nitrogen level} \\ \text{changes from 2 to 3}}}$$

Contrasts

All the quantities above are **contrasts**.

A **contrast** is a linear combination of group means μ_i's

$$C = \sum_{i=1}^{g} \omega_i \mu_i$$

where the ω's are known coefficients that add up to 0 $\sum_{i=1}^{g} \omega_i = 0$.

- **Every pairwise comparison is a contrast!** ($C = \mu_i - \mu_k$)
  $\omega_i = 1$, $\omega_k = -1$, all other ω_i's are 0, and
  $\sum_{i=1}^{g} \omega_i = 1 + (-1) + 0 + \cdots + 0 = 0$

- **A single treatment mean $C = \mu_i$ is NOT a contrast**

- Is $C = \dfrac{\mu_1 + \mu_2}{2} - \dfrac{\mu_3 + \mu_4 + \mu_5}{3}$ a contrast? Yes.
  $\omega_1 = \omega_2 = \frac{1}{2}$, $\omega_3 = \omega_4 = \omega_5 = -\frac{1}{3}$, which adds up to 0.

- Is $C = \dfrac{\mu_1 + \mu_2}{2} - \dfrac{\mu_3 + \mu_4}{2} + \mu_5$ a contrast? No.
  ($\omega_1 = \omega_2 = \frac{1}{2}$, $\omega_3 = \omega_4 = -\frac{1}{2}$, $\omega_5 = 1$, which add up to 1, not 0.

Estimator and Confidence Interval for a Contrast

A natural estimator for a contrast $C = \sum_{i=1}^{g} \omega_i \mu_i$ is

$$\widehat{C} = \sum_{i=1}^{g} \omega_i \overline{y}_{i\bullet}$$

As $\overline{y}_{1\bullet}, \overline{y}_{2\bullet}, \ldots,$ and $\overline{y}_{g\bullet}$ are indep. of each other, we know

$$\mathrm{Var}\Big(\sum_{i=1}^{g} \omega_i \overline{y}_{i\bullet} \Big) = \sum_{i=1}^{g} \mathrm{Var}(\omega_i \overline{y}_{i\bullet}) = \sum_{i=1}^{g} \omega_i^2 \mathrm{Var}(\overline{y}_{i\bullet}) = \sum_{i=1}^{g} \omega_i^2 \frac{\sigma^2}{n_i}.$$

The estimator $\widehat{C}$ has standard deviation and standard error

$$\mathrm{SD}(\widehat{C}) = \sqrt{\sigma^2 \sum_{i=1}^{g} \frac{\omega_i^2}{n_i}}\,, \qquad \mathrm{SE}(\widehat{C}) = \sqrt{\mathrm{MSE} \times \sum_{i=1}^{g} \frac{\omega_i^2}{n_i}}$$

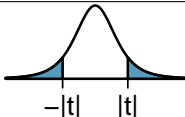A $(1 - \alpha)100\%$ confidence interval for the contrast C is
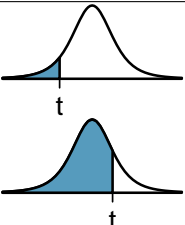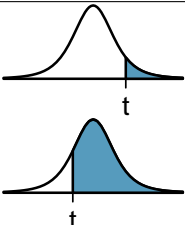
$$\widehat{C} \pm t_{\alpha/2, N-g} \times \mathrm{SE}(\widehat{C})$$

Hypothesis Testing for a Contrast

To test whether a contrast C is 0, $H_0 : C = 0$, the test statistic is

$$t = \frac{\widehat{C}}{\text{SE}(\widehat{C})} = \frac{\sum_{i=1}^{g} \omega_i \overline{y}_{i\bullet}}{\sqrt{\text{MSE} \times \sum_{i=1}^{g} \frac{\omega_i^2}{n_i}}} \sim t_{N-g}$$

The calculation of the p-value depends on H_a as follows

H_a	$C \neq 0$	$C < 0$	$C > 0$
p-value			

The bell curve above is the t-curve with df $= N - g$.

Does the Irrigation Effect Change with Nitrogen Levels?

Treatment	1N	1Y	2N	3N	4N	4Y
Mean $\bar{y}_{i\bullet}$	95	82.25	81.5	68.25	50.5	52

$\text{MSE} = 17.97$

The contrast we consider is

$$C = \underbrace{(\mu_{1N} - \mu_{1Y})}_{\substack{\text{irrigation effect when} \\ \text{nitrogen level} = 1}} - \underbrace{(\mu_{4N} - \mu_{4Y})}_{\substack{\text{irrigation effect when} \\ \text{nitrogen level} = 4}}$$

in which $(\omega_{1N}, \omega_{1Y}, \omega_{2N}, \omega_{3N}, \omega_{4N}, \omega_{4Y}) = (1, -1, 0, 0, -1, 1)$.
The contrast is estimated by

$$\widehat{C} = \bar{y}_{1N\bullet} - \bar{y}_{1Y\bullet} - (\bar{y}_{4N\bullet} - \bar{y}_{4Y\bullet}) = 95 - 82.25 - (50.5 - 52) = 14.25.$$

with the standard error

$$\text{SE}(\widehat{C}) = \sqrt{\text{MSE} \sum_{i=1}^{g} \frac{\omega_i^2}{n_i}} = \sqrt{17.97 \left(\frac{1^2}{4} + \frac{(-1)^2}{4} + \frac{(-1)^2}{4} + \frac{1^2}{4} \right)} \approx 4.24$$

Does the Irrigation Effect Change with Nitrogen Levels?

To test whether the irrigation effect changes with nitrogen level
H_0: $C = 0$ v.s. H_a: $C \neq 0$, the t-statistic is

$$t = \frac{\widehat{C}}{\mathsf{SE}(\widehat{C})} = \frac{14.25}{4.24} \approx 3.36$$

with $24 - 6 = 18$ degrees of freedom. The two-sided p-value is

```
> 2*pt(3.36,df=18, lower.tail=F)
[1] 0.003486951
```

The small P-value indicates the irrigation effects are significantly
different at the nitrogen level 1 and 4.

Does the Irrigation Effect Change with Nitrogen Levels?

The 95% confidence interval for C is

$$\widehat{C} \pm t_{0.025, N-g} \times \mathsf{SE}(\widehat{C}) \approx 14.25 \pm 2.101 \times 4.24 \approx (5.34, 23.16)$$

in which $t_{0.025, 24-6} \approx$ is found in R via the command

```
> qt(0.025,df=18, lower.tail=F)
[1] 2.100922
```

This means that the irrigation effect (% of bluestem with irrigation − without irrigation) is on average 5.34% to 23.16% higher at nitrogen level 1 than at level 4, with 95% confidence.

Is the Nitrogen Effect Linear?

Treatment	1N	1Y	2N	3N	4N	4Y
Mean $\overline{y}_{i\bullet}$	95	82.25	81.5	68.25	50.5	52

, MSE = 17.97

The contrast we consider is

$$C = (\mu_{1N} - \mu_{2N}) - (\mu_{2N} - \mu_{3N})$$
$$= \mu_{1N} - 2\mu_{2N} + \mu_{3N}$$

with the coefficients $(\omega_{1N}, \omega_{2N}, \omega_{3N}) = (1, -2, 1)$.
The contrast is estimated by

$$\widehat{C} = \overline{y}_{1N\bullet} - 2\overline{y}_{2N\bullet} + \overline{y}_{3N\bullet} = 95 - 2 \times 81.5 + 68.25 = 0.25.$$

with the standard error

$$\text{SE}(\widehat{C}) = \sqrt{\text{MSE} \sum_{i=1}^{g} \frac{\omega_i^2}{n_i}} = \sqrt{17.97 \Big(\frac{1^2}{4} + \frac{(-2)^2}{4} + \frac{1^2}{4} \Big)} \approx 5.19$$

Is the Nitrogen Effect Linear?

To test whether the nitrogen effect is linear H_0: $C = 0$ v.s. H_a: $C \neq 0$, the t-statistic is

$$t = \frac{\widehat{C}}{\text{SE}(\widehat{C})} = \frac{0.25}{5.19} \approx 0.048$$

with $24 - 6 = 18$ degrees of freedom. The two-sided p-value is

```
> 2*pt(0.048,df=18, lower.tail=F)
[1] 0.9622448
```

Conclusion: No strong evidence of nonlinear nitrogen effect (at level 1,2, and 3).

Remark: One can also test the linearity at level 2, 3, and 4

$$C = (\mu_{2N} - \mu_{3N}) - (\mu_{3N} - \mu_{4N}) = \mu_{2N} - 2\mu_{3N} + \mu_{4N}.$$

which is left as an exercise.