CS 4215: Quantitative Performance Evaluation for Computing ystems

Lydia Y. Chen
y.chen-10@tudelft.nl

Comparison questions?

Q1. You are looking for a computing cloud to host your websites so that the average latency is minimized. Which cloud provider? What kind of VMs? How many VMs?

Q2. You design a (machine learning) algorithm, which has many parameters. How do you set the parameters such that the computing time and the algorithm accuracy are maximized?

Learning objective

You will be able to design experiments which can capture the effects and their integrations

You will be able to analyze the experimental results and assess its significance

“

Design experiment is about

designing a proper set of experiments for measurement or simulation

developing a model that best describes the data obtained

analyzing the goodness of mode vis errors and variances

Last week

Characterization
Modeling
Predicting
Resource management

Terms and Definition

Terminology

- **Response variable**: outcome, e .g., response time
- **Factors (predictors)**: variables that affect the response variable, e.g., VM type, and parameter.
- **Levels (treatment)**: The values that a factor can assume, e.g., number of cores per VM
- **Replication**: Repetition of all or some experiments.
- **Design**: The number of experiments, the factor level and number of replications for each experiment.

  E.g., Full Factorial Design with 5 replications: $3\times 3 \times 4 \times 3 \times 3$ or 324 experiments, each repeated five times.

Types of Experimental Designs

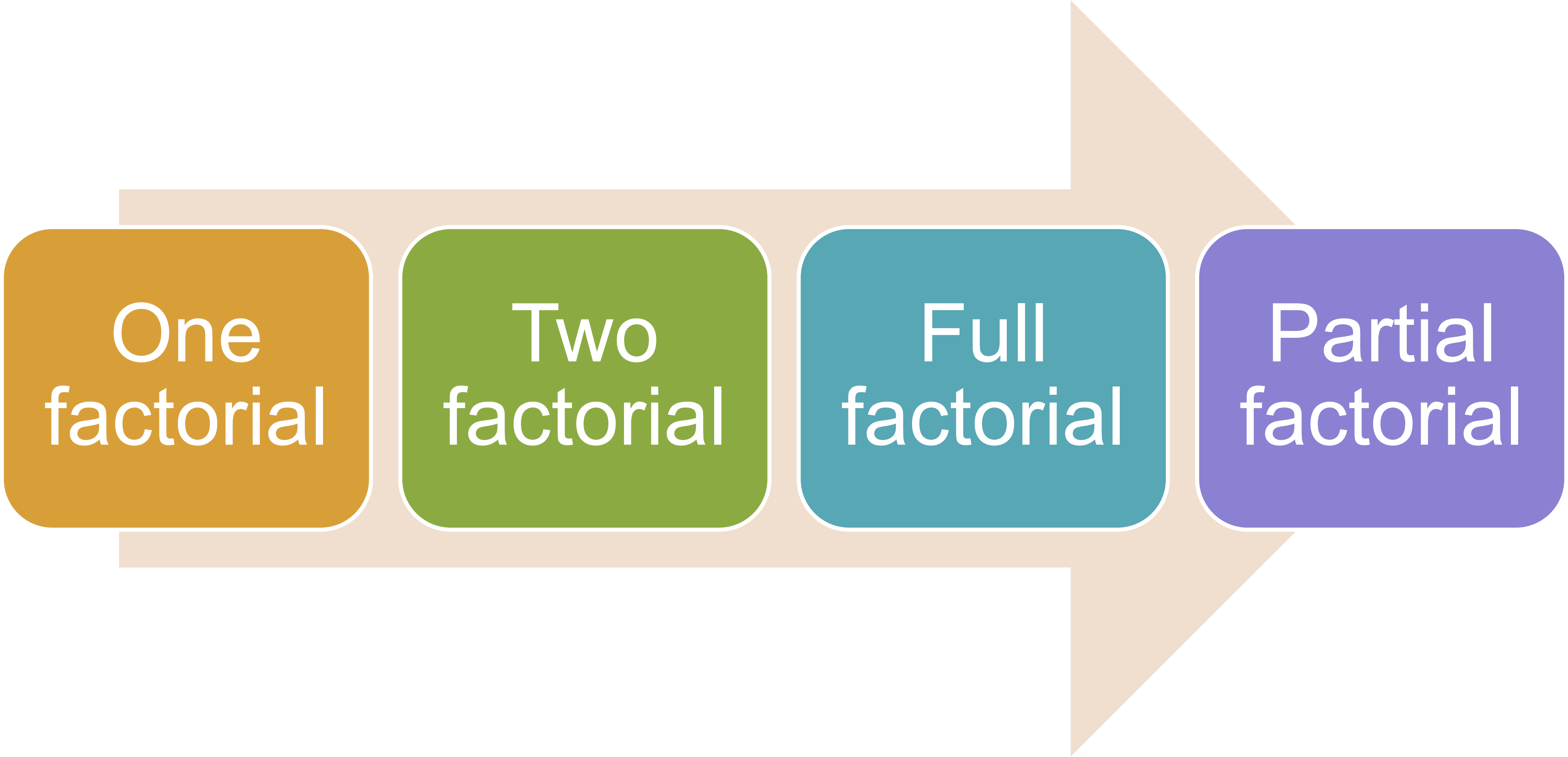
- **Simple Designs**: Vary one factor at a time
  - Not statistically efficient.
  - Wrong conclusions if the factors have interaction.

Not recommended

- **Full Factorial Design**: All combinations.
  - Can find the effect of all factors.
  - Too much time and money.
  - E.g., 2 factorial where each of n factors has 2 levels. # of experiments = 2^n/
- **Fractional Factorial Designs**: Less than Full Factorial
  - Save time and expense.
  - Less information.
  - May not get all interactions.
  - Not a problem if negligible interactions

Replication is important for all experiments!!!!

Design path


One factorial
Two factorial
Full factorial
Partial factorial

Analysis of results

- Goodness of fit:
  - Errors
  - Variation from factor
- Analysis of variation (ANOVA) and F-test
  - Importance $\neq$ Significance
  - Important $\Rightarrow$ Explains a high percent of variation
  - Significance $\Rightarrow$ High contribution to the variation compared to that by errors.
- Confidence interval of estimated effect

One Factorial Design

Examples of one factor experiment:

An algorithm has three versions: R, V, Z. Which version gives the lowest response time? What's their average performance?

➔ Run each version 5 times?

R	V	Z
144	101	130
120	144	180
176	211	141
288	288	374
144	72	302

Intuition

	R	V	Z	
	144	101	130	
	120	144	180	
	176	211	141	
	288	288	374	
	144	72	302	
Col Sum	$\Sigma\ y_{.1}$ =872	$\Sigma\ y_{.2}$ =816	$\Sigma\ y_{.3}$ =1127	$\Sigma\ y_{..}$ = 2815
Col Mean	$\bar{y}_{.1}$ =174.4	$\bar{y}_{.2}$ =163.2	$\bar{y}_{.3}$ =225.4	$\mu=\bar{y}_{..}$=187.7
Col Effect	$\alpha_1=\bar{y}_{.1}-\bar{y}_{..}$ =-13.3	$\alpha_2=\bar{y}_{.2}-\bar{y}_{..}$ =-24.5	$\alpha_3=\bar{y}_{.3}-\bar{y}_{..}$ =37.7	

$y_{i,j}$	Response of I replication of k level
$\bar{y}_{.j}$	Average response of j alternative replications
$\mu = \bar{y}_{..}$	Average response of all level experiments
$\boldsymbol{\alpha_j = \bar{y}_{.j} - \bar{y}_{..}}$	**Effect of of alternative j**
$\boldsymbol{\varepsilon_{ij} = y_{i,j} - \bar{y}_{..}}$	**Error term**

-Average algorithm requires 187.7 second of computing time.

-The effects of the R, V, and Z are -13.3, -24.5, and 37.7, respectively.

That is,
R/V/S requires -13.3/-24.5/-37.7 seconds more than an average algorithm.x

Models

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

$$\sum \alpha_j = 0$$

How to get those model parameters from "y"

$$\sum_{i=1}^{r} \sum_{j=1}^{a} y_{ij} =$$

$$=$$

$$\mu =$$

$$\bar{y}_{.j}$$

$$\alpha_j =$$

Analyzing the model:

- Allocation of variance (ANOVA):testing model accuracy and confidence
- Allocation variations to errors
  - Importance $\neq$ Significance
  - Important $\Rightarrow$ Explains a high percent of variation
  - Significance $\Rightarrow$ High contribution to the variation compared to that by errors.
- The ratio between some variances follows F-distribution.

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

Errors and variation

$$y_{ij}^2 = \mu^2 + \alpha_j^2 + e_{ij}^2 + 2\mu\alpha_j + 2\mu e_{ij} + 2\alpha_j e_{ij}$$

Total variation of y (SST) = $\sum_{i,j}\left(y_{i,j} - \bar{y}_{..}\right)^2$

=? $ar\mu^2$ =? $\mathrm{r}\sum_j \alpha_j{}^2$

$$\sum_{i,j} y_{ij}^2 = \boxed{\sum_{i,j} \mu^2} + \boxed{\sum_{i,j} \alpha_j^2} + \boxed{\sum_{i,j} e_{ij}^2}$$

$$+\text{Cross product terms}$$

goes to 0, Why?

Sum of Squared Y = Sum of Squared μ + Sum of Squared α + Sum of Squared errors e

➔ SSY =SS0+SSA+SSE

➔ Total of variance : SST= SSY-SS0=SSA+SSE

Back to examples

- SST=105357.3 , SSA=10992.13, SSE=94365.2
- What is the percentage of variation explained by the algorithms?

$$\text{Percent variation explained by processors} = 100 \times \frac{10992.13}{105357.3} = 10.4\%$$

- Is this number statistical significant?

Analysis of Variance (ANOVA)

- Degree of freedom = Number of independent values required to compute (Additive)

SSY	= SS0	+SSA	+SSE
ar	1	a-1	a(r-1)

- F-test
  - Purpose: To check if SSA is *significantly* greater than SSE
  - Errors are normally distributed ⇒ SSE and SSA have chi-square distributions.

$$\boxed{\frac{SSA/\nu_A}{SSE/\nu_e} \sim F\ distribution}$$

where ν_A=a-1 = degrees of freedom for SSA, ν_e=a(r-1) = degrees of freedom for SSE

  - Computed ratio > $F_{[1-\alpha;\ \nu_A,\ \nu_e]}$ ⇒ SSA is significantly higher than SSE.

ANOVA Table for One Factor Experiments:

Standard output

Component	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	F-Table
y	SSY=$\sum y_{ij}^2$		ar			
$\bar{y}_{..}$	SS0=$ar\mu^2$		1			
y-$\bar{y}_{..}$	SST=SSY-SS0	100	ar-1			
A	SSA $= r\Sigma\ \alpha_i^2$	$100\left(\frac{\text{SSA}}{\text{SST}}\right)$	a-1	MSA $= \frac{\text{SSA}}{a-1}$	$\frac{\text{MSA}}{\text{MSE}}$	$F_{[1-\alpha; a-1, a(r-1)]}$
e	SSE=SST- SSA	$100\left(\frac{\text{SSE}}{\text{SST}}\right)$	$a(r-1)$	MSE $= \frac{\text{SSE}}{a(r-1)}$		

Back to algorithm example

Component	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	F-Table
y	633639.00					
$y_{..}$	528281.69					
y-$y_{..}$	105357.31	100.0%	14			
A	10992.13	10.4%	2	5496.1	0.7	2.8
Errors	94365.20	89.6%	12	7863.8		

$s_e = \sqrt{\text{MSE}} = \sqrt{7863.77} = 88.68$

Is difference among algorithms significant?

Confidence of parameter

$$Parameter\ estimate \sim T\ distribution$$

Why T distribution?

What is the degree of freedom?

Parameter	Estimate	Variance
μ	$\bar{y}_{..}$	s_e^2/ar
α_j	$\bar{y}_{.j}$-$\bar{y}_{..}$	$s_e^2(a-1)/ar$
μ+α_j	$\bar{y}_{.j}$	s_e^2/r
$\sum_{j=1}^{a} h_j\alpha_j,\ \sum_{j=1}^{a} h_j$=0	$\sum_{j=1}^{a} h_j\ \bar{y}_{.j}$	$\sum_{j=1}^{a} s_e^2\ h_j^2/r$
s_e^2	$\frac{\sum e_{ij}^2}{a(r-1)}$	

Degrees of freedom for errors = a(r-1)

Example of algorithms comparison

Error variance $s_e^2 = \frac{94365.2}{12} = 7863.8$

$$\begin{aligned}\text{Std Dev of errors} &= \sqrt{(\text{Var. of errors})}\\ &= 88.7\end{aligned}$$

Std Dev of $\mu = s_e/\sqrt{ar} = 88.7/\sqrt{15} = 22.9$

$$\begin{aligned}\text{Std Dev of } \alpha_j &= s_e\sqrt{\{(a-1)/(ar)\}}\\ &= 88.7\sqrt{(2/15)} = 32.4\end{aligned}$$

How to interpret these results?
Think about where 0 is

$$\begin{aligned}\mu &= 197.7 \mp (1.782)(22.9) = (146.9, 228.5)\\ \alpha_1 &= -13.3 \mp (1.782)(32.4) = (-71.0, 44.4)\\ \alpha_2 &= -24.5 \mp (1.782)(32.4) = (-82.2, 33.2)\\ \alpha_3 &= 37.6 \mp (1.782)(32.4) = (-20.0, 95.4)\end{aligned}$$

For 90% confidence, $t_{[0.95;\,12]} = 1.782$.

Confidence Intervals For Effects

- Where to be careful?

	R	V	Z		
	144	101	130		
	120	144	180		
	176	211	141		
	288	288			
	144				
Column Sum	872	744	451	2067	
Column Mean	174.40	186.00	150.33		172.25
Column effect	2.15	13.75	-21.92		

Two Factorial Design

Examples of two factor experiment:

Understanding how code size is impacted by the workloads and processors. Run each combination 3 times.

Workloads	Processors W	X	Y	Z
I	7006	12042	29061	9903
	6593	11794	27045	9206
	7302	13074	30057	10035
J	3207	5123	8960	4153
	2883	5632	8064	4257
	3523	4608	9677	4065
K	4707	9407	19740	7089
	4935	8933	19345	6982
	4465	9964	21122	6678
L	5107	5613	22340	5356
	5508	5947	23102	5734
	4743	5161	21446	4965
W	6807	12243	28560	9803
	6392	11995	26846	9306
	7208	12974	30559	10233

Log Transformation

Workloads	Processors W	X	Y	Z
I	3.8455	4.0807	4.4633	3.9958
	3.8191	4.0717	4.4321	3.9641
	3.8634	4.1164	4.4779	4.0015
J	3.5061	3.7095	3.9523	3.6184
	3.4598	3.7507	3.9066	3.6291
	3.5469	3.6635	3.9857	3.6091
K	3.6727	3.9735	4.2953	3.8506
	3.6933	3.9510	4.2866	3.8440
	3.6498	3.9984	4.3247	3.8246
L	3.7082	3.7492	4.3491	3.7288
	3.7410	3.7743	4.3636	3.7585
	3.6761	3.7127	4.3313	3.6959
M	3.8330	4.0879	4.4558	3.9914
	3.8056	4.0790	4.4289	3.9688
	3.8578	4.1131	4.4851	4.0100

Model for two factorial

1. Model: With *r* replications

$$y_{ijk} = \mu + \boxed{\alpha_j} + \boxed{\beta_i} + \boxed{\gamma_{ij}} + e_{ijk}$$

α_j Effect of factor A
β_i Effect of factor B
γ_{ij} Effect of interaction A&B

2. How to estimate the parameters

$$\left\{\begin{array}{l} \sum_{j=1}^{a} \alpha_j = 0; \sum_{i=1}^{b} \beta_i = 0; \\ \sum_{j=1}^{a} \gamma_{1j} = \sum_{j=1}^{a} \gamma_{2j} = \cdots = \sum_{j=1}^{a} \gamma_{bj} = 0 \\ \sum_{i=1}^{b} \gamma_{i1} = \sum_{i=1}^{b} \gamma_{i2} = \cdots = \sum_{i=1}^{b} \gamma_{ia} = 0 \\ \sum_{k=1}^{r} e_{ijk} = 0 \quad \forall i, j \end{array}\right. \Rightarrow \left\{\begin{array}{c} \bar{y}_{ij.} = \mu + \alpha_j + \beta_i + \gamma_{ij} \\ \mu = \bar{y}_{...} \\ \alpha_j = \bar{y}_{.j.} - \bar{y}_{...} \\ \beta_i = \bar{y}_{i..} - \bar{y}_{...} \\ \gamma_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \end{array}\right.$$

Back to the example

Workloads	W	X	Y	Z	Row Sum	Row Mean	Row Effect
	Processors						
I	3.8427	4.0896	4.4578	3.9871	16.3772	4.0943	0.1520
J	3.5043	3.7079	3.9482	3.6188	14.7792	3.6948	-0.2475
K	3.6720	3.9743	4.3022	3.8397	15.7882	3.9470	0.0047
L	3.7084	3.7454	4.3480	3.7277	15.5295	3.8824	-0.0599
M	3.8321	4.0933	4.4566	3.9900	16.3720	4.0930	0.1507
Col Sum	18.5594	19.6105	21.5128	19.1635	78.8463		
Col Mean	3.7119	3.9221	4.3026	3.8327		3.9423	
Col effect	-0.2304	-0.0202	0.3603	-0.1096			

Interaction

Workloads	W	X	Y	Z
I	-0.0212	0.0155	0.0032	0.0024
J	0.0399	0.0333	-0.1069	0.0337
K	-0.0447	0.0475	-0.0051	0.0023
L	0.0564	-0.1168	0.1054	-0.0450
M	-0.0305	0.0205	0.0033	0.0066

How to interpret them?

- Processor W requires $10^{0.23}$ (=1.69) less code than avg processor.
- Processor X requires $10^{0.02}$ (=1.05) less than an average processor and so on.
- The ratio of code sizes of an average workload on processor W and X is $10^{0.21}$ (= 1.62)
- Workload I on processor W requires 0.02 less log code size than an average workload on processor W or equivalently 0.02 less log code size than I on an average processor

$$\hat{y}_{ij} = \mu + \alpha_j + \beta_i + \gamma_{ij} = \bar{y}_{ij.}$$

$$e_{ijk} = y_{ijk} - \bar{y}_{ij.}$$

Analyzing the model

Compare the variation ratio with F distribution

- $\frac{SSA/v_A}{SSE/v_e} \sim F\,[a-1, ab(r-1)\,]$
- $\frac{SSB/v_B}{SSE/v_e} \sim F\,[b-1, ab(r-1)\,]$
- $\frac{SSAB/v_{AB}}{SSE/v_e} \sim F\,[(a-1)(b-1), ab(r-1)\,]$

1. Decompose total of variance

From where? Derive SSY

$$\begin{array}{ccccccccccccc} \text{SST} & = & \text{SSY} & - & \text{SS0} & = & \text{SSA} & + & \text{SSB} & + & \text{SSAB} & + & \text{SSE} \\ 4.44 & = & 936.95 & - & 932.51 & = & 2.93 & + & 1.33 & + & 0.15 & + & 0.03 \\ 100\% & = & & & & = & 65.96\% & + & 29.9\% & + & 3.48\% & + & 0.66\% \end{array}$$

2. Think about degree of freedom

$$\begin{array}{ccccccccccc} \text{SSY} & = & \text{SS0} & + & \text{SSA} & + & \text{SSB} & + & \text{SSAB} & + & \text{SSE} \\ abr & = & 1 & + & (a-1) & + & (b-1) & + & (a-1)(b-1) & + & ab(r-1) \end{array}$$

ANOVA for Two Factors w Replications

Standard output

Component	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	F-Table
y	$\text{SSY} = \sum y_{ij}^2$		abr			
$\bar{y}_{\ldots}$	$\text{SS0} = abr\mu^2$		1			
$y - \bar{y}_{\ldots}$	$\text{SST} = \text{SSY} - \text{SS0}$	100	$abr - 1$			
A	$\text{SSA} = br\Sigma\alpha_j^2$	$100\left(\frac{\text{SSA}}{\text{SST}}\right)$	$a-1$	$\text{MSA} = \frac{\text{SSA}}{a-1}$	$\frac{\text{MSA}}{\text{MSE}}$	$F_{[1-\alpha;a-1,ab(r-1)]}$
B	$\text{SSB} = ar\Sigma\beta_i^2$	$100\left(\frac{\text{SSB}}{\text{SST}}\right)$	$b-1$	$\text{MSB} = \frac{\text{SSB}}{b-1}$	$\frac{\text{MSB}}{\text{MSE}}$	$F_{[1-\alpha;b-1,ab(r-1)]}$
AB	$\text{SSAB} = r\Sigma\gamma_{ij}^2$	$100\left(\frac{\text{SSAB}}{\text{SST}}\right)$	$(a-1)(b-1)$	$\text{MSAB} = \frac{\text{SSAB}}{(a-1)(b-1)}$	$\frac{\text{MSA}}{\text{MSE}}$	$F_{[1-\alpha,(a-1)(b-1),ab(r-1)]}$
e	$\text{SSE} = \text{SST} - (\text{SSA} + \text{SSB} + \text{SSAB})$	$100\left(\frac{\text{SSE}}{\text{SST}}\right)$	$ab(r-1)$	$\text{MSE} = \frac{\text{SSE}}{ab(r-1)}$		

Back to the example

Component	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	F-Table
y	936.95					
$\bar{y}_{\ldots}$	932.51					
$y - \bar{y}_{\ldots}$	4.44	100.00%	59			
Processors	2.93	65.96%	3	0.9765	1340.01	2.23
Workloads	1.33	29.90%	4	0.3320	455.65	2.09
Interactions	0.15	3.48%	12	0.0129	17.70	1.71
Errors	0.03	0.66%	40	0.0007		

$s_e = \sqrt{\text{MSE}} = \sqrt{0.0008} = 0.03$

Are differences significant?

Confidence Intervals For Effects

Parameter Estimation

Parameter	Estimate	Variance
μ	$\bar{y}_{...}$	s_e^2/abr
α_j	$\bar{y}_{i..}$-$\bar{y}_{...}$	$s_e^2(a-1)/abr$
β_i	$\bar{y}_{.j.}$-$\bar{y}_{...}$	$s_e^2(b-1)/abr$
γ_{ij}	$\bar{y}_{ij.}$-$\bar{y}_{i..}$-$\bar{y}_{.j.}$+$\bar{y}_{...}$	$s_e^2(a-1)(b-1)/abr$
$\Sigma\ h_j\ \alpha_j$, $\Sigma\ h_j$=0	$\Sigma\ h_j\ \bar{y}_{.j.}$	$\Sigma\ h_j^2 s_e^2/br$
$\Sigma\ h_i\ \beta_i$, $\Sigma\ h_i$=0	$\Sigma\ h_i\ \bar{y}_{i..}$	$\Sigma\ h_i^2 s_e^2/ar$
s_e^2	$\Sigma\ e_{ijk}^2/\{ab(r-1)\}$	

Degrees of freedom for errors = ab(r-1)

Parameter	Mean Effect	Std. Dev.	Confidence Interval
μ	3.9423	0.0035	(3.9364, 3.9482)
Processors			
W	-0.2304	0.0060	(-0.2406, -0.2203)
X	-0.0202	0.0060	(-0.0304, -0.0100)
Y	0.3603	0.0060	(0.3501, 0.3704)
Z	-0.1096	0.0060	(-0.1198, -0.0995)
Workloads			
I	0.1520	0.0070	(0.1402, 0.1637)
J	-0.2475	0.0070	(-0.2592, -0.2358)
K	0.0047	0.0070	(-0.0070, 0.0165)†
L	-0.0599	0.0070	(-0.0717, -0.0482)
M	0.1507	0.0070	(0.1390, 0.1624)

† ⇒ Not significant

- Use *t* values at *ab(r-1)* degrees of freedom for confidence intervals

Back to example

- From ANOVA table: s_e=0.03. The standard deviation of processor effects:

$$s_{\alpha_j} = \qquad = 0.0060$$

- The error degrees of freedom: *ab(r-1)* = 40 ⇒ use Normal tables
- For 90% confidence, $z_{0.95}$ = 1.645 90% confidence interval for the effect of processor W is:

- α_1 ¨ t s_{α_1} = -0.2304 +-1.645 * 0.0060
  = -0.2304 +- 0.00987
  = (-0.2406, -0.2203)

  The effect is significant.

Missing Observations

- Recommended Method:
  - Divide the sums by respective number of observations
  - Adjust the degrees of freedoms of sums of squares
  - Adjust formulas for standard deviations of effects
- Other Alternatives:
  - Replace the missing value by $\hat{y}$ such that the residual for the missing experiment is zero.
  - Use y such that SSE is minimum.

2^{k-p} Fractional Factorial Designs

2^{k-p} Fractional Factorial Designs

- Large number of factors
  - large number of experiments
  - full factorial design too expensive
  - Use a fractional factorial design
- 2^{k-p} design allows analyzing k factors with only 2^{k-p} experiments.
  - 2^{k-1} design requires only half as many experiments
  - 2^{k-2} design requires only one quarter of the experiments

Example: 2^{7-4} Design

Expt No.	A	B	C	D	E	F	G
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

- Study 7 factors with only 8 experiments!

Want big impact?

Use big image.

Thanks!

Any questions?

lydiaychen@ieee.org

Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash

Presentation design

This presentations uses the following typographies and colors:

- Titles: **Raleway**
- Body copy: **Lato**

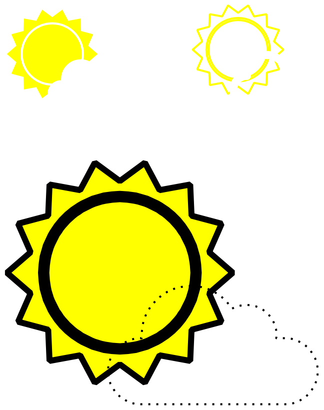
You can download the fonts on these pages:

https://www.fontsquirrel.com/fonts/raleway

https://www.fontsquirrel.com/fonts/lato

- Dark blue **#2185c5**
- Light blue **#7ecefd**
- Yellow **#ff9715**
- Magenta **#f20253**
- Dark gray **#677480**
- Light gray **#97abbc**

You don’t need to keep this slide in your presentation. It’s only here to serve you as a design guide if you need to create new slides or download the fonts to edit the presentation in PowerPoint®



Now you can use any emoji as an icon!
And of course it resizes without losing quality and you can change the color.

How? Follow Google instructions
https://twitter.com/googledocs/status/730087240156643328

and many more...