IN4390: Quantitative Performance Evaluation for Embedded Systems

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

Design of Experiments

Comparison Questions?

Q1. You are looking for a computing cloud to run your course project so that the average latency is minimized.? 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 the algorithm accuracy is maximized?

Learning Objectives

- 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, and
analyzing the goodness of mode via errors and variances.

Terms and Definitions

Terminology

Q: Why is replication important

Captures the variability.

- **Response Variable**: Outcome
- **Factors (Predictors)**: Variables that affect the ***response variable***
- **Levels (Treatment)**: The values that a ***factor*** can assume
- **Replication**: Repetition of all or some experiments
- **Design**: The number of experiments, the factor level and number of replications for each experiment

E.g., A ML algorithm can be on trained machine with 1-4 cores and two choices of RAM (2 and 4 GB). You want o replicate each experiment for 5 times. A total umber of experiments: 4*2*5

Replication is important for all experiments!!!!

Types of Experimental Designs

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

Q: Recommended?

Q: Drawback?

- **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.

Q: Advantages?

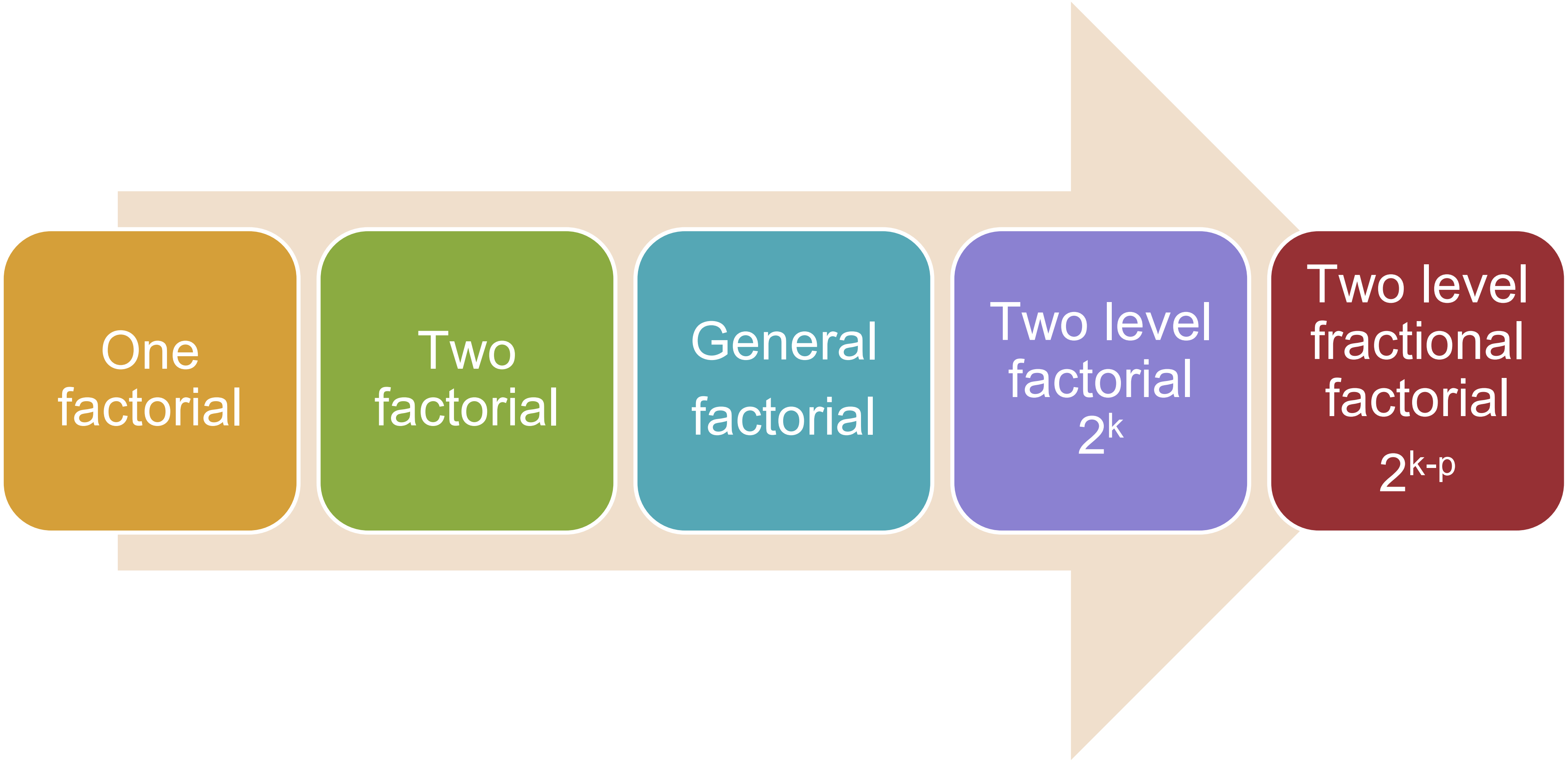
Q: Drawback?

- **Fractional factorial designs**: Less than full factorial design.
  - Saves time and expenses.
  - Less information.
  - May not get all interactions.
  - Not a problem if negligible interactions.

Q: Advantages?

Q: Drawback?

Designs You Will Learn …



1

Model +
Assumption

2

Estimate parameters from measurement

3

Goodness of fit:
Errors, ANOVA
Confidence interval

4

Check assumption

One Factorial Design

Examples of One Factor Experiment:

An algorithm has three versions: R, V, Z. Which **version** gives the **lowest** response time?

Run each version 5 times?

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

Q: Is there any difference between R,V,Z?

Q: How sure we are?

Intuition

Q: What is y_{ij} & $\bar{y}_{.j}$

	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 denotes responses

$i = \{1 \ldots r\}$, r is the number of replication

$j = \{1 \ldots a\}$, a is the number of levels

Note: "dot" subscript notation implies summation over the subscript that it replaces

$y_{i,j}$: Response of i replication for jth alternative.

$\bar{y}_{.j}$: Average response of jth alternative replications.

$\mu = \bar{y}_{..}$: Average response of all levels of experiments.

$\alpha_j = \bar{y}_{.j} - \bar{y}_{..}$: Effect of the alternative j.

$e_{ij} = y_{ij} - \bar{y}_{.j}$: Error term.

Q: How do we interpret the results

- **Average algorithm x** requires **187.7 seconds** 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**.

Model

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

$$\sum_j \alpha_j = 0$$

$$e_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Linear model

How to get those model parameters from "y" (measurement)?

Q: What does N stand for?

Normal Distribution.

Q: How to write y_{ij} when there are two levels and three replications?

$$y_{ij} = \mu + \alpha_1 + \alpha_2 + e_{i,j}, \quad j = \{1, 2\} \quad \text{and} \quad i = \{1, 2, 3\}$$

y_{ij} → **Estimator** → $\mu, \alpha_1, \ldots \alpha_a$

Measured Data → Model parameters

$$i = \{1 \ldots r\}, \; r \text{ is the number of replication}$$

$$j = \{1 \ldots a\}, \; a \text{ is the number of levels}$$

Model

Linear model

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

$$\sum_j \alpha_j = 0$$

$$e_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Estimator for model parameter

$$\hat{\mu} = \frac{1}{ar}\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij} = \bar{y}_{..}$$

$$\alpha_j = \bar{y}_{.j} - \mu = \bar{y}_{.j} - \bar{y}_{..}$$

Derivation for interested readers:

$$\begin{aligned}\sum_{i=1}^{r}\sum_{j=1}^{a} y_{ij} &= ar\mu + r\sum_{j=1}^{a}\alpha_j + \sum_{i=1}^{r}\sum_{j=1}^{a} e_{ij} \\ &= ar\mu + 0 + 0\end{aligned}$$

$$\begin{aligned}\bar{y}_{.j} &= \frac{1}{r}\sum_{i=1}^{r} y_{ij} \\ &= \frac{1}{r}\sum_{i=1}^{r}(\mu + \alpha_j + e_{ij}) \\ &= \frac{1}{r}\left(r\mu + r\alpha_j + \sum_{i=1}^{r} e_{ij}\right) \\ &= \mu + \alpha_j + 0\end{aligned}$$

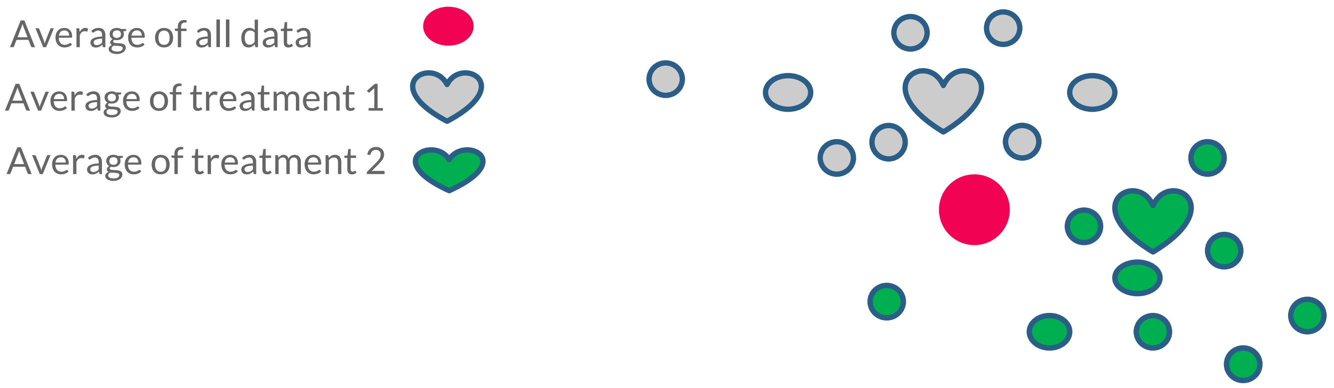
Note: Notations follow previous slides.

$$y_{ij} = \mu + \alpha_j + e_{ij}$$
$$\sum_j \alpha_j = 0$$
$$e_{ij} \sim \mathcal{N}(0, \sigma^2)$$

Analyzing Models

- Allocation of variance (ANOVA): Testing model accuracy and confidence (Test if $\alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$)
- Allocation variations to errors
  - Importance ≠ Significance
  - Important ⇒ Explains a high percent of variation
  - Significance ⇒ High contribution to the variation compared to that by errors.

Q: What is variation?



Total variation
Variation cross group
Variation within group

Q: Explanations

Decomposition of Variation

$$\sum_{ij}(y_{ij}-\bar{y}_{..})^2 = \sum_{ij} y_{ij}^2 - \sum_{ij}\bar{y}_{..}^2 = \sum_{ij}\alpha_j{}^2 + \sum_{ij} e_{ij}^2$$

SST=SSY-SS0=SSA+SSE

- Total variation of y (SST)

$$SST = \sum_{ij}(y_{ij}-\bar{y}_{..})^2$$

$$= \sum_{ij}(y_{ij}^2) - ar\bar{y}_{..}^2$$

Q: How

Model assumption

➔SST=SSY-SS0

- Sum of squared y (SSY)

Q: Why?

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

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

$=ar\mu^2$ $=r\sum_j \alpha_j{}^2$ goes to 0,

$$SSY = \sum_{ij} y_{ij}^2 = \sum_{i,j}\mu^2 + \sum_{ij}\alpha_{i,j}^2 + \sum_{i,j} e_{i,j}^2 + \text{cross product term}$$

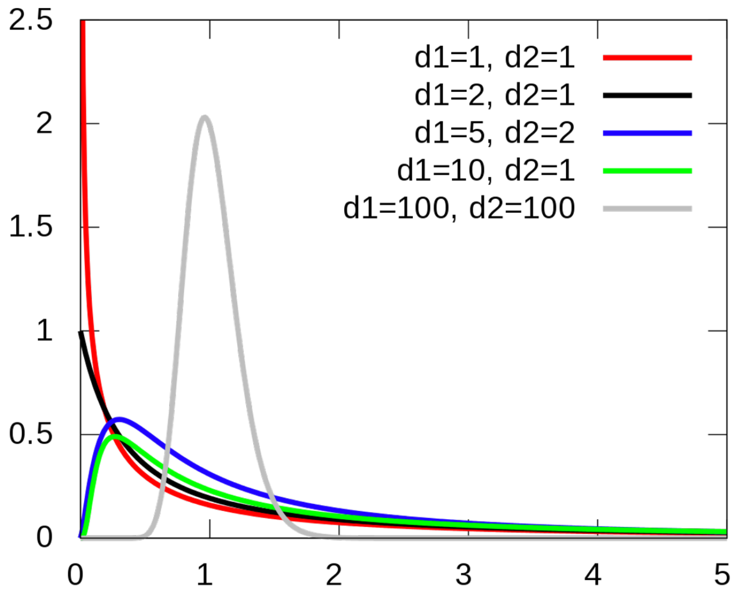
Sum of Squared Y = Sum of Squared μ + Sum of Squared α + Sum of Squared errors e

➔SSY=SS0+SSA+SSE

Analyzing Models

SST=SSY-SS0=SSA+SSE

- The ratio between some variances follows **F-distribution with df1 and df2 .**



Q: What are parameters of F-distribution

Df1 and df2

Q: Why F-distribution?
Any assumption?

Errors follow normal distribution

Back to Examples

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

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

$$\frac{SSA}{SST} = \frac{10992.13}{105357.3} = 10.4\%$$

Q: That is?

- Is this number statistical significant? Compare the variation against errors

Q: That is?

$$\frac{SSA}{SSE}$$

Q: Follows what distribution?

F-Distribution when divided by their degree of freedoms

ANOVA: ANalysis Of VAriance

Analysis of Variance (ANOVA)

- Key components in ANOVA: SSY, SS0, SSA, SSE.
- Their degree of freedom: Number of independent values required to compute (additive)

Source	SSY	SS0	SSA	SSE
Degree of freedom (v)	ar	1	a-1	a(r-1)

Think about model assumptions

Q: What does additive mean?

Q: What are the reasons of these degree of freedoms

SSY-SSO=SSA+SSE➔ $V_Y - V_0 = V_A + V_e$

Analysis of Variance (ANOVA)

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

$$\frac{SSA/v_A}{SSE/v_e} \sim F\ distribution$$

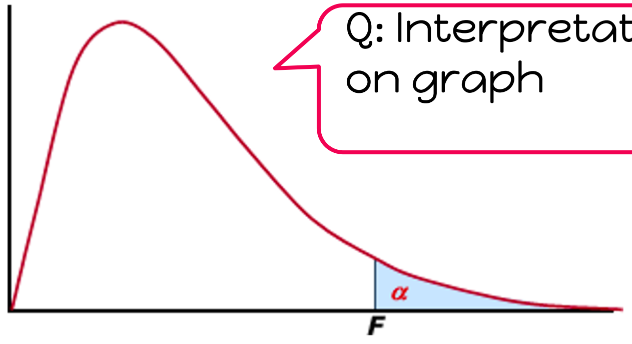
where v_A***=a-1 = degrees of freedom for SSA,*** v_e***=a(r-1) = degrees of freedom for SSE***

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

Q: What does that mean:
Stat < $F_{0.9,3,12}$

Q1: Good model?
Q2: Reject or accept the assumption?

It falls into the blue region, meaning the probability of find a value greater than the computer ratio is lower than alpha



Q: Interpretation on graph

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	$SSA = r\Sigma\ \alpha_i^2$	$100\left(\frac{SSA}{SST}\right)$	a-1	$MSA = \frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	$F_{[1-\alpha; a-1, a(r-1)]}$
e	SSE=SST- SSA	$100\left(\frac{SSE}{SST}\right)$	$a(r-1)$	$MSE = \frac{SSE}{a(r-1)}$		

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		

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

Q: Is difference among algorithms significant?

NO!

Goodness of Estimates

- ANOVA tells us goodness of overall models

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

- Goodness of estimates of model parameter
  - μ and α_j
  - Linear combination of α_j , e.g., $\alpha_1 - \alpha_2$

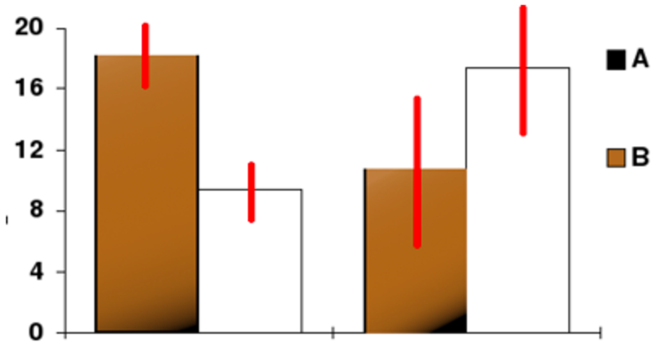
➔ Compute the Confidences Interval (CI)

➔ E.g., 95% CI of μ is [-2 2], meaning 95% chance the parameter is within the interval.

Q: Which CI tells us α_1 has impact on y
[-1 1] and [-2 -1]

Q: what CI can you get from this model?

Q: Smaller the better or bigger the better?



$\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_1 - \alpha_2, \alpha_2 - \alpha_3 \; \alpha_1 - \alpha_3, y$

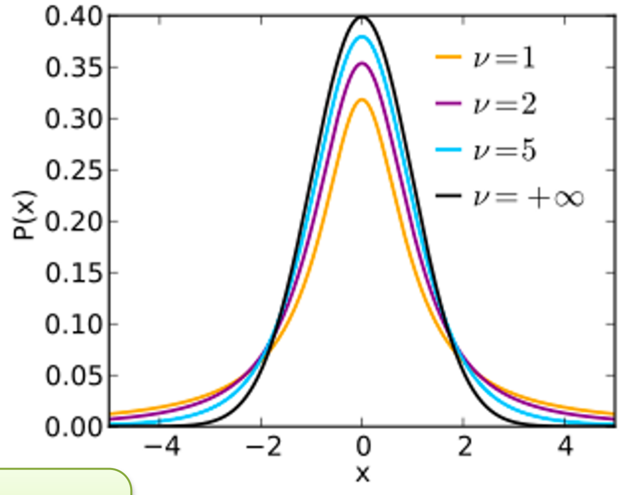
Confidence Interval of Model Parameters

$$\text{CI range of parameter} = \begin{cases} \text{estimate} + t_{1-\alpha/2,df} \cdot (\text{std of estimate of parameter}) \\ \text{estimate} - t_{1-\alpha/2,df} \cdot (\text{std of estimate of parameter}) \end{cases}$$

Q: What is $t_{\alpha/2,df}$?

Q: What is estimate?

Explaination I



Q: What is t distribution

Similar to normal distribution

Q: What is α

Significance level, but needs to check $\alpha/2$

Q: What is v here?

Degree of freedom

Confidence Interval of Model Parameters

$$\text{CI range of parameter} = \begin{cases} \text{estimate} + t_{\alpha/2,df} \cdot (\text{std of estimate of parameter}) \\ \text{estimate} - t_{\alpha/2,df} \cdot (\text{std of estimate of parameter}) \end{cases}$$

Explaination II

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

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

Q: how to derive it

Q: what is std of μ and α_j

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

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

Q: How to interpret those results

If the interval doesn't contain a zero then we know there is definite difference in the performance.

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

Q. Why .95, not .9?

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

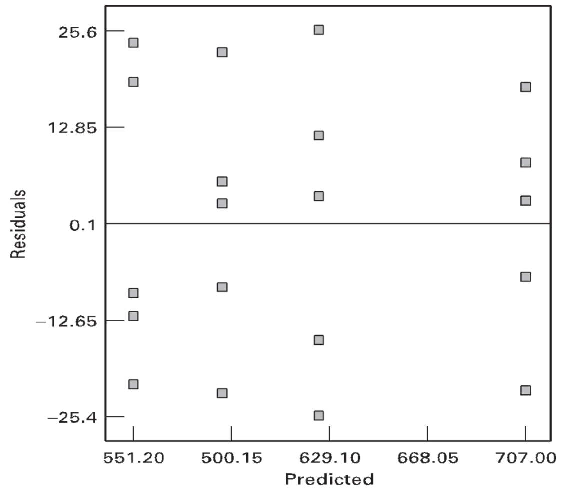
Since this is a two tailed test i.e we are looking at both sides of the curve for anomalies (5% on each side)

Checking Assumptions

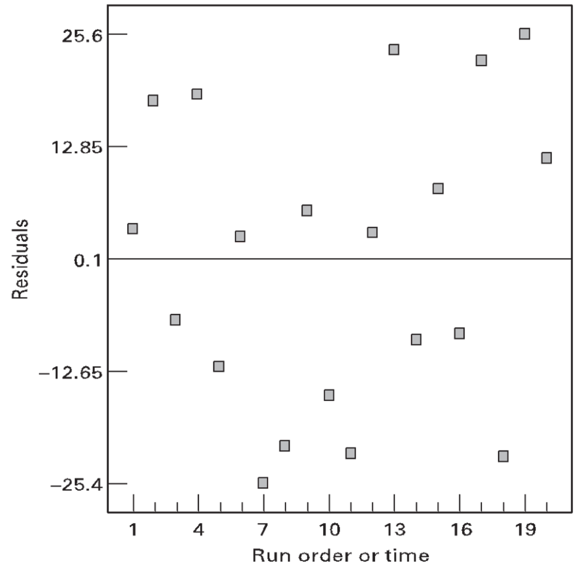
Q. What are they?

1. **Independent errors**

    a) Scatter plot of residuals versus the predicted response



    a) Plot the residuals as a function of the experiment number

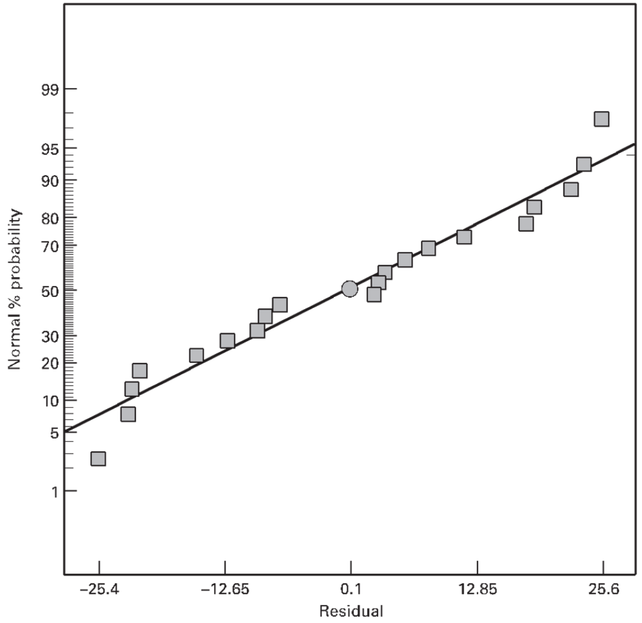


Trend up or down ⇒ other factors or side effects.

Checking Assumptions

2. **Normally distributed errors**:

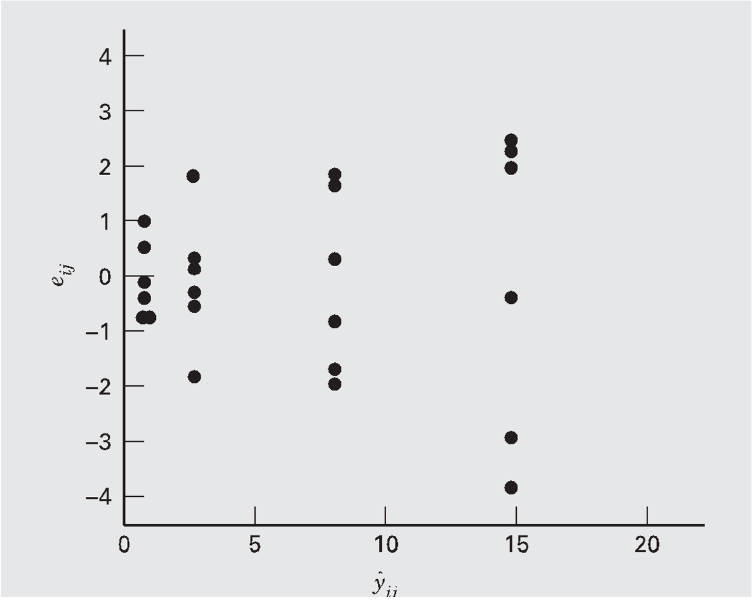
Normal quantile-quantile plot of errors



3. **Constant standard deviation of errors**:

Scatter plot of y for various levels of the factor

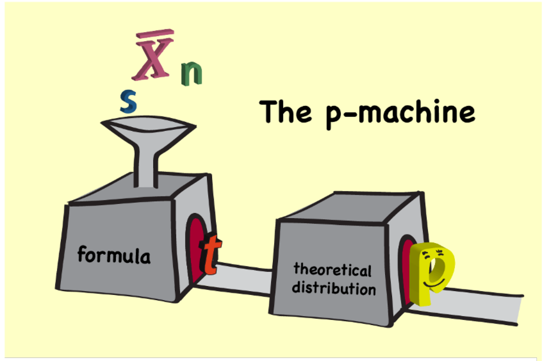
Spread at one level significantly different than other levels⇒Need transformation, e.g., log

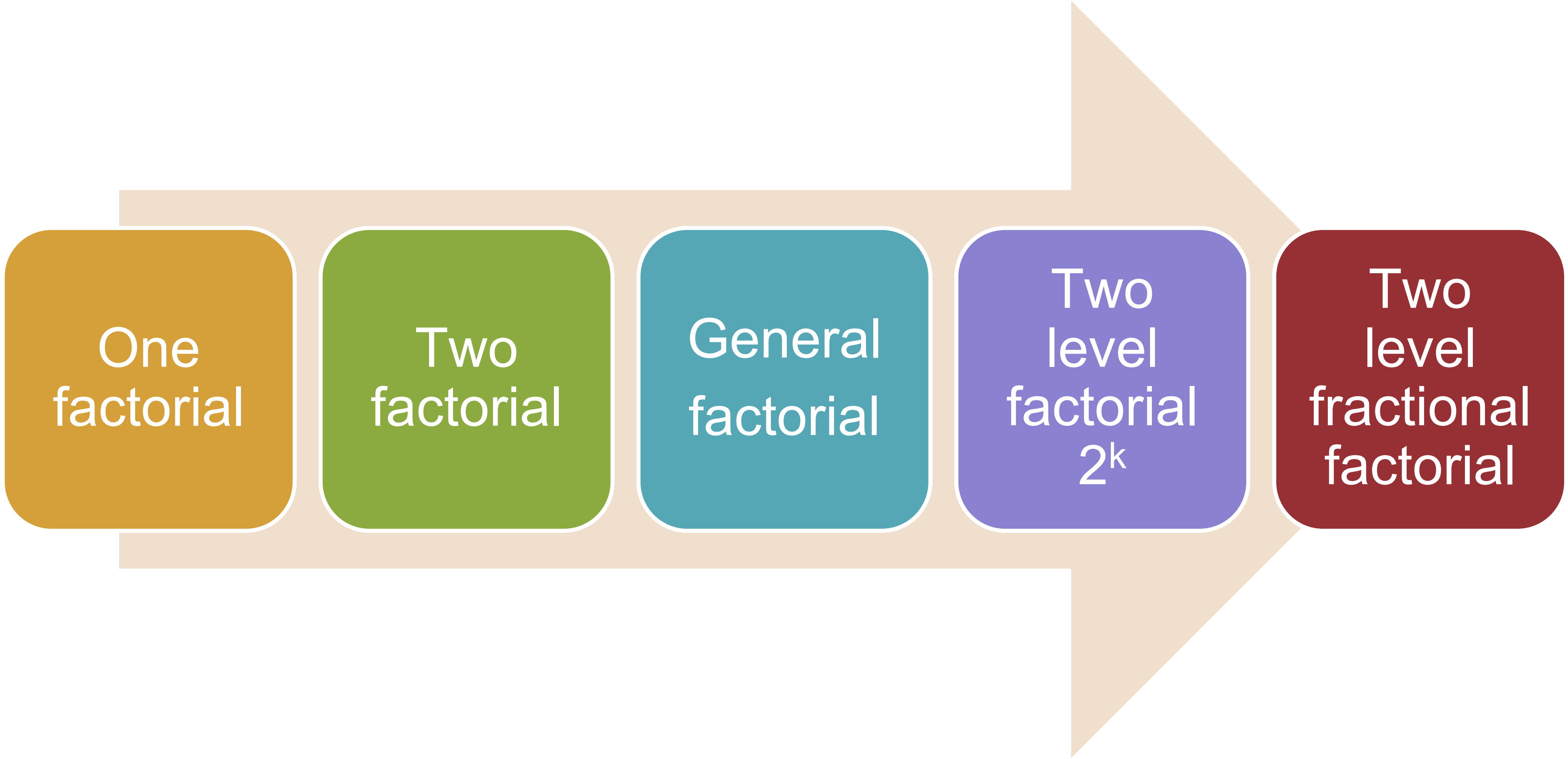


RECAP

Design of Experiments I

- One factorial design
- ANOVA Analysis, F-test
- Confidence Interval of Estimates, t-test




One factorial
Two factorial
General factorial
Two level factorial 2^k
Two level fractional factorial

1

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

2

Estimate parameters from measurement

$$SST = \sum_{ij} (y_{ij} - \bar{y}_{..}) = SSA + SSE \qquad \frac{SSA/df_A}{SSE/df_e}$$

Confidence of interval of μ and α_j

3

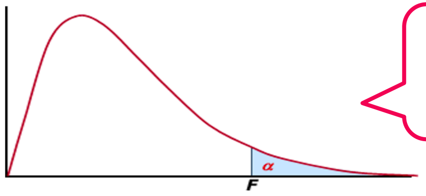
Goodness of fit:
Errors, ANOVA
Confidence interval

4

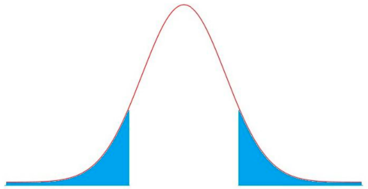
Check assumption

Q: By what?

Plotting and visual examination.



Q: A good model means?



Bigger F statistics, small P values.

Two-Factor Design

Examples of Two Factor Experiment:

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

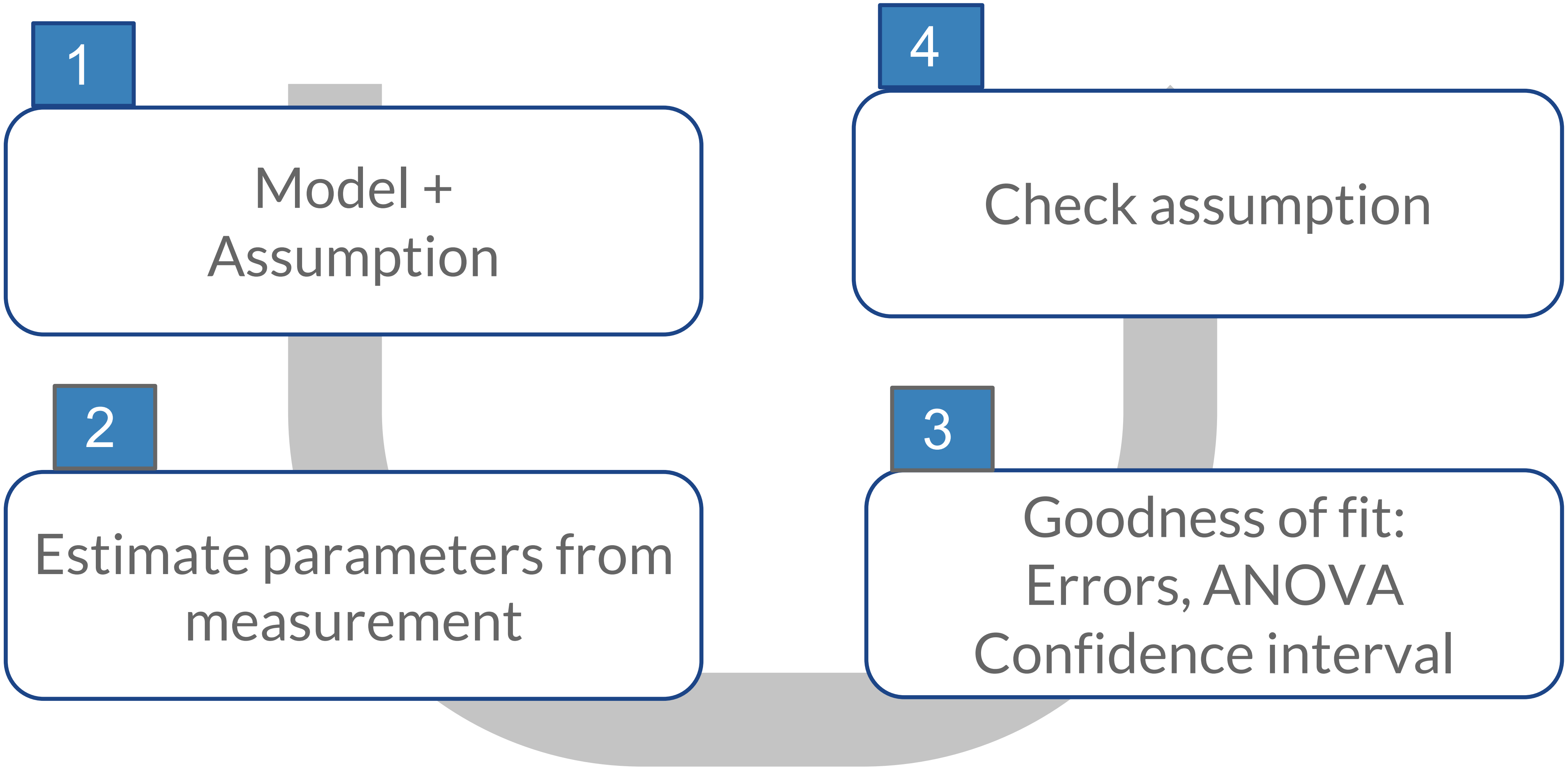
Large difference b/tw max and min values

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

Q. Why?

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


1
Model + Assumption
2
Estimate parameters from measurement
3
Goodness of fit: Errors, ANOVA Confidence interval
4
Check assumption

Model for Two Factorial

$j = \{1 \ldots a\},\ a$ is the number of factor a levels
$i = \{1 \ldots b\},\ b$ is the number of factor b levels
$k = \{1 \ldots r\},\ r$ is the number of replications per level

1. Model: With *r* replications-

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

α_j : Effect of factor A

β_j :Effect of factor B

γ_{ij} :Effect of interaction A&B.

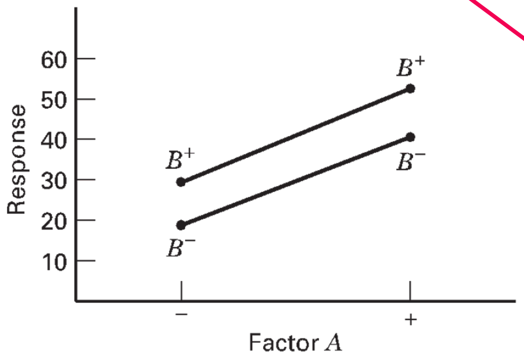
Q: Total number of experiment?

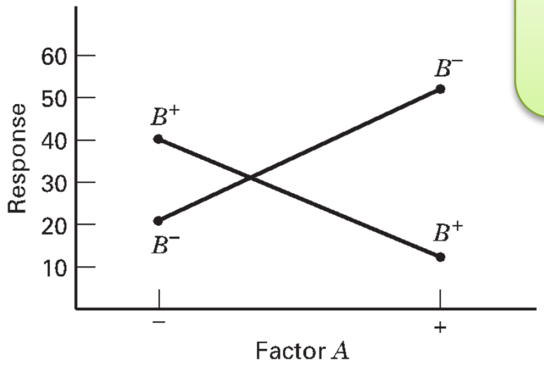
a x b x r

Q: Which figure has interactive effect?

Q: What is $\beta_2=3$?

When factor B at level 2, it will be 3 units higher than the overall average





Right figure has interaction

Model for Two Factorial

$j = \{1 \ldots a\},\ a$ is the number of factor a levels
$i = \{1 \ldots b\},\ b$ is the number of factor b levels
$k = \{1 \ldots r\},\ r$ is the number of replications per level

1. Model: With *r* replications-

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

α_j : Effect of factor A

β_j :Effect of factor B

γ_{ij} :Effect of interaction A&B.

$$\sum_{j=1}^{a} \alpha_j = 0; \sum_{i=1}^{b} \beta_i = 0;$$

$$\sum_{j=1}^{a} \gamma_{1j} = \sum_{j=1}^{a} \gamma_{2j} = \cdots = \sum_{j=1}^{a} \gamma_{bj} = 0$$

$$\sum_{i=1}^{b} \gamma_{i1} = \sum_{i=1}^{b} \gamma_{i2} = \cdots = \sum_{i=1}^{b} \gamma_{ia} = 0$$

$$\sum_{k=1}^{r} e_{ijk} = 0 \quad \forall i, j$$

Q: What about distribution of error?

$$e_{ijk} \sim N(0, \sigma)$$

Model for Two Factorial

$j = \{1 \ldots a\},\ a$ is the number of factor a levels
$i = \{1 \ldots b\},\ b$ is the number of factor b levels
$k = \{1 \ldots r\},\ r$ is the number of replications per level

1. Model: With *r* replications-

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

α_j : Effect of factor A

β_j :Effect of factor B

γ_{ij} :Effect of interaction A&B.

2. How to estimate the parameters

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

$$\alpha_j = \bar{y}_{.j.} - \bar{y}_{...}$$

$$\beta_i = \bar{y}_{i..} - \bar{y}_{...}$$

$$\gamma_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

Check the textbook for derivation.

Back to the Example

Q. How to interpret those values

Workloads	Processors W	X	Y	Z	Row Sum	Row Mean	Row Effect
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			

Processor W requires $10^{0.23}$ (=1.69) less code than avg processor.

Processor X requires $10^{0.02}$ (=1.05) less than an avg. processor.

Estimated interaction terms

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

Workload I on processor W requires 0.02 less log code size than an average workload on processor W. Equivalently 0.02 less log code size than workload I on an average processor.
.

$j = \{1 \ldots a\},\ a$ is the number of factor a levels

$i = \{1 \ldots b\},\ b$ is the number of factor b levels

$k = \{1 \ldots r\},\ r$ is the number of replications per level

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

Back to the Example: Analysis

- Overall model

ANOVA

Q: What analysis and what to check for them?

$$\alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$$

$$\beta_1 = \beta_2 = \cdots = \beta_b = 0$$

$$\gamma_{11} = \cdots = \gamma_{1a} = \gamma_{21} = \cdots = \gamma_{2a} \cdots = \gamma_{b1} = \cdots = \gamma_{ba} = 0$$

- Individual parameters

Confidence Interval

Q. How many parameters

a+b+ab

Analyzing the Model by ANOVA

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

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

Q: What are those terms

Variations explained by factor A, B, AB.

$$SST = SSY - SS0 = SSA + SSB + SSAB + SSE$$

$$\sum_{ijk}(y_{ijk} - \bar{y}_{...})^2 = \sum_{ijk} y_{ijk}^2 - \sum_{ijk} \bar{y}_{...} = \sum_{ijk} \alpha_i^2 + \sum_{ijk} \beta_j^2 + \sum_{ijk} \gamma_{ij}^2 + \sum_{ijk} e_{ijk}^2$$

Source	SSY	SS0	SSA	SSB	SSAB	SSE
Degree of freedom (V)	abr	1	a-1	b-1	(a-1)(b-1)	ab(r-1)

Q: What about their degree of freedom?

Q: how additive property work here?

Analyzing the Model by ANOVA

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

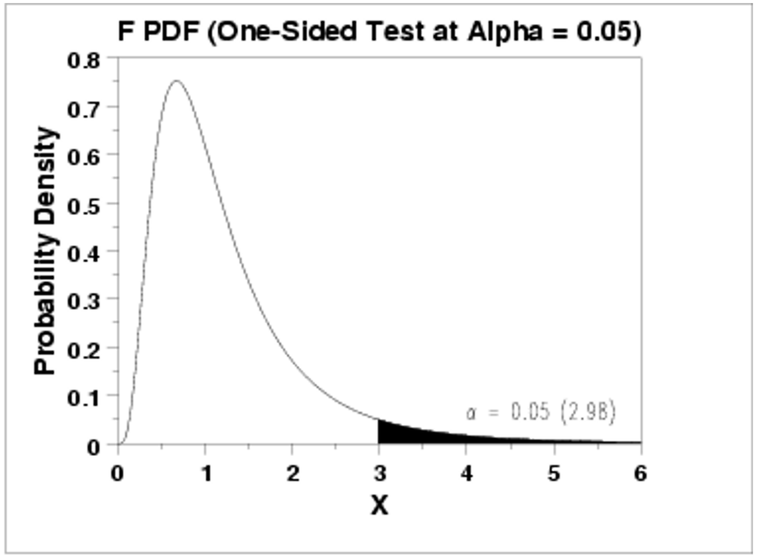
$$e_{ijk} = y_{ijk} - \bar{y}_{ij.}$$

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

Q. Those values tell us: importance or significance

Q. How should we use these values?

Compare it with $F_{[1-\alpha;\, vA,\, ve]}$, $F_{[1-\alpha;\, vB,\, ve]}$ & $F_{[1-\alpha;\, vAB,\, ve]}$



ANOVA for Two Factors w Replications

Standard output

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

Back to the example

Compo-nent	Sum of Squares	%Variation	DF	Mean Square	F-Comp.	F-Table
y	936.95					
$\bar{y}_{...}$	932.51					
$y - \bar{y}_{...}$	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		

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

Q. Are they significant

YES

Confidence Intervals For Effects

$$\text{CI range of parameter} = \begin{cases} \text{estimate} + t_{1-\alpha/2,df} \cdot (\text{std of estimate of parameter}) \\ \text{estimate} - t_{1-\alpha/2,df} \cdot (\text{std of estimate of parameter}) \end{cases}$$

Parameter Estimation

Be careful

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}	$\bar{y}_{ij.}$-$\bar{y}_{i..}$-$\bar{y}_{.j.}$+$\bar{y}_{...}$	$s_e^2(a-1)(b-1)/abr$
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)

Q. Which factors are not significant?

Yes, except K.

Back to example: CI for Processor W

When sample size is greater than 32, T distribution is similar to Normal distribution.

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

$$s_{\alpha_j} = s_e \sqrt{\frac{a-1}{abr}} = 0.03\sqrt{\frac{4-1}{4\times5\times3}} = 0.0060$$

Q: Why not T distribution table

- 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:

Q. Why 0.95 not 0.9?

T distribution is two sided

CI = {-0.230 +1.645 * .0060, -.2304 -1.645 * .0060} = (-0.240, -0.220)

Q. significant

Yes. 0 is not included

Recommended Readings

- The Art of Computer Systems Performance Analysis: Chapter 16, 20, 22 and 23
- Design and Analysis of Experiments: Chapter 3

Homework: What is the difference with
Regression models?

Thanks!

Any questions?

lydiaychen@ieee.org