Comparing Alternatives and Experiental Design

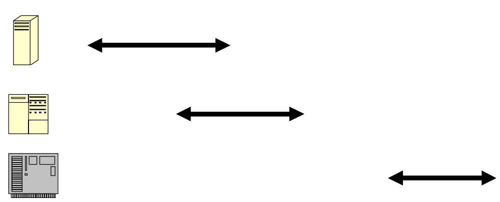
Performance Modeling Lecture #4

Repetition

- Average Performance and Variability
  - Different mean values
  - Variance
- Errors in measurements
  - Different types of errors
  - How to deal with them
- Comparing two sets of measurements

Comparing More Than Two Alternatives

- Naïve approach
  - Compare confidence intervals



One-Factor Analysis of Variance (ANOVA)

- Very general technique
  - Look at total *variation* in a set of measurements
  - Divide into meaningful components
- Also called
  - One-way classification
  - One-factor experimental design
- Introduce basic concept with one-factor ANOVA
- Generalize later with *design of experiments*

One-Factor Analysis of Variance (ANOVA)

- Separates total variation observed in a set of measurements into:
  1. Variation within one system
     - Due to random measurement errors
  2. Variation between systems
     - Due to real differences + random error
- Is variation(2) statistically > variation(1)?

ANOVA

- Make *n* measurements of *k* alternatives
- y_{ij} = *i*th measurment on *j*th alternative
- Assumes errors are:
  - Independent
  - Gaussian (normal)

Measurements for All Alternatives

	Alternatives					
Measure ments	1	2	…	*j*	…	*k*
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Col mean	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Effect	α_1	α_2	…	α_j	…	α_k

Overall Mean

- Average of all measurements made of all alternatives

$$\bar{y}_{..} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} y_{ij}}{kn}$$

Overall Mean

	Alternatives					
Measure ments	1	2	…	*j*	…	*k*
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Col mean	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Effect	α_1	α_2	…	α_j	…	α_k

Column Means

- Column means are average values of all measurements within a single alternative
  - Average performance of one alternative

$$\bar{y}_{.j} = \frac{\sum_{i=1}^{n} y_{ij}}{n}$$

Column Means

	Alternatives					
Measure ments	1	2	…	*j*	…	*k*
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Col mean	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Effect	α_1	α_2	…	α_j	…	α_k

Error = Deviation From Column Mean

	Alternatives					
Measure ments	1	2	…	*j*	…	*k*
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Col mean	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Effect	α_1	α_2	…	α_j	…	α_k

Effect = Deviation From Overall Mean

Measurements	Alternatives					
	1	2	…	*j*	…	*k*
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Col mean	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Effect	α_1	α_2	…	α_j	…	α_k

Effects and Errors

- *Effect* is distance from overall mean
  - Horizontally across alternatives
- *Error* is distance from column mean
  - Vertically within one alternative
  - Error across alternatives, too
- Individual measurements are then:

$$y_{ij} = \bar{y}_{..} + \alpha_j + e_{ij}$$

Sum of Squares of Differences

$$SSA = n\sum_{j=1}^{k}\left(\bar{y}_{.j} - \bar{y}_{..}\right)^2$$

$$SSE = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(y_{ij} - \bar{y}_{.j}\right)^2$$

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n}\left(y_{ij} - \bar{y}_{..}\right)^2$$

Sum of Squares of Differences

- ***SST*** = differences between each measurement and overall mean
- ***SSA*** = variation due to effects of alternatives
- ***SSE*** = variation due to errors in measurments

$$SST = SSA + SSE$$

ANOVA – Fundamental Idea

- Separates variation in measured values into:

1. Variation due to effects of alternatives
   - SSA – variation across columns
2. Variation due to errors
   - SSE – variation within a single column

- If differences among alternatives are due to real differences,
  - SSA should be statistically > SSE

Comparing SSE and SSA

- Simple approach
  - *SSA* / *SST* = fraction of total variation explained by differences among alternatives
  - *SSE* / *SST* = fraction of total variation due to experimental error
- But is it statistically significant?

Statistically Comparing SSE and SSA

$$\text{Variance} = \text{mean square value}$$

$$= \frac{\text{total variation}}{\text{degrees of freedom}}$$

$$s_x^2 = \frac{SSx}{df}$$

Degrees of Freedom

- $df(SSA) = k - 1$, since k alternatives
- $df(SSE) = k(n - 1)$, since k alternatives, each with $(n - 1)$ df
- $df(SST) = df(SSA) + df(SSE) = kn - 1$

Degrees of Freedom for Effects

	Alternatives					
Measurements	1	2	…	j	…	k
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Col mean	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Effect	α_1	α_2	…	α_j	…	α_k

Degrees of Freedom for Errors

	Alternatives					
Measurements	1	2	…	j	…	k
1	y_{11}	y_{12}	…	y_{1j}	…	y_{k1}
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}
…	…	…	…	…	…	…
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}
…	…	…	…	…	…	…
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}
Col mean	$y_{.1}$	$y_{.2}$	…	$y_{.j}$	…	$y_{.k}$
Effect	α_1	α_2	…	α_j	…	α_k

Variances from Sum of Squares (Mean Square Value)

$$s_a^2 = \frac{SSA}{k-1}$$

$$s_e^2 = \frac{SSE}{k(n-1)}$$

Comparing Variances

- Use F-test to compare ratio of variances

$$F = \frac{s_a^2}{s_e^2}$$

$$F_{[1-\alpha; df(num), df(denom)]} = \text{tabulated critical values}$$

F-test

- If $F_{computed} > F_{table}$
  → We have $(1 - \alpha)$ * 100% confidence that variation due to actual differences in alternatives, SSA, is statistically greater than variation due to errors, SSE.

ANOVA Summary

Variation	Alternatives	Error	Total
Sum of squares	SSA	SSE	SST
Deg freedom	$k-1$	$k(n-1)$	$kn-1$
Mean square	$s_a^2 = SSA/(k-1)$	$s_e^2 = SSE/[k(n-1)]$	
Computed F	s_a^2/s_e^2		
Tabulated F	$F_{[1-\alpha;(k-1),k(n-1)]}$		

ANOVA Example

	Alternatives			
Measurements	**1**	**2**	**3**	**Overall mean**
1	0.0972	0.1382	0.7966	
2	0.0971	0.1432	0.5300	
3	0.0969	0.1382	0.5152	
4	0.1954	0.1730	0.6675	
5	0.0974	0.1383	0.5298	
Column mean	0.1168	0.1462	0.6078	0.2903
Effects	-0.1735	-0.1441	0.3175	

ANOVA Example

Variation	Alternatives	Error	Total
Sum of squares	$SSA = 0.7585$	$SSE = 0.0685$	$SST = 0.8270$
Deg freedom	$k-1=2$	$k(n-1)=12$	$kn-1=14$
Mean square	$s_a^2 = 0.3793$	$s_e^2 = 0.0057$	
Computed F	$0.3793/0.0057 = 66.4$		
Tabulated F	$F_{[0.95;2,12]} = 3.89$		

Conclusions from example

- SSA/SST = 0.7585/0.8270 = 0.917
  → 91.7% of total variation in measurements is due to differences among alternatives
- SSE/SST = 0.0685/0.8270 = 0.083
  → 8.3% of total variation in measurements is due to noise in measurements
- Computed F statistic > tabulated F statistic
  → 95% confidence that differences among alternatives are statistically significant.

Important Points

- Use one-factor ANOVA to separate total variation into:
  - Variation within one system
    - Due to random errors
  - Variation between systems
    - Due to real differences (+ random error)
- Is the variation due to real differences *statistically* greater than the variation due to errors?
- Use contrasts to compare effects of subsets of alternatives

Design of Experiments

- **Goals**
- **Terminology**
- **Full factorial designs**
  - ***m*-factor ANOVA**
- **Fractional factorial designs**
- **Multi-factorial designs**

Generalized *Design of Experiments*

- Goals
  - Isolate effects of each input variable.
  - Determine effects of interactions.
  - Determine magnitude of experimental error
  - Obtain maximum information for given effort
- Basic idea
  - Expand 1-factor ANOVA to *m* factors

Terminology

- Response variable
  - Measured output value
    - E.g. total execution time
- Factors
  - Input variables that can be changed
    - E.g. cache size, clock rate, bytes transmitted
- Levels
  - Specific values of factors (inputs)
    - Continuous (~bytes) or discrete (type of system)

Terminology

- Replication
  - Completely re-run experiment with same input levels
  - Used to determine impact of measurement error
- Interaction
  - *Effect* of one input factor depends on *level* of another input factor

Two-factor Experiments

- Two factors (inputs)
  - A, B
- Separate total variation in output values into:
  - Effect due to A
  - Effect due to B
  - Effect due to interaction of A and B (AB)
  - Experimental error

Example – User Response Time

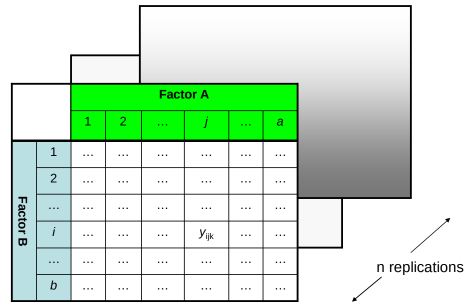
- A = degree of multiprogramming
- B = memory size
- AB = interaction of memory size and degree of multiprogramming

	B (Mbytes)		
A	32	64	128
1	0.25	0.21	0.15
2	0.52	0.45	0.36
3	0.81	0.66	0.50
4	1.50	1.45	0.70

Two-factor ANOVA

- Factor A – *a* input levels
- Factor B – *b* input levels
- *n* measurements for each input combination
- *abn* total measurements

Two Factors, *n* Replications



Recall: One-factor ANOVA

- Each individual measurement is composition of
  - Overall mean
  - Effect of alternatives
  - Measurement errors

$$y_{ij} = \bar{y}_{..} + \alpha_i + e_{ij}$$

$$\bar{y}_{..} = \text{overall mean}$$

$$\alpha_i = \text{effect due to A}$$

$$e_{ij} = \text{measurement error}$$

Two-factor ANOVA

- Each individual measurement is composition of
  - Overall mean
  - Effects
  - Interactions
  - Measurement errors

$$y_{ijk} = \bar{y}_{...} + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

$$\bar{y}_{...} = \text{overall mean}$$

$$\alpha_i = \text{effect due to A}$$

$$\beta_j = \text{effect due to B}$$

$$\gamma_{ij} = \text{effect due to interaction of A and B}$$

$$e_{ijk} = \text{measurement error}$$

Sum-of-Squares

- As before, use sum-of-squares identity

SST = SSA + SSB + SSAB + SSE

- Degrees of freedom
  - *df*(SSA) = *a* – 1
  - *df*(SSB) = *b* – 1
  - *df*(SSAB) = (*a* – 1)(*b* – 1)
  - *df*(SSE) = *ab*(*n* – 1)
  - *df*(SST) = *abn* - 1

Two-Factor ANOVA

	A	B	AB	Error
Sum of squares	SSA	SSB	$SSAB$	SSE
Deg freedom	$a-1$	$b-1$	$(a-1)(b-1)$	$ab(n-1)$
Mean square	$s_a^2 = SSA/(a-1)$	$s_b^2 = SSB/(b-1)$	$s_{ab}^2 = SSAB/[(a-1)(b-1)]$	$s_e^2 = SSE/[ab(n-1)]$
Computed F	$F_a = s_a^2/s_e^2$	$F_b = s_b^2/s_e^2$	$F_{ab} = s_{ab}^2/s_e^2$	
Tabulated F	$F_{[1-\alpha;(a-1),ab(n-1)]}$	$F_{[1-\alpha;(b-1),ab(n-1)]}$	$F_{[1-\alpha;(a-1)(b-1),ab(n-1)]}$	

Need for Replications

- If n=1
  - Only one measurement of each configuration
- Can then be shown that
  - SSAB = SST – SSA – SSB
- Since
  - SSE = SST – SSA – SSB – SSAB
- We have
  - SSE = 0

Need for Replications

- Thus, when n=1
  - SSE = 0
  - → No information about measurement errors
- Cannot separate effect due to interactions from measurement noise
- Must *replicate* each experiment at least twice

Example

- Output = user response time (seconds)
- Want to separate effects due to
  - A = degree of multiprogramming
  - B = memory size
  - AB = interaction
  - Error
- Need replications to separate error

	B (Mbytes)		
A	32	64	128
1	0.25	0.21	0.15
2	0.52	0.45	0.36
3	0.81	0.66	0.50
4	1.50	1.45	0.70

Example

	B (Mbytes)		
A	**32**	**64**	**128**
1	0.25	0.21	0.15
	0.28	0.19	0.11
2	0.52	0.45	0.36
	0.48	0.49	0.30
3	0.81	0.66	0.50
	0.76	0.59	0.61
4	1.50	1.45	0.70
	1.61	1.32	0.68

Example

	A	B	AB	Error
Sum of squares	3.3714	0.5152	0.4317	0.0293
Deg freedom	3	2	6	12
Mean square	1.1238	0.2576	0.0720	0.0024
Computed F	460.2	105.5	29.5	
Tabulated F	$F_{[0.95;3,12]} = 3.49$	$F_{[0.95;2,12]} = 3.89$	$F_{[0.95;6,12]} = 3.00$	

Conclusions From the Example

- 77.6% (SSA/SST) of all variation in response time due to degree of multiprogramming
- 11.8% (SSB/SST) due to memory size
- 9.9% (SSAB/SST) due to interaction
- 0.7% due to measurement error
- 95% confident that all effects and interactions are statistically significant

A Problem

- *Full factorial design with replication*
  - Measure system response with all possible input combinations
  - Replicate each measurement n times to determine effect of measurement error
- m factors, v levels, n replications

  $\rightarrow n\, v^m$ experiments
- m = 5 input factors, v = 4 levels, n = 3
  - $\rightarrow 3(4^5)$ = 3,072 experiments!

Fractional Factorial Designs: $n2^m$ Experiments

- Special case of generalized m-factor experiments
- Restrict each factor to two possible values
  - High, low
  - On, off
- Find factors that have largest impact
- Full factorial design with only those factors

$n2^m$ Experiments

	A	B	AB	Error
Sum of squares	SSA	SSB	$SSAB$	SSE
Deg freedom	1	1	1	$2^m(n-1)$
Mean square	$s_a^2 = SSA/1$	$s_b^2 = SSB/1$	$s_{ab}^2 = SSAB/1$	$s_e^2 = SSE/[2^m(n-1)]$
Computed F	$F_a = s_a^2/s_e^2$	$F_b = s_b^2/s_e^2$	$F_{ab} = s_{ab}^2/s_e^2$	
Tabulated F	$F_{[1-\alpha;1,2^m(n-1)]}$	$F_{[1-\alpha;1,2^m(n-1)]}$	$F_{[1-\alpha;1,2^m(n-1)]}$	

Important Points

- Experimental design is used to
  - Isolate the effects of each input variable.
  - Determine the effects of interactions.
  - Determine the magnitude of the error
  - Obtain maximum information for given effort
- Expand 1-factor ANOVA to m factors
- Use $n2^m$ design to reduce the number of experiments needed
  - But loses some information

Summary

- Design of experiments
  - Isolate effects of each input variable.
  - Determine effects of interactions.
  - Determine magnitude of experimental error
- m-factor ANOVA (*full factorial design*)
  - All effects, interactions, and errors

Summary

- $n2^m$ designs
  - Fractional factorial design
- All effects, interactions, and errors
- But for only 2 input values
  - high/low
  - on/off

Contrasts

- ANOVA tells us that there is a statistically significant difference among alternatives
- But it does *not* tell us *where* difference is
- Use method of contrasts to compare subsets of alternatives
  - A vs B
  - {A, B} vs {C}
  - Etc.

Contrasts

- Contrast = linear combination of *effects* of alternatives

$$c = \sum_{j=1}^{k} w_j \alpha_j$$

$$\sum_{j=1}^{k} w_j = 0$$

Contrasts

- E.g. Compare effect of system 1 to effect of system 2

$$w_1 = 1$$
$$w_2 = -1$$
$$w_3 = 0$$
$$c = (1)\alpha_1 + (-1)\alpha_2 + (0)\alpha_3$$
$$= \alpha_1 - \alpha_2$$

Construct confidence interval for contrasts

- Need
  - Estimate of variance
  - Appropriate value from *t* table
- Compute confidence interval as before
- If interval includes 0
  - Then no statistically significant difference exists between the alternatives included in the contrast

Variance of random variables

- Recall that, for independent random variables X_1 and X_2

$$\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2]$$
$$\mathrm{Var}[aX_1] = a^2\,\mathrm{Var}[X_1]$$

Variance of a contrast *c*

$$\mathrm{Var}[c] = \mathrm{Var}[\textstyle\sum_{j=1}^{k} (w_j \alpha_j)]$$
$$= \textstyle\sum_{j=1}^{k} \mathrm{Var}[w_j \alpha_j]$$
$$= \textstyle\sum_{j=1}^{k} w_j^2\,\mathrm{Var}[\alpha_j]$$

$$s_c^2 = \frac{\sum_{j=1}^{k} (w_j^2 s_e^2)}{kn}$$
$$s_e^2 = \frac{SSE}{k(n-1)}$$
$$df\,(s_c^2) = k(n-1)$$

- Assumes variation due to errors is equally distributed among *kn* total measurements

Confidence interval for contrasts

$$(c_1, c_2) = c \mp t_{1-\alpha/2;k(n-1)} s_c$$

$$s_c = \sqrt{\frac{\sum_{j=1}^{k}(w_j^2 s_e^2)}{kn}}$$

$$s_e^2 = \frac{SSE}{k(n-1)}$$

Example

- 90% confidence interval for contrast of [Sys1-Sys2]

$$\alpha_1 = -0.1735$$

$$\alpha_2 = -0.1441$$

$$\alpha_3 = 0.3175$$

$$c_{[1-2]} = -0.1735 - (-0.1441) = -0.0294$$

$$s_c = s_e \sqrt{\frac{1^2 + (-1)^2 + 0^2}{3(5)}} = 0.0275$$

$$90\% : (c_1, c_2) = (-0.0784, 0.0196)$$