CS 147:
Computer Systems Performance Analysis
Two-Factor Designs

CS 147:
Computer Systems Performance Analysis
Two-Factor Designs

Overview

Two-Factor Designs
- No Replications
- Adding Replications

Overview

Two-Factor Designs
No Replications
Adding Replications

Two-Factor Design Without Replications

- Used when only two parameters, but multiple levels for each
- Test all combinations of levels of the two parameters
- One replication (observation) per combination
- For factors A and B with a and b levels, ab experiments required

Two-Factor Design Without Replications

- Used when only two parameters, but multiple levels for each
- Test all combinations of levels of the two parameters
- One replication (observation) per combination
- For factors A and B with a and b levels, ab experiments required

When to Use This Design?

- System has two important factors
- Factors are categorical
- More than two levels for at least one factor
- Examples:
  - Performance of different processors under different workloads
  - Characteristics of different compilers for different benchmarks
  - Performance of different Web browsers on different sites

When to Use This Design?
- System has two important factors
- Factors are categorical
- More than two levels for at least one factor
- Examples:
  - Performance of different processors under different workloads
  - Characteristics of different compilers for different benchmarks
  - Performance of different Web browsers on different sites

When to Avoid This Design?

- Systems with more than two important factors
  - Use general factorial design
- Non-categorical variables
  - Use regression
- Only two levels per factor
  - Use 2^2 designs

When to Avoid This Design?

- Systems with more than two important factors
  - Use general factorial design
- Non-categorical variables
  - Use regression
- Only two levels per factor
  - Use 2^2 designs

Model For This Design

- $y_{ij} = \mu + \alpha_j + \beta_i + e_{ij}$
- y_{ij} is observation
- μ is mean response
- α_j is effect of factor A at level j
- β_i is effect of factor B at level i
- e_{ij} is error term
- Sums of α_j's and β_i's are both zero

Model For This Design

- $y_{ij} = \mu + \alpha_j + \beta_i + e_{ij}$
- y_{ij} is observation
- μ is mean response
- α_j is effect of factor A at level j
- β_i is effect of factor B at level i
- e_{ij} is error term
- Sums of α_j's and β_i's are both zero

Assumptions of the Model

- Factors are additive
- Errors are additive
- Typical assumptions about errors:
  - Distributed independently of factor levels
  - Normally distributed
- Remember to check these assumptions!

Assumptions of the Model

- Factors are additive
- Errors are additive
- Typical assumptions about errors:
  - Distributed independently of factor levels
  - Normally distributed
- Remember to check these assumptions!

Computing Effects

- Need to figure out μ, α_j, and β_i
- Arrange observations in two-dimensional matrix
  - *b* rows, *a* columns
- Compute effects such that error has zero mean
  - Sum of error terms across all rows and columns is zero

Computing Effects

- Need to figure out μ, α_j, and β_i
- Arrange observations in two-dimensional matrix
  - *b* rows, *a* columns
- Compute effects such that error has zero mean
  - Sum of error terms across all rows and columns is zero

Two-Factor Full Factorial Example

- Want to expand functionality of a file system to allow automatic compression
- Examine three choices:
  - Library substitution of file system calls
  - New VFS
  - Stackable layers
- Three different benchmarks
- Metric: response time

Two-Factor Full Factorial Example
- Want to expand functionality of a file system to allow automatic compression
- Examine three choices:
  - Library substitution of file system calls
  - New VFS
  - Stackable layers
- Three different benchmarks
- Metric: response time

Data for Example

	Library	VFS	Layers
Compile Benchmark	94.3	89.5	96.2
Email Benchmark	224.9	231.8	247.2
Web Server Benchmark	733.5	702.1	797.4

Data for Example

	Library	VFS	Layers
Compile Benchmark	94.3	89.5	96.2
Email Benchmark	224.9	231.8	247.2
Web Server Benchmark	733.5	702.1	797.4

Computing μ

- Averaging the j^{th} column,

$$\overline{y}_{\cdot j} = \mu + \alpha_j + \frac{1}{b}\sum_i \beta_i + \frac{1}{b}\sum_i e_{ij}$$

- By assumption, error terms add to zero
- Also, the β_j's add to zero, so $\overline{y}_{\cdot j} = \mu + \alpha_j$
- Averaging rows produces $\overline{y}_{i\cdot} = \mu + \beta_i$
- Averaging everything produces $\overline{y}_{\cdot\cdot} = \mu$

Computing μ

- Averaging the j^{th} column,

$$\overline{y}_{\cdot j} = \mu + \alpha_j + \frac{1}{b}\sum_i \beta_i + \frac{1}{b}\sum_i e_{ij}$$

- By assumption, error terms add to zero
- Also, the β_j's add to zero, so $\overline{y}_{\cdot j} = \mu + \alpha_j$
- Averaging rows produces $\overline{y}_{i\cdot} = \mu + \beta_i$
- Averaging everything produces $\overline{y}_{\cdot\cdot} = \mu$

Model Parameters

Using same techniques as for one-factor designs, parameters are:

- $\overline{y}_{..} = \mu$
- $\alpha_j = \overline{y}_{.j} - \overline{y}_{..}$
- $\beta_i = \overline{y}_{i.} - \overline{y}_{..}$

Model Parameters

Using same techniques as for one-factor designs, parameters are:

- $\overline{y}_{..} = \mu$
- $\alpha_j = \overline{y}_{.j} - \overline{y}_{..}$
- $\beta_i = \overline{y}_{i.} - \overline{y}_{..}$

Calculating Parameters for the Example

- μ = grand mean = 357.4
- $\alpha_j = (-6.5, -16.3, 22.8)$
- $\beta_i = (-264.1, -122.8, 386.9)$
- So, for example, the model predicts that the email benchmark using a special-purpose VFS will take $357.4 - 16.3 - 122.8 = 218.3$ seconds

Calculating Parameters for the Example

- μ = grand mean = 357.4
- $\alpha_j = (-6.5, -16.3, 22.8)$
- $\beta_i = (-264.1, -122.8, 386.9)$
- So, for example, the model predicts that the email benchmark using a special-purpose VFS will take $357.4 - 16.3 - 122.8 = 218.3$ seconds

Estimating Experimental Errors

- Similar to estimation of errors in previous designs
- Take difference between model's predictions and observations
- Calculate Sum of Squared Errors
- Then allocate variation

Estimating Experimental Errors

- Similar to estimation of errors in previous designs
- Take difference between model's predictions and observations
- Calculate Sum of Squared Errors
- Then allocate variation

Allocating Variation

- Use same kind of procedure as on other models
- $SSY = SS0 + SSA + SSB + SSE$
- $SST = SSY - SS0$
- Can then divide total variation between SSA, SSB, and SSE

Allocating Variation

- Use same kind of procedure as on other models
- $SSY = SS0 + SSA + SSB + SSE$
- $SST = SSY - SS0$
- Can then divide total variation between SSA, SSB, and SSE

Calculating SS0, SSA, SSB

- $\text{SS0} = ab\mu^2$
- $\text{SSA} = b\sum_j \alpha_j^2$
- $\text{SSB} = a\sum_i \beta_i^2$
- Recall that *a* and *b* are numbers of levels for the factors

Calculating SS0, SSA, SSB

- $\text{SS0} = ab\mu^2$
- $\text{SSA} = b\sum_j \alpha_j^2$
- $\text{SSB} = a\sum_i \beta_i^2$
- Recall that *a* and *b* are numbers of levels for the factors

Allocation of Variation for Example

- $SSE = 2512$
- $SSY = 1,858,390$
- $SS0 = 1,149,827$
- $SSA = 2489$
- $SSB = 703,561$
- $SST = 708,562$
- Percent variation due to A: 0.35%
- Percent variation due to B: 99.3%
- Percent variation due to errors: 0.35%

Allocation of Variation for Example
- $SSE = 2512$
- $SSY = 1,858,390$
- $SS0 = 1,149,827$
- $SSA = 2489$
- $SSB = 703,561$
- $SST = 708,562$
- Percent variation due to A: 0.35%
- Percent variation due to B: 99.3%
- Percent variation due to errors: 0.35%

Analysis of Variation

- Again, similar to previous models, with slight modifications
- As before, use an ANOVA procedure
  - Need extra row for second factor
  - Minor changes in degrees of freedom
- End steps are the same
  - Compare F-computed to F-table
  - Compare for each factor

Analysis of Variation

- Again, similar to previous models, with slight modifications
- As before, use an ANOVA procedure
  - Need extra row for second factor
  - Minor changes in degrees of freedom
- End steps are the same
  - Compare F-computed to F-table
  - Compare for each factor

Analysis of Variation for Our Example

- $\text{MSE} = \text{SSE}/[(a-1)(b-1)] = 2512/[(2)(2)] = 628$
- $\text{MSA} = \text{SSA}/(a-1) = 2489/2 = 1244$
- $\text{MSB} = \text{SSB}/(b-1) = 703,561/2 = 351,780$
- F-computed for A $= \text{MSA}/\text{MSE} = 1.98$
- F-computed for B $= \text{MSB}/\text{MSE} = 560$
- 95% F-table value for A & B is 6.94
- So A is not significant, but B is

Analysis of Variation for Our Example

- $\text{MSE} = \text{SSE}/[(a-1)(b-1)] = 2512/[(2)(2)] = 628$
- $\text{MSA} = \text{SSA}/(a-1) = 2489/2 = 1244$
- $\text{MSB} = \text{SSB}/(b-1) = 703,561/2 = 351,780$
- F-computed for A $= \text{MSA}/\text{MSE} = 1.98$
- F-computed for B $= \text{MSB}/\text{MSE} = 560$
- 95% F-table value for A & B is 6.94
- So A is not significant, but B is

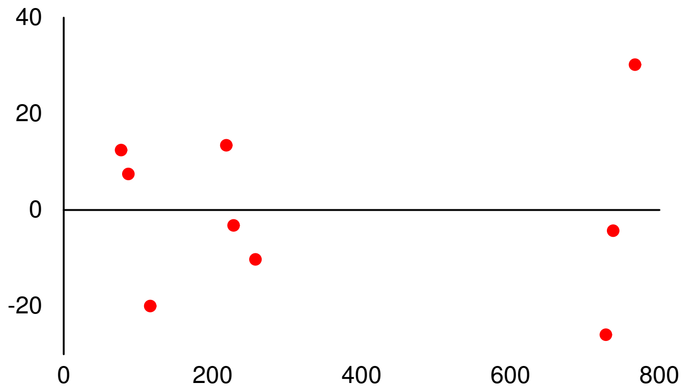
Checking Our Results with Visual Tests

- As always, check if assumptions made in the analysis are correct
- Use residuals vs. predicted and quantile-quantile plots

Checking Our Results with Visual Tests

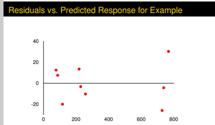
- As always, check if assumptions made in the analysis are correct
- Use residuals vs. predicted and quantile-quantile plots

Residuals vs. Predicted Response for Example



2015-06-15

CS147
└Two-Factor Designs
└No Replications
└Residuals vs. Predicted Response for Example



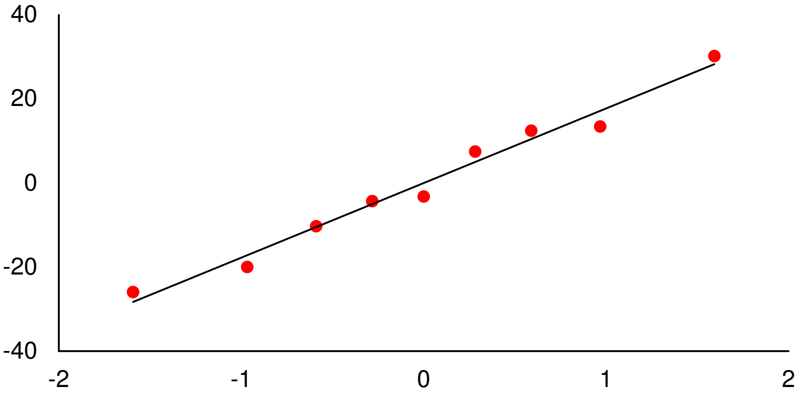
What Does the Chart Reveal?

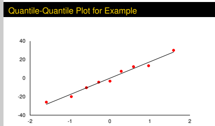
- Do we or don't we see a trend in errors?
- Clearly they're higher at highest level of the predictors
- But is that alone enough to call a trend?
  - Perhaps not, but we should take a close look at both factors to see if there's reason to look further
  - Maybe take results with a grain of salt

What Does the Chart Reveal?

- Do we or don't we see a trend in errors?
- Clearly they're higher at highest level of the predictors
- But is that alone enough to call a trend?
  - Perhaps not, but we should take a close look at both factors to see if there's reason to look further
  - Maybe take results with a grain of salt

Quantile-Quantile Plot for Example





Confidence Intervals for Effects

- Need to determine standard deviation for data as a whole
- Then can derive standard deviations for effects
  - Use different degrees of freedom for each
- Complete table in Jain, p. 351

Confidence Intervals for Effects
- Need to determine standard deviation for data as a whole
- Then can derive standard deviations for effects
  - Use different degrees of freedom for each
- Complete table in Jain, p. 351

Standard Deviations for Example

- $s_e = 25$
- Standard deviation of μ:

$$s_\mu = s_e/\sqrt{ab} = 25/\sqrt{3 \times 3} = 8.3$$

- Standard deviation of α_j:

$$s_{\alpha_j} = s_e\sqrt{(a-1)/ab} = 25\sqrt{2/9} = 11.8$$

- Standard deviation of β_i:

$$s_{\beta_i} = s_e\sqrt{(b-1)/ab} = 25\sqrt{2/9} = 11.8$$

Standard Deviations for Example

- $s_e = 25$
- Standard deviation of μ:

$$s_\mu = s_e/\sqrt{ab} = 25/\sqrt{3 \times 3} = 8.3$$

- Standard deviation of α_j:

$$s_{\alpha_j} = s_e\sqrt{(a-1)/ab} = 25\sqrt{2/9} = 11.8$$

- Standard deviation of β_i:

$$s_{\beta_i} = s_e\sqrt{(b-1)/ab} = 25\sqrt{2/9} = 11.8$$

Calculating Confidence Intervals for Example

- Only file system alternatives shown here
- We'll use 95% level
- 4 degrees of freedom
- CI for library solution: $(-39, 26)$
- CI for VFS solution: $(-49, 16)$
- CI for layered solution: $(-10, 55)$
- So none of the solutions are significantly different from mean at 95% confidence

Calculating Confidence Intervals for Example
- Only file system alternatives shown here
- We'll use 95% level
- 4 degrees of freedom
- CI for library solution: $(-39, 26)$
- CI for VFS solution: $(-49, 16)$
- CI for layered solution: $(-10, 55)$
- So none of the solutions are significantly different from mean at 95% confidence

Looking a Little Closer

- Do zero CI's mean that none of the alternatives for adding functionality are different?
- Not necessarily
- Use contrasts to check (see Section 18.5 & p. 366)

Looking a Little Closer

- Do zero CI's mean that none of the alternatives for adding functionality are different?
- Not necessarily
- Use contrasts to check (see Section 18.5 & p. 366)

Comparing Contrasts

- Is library approach significantly better than layers?
- Define a contrast: $u = \sum_{j=1}^{a} h_j \alpha_j$ where h_j's are chosen so that $\sum_{j=1}^{a} h_j = 0$
- To compare library vs. layers, set $h = (1, 0, -1)$
- Contrast mean $= \sum_{j=1}^{a} h_j \bar{y}_{\cdot j} = 350.9 - 380.267 = -29.367$
- Contrast variance $= s_e^2 (\sum_{j=1}^{a} h_j^2)/b = 25 \times 2/3 = 16.667$, so contrast s.d. $= 4.082$
- Using $t_{[1-\alpha/2;(a-1)(b-1)]} = t_{[.975;4]} = 2.776$, confidence interval is $-29.367 \mp 4.082 \times 2.776 = (-40.7, -18.0)$
- So library approach is better, at 95%

Comparing Contrasts

- Is library approach significantly better than layers?
- Define a contrast: $u = \sum_{j=1}^{a} h_j \alpha_j$ where h_j's are chosen so that $\sum_{j=1}^{a} h_j = 0$
- To compare library vs. layers, set $h = (1, 0, -1)$
- Contrast mean $= \sum_{j=1}^{a} h_j \bar{y}_{\cdot j} = 350.9 - 380.267 = -29.367$
- Contrast variance $= s_e^2 (\sum_{j=1}^{a} h_j^2)/b = 25 \times 2/3 = 16.667$, so contrast s.d. $= 4.082$
- Using $t_{[1-\alpha/2;(a-1)(b-1)]} = t_{[.975;4]} = 2.776$, confidence interval is $-29.367 \mp 4.082 \times 2.776 = (-40.7, -18.0)$
- So library approach is better, at 95%

Missing Observations

- Sometimes experiments go awry
- You don't want to discard an entire study away just because one observation got lost
- Solution:
  - Calculate row/column means and standard deviations based on actual observation count
  - Degrees of freedom in SS* also must be adjusted
  - See book for example
- Alternatives exist but are controversial
- If lots of missing values in a column or row, throw it out entirely
  - Best is to have only 1–2 missing values

Missing Observations

- Sometimes experiments go awry
- You don't want to discard an entire study away just because one observation got lost
- Solution:
  - Calculate row/column means and standard deviations based on actual observation count
  - Degrees of freedom in SS* also must be adjusted
  - See book for example
- Alternatives exist but are controversial
- If lots of missing values in a column or row, throw it out entirely
  - Best is to have only 1–2 missing values

Replicated Two-Factor Designs

- For r replications of each experiment, model becomes

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

- γ_{ij} represents interaction between factor A at level j and B at level i
- As before, effect sums $\sum \alpha_j$ and $\sum \beta_i$ are zero
- Interactions are zero for both row and column sums:

$$\forall i \quad \sum_{j=1}^{a} \gamma_{ij} = 0 \qquad \forall j \quad \sum_{i=1}^{b} \gamma_{ij} = 0$$

- Per-experiment errors add to zero:

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

Replicated Two-Factor Designs

- For r replications of each experiment, model becomes

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

- γ_{ij} represents interaction between factor A at level j and B at level i
- As before, effect sums $\sum \alpha_j$ and $\sum \beta_i$ are zero
- Interactions are zero for both row and column sums:

$$\forall i \quad \sum_{j=1}^{a} \gamma_{ij} = 0 \qquad \forall j \quad \sum_{i=1}^{b} \gamma_{ij} = 0$$

- Per-experiment errors add to zero:

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

Calculating Effects

Same as usual:

- Calculate grand mean $\overline{y}_{...}$, row and column means $\overline{y}_{i..}$ and $\overline{y}_{.j.}$ and per-experiment means $\overline{y}_{ij.}$
- $\mu = \overline{y}_{...}$
- $\alpha_j = \overline{y}_{.j.} - \mu$
- $\beta_i = \overline{y}_{i..} - \mu$
- $\gamma_{ij} = \overline{y}_{ij.} - \alpha_j - \beta_i - \mu$
- $e_{ijk} = y_{ijk} - \overline{y}_{ij.}$

Calculating Effects

Same as usual:

- Calculate grand mean $\overline{y}_{...}$, row and column means $\overline{y}_{i..}$ and $\overline{y}_{.j.}$ and per-experiment means $\overline{y}_{ij.}$
- $\mu = \overline{y}_{...}$
- $\alpha_j = \overline{y}_{.j.} - \mu$
- $\beta_i = \overline{y}_{i..} - \mu$
- $\gamma_{ij} = \overline{y}_{ij.} - \alpha_j - \beta_i - \mu$
- $e_{ijk} = y_{ijk} - \overline{y}_{ij.}$

Analysis of Variance

- Again, extension of earlier models
- See Table 22.5, p. 375, for formulas
- As usual, must do visual tests

Analysis of Variance

- Again, extension of earlier models
- See Table 22.5, p. 375, for formulas
- As usual, must do visual tests

Why Can We Find Interactions?

- Without replications, two-factor model didn't give interactions
- Why not?

Why Can We Find Interactions?
- Without replications, two-factor model didn't give interactions
- Why not?

This slide has animations.

In unreplicated experiment, we could have assumed no experimental errors and attributed variation to interaction instead (but that wouldn't be wise).

Why Can We Find Interactions?

- Without replications, two-factor model didn't give interactions
- Why not?
- Insufficient data
- Variation from predictions was attributed to errors, not interaction
  - Interaction is confounded with errors
- Now, we have more info
  - For given A, B setting, errors are assumed to cause variation in r replicated experiments
  - Any remaining variation must therefore be interaction

Why Can We Find Interactions?

- Without replications, two-factor model didn't give interactions
- Why not?
- Insufficient data
- Variation from predictions was attributed to errors, not interaction
  - Interaction is confounded with errors
- Now, we have more info
  - For given A, B setting, errors are assumed to cause variation in r replicated experiments
  - Any remaining variation must therefore be interaction

This slide has animations.

In unreplicated experiment, we could have assumed no experimental errors and attributed variation to interaction instead (but that wouldn't be wise).

General Full Factorial Designs

- Straightforward extension of two-factor designs
- Average along axes to get effects
- Must consider all interactions (various axis combinations)
- Regression possible for quantitative effects
  - But should have more than three data points
- If no replications, errors confounded with highest-level interaction

General Full Factorial Designs

- Straightforward extension of two-factor designs
- Average along axes to get effects
- Must consider all interactions (various axis combinations)
- Regression possible for quantitative effects
  - But should have more than three data points
- If no replications, errors confounded with highest-level interaction