Before-and-After Comparison

$$b_i = \text{before measurement}$$

$$a_i = \text{after measurement}$$

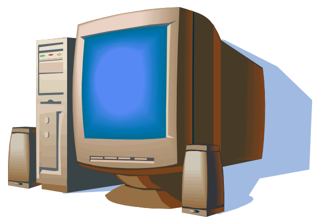
$$d_i = a_i - b_i$$

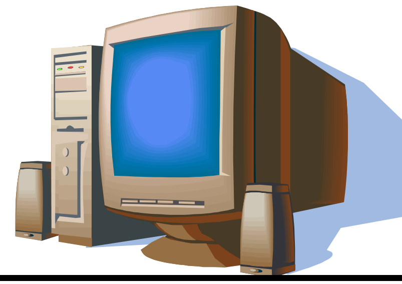
$$\bar{d} = \text{mean value of } d_i$$

$$s_d = \text{standard deviation of } d_i$$

$$(c_1, c_2) = \bar{d} \mp t_{1-\alpha/2;n-1} \frac{s_d}{\sqrt{n}}$$

Before-and-After Comparison





Measurement (i)	Before (b_i)	After (a_i)	Difference ($d_i = b_i - a_i$)
1	85	86	-1
2	83	88	-5
3	94	90	4
4	90	95	-5
5	88	91	-3
6	87	83	4

Before-and-After Comparison

$$\text{Mean of differences} = \bar{d} = -1$$

$$\text{Standard deviation} = s_d = 4.15$$

- From mean of differences, appears that change reduced performance.
- However, standard deviation is large.

95% Confidence Interval for Mean of Differences

$$t_{1-\alpha/2;n-1} = t_{0.975;5} = 2.571$$

	a		
n	0.90	0.95	0.975
…	…	…	…
5	1.476	2.015	2.571
6	1.440	1.943	2.447
…	…	…	…
∞	1.282	1.645	1.960

95% Confidence Interval for Mean of Differences

$$c_{1,2} = \bar{d} \mp t_{1-\alpha/2;n-1} \frac{s_d}{\sqrt{n}}$$

$$t_{1-\alpha/2;n-1} = t_{0.975;5} = 2.571$$

$$c_{1,2} = -1 \mp 2.571\left(\frac{4.15}{\sqrt{6}}\right)$$

$$c_{1,2} = [-5.36, 3.36]$$

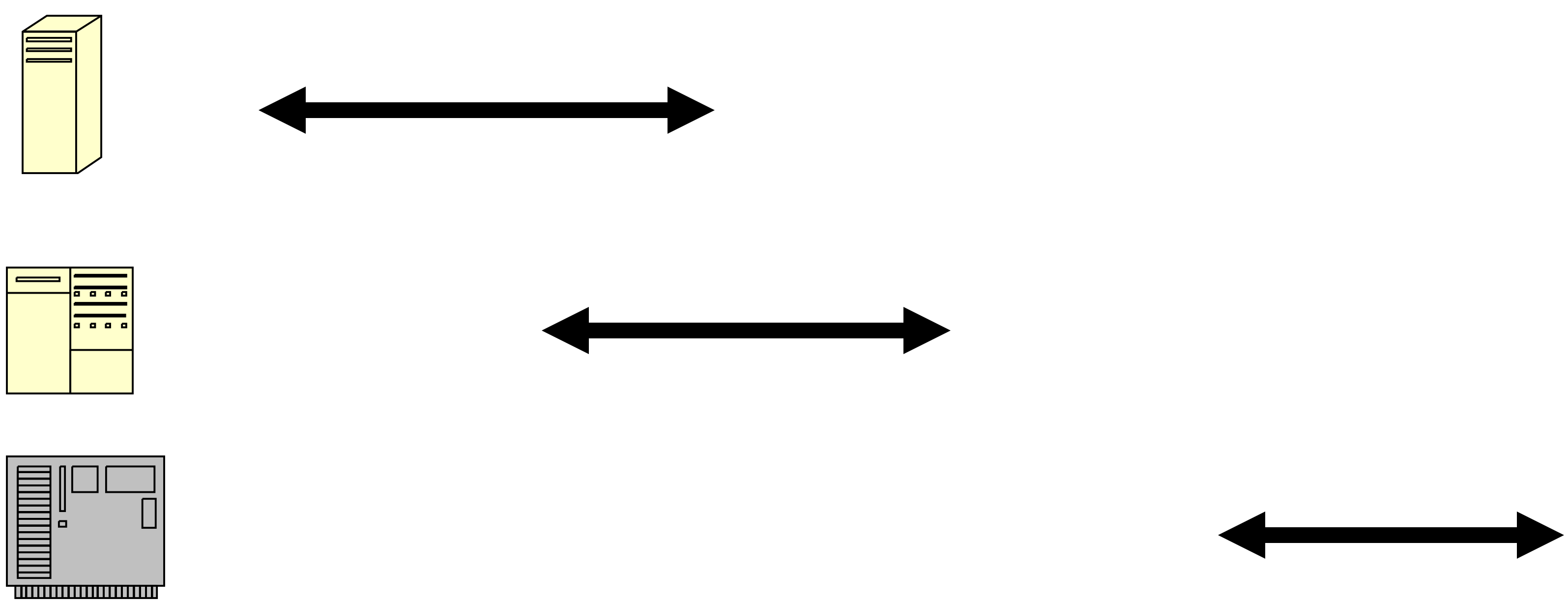
95% Confidence Interval for Mean of Differences

- $c_{1,2}$ = [-5.36, 3.36]
- Interval includes 0

→ With 95% confidence, there is *no statistically significant difference* between the two systems.

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*

Design of Experiments - 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)

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 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	

Sum of Squares of Differences

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

Difference of column mean with overall mean

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

Difference of each with column mean

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

Difference of each with Overall mean

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	$F_{[0.95;2,12]} = 3.89$		

Sa^2 = SSA/Deg freedom
Se^2 = SSE/Deg freedom

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.

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.

Example

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

$$\alpha_1 = -0.1735$$

$$\alpha_2 = -0.1441$$

$$\alpha_3 = 0.3175$$

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

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

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

Since this includes 0, no statistical difference

Example

- 90% confidence interval for contrast of [Sys1- Sys3]

- C[1-3] = -0.4910
- 90% Conf Interval = [-0.540, -0.442]

Since this does not include 0,
We cannot say there is no statistical difference

Full set of ANOVA Slides follow

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					
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

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					
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

Deviation From Column Mean

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

$$e_{ij} = \text{deviation of } y_{ij} \text{ from column mean}$$

$$= \text{error in measurements}$$

Error = Deviation From Column Mean

	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

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					
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

Deviation From Overall Mean

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

$$\alpha_j = \text{deviation of column mean from overall mean}$$

$$= \text{effect of alternative } j$$

Effect = Deviation From Overall Mean

	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

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: SSE

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

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

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

Sum of Squares of Differences: SSA

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

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

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

Sum of Squares of Differences: SST

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

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

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

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					
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

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

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 – α) * 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.

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*

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

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

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

$$df(s_c^2) = k(n-1)$$

- 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]

$$\alpha_1 = -0.1735$$

$$\alpha_2 = -0.1441$$

$$\alpha_3 = 0.3175$$

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

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

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

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?

Important Points

- Use contrasts to compare effects of subsets of alternatives