Comparing Systems Using Sample Data

1

Comparing alternatives

- Next: comparing two alternatives
  - use confidence intervals
- Comparing more than two alternatives
  - ANOVA
    - Analysis of Variance
  - Will discuss later this semester

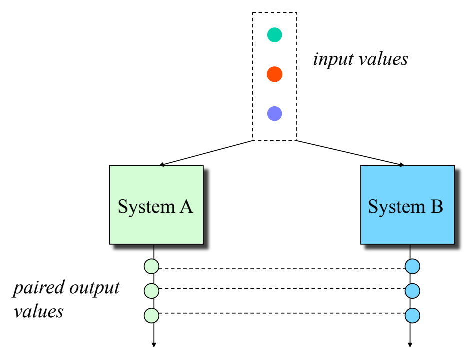
2

Comparing Two Alternatives

- Suppose you want to compare two cache replacement policies under similar workloads.
- Metric of interest: cache hit ratio.
- Types of comparisons:
  - Paired observations
  - Unpaired observations.

3

Paired Observations



4

Example of Paired Observations

- Six similar workloads were used to compare the cache hit ratio obtained under object replacement policies A and B on a Web server. Is A better than B?

Workload	Cache Hit Ratio		
	Policy A	Policy B	A-B
1	0.35	0.28	0.07
2	0.46	0.37	0.09
3	0.29	0.34	-0.05
4	0.54	0.60	-0.06
5	0.32	0.22	0.10
6	0.15	0.18	-0.03
	Sample mean		0.02000
	Sample variance		0.00552
	Sample standard dev.		0.07430

5

Example of Paired Observations

Sample mean	0.02000
Sample variance	0.00552
Sample standard dev.	0.07430

In Excel: TINV(1-0.9,5)

0.95 quantile of t-variable with 5 degrees of freedom 2.015
90% confidence interval
lower bound -0.0411
upper bound 0.0811

$$(\bar{x} - t_{[1-\alpha/2;n-1]}\frac{s}{\sqrt{n}}, \bar{x} + t_{[1-\alpha/2;n-1]}\frac{s}{\sqrt{n}})$$

0.0743 → s; 0.02 → $\bar{x}$; 2.015 → $t_{[1-\alpha/2;n-1]}$; 6 → n

6

Example of Paired Observations

Sample mean	0.02000
Sample variance	0.00552
Sample standard dev.	0.07430

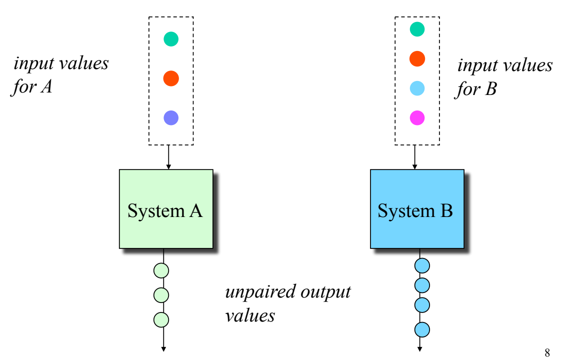
In Excel: TINV(1-0.9,5)

0.95 quantile of t-variable with 5 degrees of freedom 2.015
90% confidence interval
lower bound -0.0411
upper bound 0.0811

The interval includes zero, so we cannot say that policy A is better than policy B.

7

Unpaired Observations



8

Inferences concerning two means

- For large samples, we can statistically test the equality of the means of two samples by using the statistic

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1}} + \sqrt{\frac{\sigma_2^2}{n_2}}}$$

  - Z is a random variable having the standard normal distribution.
  - We need to check if the confidence interval of Z at a given level includes zero
  - We can approximate the population variances above with sample variances when n_1 and n_2 are greater than 30

9

Inferences concerning two means (cont'd)

- For small samples, if the population variances are unknown, we can test for equality of the two means using the t-statistic below, provided we can assume that both populations are normal with equal variances

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

  - t is a random variable having the t-distribution with $n_1 + n_2 - 2$ degrees of freedom and S_p is the square root of the pooled estimate of the variance of the two samples

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

10

Inferences concerning two means (cont'd)

- The pooled-variance t test can be used if we assume that the two population variances are equal
  - In practice, we can use it if one sample variance is less than 4 times the variance of the other sample
- If this is not true, we need another test
  - Smith-Satterthwaite test described on the following slides

11

Unpaired Observations (t-test)

1. Size of samples for A and B: n_A and n_B
2. Compute sample means:

$$\overline{x}_A = \frac{1}{n_A}\sum_{i=1}^{n_A} x_{iA}$$

$$\overline{x}_B = \frac{1}{n_B}\sum_{i=1}^{n_B} x_{iB}$$

12

Unpaired Observations (t-test)

3. Compute the sample standard deviations:

$$s_A = \sqrt{\frac{\left(\sum_{i=1}^{n_A} x_{iA}^2\right) - n_A(\bar{x}_A)^2}{n_A - 1}}$$

$$s_B = \sqrt{\frac{\left(\sum_{i=1}^{n_B} x_{iB}^2\right) - n_B(\bar{x}_B)^2}{n_B - 1}}$$

13

Unpaired Observations (t-test)

4. Compute the mean difference: $\bar{x}_a - \bar{x}_b$
5. Compute the standard deviation of the mean difference: $s = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$
6. Compute the effective number of degrees of freedom.

$$\nu = \frac{\left(s_a^2/n_a + s_b^2/n_b\right)^2}{\frac{1}{n_a - 1}\left(\frac{s_a^2}{n_a}\right)^2 + \frac{1}{n_b - 1}\left(\frac{s_b^2}{n_b}\right)^2}$$

14

Unpaired Observations (t-test)

7. Compute the confidence interval for the mean difference:

$$(\bar{x}_a - \bar{x}_b) \pm t_{[1-\alpha/2;\nu]} \times s$$

8. If the confidence interval includes zero, the difference is not significant at $100(1-\alpha)\%$ confidence level.

15

Example of Unpaired Observations

- Two cache replacement policies A and B are compared under similar workloads. Is A better than B?

Workload	Cache Hit Ratio	
	Policy A	Policy B
1	0.35	0.49
2	0.23	0.33
3	0.29	0.33
4	0.21	0.55
5	0.21	0.65
6	0.15	0.18
7	0.42	0.29
8		0.35
9		0.44
Mean	0.2657	0.4011
St. Dev	0.0934	0.1447

16

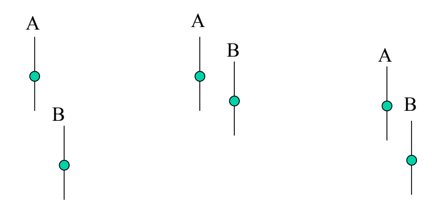
Example of Unpaired Observations

na	7		
nb	9		
mean diff	-0.135		
st.dev diff.	0.059776		
Eff. Deg. Freed.	13		
alpha	0.1	for	90% confidence interval
1-alpha/2	0.95		
t[1-alpha/2,v]	1.782287	← In Excel: TINV(1-0.9,13-1)	
90% Confidence Interval			
lower bound	-0.24193		
upper bound	-0.02886		

At a 90% confidence level the two policies are not identical since zero is not in the interval. With 90% confidence, the cache hit ratio for policy A is smaller than that for policy B. So, policy B is better at that confidence level.

17

Approximate Visual Test



CIs do not overlap: A is higher than B

CIs overlap and mean of A is in B's CI: A and B are similar

CIs overlap and mean of A is not in B's CI: need to do t-test

18

Example of Visual Test

Workload	Cache Hit Ratio	
	Policy A	Policy B
1	0.35	0.49
2	0.23	0.33
3	0.29	0.33
4	0.21	0.55
5	0.21	0.65
6	0.15	0.18
7	0.42	0.29
8		0.35
9		0.44
Mean	0.2657	0.4011
St. Dev	0.0934	0.1447

na	7		
nb	9		
alpha	0.1	for	90% confidence interval
1-alpha/2	0.95		
	Policy A	Policy B	
t[1-alpha/2,v]	1.9432	1.8595	
90% Confidence Interval			
lower bound	0.197	0.311	
upper bound	0.334	0.491	

CIs overlap but mean of A is not in CI of B and vice-versa. Need to do a t-test.

19

Non-parametric tests

- The unpaired t-tests can be used if we assume that the data in the two samples being compared are taken from normally distributed populations
- What if we cannot make this assumption?
  - We can make some normalizing transformations on the two samples and then apply the t-test
  - Some non-parametric procedure such as the Wilcoxon rank sum test that does not depend upon the assumption of normality of the two populations can be used

20

Rank-sum (Wilcoxon test)

- Non-parameteric test, i.e., does not depend upon distribution of population, for comparing two samples
- Example:
  - Suppose the time between two successive crashes are recorded for two competing computer systems as follows (time in weeks):
    System I: 0.63 0.17 0.35 0.49 0.18 0.43 0.12 0.20 0.47 1.36 0.51 0.45 0.84 0.32 0.40
    System II: 1.13 0.54 0.96 0.26 0.39 0.88 0.92 0.53 1.01 0.48 0.89 1.07 1.11 0.58
  - The problem is to determine if the two populations are the same or if one is likely to produce larger observations than the other

21

Rank-sum test (cont'd)

- U-test is a non-parameteric alternative to the paired and unpaired t-tests
- First step in the U-test is to rank the data jointly, in increasing order of magnitude

0.12	0.17	0.18	0.20	0.26	0.32	0.35	0.39	0.40	0.43
I	I	I	I	II	I	I	II	I	I
0.45	0.47	0.48	0.49	0.51	0.53	0.54	0.58	0.63	0.84
I	I	II	I	I	II	II	II	I	I
0.88	0.89	0.92	0.96	1.01	1.07	1.11	1.13	1.36	
II	II	II	II	II	II	II	II	I	

- Assign each data item a rank in this order
  - If there are ties among values, the rank assigned to each observation is the mean of the ranks which they jointly occupy

22

Rank-sum test (cont'd)

- The values in the first sample occupy ranks 1, 2,3,4,6,7,9,10,11,12,14,15,19,20 and 29
- The sum of the ranks for the two samples, W_1 = 162 and W_2 = 273
- The U-test is based on the statistics

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2}$$

or

$$U_2 = W_2 - \frac{n_2(n_2+1)}{2}$$

or on the statistic U which is the smaller of the two

23

Rank-sum test (cont'd)

- Under the null hypothesis that the two samples come from identical populations, it can be shown that the mean and variance of the sampling distribution of U_1 are

$$\mu_{U_1} = \frac{n_1 n_2}{2}$$

and

$$\sigma_{U_1}^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

- Numerical studies have shown that the sampling distribution of U1 can be approximated closely by the normal distribution when n1 and n2 are both greater than 8

24

Rank-sum test (cont'd)

- Thus, the test of the null hypothesis that both samples come from identical populations can be based on

$$Z = \frac{U_1 - \mu_{U_1}}{\sigma_{U_1}}$$

which is a random variable having approximately the standard normal distribution

- The alternative hypothesis is either:
  - Two-sided test (Populations are not identical)
    - We reject the null hypothesis if $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$
  - One-sided test
    - Population 2 is stochastically larger than Population 1
      - We reject the null hypothesis if $Z < -z_{\alpha}$
    - Or, Population 1 is stochastically larger than Population 2
      - We reject the null hypothesis if $Z > z_{\alpha}$

25

Example cont'd

- At the 0.01 level of significance, test the null hypothesis that the two samples in our example come from the same population
  - Alternative hypothesis, populations are not identical
  - For α = 0.01, we can reject the null hypothesis if Z < -2.575 or Z > 2.575
    - Calculations: n1 = 15, n2 = 14, W1 = 162
U1 = 162 - 15x16/2 = 42
Z = (42 - 15x14/2)/√((15x14x30)/12) = -2.75
  - Since Z is less than -2.575, we reject the null hypothesis; we conclude there is a difference between the two systems

26

ANOVA- Analysis of Variance

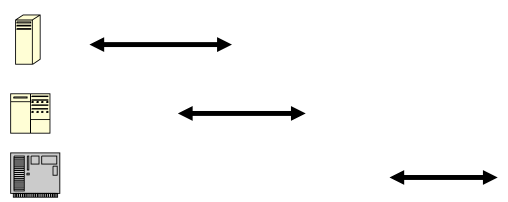
Comparing alternatives

- Comparing two alternatives
  - use confidence intervals
- Comparing more than two alternatives
  - ANOVA
    - Analysis of Variance

28

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*

30

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)?
- Want to determine whether variation on component (1) is larger then component (2)

31

ANOVA

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

32

Measurements for All Alternatives

Organize in a table

	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

33

Column Means

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

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

34

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

35

Overall Mean

- Average of all measurements made of all alternatives

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

36

Deviation From Column Mean

- For each column, we can write deviation from its that alternative's mean

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

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

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

37

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

38

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

39

Deviation From Overall Mean

- For each column mean, we can write deviation from it's the total mean

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

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

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

40

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

41

Effects and Errors

- Combining the two we can write each measurements as

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

- *Effect* is distance from overall mean
  - Horizontally across alternatives
- *Error* is distance from column mean
  - Vertically within one alternative
  - Error across alternatives, too

42

Sum of Squares of Differences: SSE

- We can split the measurements due to the total variation into two components – effect of alternatives and variation due to errors
- Variation due to errors

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

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

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

43

Sum of Squares of Differences: SSA

- Variation due to alternatives

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

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

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

44

Sum of Squares of Differences: SST

- Total variation

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

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

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

45

Sum of Squares of Differences

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

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

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

46

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

$$SST = SSA + SSE$$

47

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

48

Comparing SSE and SSA

- Simple approach – find the ratios
  - *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?
- Need a statistical test F-test

49

Variances from Sum of Squares (Mean Square Value)

- Estimate variances of SSA and SSE

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

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

50

Comparing Variances

- Use F-test to compare ratio of variances

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

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

51

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.

52

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.

53

Degrees of Freedom

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

54

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

55

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

56

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

57

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	$F_{[1-\alpha;(k-1),k(n-1)]}$		

58

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	

59

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

60

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.

61

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.

62

Contrasts

- Contrast = linear combination of *effects* of alternatives
- Contrast can be used to compare the effects of alternatives

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

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

63

Contrasts

- E.g. Compare effect of system 1 to effect of system 2 – choose the weights appropriately

$$w_1 = 1$$

$$w_2 = -1$$

$$w_3 = 0$$

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

$$= \alpha_1 - \alpha_2$$

64

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

65

Variance of random variables

- Recall that, for independent random variables X_1 and X_2

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

66

Variance of a contrast *c*

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

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

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

67

Confidence interval for contrasts

$$(c_1, c_2) = c \mp t_{1-\alpha/2;k(n-1)} s_c$$
$$s_c = \sqrt{\frac{\sum_{j=1}^{k}(w_j^2 s_e^2)}{kn}}$$
$$s_e^2 = \frac{SSE}{k(n-1)}$$

68

Example

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

$$\alpha_1 = -0.1735$$
$$\alpha_2 = -0.1441$$
$$\alpha_3 = 0.3175$$
$$c_{[1-2]} = -0.1735 - (-0.1441) = -0.0294$$
$$s_c = s_e \sqrt{\frac{1^2 + (-1)^2 + 0^2}{3(5)}} = 0.0275$$
$$90\% : (c_1, c_2) = (-0.0784, 0.0196)$$

69

Summary

- 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

70