UCLA STAT 251

Statistical Methods for the Life and Health Sciences

•**Instructor**: **Ivo Dinov,**
Asst. Prof. In Statistics and Neurology

University of California, Los Angeles, Winter 2002
http://www.stat.ucla.edu/~dinov/

ANOVA. The F-test.

- One-sample issues
- Two independent samples
- More than 2 samples
- Blocking, stratification and related samples

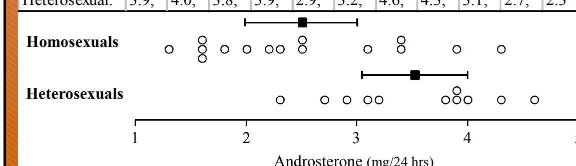
Paired Comparisons

1. What is a paired-comparison experiment? (obs'd data are matched in pairs).
2. In a paired-comparison experiment, why is it wrong to treat the two sets of measurements as independent data sets? (data are usually taken from the same unit under diff. Treatments, so obs's should be related).
3. How do you analyze the data from a paired-comparison experiment? (analyze the difference).
4. What situations is appropriate to use the paired-comparison method to analyze the data? (pre- and post-metrifonate study using FDG PET imaging).

Analysis of two independent samples

Urinary androsterone levels – data, dot-plots and 95% CI. Relations between hormonal levels and homosexuality, Margolese, 1970. Hormonal levels are lower for homosexuals. Samples are independent, as unrelated. Results, P-value of t-test 0.004 with a CI (μ_{Het}-μ_{Hom})=[0.4:1.7]. Normal hypothesis satisfied? Skewed?

Urinary Androsterone Levels(mg/24 hr)											
Homosexual:	2.5,	1.6,	3.9,	3.4,	2.3,	1.6,	2.5,	3.4,	1.6,	4.3,	2.0,
	1.8,	2.2,	3.1,	1.3							
Heterosexual:	3.9,	4.0,	3.8,	3.9,	2.9,	3.2,	4.6,	4.3,	3.1,	2.7,	2.3



Urinary androsterone levels cont.

Two Sample T-Test and Confidence Interval

```
Two sample T for androsterone
                N      Mean     StDev   SE Mean
hetero         11     3.518     0.721      0.22
homose         15     2.500     0.923      0.24
95% CI for mu (hetero) - mu (homose): ( 0.35,  1.69)
T-Test mu (hetero) = mu (homose) (vs not=):
                                    T=3.16  P=0.0044  DF=23
```

Confidence interval
t-test statistic *P-value*

Minitab 2-sample *t*-output for the androstenone data

Comparing two means for independent samples

Suppose we have 2 samples/means/distributions as follows: $\{\bar{x}_1, N(\mu_1,\sigma_1)\}$ and $\{\bar{x}_2, N(\mu_2,\sigma_2)\}$. We've seen before that to make inference about $\mu_1-\mu_2$ we can use a T-test for H_0: $\mu_1-\mu_2=0$ with $t_0 = \frac{(\bar{x}_1-\bar{x}_2)-0}{SE(\bar{x}_1-\bar{x}_2)}$

And CI$(\mu_1-\mu_2) = \bar{x}_1-\bar{x}_2 \pm t \times SE(\bar{x}_1-\bar{x}_2)$

If the 2 samples are independent we use the SE formula

$SE = \sqrt{s_1^2/n_1 + s_2^2/n_2}$ with $df = Min(n_1-1; n_2-1)$

This gives a conservative approach for hand calculation of an approximation to the what is known as the Welch procedure, which has a complicated exact formula.

Means for independent samples – equal or unequal variances?

Pooled T-test is used for samples with assumed equal variances. Under data Normal assumptions and equal variances of $(\bar{x}1-\bar{x}2-0)/SE(\bar{x}1-\bar{x}2)$, where

$$SE = sp\sqrt{1/n_1 + 1/n_2}\,;\; s_p^2 = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

is exactly Student's t distributed with $df = (n_1 + n_2 - 2)$

Here s_p is called the pooled estimate of the variance, since it pools info from the 2 samples to form a combined estimate of the single variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Another technique is to use the Welch unequal variance method.

Comparing two means for independent samples

1. How sensitive is the two-sample *t*-test to non-Normality in the data? (The 2-sample T-tests and CI's are even more robust than the 1-sample tests, against non-Normality, particularly when the shapes of the 2 distributions are similar and $n_1=n_2=n$, even for small n, remember $df= n_1+n_2-2$.
3. Are there nonparametric alternatives to the *two-sample t-test*? (Wilcoxon rank-sum-test, Mann-Witney test, equivalent tests, same P-values.)
4. What difference is there between the quantities tested and estimated by the two-sample *t*-procedures and the nonparametric equivalent? (Non-parametric tests are based on ordering, not size, of the data and hence use **median, not mean**, for the average. The equality of 2 means is tested and CI($\tilde{\mu}_1$ - $\tilde{\mu}_1$).

We know how to analyze 1 & 2 sample data. How about if we have than 2 samples – One-way ANOVA, *F*-test

One-way ANOVA refers to the situation of having one factor (or categorical variable) which defines group membership – e.g., comparing 4 reading methods, effects of different reading methods on reading comprehension, data: 50 – 13/14 y/o students tested.

Hypotheses for the one-way analysis-of-variance F-test

Null hypothesis: All of the underlying true means are identical.

Alternative: Differences exist between some of the true means.

Comparing 4 reading methods

Comparing 4 reading methods, effects of different reading methods on reading comprehension, data: 50 – 13/14 y/o students tested.

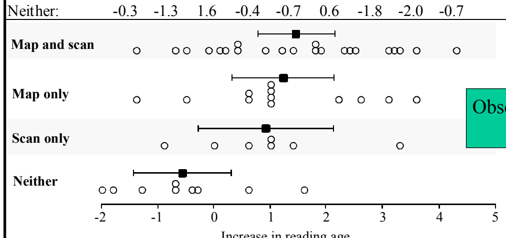
-Mapping: using diagrams to relate main points in text;
-Scanning: reading the intro and skimming for an overview before reading details;
-Mapping and Scanning;
-Neither.

Table below shows increases in test scores, of 4 groups of students taking similar exams twice, w/ & w/o using a reading technique.

Research question: Are the results better for students using mapping, scanning or both?

Increase in Reading Age

Both:	0.1	3.2	4.3	-0.5	1.9	3.3	2.5	3.6	0.4	2.3	-1.4	-0.7
	-0.1	0.2	0.4	0.9	1.2	1.4	1.8	1.8	2.4	3.1		
Map Only:	1.0	-0.5	1.0	0.6	0.6	1.0	1.0	-1.4	2.2	3.6	3.1	2.6
Scan Only:	1.0	3.3	1.4	-0.9	1.0	0.0	0.6					
Neither:	-0.3	-1.3	1.6	-0.4	-0.7	0.6	-1.8	-2.0	-0.7			

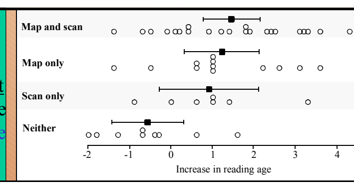


Observational study

Increases in reading ages with individual 95% CIs.

From *Chance Encounters* by C.J. Wild and G.A.F. Seber, © John Wiley & Sons, 2000.

The F-test indicates that there's real evidence true differences exist it does not give indication of where the differences are or how large they are.



One-way Analysis of Variance

```
Analysis of Variance for Increase
Source    DF      SS      MS      F       P
Grp        3   27.06    9.02   4.45   0.008
Error     46   93.35    2.03
Total     49  120.41
                           Individual 95% CIs For Mean
                           Based on Pooled StDev
Level      N    Mean   StDev  ------+---------+---------+---------+
MapScan   22   1.459   1.544                       (------*-----)
MapOnly   12   1.233   1.441                    (-------*--------)
ScanOnly   7   0.914   1.302               (----------*---------)
Neither    9  -0.556   1.135   (--------*---------)
                                ------+---------+---------+---------+
Pooled StDev =    1.425           -1.0       0.0       1.0       2.0
```

F-statistic *P-value* *Anova Table*

Show SYSTAT: DATA:
C:\Ivo.dir\Research\Data.dir\ReadingTechniquesStatData.sys

Interpreting the *P*-value from the *F*-test

(The null hypothesis is that all underlying true means are identical.)

- A ***large P-value*** indicates that the differences seen between the sample means could be explained simply in terms of sampling variation.
- A ***small P-value*** indicates evidence that real differences exist between **at least some** of the true means, but gives *no indication* of where the differences are or how big they are.
- ***To find out how big*** any differences are we need confidence intervals.

Form of a typical ANOVA table

Typical Analysis-of-Variance Table for One-Way ANOVA

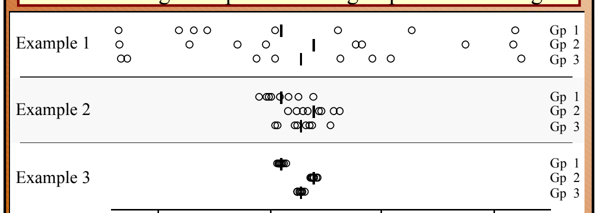
Source	Sum of squares	df	Mean sum of Squares[a]	F-statistic	P-value
Between	$\sum n_i(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$	k -1	s_B^2	$f_0 = s_B^2/s_W^2$	pr($F \geq f_0$)
Within	$\sum (n_i - 1)s_i^2$	n_{tot} - k	s_W^2		
Total	$\sum\sum (x_{ij}-\bar{x}_{\cdot\cdot})^2$	n_{tot} - 1			

[a] Mean sum of squares = (sum of squares)/*df*

- The *F*-test statistic, f_0, applies when we have independent samples each from ***k*** Normal populations, $N(\mu_i, \sigma)$, note same variance is assumed.

Where did the F-statistics came from?

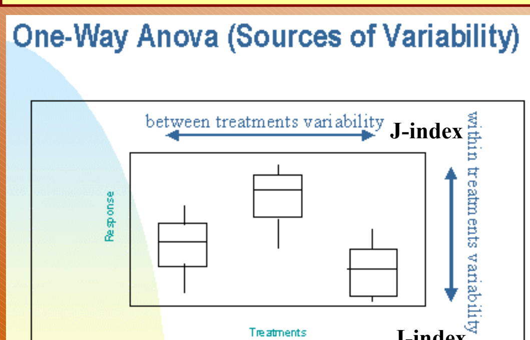
- Let's look at this example comparing groups. How do we obtain intuitive evidence against H_0? Far separated sample means + differences of sample means are large compared to their internal (within) variability! Which of the following examples indicate group diff's are "large"?



More about the *F*-test

- s^2_B is a measure of variability of sample means, how far apart they are.
$$s_B^2 = \frac{\sum n_i(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2}{k-1}$$
- s^2_W reflects the avg. internal Variability within the samples.
$$s_W^2 = \frac{\sum (n_i - 1)s_i^2}{n_{tot}-k}$$
- The *F*-test statistic, f_0, tests H_0 by comparing the variability of the sample means (numerator) with the variability within the samples (denominator).
- Evidence against H_0 is provided by values of f_0 which would be unusually large if H_0 was true.

What are x_i, $x_{\cdot\cdot}$, $x_{\cdot j}$, etc.?



ANOVA – the WM, GM, CSF volumes Manual vs. Automated extraction techniques.

- We have two ways of computing the WM, GM CSF volumes for MRI brain data:
  - Manual method – extremely labor intensive
  - Semi-automated – atlas based
- Ten individual's MRI volumes were segmented into the three different tissue types using methods 1 & 2.
- Results are in: C:\Ivo.dir\Research\Data.dir\WM_GM_CSF_tissueMaps.dir
- SYSTAT: ATLAS_IVO_all.xls (all 3 tissue types)
- DIR: C:\Ivo.dir\Research\Data.dir\WM_GM_CSF_tissueMaps.dir

What are x_i, $x_{..}$, $x_{.j}$, etc.?
Do the WM, GM, CSF volumetric measures!

Apple juice sales (units per week)

City 1 Convenience	City 2 Quality	City 3 Price
529	804	672
658	630	531
793	774	443
514	717	596
663	679	602
719	604	502
711	620	659
606	697	689
461	706	675
529	615	512
498	492	691
663	719	733
604	787	698
495	699	776
485	572	561
557	523	572
353	584	469
557	634	581
542	580	679
614	624	532

H_0: $\mu_1 = \mu_2 = \mu_3$

H_A: at least 2 means differ

$x_{i,j}$, $1 <= i <= n_j$; $1 < j <= 3$

What are x_i, $x_{..}$, $x_{.j}$, etc.?
Sum of Squares for treatments (cities)

$$SST = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2$$

$$SST = 20(577.55 - 613.07)^2$$
$$+ 20(653.00 - 613.07)^2$$
$$+ 20(608.65 - 613.07)^2$$
$$= 57{,}512.23$$

What are x_i, $x_{..}$, $x_{.j}$, etc.?
Sum of squares for the Error

Sum of Squares for Error: $SSE = \sum_{j=1}^{k} \left(\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \right)$

$$SSE = 19(10{,}774.44) + 19(7{,}238.61) + 19(8{,}669.47)$$
$$= 506{,}967.88$$

What are x_i, $x_{..}$, $x_{.j}$, etc.?
F-test

Test Statistic: $F = \frac{MST}{MSE} = \frac{SST/(k-1)}{SSE/(n-k)}$

$$= \frac{57{,}512.23/(3-1)}{506{,}967.88/(60-3)}$$

$$= 3.23$$

Rejection Region: $F > F_{\alpha, k-1, n-k} = F_{.05,2,57} = 3.15$

Conclusion: Reject H_0

What are x_i, $x_{..}$, $x_{.j}$, etc.?
One-Way Design ANOVA Table

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Statistic
Treatments	k-1	SST	MST	MST/MSE
Error	n-k	SSE	MSE	
Total	n-1	SS(Total)		

Note: MST=SST/(k-1)

MSE=SSE/(n-k)

F-test assumptions

1. Samples are independent, physically independent subjects, units, objects are being studies.
2. Sample Normal distributions, especially sensitive for small n_i, number of observations, $N(\mu_i, \sigma)$.
3. Standard deviations should be equal within all samples, $\sigma_1 = \sigma_2 = \sigma_3 = \ldots \sigma_{n_k} = \sigma$. $(1/2 <= \sigma_k/\sigma_j <= 2)$

How to check/validate these assumptions for your data?

For the reading-score improvement data:

- independence is clear since different groups of students are used.
- Dot-plots of group data show no evidence of non-Normality.
- Sample SD's are very similar, hence we assume population SD's are similar.

Bonferroni Correction

1. What if the number of comparisons, a positive integer number without decimals, is large? Bonferroni correction concerns the question if, in the case of more than one test in a particular study, the alpha level should be adjusted downward to consider chance capitalization/accumulation.
2. The alpha level is the chance taken by researchers to make a Type I error. The Type I (false-positive) error is the error of incorrectly declaring a difference, effect or relationship to be true due to chance producing a particular state of events.

Bonferroni Correction

1. Customarily the alpha level is set at 0.05, or, in no more than one in twenty statistical tests the test will show 'something' while in fact there is nothing. In the case of more than one statistical test the chance of finding at least one test statistically significant due to chance fluctuation, and to incorrectly declare a difference or relationship to be true, increases.
2. In five tests the chance of finding at least one difference or relationship significant due to chance fluctuation equals 0.22, or one in five. In ten tests this chance increases to 0.40, which is about one in two. Using the Bonferroni method the alpha level of each individual test is adjusted downwards to ensure that the overall risk for a number of tests remains 0.05. Even if more than one test is done the risk of finding a difference or effect incorrectly significant continues to be 0.05.

Bonferroni Correction

1. Although the logic is beautiful, there is a serious drawback. If the chance of incorrectly producing a difference, making a Type I error, on an individual test is reduced, the chance of making a Type II error is increased, that no effect or difference is declared, while in fact there is an effect. Thus, by reducing for individual tests the chance on type one errors, i.e. the chance of introducing ineffective medical treatments or ineffective improvements; the chance on a Type II errors is increased, i.e. the chance that effective treatments, effective educational methods, or improved production methods, are not discovered. So, when is Bonferroni correction used correctly and when is it used incorrectly? There are three basic scenarios.
2. Perneger, TV. What is wrong with Bonferroni adjustments. British Medical Journal 1998;136:1236-1238.
3. Sankoh AJ, Huque MF, Dubey SD. Some comments on frequently used multiple endpoint adjustments methods in clinical trials. Statistics in Medicine 1997;16:2529-2542.

Nonparametric (distribution-free) methods

- less sensitive to outliers
- do not assume any particular distribution for the original observations
- do assume random samples from the populations of interest
- measure of center is the median rather than the mean
- tend to be somewhat less effective at detecting departures from a null hypothesis and tend to give wider confidence intervals

Normal Theory Techniques – One sample methods

Two-sided *t*-tests and *t*-intervals for a single mean are
- quite robust against non-Normality
- can be sensitive to presence of outliers in small to moderate-sized samples

- One-sided tests are reasonably sensitive to skewness.
- Normality can be checked
  - Graphically: Normal quantile–quantile (Q-Q) plots
  - formally, e.g. the Kolmogoroff-Smirnof, Wilk-Shapiro tests.

Paired data

- We have to distinguish between independent and related samples because they require different methods of analysis.
- Paired data (Section 10.1.2) is an example of related data.
- With paired data, we analyze the differences
  - this converts the initial problem into a one-sample problem.
- The ***sign test*** and ***Wilcoxon rank-sum*** test are nonparametric alternatives to the one-sample or paired *t*-test.

2-sample *t*-tests and intervals for differences between means $\mu_1 - \mu_2$

Assume

- statistically independent random samples from the two populations of interest
  - both samples come from Normal distributions
- Pooled method also assumes that $\sigma_1 = \sigma_2$
  Welch method (unpooled) does not

Two-sample *t*-methods are

- remarkably robust against non-Normality
- can be sensitive to the presence of outliers in small to moderate-sized samples
- One-sided tests are reasonably sensitive to skewness.

- The ***Wilcoxon*** or ***Mann-Whitney*** test is a nonparametric alternative to the two-sample *t*-test.

More than two samples and the *F*-test

- For testing whether more than two means are different we use the *F*-test.
- The method of comparing several means is referred to as a ***one-way analysis of variance***.
- The formal null hypothesis (H_0) tested is that all k ($k \geq 2$) underlying population means μ_i are identical.
- The alternative hypothesis (H_1) is that differences exist between at least some of the μ_i's.

The *F*-test cont.

- The numerator of the *F*-statistic f_0 reflects how far apart the sample means are. The denominator reflects average variability within the samples
- Evidence against H_0 is provided by
  - sample means that are further apart than expected from the internal variability of the samples.
  - large values of the *F*-statistic.
- A small *P*-value demonstrates evidence that differences exist between some of the true means
  - To estimate the size of any differences we use confidence intervals

Assumptions of the *F*-test cont.

- Assumptions of the *F*-test
  - independent samples;
  - Normality;
  - equal population standard deviations.
- The test
  - is robust to non-Normality
  - is reasonably robust to differences in the standard deviations when there are equal numbers in each sample, but not so robust if the sample sizes are unequal
  - can be used if the usual plots are satisfactory and the largest sample standard deviation is no larger than twice the smallest
  - is not robust to any dependence between the samples.

2-Way ANOVA analysis

- Contrasts
- Multiple comparisons for means
- Multiple comparisons for pair-wise comparisons
- Simultaneous confidence intervals
- Sample size computations

2-Way ANOVA analysis

- **Definition**: In the one-way ANOVA layout, a **linear function of the sample means** $\mu_1, \mu_2, \ldots, \mu_n$ is

$$\theta = c_1\mu_1 + c_2\mu_2 + \ldots + c_n\mu_n$$

2-Way ANOVA analysis

- **Sampling distribution of linear function of sample means**: Let $\overline{Y}_1, \overline{Y}_2, \overline{Y}_3, \ldots, \overline{Y}_k$, be the means of independent random samples of sizes $n_1, n_2, n_3, \ldots, n_k$, with mean $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_1^2, \ldots, \sigma_k^2$,

Then let $\theta = c_1\mu_1 + c_2\mu_2 + \ldots + c_n\mu_n$

where $c_1, c_2, \ldots, c_k$, are known constants and

$$\hat{\theta} = c_o + c_1\overline{Y}_1 + c_2\overline{Y}_2 + \ldots + c_k\overline{Y}_k,$$

The sampling distribution of $\hat{\theta}$ is:

2-Way ANOVA analysis

- **Sampling distribution of linear function of sample means**:

$$\hat{\theta} = c_o + c_1\overline{Y}_1 + c_2\overline{Y}_2 + \ldots + c_k\overline{Y}_k,$$

Mean: $\mu_{\hat{\theta}} = \theta$

Variance: $\sigma_{\hat{\theta}}^2 = \frac{c_1^2 s_1^2}{n_1} + \frac{c_2^2 s_2^2}{n_2} + \ldots + \frac{c_k^2 s_k^2}{n_k}$

If target popul's are Normal, $\hat{\theta}$ is Normal, too.

2-Way ANOVA analysis

- **Inference about linear function of population means**:

$N = n_1 + n_2 + \ldots + n_k$

CI's: 100(1-a)% CI(θ), when common variances, σ.

$$\boxed{\hat{\theta} - t_{(N-k,a/2)}\hat{\sigma}_{\hat{\theta}} \le \theta \le \hat{\theta} + t_{(N-k,a/2)}\hat{\sigma}_{\hat{\theta}}}$$

where: $\hat{\theta} = c_o + c_1\overline{Y}_1 + c_2\overline{Y}_2 + \ldots + c_k\overline{Y}_k,$

$$\hat{\sigma}_{\hat{\theta}} = \sqrt{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \ldots + \frac{c_k^2}{n_k}\right) \times Mean\_S_{Within}^2}$$

2-Way ANOVA analysis

- **Inference about linear function of population means**:

Hypothesis Testing: H_o: $\theta = \theta_o$ can be tested by:

$$t = \frac{\theta - \theta_o}{\hat{\sigma}_{\hat{\theta}}} \sim t_{(N-k,\alpha/2)}$$

2-Way ANOVA analysis

- **Example linear function of population means**:

The following data come from a study investigating the fraction of antibiotics injected into the bloodstream which bind to serum proteins. (Bovine serum was used.)

Antibiotic	Binding Percentage	Sample mean
Penicillin G	29.6 24.3 28.5 32	28.6
Tetracyclin	27.3 32.6 30.8 34.8	31.4
Streptomycin	5.8 6.2 11 8.3	7.8
Erythromycin	21.6 17.4 18.3 19	19.1
Chloramphenicol	29.2 32.8 25 24.2	27.8

2-Way ANOVA analysis

- **Example linear function of population means**:

In the study, $n_T = 20$ independent samples of bovine serum were used. These were assigned at random to one of 5 antibiotic treatments in such a way that there would be **n=4** samples for each antibiotic. This experimental design is called a completely randomized design (CRD).

The idea is to compare the variability among these treatment means: (28.6; 31.4; 7.8; 19.1; 27.8)

2-Way ANOVA analysis

- **Example linear function of population means**:

For the binding fraction data, consider a test of the equality of the binding fractions of the first two antibiotics: Penicillin and Tetracyclin. This can be carried out by estimating the appropriate simple contrast:

$$\theta = \mu_1 - \mu_2 = (1)\mu_1 + (-1)\mu_2 + (0)\mu_3 + (0)\mu_4 + (0)\mu_5$$

$\theta = 28.6 - 31.4;$

ANOVA Table

Source	d.f.	Sum Square	Mean Square	F
Treatments	4	1481	370	41
Error	15	136	9.05	
Total	19	1617		

2-Way ANOVA analysis

- **Example linear function of population means**:

$$\theta = \mu_1 - \mu_2 = (1)\mu_1 + (-1)\mu_2 + (0)\mu_3 + (0)\mu_4 + (0)\mu_5$$

$\theta = 28.6 - 31.4$; $\theta\hat{} = -2.8$; Testing H_o: $\theta = \theta_o = 0$

$$\hat{\sigma}_{\hat{\theta}} = \sqrt{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \ldots + \frac{c_k^2}{n_k}\right) \times Mean\_S^2_{Within}} = \sqrt{\left(\frac{1^2}{4} + \frac{(-1)^2}{4}\right) 9.05} = 2.127$$

$$t = \frac{\theta - \theta_o}{\hat{\sigma}_{\hat{\theta}}}$$

$$t \sim t_{(20-4,\, 0.05/2)}$$

$$t = \frac{-2.8}{2.127} = -1.32$$

ANOVA Table

Source	d.f.	Sum Square	Mean Square	F
Treatments	4	1481	370	41
Error	15	136	9.05	
Total	19	1617		

2-Way ANOVA analysis

- **Definition**: In the one-way ANOVA layout, a linear function of the group means $\mu_1, \mu_2, \ldots, \mu_n$ of the form

$$\theta = c_1\mu_1 + c_2\mu_2 + \ldots + c_n\mu_n$$

where $c_1 + c_2 + \ldots + c_n = 0$ is called a **contrast**.

- **Definition**: c_k's are called **coefficients** in the contrast.
- **Definition**: Contrasts in which only two of the coefficients are nonzero (and are often **–½; +½**) are called **simple contrasts**.

2-Way ANOVA analysis

- **Definition**: An estimator for a contrast of interest can be obtained by substituting treatment group sample means $\bar{y}_i$ for treatment population means μ_i in the contrast :

$$\hat{\theta} = c_1\bar{y}_1 + c_2\bar{y}_2 + \ldots + c_n\bar{y}_n$$

- Example:

$$\hat{\theta} = \bar{y}_1 - \bar{y}_2; \quad \text{for} \quad \mu_1 - \mu_2 = 0.$$

Orthogonal contrasts

- **Definition**: Suppose we have 2 contrasts ($n_1 = n_2 = \ldots = n_k$):

$$\theta_1 = c_1\mu_1 + c_2\mu_2 + \ldots + c_n\mu_n$$

$$\theta_2 = d_1\mu_1 + d_2\mu_2 + \ldots + d_n\mu_n$$

The two contrasts θ_1 and θ_2 are **mutually orthogonal** if the products of their coefficients sum to zero: $c_1d_1 + c_2d_2 + \ldots + c_nd_n = 0$

- Consider several contrasts, say k of them: $\theta_1, \theta_2, \ldots, \theta_k$. The set is **mutually orthogonal** if all pairs are mutually orthogonal.

Orthogonal contrasts

- Examples : Which of these are orthogonal?

$(-1,1,0,0,0)$ and $(0,0,-1,1,0)$

$(1,-1/2,-1/2,0,0)$ and $(0,0,0,-1,1)$

$(-1,1,0,0,0)$ and $(0,-1,1,0,0)$

Orthogonal contrasts - importance

- Why are orthogonal contrasts of interest?
- Let $\{\theta_1\hat{}, \theta_2\hat{}, \ldots, \theta_k\hat{}\}$ be a set of (k-1) orthogonal contrasts (comparisons) between k smaple means and let SST be the treatment-sum-os-squares (between variability). Then

  $SST = SS[\theta_1\hat{}] + SS[\theta_2\hat{}] + \ldots + SS[\theta_{k-1}\hat{}]$
- I.E. between-treatment-sum-of-squares is subdivided (decomposed) into (k-1) terms which each provide variability info about observed diff's between 2 specific subgroups of treatment means.

Orthogonal contrasts - importance

- $SST = SS[\theta_1\hat{}] + SS[\theta_2\hat{}] + \ldots + SS[\theta_{k-1}\hat{}]$
- I.E. between-treatment-sum-of-squares is subdivided (decomposed) into (k-1) terms which each provide variability info about observed diff's between 2 specific subgroups of treatment means.
  - where $$SS[\hat{\theta}_i] = \frac{\hat{\theta}_i^{\,2}}{\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \ldots + \frac{c_k^2}{n_k}}$$

Contrasts

- Sums of squares for contrasts
- Multiple Comparisons
  - Scheffe
  - Bonferroni
  - Tukey

  Present from: ANOVA_Ch9.pdf

  C:\Ivo.dir\UCLA_Classes\Winter2002\Stat_M251\PDF_lectures

2-Way ANOVA

- Factorial designs: study designs where responses are measured at different combinations of levels of one or more experimental factors.
- Ex. Treatments {A, B, C} with levels $\{a_1, a_2, \ldots a_a\}$ $\{b_1, b_2, \ldots b_b\}$ and $\{c_1, c_2, \ldots c_c\}$, respectively – axbxc factorial experiment.
- Ex. {H=Hemisphere, T=TissueType, M=Method} for the human brain manual vs. automated delineations. H={L,R}; T={WM, GM, CSF}; M={Manual, Auto}.

2-Way ANOVA

- 3 types of Factorial Effects: simple, interaction, main.
- Ex. {H=Hemisphere, M=Method} for the human brain manual vs. automated delineations. H={L,R}; M={Manual, Auto}.
- Simple effects: Let μ_{ij} denote the expected response to treatment $h_i m_j$. Simple effect of **H** at level m_1 of **M** is defined by: $m[HM_1] = \mu_{21} - \mu_{11}$. This is the amount of change in the expected response when the level of **H** is changed from h_2 to h_1, and the level of **M** is fixed at m_1.

2-Way ANOVA

- Interaction effects: $\mu[HM] = 1/2(\mu[HM_2] - \mu[HM_1])$.
- Note: $\mu[HM] == 1/2(\mu[H_2M] - \mu[H_1M])$.
- There's no interaction between H & M ←→ $\mu[HM]=0$. $|\mu[HM]|$ measures the intensity-degree of interaction.
- Testing for interactions: H_o: $\mu[HM]=0$ vs. H_1: $\mu[HM]!=0$ E.Q. $\mu[HM] = ½\mu_{22} - ½\mu_{12} - ½\mu_{21} + ½\mu_{11}$;
- This contrast is estimated by:
  - $\mu\hat{}[HM] = ½\bar{Y}_{22} - ½\bar{Y}_{12} - ½\bar{Y}_{21} + ½\bar{Y}_{11}$;

2-Way ANOVA

- Ex. {H=Hemi, M=Method} for the human brain manual vs. automated delineations. H={L,R}; M={Manual, Auto}.
- Simple effects: Let μ_{ij} denote the <u>expected response</u> to treatment $h_i m_j$. Simple effect of **H** at

	Level of –	–Factor M	Simple Effects of M
Level of H	m_1	m_2	$\mu[H_i M]$
h_1	μ_{11}	μ_{12}	$\mu[H_1 M]=\mu_{12}-\mu_{11}$
H_2	μ_{21}	μ_{22}	$\mu[H_2 M]=\mu_{22}-\mu_{21}$
Simple effects of H	$\mu[HM_1] = \mu_{21}-\mu_{11}$	$\mu[HM_2] = \mu_{22}-\mu_{12}$	

2-Way ANOVA

- <u>**Main effects**</u>: $\mu[H] = ½(\mu[HM_2]+\mu[HM_1]) = =½\mu_{22}-½\mu_{12}+½\mu_{21}-½\mu_{11}$;
- Similarly: $\mu[M] = ½(\mu[H_2M]+\mu[H_1M]) = =½\mu_{22}+½\mu_{12}-½\mu_{21}-½\mu_{11}$;
- $\mu[H]$ is the avg. change in the expected response (population mean response) when the level of H goes from L→R.

Orthogonal contrasts

- **<u>Definition</u>**: Suppose we have 2 contrasts:

$$\theta_1 = c_1\mu_1 + c_2\mu_2 + ... + c_n\mu_n$$

$$\theta_2 = d_1\mu_1 + d_2\mu_2 + ... + d_n\mu_n$$

The two contrasts θ_1 and θ_2 are **mutually orthogonal** if the products of their coefficients sum to zero: $c_1 d_1 + c_2 d_2 + ... + c_n d_n = 0$

- Consider several contrasts, say k of them: $\theta_1, \theta_2, \ldots, \theta_k$. The set is **mutually orthogonal** if all pairs are mutually orthogonal.

Analysis of 2x2 Factorial Design

- <u>**First test if there is interaction between the 2 factors**</u>:
  - If there's statistically significant interaction → examine separately the simple effects for each factor;

    H_0: $\mu[HM]=0$ vs. H_1: $\mu[HM]$!= 0, where the interaction effect is measured by the contrast:

    $$\hat{\mu}[HM]= ½\bar{Y}_{22} -½\bar{Y}_{12} -½\bar{Y}_{21} + ½\bar{Y}_{11};$$

    - ❑ If there is interaction present (effects of **Hemi** on the **Methods** are significant) ➔ study the simple effects of the **Hemi** on each of the 2 **Methods**

$$\hat{\mu}[H_1M]=\bar{Y}_{12}-\bar{Y}_{11}; \quad \hat{\mu}[H_2M]=\bar{Y}_{22}-\bar{Y}_{21};$$

Analysis of 2x2 Factorial Design

- <u>**First test if there is interaction between the 2 factors**</u>:
  - If there's statistically significant interaction → examine separately the simple effects for each factor;
  - If there is <u>**no interaction**</u> make inference about each of the 2 main effects, using the following <u>**contrasts**</u>.

$$\hat{\mu}[H] = ½(\hat{\mu}[HM_2]+\hat{\mu}[HM_1]) = ½\bar{Y}_{22}-½\bar{Y}_{12}+½\bar{Y}_{21}-½\bar{Y}_{11};$$

$$\hat{\mu}[M]=½(\hat{\mu}[H_2M]+\hat{\mu}[H_1M]) = ½\bar{Y}_{22}+½\bar{Y}_{12}-½\bar{Y}_{21}-½\bar{Y}_{11};$$

Analysis of 2x2 Factorial Design

- <u>**How do we actually test these contrasts for significance?**</u>
  - As we've seen:
  - Two-sided T-test

$$\hat{\sigma}_{\hat{\theta}} = \sqrt{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + ... + \frac{c_k^2}{n_k}\right) \times Mean\_S^2_{Within}}$$

$$t = \frac{\theta - \theta_o}{\hat{\sigma}_{\hat{\theta}}} \sim t_{(N-k,\alpha/2)}$$

  - where $\theta = c_1\mu_1 + c_2\mu_2 + ... + c_k\mu_k$, and

$$\hat{\theta} = c_1\bar{Y}_1 + c_2\bar{Y}_2 + ... + c_k\bar{Y}_k$$

Analysis of 2x2 Factorial Design

- How do we actually test these contrasts for significance?
  - Two-sided T-test **E.Q.** to
  - One-sided F-test $\hat{\theta} = c_1\overline{Y}_1 + c_2\overline{Y}_2 + \ldots + c_k\overline{Y}_k$

$$\hat{\sigma}_{\hat{\theta}} = \sqrt{\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \ldots + \frac{c_k^2}{n_k}\right) \times Mean\_S^2_{Within}}$$

$$F_c = t_c^2 = \left(\frac{\hat{\theta}}{\hat{\sigma}_{\hat{\theta}}}\right)^2$$

$$F_c \sim F_{(df\_num=1,\ df\_deno=N-k-1,\ \alpha)}$$

ANOVA of 2x2 Factorial Design

- The significance of these contrasts? Use the F-test:
  - Effects coding used for categorical variables in model. Categorical values encountered during processing are:
  - METHOD (2 levels) 1, 2
  - HEMISPH (2 levels) 1, 2
  - Dep Var: VALUE N: 119

Analysis of Variance

Source	Sum-of-Sq's	df	Mean-Square	F-ratio	P
METHOD	2.97424E+08	1	2.97424E+08	0.39813	0.52931
HEMISPH	8.65479E+06	1	8.65479E+06	0.01159	0.91447
METH*HEMI	7.11598E+06	1	7.11598E+06	0.00953	0.92242
Error	8.59114E+10	115	7.47056E+08		

Not-Signif.
→ Main eff's

ANOVA of 2x2 Factorial Design

- The significance of these contrasts? Use the F-test:
  - Effects coding used for categorical variables in model. Categorical values encountered during processing are:
  - METHOD(2 levels);HEMISPH(2 levels); Dep Var: VALUE