EVALUATION OF MEDICAL TESTS AND TREATMENT

COMPARISONS

Kenneth Alonso, MD, FACP

When do serial values differ?

- The COEFFICIENT OF VARIATION is the standard deviation of the sample divided by the mean.
- A coefficient of variation (CV) of 10% is common in laboratory testing.)
- Electrolyte determinations employing ion specific electrodes have a CV of 1%.
- Automated cell counts and cell size determinations also have a CV of 1%.
- Thus, a difference in serial values greater than 2CV is generally regarded as significant.

- <u>Laboratory values are not absolute numbers.</u>

Comparison of groups

- Similar experiments, with similar null and alternative hypotheses, will be analyzed differently depending upon the property examined.
- If the property can be measured, it should be analyzed with a t-test or with an ANOVA.
- If the property is an attribute or a category, it should be analyzed with a Chi-square test. For a 2 x 2 contingency table, the formula follows:

$$\chi^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

Comparison of groups

- For a group of n=1, the paired t-test is chosen.
- To compare 2 groups, the unpaired t-test is often chosen.
- If 3 or more groups are to be compared, ANOVA is the procedure of choice.
- If multiple comparisons are to be made, ANOVA is repeated with each comparison.

Wilcoxson t-test

- Normal data distribution is assumed.
- Samples may be Independent (two randomly selected unrelated groups), or
- Dependent (two groups matched for some variable or repeated measurements on the same group).
- The degrees of freedom for the test are 2(n-1) where n is the sum of the number of participants. SDp is the pooled standard deviation.

$$t_{(n_1+n_1-2)} = \frac{(\bar{X}_1 - \bar{X}_2)}{SD_p\sqrt{[(1/n_1)+(1/n_2)]}}$$

$$SD_p = \sqrt{\frac{(n_1-1)\,SD_1^2+(n_2-1)\,SD_2^2}{n_1+n_2-2}}$$

ANOVA

- Normal distribution is assumed.
- Alternatively, each response can be ranked and ANOVA performed on rank-transformed data.
- This can reduce error in comparing samples that are not normally distributed.
- The degrees of freedom are 2n-1 where n is the sum of the number of participants.
- The F ratio is found by dividing the mean square among groups by the error mean square.

Pearson product moment correlation coefficient

- Indicates the strength and relationship of two random variables.
- Requires a normal distribution.
- A value of +1 means there is a perfect positive relationship between the two variables; -1, negative relationship; 0, no relationship.

$$r = \frac{\Sigma(X - \bar{X})\,(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2\,\Sigma(Y - \bar{Y})^2}}$$

If the distribution is not normal

- A CHI-SQUARE (χ^2) test is a better test to evaluate the strength of relationship between two variables if the distribution is not normal.
- The square of the correlation coefficient is known as the COEFFICIENT OF DETERMINATION, and is the fraction of the variance in y that is accounted for by a linear fit of x to y.

Size of study group

- Determine the p-value, probability, or Type I error rate (α).
- Determine the number of predictors.
- Determine the anticipated effect size, δ.
- By convention, 0.02, 0.15, and 0.35 are small, medium, and large, respectively.
- Determine the desired statistical power level. (Power, φ, is (1-Type II error rate, β)).
- By convention, this should be 0.80 or higher.
- The standard deviation between means, s, is roughly one-quarter of the mean difference.

Size of study group

- For one predictor the sample size needed, is:

  $n = (\Phi \times s / \delta \times \alpha)^2$

- A study with a yes/no outcome measure needs approximately 50 events to occur in the control group to have an 80% power of detecting a 50% relative risk reduction.
- If the control group risk is 20%, two groups of roughly 250 are required; if a 10% risk, 500; if a 5% risk, 1000; if a 1% risk, 5000.

Size of study group

- A study with a continuous outcome measure needs about 50 persons per group.
- The sample size required to detect minimum differences varies from 17 for 1 standard deviation; 33 for 0.7 standard deviation; 64 for 0.5 standard deviation; 175 for 0.3 standard deviation; 1571 for 0.1 standard deviation.

Is it useful?

- MEANS generally differ between groups and may be so as a result of chance.
- If confidence intervals do not overlap between groups, the two groups differ.
- Increasing the size of the group increases the likelihood that small differences will be detected.
- A twenty percent difference between groups can be demonstrated with a sample size of 50 patients.

Is it useful?

- The right question must be asked.
- Zidovudine was approved for AIDS treatment because of a difference of 11 AIDS defining events between the treated group of five hundred patients and that of the untreated group of similar size.
- Survival was not affected, not surprisingly, as mortality was not chosen as an end point.
- Quality of life diminished. That was not chosen as an endpoint.

HIV treatment results

- HAART treatment of HIV infection is very expensive. When does one initiate treatment?
- Early treatment of HIV infection with HAART is associated with increased life expectancy regardless of viral load if the patient is <30 years old and CD4 count is >200 cells/mm^3.
- Life expectancy ranges from 14.5 years if viral load >300,000 copies/ml
- (and rises) to 18.2 years if viral load <10,000 copies/ml and CD4 >500 cells/mm^3.

HIV treatment results

- Early treatment of HIV infection with HAART in patients OVER 40 YEARS OF AGE is associated with a life expectancy of 11.4 years if CD4 counts are >200cells/mm^3 AND viral loads are >300,000 copies/ml, rising to 12.9 years if CD4 counts are >500cells/mm^3 AND viral loads are <10,000 copies/ml.
- Little improvement in the life expectancy of 9.2 years of those patients OLDER THAN 50 YEARS is seen with early treatment with HAART.
- Whom do you treat? At what cost?