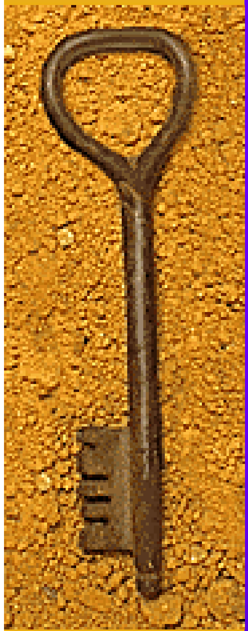
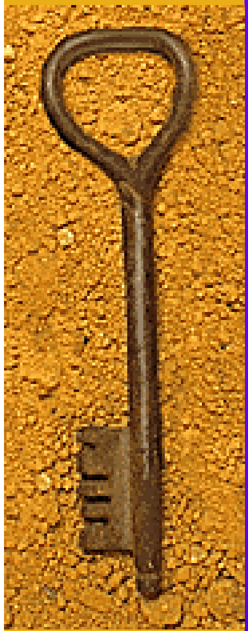


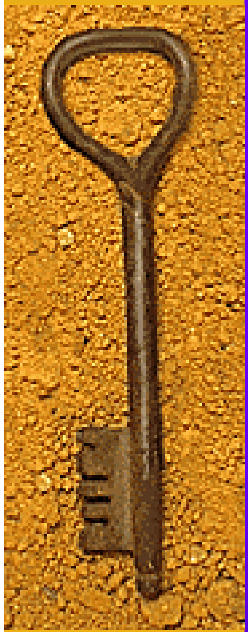
Inferential Statistics

- Allow researchers to generalize results to a population
- Results are based on a sample from the population
- The more the sample represents the population the more generalizable are the results



Why Inferential Statistics

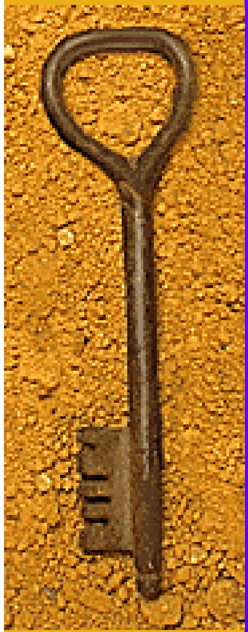
- To determine if results obtained from the sample are the same as those would have been by using the population



Statistics vs. Parameters

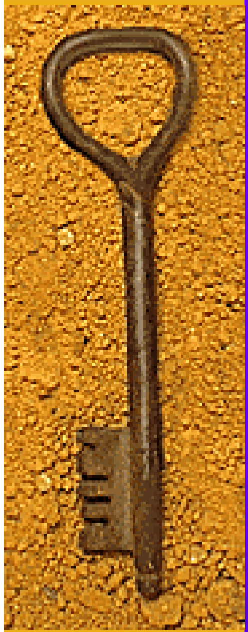
- Statistics come from samples
- Parameters are from population

- A mean from a sample is a statistic
- A mean from a population is a parameter



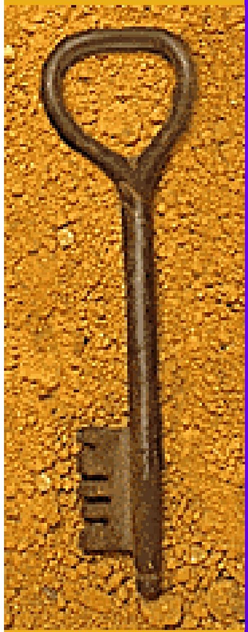
The Heart of Inferential Statistics

- How likely is it?
- Produces only probability statements about the population



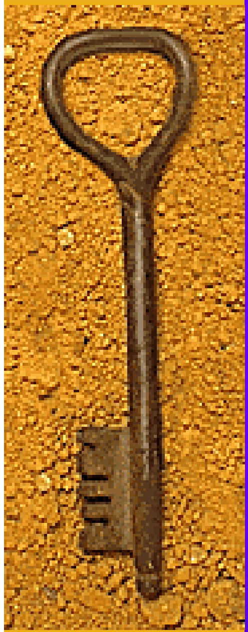
Sampling Error

- The expected random variation between the sample mean and the population mean
- The sample or samples and the population are almost never identical
- Not a result of researchers bias, design, or error



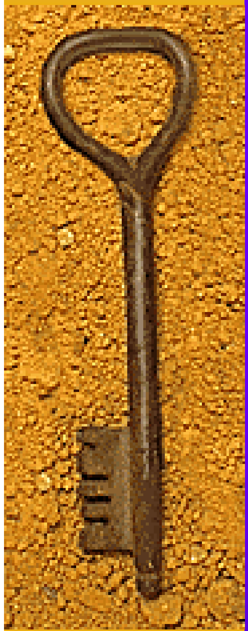
Mean of Means

- In a normal distribution of sample means there is also a mean (a mean of means)



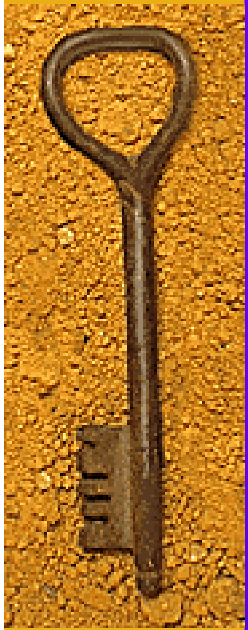
The Standard Error of the Mean

- The standard deviation of means in the distribution of sample means



Error

- An indication that in the sample means making up the distribution contain some estimate of the population mean

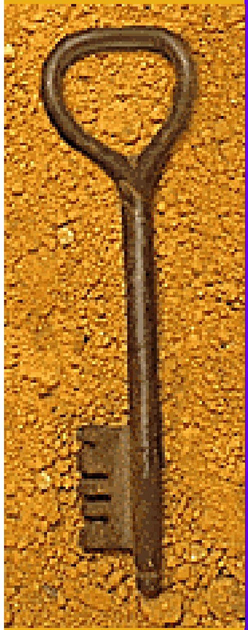


Standard Error of Mean

- $SE_{\bar{x}}$ tells how much the expected sample means would differ by using different samples from the population

Normal Curve Percentages

- Approximately 68% of sample means will be between =1 and –1 standard error of the mean (like standard deviation)
- 95% will be between +2 and –2 standard error of the mean
- 99+% will be between +3 and -3standard error of the mean



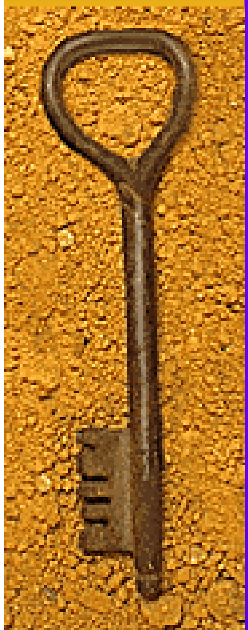
The Formula

- $$SE_{\bar{X}} = \frac{SD}{\sqrt{N-1}}$$

  $SE_{\bar{X}}$ = Standard Error of mean

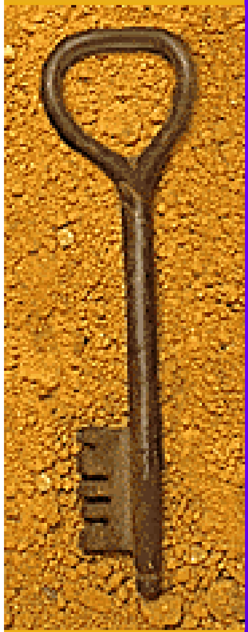
  SD = Standard Deviation for a sample

  N = Sample size



Confidence Limits

- The estimate of probable limits within which the population mean falls
- Use the estimate of the $SE_{\bar{X}}$ the sample mean, $\bar{X}$, and the normal curve to estimate
- As limits get further apart, confidence levels increase

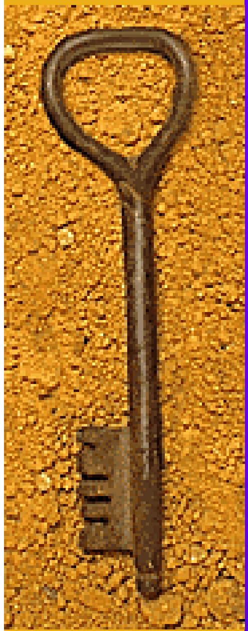


Sample Size

- The major factor affecting the standard error of the mean
- As sample size increases the standard error of the mean decreases
- A large sample more likely represents the population

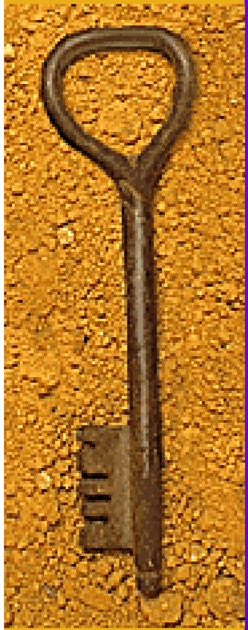
Population Standard Deviation

- Affects the standard error of the mean
- If deviation is large then population is spread out on variables – so should sample size
- Researchers can not control this deviation



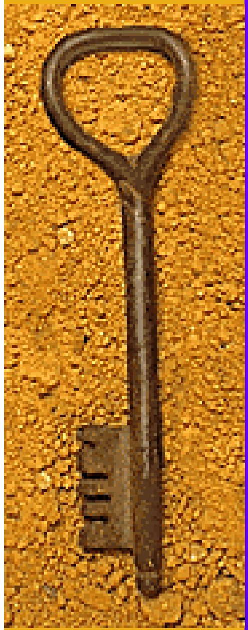
Estimate of Standard Error

- Can calculate for:
  - Mean
  - Variability
  - Relationships
  - Relative position



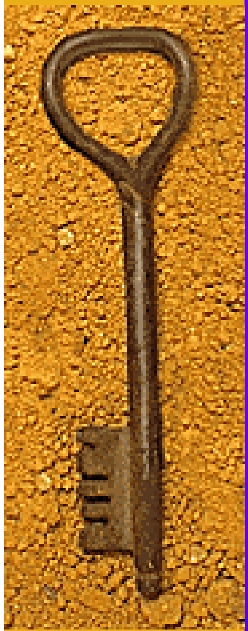
The Null Hypothesis

- The chance explanation for the difference between the two samples
- States that no true difference or relationship between parameters exists - difference is only in sampling error



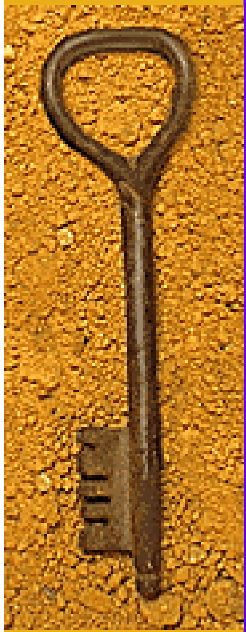
Research Hypothesis

- States that one method is expected to be better than another



The Null Hypothesis

- Rejection of null hypothesis is more conclusive support for a positive research hypothesis
- The test of significance to determine if the difference between the means is a true difference provides a test of the null hypothesis
- Null hypothesis is rejected as being probably false or not rejected as being probably true



Research Hypothesis

- If null hypothesis is rejected (mean for A greater than B) the research hypothesis is supported
- If null hypothesis is not rejected (A not greater that B) then research hypothesis not supported

Testing the Null Hypothesis

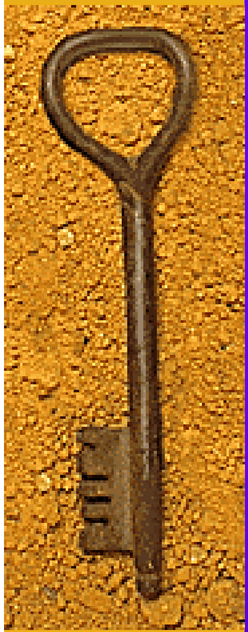
- Select a test of significance
- Select a probability level

Tests of Significance

- Helps decide if the null hypothesis can be rejected and infer that differences is significantly greater than that of chance
- Use a pre-selected probability level to determine to reject or fail to reject the null hypothesis
- Usually pre-selected levels are 5 out of 100, or 1 out of 100 chances

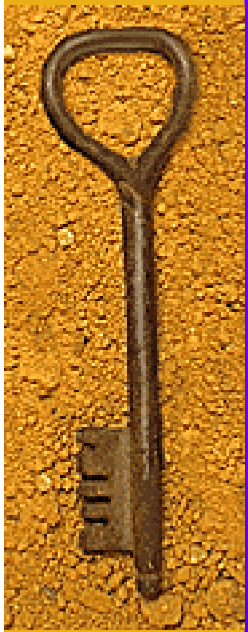
Statistical Tests of significance

- Depend on:
  - Scale of measurement represented by data
  - Method of participant selection
  - Number of groups
  - Number of independent variables



Type I and Type II Errors

- Four possible decisions. The null hypothesis is:
- True (A=B) and researcher agrees (Good!)
- False (A≠B) and researcher agrees (Good!)
- True (A=B) and researcher disagrees (Not Good!) Type I
- False (A≠B) and researcher disagrees (Not Good!) Type II



Probability Level

- Determines the probability of committing a Type I error
- $\alpha = .05$ means 5% probability of making a Type I error
- $\alpha = .01$ means 1% probability of making a Type I error

Type I vs. type II error

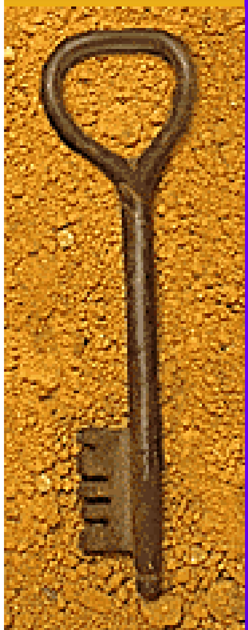
- As the chance of committing a Type I decreases the chance of committing a Type II error increases

Deciding Probability Levels

- Made before executing the study
- Made by carefully considering the relative seriousness of committing either of these types of errors

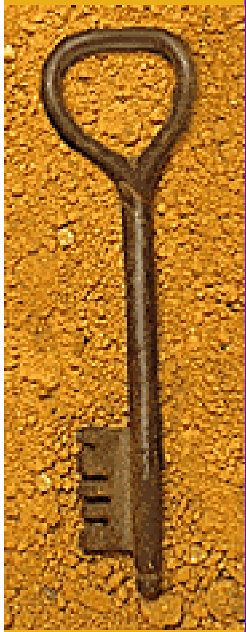
Two Tailed Tests

- Tests of significance are almost always two-tailed
- Null hypothesis states that there is no difference between groups
- Two-tailed allows for a possibility for difference in either direction – either group mean can be higher



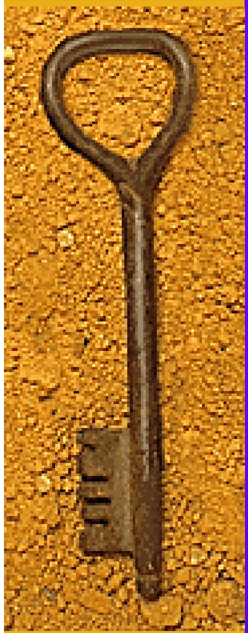
One Tailed Tests

- Allow for difference in only one direction
- Researcher must be reasonably certain a difference will occur in only one direction
- Has a major advantage: “easier” to obtain a significant difference
- See Figure 13.3 page 481



Degrees of Freedom

- Dependent on number of participants and number of groups
- The “control” to choose
- A choice of any 5 numbers is 5 degrees of freedom
- By limiting or controlling choice researcher limits or controls degrees of freedom

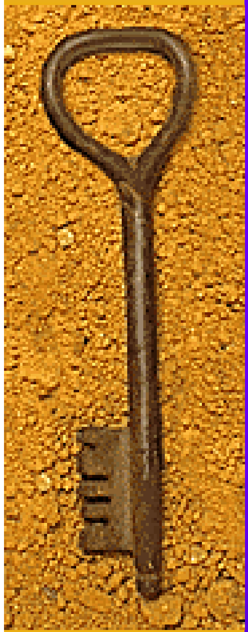


Tests of Significance (Types)

Depend on type of data

- Parametrics
  - Variable is normally distributed
  - Data represents interval or ratio scale
  - Random sample ratio of variance is known OR groups are equal

- Non-Parametrics
  - No assumption or not known
  - Data is ordinal or nominal
  - When parametric assumption has been greatly violated

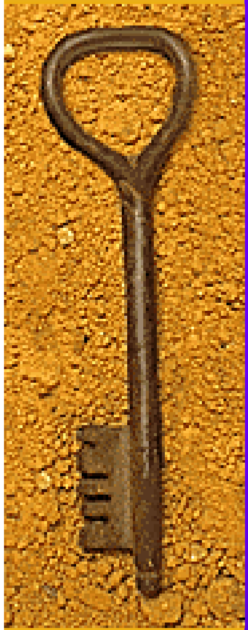


Parametric vs Nonparametric

- Parametric tests are more powerful – more likely to reject a null hypothesis
- Nonparametric uses a larger sample size to reach same level of significance
- Parametric tests hypothesis that nonparametric tests cannot

The *t* Test

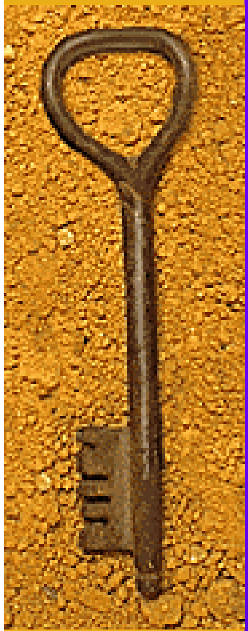
- Parametric – Nonparametric
- Used to determine if two means are significantly different at a selected probability level
- Adjusts scores of small sample sizes – *t* values required to reject a null hypothesis are higher for small samples



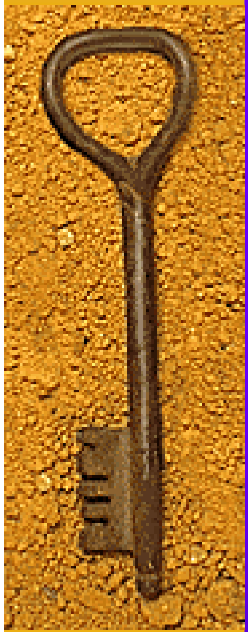
The *t* Test

- Compares actual mean difference to the difference expected by chance ($\bar{x}_1 - \bar{x}_2$)
- Involves forming a ratio of the means: the numerator is the between sample means and the denominator is the chance difference if the null hypothesis were true
- The denominator becomes the standard error of the difference between means

The *t* test for Independent Samples

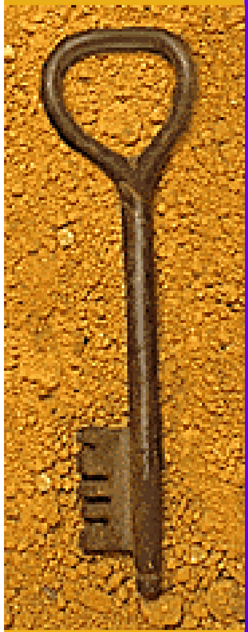


- Samples are not matched or related – except they are selected from same population
- The two means will be different
- *t* test helps determine if the difference a probable significant difference between the two means



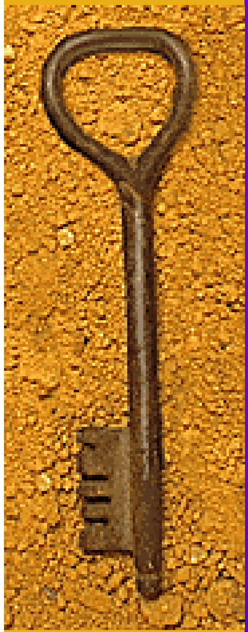
Statistics Symbols

α	Probability level
X	Score
t	*t* test result
$\bar{X}$	mean
SS	sum of squares
Σ	sum of
ΣX	add (or sum of) scores
N	number of participants
n	(w/ subscript) = the groups
SD	standard deviation
z	z score
p	Probability values
df	degrees of freedom



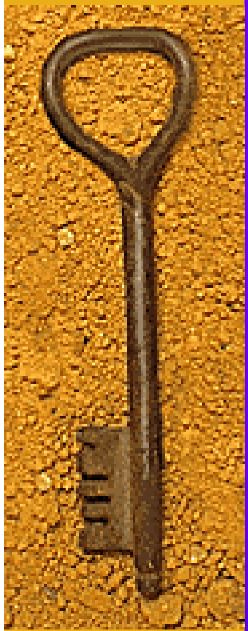
t Test

- Turn to pages 485 – 9 to calculate *t* test
- Use table A.4 to help determine the rejection of the null hypothesis



T Test Significance

- *t* test is either significant or not significant – not almost
- Negative values are not treated differently – look up number without regarding + or –
- The table is two tailed: knowing which group is which is important



t Test for Nondependent Samples

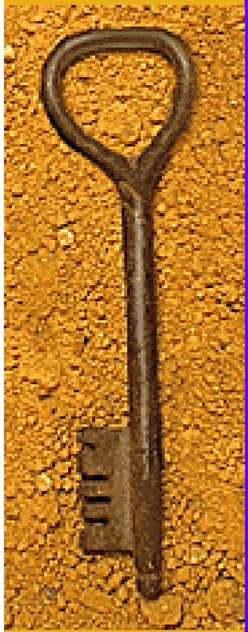
- Used to compare groups that are matched or a single group's scores on two treatments or on pre or post tests
- Scores are expected to be correlated
- The error term of the *t* test tends to be smaller
- D = Difference $\overline{D}$ = Difference of means
- Turn to page 489-90 for an example of the calculations
- Use Table A.4 to help determine statistical significance

Analysis of Gain

- To calculate
  - Subtract each participants scores on pretest from post test
  - Compute mean gain (or difference) for the group
  - Calculate a *t* value between the two average mean differences

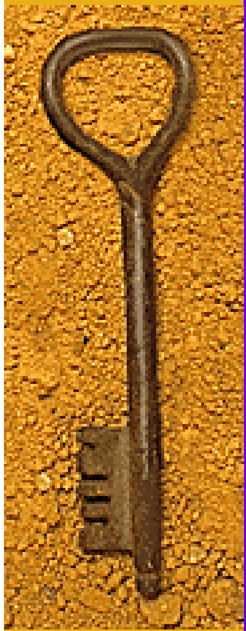
Analysis of Gain

- Problems
  - Not every participant has the same opportunity to gain - high scores on pretest cannot gain like lower scores
  - Gain or difference scores are less reliable



Analysis for Pretest – Post test Groups

- Depends on group's performance on pretest
- If scores on pretest are similar and neither group has received treatment use a *t* test
- If there is a difference use an analysis of covariance

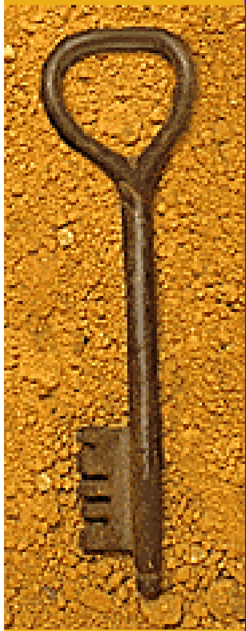


Simple Analysis of Variance (ANOVA)

- Helps determine whether a significant difference exists between two or more means at a selected probability level
- Most appropriate for scores of three or more groups
- Keeps error rate under control by using one test rather than several *t* tests

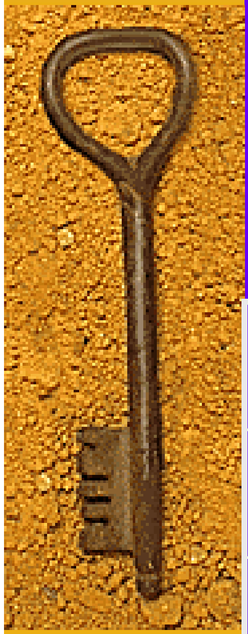
ANOVA

- Concept – the total variation (variance) comes from two sources
  - – Treatment (difference between groups)
  - – Error variance (variance within groups)
- The F ratio is used
  - – Treatment variance – numerator
  - – Error variance – denominator



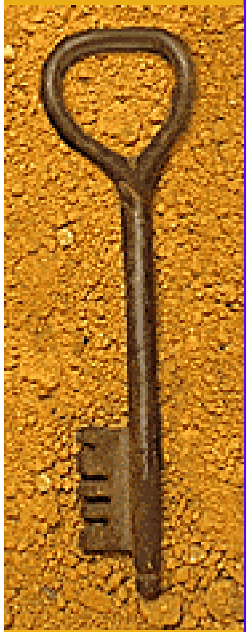
ANOVA Assumptions

- Groups are randomly formed
- Essentially same on dependent variable at beginning of study
- If treatment variance is significantly larger than error variance, there is a significant F ratio – the null hypothesis is rejected
- Turn to pages 492 – 5 to calculate ANOVA and Table A.5



More Symbols

K	number of treatment groups
ins	mean squares
MS_B	mean squares between groups
MS_W	mean squares within groups

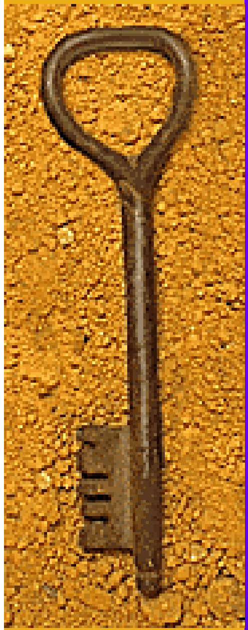


Multiple Comparisons

- All are special forms of t test
- Scheffé Multiple comparison
- Tukey’s HSD
- Duncan’s multiple range

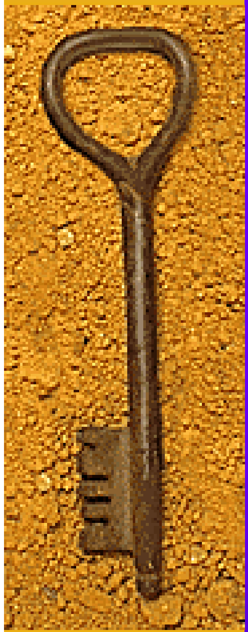
Scheffé Multiple Comparison

- Conservative test – less likely to commit a Type I error
- It is possible to find no significant difference even though the F ratio was significant
- Used to discover where the significant difference is between means
- Used to compare combinations of means



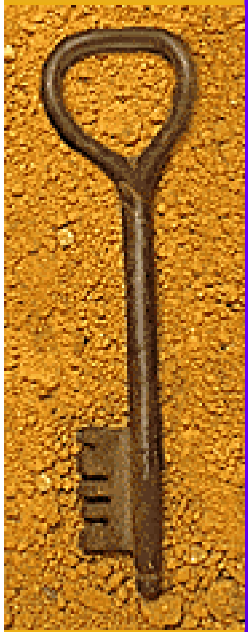
Scheffé Multiple Comparison

- Turn to pages 497 – 9 to calculate Scheffé Multiple Comparison



Factorial Analysis of Variance

- Provides a separate F ratio for each independent variable and for each interaction

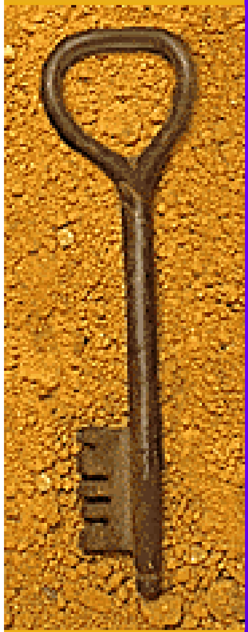


Analysis of Covariance

- Used to control extraneous variables
- Used to increase power

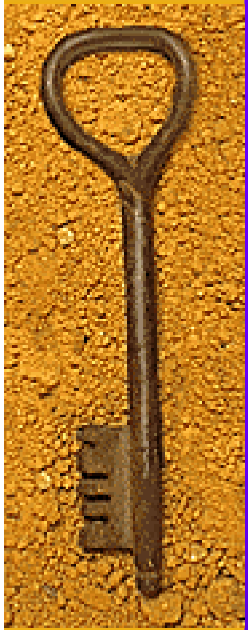
ANCOVA

- A form of ANOVA
- Statistical rather than experimental method to equate groups on one or more variables
- Adjusts post test scores for initial differences on a variable and compares the adjusted scores
- Used in causal – comparative and in experimental studies



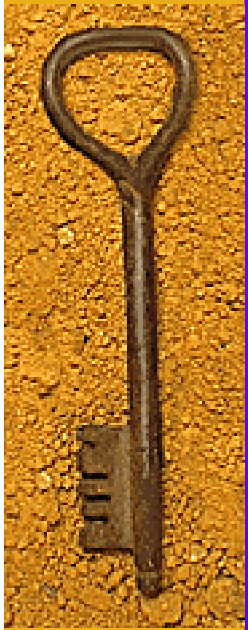
ANCOVA (cont.)

- If used with existing or already formed groups – results must be interpreted cautiously
- Increases in the power of statistical test by reducing within group variance
- Remember power means the ability to reject a false null hypothesis



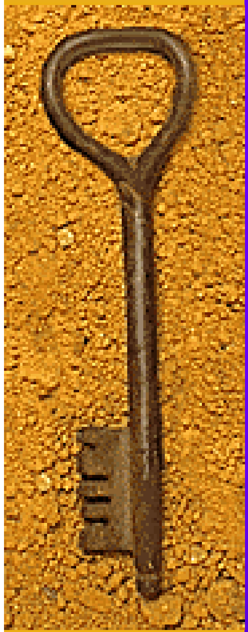
Multiple Regression

- A prediction equation that includes more than one predictor
- Uses variables that are known to individually predict (correlate) the criterion
- Can be used with data representing and scale of measurement
- Can be used to analyze results of experimental causal – comparative, and correlational studies
- Helps determine whether variables are related **and** degree of the relationship



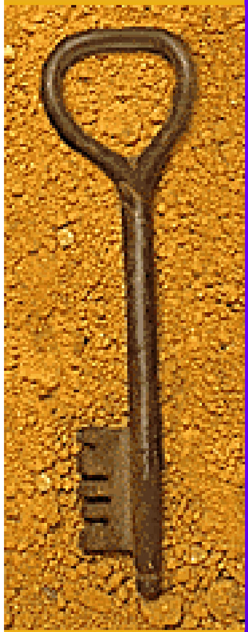
Steps in Multiple Regression

- Identify variable that best predicts criterion
- Identify variable that will most improve prediction
- The sign of the relationship (+ or -) does not indicate how good (the strength) of the prediction
- The larger the number of variables the greater the sample size needs to be



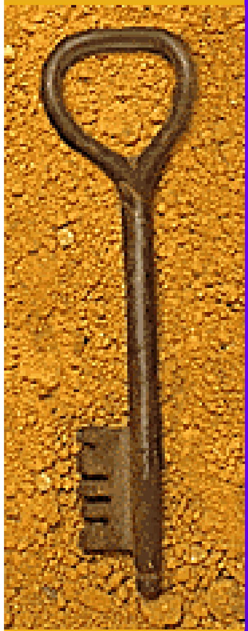
Chi Square

- Symbolized as X^2
- A nonparametric test
- Appropriate when data is in form of frequency counts and percentages
- Two or more mutually exclusive categories are required
- Appropriate for nominal data
- Appropriate when categories are true or artificial



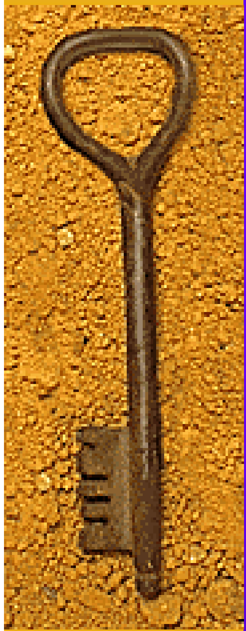
Chi Squared Test

- Compares proportions actually observed to those that were expected



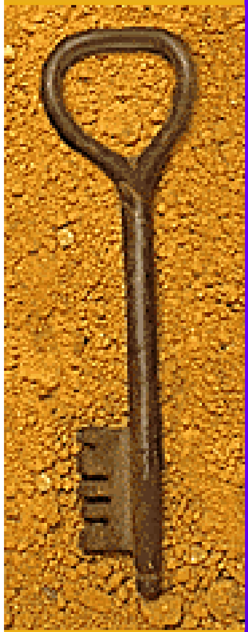
One Dimensional Chi Square

- Used to compare frequencies in different categories
- Data is presented in contingency table



Two Dimensional Chi Square

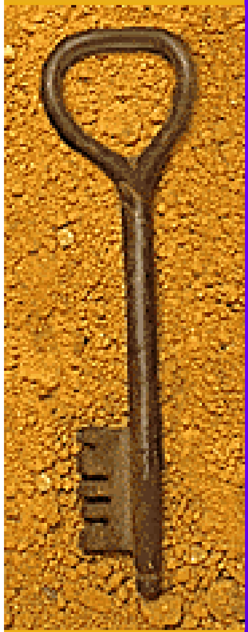
- Used when frequencies are categorized along more than one dimension
- A factorial chi square



Chi Square (cont.)

- Study pages 503 – 9 for explanation to calculations for chi square

Parametric and Nonparametric Tests



- Study Figure 13.5 page 509 for types of parametric and nonparametric tests