# Statistical Analyses

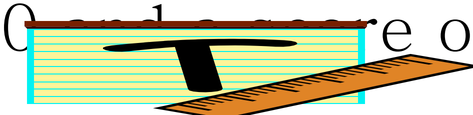
Dr. Saeid Eslami
EslamiS@mums.ac.ir

# The Scientific Methodologies assignment

- Devise short, multi-item, instruments (knowledge-attitude)
- Item-analysis
- Reliability-validity
- Using statistics "these research tools" to test some assumptions.
- Descriptive statistics
- Normality
- Parametric - Nonparametric Stats
- Correlation-Regression
- T-test – ANOVA-etc.

# Scales of Measurement

- Nominal scale:A nominal scale is really a list of categories to which objects can be classified.
- Ordinal scale:is a measurement scale that assigns values to objects based on their ranking with respect to one another.
- Interval scale: one unit on the scale represents the same magnitude on the trait or characteristic being measured across the whole range of the scale. For example, if anxiety were measured on an interval scale, then a difference between a score of 10 and a score of 11 would represent the same difference in anxiety as would a difference between a score of 50 and a score of 51.



- Ratio scale:Ratio scales are like interval scales except they have true zero points.

- **Nominal Data**
  - Sex-Nationality-
- **Ordinal Data**
  - ordered but differences between values are not important
  - e.g., political parties on left to right spectrum given labels 0, 1, 2
  - e.g., Likert scales, rank on a scale of 1..5 your degree of satisfaction
  - e.g., restaurant ratings
- **Interval Data**
  - ordered, constant scale, but no natural zero
  - Achievement scale
  - differences make sense, but ratios do not (e.g., 30°-20°=20°-10°, but 20°/10° is not twice as hot!
  - e.g., temperature (C,F), dates
- **Ratio Data**
  - ordered, constant scale, natural zero

# Parametric versus Nonparametric Statistics – When to use them and which is more powerful?

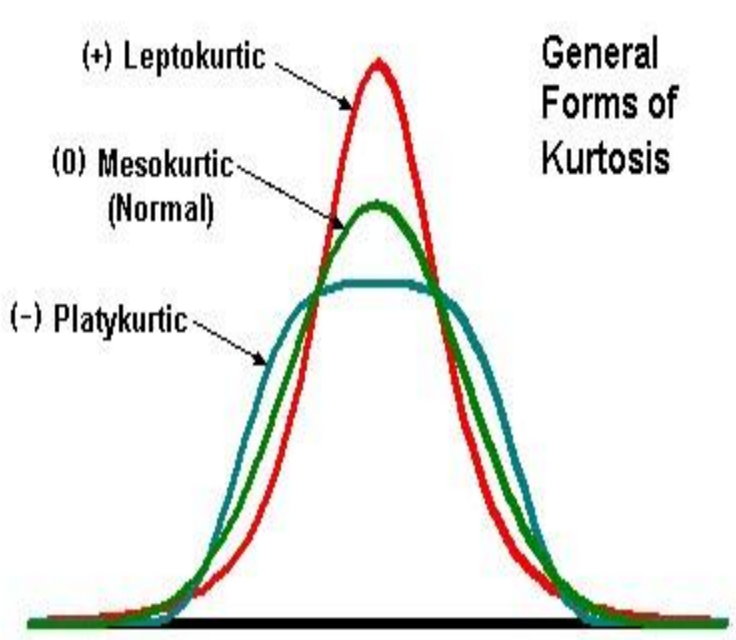
# Parametric Assumptions

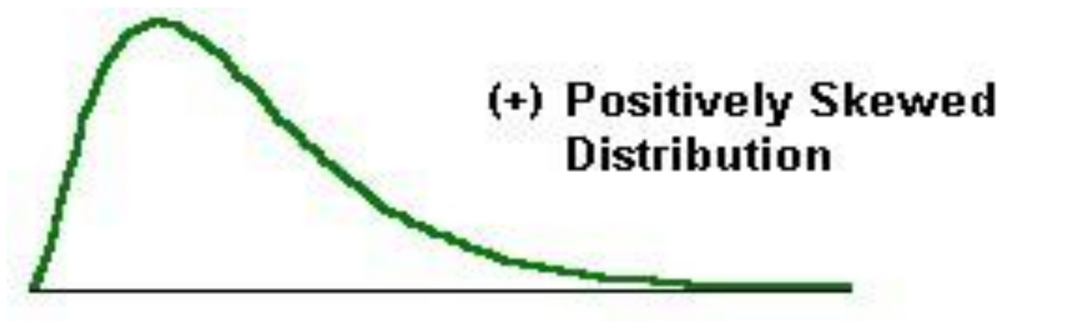
- The observations must be independent (For example participants need to have completed the dependent variable separately, not in groups).
- The observations must be drawn from normally distributed populations
- These populations must have the same variances

- **parametric test**, of course, is a test that requires a parametric assumption, such as normality. A **nonparametric test does not rely on parametric assumptions** like normality.
- a nonparametric test protects against some violations of assumptions and not others. The two sample t-test requires three assumptions, normality, equal variances, and independence. The non-parametric alternative, the Mann-Whitney-Wilcoxon test, does not rely on the normality assumption,

# Measures of Skewness and Kurtosis

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A statistical measure used to describe the distribution of observed data around the mean




# Differences between independent groups

- Two samples – compare mean value for some variable of interest

| Parametric | Nonparametric |
|---|---|
| t-test for independent samples | Wald-Wolfowitz runs test |
| | Mann-Whitney U test |
| | Kolmogorov-Smirnov two sample test |

# Differences between independent groups

- Multiple groups

| Parametric | Nonparametric |
| --- | --- |
| Analysis of variance (ANOVA/ MANOVA) | Kruskal-Wallis analysis of ranks |
| | Median test |

# Differences between dependent groups

- Compare two variables measured in the same sample
- If more than two variables are measured in same sample

| Parametric | Nonparametric |
|---|---|
| t-test for dependent samples | Sign test |
| | Wilcoxon's matched pairs test |
| Repeated measures ANOVA | Friedman's two way analysis of variance |

# Relationships between variables

- Two variables of interest are categorical

| Parametric | Nonparametric |
|---|---|
| Correlation coefficient | Spearman R |
| | Kendall Tau |
| | Coefficient Gamma |
| | Chi square |
| | Phi coefficient |
| | Fisher exact test |
| | Kendall coefficient of concordance |

# Summary Table of Statistical Tests

| Level of Measurement | Sample Characteristics | | | | | Correlation |
|---|---|---|---|---|---|---|
| | 1 Sample | 2 Sample | | K Sample (i.e., >2) | | |
| | | Independent | Dependent | Independent | Dependent | |
| Categorical or Nominal | X2 or bi-nomial | X2 | Macnarmar's X2 | X2 | Cochran's Q | |
| Rank or Ordinal | | Mann Whitney U | Wilcoxin Matched Pairs Signed Ranks | Kruskal Wallis H | Friendman's ANOVA | Spearman's rho |
| Parametric (Interval & Ratio) | z test or t test | t test between groups | t test within groups | 1 way ANOVA between groups | 1 way ANOVA (within or repeated measure) | Pearson's r |

(Plonskey, 2001)

# Parametric correlation

- Pearson Correlation. The most widely-used type of correlation coefficient is *Pearson r* (Pearson, 1896), also called *linear* or *product-moment* correlation (the term *correlation* was first used by Galton, 1888). Using non technical language, one can say that the correlation coefficient determines the extent to which values of two variables are "proportional" to each other.

# Assumptions of the Single Sample T-Test

- Normality: Assume that the population is distributed normally, ANOVA is quite robust over moderate violations of this assumption. Check for normality by creating a histogram.
- Independent Observations: The observations within each treatment condition must be independent of each other. For example participants need to have completed the dependent variable

# Paired-Samples *t* Test

Paired-Samples t Test

- Also known as the *t* test for dependent means
- Also known as the Dependent *t*-test

# Definitions for Paired-Samples *t* Test

<u>Definitions of Test</u>

- <u>Sample mean</u> is what researcher will find
  - The value (score) by using statistical analysis (mean)

- <u>Paired sample</u> = <u>paired scores</u>
- Paired = <u>matched</u> (they go together)

# Interpreting SPSS Output for *t* Test

## Interpreting Output

Ch. 12 Holcomb Paired-Samples *t* Test Output – Unformatted

**Paired Samples Statistics**

| | | Mean | N | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Pair 1 | Pretest | 11.5556 | 9 | 2.29734 | .76578 |
| | Posttest | 9.6667 | 9 | 2.23607 | .74536 |

**Paired Samples Correlations**

| | | N | Correlation | Sig. |
|---|---|---|---|---|
| Pair 1 | Pretest & Posttest | 9 | .527 | .145 |

**Paired Samples Test**

| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | 95% Confidence Interval of the Difference | | | | |
| | | Mean | Std. Deviation | Std. Error Mean | Lower | Upper | | | |
| Pair 1 | Pretest - Posttest | 1.88889 | 2.20479 | .73493 | .19414 | 3.58364 | 2.570 | 8 | .033 |

## One-Tailed and Two-Tailed Significance Tests

- When do you use a one-tailed or two-tailed test of significance?
- The answer is that it depends on your hypothesis.
- When your research hypothesis states the direction of the difference or relationship, then you use a one-tailed probability. For example, a one-tailed test would be used to test these null hypotheses: Females will not score significantly higher than males on an IQ test. In each case, the null hypothesis (indirectly) predicts the direction of the difference.
- A two-tailed test would be used to test these null hypotheses (no direction): There will be no significant difference in IQ scores between males and females. There will be no significant difference in the amount of product purchased

# Test for Significance

- <u>If your statistic is higher than the critical value from the table:</u>
- Your finding is significant.
- You reject the null hypothesis.
- The probability is small that the difference or relationship happened by chance, and p is less than the critical alpha level ($p < \alpha$).
- <u>If your statistic is lower than the critical value from the table:</u>
- Your finding is not significant.
- You fail to reject the null hypothesis.
- The probability is high that the difference or relationship happened by chance, and p is greater than the critical alpha level ($p > \alpha$).

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | 95% Confidence Interval of the Difference | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| Number of Older Siblings | Equal variances assumed | 1.669 | .203 | -1.461 | 44 | .151 | -.580 | .397 | -1.381 | .220 |
| | Equal variances not assumed | | | -1.612 | 31.607 | .117 | -.580 | .360 | -1.314 | .153 |

The columns labeled "Levene's Test for Equality of Variances" tell us whether an assumption of the t-test has been met. The t-test assumes that the variability of each group is approximately equal. If that assumption isn't met, then a special form of the t-test should be used. Look at the column labeled "Sig." under the heading "Levene's Test for Equality of Variances". In this example, **the significance (p value) of Levene's test is .203**. If this value is less than or equal to your **α level for the test (usually .05),** then you can **reject the null hypothesis** that the variability of the two groups is equal, implying that the variances are unequal. then you should use the bottom row of the output (the row labeled "Equal variances not assumed.") If the p value is greater than your α level, then you should use the middle row of the output (the row labeled "Equal variances assumed.") In this example, .203 is larger than α, so we will assume that the variances are equal and we will use the middle row of the output.

**Independent Samples Test**

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Confidence Interval of the Difference | |
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | Lower | Upper |
| Number of Older Siblings | Equal variances assumed | 1.669 | .203 | -1.461 | 44 | .151 | -.580 | .397 | -1.381 | .220 |
| | Equal variances not assumed | | | -1.612 | 31.607 | .117 | -.580 | .360 | -1.314 | .153 |

The column labeled "Sig. (2-tailed)" gives the two-tailed p value associated with the test. In this example, the p value is .151. the decision rule is given by: **If $p \leq \alpha$**, then reject H0. In this example, .151 is not less than or equal to .05, so we fail to reject H0. That implies that we failed to observe a difference in the number of older siblings between the two sections of this class.

If we were writing this for publication in an APA journal, we would write it as:

A *t* test failed to reveal a statistically reliable difference between the mean number of older siblings that the 10 AM section has (M = 0.86, s = 1.027) and that the 11 AM section has (M = 1.44, s = 1.318), *t*(44) = 1.461, *p* = .151, α = .05.

# Nonparametric Correlations

- The following are three types of commonly used nonparametric correlation coefficients (Spearman R, Kendall Tau, and Gamma coefficients). Note that the chi-square statistic computed for two-way frequency tables, also provides a careful measure of a relation between the two (tabulated) variables, and unlike the correlation measures listed below, it can be used for variables that are measured on a simple nominal scale.

# Reading regression output

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .629[a] | .395 | .345 | 3.360 |

a. Predictors: (Constant), words spoken per minute, IQ score, confidence in speaking time 1

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 265.441 | 3 | 88.480 | 7.837 | .000[a] |
| | Residual | 406.459 | 36 | 11.291 | | |
| | Total | 671.900 | 39 | | | |

a. Predictors: (Constant), words spoken per minute, IQ score, confidence in speaking time 1

b. Dependent Variable: attitude to school

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 3.379 | 3.073 | | 1.100 | .279 |
| | IQ score | .097 | .024 | .532 | 4.055 | .000 |
| | confidence in speaking time 1 | -.958 | .448 | -.355 | -2.138 | .039 |
| | words spoken per minute | .112 | .086 | .213 | 1.295 | .204 |

a. Dependent Variable: attitude to school

# Reporting Statistics in APA Style
## A Short Guide to Handling Numbers and Statistics in APA Format

- http://my.ilstu.edu/~mshesso/apa_stats_format.html
- Description: The 16 teenagers who volunteered for the pilot study were younger than expected, $M$ = 14.2 years, $SD$ = 1.3.
- Correlation: The correlation of peer reports ($M$ = 4.2, $SD$ = 2.1, $N$ = 367) and self reports ($M$ = 5.8, $SD$ = 2.3) of victimization was highly significant, $r(365)$ = .32, $p$ = .008.
- Regression: A linear regression analysis revealed that social skills was a highly significant predictor of aggression scores (b = .40, $p$ = .008), accounting for 16% of the variance in aggressive behavior.
- Achievement test scores were regressed on class size and number of writing assignments. These two predictors accounted for just under half of the variance in test scores ($R2$ = .49), which was highly significant, $F$(2,289) = 12.5, $p$ = .005. Both the writing assignment (b = .46, $p$ = .001) and

- t Tests: The 36 study participants had a mean age of 27.4 ($SD$ = 12.6) and were significantly older than the university norm of 21.2 years, $t(35) = 2.95$, $p = 0.01$.
- The 25 participants had an average difference from pre-test to post-test anxiety scores of −4.8 ($SD$ = 5.5), indicating the anxiety treatment resulted in a highly significant decrease in anxiety levels, $t(24) = -4.36$, $p = .005$ (one-tailed).
- The 36 participants in the treatment group ($M$ = 14.8, $SD$ = 2.0) and the 25 participants in the control group ($M$ = 16.6, $SD$ = 2.5), demonstrated a significant difference in performance ($t[59] = -3.12$, $p = .01$); as expected, the visual priming treatment inhibited performance on the phoneme