Basic Statistics Quik-Krib™ for Scientists, Engineers, and Healthcare Providers

Recommended Reading: **Naked Statistics: Stripping the Dread from the Data** by Charles Wheelan

BASICS

A POPULATION is a set of entities that share one or more attributes in common. When we measure something in a population, it is called a **PARAMETER**. When we measure something in a SAMPLE of that population, it is called a **STATISTIC**. Thus, *a parameter is to a population as a statistic is to a sample*.

Statistics are separated into two main areas: descriptive statistics and inferential statistics.

DESCRIPTIVE STATISTICS - These are **numbers** used to represent a **large amount of information**. Examples are: mean, mode, median, standard deviation, confidence interval, correlation, etc.

INFERENTIAL STATISTICS - These help us to draw **conclusions**. We use samples from the population to **INFER** conclusions about the population. If we want to determine if some treatment (Rx) is better/worse than another or if there are differences in how two groups perform, we use inferential statistics. The workhorse of inferential statistics is hypothesis testing. Every conjecture (**NULL** hypothesis) has a contradictory alternate (**ALTERNATE** hypothesis). Hypothesis testing is accomplished by designing experiments that are intended to DISPROVE a conjecture (drugs effects are same OR device is defective). Experimental designs that attempt to PROVE the conjecture (this drug is better OR device works fine) are <u>defective designs</u> that suffer from confirmation bias.

DESCRIPTIVE STATISTICS

STATISTIC	EXPLANATION
MEAN	The average of a set of values; only applies to interval and ratio data
MEDIAN	The middle value of an ordered set of values
MODE	The value that is seen most often in the ordered set
VARIANCE	A measure of how far the set of values are spread out; it shows dispersion
STANDARD DEVIATION	SD is square root of variance; it has the same physical units as the data values
CONFIDENCE INTERVAL	A range of values that are good estimates of the unknown population parameter
CORRELATION	A measure of observed dependence of two variables; unrelated to causation!
CHRONBACH'S alpha	A measure of inter-correlation among test items intended to measure same thing

ADMONITION: When confidence intervals DO NOT overlap, we can say that results ARE different. When standard deviations overlap, results are NOT different, but when confidence intervals overlap we cannot say that!!!

MEASUREMENT: Four (4) types of measurement SCALES. Very important to be aware of the correct measurement scale for your data; the type of measurement scale determines the permissible descriptive and inferential statistics (see table below). Using statistics that are not permissible will result in wrong conclusions about your data and bad things may happen, if you and others rely on those conclusions.

SCALE TYPE	ATTRIBUTES	EXAMPLE	PERMISSIBLE STATISTICS
NOMINAL	Labels ONLY	subject names, demographics	Mode, Chi-square (X^2), Contingency coefficient
ORDINAL	Labels, Order,	Likert-like scales	Median, Mode, Percentile, Correlation (e.g., Spearman), Analysis of Variance (e.g., Friedman)
INTERVAL	Labels, Order, Defined Intervals, but an ARBITRARY Zero	temperature (°C/F), calendar dates	Mean, Median, Mode, Standard Deviation, Correlation (e.g., Pearson), Regression, Analysis of Variance
RATIO	Labels, Order, Defined Intervals, and a TRUE Zero	length, weight subject ages	All the ABOVE plus: Geometric & Harmonic Means, Coefficient of Variation, Logarithms

Measurements are never perfect; they always have some uncertainty that results in variation of the data. When we use measurements as a test for a yes/no or presence/absence (binomial) result, test may be characterized by two important statistics:

- **Specificity**: the proportion of things that you were looking for that were actually there;
- **Sensitivity**: the proportion of things that you were NOT looking for that were excluded.

		TEST RESULTS	
		Test = Positive	**Test = Negative**
REALITY	**Positive**	**True Positive (TP)**	*False Negative (FN)* <u>- Test Failed -</u>
	Negative	*False Positive (FP)* <u>- Test Failed -</u>	**True Negative (TN)**

Sensitivity = $\frac{\text{Found by Test}}{\text{All that were really there}}$ = TP / (TP + FN)

Specificity = $\frac{\text{Excluded by Test}}{\text{All that were NOT there}}$ = TN / (TN + FP)

Basic Statistics Quik-Krib™ for Scientists, Engineers, and Healthcare Providers
Recommended Reading: **Naked Statistics: Stripping the Dread from the Data** by Charles Wheelan

INFERENTIAL STATISTICS

When we want to make a statement about the POPULATION, but can only access one or more sets of SAMPLES of the population, we use inferential statistical test to draw conclusions from the DATA. Four (4) interrelated study attributes influence statistical inference:

- Sample Size (**n**): number of units accessible for study
- Effect Size (**SNR**): signal-noise ratio, size of Rx-effect relative to noise
- Statistical Significance LEVEL (**α**): probability of false positive
- Statistical Power LEVEL (**β**): probability of false negative

Typically use **α**=0.05, **β**=0.2, and **SNR** to get minimum required **n**.
Estimate of the SNR typically comes from PILOT study or PRIOR knowledge.
Measurements **DO NOT ALWAYS** tell us what we want!

		Hypothesis Chosen	
		Accept H_0 (NULL Hypothesis)	Accept H_1 (ALTERNATE Hypothesis)
REALITY	H_0 True (Drug/Device Defective)	**CORRECT! (1-α)** 1-α=95% iff α = 0.05; 1-α = significance	Type I (α) Error; Consumer's Risk; FALSE POSITIVE; ACCEPT H_1 when H_1 FALSE
	H_1 True (Drug/Device Works)	Type II (β) Error; Provider's Risk; FALSE NEGATIVE; ACCEPT H_0 when H_0 FALSE	CORRECT! (1-β) 1-β=80% iff β=0.2; 1-β = power

Assumption always is Drug, Device, Procedure Does NOT work; Need to prove it does!

- **Sensitivity (1-β)** tells us how good test is at finding what we are looking for;
- **Specificity (1-α)** tells us how good test is at excluding things we do not want (artifacts, noise, wrong answers, etc.)

SELECTING INFERENTIAL TESTS	**TYPE OF STUDY DATA** Reference: www.graphpad.com/www/book/Choose.htm			
GOAL	**NORMAL POPULATION**	**RANK, SCORE, OR NON-NORMAL POPULATION**	**BINOMIAL (Y/N or T/F)**	**SURVIVAL TIME**
Describe 1 group	Mean, SD	Median, interquartile range	Proportion	Kaplan-Meier survival curve
Compare 1 group to a hypothetical value	1-sample *t* test	Wilcoxon signed-rank test	Chi-square or Binomial test	
Compare 2 UNPAIRED groups	UNPAIRED *t* test	Mann-Whitney U test	Fisher's test (chi-square for large samples)	Log-rank test or Mantel-Haenszel
Compare 2 PAIRED groups	PAIRED *t* test	Wilcoxon signed-rank test	McNemar's paired Δ test	Conditional proportional hazards regression
Compare 3 or more UNMATCHED groups	1-way ANOVA (F-test)	Kruskal-Wallis 1-way ANOVA test	Chi-square test	Cox proportional hazard regression
Compare 3 or more MATCHED groups	1-way repeated-measures ANOVA	Friedman test	Cochrane's Q test	Conditional proportional hazards regression
Quantify association between two variables	Pearson's product-moment correlation	Spearman's rank correlation	Contingency coefficients	
Predict value from another measured variable	Linear or nonlinear regression	Nonparametric regression	Simple logistic regression	Cox proportional hazard regression
Predict value from several measured or binomial variables	Multiple linear or nonlinear regression		Multiple logistic regression	Cox proportional hazard regression

NOTE: Nominal and ordinal data usually require "non-parametric" (i.e., distribution-independent) statistics; interval and ratio data, IF they are normally-distributed, can use parametric statistics. You can always use non-parametric statistics; BUT, if you use parametric statistics on non-normal data, you are violating fundamental assumptions used to derive the statistical test and you can get erroneous (& really **WEIRD**) results!

TYPES OF HYPOTHESIS TESTING STUDY VALIDITY: **INTERNAL** (degree to which causal conclusion is warranted; bias has been minimized); **CONCLUSION** (degree to which conclusions about relationships in data are reasonable); **CONSTRUCT** (degree to which inferences made be legitimately made FROM operationalizations TO theoretical constructs); **EXTERNAL** (degree to which study results may be generalized FROM sample(s) TO population).