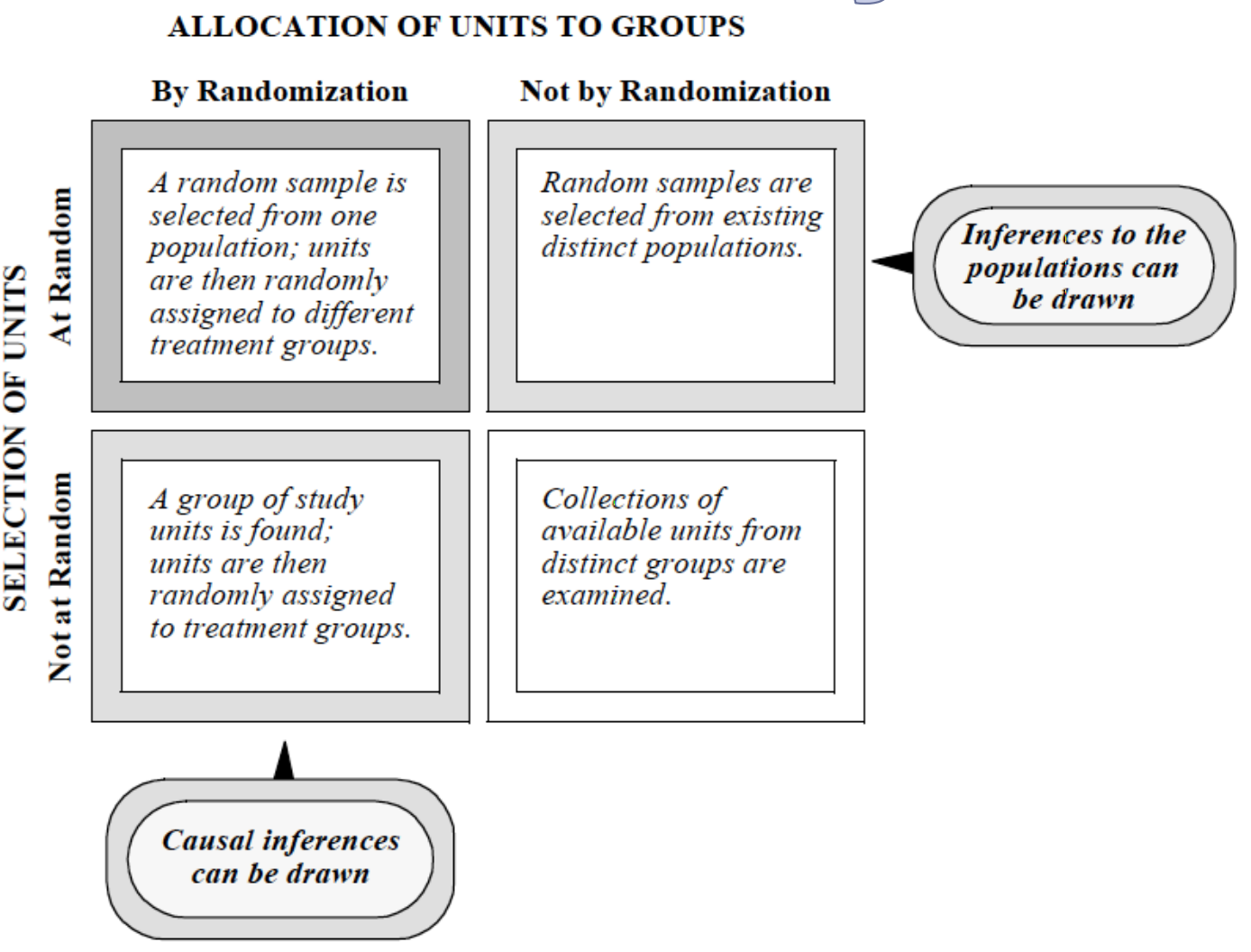
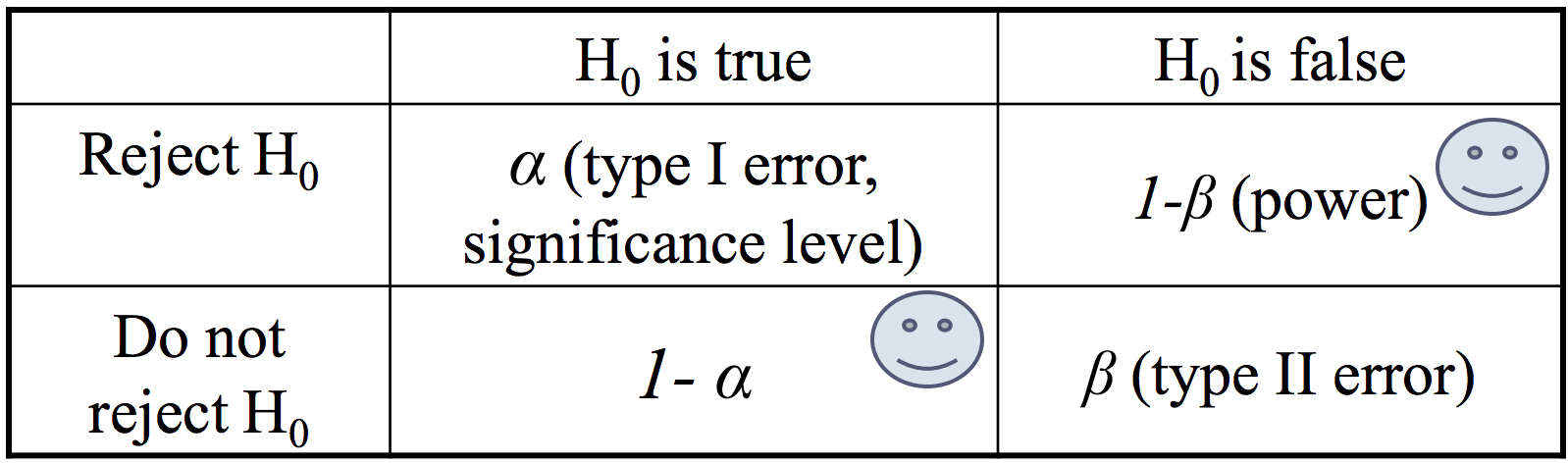
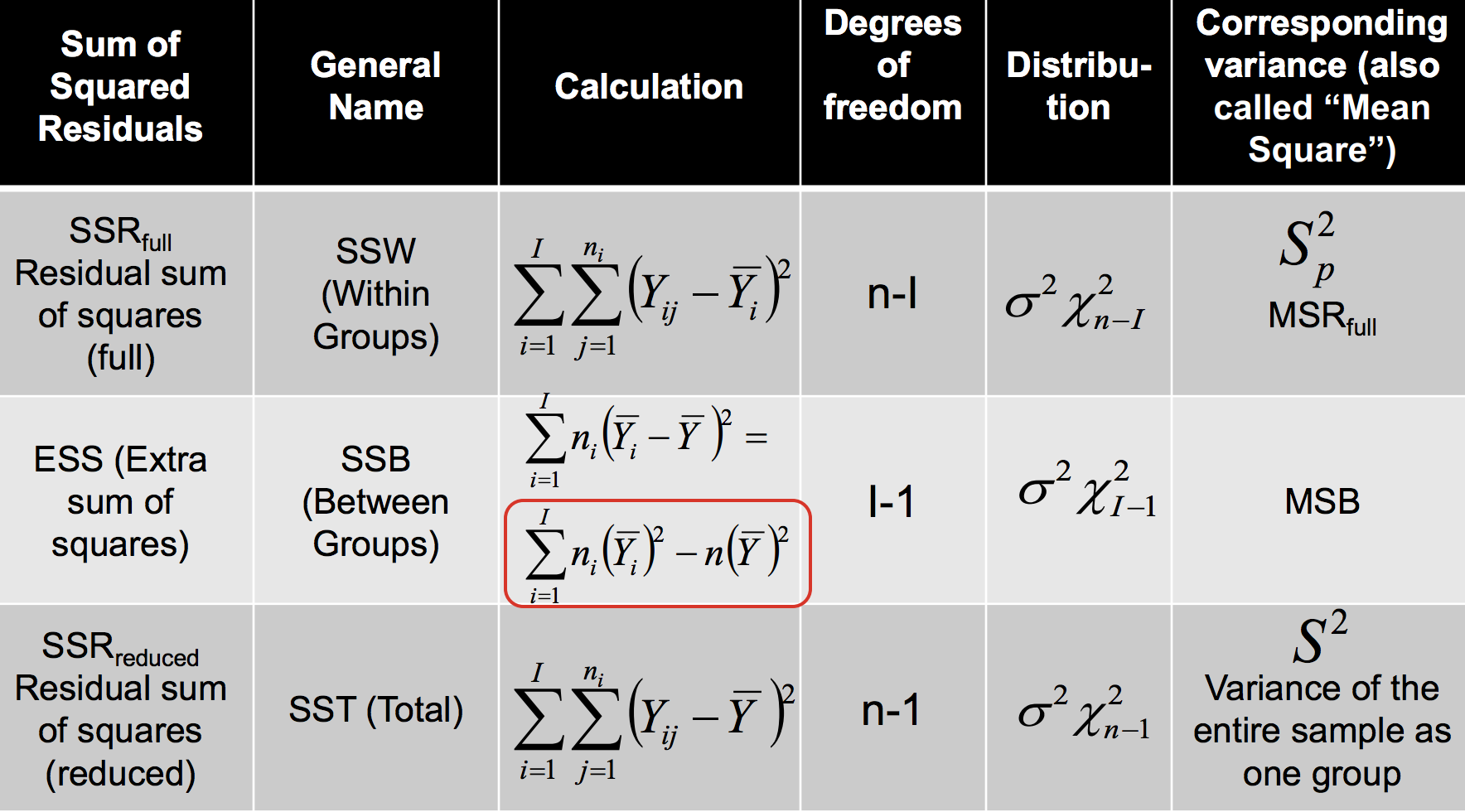
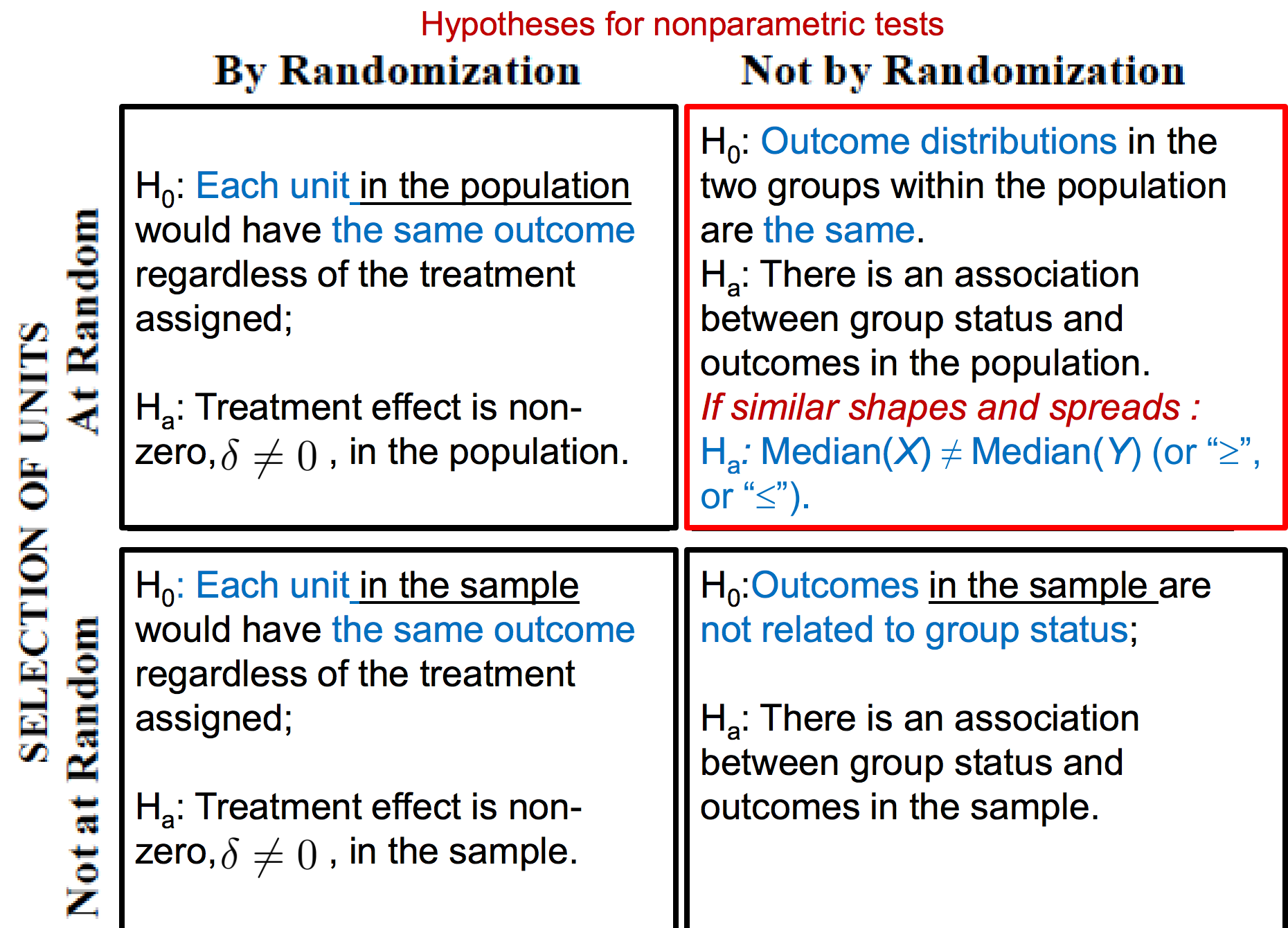
Definitions:

|  |  |
| --- | --- |
| **Study (experimental) unit (subject)** - one member of a set of entities being studied.  **Parameter (also, estimand)** -proportion of childless households in the population.  **Estimate** - proportion of childless households in the sample.  **Parameter** is a population characteristic.  **Estimand** is a parameter that is being estimated (**μ**).  **Statistic** is a function of the data (therefore, it is random).  **Statistic:** A function of the data, **μ** **μ** (y ). Test Statistic: Statistic used to weigh evidence  supporting and contradicting the null hypothesis.  **Reference Distribution**: Probability distribution of the test statistic, assuming that the null hypothesis istrue, f(**μ** (Y)|H0).  **Estimator** is a statistic used as a guess for the value of the estimand ( ).  **Estimate** is a quantity, which is a particular realization of the estimator (4/12 = 0.33).  **Target Population:** A collection of units a researcher is interested in; a group about which the researcher wishes to draw conclusions.  **Sampling frame:** Collection of units that are potential members of the sample.  **Overcoverage:** selecting units that are outside your sampling frame.  **Undercoverage**: excluding units that are within your sampling frame.  **Sample:** A [randomly selected] subset of a sampling frame  **Simple Random Sampling (SRS)** – every subset of n units has equal chance to be selected  **Stratified Sampling** – split the population into homogeneous subpopulations and use SRS (or another method) within a sampling frame of each subpopulation.  **Systematic Random Sampling** - select every k’th unit from the ordered sampling frame, starting randomly from of the first k positions.  **Variable probability sampling** – allow units to have nonequal probabilities of being sampled.  **p-value** - probability that the test statistic would be at least as extreme as observed, under the null hypothesis. (or “strength of evidence against the H0”)  **Significance level** (α) is the criterion compared against the p-value. The null hypothesis is rejected if p-value is lower than α.  α reflects the probability of rejecting the null hypothesis given that it is true (**Type I error**).  **Scope of inference**  **Internal validity**: are assumptions of the test satisfied?  **External validity**: possible to make inference to a broader population?  **Sampling distribution** of a statistic is a (reference) distribution that arises from a chance mechanism used to select a random sample from a population.  **Type I Error** (or significance level α): Probability of rejecting the null hypothesis, when the null hypothesis is true.  **Type II Error** (β): Probability of failing to reject the null hypothesis, when the null hypothesis is false.  **Power (1-β):** Probability of rejecting the null hypothesis, when a particular alternative hypothesis is true.  If alpha = 0, beta = 1; if alpha = 1, beta = 0; if alpha increases, beta decreases (power increases).  Increase power by: increase alpha; do one-tailed test; increase effect size; increase sample size  Most to least power, if assumption of t-tools is met  For small samples:  1. T-tools  2. Permutation test  3. Rank-sum test (or signed-rank test for paired data)  4. (Sign test for paired data)  Large samples rejection rates are basically equivalent. | Population Variance: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.14.46 AM.png  Sample Variance:  Sample variance is chi-sq distributed: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.17.15 AM.png  Let Z0~N(0,1), independent of variance X^2~X2n,, thenMacintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.19.56 AM.png  Chi-sq distribution with n df: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.54.02 AM.png  Pooled sample variance: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.59.25 AM.png, where  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.00.13 AM.png  Sampling distribution of pooled sample variance: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.02.44 AM.png  F-distribution: if Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.18.46 AM.png, independent of each other, then  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.17.47 AM.png(F-distribution with nx and ny d.f) |
| **No** |  |

|  |  |  |
| --- | --- | --- |
| **(Fisher’s) Randomization Test**: is a distribution-free test for treatment effect in **randomized experiments.**  **H0** : Zero treatment effect for all units, δ=0. Each unit’s outcome is the same, regardless of the treatment assigned. Consequently, the distribution of outcomes is identical in two groups  **Ha** : Non-zero treatment effect for ALL units, δ≠0 .  Assumptions:   1. Random assignment to groups. 2. Under the H0, independence of study units.   **Test statistic:** Difference between average outcomes in the two groups\*.Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 9.51.45 AM.png  \*In fact, any summary of the data can represent a test statistic.  **Randomization distribution** is a reference distribution of a test statistics in a randomization test, where variation is due to random assignment of the treatment.  **Permutation test** is a distribution-free nonparametric test for association between group status and outcome in **observational studies.**  **H0 :** Outcomes not related to group status.  **Ha :** Outcomes are related to (associated with) group status.  Assumptions: Independence of study units.  **Test statistic:** Difference in average observed outcomes  between the two groups (or, any other statistic),  Permutation distribution is a reference distribution of a test statistics in a permutation test.  *NOTE: Permutation tests are actually evaluating whether the distributions are equivalent, not just the equality of the means:*  **One Sample z-Test**  H0: **μ = μ0**  HA: **μ != μ0**  **Test Statistic:** Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.09.10 AM.png (make sure to double p-value if doing 2-sided test, of ½ alpha)  Exact sampling distribution of Z under H0: Z~N(0,1)  **Assumptions:**  1. Independence  2. Known population variance, σ^2  3. Normality (CLT allows for deviations!)  4. Random sampling from population  **One Sample t-Test**  H0: **μ = μ0**  HA: **μ != μ0**  Test Statistic: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.21.43 AM.png  Exact sampling distribution of T under H0: T ~tn-1 (t-distribution with df = n-1)  **Assumptions:**  1. Independence  2. Normality (CLT allows for deviations!)  3. Random sampling from population  **Two Sample z-Test**  H0: **μx - μy = μ0**  **HA: μx - μy != μ0**  Test Statistic: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.33.14 AM.png  **Pooled Two-Sample t-Test**  H0: μx - μy = μ0  HA: μx - μy != μ0  Test Statistic: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.37.39 AM.png  Exact sampling distribution of T under H0: T ~tnx+ny-2 (t-distribution with df = nx + ny -2)  Note:s Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.40.24 AM.png,  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.41.22 AM.pngis an unbiased estimator of Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.41.29 AM.png  Assumptions: same as Welch’s, but  If using pooled variance – assume equal variances between populations.  **Unpooled (Welch) Two-Sample t-Test**  H0: μx - μy = μ0  HA: μx - μy != μ0  Test Statistic: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.43.26 AM.png  Approximate Sampling distribution:  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.44.34 AM.png  Assumptions:  1. Independence between units:  within each population (Yi with Yj and Xi with Xj)  between populations (Yi with Xj)  2. Homogeneity of units within each population:  Equal means within each population;  Equal variances within each population;  If using pooled variance – assume equal variances between populations.  3. Population distributions are Normal:  CLT allows for deviations.   1. Random sampling from populations.   **Paired t-Test**  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.55.48 AM.png  H0: μD = μ0  HA: μD != μ0  Test Statistic: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 10.56.45 AM.png  Exact sampling distribution of T under H0: T~tn-1 (where n is # of pairs)  Assumptions same as one-sample t-test  **F-test for equal variance** (could also use levene’s test): (caution: will reject for large samples when ratio is close to 1)  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.16.19 AM.png  Same assumptions as t-test:  In testing the equality of population variances, two assumptions are required: independent samples and normally distributed populations.  2-sided p-value is just 2\*the one-sided value. | **How does CLT apply?**  **# checking assumptions:**  **Independence:** plot subgroups, and see if relationship still holds (cluster effect)  Plot data over time (serial effects)  Plot data vs space (spatial effect)  **Normality:** Histograms should look normal – overlay with normal curve and kernel density plot  Boxplot or Q-Q plot will show outliers  **Equal Variance:** divide variances to see if ratio is <0.5 or > 2  **T-test is robust:**  -if populations are symmetrically skewed  -if sample sizes are equal (and large)  -when sample sizes are equal, pooled-ttest is robust to unequal variances  **Sensitive:** When the sample sizes are not equal, t-tests are more sensitive to skewedness and long-tailedness.  For small samples, t-tests are somewhat sensitive to markedly different skewedness in two groups.  Watch out for outliers.  **Robustness of F-tests:**  Not resistant to outliers. Fairly robust to non-normality. Independence w/I and b/w groups is necessary. Equal variance is necessary.  **F-Test Assumptions:**  Equal variances is crucial  Not resistant to outliers  Normality is not critical.  **Confidence intervals**: Yi ~iid N(mu, sigma^2)  One Sample t-test:Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.31.48 AM.png  Two Sample t-test: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.33.41 AM.png  Two-Sample t-test for I samples (I > 2)  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.06.24 PM.png  **Transformations:**  Log Transformation:   1. **Randomized Experiments** (T=Treatment, C = Control)   Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.36.30 AM.png Responses: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.37.33 AM.png  Interpretation: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.38.30 AM.png -- The response of an experimental unit to the treatment will be exp(ZbarT – ZbarC) times as large as its response to the control   1. Observational studies with **random sampling**:   Interpretation: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.40.25 AM.png -- The median of the second population is exp(Zbar2 – Zbar1) times as large as the median of the first population.   1. In paired t-test: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.45.17 AM.png   **Randomized Experiments**:Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.43.35 AM.png  **Observation studies w/ rand. Sampling**: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.44.48 AM.png  Square Root Transformation sqrt(Y): Good for moderately skewed data, but hard to interpret results  Reciprocal Transformation 1/Y: good for severely skewed data, can be used with negative data  Logit log(Y/(1-Y)) or arcsin(2Y-1): good for proportions  **Nonparametric alternatives to t-tools:**  **Rank Sum test:**  Hypotheses:  H0: (if treatments are randomized): Same as fisher’s randomization test  H0: (if treatments are not randomized): same as permutation test  \* If shapes and spreads of the two populations are similar: HA: There is a difference in medians between the two populations.  Transform data:  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.55.31 AM.png  Ties are averaged  Test statistic: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 11.56.12 AM.png  Assumptions: Independence  Normal approximation to rank Sum test (if no ties, and samples greater than 10):Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.27.17 PM.png  With ties, Rbar is sample mean (mean of all ranks) and Sr^2 is sample variance for combined set of nx + ny ranks: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.28.13 PM.png  We can use this Z-statistic: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.28.53 PM.png  **Sign test for paired data** (lower power than Wilcoxon signed-rank test)  **General formuation: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.33.50 PM.png**  **If differences are continuous rv’s: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.33.59 PM.png**  **If differences are symmetric: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.34.03 PM.png**  **Test Statistic: K = number of positive differences.**  **Exact distribution of K: K ~Binom(m, 0.5), where m is final number of pairs with nonzero Di.**  **Normal approximation (for large samples): Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.38.53 PM.png**  **Assumptions:**   1. The differences *Di* are assumed to be independent. 2. Each *Di* comes from the same continuous population. 3. The values *X*i and *Y*i represent are ordered (at least the [ordinal scale](http://en.wikipedia.org/wiki/Ordinal_scale)), so the comparisons "greater than", "less than", and "equal to" are meaningful. | **Wilcoxon Signed-Rank Test for Paired Data:**  (more assumptions, but higher power than sign test)  **Hypotheses: General formulation:**  **H0:**The rank of the magnitude of within-pair difference is unrelated to the sign of the difference.  **Ha:** The rank of the magnitude of within-pair  difference is related to the sign of the difference.  **If differences are continuous r.v.’s:**  **H0**: Median(Di)=0  **Ha**: Median(Di)≠0 (or >0, or <0)  **If the distribution of differences is symmetric :**  **H0**: E(Di)=0  **Ha:** E(Di)≠0 (or >0, or <0)  1. Calculate differences (+ or -).  2. Rank the absolute values of differences |Di|  3. Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.43.03 PM.png  Test Statistic: (sum of ranks for positive differences)  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.43.08 PM.png  **Exact sampling distribution**: do permutation test, and switch group status for each pair.  **Normal approximation** (for large sample sizes, m is >= 20, where m is the number of non-zero differences):  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 1.49.05 PM.png  Assumptions:   1. Data are paired and come from the same population. 2. Each pair is chosen randomly and independently. 3. The data are measured at least on an [ordinal scale](http://en.wikipedia.org/wiki/Ordinal_scale), but need not be normal.   **Two-Sample t-Test for I samples**  H0: μm - μk = μ0  HA: μm - μk != μ0  Test Statistic: Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 2.20.19 PM.png  Exact sampling distribution of T under H0 is:  T~tn1 + n2 + ….+nI – I, where I is number of groups.  **Assumptions:**  Independence of units within each group;  Independence of units between groups;  Normality of all I populations;  Equal variances for all I populations;  Random sampling.  **Anova F-test for Equality of All means in I samples**  H0: μ1 = μ2 = …=μi (reduced model – equal means model)  Ha: at least one mean is different from the others (full model – separate means model)  **SSR**full = SSW (within groups)  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.15.31 PM.pngMacintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.15.55 PM.png  **SSR**reduced = SST (**total error**)  S^2 is the sample variance of the entire sample, taken as one group:  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.17.21 PM.png  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.18.01 PM.png  **ESS** = SSB (between groups)  ESS = SSRreduced - SSRfull  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.20.17 PM.png  Test Statistic:  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.20.50 PM.png  MSB = Mean square between  MSRfull=Mean square residuals  Easy way to calculate: MSRfull == Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.25.15 PM.png  Exact sampling distribution of R under H0:  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.23.21 PM.png  **Assumptions (check w/ Q-Q or residual plot):**  1. Normality of populations;  2. Equal population variances for all groups;  3. Independence within each group;  4. Independence between groups;  5. Homogeneity within each group;  6. Random sampling  **Kruskal-Wallis Test** (Nonparametric ANOVA)  H0: Median(Y1) = Median(Y2) = …=Median(YI)  HA: at least one median is different from the others  Test Statistic – convert to ranks, and do ANOVA on ranks:  Macintosh HD:Users:callinswitzer:Desktop:Screen Shot 2014-10-12 at 3.36.59 PM.png  SR^2 is the sample variance of **all** ranks (not pooled)  Assumptions:  1. Independence.  2. Random Sampling  F-test for Equality in a subset of groups  Linear Combinations of Means |
| **Multiple Comparisons** | **Test Normality:**  Shapiro-wilk test, Anderson-Darling test, Kolmogorov–Smirnov Test |  |



Derive sampling distribution: Q2d on Hw 3

Review “Overview” in lecture 13

Central Limit Theorem

Law of large numbers

Internal vs. external validity

Study design concepts (units, target population, SRS vs. stratified

random sampling, etc.);

 Sample statistics vs. population parameters;

 Rules for expectations and variances of r.v.;

 Mechanics of inference using randomization (i.e., random

permutation) vs. random sampling;

 Tests for 1, 2, and 3+ samples:

 Hypotheses that they are testing (depending on a question of interest,

study design, and data transformation used, if any);

 Assumptions for each test and methods for checking them;

 Test-statistics for each test and their sampling distributions, looking up

p-values using a table, calculating confidence intervals;

 Ability to interpret output of R functions that perform these tests;

 Ability to state conclusions and comment on a scope of inference.

 Definition of distributions by representation;

Conceptual, derivations, other non-simulation questions,

similar to those on homeworks.

 Motivation for and interpretation of simulations on

homeworks.

 Tests and confidence-interval calculation using

distribution tables, comments on assumptions.

 Explanation of what happens behind any computer

function you’ve used.

 Suggestion: verify the calculations behind t-tests and

other tests for some of the homework problems.

 Ramsey and Schafer has lots of problems to work

through (including conceptual ones)