D. Two-Sample Estimation and Confidence Intervals

1. $T(\theta) = \sum_{j=1}^{m+n} a_j Z_j(\theta)$,
   a. where $Z_j(\theta)$ is 1 if item ranked $j$ among $X_1, \ldots, X_m, Y_1 - \theta, \ldots, Y_n - \theta$ came from $Y$, and 0 otherwise.

2. Can define an estimator as that value of $\theta$ that makes test statistic equal to its null expectation
   a. $\hat{\theta}$ solves $T(\hat{\theta}) = n \tilde{a}$.
   b. CI is $\{ \theta | t_l \leq T(\theta) \leq t_u \}$, for $t_l$ and $t_u$ such that $P_0 [T(0) \leq t_l] \leq \alpha/2$, $P_0 [T(0) \geq t_u] \leq \alpha/2$.
      i. $t_l, t_u \approx n \tilde{a} \pm z_{\alpha/2} \sqrt{\text{Var}[T(0)]}$.
   c. Hence estimator is median of differences:
      i. Hodges-Lehmann estimator in terminology of Higgins and SAS.
      • More standard definitions in one-sample case follows later.
      ii. Use mean of middle two if $n \times m$ even, even though not technically required by the theory.

3. When $a_j$ are ranks $j$, $T(\theta)$ is the Mann Whitney Wilcoxon test
   a. Work with Mann-Whitney-Wilcoxon test
      $W(\theta) = \sum_i \sum_j I(X_i < Y_j - \theta)$.
   b. $W(\theta) = 0$ if and only if $n \times m$ even, and exactly $n \times m/2$ of $Y_i - X_i$ are greater than $\theta$, or $n \times m$ odd, and $(n \times m - 1)/2$ of $Y_i - X_i$ are $> \theta$, $(n \times m - 1)/2$ are $\leq \theta$, and one is $\theta$.
   c. Hence estimator is median of differences:
      i. Hodges-Lehmann estimator in terminology of Higgins and SAS.
      • More standard definitions in one-sample case follows later.
      ii. Use mean of middle two if $n \times m$ even, even though not technically required by the theory.

E. Mann-Whitney-Wilcoxon tests whether $P[X_k \leq Y_j] = \frac{1}{2}$

1. Clearly true if $F = G$
2. Also true if $F$ and $G$ are symmetric about same point.
3. Also true if $\int_{-\infty}^{\infty} F(y)g(y) \, dy = \frac{1}{2}$.
   a. Ex. $Y \sim \mathcal{E}(1), X \sim N(\theta, 1)$.
   b. Solve $1/2 = \int_{0}^{\infty} (1 - \exp(-y)) \exp(-y - \theta)^2/2) (2\pi)^{-1/2} \, dy$
   c. Obtain $\theta = .876$. [Mark A R]

F. Tests of equivalence of CDFs.

1. Calculate empirical CDFs as we saw before.
2. Kolmogorov-Smirnov:
   a. Use largest difference between these as test statistic
   b. Use all permutations of data between the two samples, and count portion with as large or larger difference as in $p$-value
      i. Sometimes do this with a random sample
      ii. Sometimes use an approximation to this distribution from 960:554. [Mark E R] [Mark E sas]