Binomial Test : tests of hypotheses for medians. $Z_B = \frac{B - 0.5n}{\sqrt{0.25n}}$, B = # of obs greater than or equal to $\theta_{0.5}$

Confidence Interval : $P(X_a < \theta_{0.5} < X_b) = 1-\alpha = \sum_{x=a}^{b-1}\binom{n}{x}(0.5)^n$ = the probability that at least "a" and at most "b-1" of the obs fall less than $\theta_{0.5}$.

$\hat{F}(x) \pm Z_{(1-\alpha/2)}\sqrt{\frac{\hat{F}(x)[1-\hat{F}(x)]}{n}}$

Large Sample Approximation : $\frac{a-np}{\sqrt{np(1-p)}} = -Z_{(1-\alpha/2)}$, $\frac{b-1-np}{\sqrt{np(1-p)}} = Z_{(1-\alpha/2)}$

Power : In general, the binomial test will have higher power than the CLT test for heavy-tailed distributions.

Power of CLT Test : $1 - \Phi\left(Z_{(1-\alpha/2)} - \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)$

Power of Binomial Test : $1 - \Phi\left(Z_{(1-\alpha/2)}\sqrt{\frac{0.25}{P(1-P)}} - \frac{P-0.5}{\sqrt{P(1-P)/n}}\right)$

↑ CH 1 ↓ CH 2

Permutation Test : the test of the distribution of $\binom{N}{m}$ differences of means.

Process of Permutation Test :
- Calculate $\binom{N}{m}$, m = # of observations designated to new method. (# of qualified obs / $\binom{N}{m}$)
- Compute the difference, or any desirable computations, of all combinations.
- Compute # of cases where the differences of a combination is greater than or equal to that of the targeted combination.

Large Sample Approximation : $D_{obs} \pm Z_{(1-\alpha/2)}\sqrt{\frac{P(1-P)}{K}}$

Wilkoxon Rank-Sum Test : Let W_i be the sum of the ranks of the observations from treatment i. The test is a 2-sample permutation test based on W_i

Process of Wilcoxon Test :
- Combine $m+n$ obs and assign ranks to each.
- Compute W_1 and W_2 according to the preassigned ranks.
- Find all possible permutations of ranks. $\binom{N}{m}$ combinations
- Compute W_i's for each combination.
- $P_{upper\text{-}tail} = \frac{\#\text{ of rank sums} \geq \text{observed rank sum } W}{\binom{m+n}{m}}$

Differences of mean rank $= W_i\left(\frac{1}{m} + \frac{1}{n}\right) - \frac{N(N+1)}{2n}$

Wilcoxon Rank-Sum Test Adjusted for ties :
- process stays the same except adjusting ranks with some numbers to average

Large Sample Approximation : $E(T_i) = m\mu$, $Var(T_i) = \frac{mn\sigma^2}{N-1}$ $\Rightarrow$ $\mu = \frac{\sum A_i}{N}$, $\sigma^2 = \frac{\sum (A_i-\mu)^2}{N}$ $\Rightarrow$ $Z = \{T_i - E(T_i)\}/\sqrt{Var(T_i)}$

Mann-Whitney Statistics "U" : U = # of pairs (X_i, X_j) for which $X_i < Y_j$. * i and j are from 2 different samples. $H_0: F(x) = F(Y)$

Process of Mann-Whitney test :
- make pairs of (X_i, Y_j) and count the # of pairs $X_i < Y_j$
- Using the Mann-Whitney CI table, confirm whether U is within the CI.

$R(Y_j) = (\#\text{ of } Y\text{'s} \leq Y_j) + (\#\text{ of } X_i \leq Y_j)$

$\Rightarrow W_2 = \sum_{j=1}^{n} R(Y_j) = \frac{n(n+1)}{2} + U$

Hodges-Lehmann Estimate of Δ : the median of all the pairwise differences of $X_i - Y_j$

Using "pooled" t-test $\Rightarrow \bar{X}_1 - \bar{X}_2 \pm t_{(1-\alpha)} S_e\sqrt{\frac{1}{m}+\frac{1}{n}}$, with df. $= m+n-2$, $S_e = \sqrt{\frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}}$

Process :
- form all mn pairwise differences of $X_i - Y_j$
- arrange all pair-wise differences in an ascending order, and find K_a and K_b that satisfy, $pwd(k_a) < \Delta \leq pwd(k_a) \approx 1-\alpha\%$, referring to Mann-Whitney table

Siegel-Tukey Test : Test of equality of variations of two populations. EX) $H_0: \sigma_1 = \sigma_2$

Process of Siegel-Tukey Test :
- arrange the observations in an ascending order.
- assign rank 1 to the smallest obs, rank 2 to the largest obs, rank 3 to the next largest obs, rank 4 to the next smallest, ... so on.
- apply Wilcoxon rank-sum test. The smaller rank sums are associated with the treatment that has the larger variability.

Test on Deviances :
- obtain dev_{iX} and dev_{jY} and compute RMD from the original data, $dev_{iX} = X_i - \mu_1$, $dev_{jY} = Y_j - \mu_2$ $\Rightarrow$ $RMD = \frac{\sum_{i=1}^{m}|dev_{iX}|/m}{\sum_{j=1}^{n}|dev_{jY}|/n}$

alternatively, $RMD = \frac{\max\left(\sum_{i=1}^{m}|dev_{iX}|/m, \sum_{j=1}^{n}|dev_{jY}|/n\right)}{\min\left(\sum_{i=1}^{m}|dev_{iX}|/m, \sum_{j=1}^{n}|dev_{jY}|/n\right)}$

- Permute the deviances among 2 treatments and obtain RMD for each permutation
- Compute p-values, $RMD_i \geq RMD_{obs}$ for upper-tail, $RMD_i \leq RMD_{obs}$ for lower-tailed

Kolmogorov-Smirnov Test : can be applied even when you don't have any information about the population distributions. $T_{KS} = \max_z |\hat{F}_1(z) - \hat{F}_2(z)|$

Process of K-S Test :
- allign the data in an ascending order and specify where each obs belongs to, and count one by one and check where the obs belong.
- Calculate the absolute difference for each step and find the maximum absolute difference.

Sampling Formulas : let A_i be the ranks and $\mu = \frac{\sum A_i}{N}$, $E(T_i) = m\mu$, $Var(T_i) = \frac{mn\sigma^2}{N-1}$, $\sigma^2 = \frac{\sum A_i^2}{N} - \mu^2$, $Z = \frac{T_i - E(T_i)}{\sqrt{Var(T_i)}}$

Sampling Formulas for Wilcoxon Rank-Sum Test : $\sum i = \frac{N(N+1)}{2}$, $\sum i^2 = \frac{N(N+1)(N+2)}{6}$, $\sigma^2 = \frac{(N-1)(N+1)}{12}$ $\Rightarrow$ $E(W) = \frac{m(N+1)}{2}$, $Var(W) = \frac{mn(N+1)}{12}$

CH-3

K-Sample Permutation Tests : comparing the distributions of K different samples. EX) $H_0: F_1(x) = F_2(x) = \cdots = F_K(x)$ ~ Using F-statistic

Sum of Squares for Treatments SST : $\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$, where $\bar{x} = \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)/N$ $\Rightarrow$ **Mean Squares for Treatments MST** : $\frac{SST}{K-1}$ & **Sum of Square for Error SSE** : $\sum_{i=1}^{k}(n_i-1)S_i^2$

Mean Squares for Error : $\frac{SSE}{N-K}$ $\Rightarrow$ $F = \frac{MST}{MSE}$, with df = K-1 for numerator and df = N-K for denominator

Process of Permutation F-Test :
- Obtain the F-statistic for the original data
- Obtain all possible permutations, and compute the F-statistic for each permutation.

Compute $\frac{\#\text{ of } F_i(x) \geq F_{obs}}{\text{all permutations}}$

$\binom{N}{n_1, n_2, \ldots n_K} = \frac{N!}{n_1!\,n_2!\cdots n_k!}$

Alternative Forms of the Permutation F statistic : $SS_{total} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2 = SST + SSE = C$, $F = \frac{SST/K-1}{(C-SST)/(N-K)}$, where $SST = \sum_{i=1}^{k} n_i\bar{x}_i^2 - N\bar{x}^2$, $SSX = \sum_{i=1}^{k} n_i\bar{x}_i^2$

t_i = # of tied obs in ith group

Kruskal-Wallis Statistic : $KW = \frac{12}{N(N+1)}\sum_{i=1}^{k} n_i\left(\bar{R}_i - \frac{N+1}{2}\right)^2$, most of KW critical values will be smaller than χ^2's critical values.

Adjustment for Ties : $KW = \frac{1}{S_R^2}\sum_{i=1}^{k} n_i\left(\bar{R}_i - \frac{N+1}{2}\right)^2$

Alternative : $KW_{ties} = KW \Big/ \left\{\frac{1 - \sum_{i=1}^{k}(t_i^3 - t_i)}{N^3 - N}\right\}$

Multiple Comparisons : multiple comparisons using pairwise tests allow us to verify which treatment differs from the others, if any.

EX : Bonferroni Adjustment, Fisher's Protected Least Significant Difference, Tukey's Honest Significant Difference

Family Wise Error Rate :
- the probability of making one or more false rejections (FWER > α)

FWER = 1 − P(no false rejection for K treatments)
= 1 − P(no rejection of H_{0K})$^{K(K-1)/2}$
= 1 − $(1-\alpha)^{K(K-1)/2}$

Bonferroni Adjustment : uses α' rather than α, $\alpha' = \frac{2\alpha}{K(K-1)}$

- If the data are normally distributed, we may use the t-test; otherwise, use the Wilcoxon rank-sum test or any other nonparametric approach.

Fisher's Protected LSD :
- run F-test for equality of means.
- if the F-statistic is significant at level α, run all pairwise t-tests at level α.

$|\bar{X}_i - \bar{X}_j| \geq t_{(\alpha/2, df)}\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$

Rank-based analogue of LSD for testing the equality of distributions $\Longrightarrow$ $|\bar{R}_i - \bar{R}_j| \geq z_{(\alpha/2)}\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$

Tukey's HSD Procedure : measures the largest difference between sample means. $Q = \sqrt{n}(\max_i(\bar{x}_i) - \min_i(\bar{x}_i))/\sqrt{MSE}$ $\Rightarrow$ $|\bar{x}_i - \bar{x}_j| \geq q(\alpha, k, k(n-1))\sqrt{\frac{MSE}{n}}$ $\Longleftrightarrow$ $|\bar{R}_i - \bar{R}_j| \geq q(\alpha, K, \infty)\sqrt{\frac{N(N+1)}{12n}}$

Tukey-Kramer Procedure : to declare treatments i and j to be statistically different. $|\bar{x}_i - \bar{x}_j| \geq q(\alpha, K, N-k)\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$ $\Longleftrightarrow$ $|\bar{R}_i - \bar{R}_j| \geq q(\alpha, k, \infty)\sqrt{\frac{N(N+1)}{24}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$

Bonus

Ordered Alternatives : if treatments are not equal, it may be possible to anticipate the direction $H_a: F_1(x) \geq F_2(x) \geq F_3(x) \geq \cdots \geq F_k(x)$

Jonckheere-Terpstra Test : a test statistic of the form T in which T_{ij}'s are one-sided Mann-Whitney statistics, $T_{JT} = \sum_{i<j} T_{ij}$

- Compute JT_{obs}, and obtain all possible samples without replacement of sizes n_i and n_j, or randomly select if N is large.
- To obtain the upper-tail p-value, compute fraction of the JT's greater than or equal to JT_{ob}

$\Rightarrow$ Large Sample Approximation : $E(JT) = \frac{N^2 - \sum_{k=1}^{K} n_k^2}{4}$, $Var(JT) = \{N^2(2N+3) - \sum_{k=1}^{K} n_k^2(2n_k+3)\}/72$ $\Rightarrow$ $Z = \frac{JT - E(JT)}{\sqrt{Var(JT)}}$ | $E(JT_{Wilcoxon}) = \sum_{i<j} \frac{n_j(n_i+n_j+1)}{2}$

CH 4 Paired-Comparison Permutation Test : $\rightarrow$ Compute D_i's and $\bar{D}_{obs}$ $\rightarrow$ Then compute $\bar{D}_i$ for all 2^n permutations. $\rightarrow$ Upper-tail p-value $= \frac{\#\text{ of } \bar{D}_i\text{'s} \geq \bar{D}_{obs}}{2^n}$, $H_0 : F(x) = 1 - F(x)$

↳ Large-Sample Approximations : $E(\bar{D}) = 0$, $Var(\bar{D}) = \frac{1}{n}\sum_{i=1}^{n} |D_i|^2$, $Z = \bar{D}/\sqrt{Var(\bar{D})}$ | $E(S_+) = \frac{1}{2}\sum_{i=1}^{n} |D_i|$, $Var(S_+) = \frac{1}{4}\sum_{i=1}^{n} |D_i|^2$, $Z = (S_+ - E(S_+))/\sqrt{Var(S_+)}$

Signed-Rank Test : nonparametric test for paired comparisons based on ranks. We rank the absolute values of the differences, and then attach the signs to the ranks. $\bar{D} = \frac{\sum_{i=1}^{n} V_i|D_i|}{n}$, $V_i = -1$ or 1

Wilcoxon Signed-Rank Test : - Obtain SR_+^{obs} and all 2^n possible SR_+^i. $\rightarrow$ $P_{upper} = (\#\text{ of } SR_+ \geq SR_+^{obs})/2^n$ | $P_{Lower} = (\#\text{ of } SR_+ \leq SR_+^{obs})/2^n$ * always one-tailed | $SR_+ = \sum_{i=1}^{n} V_i|D_i|$, $V_i = 0$ or 1

↳ Large-Sample Approximation : $SR_+ = \sum_{i=1}^{n} i\,V_i$ ---> $E(SR_+) = \frac{n(n+1)}{4}$ $\rightarrow$ $Var(SR_+) = \frac{n(n+1)(2n+1)}{24}$ $\rightarrow$ $Z = \{SR_+ - E(SR_+)\}/\sqrt{Var(SR_+)}$ | Selecting Among Paired-Comparison Tests : - Shifting by $\Delta \Rightarrow$ signed-rank test

↳ Ranking with Zeros : the absolute values of the differences, including zeros, are ranked. - If the distributions have lighter tails $\Rightarrow$ t-test

Features of Randomized Complete Block Design : - Experimental units are divided into blocks in a way that units or conditions within blocks are homogeneous.
- Blocks have the same # of units as there are treatments, and the treatments are randomly assigned to experimental units within blocks.

Permutation F-test for RCBD : - Compute F_{obs} and F-statistics for all $(K!)^b$ permutations. $\Rightarrow$ $P_{upper} = (\#\text{ of } F\text{'s} \geq F_{obs})/(K!)^b$ | $SST^* = \sum_{k=1}^{K} (\bar{x}_{i\cdot} - \bar{x})^2$ or $SSX^* = \sum_{k=1}^{K} (\bar{x}_{i\cdot})^2$

Friedman's Test for RCBD : - assign ranks within blocks and obtain mean rank for each treatment. $\Rightarrow$ $FM = \frac{12b}{K(K+1)} \sum_{i=1}^{K} (\bar{R}_i - \frac{K+1}{2})^2$, and refer to χ^2 table with k-1 degrees of freedom.

Cochran's Q : used for experiments with binary outcomes | both use FM test | Ordered Alternatives for RCBD $\Rightarrow$ Page's Test : - measure of the association b/w the presumed order and the rank of the treatments.

Kendall's W : level of agreement (Concordance) | $E(SSR) = \sum_{k=1}^{K} E(\bar{R}_i - \frac{K+1}{2})^2 = \frac{K}{b^2}\sum_{k=1}^{b} \sum_{B_j} v_{B_j}^2$ | Assume the obs are ranked, $PG = \sum_{k=1}^{b} i\,R_i$, $\rightarrow$ has the same process with the permutation test

Large-Sample Approximation : $E(PG) = \frac{bk(K+1)^2}{4}$, $Var(PG) = \frac{b(k-1)K^2(K+1)^2}{144}$ $\Rightarrow$ $Z = \frac{PG - E(PG)}{\sqrt{Var(PG)}}$

CH 5 For bivariate sampling, $\beta_1 = \rho\frac{\sqrt{V_Y}}{\sqrt{V_X}} \sim \hat{\beta}_1 = r\frac{S_Y}{S_X}$

Pearson Correlation : $r = \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) / \sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}$ $\Rightarrow$ $t_{cor} = \sqrt{\frac{n-2}{1-r^2}}\, r$ $\leftarrow$ t-distribution with n-2 df. $H_0 : \rho = 0$

Slope of Least Squares Line : $SSE = \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$ $\Leftarrow$ $\hat{\beta}_1 = \{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})\} / \{\sum_{i=1}^{n}(X_i - \bar{X})^2\}$ $\Leftarrow$ $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$ | $t_{slope} = \hat{\beta}_1\sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{MSE}}$ $\Leftarrow$ $MSE = \frac{SSE}{n-2}$ | r_i's or r_{obs} | $\frac{\#\text{ of } \hat{\beta}_1\text{'s} \geq \hat{\beta}_1^{(obs)}}{n!}$

Process of Permutation test for Slope / Correlation : - Compute $\hat{\beta}_1^{obs}$ or r_{obs}, and compute $\hat{\beta}_1$'s or r_{obs}'s for all n! permutation (Permute Y's among the X's) $\Rightarrow$

↳ Large Sample Approximation : $Z_r = r\sqrt{n-1}$ | Spearman Rank Correlation : Same Process with Permutation test for Correlation except that it measures correlation with ranks. $Z = \frac{r_T}{\sqrt{Var(r_T)}}$

Kendall's τ : - Concordant means $X_i < X_j$ implies $Y_i < Y_j$ for $i<j$ | $\tau = 2P[(X_1 - X_3)(Y_1 - Y_3)] > 0 - 1$ * $\tau = \frac{1}{2}$ implies no association | Large-Sample Approximation : $E(r_T) = 0$, $Var(r_T) = \frac{4n+10}{9(n^2-n)}$

Permutation Tests for Contingency Tables | Permutation χ^2 Test : - Compute all $\binom{n}{r}$ χ^2-statistics, where n = # of units, r = # of rows. $\chi^2 = \sum_{i}^{r}\sum_{j}^{c}(n_{ij} - e_{ij})^2/e_{ij}$, $e_{ij} = \frac{n_{i\cdot}\,n_{\cdot j}}{n}$ $\Rightarrow$

$\Rightarrow$ and count the # of χ^2-statistics greater than or equal to χ^2_{obs} | Fisher's Exact Test : let X denote n_{11}, then $P(X \leq n_{11}) = \sum_{x=0}^{n_{11}} \binom{n_{1\cdot}}{n_{11}}\binom{n_{2\cdot}}{n_{1\cdot} - n_{11}} / \binom{N}{n_{1\cdot}}$, red-colored are fixed

McNemar's Test : $P(T_{MN} \geq T_{MN}^{obs}) = \sum_{i=obs}^{n}\binom{n}{i}(0.5)^n$ $\Rightarrow$ $Z_{MN} = \frac{|n_{MN} - 0.5n|}{\sqrt{0.25n}}$ | Mantel-Haenszel Test : $E(X_k) = \frac{r_{1k}c_{1k}}{N_k}$, $Var(X_k) = \frac{r_{1k}r_{2k}c_{1k}c_{2k}}{N_k^2(N_k-1)}$ $\Rightarrow$ $MH = \left(\sum_{k=1}^{s}[X_k - E(X_k)]\right)^2 / \sum_{k=1}^{s} Var(X_k)$

MH Estimate of the common odds ratio : $\hat{\theta} = \frac{A}{B}$, $A = \sum_{k=1}^{s} n_{11k}n_{22k}/N_k$, $B = \sum_{k=1}^{s} n_{12k}n_{21k}/N_k$ $\Rightarrow$ $Var(\log\hat{\theta}) = \{\sum_{k=1}^{s}(n_{11k}+n_{22k})(n_{11k}n_{22k})/N_k^2\}/2A^2 + \{\sum_{k=1}^{s}(n_{11k}+n_{22k})(n_{21k}n_{21k}) + (n_{12k}+n_{21k})(n_{11k}n_{22k})/N_k^2\}/2AB + \cdots$

$\cdots + \{\sum_{k=1}^{s}(n_{12k}+n_{21k})(n_{12k}n_{21k})/N_k^2\}/2B^2$

CH 8

Bootstrap Variance and Bias : $B = E(\hat{\theta}) - \theta$, $MSE = Var + B^2$, $\hat{E} = \frac{1}{REP}\sum^{REP}\hat{\theta}_{bi}$, $\hat{B} = \hat{E} - \hat{\theta}$, $Var = \frac{1}{REP}\sum^{REP}(\hat{\theta}_{bi} - \hat{E})^2$

* For symmetric dist, just use t-dist with df = n-1

Process of Obtaining a Bootstrap CI for μ : - Compute $\bar{X}_{obs}$ and S_{obs}, and compute all $\bar{X}_b$'s, S_b's, and t_b's $\Rightarrow$ $\bar{X} - t_{b,0.975}(\frac{S}{\sqrt{n}}) < \mu < \bar{X} - t_{b,0.025}(\frac{S}{\sqrt{n}})$ The process is the same except for using χ^2-statistics

CI for the Variance and Standard Deviation : $\{(n-1)S^2\}/\chi^2_{0.975} < \sigma^2 < \{(n-1)S^2\}/\chi^2_{0.025}$ $\Leftrightarrow$ $P(\chi^2_{0.025} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{0.975})$, if normal, * $\chi^2 = \{(n-1)S^2\}/\sigma^2$

Percentile : - draw bootstrap samples and compute b number of $\hat{\theta}_b$. $\Rightarrow$ Find $(1-\alpha/2)100^{th}$ and $(\alpha/2)100^{th}$ percentiles of the bootstrap distribution.

Residual Method : - draw bootstrap samples and compute $e_b = \hat{\theta}_b - \hat{\theta}$. $\Rightarrow$ Find $(1-\alpha/2)100^{th}$ and $(\alpha/2)100^{th}$ percentiles of the bootstrap distribution. $\Rightarrow$ $\bar{z}_0 = \Phi^{-1}\{\frac{1}{REP}\sum^{REP}(\hat{\theta}_b < \hat{\theta})\}$

Bias-Corrected and Accelerated Method BCA : BCA adjusts for bias and skewness. If the bootstrap distribution is symmetric, then BCA, percentile, and residual method give the same conclusions.

- given $a = \{\sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(\cdot)})^3\}/6[\sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(\cdot)})^2]^{\frac{3}{2}}$, where $\hat{\theta}_{(\cdot)}$ is the mean of $\hat{\theta}_{(\cdot)}$'s. $\Rightarrow$ $E[T(\hat{\theta})] = T(\theta) - z_0[1 + aT(\theta)]$, $\nabla[T(\hat{\theta})] = 1 + aT(\theta)$ $\Rightarrow$ $\frac{T(\hat{\theta}) + z_0 - z_P}{1 - a(z_0 - z_P)} < T(\theta) < \frac{T(\hat{\theta}) + z_0 - z_P}{1 - a(z_0 + z_P)}$ $\Rightarrow$ $\alpha_U = \frac{z_0 + z_P}{1 - a(z_0 + z_P)} + z_0$, $\alpha_L = \frac{z_0 - z_P}{1 - a(z_0 - z_P)} + z_0$

Bivariate Bootstrap Sampling : random sample with replacement of the pairs (X_i, Y_i), $i = 1, 2, \cdots, n$. * n^n equally likely possible bootstrap samples

- draw a specified # of bivariate samples of size n and compute r_b for each bootstrap sample. $\Rightarrow$ Obtain CI from r_b's using BCA or percentile method.

Fixed-X Bootstrap Sampling : assume $h(x)$ is a function of the independent variable, and Y may be expressed in terms of the regression model. $Y = h(x) + \varepsilon$

- Compute an estimate $\hat{h}(x)$ and obtain $e_i = Y_i - \hat{h}(X_i)$. $\Rightarrow$ Select n values of the e_i's at random with replacement. $\Rightarrow$ Compute $Y_i = \hat{h}(x_i) + e_{i,b}$

Bootstrap Inferences for the Slope of a Regression Line : - Compute the OLS estimates and the residuals $e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$. $\Rightarrow$ $SE(\hat{\beta}_1) = \sqrt{\frac{MSE}{(n-1)S_X^2}}$, $t = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)}$, $-t_{0.975} < \frac{\hat{\beta}_{1,b} - \hat{\beta}_1}{SE(\hat{\beta}_1)} < t_{0.975}$

$\Rightarrow$ - take samples of size n of the e_i's with replacement, and form the bootstrap sample $(x_i, e_{i,b})$. Compute the least squares estimate of the slope $\hat{\beta}_{1,e}$, the estimated $SE(\hat{\beta}_{1e})$, and $t_e = \hat{\beta}_{1e}/SE(\hat{\beta}_{1e})$

CH 10 - Repeat above to obtain the bootstrap distribution of the t_e's, and construct a CI or conduct a test of hypothesis for β_1.

Histogram : suggested width, $d = 3.5(S/\sqrt[3]{n})$, where S = sample σ | Kernel Method : $\hat{f}(x) = \frac{1}{n\Delta}\sum^{n} w(\frac{x - x_i}{\Delta})$, where $w(z)$ = kernel, $\Delta = 1.06(S/5\sqrt{n})$

KNN Regression : For any X_0, we can always find K nearest values of X_i to X_0. We take average of those values and set it as the value of X_0. $\Rightarrow$ $\hat{f}(X_0) = \frac{1}{K}\sum Y_i$ | K $\uparrow$ Smoothness $\uparrow$ | K $\downarrow$ Smoothness $\downarrow$

Loess Method : $\sum_{X_i \in N_K(X_0)} [Y_i - l(x_i)]^2 w(|X_0 - X_i|/\Delta_{X_0})$, where $l(x)$ = linear/quadratic/higher, $N_K(X_0) = K$ nearest X_i's, $w(u) = (1-u^3)^3$, $\Delta_{X_0} = \max_{X_i \in N_K(X_0)} |X_i - X|$ | * span $= \frac{K}{n}$. Span $\uparrow$ complexity $\downarrow$, Span $\downarrow$ complexity $\uparrow$

Kernel Method : uses points within the range of $X_0 \pm h$, h = bandwidth $\Rightarrow$ $h \uparrow$ complexity $\downarrow$, $h \downarrow$ complexity $\uparrow$ | $\hat{f}(X,Y) = \frac{1}{n\Delta_X\Delta_Y}\sum_{i=1}^{n} w(\frac{X - X_i}{\Delta_X}) w(\frac{Y - Y_i}{\Delta_Y})$ $\Rightarrow$ $\hat{\theta}(x) = \int_c^d y\,f(x,y)\,dy \approx \{\sum_{i=1}^{n} Y_i \cdot w(\frac{x - X_i}{\Delta_X})\} / \{\sum_{i=1}^{n} w(\frac{x - X_i}{\Delta_X})\}$

$MSE_i = (Y_i - \hat{Y}_i)^2$ $\Rightarrow$ $CV_{(n)} = \frac{1}{n}\sum^{n} MSE_i$ * using n is LOOCV and using $n \to K$ is K-Fold CV. Knots : the points where the coefficients change. (usually set as quantiles) * # of knots $\uparrow$ complexity $\uparrow$

Regression Splines (Piecewise Polynomials) : fitting separate low-degree polynomials over different regions of X. $\Rightarrow$ $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_k x_i^k$, where β_i's differ in different parts of the range of X.

↳ Cubic Splines, Natural Splines | Smoothing Splines : giving penalty to models with high variances. $\sum^{n}(y_i - g(x_i))^2 + \lambda\int g''(t)^2\,dt$

Generalized Additive Models : $y_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \cdots + f_p(x_{ip}) = \beta_0 + \sum_{j=1}^{p} f_j(x_{ij}) + \varepsilon_i$

Regression Trees : We choose to divide the regions (predictor space) into high-dimensional rectangles or boxes. The goal is to find boxes R_j's that minimizes $RSS = \sum_{j=1}^{J}\sum(y_i - \hat{y}_{R_j})^2$

Tree-Pruning : growing a very large tree and pruning it back in order to obtain a subtree. | Bagging : generating B different bootstrap datasets $\Rightarrow$ $\hat{f}(x) = \frac{1}{B}\sum^{B}\hat{f}^b(x)$ $\Rightarrow$ Variable Importance

Random Forests : Unlike Bagging, each time a split in a tree is considered, a random sample of m predictors is chosen as split candidates from the full set of p predictors.

- In building a random forest, the algorithm is not even allowed to consider a majority of the available predictors.
- Suppose there is one very strong predictor X_1 in the data. In the collection of trees, most or all of the trees will use X_1 in the top split.
- We've seen that averaging highly correlated quantities doesn't lead to as large of reduction in variance as averaging many uncorrelated quantities

Boosting : each tree is grown using information from previously grown trees; each tree is fit on a modified version of the original data set. d) $\hat{f}(x) = \sum^{B}\lambda\hat{f}^b(x)$

- Set $\hat{f}(x) = 0$ and $r_i = y_i$ for all training data set. $\Rightarrow$ Repeat B times : a) fitting $\hat{f}^b$ with d+1 terminal nodes. b) update $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda\hat{f}^b(x)$ c) Update $r_i \leftarrow r_i - \lambda\hat{f}^b(x_i)$

* - boosting can overfit if B is too large. - Very small λ can require using a very large B for good performance

- The number d of splits in each tree controls the complexity of the boosted model.