△ **ONE-WAY ANOVA** (one cate. predictor) ▲$Y_{ij} = \mu + \tau_i + \epsilon_i$ with constraint (either $\tau_1 = 0$ or $\sum_{i=1}^{k} n_i \tau_i = 0$), τ_i is the ith level effect. ▲ para est. $\hat{\mu}_i = \bar{Y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}, Var(\hat{\mu}_i) = Var(\bar{Y}_i) = \sigma^2/n_i$ ▲$s^2 = (\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2)/n - k = SSE/(n-k) = MSE = s_w^2$ ▲ within group var: $s_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2$ if homo, then all s_i^2 should be similar. rule of thumb (for homo): the ratio of the largest to the smallest within group variance is no more that about k (num of levels) ▲ ci for μ_i's based on $T = (\bar{Y}_i - \mu_i)/(s/\sqrt{n_i}) \sim t_{n-k}$, ci is $\bar{Y}_i \pm t_{n-k}(1-\alpha/2)s/\sqrt{n_i}$ ▲ ci for difference $\mu_i - \mu$: $(\bar{Y}_i - \bar{Y}) \pm t_{n-k}(1-\alpha/2)s\sqrt{1/n_i + 1/n}$ ▲ ci for difference $\mu_i - \mu_1$: $(\bar{Y}_i - \bar{Y}_1) \pm t_{n-k}(1-\alpha/2)s\sqrt{1/n_i + 1/n_1}$ ▲ general testing - ci of a contrast h (combination): $\sum_{i=1}^{k}h_i\bar{Y}_i \pm t_{n-k}(1-\alpha/2)s\sqrt{\sum_{i=1}^{k}(h_i^2/n_i)}$, with $H_0 : \sum_{i=1}^{k}h_i\mu_i = c_0$ vs $H_a :\neq c_0$, using test statistic $T = (\sum_{i=1}^{k}h_i\bar{Y}_i - c_0)/(s\sqrt{\sum_{i=1}^{k}(h_i^2/n_i)}$ ▲ anova table (source, df, sum of squares, mean squares): factor or treatment, df= $k-1, SSR = \sum_{i=1}^{k}n_i(\bar{Y}_i - \bar{Y})^2, MSR = SSR/(k-1) = s_b^2$; residual or error, df= $n-k, SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2, MSE = SSE/(n-k) = s_w^2$; total, df= $n-1, SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y})^2$ ▲$F = MSR/MSE = s_b^2/s_w^2 \sim F_{k-1,n-k}(1-\alpha)$ ▲ Bonferroni method: $1-\alpha$ simultaneous g ci, use individual ci which have confidence level $1-\alpha/g$ ▲ anova on a factor with k levels is equiv to a mlr of $Y = \beta_0 + \beta_1 z_1 + \cdots + \beta_{k-1}z_{k-1} + \epsilon$, z are indicators

△ **RANDOM EFFECT** ▲ a one-way anova model with random effects $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ where $\alpha_i \sim iid\ N(0, \sigma_\alpha^2)$ and $\epsilon_{ij} \sim iid\ N(0, \sigma_\epsilon^2)$. Random effects are ususally nuisance variable that has influence on response but we are not interested in. Intraclass correlation coefficient $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_\epsilon^2)$ ▲ randomized block experiment: by blocking we can control for the effect of different companies (a nuisance parameter), and isolate the treatment effect within a block.

△ **TWO-WAY ANOVA** (a cont. numerical response and 2 catg. predictors) ▲$Y_{ijk} = \mu_i + \nu_j + \epsilon_{ijk} = \mu + \tau_i + \alpha_j + \epsilon_{ijk}$ (additive model), constraint (either $\tau_1 = \alpha_1 = 0$ or $\sum_{i=1}^{I}\tau_i = \sum_{j=1}^{J}\alpha_j = 0$) ▲ sequential F-stat: $H_0 : \beta_{(2)} = 0, F = (SSR(\beta_{(2)} \mid \beta_{(1)}, \beta_0)/(J-1))/(MSE_{full})$ where denominator has $nIJ-(I+J-1)$ df. ▲ use cell-means plot (observed averages within each factor level combinations vs. levels of one of the factors) to see if additive model is appropriate (if parallel)

△ **ANCOVA** (basically cont+cate predictors) ▲$Y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \epsilon_{ij}$ (a cont. predictor and a factor) ▲$Y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \gamma_i x_{ij} + \epsilon_{ij}$ (a cont. predictor, a factor and an interaction) ▲ sequential $F = (SSR(\beta_{(2)} \mid \beta_{(1)}, \beta_k)/(k-1))/(MSE_{full})$, with $k-1$ numerator and $n-(k+1+(k-1)) = n-2k$ denominator df.

△ **GLMs** ▲ disadvantange of lm: extrapolation; ▲ weighted lr fixes hetero $w_i^2 = n_i/fitted(1-fitted) = Var(Y_i)^{-1}$ but wlr has its limitations too.

△ **LINK FUNCTION** ▲ logit $g(p) = \log(p/(1-p)), g^{-1}(p) = \exp p/(1+\exp p)$, probit $g(p) = \Phi^{-1}(p), g^{-1}(p) = \Phi(p)$, complementary log-log $g(p) = \log(-\log(1-p)), g^{-1}(p) = 1 - \exp(-\exp(u))$; ▲ default link function binomial (logit), Gamma (inverse), Poisson (log).

△ **LOGISTIC REG** ▲$g(p) = \log(p/(1-p)) = \beta_0 + \beta_1 X_1 + \cdots + \beta_q X_q$. response $Y \sim Bern(p)$ ▲ MLE does not require homo or normal dist, more flexible ▲ form of exponential families with dispersion (dist, $E(Y) = \mu, Var(Y) = \phi V(\mu), b(\mu), V(\mu), \phi$): $N(\mu, \sigma^2), \mu, \sigma^2, \mu, 1, \sigma^2$; $Bin(n,p), np, np(1-p), \log(\mu/(n-\mu)), \mu(n-\mu)/n, 1$; $Bin(n,p)/n, p, p(1-p)/n, \log(\mu/(1-\mu)), \mu(1-\mu), 1/n$; $Poisson(\lambda), \lambda, \lambda, \log(\mu), \mu, 1$; $Poisson(\lambda T)/T, \lambda, \lambda/T, \log(\mu), \mu, 1/T$; $Gamma(\alpha, \beta), \alpha/\beta, \alpha/\beta^2, -\mu^{-1}, \mu^2, 1/\alpha$

△ **DROP-IN DEVIANCE TEST** ▲ the likelihood ratio test $LRT = deviance_{reduced} - deviance_{full}$, compare with χ_d^2, d = difference in number of parameters

△ **DELTA METHOD** ▲ a stats approach to derive an approx prob dist for a function of an asymptotically normal estimator using the Taylor series approximation. If a seq of rv $Y_1, \ldots, Y_n$ satisfying $\sqrt{n}(Y_i - \theta) \to D\ N(0, \sigma^2)$ where θ, σ^2 are finite valued constants, then $\sqrt{n}(g(Y_i) - g(\theta)) \to D\ N(0, [g'(\theta)]^2\sigma^2)$. ▲ $w_i^2 = 1/(V(\mu_i)g'(\mu_i)^2)$ ▲ $Y \sim Poisson(\mu, \sigma^2 = \lambda)$ and $U = g(Y) = \sqrt{Y}, g'(Y) = 1/(2\sqrt{Y})$, then by delta method $E(U) = E[g(Y)] \approx g(\mu) = \sqrt{\lambda}$ and $Var(U) = Var[g(Y)] \approx [g'(\mu)]^2\sigma^2 = 1/4$ constant.

△ **CI FOR** $g^{-1}(X^T\beta)$ ▲ compute 95% ci for $X^T\beta$ as $\{L, U\}$ then apply the function $g^{-1}()$ to both L and U

△ **DEVIANCE** ▲ deviance/residual deviance $D(\hat{Y}, Y) = 2\phi\{l(Y, \phi) - l(\hat{Y} - \phi)\}$ measures the scaled difference between the log-likelihood for the observed data and hte ll of the fitted values, smaller is better

△ **SCALED DEVIANCE** ▲ exponential family errors, $D(\hat{Y}, Y) = 2\sum_{i=1}^{n}\{Y_i(\hat{\theta}_{saturated} - \hat{\theta} - b(\hat{\theta}_{saturated}) + b(\hat{\theta})\}$ ▲ likelihood ratio stat $= D^* = D(\hat{Y}, Y)/\phi \sim \chi_d^2(1-\alpha)$ ▲ two ways to test whether a coefficient is zero either by examining a t-stat for the est or by using the appropriate chi-squared test for the drop in deviance

△ **DISPERSION** ▲ dispersion parameter ϕ indicates if we have more or less than the expected variance. $\phi_{assumed} = MSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/(n-p)$ for normal, $= 1$ for binomial and Poisson, $= CV = (\sum_{i=1}^{n}((Y_i - \hat{Y}_i)/\hat{Y}_i)^2)/(n-p)$ for Gamma. CV is the estiamted coefficient of variation (relative standard deviation) for gamma dist ▲ alternative est of ϕ for all glms is $\phi_{alt} = D(\hat{Y}, Y)/(n-p)$. if $\phi_{alt} = \phi_{assumed}$, good; $\phi_{alt} < \phi_{assumed}$, under; $\phi_{alt} > \phi_{assumed}$, over. ▲ rule of thumb, for binomial or poisson, $\phi = 1$, if $(\sum_{i=i}^{n}d_i^2)/(n-p)$ is much larger than 1 then overdispersion. significant rule of thumb: $(\sum_{i=1}^{n}d_i^2)/(n-p) > 1+3\sqrt{2/(n-p)}$, similarly underdispersion if $(\sum_{i=1}^{n}d_i^2)/(n-p) < 1-3\sqrt{2/(n-p)}$ R: `glm$deviance/glm$df.residual`

△ **GOODNESS OF FIT TEST** ▲$D(\hat{Y}, Y)/\phi \sim \chi_{n-p}^2$

△ **PEARSON RESIDUAL** ▲ residual $e_i = Y_i - \hat{Y}_i$. but the variance of the Y_i's is not constant, but proportional to the variance function $V(\mu_i)$ ▲ Pearson residual as $r_i = e_i/\sqrt{V(\hat{Y}_i)}$, if the plot shows a hetero trend, there is overdispersion in the data. fix: change of link function, but often we pick a new error structure with a variance function $V(\mu)$.

△ **DEVIANCE RESIDUAL** ▲ deviance residual d_i is defined by $\sum_{i=1}^{n}d_i^2 = D(\hat{Y}, Y)$ The plot of d_i vs. linear predictor values should be examined for non-constant spread, which would indicate an overdispersion problem, and for trends, which would indicate an incorrect model structure of some kind. In addition, plot the deviance res (and Pearson res as well) vs predictors to look for trends which might indicate an incorrect link or the need for transformation on the predictors. but for discrete dist like Poisson, any residual associated with a fitted value less than 2 will cuase distortion, even when the model is appropriate; similarly, for binomial, residuals with fitted values near 0 or 1 cause distortion too.

△ **ABS RESIDUAL** ▲ absolute residuals vs. either fitted or linear predictor values can be used to check whether the variance function $V(\mu)$ of the chosen error structure is appropriate for the data. if adequate, no upward/downward trend. $V(\mu)$ models how the spread in the data changes as the expected value changes, and trends in the abs res plot indicate that the chosen relationship between the variance and the expectation is not correct. if possion is used, and

increasing trend in the plot, that means $V(\mu) = \mu$ is not fast enough, maybe $V(\mu) = \mu^2$ is needed, indicating a gamma model.

△ **PARTIAL RESIDUAL PLOT** ▲ before checking link function, we check for appropriate scale for predictors, jth predictor variable X_j's partial residual value $u_i = g(Y_i) - x_i^T \hat{\beta} + x_{i,j}\hat{\beta}_j$ vs the values of the predictor variable under investigation ($x_{i,j}$'s). expect a straight line ▲ link function adequacy can be checked by plotting $g(Y_i)$ vs fitted value of the linear predictor, straight line suggests adequacy. ▲ first scale, then link function

△ **INFLUENCE** ▲ leverage cut-off $h_{ii} > 2p/n$, p is num of parameters (including intercept) ▲ also Cook's distance, large, influential, further examined

△ **STUDENTIZED PEARSON RESIDUAL** ▲$Var(r_i) \approx Var(d_i) \approx \hat{\phi}(1 - h_{ii})$, rarely used

△ **DELETION RESIDUAL** ▲$r_i^* = (Y_i - \hat{Y}_{i,-i})/\sqrt{V(\hat{Y}_{i,-i})}$ similar idea to PRESS residuals of mlr, used for assessing outliers. rarely used too.

▲ usually use deviance residuals themselves to detect outliers

△ **TWO TESTS FOR CONTINGENCY TABLES** (test H_0 : no association between factor 1&2) ▲ likelihood ratio test: $2\sum_{i=1}^{R}\sum_{j=1}^{C} O_{ij}\log(O_{ij}/E_{ij})$ ▲ Pearson Chi-squared test: $\sum_{i=1}^{R}\sum_{j=1}^{C}((O_{ij} - E_{ij})^2/E_{ij})$ ▲ both compared with $\chi^2_{(R-1)(C-1)}$. If our stats $> \chi^2_{(R-1)(C-1)}$ then reject H_0, there is association. ▲ Pearson Chi-squared stat is only an approx, it is reliable only when a majority of E_{ij} is greater than 5. If too many under 5, either combine categories within the variables to make new E_{ij}'s larger, or rows or columns with small E_{ij}'s just to be ignored ▲ Possion Pearons res $r_{ij} = (O_{ij} - E_{ij})/\sqrt{E_{ij}}$ so perason chi-squared stat $X^2 = \sum_{i=1}^{R}\sum_{j=1}^{C} r_{ij}^2$

△ **INTERPRETATION OF PEARSON RESIDUALS** ▲(i,j)th cell with a large positive residual implies $O_{ij} \gg E_{ij}$ E_{ij} is the expected value under the independence assumption, indicating that the individuals in the jth column are more likely to be in the ith row than individuals in the other columns; large negative res indicate that $O_{ij} \ll E_{ij}$, implying individuals in the jth column are less likely to be in the ith row than individuals in the other columns

△ **MULTI-WAY CONTINGENCY TABLE** ▲ deviance stat $D^*(\tilde{\mu}, \hat{\mu}) = 2\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}(O_{ijk}\log(O_{ijk}/E_{ijk}))$ ▲ pearson chi-sq stat $X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}((O_{ijk} - E_{ijk})^2/E_{ijk})$ both compared with $\chi^2_{IJK-I-J-K+2}(1 - \alpha)$ ▲ Simpson's paradox (confounding): should test for conditional independence instead ▲ test the hypothesis of "conditional independence", $H_0 : \pi_{ijk} = \pi_{iok}\pi_{ojk}/\pi_{ook}$, expected cell counts $E_{ijk} = Y_{iok}Y_{ojk}/Y_{ook}$, df is $(I - 1)(J - 1)K$.