Stat 534: formulae referenced in lecture, week 2: randomization and residuals

Permutation for simple study designs: e.g. comparing 2 groups or simple linear regression

- The procedure:
  - Compute test statistic (e.g., drop in deviance) from observed data = T_o
  - Randomly permute group labels to the observations, compute test statistic using permuted labels
  - Repeat above many times (1000, 10000) to get many T_i
  - Calculate P[T_o more extreme than T_i], usually P[$T_o \geq T_i$]
- Assumes null hypothesis that the two groups are identical in all respects
- Can be applied to simple linear regression, assumes slope = 0
- Can be applied to block designs by randomizing within the block

Permutation for more complicated models

- Examples:
  - testing effect of factor A or factor B in a model with A + B (regression or ANOVA)
  - testing interaction, i.e. A*B, in a factorial model, A + B + A*B
- Permuting labels doesn't separate effects of each factor
- Want to allow A to have an effect, then test whether B has an effect
- Active statistical research area
- One approach that seems (where studied) to work well: Permute residuals

Residual permutation:

- Fit model, estimate residuals
- Generate permuted data that satisfies the null hypothesis you want to test
  - by setting parameters associated with the null hypothesis to 0
  - calculating predicted values under the null hypothesis
  - permuting the residuals and generating the permuted data
- Example: testing interaction = 0
  - Fit model Y $\sim$ A + B + A:B, and estimate residuals

  - Generate $\hat{Y}_0$ assuming A:B = 0 by $\hat{Y}_0$ = A + B
  - New data set is $\hat{Y}_0$ + permuted residuals
- Aside: If you generate $\hat{Y}$ from full model, A + B + A:B, you're doing a residual bootstrap
  - Very useful for confidence intervals when usual theory fails
- Assumption that makes residual permutation and residual bootstrap work:
  - Residuals are exchangeable
  - Essentially: all residuals have the same distribution

Ingredients for residuals, not all used for all residuals:

- y_i: observed value for i'th observation
- $\hat{y}_i$: predicted value for i'th observation
- s_i: standard deviation of i'th observation, $= \sqrt{\text{Var } y_i}$
- h_{ii}: leverage of the i'th observation. If y_i changes, how much does $\hat{y}_i$ change? when h_{ii} close to 1, $\hat{y}_i$ determined (essentially) only by the single observation at y_i
- d_i: deviance contribution from the i'th observation
- D: deviance $= 2(\text{lnL} - \text{lnL}_S)) = \sum d_i$
  lnL is the log likelihood a model fit to data.
  lnL_S is the log likelihood for a "saturated" model: has n parameters for n observations and fits the data perfectly
- $F(y_i \mid \theta)$: cumulative probability of y_i with distribution F and parameters $\theta = \text{P}[\text{Y} \leq y_i\ ]$
- $F^-(y_i \mid \theta)$: cumulative probability of $y_i - 1$

Residuals from simpler to more complex:

- Usual residual
$$y_i - \hat{y}_i$$
- "Simple" standardized residual
$$\frac{y_i - \hat{y}_i}{s_i}$$
- Standardized residual = Pearson residual
$$\frac{y_i - \hat{y}_i}{s_i\sqrt{1 - h_{ii}}}$$

- Deviance residual
$$\text{sign}(y_i - \hat{y}_i)\sqrt{d_i}$$
- PIT residual = Dunn-Smyth residual = randomized quantile residual
$$u_i \text{F}(y_i \mid \theta) + (1 - u_i)F^{-}(y_i \mid \theta)$$
where $u_i \sim U(0, 1)$, a uniform distribution between 0 and 1

Two examples of d_i:

- Normal distribution: $d_i = (y_i - \hat{y}_i)^2$
- Poisson distribution: $d_i = 2\left[y_i \log(y_i/\hat{\lambda}_i) - y_i + \hat{\lambda}_i\right]$

Some points on these:

- Names are not consistently used: standardized and studentized defined multiple ways
- Logic behind Pearson residual is to divide by the sd of the residual
- Which leads to residuals with constant variance
- Deviance residuals from continuous data have a standard normal distribution (mean = 0, sd 1) if model correct
- With discrete data, everything up to PIT residuals creates bands and can be very hard to interpret for 0/1 data or counts close to 0. See graphs.

Properties of PIT residuals and some details:

- PIT residuals include a random component, u_i, to smear between the band for y_i and the one "below" it
- PIT residuals, as defined above, are $U(0, 1)$ when model correct, so centered at 0.5 and ranging from 0 to 1
- Can make them more like typical residuals by applying a probit = inverse normal transformation
- Transformed residuals are standard normal when model correct and no parameters need to be estimated
- Estimated parameters tend to overfit small samples.
  - Can "spread then out" by scaling by sample standard deviation of the residuals
  - Improves performance in small samples, no effect in large samples