Some old and new tests in genetic association analysis: an introduction

Wei Pan\textsuperscript{1}

\textsuperscript{1}Division of Biostatistics, School of Public Health
University of Minnesota
July 22, 2015
Outline

• Introduction: problem

• New method: SSU test
  Some theory, connections with others, numerical results...

• Discussion

• Main refs:
  Pan (2009, *Genet Epi*), Han and Pan (2010, *Genet Epi*), Pan
  (2011, *Genet Epi*), ...
Introduction

- Single Nucleotide Polymorphisms (SNP)
  DNA seq 1 – AAGC\text{C}TA
  DNA seq 2 – AAGC\text{T}TA
  two alleles, C and T; 3 genotypes: CC, TT, CT;
  SNP: a minor allele freq (MAF) $\geq$ 5% (or 1%).
  GWAS: Genome-wide SNPs are measured as markers for each subject;

- Problem: Genome-wide \textit{association} studies (GWAS)
  Goal: to detect assoc b/w a phenotype (e.g. disease status)
  and genome-wide SNPs;
  Ultimate goal: to detect \textit{causal} genetic variants.

- The NIH Catalog of Published GWAS includes thousands of SNPs that are associated with some phenotypes, such as prostate cancer, diabetes, bipolar disorder...
• Most common study design: case-control; 
n in hundreds, then thousands, then ?
hundreds of thousands SNPs (e.g. 500K Affy arrays);
\( OR : < 1.5 \), typically, even only 1.1-1.2.
• Data:

\[
\begin{array}{cccccccc}
\text{Obs} & Y & \text{SNP1} & \ldots & \text{SNP2} & \ldots & (\text{SNP0}) & \ldots & \text{SNP}k \\
1 & 1 & \text{CT} & \ldots & \text{AG} & \ldots & \text{CG} & \ldots & \text{AC} \\
2 & 1 & \text{TT} & \ldots & \text{AG} & \ldots & \text{GG} & \ldots & \text{AA} \\
3 & 1 & \text{CT} & \ldots & \text{AA} & \ldots & \text{CG} & \ldots & \text{CC} \\
\ldots \ldots \\
1001 & 0 & \text{CT} & \ldots & \text{AG} & \ldots & \text{CC} & \ldots & \text{AC} \\
1002 & 0 & \text{TT} & \ldots & \text{GG} & \ldots & \text{CC} & \ldots & \text{AC} \\
1003 & 0 & \text{CC} & \ldots & \text{GG} & \ldots & \text{CC} & \ldots & \text{CC} \\
\ldots \ldots \\
\end{array}
\]

• A binary response: \( Y = 0 \) or 1;
each SNP \( j \) has up to 3 possible values; coded as \( X_j = 0, 1 \) or 2, though other codings are possible.

• The causal SNP0 may not be observed.

• Linkage disequilibrium (LD): SNP0 and its nearby SNPs are
correlated (and form an LD block).

⇒ If SNP0 is causal, then its nearby SNPs are associated with Y!

• Statistical question: any SNP associated with Y? univariate or multivariate?

• Here we consider \( k > 1 \) SNPs inside a given LD block or sliding window.
  Selection of LD block or window size: relevant, not trivial.

• GxG and GxE can be similarly formulated.
Existing methods

- Single-locus (or SNP-by-SNP or univariate) analysis:
  
  - Model: $Y \sim SNP_j$
    
    $\text{Logit Pr}(Y_i = 1) = \beta_{M,0j} + X_{ij}\beta_{M,j}$, \hspace{1cm} (1)
  
  - $H_{0,j}$: $\beta_{M,j} = 0$ for each $j = 1, ..., k$
    
    $\implies p_j$.
  
  - Combining: $U \min P = \min(p_1, p_2, ..., p_k)$ or ...
    
    Need to do multiple test adjustment!
    
    Time-consuming with permutation, or conservative with Bonferroni method.
    
    Analytical: sometimes; numerical integration.
  
  - Model (1): as a $2 \times 3$ table; Cochran-Armitage trend test.
• Multivariate (or global or joint) analysis:
  - Model: \( Y \sim SNP_1 + \ldots + SNP_k \)

\[
\text{Logit } \Pr(Y_i = 1) = \beta_0 + \sum_{j=1}^{k} X_{ij} \beta_j, \tag{2}
\]

- \( H_0: \beta_1 = \ldots = \beta_k = 0 \)

- Use the score, Wald or LR test:
  \( T_W = \hat{\beta}' V^{-1} \hat{\beta}, \quad T_S = U' V_U^{-1} U \sim \chi^2_k \text{ under } H_0; \)
  \( V = \text{Cov}(\hat{\beta}), \quad V_U = \text{Cov}(U); \)
  Possibly large \( DF = k. \)

- Hotelling’s \( T^2 \) test: closely related to the score test.
• Sum test
  – Working assumption: $\beta_1 = \ldots = \beta_k \equiv \beta_c$.
    in general, incorrect!
  – Model:
    \[
    \text{Logit } \Pr(Y_i = 1) = \beta_{0,c} + \sum_{j=1}^{k} X_{ij} \beta_c = \beta_{0,c} + X_{i,c} \beta_c, \tag{3}
    \]
  – $H_{0,c}: \beta_c = 0$
  – Apply the score, Wald or LR test:
    \[
    T_W = \hat{\beta}_c^2 / V_c \sim \chi_1^2 \text{ under } H_{0,c}.
    \]
  – Feature: DF=1; no multiple test!
  – Correct test size:
    \[
    H_0 \implies H_{0,c}!
    \]
  – Power: simulation results; $n = 500 + 500$
• Chapman and Whittaker (2008, *Genetic Epi*):
The UminP and a test by Goeman et al (2006, JRSS-B) work best.

• Goeman’s test:
  – Set-up: “large $k$, small $n$” as for microarray data;
  – Main idea:
    Prior for $\beta = (\beta_1, ..., \beta_k)'$: $E(\beta) = 0$, $\text{Cov}(\beta) = \tau^2 I$.
    Now test $H_{0,\tau^2}$: $\tau^2 = 0$.
  – For logistic regression:
    $T_{Go} = \frac{1}{2}(U'U - \text{Trace}(I_F))$, where $U = X'(Y - \bar{Y})$, and $I_f = \text{Cov}(U) = \bar{Y}(1 - \bar{Y})(X - \bar{X})'(X - \bar{X})$.
    – Null distribution unknown; use simulation or permutation.

• Why does Goeman’s test work here ("large $n$, small $k$")?
<table>
<thead>
<tr>
<th>Corr</th>
<th>OR</th>
<th>Sum</th>
<th>LRT</th>
<th>$T^2$</th>
<th>UminP</th>
<th>Goeman</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rand</td>
<td>1.0</td>
<td>.044</td>
<td>.048</td>
<td>.051</td>
<td>.050</td>
<td>.048</td>
</tr>
<tr>
<td>1.2</td>
<td></td>
<td>.134</td>
<td>.078</td>
<td>.079</td>
<td>.087</td>
<td>.121</td>
</tr>
<tr>
<td>1.4</td>
<td></td>
<td>.320</td>
<td>.148</td>
<td>.153</td>
<td>.200</td>
<td>.290</td>
</tr>
<tr>
<td>1.6</td>
<td></td>
<td>.546</td>
<td>.243</td>
<td>.246</td>
<td>.360</td>
<td>.523</td>
</tr>
<tr>
<td>1.8</td>
<td></td>
<td>.753</td>
<td>.383</td>
<td>.391</td>
<td>.537</td>
<td>.729</td>
</tr>
<tr>
<td>2.0</td>
<td></td>
<td>.863</td>
<td>.530</td>
<td>.540</td>
<td>.688</td>
<td>.848</td>
</tr>
</tbody>
</table>
HapMap CEU data for gene IL21R; \#SNP=27:

<table>
<thead>
<tr>
<th>$n$</th>
<th>OR</th>
<th>Sum</th>
<th>LRT</th>
<th>$T^2$</th>
<th>UminP</th>
<th>Goeman</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>1.0</td>
<td>.046</td>
<td>.098</td>
<td>.063</td>
<td>.057</td>
<td>.052</td>
</tr>
<tr>
<td>200</td>
<td>1.2</td>
<td>.078</td>
<td>.107</td>
<td>.078</td>
<td>.087</td>
<td>.087</td>
</tr>
<tr>
<td>200</td>
<td>1.4</td>
<td>.204</td>
<td>.200</td>
<td>.148</td>
<td>.256</td>
<td>.265</td>
</tr>
<tr>
<td>200</td>
<td>1.6</td>
<td>.351</td>
<td>.344</td>
<td>.275</td>
<td>.500</td>
<td>.474</td>
</tr>
<tr>
<td>500</td>
<td>1.0</td>
<td>.050</td>
<td>.054</td>
<td>.031</td>
<td>.055</td>
<td>.047</td>
</tr>
<tr>
<td>500</td>
<td>1.2</td>
<td>.165</td>
<td>.142</td>
<td>.107</td>
<td>.183</td>
<td>.204</td>
</tr>
<tr>
<td>500</td>
<td>1.4</td>
<td>.432</td>
<td>.408</td>
<td>.333</td>
<td>.652</td>
<td>.600</td>
</tr>
<tr>
<td>500</td>
<td>1.6</td>
<td>.607</td>
<td>.717</td>
<td>.667</td>
<td>.908</td>
<td>.831</td>
</tr>
</tbody>
</table>
New method

- Recall \( LRT \approx Wald's \approx \text{Score} = U' V^{-1} U \),
  \[
  U = \sum_{i=1}^{m} X_i (Y_i - \bar{Y}),
  \]
  \[
  V = \text{Cov}(U) = I_F = \bar{Y} (1 - \bar{Y}) (X - \bar{X})' (X - \bar{X}).
  \]

- New tests:
  \[
  SSU = U' U, \quad SSU_w = U' \text{diag}(V)^{-1} U.
  \]

- Null distributions for \( Q = U' W^{-1} U \):
  1) \( W = I \) and \( W = \text{Diag}(V_M) \) in the above;
  2) \( Q \sim \sum_{j=1}^{k} c_j \chi_1^2 \), where \( c_j \)'s are the eigen values of \( V_M W^{-1} \);
  3) Zhang (2005, JASA): approximate by \( a \chi_d^2 + b \) with

  \[
  a = \frac{\sum_{j=1}^{k} c_j^3}{\sum_{j=1}^{k} c_j^2}, \quad b = \sum_{j=1}^{k} c_j - \left( \frac{\sum_{j=1}^{k} c_j^2}{\sum_{j=1}^{k} c_j^3} \right)^2, \quad d = \frac{\left( \frac{\sum_{j=1}^{k} c_j^2}{\sum_{j=1}^{k} c_j^3} \right)^3}{\left( \frac{\sum_{j=1}^{k} c_j^3}{\sum_{j=1}^{k} c_j^3} \right)^2}.
  \]
4) $\Pr(SSU > s | H_0) \approx \Pr(\chi_d^2 > (s - b)/a)$.

- Wald’s versions of SSU and SSUw ...
Simulation with corr randomly b/w 0.2–0.7; #SNP=10;

\(n = 500 + 500:\)

<table>
<thead>
<tr>
<th>OR</th>
<th>Sum</th>
<th>LRT</th>
<th>UminP</th>
<th>Goeman</th>
<th>SSUw</th>
<th>SSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>.044</td>
<td>.048</td>
<td>.050</td>
<td>.048</td>
<td>.044</td>
<td>.046</td>
</tr>
<tr>
<td>1.2</td>
<td>.134</td>
<td>.078</td>
<td>.087</td>
<td>.121</td>
<td>.116</td>
<td>.114</td>
</tr>
<tr>
<td>1.4</td>
<td>.320</td>
<td>.148</td>
<td>.200</td>
<td>.290</td>
<td>.281</td>
<td>.284</td>
</tr>
<tr>
<td>1.6</td>
<td>.546</td>
<td>.243</td>
<td>.360</td>
<td>.523</td>
<td>.505</td>
<td>.500</td>
</tr>
<tr>
<td>1.8</td>
<td>.753</td>
<td>.383</td>
<td>.537</td>
<td>.729</td>
<td>.718</td>
<td>.721</td>
</tr>
<tr>
<td>2.0</td>
<td>.863</td>
<td>.530</td>
<td>.688</td>
<td>.848</td>
<td>.837</td>
<td>.836</td>
</tr>
</tbody>
</table>
HapMap CEU data for gene IL21R; \#SNP=27:

<table>
<thead>
<tr>
<th>OR</th>
<th>Sum</th>
<th>LRT</th>
<th>UminP</th>
<th>Goeman</th>
<th>SSUw</th>
<th>SSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>.046</td>
<td>.098</td>
<td>.057</td>
<td>.052</td>
<td>.047</td>
<td>.047</td>
</tr>
<tr>
<td>1.2</td>
<td>.078</td>
<td>.107</td>
<td>.087</td>
<td>.087</td>
<td>.079</td>
<td>.084</td>
</tr>
<tr>
<td>1.4</td>
<td>.204</td>
<td>.200</td>
<td>.256</td>
<td>.265</td>
<td>.265</td>
<td>.261</td>
</tr>
<tr>
<td>1.6</td>
<td>.351</td>
<td>.344</td>
<td>.500</td>
<td>.474</td>
<td>.457</td>
<td>.464</td>
</tr>
</tbody>
</table>

(n = 200)

<table>
<thead>
<tr>
<th>OR</th>
<th>Sum</th>
<th>LRT</th>
<th>UminP</th>
<th>Goeman</th>
<th>SSUw</th>
<th>SSU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>.050</td>
<td>.054</td>
<td>.055</td>
<td>.047</td>
<td>.044</td>
<td>.042</td>
</tr>
<tr>
<td>1.2</td>
<td>.165</td>
<td>.142</td>
<td>.183</td>
<td>.204</td>
<td>.208</td>
<td>.202</td>
</tr>
<tr>
<td>1.4</td>
<td>.432</td>
<td>.408</td>
<td>.652</td>
<td>.600</td>
<td>.589</td>
<td>.594</td>
</tr>
<tr>
<td>1.6</td>
<td>.607</td>
<td>.717</td>
<td>.908</td>
<td>.831</td>
<td>.836</td>
<td>.828</td>
</tr>
</tbody>
</table>

(n = 500)
• $SSU \approx SSU_w$ if $\text{diag}(V_M) \approx v1$.

• Connection b/w SSU and Goeman’s test:

$$
T_{Go} = \frac{1}{2}(Y - \bar{Y})'XX'(Y - \bar{Y}) - \frac{1}{2}\bar{Y}(1 - \bar{Y})\text{Trace}((X - \bar{X})'(X - \bar{X})),
$$

Conditional on $Y$ the second term is fixed (i.e. non-random) and can be dropped:

$$
T_{Go} = \frac{1}{2}U'U + c_0 = \frac{1}{2}U'MU_M + c_0 \propto SSU.
$$

• Why do SSU/SSUw work?

  How could they beat “optimal” score, Wald and LR tests???

• Cox and Hinkley, *Theoretical Statistics*, 1974:
  
  – Optimality of the score, Wald and LR tests: locally most powerful, but only for ...;
o/w, no uniformly most power (unbiased) (UMPU) test!

- If we knew $\beta$, then
  $$T_{MP} = \beta'U, \textbf{but} \ldots$$
- Try $\max_b b'U$ s.t. $\text{Var}(b'U) = b'I_F b = 1$?

- We estimate $T_{MP}$ by
  $$T_{EMP} = \hat{\beta}'_M U.$$

- $T_{EMP} \approx SSUw = U'Diag(I_F)^{-1}U$ because
  $$\hat{\beta}_M = I^{-1}_{M,d}U_M + O_p(m^{-1}), \quad U = U_M. \quad (4)$$

- How about estimating $\beta$ by $\hat{\beta}$?
  $$T_{EMP,J} = \hat{\beta}'U \approx U'I_F^{-1}U, \text{ which is } \ldots$$
• Connection b/w SSU and kernel machine regression (KMR):
  – KMR (Kwee et al 2008, AJHG; Wu et al 2010, AJHG): use a semi-parametric regression model

\[
\text{Logit Pr}(Y_i = 1) = \beta_0 + h(X_{i1}, \ldots, X_{ik}), \quad (5)
\]

\(h(\cdot)\) is an unknown function to be estimated. The form of \(h(\cdot)\) is determined by a user-specified positive and semi-definite (psd) kernel function \(K(\cdot, \cdot)\): by the representer theorem (Kimeldorf and Wahba 1971),

\[h_i = h(X_i) = \sum_{j=1}^{n} \gamma_j K(X_i, X_j) \text{ with some } \gamma_1, \ldots, \gamma_n.\]

– To test \(H_0: h = (h_1(X_1), \ldots, h_n(X_n))' = 0.\)
  let \(K = (K(X_i, X_j)), \gamma = (\gamma_1, \ldots, \gamma_n)',\) then \(h = K\gamma.\)

**Assume \(h\) as subject-specific random effects:**

\(E(h) = 0, Cov(h) = \tau K.\)

\(H_0 = H_0': \tau = 0.\)
Score test for $H'_0$:

$$Q = (Y - \tilde{Y}1)'K(Y - \tilde{Y}1) = SSU$$

for $H''_0$: $b = 0$ in

$$\text{Logit } \Pr(Y = 1) = b_0 + Zb$$

with $K = ZZ'$. 
• Genomic distance based regression (GDBR) (Wessel and Schork 2006, AJHG), a nonparametric MANOVA:

\[
F = \frac{tr(\hat{Y}'\hat{Y})}{tr(R'R)} = \frac{tr(\hat{Y}'\hat{Y}')}{tr(RR')} = \frac{tr(HYY'H)}{tr((I - H)YY'(I - H))} = \frac{tr(HGH)}{tr((I - H)G(I - H))} \propto SSU
\]

for \( H_0'' : b = 0 \) in

\[
\text{Logit } Pr(Y = 1) = b_0 + Zb
\]

with \( G = ZZ' \).

• A side-product (Pan 2011, Genet Epî):

KMR=GDBR=SSU if \( K = G = ZZ' \).
Application to Rare Variants

- RV: $X$ is sparse with most ($>95\%$ or $99\%$) elements as 0’s.
- Some dim reduction is necessary, e.g. variable selection;
  Most popular: pooling/collapsing SNP/SNV together, as done in the Sum test.
- Problems:
  Pooled assoc tests: bad with 1) opposite assoc directions; 2) large # neutral RVs.
- How about the SSU/SSUw and related tests?
- Some simulation results:
8 causal RVs with a common $OR = 2$; and a number of non-functional RVs. no LD.
<table>
<thead>
<tr>
<th>Test/#nfRVs</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>UminP</td>
<td>.441</td>
<td>.336</td>
<td>.296</td>
<td>.222</td>
<td>.175</td>
<td>.117</td>
</tr>
<tr>
<td>Score</td>
<td>.746</td>
<td>.632</td>
<td>.595</td>
<td>.471</td>
<td>.332</td>
<td>.245</td>
</tr>
<tr>
<td>SSU</td>
<td>.756</td>
<td>.702</td>
<td>.694</td>
<td>.626</td>
<td>.499</td>
<td>.423</td>
</tr>
<tr>
<td>SSUw</td>
<td>.743</td>
<td>.638</td>
<td>.593</td>
<td>.477</td>
<td>.339</td>
<td>.268</td>
</tr>
<tr>
<td>Sum</td>
<td>.951</td>
<td>.875</td>
<td>.808</td>
<td>.673</td>
<td>.484</td>
<td>.313</td>
</tr>
<tr>
<td>KMR(Linear)</td>
<td>.762</td>
<td>.711</td>
<td>.699</td>
<td>.631</td>
<td>.509</td>
<td>.438</td>
</tr>
<tr>
<td>KMR(Quad)</td>
<td>.755</td>
<td>.707</td>
<td>.699</td>
<td>.629</td>
<td>.501</td>
<td>.410</td>
</tr>
<tr>
<td>CMC</td>
<td>.938</td>
<td>.853</td>
<td>.777</td>
<td>.616</td>
<td>.399</td>
<td>.211</td>
</tr>
<tr>
<td>wSum</td>
<td>.940</td>
<td>.846</td>
<td>.782</td>
<td>.618</td>
<td>.424</td>
<td>.267</td>
</tr>
<tr>
<td>aSum-P</td>
<td>.933</td>
<td>.858</td>
<td>.780</td>
<td>.669</td>
<td>.499</td>
<td>.313</td>
</tr>
<tr>
<td>C-alpha-P</td>
<td>.771</td>
<td>.712</td>
<td>.688</td>
<td>.627</td>
<td>.484</td>
<td>.378</td>
</tr>
<tr>
<td>Step-up</td>
<td>.859</td>
<td>.801</td>
<td>.769</td>
<td>.679</td>
<td>.521</td>
<td>.335</td>
</tr>
</tbody>
</table>
\( OR = (3, 3, 2, 2, 1/2, 1/2, 1/2); \) no LD.
<table>
<thead>
<tr>
<th>Test/ #nfRVs</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>UminP</td>
<td>.607</td>
<td>.532</td>
<td>.481</td>
<td>.417</td>
<td>.346</td>
</tr>
<tr>
<td>Score</td>
<td>.869</td>
<td>.772</td>
<td>.721</td>
<td>.632</td>
<td>.483</td>
</tr>
<tr>
<td>SSU</td>
<td>.895</td>
<td>.835</td>
<td>.815</td>
<td>.774</td>
<td>.696</td>
</tr>
<tr>
<td>SSUw</td>
<td>.867</td>
<td>.773</td>
<td>.732</td>
<td>.633</td>
<td>.501</td>
</tr>
<tr>
<td>Sum</td>
<td>.682</td>
<td>.566</td>
<td>.465</td>
<td>.365</td>
<td>.258</td>
</tr>
<tr>
<td>KMR(Linear)</td>
<td>.897</td>
<td>.842</td>
<td>.824</td>
<td>.783</td>
<td>.707</td>
</tr>
<tr>
<td>KMR(Quad)</td>
<td>.893</td>
<td>.835</td>
<td>.815</td>
<td>.781</td>
<td>.698</td>
</tr>
<tr>
<td>CMC</td>
<td>.661</td>
<td>.544</td>
<td>.456</td>
<td>.336</td>
<td>.204</td>
</tr>
<tr>
<td>wSum</td>
<td>.659</td>
<td>.548</td>
<td>.459</td>
<td>.335</td>
<td>.228</td>
</tr>
<tr>
<td>aSum-P</td>
<td>.854</td>
<td>.745</td>
<td>.684</td>
<td>.574</td>
<td>.430</td>
</tr>
<tr>
<td>C-alpha-P</td>
<td>.906</td>
<td>.844</td>
<td>.823</td>
<td>.775</td>
<td>.674</td>
</tr>
<tr>
<td>Step-up</td>
<td>.839</td>
<td>.767</td>
<td>.724</td>
<td>.640</td>
<td>.527</td>
</tr>
</tbody>
</table>
\[ OR = (3, 1/3, 2, 2, 2, 1/2, 1/2, 1/2); \text{ with LD.} \]
<table>
<thead>
<tr>
<th>Test/#nfRVs</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>UminP</td>
<td>.489</td>
<td>.479</td>
<td>.452</td>
<td>.365</td>
<td>.318</td>
</tr>
<tr>
<td>Score</td>
<td>.599</td>
<td>.538</td>
<td>.491</td>
<td>.380</td>
<td>.276</td>
</tr>
<tr>
<td>SSU</td>
<td>.603</td>
<td><strong>.624</strong></td>
<td><strong>.635</strong></td>
<td><strong>.581</strong></td>
<td><strong>.574</strong></td>
</tr>
<tr>
<td>SSUw</td>
<td>.532</td>
<td>.561</td>
<td>.574</td>
<td>.506</td>
<td>.493</td>
</tr>
<tr>
<td>Sum</td>
<td>.342</td>
<td>.312</td>
<td>.315</td>
<td>.258</td>
<td>.239</td>
</tr>
<tr>
<td>KMR(Linear)</td>
<td>.611</td>
<td><strong>.630</strong></td>
<td><strong>.644</strong></td>
<td><strong>.597</strong></td>
<td><strong>.590</strong></td>
</tr>
<tr>
<td>KMR(Quad)</td>
<td>.545</td>
<td>.563</td>
<td>.565</td>
<td>.493</td>
<td>.474</td>
</tr>
<tr>
<td>CMC</td>
<td>.296</td>
<td>.283</td>
<td>.189</td>
<td>.182</td>
<td>.365</td>
</tr>
<tr>
<td>wSum</td>
<td>.369</td>
<td>.297</td>
<td>.287</td>
<td>.191</td>
<td>.200</td>
</tr>
<tr>
<td>aSum-P</td>
<td>.350</td>
<td>.323</td>
<td>.325</td>
<td>.258</td>
<td>.243</td>
</tr>
<tr>
<td>C-alpha-P</td>
<td><strong>.629</strong></td>
<td><strong>.650</strong></td>
<td><strong>.668</strong></td>
<td><strong>.607</strong></td>
<td><strong>.598</strong></td>
</tr>
<tr>
<td>Step-up</td>
<td>.524</td>
<td>.516</td>
<td>.532</td>
<td>.429</td>
<td>.409</td>
</tr>
</tbody>
</table>
Discussion

• No UMPU test!
  Test selection? selecting the most powerful one (Pan et al 2009, *Hum Hered*).
  Highly adaptive tests, e.g. aSPU (Pan et al 2014, *Genetics*).

• SSU: Applied to detect gene-gene and gene-environment interactions (Pan 2010 *Hum Hered*).
  aSPU?

• Main results applicable to other GLMs or regressions in general!
  Why do we always use the score/Wald/LR test in regression?
  They are not UMPU (though they are UMPI).
  Ignore correlations, as in the SSU test?
  Reduce # parameters, as in the Sum test? Tukey’s 1-DF test!
Acknowledgement: I’d like to thank my collaborators and especially my current and former students. This research was supported by NIH.

You can download our papers from
http://www.biostat.umn.edu/rrs.php

Thank you!