

# CHECKLIST FOR KNOWLEDGE DISCOVERY

c Arnold Goodman, Associate Director, UCI Center for Statistical Consulting, agoodman@uci.edu

	<b>DATA MINING</b> <i><u>databases &amp; algorithms</u></i>	<b>DATA PREDICTION</b> <i><u>processes &amp; their models</u></i>	<b>DATA KNOWLEDGE</b> <i><u>current &amp; new knowledge</u></i>
<b>OBJECTIVE</b>	explore the data for interesting patterns	exploit interesting patterns for accurate process models	explain process models for basic understanding
<b>APPROACH</b>	size is overcome to search for information	process is utilized to suggest implications	rigor is required to evaluate new knowledge
<b>CRITERIA</b>	data become findings by obtaining questions	findings become conclusions by obtaining answers	conclusions become theory by obtaining reasons
<b>FOCUS</b>	association is sought through coincidence	association is described through correlation	association is codified through relationship
<b>STATS</b>	sampling & design aid start? sequential analysis aids end?	model building aids start model estimation aids end	knowledge synthesis aids start hypothesis tests aid end
<b>PROCESS</b>	reduce the data with maximum gain	generalize the data with gain vs loss trade-off	understand the data with minimum loss
<b>BEHAVIOR</b>	break old rules powered by technology	propose new rules powered by mathematics	establish new rules powered by reason
<b>ATTITUDE</b>	seek local effects proactively & persistently	seek global potential actively & patiently	seek global reality reactively & prudently

Data mining data might be opportunistic, diverse, heterogeneous, overwhelming and perhaps non-stationary in space or time and their analysis is application-oriented and technology-driven. This situation is outside statistical traditions, but not statistical thinking. Data mining needs to advance toward statistics, if it is to expand beyond exploring for the interesting, and statistics needs to advance toward data mining if it is to expand beyond explaining the already-discovered. Where one is determines what is and should be accomplished.