

Sampling Methods and Inference Opportunities, and SAS procedures in Discrete – Categorical Response Data

In a 2x2 tabulation, three possibilities are there, data collected is based on SRS, data collected is based on StSRS, and data collected is based on randomized experiment (RE).

Method of sampling	Which margins are fixed?	Where used?	What probability distributions are represented?	What does “No association” mean?	Inferences are made to what?	Usable SAS procedures
SRS	Total sample	Cross-sectional studies	Multinomial	Independence between two random variables	To “target population”	CATMOD, GENMOD, LOGISTIC
StRS	Independent variables margins are fixed	Prospective and retrospective studies	Product multinomial	$P[\text{indept}=1]=Pr[\text{indept}=0]$, Where indep is the variable where the margins are fixed (matched samples are done)	To “ <i>target strata</i> ”	CATMOD, GENMOD, LOGISTIC
RE	Both margins are fixed (row and colum)	Market research studies	hypergeometric		To “target cells”	Fisher’s exact test, CMH in FREQ procedure

Often, converting the dependent variable in RE as a covariate, analysts use Fishers/CMH. Also, assuming that the data sampled is truly RS (with appropriate marginals fixed if any), analysts also use PROC CATMOD; the extent of bias in results is dependent on the extent of the deviation of the “truly”.