Concepts of statistical inference are often misunderstood. There has been much recent interest in developing student’s inferential reasoning in statistics. However, research on how to assess students statistical inference is scarce. This study describes the development and validation of an instrument to assess college student’s inferential reasoning in statistics. As validity is the most important aspect of psychological assessment, this study subsequently collects and integrates different sources of validity evidence of the instrument. Each of the sources is investigated to support validity arguments made about test score interpretations and uses. The analysis results from content expert reviews, student cognitive interviews, and pilot testing are presented. An argument-based approach to validity is used as a theoretical framework to examine strength of validity arguments supported by the different validity evidence. (Received September 19, 2011)