Counts and Proportions

- Often in medical and public health studies, our endpoint of interest is binary or dichotomous
  - Examples
    - disease vs. no disease
    - response vs. no response
    - death vs. no death
    - Success vs. failure
Binary Data

- Only 2 possible responses ➔
  Use proportion to summarize the data
- Often, continuous endpoints are dichotomized into a binary endpoint
  - For example, in a study of the effect of a drug on LDL levels, for each subject, the LDL measurement at the end of the study (a continuous measure) may be dichotomized into “response” vs. “no response” based on a cutpoint defining whether the LDL level has been reduced to acceptable, normal, or safe levels.
Binary Data

• Similar to hypothesis testing with continuous data, one may perform hypothesis tests on binary data:
  – 1-sample test of a proportion
    • $H_0: \ p=p_0$
    • $H_A: \ p\neq p_0$
  – 2-sample test comparing proportions
    • $H_0: \ p_1=p_2$
    • $H_A: \ p_1\neq p_2$
Binary Data

- Similar to continuous data, we may derive confidence intervals for
  - A single proportion
  - The difference between two proportions
Note

• We may use the CLT for binary data also (as the CLT applies to all distributions)
  – But note that the CLT is an asymptotic result (as $n \rightarrow \infty$)
    • Thus, we must be careful when $n$ is small
Stat-e102: Introduction to Biostatistics

Inference on Proportions

• Estimation of a population proportion
• Sampling distribution of a proportion
• Confidence intervals
• Hypothesis testing for a proportion
• Sample size estimations
• Comparison of two proportions
Inference on proportions

• Insert InfonProps.pdf