Binary data – 1 group testing, power, and sample size

Fundamental Rule of Data Analysis

Different types of data require different statistical analyses.

Notation

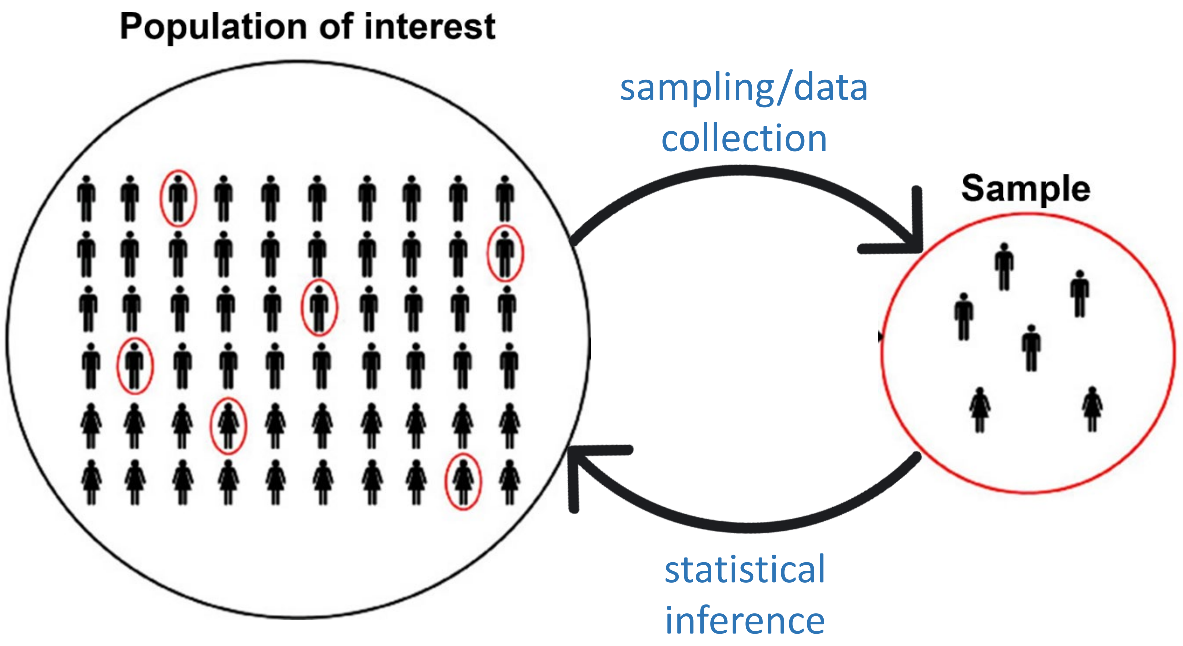
- p = population proportion

  (of subjects in the category of interest)

- $\hat{p}$ = sample proportion

  (of subjects in the category of interest)

p is fixed



$\hat{p}$ varies (would be different with a different sample of the population)

* p is not to be confused with p-value!

Goal: Does the population proportion equal a certain value?

1. State hypotheses
   - Null hypothesis (H_0): p = a certain value
   - Alternative hypothesis (H_A): $p \neq$ a certain value
2. Collect a sample
3. Assess how much the information in the sample supports the hypotheses
   - P-value
4. Reject or fail to reject the null hypothesis

Goal: Does the population proportion equal a certain value?

population proportion

1. State hypotheses
   - Null hypothesis (H_0): p = a certain value
   - Alternative hypothesis (H_A): $p \neq$ a certain value

   $\boldsymbol{p_0}$

2. Collect a sample ← **Gives you $\hat{p}$ (sample proportion)**
3. Assess how much the information in the sample supports the hypotheses
   - P-value
4. Reject or fail to reject the null hypothesis

   If p-value ≤ α, reject H_0
   If p-value > α, fail to reject H_0

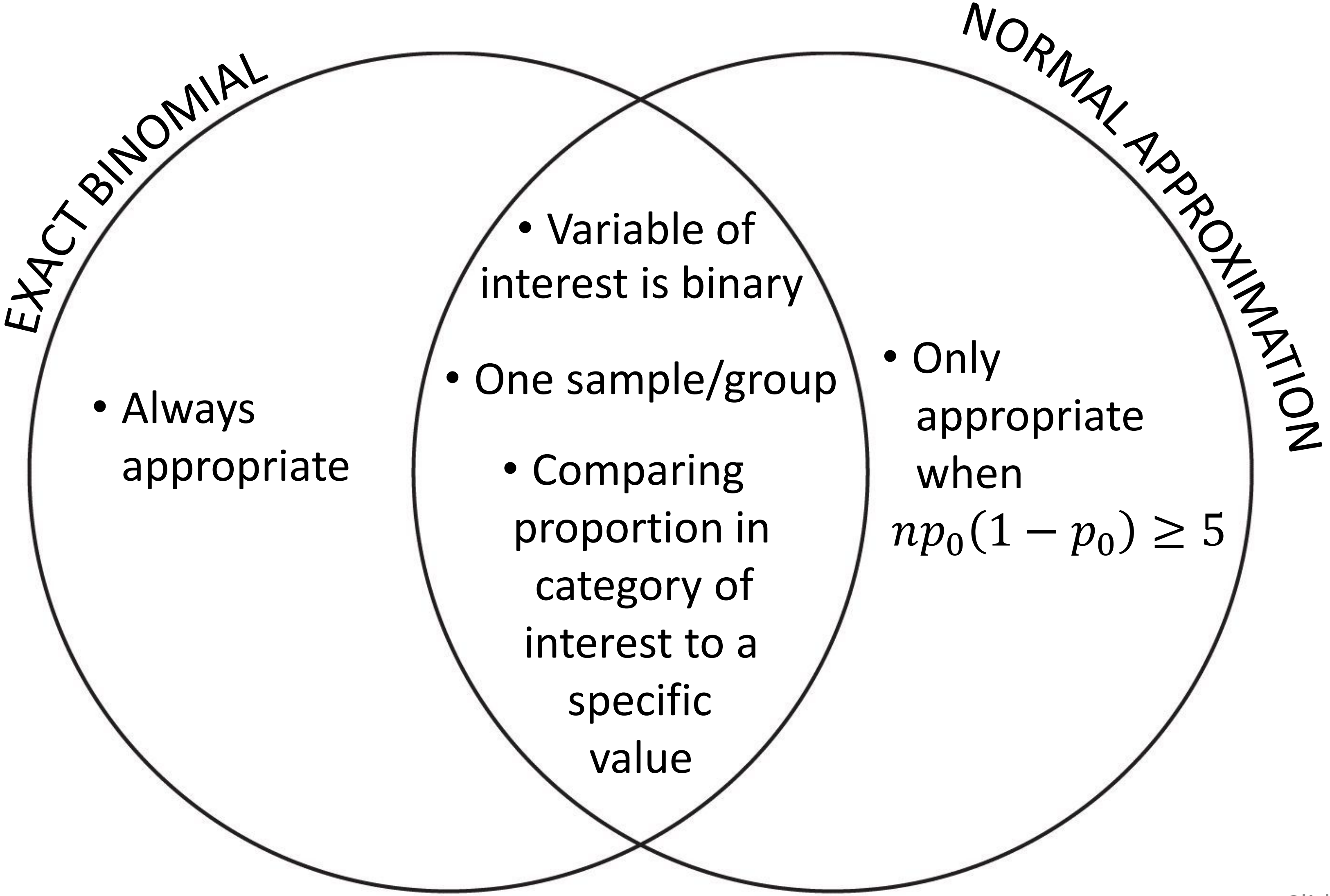
Basics of Hypothesis Testing

- To review information about…
  - Significance level (α)
  - p-value definition
  - Reject vs. fail to reject decision
  - One-sided vs. two-sided test

… see “Continuous data – 1 group testing” lecture

One-group Test of a Proportion

- Hypothesis test for a proportion in one group can be performed using one of two methods:
  - Exact binomial test
  - Normal approximation of the binomial test



Low Birth Weight Data

- Information on 100 low birth weight infants born in two teaching hospitals in Boston, Massachusetts

Variable	Description
sex	Sex of the baby (Male, Female)
birthwt	Birth weight of the baby (g)
gestage	Gestational age (weeks)
hemorrhage	Germinal matrix hemorrhage in the baby (Yes, No)
toxemia	Toxemia diagnosis for the mother (Yes, No)
apgar	Apgar score (integers, min=0, max=10). This is a scoring system used for assessing the clinical status of a newborn. 7 or higher is generally considered normal, 4-6 is low, and 3 or below is critically low.

Find the dataset (lowbwt.xlsx) and the full data dictionary (lowbwt Data Dictionary.pdf) in the Data Module on the Canvas site

Example: Toxemia Diagnosis

- Is the proportion of toxemia diagnoses among all mothers of low birth rate infants equal to 7%, which is the rate of toxemia diagnoses among mothers of normal birth weight infants? Use $\alpha = 0.05$.

Example: Toxemia Diagnosis

Exact binomial test:

```
        Exact binomial test

data:  xtabs(~toxemia, data = dat)
number of successes = 21, number of trials = 100, p-value = 5.024e-06
alternative hypothesis: true probability of success is not equal to 0.07
95 percent confidence interval:
 0.1349437 0.3029154
sample estimates:
probability of success
                  0.21
```

$H_0: p = 0.07 \quad H_A: p \neq 0.07$

p-value < 0.001

Since the p-value is less than 0.05, we have evidence to suggest that the probability of toxemia diagnosis among mothers of low birth weight infants is different from that among mothers of normal birth weight infants.

Example: Toxemia Diagnosis

Normal approximation of the binomial test:

$np_0(1 - p_0) = 100(0.07)(0.93) = 6.51$ ✅

```
        1-sample proportions test with continuity correction

data:  xtabs(~toxemia, data = dat), null probability 0.07
X-squared = 27.995, df = 1, p-value = 1.216e-07
alternative hypothesis: true p is not equal to 0.07
95 percent confidence interval:
 0.1375032 0.3052597
sample estimates:
   p
0.21
```

$H_0: p = 0.07 \quad H_A: p \neq 0.07$

p-value < 0.001

Since the p-value is less than 0.05, we have evidence to suggest that the probability of toxemia diagnosis among mothers of low birth weight infants is different from that among mothers of normal birth weight infants.

Power and Sample Size

- Each type of hypothesis test has a formula for calculating the power of the test for a given sample size OR the necessary sample size to achieve a certain power
  - For one-group test of a proportion (binary data), the <u>formula</u> is different than when we were working with t-tests of means (continuous data), but the main <u>underlying concepts</u> are the same

Basics of Power and Sample Size

- To review information about…
  - Types of errors
  - Power, consequences of an underpowered study
  - Sample size calculations

… see “Continuous data – 1 group power and sample size” lecture

Factors Affecting Power

- Sample size
  - Larger sample size → have more information → higher power
- Effect size (distance between true population proportion and hypothesized proportion (null value))
  - Bigger difference → easier to detect the difference → higher power
- Significance level
  - Larger α → threshold for rejection isn't as stringent → reject more often → higher power (and also more false positives)

Factors Affecting Necessary Sample Size

- Power
  - Want higher power → need more information → larger sample size necessary
- Effect size (distance between true population proportion and hypothesized proportion (null value))
  - Smaller difference → harder to detect the difference → larger sample size necessary
- Significance level
  - Smaller α → threshold for rejection is more stringent → reject less often → need more subjects to maintain given power

Example: Toxemia Diagnosis

- We were interested in testing whether the proportion of toxemia diagnoses among mothers of low birth weight infants equal to 7%, which is the rate of toxemia diagnoses among mothers of normal birth weight infants.
- If the proportion of toxemia diagnoses among mothers of low birth weight infants really is 21%, what power do we have with our sample to detect a difference in the proportion of toxemia diagnoses between mothers of low birth weight infants and mothers of normal birth weight infants? Use $\alpha = 0.05$.

Hypothesized p (p_0) =

True (assumed) p =

Sample size =

Significance level =

Example: Toxemia Diagnosis

Output from one-group proportion test power calculation:

```
p0 = 0.07
p_true = 0.21
n = 100
power = 0.9864278
sig.level= 0.05
alternative = two.sided
```

We have 98.6% power to detect that the proportion of toxemia diagnoses in mothers of low birth weight infants is not equal to 7%.

Example: Hemorrhage

- It has been previously reported that 20% of low birth weight infants in South America experience germinal matrix hemorrhage. Researchers hypothesize that this rate is different in North America.
- Using the low birth weight dataset as pilot data, how many subjects should they recruit in order to have 80% power to show that the rate of germinal matrix hemorrhage in low birth weight infants in North America is different from 20% (using a significance level of $\alpha = 0.05$)?

Hypothesized p (p_0) =

True (assumed) p =

Power =

Significance level =

LBW data, *hemorrhage* percentages:

No	Yes
0.85	0.15

Example: Hemorrhage

Output from one-group proportion test sample size calculation:

```
p0 = 0.2
p_true = 0.15
n = 470.46
power = 0.8
sig.level= 0.05
alternative = two.sided
```

We need 471 subjects to have 80% power to detect that the proportion of low birth weight infants with germinal matrix hemorrhage in North America is not 20%.

Example: Hemorrhage

- It has been previously reported that 20% of low birth weight infants in South America experience germinal matrix hemorrhage. Researchers hypothesize that this rate is different in North America.
- Using the low birth weight dataset as pilot data, how many subjects should they recruit in order to have **90%** power to show that the rate of germinal matrix hemorrhage in low birth weight infants in North America is different from 20% (using a significance level of $\alpha = 0.05$)?

Hypothesized p (p_0) =

True (assumed) p =

Power =

Significance level =

LBW data, *hemorrhage* percentages:

No	Yes
0.85	0.15

Example: Hemorrhage

Output from one-group proportion test sample size calculation:

```
p0 = 0.2
p_true = 0.15
n = 616.6193
power = 0.9
sig.level= 0.05
alternative = two.sided
```

We need 617 subjects to have 90% power to detect that the proportion of low birth weight infants with germinal matrix hemorrhage in North America is not 20%.

Example: Hemorrhage

- It has been previously reported that **18%** of low birth weight infants in South America experience germinal matrix hemorrhage. Researchers hypothesize that this rate is different in North America.
- Using the low birth weight dataset as pilot data, how many subjects should they recruit in order to have 90% power to show that the rate of germinal matrix hemorrhage in low birth weight infants in North America is different from **18%** (using a significance level of $\alpha = 0.05$)?

Hypothesized p (p_0) =

True (assumed) p =

Power =

Significance level =

LBW data, *hemorrhage* percentages:

No	Yes
0.85	0.15

Example: Hemorrhage

Output from one-group proportion test sample size calculation:

```
p0 = 0.18
p_true = 0.15
n = 1628.389
power = 0.9
sig.level= 0.05
alternative = two.sided
```

We need 1629 subjects to have 90% power to detect that the proportion of low birth weight infants with germinal matrix hemorrhage in North America is not 18%.

Important Points

- Hypothesis test for proportion in one group
  - Set up and interpretation
  - Exact binomial test vs. normal approximation of the binomial test
- Power and sample size calculations for one-group proportion test
  - Computation
  - Interplay between power, sample size, effect size, and significance level