STAT 518 --- Section 3.1: The Binomial Test

• Many studies can be classified as binomial experiments.

Characteristics of a binomial experiment

(1) The experiment consists of a number (denoted n) of identical trials.
(2) There are only two possible outcomes for each trial – denoted "Success" (O_1) or "Failure" (O_2)
(3) The probability of success (denoted p) is the same for each trial.
(Probability of failure = $q = 1 - p$.)
(4) The trials are independent.

Example 1: We want to estimate the probability that a pain reliever will eliminate a headache within one hour.
Example 2: We want to estimate the proportion of schools in a state that meet a national standard for excellence.
Example 3: We want to estimate the probability that a drug will reduce the chance of a side effect from cancer treatment.

• Consider a specific value of p, say p^* where $0 < p^* < 1$.

• For a test about p, our null hypothesis will be:

$$H_0: p = p^*$$

• **The alternative hypothesis could be one of:**

Two-tailed	**Lower-tailed**	**Upper-tailed**
$H_1: p \neq p^*$	$H_1: p < p^*$	$H_1: p > p^*$

• **The <u>test statistic</u> is $T =$** # of "successes" out of the n trials.

• **The null distribution of T is simply the** binomial **distribution with parameters** n and p^*.

• **Table A3 tabulates this distribution for selected parameter values (for $n \leq 20$).**

• **For examples with $n > 20$, a normal approximation may be used, or better yet, a computer can perform the exact binomial test even with large sample sizes.**

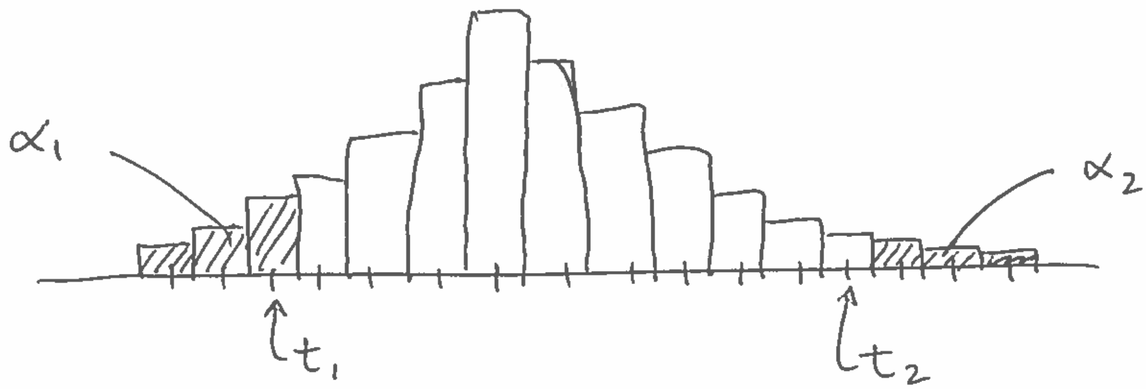
<u>Decision Rules</u>

• **Two-tailed test: We reject H_0 if T is very** small **or very** large**.**

Reject H_0 if $T \leq t_1$ or $T > t_2$.

• **How to pick the numbers t_1 and t_2?**

Picture of null distribution: Binomial (n, p^*) might look like:



• **From Table A3, using *n* and *p**, find t_1 and t_2 such that**

$$P(Y \le t_1) = \alpha_1 \quad \text{and} \quad P(Y \le t_2) = 1 - \alpha_2$$

$$\Leftrightarrow P(Y > t_2) = \alpha_2$$

where $\alpha_1 + \alpha_2 \le \alpha$.

• **Note we need P(Type I error) $\le \alpha$.** stated (nominal) significance level.

• **The P-value of the test, for an observed test statistic T_{obs}, is defined as:**

$$2 \times \left[\min\{P(Y \le t_{obs}), P(Y \ge t_{obs})\}\right]$$

where $Y \sim$ Binomial(n, p^*).

Use Table A3 (or computer) to find P-value.

• **Lower-tailed test: We reject H_0 if T is very** small.

Reject H_0 if $T \le t$.

- **We pick the critical value t such that**

$$P(Y \le t) \approx \alpha, \quad \text{where } Y \sim Bin(n, p^*)$$

- **From Table A3, using n and p^*, find t such that**

$$P(Y \le t) \le \alpha$$

- **The <u>P-value</u> of the test, for an observed test statistic T_{obs}, is:**

$$P(Y \le t_{obs})$$

where $Y \sim$ Binomial(n, p^*).

- **Upper-tailed test: We reject H_0 if T is very** <u>large</u>**.**

Reject H_0 if $T > t$.

- **We pick the critical value t such that**

$$P(Y \le t) \approx 1 - \alpha$$

- **From Table A3, using n and p^*, find t such that**

$$P(Y \le t) \ge 1 - \alpha, \quad \text{so that } P(Y > t) \le \alpha$$

- **The <u>P-value</u> of the test, for an observed test statistic T_{obs}, is:**

$$P(Y \ge t_{obs})$$

where $Y \sim$ Binomial(n, p^*).

Example 1: The standard pain reliever eliminates headaches within one hour for 60% of consumers. A new pill is being tested, and on a random sample of 17 people, the headache is eliminated within an hour for 14 of them. At $\alpha = .05$, is the new pill significantly better than the standard?

p = probability of eliminating headaches with 1 hour for new pill

Hypotheses:

H_0: $p \leq 0.6$

H_1: $p > 0.6$

$n = 17$

$p^* = 0.6$

Decision rule: Reject H_0 if $T > 13$

$P(Y \leq 12) = 0.8740$

$P(Y \leq 13) = 0.9536$

Table A3

Test statistic T = $14 > 13$, so reject H_0.

P-value = $P(T \geq 14) = 1 - P(T \leq 13)$

$= 1 - .9536 = \boxed{.0464}$

$\Rightarrow$ P-value $\leq \alpha$, so reject H_0.

Conclusion:

We conclude at $\alpha = .05$ that the new pill has a significantly higher success probability than 0.6 (standard pill's probability).

On computer: Use `binom.test` function in R (see example code on course web page)

Example 2: In the past, 35% of all high school seniors have passed the state science exit exam. In a random sample of 19 students from one school, 8 passed the exam. At $\alpha = .05$, is the probability for this school significantly different from the overall probability?

p = true proportion passing from this school

Hypotheses: $H_0: p = 0.35$

$n = 19$

$p^* = .35$

$H_1: p \neq 0.35$

Decision rule: Reject H_0 if $T \le 2$ or $T > 11$

From Table A3, $P(Y \le 2) = .0170 \rightarrow \alpha_1 = .0170$

$P(Y \le 11) = .9886 \rightarrow \alpha_2 = .0114$

$\Rightarrow \alpha_1 + \alpha_2 = .0284 \le \alpha$ ✓

Test statistic T = 8. Note $8 \not\le 2$ and $8 \not> 11$, so we do <u>not</u> reject H_0.

P-value =

$2 \times [\min\{P(T \le 8), P(T \ge 8)\}]$

$= 2 \times [\min\{.8145, (1 - .6656)\}] = 2[.3344] = \boxed{.6688}$

Conclusion: We fail to reject H_0. At $\alpha = .05$, we cannot conclude that the passing probability for this school differs from 0.35.

On computer: Use `binom.test` function in R (see example code on course web page)

Interval Estimation of p

• The binomial distribution can be used to construct exact (even for small samples) confidence intervals for a population proportion or binomial probability.

• The Clopper-Pearson CI method inverts the test of H_0: $p = p^*$ vs. H_1: $p \neq p^*$.

• This CI consists of all values of p^* such that the above null hypothesis would not be rejected, for our given observed data set.

Example 2:

• You can verify that a p^* of 0.40 would not be rejected based on our exit-exam data.

• So 0.40 would be inside the CI for p.

• But a value for p^* like 0.90 would have been rejected, so the CI for p would not include 0.90.

• In general, finding all the values that make up the CI requires a table or computer.

• Table A4 gives two-sided confidence intervals (either 90%, 95%, or 99% CIs) for p when $n \leq 30$.

• For larger samples, for one-sided CIs, or for other confidence levels, the `binom.test` function in R gives the Clopper-Pearson CI.

<u>Example 2 again</u>: Find a 95% CI for the probability that a random student for this school passes the exam.

Table A4: $n = 19, \quad Y = 8$

95% CI for p: $(.203, .665)$

• **Using R, find a 98% CI for *p*.**

98% CI: $(.173, .702)$

<u>Example 1 again</u>: Find a 90% CI for the proportion of headaches relieved by the new pill.

Table A4: $n = 17, \quad Y = 14$

90% CI: $(.604, .950)$

• **Using R, find a 90% <u>one-sided lower confidence bound</u> for *p*.**

$p \geq 0.648$

$[.648, 1] \leftarrow$ one-sided CI

• **Note: The Clopper-Pearson method guarantees coverage probability of <u>at least</u> the nominal level. It may result in an excessively wide interval.**

• **The Wilson score CI approach (use `prop.test` in R) typically gives shorter intervals, but <u>could</u> have coverage probability <u>less than</u> the nominal level.**