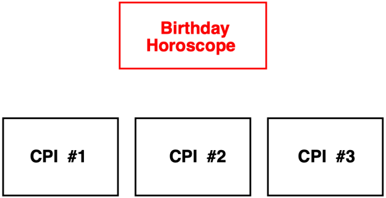
IMS 16 (ish): Inference for a single proportion

Are Astrologers' Predictions Better Than Guessing?

An astrologer prepares horoscopes for 116 adult volunteers. Each subject also filled out a California Personality Index (CPI) survey. For a given adult, his or her horoscope and CPI survey are shown to the astrologer as well as the CPI surveys for two other randomly selected adults. The astrologer is asked which survey is the correct one for that adult. The experiment was double-blind, and multiple astrologers participated in the experiment. Their accuracy rates where recorded.



National Council for Geocosmic Research (NCGR), claimed that the probability of a correct guess on any given trial in the experiment was larger than 1/3.

What is the parameter of interest? Should we consider a one-sided/two-sided test?

Hypothesis Testing with a Single Proportion

A significance test about a hypothesis has five steps.

1) Frame the research question in terms of hypotheses.
2) Collect data (check conditions)
3) Model the randomness that would occur if the null hypothesis was true.
4) Analyze the data.
5) Form a conclusion.

Step 1: Frame the research question in terms of hypotheses

Two-Sided	*One-sided*	
$H_0 : p = \frac{1}{3}$	$H_0 : p \leq \frac{1}{3}$	$H_0 : p \geq \frac{1}{3}$
$H_A : p \neq \frac{1}{3}$	$H_A : p > \frac{1}{3}$	$H_A : p < \frac{1}{3}$

We want to test the claim

$$H_0 : p \leq \frac{1}{3}$$

In other words, H_0: The rate at which astrologers are able to match a birthday/horoscope to the correct CPI result is about as good as random guessing.

Step 2: Collect Data (or Check Conditions)

- Make sure data collected meets conditions.
- Recall, by FTS the sampling distribution of $\hat{p}$ based on a sample of size n from a population with a true proportion p is nearly normal when:
  - The sample's observations are **independent**, e.g., are from a simple random sample.
  - We expected to see at least 10 successes and 10 failures in the sample, which we refer to as the **success-failure condition**. i.e.

$$np \geq 10 \text{ and } n(1-p) \geq 10$$

Step 2: Collect Data (or Check Conditions)

Example

- **Independent**: experiment wad double blind. It is reasonable to expect independence.
- **Success-Failure conditions**:
  - $np = 116(1/3) \approx 38.67 \geq 10$
  - $n(1-p) = 116(2/3) \approx 77.33 \geq 10$

Step 3: Model the randomness that would occur if the null hypothesis was true (with parametric bootstrap)

Recall, for trial $i = 1, \ldots, n$,

$$X_i = \begin{cases} 1 & \text{if condition met} \\ 0 & \text{if otherwise} \end{cases}$$

We estimate the proportion of times the

$$\hat{p} = \frac{\#\text{trials condition met}}{\#\text{ trials total}} = \frac{\sum_{i=1}^{n} X_i}{n}$$

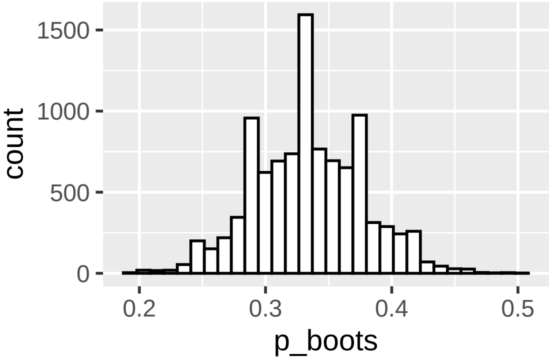
Step 3: Model the randomness that would occur if the null hypothesis was true (with parametric bootstrap)

Definition

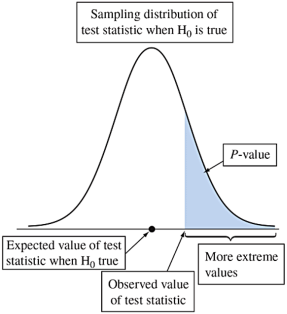
Simulating observations using a hypothesized null parameter value is often called a **parametric bootstrap simulation**.

- We generate B data sets using the same sample size $n = 116$ under the assumption that $p = 1/3$.
- We calculate $\hat{p}_{boot,1}, \hat{p}_{boot,2}, \cdots, \hat{p}_{boot,B}$.
- Plot the $\hat{p}_{boot,i}$ values in a histogram to see the sampling distribution of $\hat{p}$ under H_0.

Step 3: Model the randomness that would occur if the null hypothesis was true (with parametric bootstrap)



Step 4: Analyze the data.

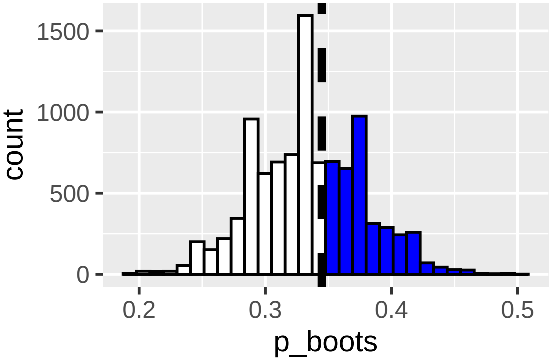


Definition

A **p-value** is the probability of observing our point estimate or something more extreme given H_0 is true.

A small p-value indicates a small likelihood of observing $\hat{p}$ if the null hypothesis were true.

Step 4: Analyze the data.

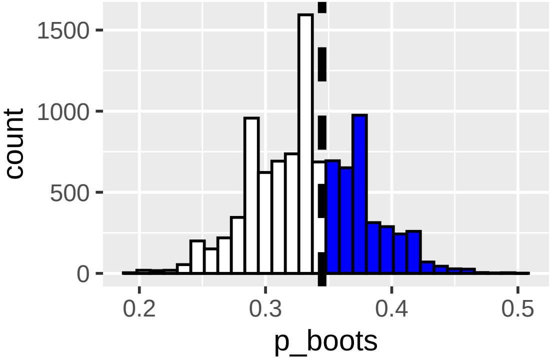


The observed 40 correct guesses out of 116, thus: $\hat{p} = 40/116 \approx .345$

The proportion of $\hat{p}_{boot}$ values simulated that where greater than $\hat{p}$ are 0.3604. Thus, our p-value is 0.3604.

Step 5: Form a conclusion.

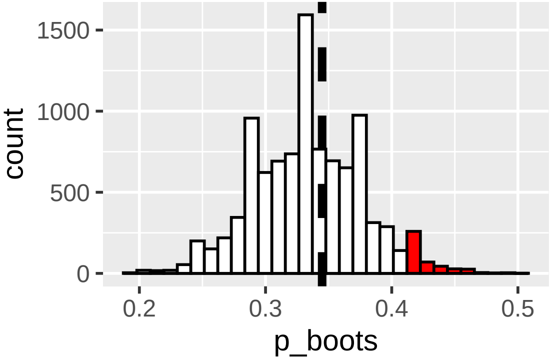
GUT CHECK:



Does the observed statistic $\hat{p}$ seem like a reasonable value to expect under the H_0?

Step 5: Form a conclusion.

- We use a **significance level** (α) to determine if we reject H_0 if the p-value is less than or equal to that number.
- Typical values are $\alpha = 0.01, 0.05, 0.10$.
- Only reject H_0 if $\hat{p}$ is in the top $\alpha 100\%$ extreme/unlikely values for the sampling distribution under H_0.



Step 5: Form a conclusion.

- What did you check?
- Are you in favor or against H_0?
- State the conclusion in context in terms of H_0.

Check	Conclusion
p-value $\leq \alpha$	Against H_0
p-value $> \alpha$	Favor H_0

Step 5: Form a conclusion.

Example

Our p-value is 0.3604 which is larger than $\alpha = 0.05$. We therefore conclude in favor of H_0. The true rate (p) at which astrologists can accurately match birthdays/horoscopes to the CPI results could be rate of random guessing.

Practice Problem

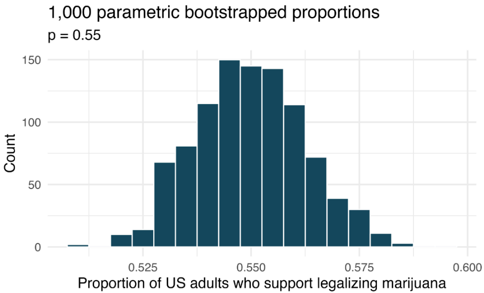
[IMS 16.6] Legalization of marijuana, bootstrap test. The 2018 General Social Survey asked a random sample of 1,563 US adults: *Do you think the use of marijuana should be made legal, or not?* 60% of the respondents said it should be made legal. Consider a scenario where, in order to become legal, 55% (or more) of voters must approve.

a) What are the null and alternative hypotheses for evaluating whether these data provide convincing evidence that, if voted on, marijuana would be legalized in the US.

b) A parametric bootstrap simulation (with 1,000 bootstrap samples) was run and the resulting null distribution is displayed in the histogram below. Find the p-value using this distribution and conclude the hypothesis test in the context of the problem.

Practice Problem

[IMS 16.6] Legalization of marijuana, bootstrap test.



a) What are the null and alternative hypotheses for evaluating whether these data provide convincing evidence that, if voted on, marijuana would be legalized in the US.

b) A parametric bootstrap simulation (with 1,000 bootstrap samples) was run and the resulting null distribution is displayed in the histogram below. Find the p-value using this distribution and conclude the hypothesis test in the context of the problem.

Hypothesis Testing with a Single Proportion using Math

Hypothesis Testing with a Single Proportion

1. Frame the research question in terms of hypotheses.
2. Collect data (check conditions)
3. Model the randomness that would occur if the null hypothesis was true.
4. Analyze the data.
5. Form a conclusion.

Steps largely remain the same. Now we model the sampling distribution under H_0 using FTS and the normal distribution, instead of parametric bootstrap distribution.

Step 3: Model the randomness that would occur if the null hypothesis was true (with math)

Recall, for trial $i = 1, \ldots, n$,

$$X_i = \begin{cases} 1 & \text{if condition met} \\ 0 & \text{if otherwise} \end{cases}$$

We estimate the proportion of times the

$$\hat{p} = \frac{\#\text{trials condition met}}{\#\text{ trials total}} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Provided we have the independence and the success-failure condition, by FTS:

$$\hat{p} \sim N\left(p, SE\right)$$

Step 3: Model the randomness that would occur if the null hypothesis was true (with math)

Standard Error for a proportion

The standard deviation of the $\widehat{p}$, otherwise known as the **standard error** for a proportion can be estimated via the shortcut:

$$SE = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$

Thus, under $H_0 : p = p_0$ the standard error of $\widehat{p}$ is

$$SE_{H_0} = \sqrt{\frac{p_0(1-p_0)}{n}}$$

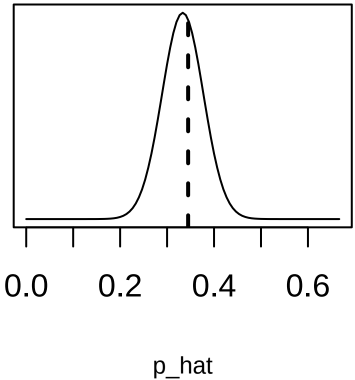
Step 3: Model the randomness that would occur if the null hypothesis was true (with math)

The standard error under $H_0 : p = 1/3$,

$$SE_{H_0} = \sqrt{\frac{\frac{1}{3}(1 - \frac{1}{3})}{116}} \approx 0.044$$

Step 3: Model the randomness that would occur if the null hypothesis was true (with math)

Sampling distribution under H_0 using math

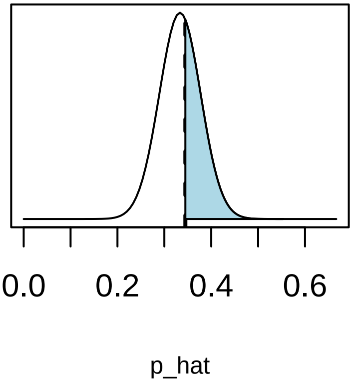


Step 4: Analyze the data.

Still calculate the p-value using the probability we observe our data or something more extreme.

```
1-pnorm(40/116, mean = 1/3, sd = 0.044)
```

```
[1] 0.3969564
```



Step 5: Form a conclusion.

- Conclusion Procedure Stays the Same
- We use a **significance level** (α) to determine if we reject H_0 if the p-value is less than or equal to that number.

Check	Conclusion
p-value $\leq \alpha$	Against H_0
p-value $> \alpha$	Favor H_0

Practice Problem

[IMS 16.8 Adjusted] Legalization of marijuana, bootstrap test. The 2018 General Social Survey asked a random sample of 1,563 US adults: *Do you think the use of marijuana should be made legal, or not?* 60% of the respondents said it should be made legal. Consider a scenario where, in order to become legal, 55% (or more) of voters must approve.

a) Check the conditions for using a mathematical model.
b) Compute the standard error using the mathematical model.
c) Draw the sampling distribution under H_0, and indicate the point estimate $(\hat{p})$ with a vertical line.
d) What code would you use in R to find the p-value using the mathematical model?
e) Do the mathematical model and parametric bootstrap give similar p-values?
f) In this setting (to test whether the true underlying population proportion is greater than 0.55), would there be a strong reason to choose the mathematical model over the parametric bootstrap (or vice versa)?

Practice Problem

Vegetarian college students. Suppose that a recent newspaper claimed that 1 in 10 college students are vegetarians. You want to test if this number is correct.

a) What are the null and alternative hypotheses to test the newspapers claim.
b) You decide to collect your own random sample of 200 students and observe that 16 are vegetarians. Check the conditions for using a mathematical model.
c) Compute the standard error using the mathematical model.
d) Draw the sampling distribution under H_0, and indicate the point estimate $(\hat{p})$ with a vertical line.
e) What code would you use in R to find the p-value using the mathematical model?