Data Analytics with R

Sumit Mishra

Institute for Financial Management and Research, Sri City

Foundations of Inference

01 December 2020

Point estimates and sampling variability

Point estimates and error

- We are often interested in *population parameters*.
- Complete populations are difficult to collect data on, so we use *sample statistics* as *point estimates* for the unknown population parameters of interest.
- *Error* in the estimate = difference between population parameter and sample statistic
- *Bias* is systematic tendency to over- or under-estimate the true population parameter.
- *Sampling error* describes how much an estimate will tend to vary from one sample to the next.
- Much of statistics is focused on understanding and quantifying sampling error, and *sample size* is helpful for quantifying this error.

Margin of error

The Lokniti-CSDS Post poll, on which this analysis is based, got the vote shares of the two major alliances reasonably close. We projected a vote share of 39% for the MGB and 36% for the NDA with an error margin of +/-3%. Eventually, the NDA got 37.3% and the MGB 37.2%. All numbers reported here have been weighted by the final outcome.

Late swing in last phase

The post-poll survey indicates a substantive last-minute swing in favour of the NDA. This was a vital factor in explaining the direction of the final verdict. One of every four respondents said they decided their vote only on the day of voting. Close to half of them voted for an NDA candidate. This trend was stronger in the last phase of voting, when more than two-thirds of those who decided on the day of voting, chose NDA candidates (Chart 1).

Margin of error

Modi tells CMs

Lockdown 3.0: Rural India may drive demand recovery amid Covid-19 crisis

About 78 per cent respondents said they were also satisfied with the steps taken by their own state government, according to the findings of the survey.

A total of 25,371 respondents were interviewed between May 30 and July 16, 2020 during the exercise and all of them were main earners of their households, and thus primarily men, the Gaon Connection said.

The survey was designed and data analysed by the Lokniti-CSDS team at the New Delhi-based Centre for Study of Developing Societies (CSDS).

It was carried out through face-to-face interviews following social distancing in 179 districts across 23 states and union territories by Gaon Connection Insights, the data and insights arm of the media platform.

ALSO READ: **Coronavirus LIVE updates: Odisha's Covid-19 cases cross 47,000-mark**

On the margin of error in the survey, Sanjay Kumar, professor at CSDS, said: "If the entire sample had been selected using the probability sampling methods, the overall margin of error for a sample as large as this (25,300 respondents) would have been +/- 1 per cent at a 95 per cent confidence level."

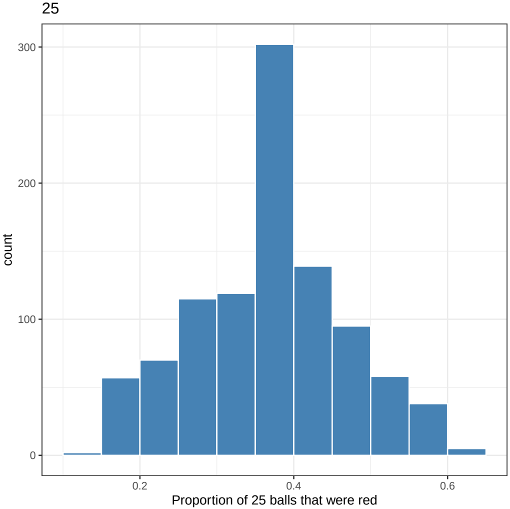
- $78\% \pm 1\%$: We are 95% confident that 77% to 79% of the public believe that the effort made by their respective state governments in curbing the coronavirus has been satisfactory.

Application

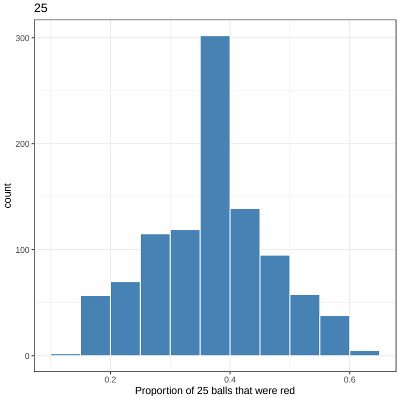
Run lines 1-26 in `R`

Sampling Distribution

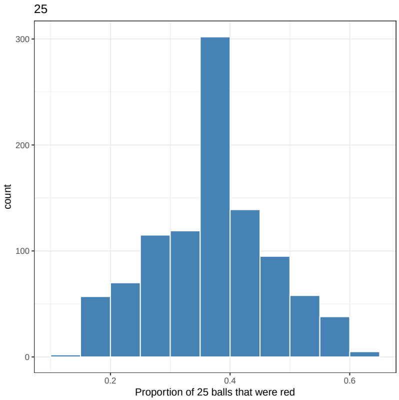
Suppose you were to repeat this process many times and obtain many $\hat{p}$s. This distribution is called a *sampling distribution*.



What is the shape and center of this distribution? Based on this distribution, what do you think is the true population proportion?



What is the shape and center of this distribution? Based on this distribution, what do you think is the true population proportion?



The distribution is unimodal and roughly symmetric. A reasonable guess for the true population proportion is the center of this distribution, approximately 0.375.

Sampling distributions are never observed

- In real-world applications, we never actually observe the sampling distribution, yet it is useful to always think of a point estimate as coming from such a hypothetical distribution.
- Understanding the sampling distribution will help us characterize and make sense of the point estimates that we do observe.

Central Limit Theorem

Central limit theorem

Sample proportions will be nearly normally distributed with mean equal to the population proportion, p, and standard error equal to $\sqrt{\frac{p\,(1-p)}{n}}$.

$$\hat{p} \sim N\left(mean = p, SE = \sqrt{\frac{p\,(1-p)}{n}}\right)$$

- It wasn't a coincidence that the sampling distribution we saw earlier was symmetric, and centered at the true population proportion.
- We won't go through a detailed proof of why $SE = \sqrt{\frac{p\,(1-p)}{n}}$, but note that as n increases SE decreases.
  - As n increases samples will yield more consistent $\hat{p}$s, i.e. variability among $\hat{p}$s will be lower.

CLT - conditions

Certain conditions must be met for the CLT to apply:

1. *Independence:* Sampled observations must be independent.
   This is difficult to verify, but is more likely if
   - random sampling/assignment is used, and
   - if sampling without replacement, $n < 10\%$ of the population.

CLT - conditions

Certain conditions must be met for the CLT to apply:

1. *Independence:* Sampled observations must be independent.
   This is difficult to verify, but is more likely if
   - random sampling/assignment is used, and
   - if sampling without replacement, $n < 10\%$ of the population.
2. *Sample size:* There should be at least 10 expected successes and 10 expected failures in the observed sample.
   This is difficult to verify if you don't know the population proportion (or can't assume a value for it). In those cases we look for the number of observed successes and failures to be at least 10.

When p is unknown

- The CLT states $SE = \sqrt{\frac{p\,(1-p)}{n}}$, with the condition that np and $n(1-p)$ are at least 10, however we often don't know the value of p, the population proportion
- In these cases we substitute $\hat{p}$ for p

When p is low

Suppose we have a population where the true population proportion is $p = 0.05$, and we take random samples of size $n = 50$ from this population. We calculate the sample proportion in each sample and plot these proportions. Would you expect this distribution to be nearly normal? Why, or why not?

When p is low

Suppose we have a population where the true population proportion is $p = 0.05$, and we take random samples of size $n = 50$ from this population. We calculate the sample proportion in each sample and plot these proportions. Would you expect this distribution to be nearly normal? Why, or why not?

No, the success-failure condition is not met ($50 * 0.05 = 2.5$)*, hence we would not expect the sampling distribution to be nearly normal.*

When the conditions are not met...

- When either np or $n(1-p)$ is small, the distribution is more discrete.
- When np or $n(1-p) < 10$, the distribution is more skewed.
- The larger both np and $n(1-p)$, the more normal the distribution.
- When np and $n(1-p)$ are both very large, the discreteness of the distribution is hardly evident, and the distribution looks much more like a normal distribution.

Extending the framework for other statistics

- The strategy of using a sample statistic to estimate a parameter is quite common, and it's a strategy that we can apply to other statistics besides a proportion.
  - Take a random sample of students at a college and ask them how many extracurricular activities they are involved in to estimate the average number of extra curricular activities all students in this college are interested in.
- The principles and general ideas are from this chapter apply to other parameters as well, even if the details change a little.

Confidence intervals for a proportion

Confidence intervals

- A plausible range of values for the population parameter is called a *confidence interval*.
- Using only a sample statistic to estimate a parameter is like fishing in a murky lake with a spear, and using a confidence interval is like fishing with a net.

We can throw a spear where we saw a fish but we will probably miss. If we toss a net in that area, we have a good chance of catching the fish.

- If we report a point estimate, we probably won't hit the exact population parameter. If we report a range of plausible values we have a good shot at capturing the parameter.

Photos by Mark Fischer (http://www.flickr.com/photos/fischerfotos/7439791462) and Chris Penny (http://www.flickr.com/photos/clearlydived/7029109617) on Flickr.

Most preferred social media platform

A recent survey of around 1100 young Indians reports Instagram as their most preferred social media platform. Estimate the true proportion of young people whose first choice is Instagram.

https://www.business-standard.com/article/technology/instagram-most-preferred-platform-among-indian-youth-survey-120100900588_1.html/

Most preferred social media platform

$$\hat{p} = 0.7 \qquad n = 1100$$

Most preferred social media platform

$$\hat{p} = 0.7 \qquad n = 1100$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 1.96 \times SE$$

Most preferred social media platform

$$\hat{p} = 0.7 \qquad n = 1100$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 1.96 \times SE$$

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7 \times 0.3}{1100}} \approx 0.014$$

Most preferred social media platform

$$\hat{p} = 0.7 \qquad n = 1100$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 1.96 \times SE$$

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7 \times 0.3}{1100}} \approx 0.014$$

$$\hat{p} \pm 1.96 \times SE \quad = \quad 0.7 \pm 1.96 \times 0.013$$

Most preferred social media platform

$$\hat{p} = 0.7 \qquad n = 1100$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 1.96 \times SE$$

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7 \times 0.3}{1100}} \approx 0.014$$

$$\begin{aligned} \hat{p} \pm 1.96 \times SE &= 0.7 \pm 1.96 \times 0.013 \\ &= (0.7 - 0.03, 0.7 + 0.03) \end{aligned}$$

Most preferred social media platform

$$\hat{p} = 0.7 \qquad n = 1100$$

The approximate 95% confidence interval is defined as

$$point\ estimate \pm 1.96 \times SE$$

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7 \times 0.3}{1100}} \approx 0.014$$

$$\begin{aligned}
\hat{p} \pm 1.96 \times SE &= 0.7 \pm 1.96 \times 0.013 \\
&= (0.7 - 0.03, 0.7 + 0.03) \\
&= (0.67, 0.73)
\end{aligned}$$

What does 95% confident mean?

- Suppose we took many samples and built a confidence interval from each sample using the equation $\textit{point estimate} \pm 1.96 \times SE$.
- Then about 95% of those intervals would contain the true population proportion (p).

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

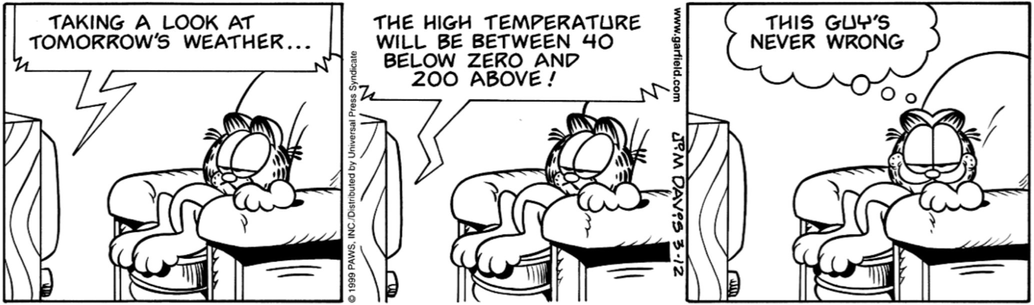
A wider interval.

Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

A wider interval.

Can you see any drawbacks to using a wider interval?

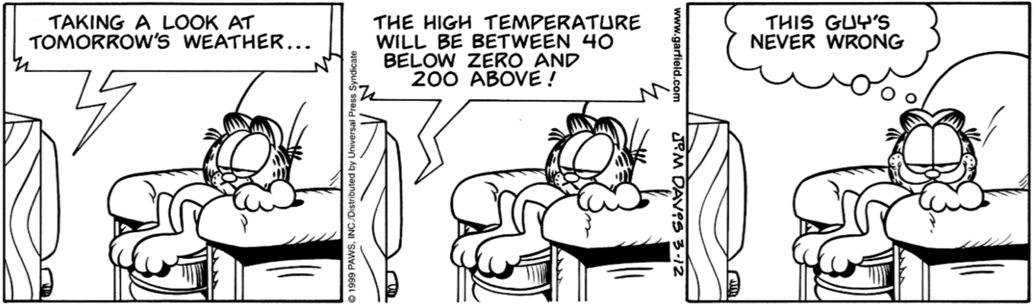


Width of an interval

If we want to be more certain that we capture the population parameter, i.e. increase our confidence level, should we use a wider interval or a smaller interval?

A wider interval.

Can you see any drawbacks to using a wider interval?



If the interval is too wide it may not be very informative.

Image source: http://web.as.uky.edu/statistics/users/earo227/misc/garfield_weather.gif

Changing the confidence level

$$\textit{point estimate} \pm z^{\star} \times SE$$

- In a confidence interval, $z^{\star} \times SE$ is called the *margin of error*, and for a given sample, the margin of error changes as the confidence level changes.
- In order to change the confidence level we need to adjust $z^{\star}$ in the above formula.
- Commonly used confidence levels in practice are 90%, 95%, 98%, and 99%.
- For a 95% confidence interval, $z^{\star} = 1.96$.
- However, using the standard normal (z) distribution, it is possible to find the appropriate $z^{\star}$ for any confidence level.

Which of the below Z scores is the appropriate $z^\star$ when calculating a 98% confidence interval?

(a) $Z = 2.05$

(b) $Z = 1.96$

(c) $Z = 2.33$

(d) $Z = -2.33$

(e) $Z = -1.65$

Which of the below Z scores is the appropriate $z^\star$ when calculating a 98% confidence interval?

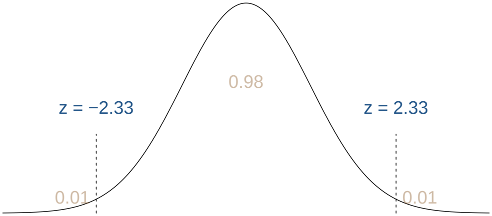
(a) $Z = 2.05$

(b) $Z = 1.96$

(c) $Z = 2.33$

(d) $Z = -2.33$

(e) $Z = -1.65$



Interpreting confidence intervals

Confidence intervals are ...

- always about the population
- not probability statements
- only about population parameters, not individual observations
- only reliable if the sample statistic they're based on is an unbiased estimator of the population parameter

Hypothesis testing for a proportion

Remember when...

Gender discrimination experiment:

		Callback		
		No	Yes	Total
Race	Black	2278	157	2435
	White	2200	235	2435
	Total	4478	392	4870

Remember when...

Gender discrimination experiment:

		Callback		
		No	Yes	Total
Race	Black	2278	157	2435
	White	2200	235	2435
	Total	4478	392	4870

$$\hat{p}_{black} \approx 0.03 \text{ and } \hat{p}_{white} \approx 0.05$$

Remember when...

Gender discrimination experiment:

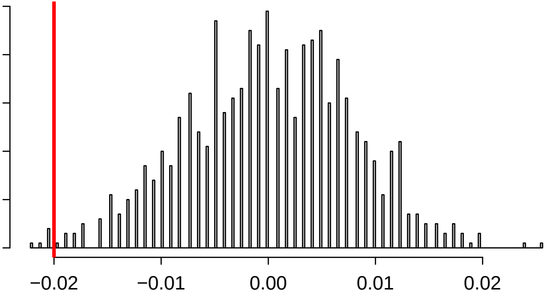
		Callback		
		No	Yes	Total
Race	Black	2278	157	2435
	White	2200	235	2435
	Total	4478	392	4870

$$\hat{p}_{black} \approx 0.03 \text{ and } \hat{p}_{white} \approx 0.05$$

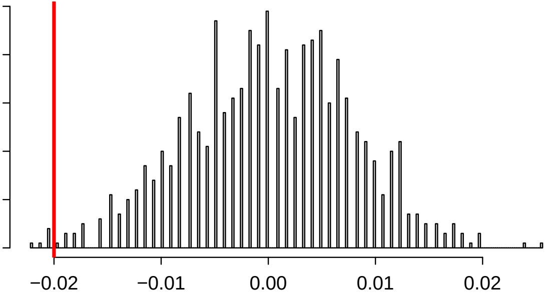
Possible explanations:

- Callback and race are *independent*, no racial discrimination, observed difference in proportions is simply due to chance. $\rightarrow$ *null* - (nothing is going on)
- Callback and race are *dependent*, there is racial discrimination, observed difference in proportions is not due to chance. $\rightarrow$ *alternative* - (something is going on)

Result



Result



Since it was quite unlikely to obtain results like the actual data or something more extreme in the simulations (callback for the African-American sounding-names 2% or lower than those for White sounding names), we decided to reject the null hypothesis in favor of the alternative.

Recap: hypothesis testing framework

- We start with a *null hypothesis* (H_0) that represents the status quo.

Recap: hypothesis testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We also have an *alternative hypothesis (H_A)* that represents our research question, i.e. what we're testing for.

Recap: hypothesis testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We also have an *alternative hypothesis (H_A)* that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem (coming up next...).

Recap: hypothesis testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We also have an *alternative hypothesis (H_A)* that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem (coming up next...).
- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

Recap: hypothesis testing framework

- We start with a *null hypothesis (H_0)* that represents the status quo.
- We also have an *alternative hypothesis (H_A)* that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or traditional methods based on the central limit theorem (coming up next...).
- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

We'll formally introduce the hypothesis testing framework using an example on testing a claim about a proportion.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true		
	H_A true		

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$\checkmark$	
	H_A true		

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$\checkmark$	
	H_A true		$\checkmark$

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$\checkmark$	*Type 1 Error*
	H_A true		$\checkmark$

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$\checkmark$	*Type 1 Error*
	H_A true	*Type 2 Error*	$\checkmark$

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.

Decision errors (cont.)

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✓	*Type 1 Error*
	H_A true	*Type 2 Error*	✓

- A *Type 1 Error* is rejecting the null hypothesis when H_0 is true.
- A *Type 2 Error* is failing to reject the null hypothesis when H_A is true.
- We (almost) never know if H_0 or H_A is true, but we need to consider all possibilities.

Hypothesis Test as a trial

If we think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

$$\begin{aligned} H_0 &: \text{ Defendant is innocent} \\ H_A &: \text{ Defendant is guilty} \end{aligned}$$

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty
- Declaring the defendant guilty when they are actually innocent

Hypothesis Test as a trial

If we think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

$$H_0 : \text{ Defendant is innocent}$$
$$H_A : \text{ Defendant is guilty}$$

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

  Type 2 error

- Declaring the defendant guilty when they are actually innocent

Hypothesis Test as a trial

If we think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

$$H_0: \text{ Defendant is innocent}$$
$$H_A: \text{ Defendant is guilty}$$

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

  Type 2 error

- Declaring the defendant guilty when they are actually innocent

  Type 1 error

Hypothesis Test as a trial

If we think of a hypothesis test as a criminal trial then it makes sense to frame the verdict in terms of the null and alternative hypotheses:

$$\begin{aligned} H_0 &: \text{ Defendant is innocent} \\ H_A &: \text{ Defendant is guilty} \end{aligned}$$

Which type of error is being committed in the following circumstances?

- Declaring the defendant innocent when they are actually guilty

  Type 2 error

- Declaring the defendant guilty when they are actually innocent

  Type 1 error

Which error do you think is the worse error to make?

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error if the null hypothesis is true.

$$P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$$

Type 1 error rate

- As a general rule we reject H_0 when the p-value is less than 0.05, i.e. we use a *significance level* of 0.05, $\alpha = 0.05$.
- This means that, for those cases where H_0 is actually true, we do not want to incorrectly reject it more than 5% of those times.
- In other words, when using a 5% significance level there is about 5% chance of making a Type 1 error if the null hypothesis is true.

$$P(\text{Type 1 error} \mid H_0 \text{ true}) = \alpha$$

- This is why we prefer small values of α – increasing α increases the Type 1 error rate.

Example

Lines 174-193 in R

Setting the hypotheses

- The *parameter of interest* is the proportion of red balls.

Setting the hypotheses

- The *parameter of interest* is the proportion of red balls.
- There may be two explanations why our sample proportion is lower than 0.375.
    - The true population proportion is different than 0.375.
    - The true population mean is 0.375, and the difference between the true population proportion and the sample proportion is simply due to natural sampling variability.

Setting the hypotheses

- We start with the assumption that 37.5% of balls are red.

$$H_0 : p = 0.375$$

Setting the hypotheses

- We start with the assumption that 37.5% of balls are red.

$$H_0 : p = 0.375$$

- We test the claim that the proportion of red balls is different than 37.5%

$$H_A : p \neq 0.375$$

Test statistic

In order to evaluate if the observed sample proportion is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

Test statistic

In order to evaluate if the observed sample proportion is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

$$\hat{p} \sim N\left(\mu = 0.375, SE = \sqrt{\frac{0.375 \times 0.625}{50}}\right)$$

Test statistic

In order to evaluate if the observed sample proportion is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

$$\hat{p} \sim N\left(\mu = 0.375, SE = \sqrt{\frac{0.375 \times 0.625}{50}}\right)$$

$$Z = \frac{0.3 - 0.375}{0.07} = -1.1$$

Test statistic

In order to evaluate if the observed sample proportion is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

$$\hat{p} \sim N\left(\mu = 0.375, SE = \sqrt{\frac{0.375 \times 0.625}{50}}\right)$$

$$Z = \frac{0.3 - 0.375}{0.07} = -1.1$$

The sample proportion is -1.1 standard errors away from the hypothesized value. Is this considered unusually low? That is, is the result *statistically significant*?

Test statistic

In order to evaluate if the observed sample proportion is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.

$$\hat{p} \sim N\left(\mu = 0.375, SE = \sqrt{\frac{0.375 \times 0.625}{50}}\right)$$

$$Z = \frac{0.3 - 0.375}{0.07} = -1.1$$

The sample proportion is -1.1 standard errors away from the hypothesized value. Is this considered unusually low? That is, is the result *statistically significant*?

Yes, and we can quantify how unusual it is using a p-value.

p-values

- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.

p-values

- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.
- If the p-value is *low* (lower than the significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject* H_0.

p-values

- We then use this test statistic to calculate the *p-value*, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis were true.
- If the p-value is *low* (lower than the significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence *reject H_0*.
- If the p-value is *high* (higher than α) we say that it is likely to observe the data even if the null hypothesis were true, and hence *do not reject H_0*.

Virtual bowl example - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample proportion lower than 0.3), if in fact H_0 were true (the true population proportion was 0.375).

Virtual bowl example - p-value

p-value: probability of observing data at least as favorable to H_A as our current data set (a sample proportion lower than 0.3), if in fact H_0 were true (the true population proportion was 0.375).

$$P(|Z| > 1.1) < 0.14$$

The Bowl Example

- p-value > 0.05

The Bowl Example

- p-value > 0.05

The Bowl Example

- p-value > 0.05
- Since p-value is *high* (higher than 5%) we *don't reject H_0*.

The Bowl Example

- p-value > 0.05
- Since p-value is *high* (higher than 5%) we *don't reject H_0*.
- The difference between the null value of 0.375 and observed sample proportion of 0.3 is *due to chance* or sampling variability.

Choosing a significance level

- While the the traditional level is 0.05, it is helpful to adjust the significance level based on the application.
- Select a level that is smaller or larger than 0.05 depending on the consequences of any conclusions reached from the test.
- If making a Type 1 Error is dangerous or especially costly, we should choose a small significance level (e.g. 0.01). Under this scenario we want to be very cautious about rejecting the null hypothesis, so we demand very strong evidence favoring H_A before we would reject H_0.
- If a Type 2 Error is relatively more dangerous or much more costly than a Type 1 Error, then we should choose a higher significance level (e.g. 0.10). Here we want to be cautious about failing to reject H_0 when the null is actually false.

One vs. two sided hypothesis tests

- In two sided hypothesis tests we are interested in whether p is either above or below some null value p_0: $H_A : p \neq p_0$.
- In one sided hypothesis tests we are interested in p differing from the null value p_0 in one direction (and not the other):
  - If there is only value in detecting if population parameter is less than p_0, then $H_A : p < p_0$.
  - If there is only value in detecting if population parameter is greater than p_0, then $H_A : p > p_0$.
- Two-sided tests are often more appropriate as we often want to detect if the data goes clearly in the opposite direction of our alternative hypothesis as well.