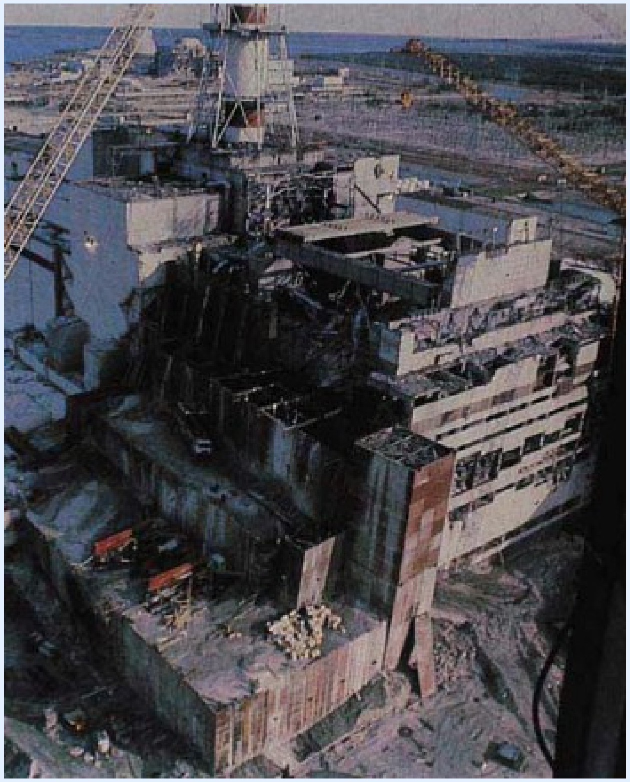
Math 140
Introductory Statistics

Next midterm May 1st

8.2 P-values

In 1986 the nuclear reactor at Chernobyl, Ukraine leaked radioactive material, generating concerns about DNA mutations in humans and animals.



Chernobyl

Barn swallows usually have red or blue feathers. Normal genetic mutations occur at a 2% rate, giving rise to white feathers or other abnormalities instead.

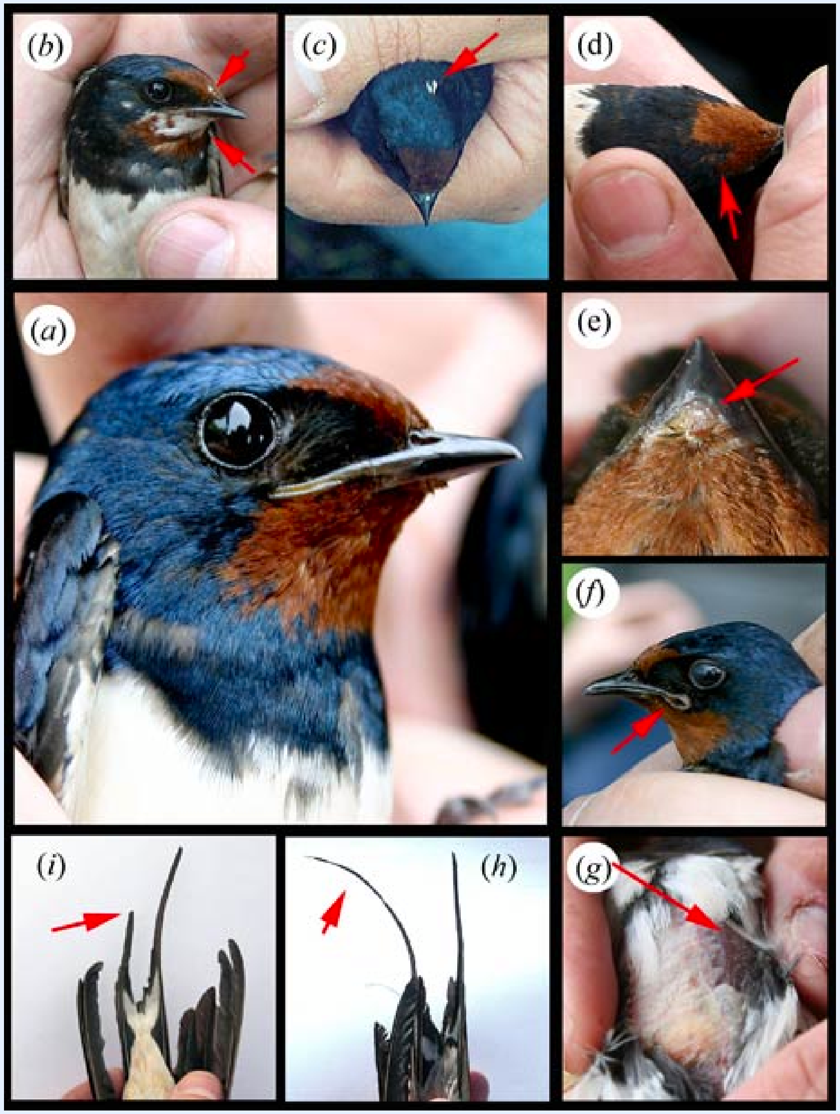
In Chernobyl between 1991 and 2000 it was observed that on a sample of 266 birds, 16% of them had white feathers.

Chernobyl

Normal genetic mutations : 2% rate

Chernobyl (266 birds) genetic mutations: 16%

What can we conclude from our survey about correlations between the leak and genetic mutations?



Chernobyl birds

p for the entire population is 0.02
(the proportion of mutated birds)

Our sampling proportion (what we measured) is

$\hat{p}$ = 0.16 For n = 266

Does the survey present convincing evidence that there was a higher mutation rate in Chernobyl birds?

Chernobyl birds

p for the entire population is 0.02
(the proportion of mutated birds)

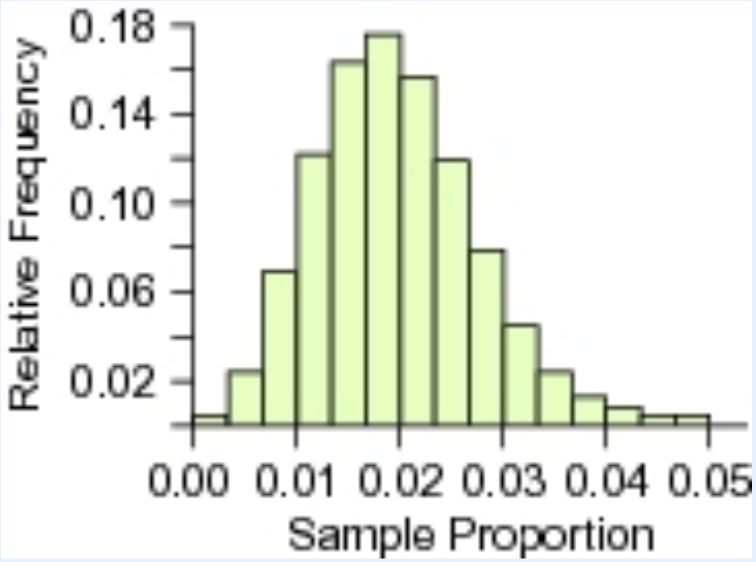
Our sampling proportion (what we measured) is

$\hat{p}$ = 0.16 For n = 266

Lets look at the sampling distribution for $\hat{p}$
from samples of n = 266

That is, let's go out there and sample groups
of n=266 birds and see their mutation rates

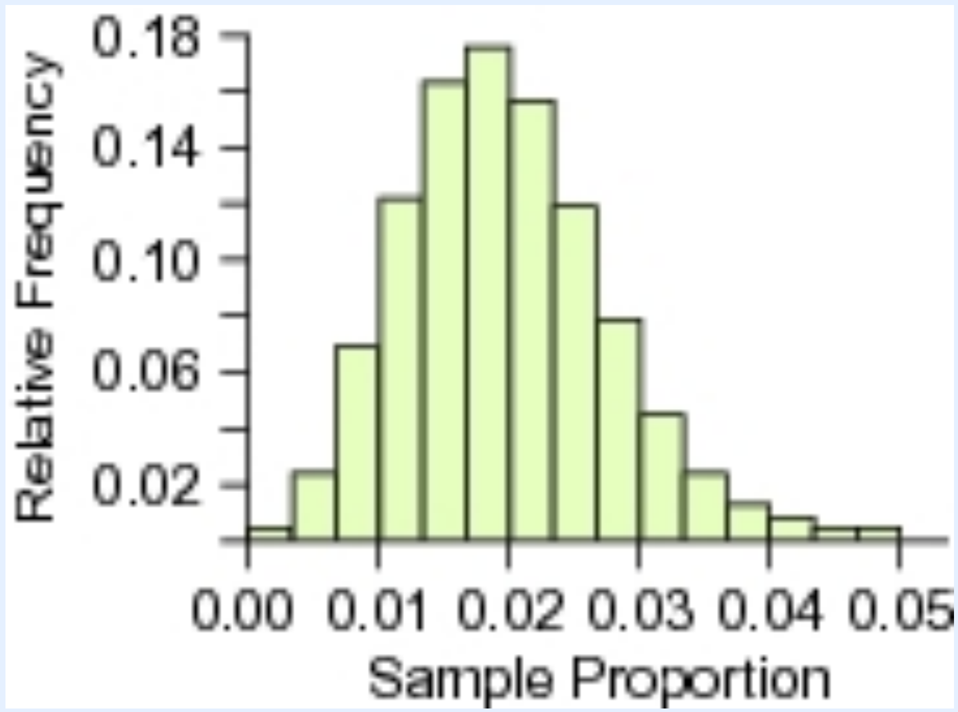
Mutated birds sampling distribution



The mean is 0.02. Is this unexpected?
The distribution is approximately normal.

The value 0.16 is way out there from this chart!

Most likely mutations were due to radiation



Since from the sampling distribution
the value of 0.16 is extremely unlikely

Astrology and Psychics

Do the positions of stars and planets on our birth dates really affect our lives?

Astrology and Psychics

Natal charts (horoscopes based on birth dates and times) were prepared for 83 people

People were given 3 of them, their own and that of two other people at random.

They were then told to pick which one most adequately described them

Astrology and Psychics

28 out of 83 picked the correct one,
made for their own birth dates.

Does the experiment provide convincing evidence
That a person's natal chart describes them better
Than a random one?

What should we compare our results to?

Astrology and Psychics

28 out of 83 picked the correct one, made for their own birth dates.

What should we compare this to?

$p = 1/3$

This is the population proportion we would expect if the horoscope selection were totally random.

This is the standard we want to compare to

What is our success rate?

$$\hat{p} = ?$$

Here we use samples of size n = 83 and
Want to judge whether our observation of $\hat{p}$
is likely or not.

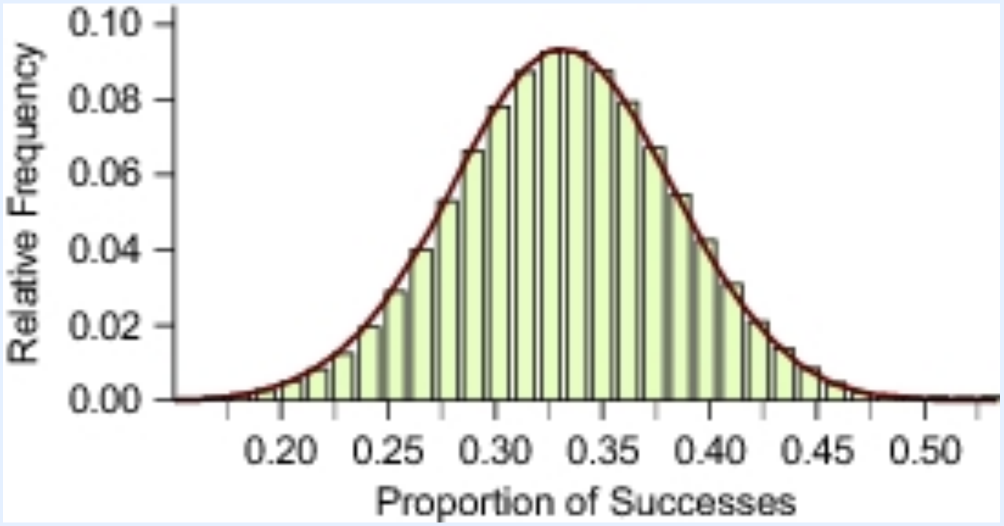
What is our success rate?

Well, we found that 28 out of 83 selected the right answer, so our sample proportion is

$$\hat{p} = 28/83 = 0.337$$

We need to compare this to p = 1/3 = 0.333

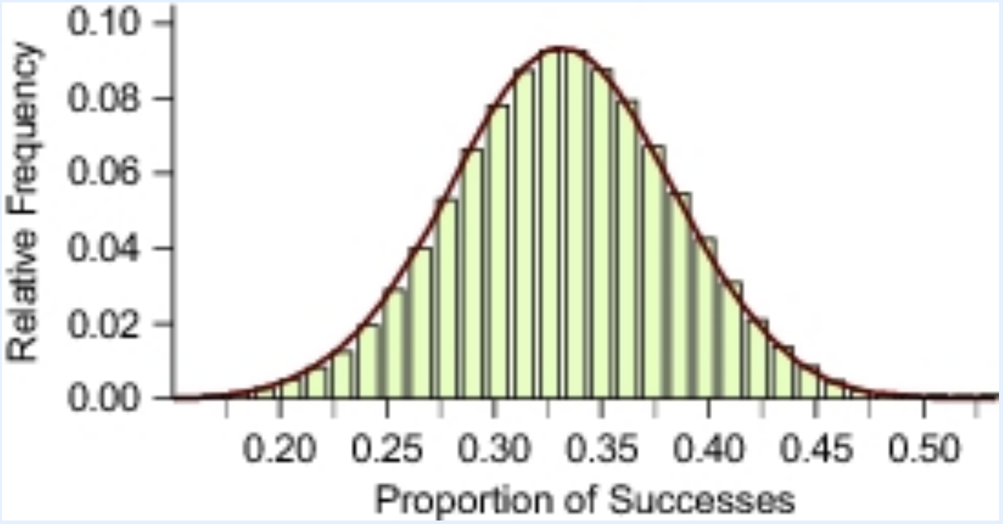
Construct the sampling distribution



Here we use samples of size n = 83
Our observed value of 0.337 is very close to the middle of this normal distribution

Most likely horoscope selection is by chance and astrological charts don't really tell you anything

Construct the sampling distribution



This study was approved by the National Council of Geocosmic Reseacrh, an organization of US astrologers.

Burdens of proof

Astrologers need to show that they do a much much better job than just guessing

In the case of Chernobyl, the sampling distribution value of 16% was so much higher than 2% that we can conclude that mutations are not by chance.

Burdens of proof

Astrologers failed but ONLY within the cohort of people they surveyed. We don’t know anything about the larger population.

In the case of Chernobyl since we had population data we could safely conclude that the mutations are not happening at their natural occurrence rate.

Null and alternative hypothesis

In formal research the null hypothesis is when nothing has changed, all populations are equal, everything is fair and all outcomes are likely

This is what you'd expect if there was no intervening, result altering process
It is the standard case

The alternative hypothesis is where we state the nature of the change, we expect the outcomes to be larger or smaller than what the null hypothesis predicts due to intervening, result altering processes

The Null Hypothesis and the Alternative Hypothesis

If you are doing a test of significance for a proportion and wish to compare the results from a sample to a standard value, p_0, begin by writing the **null hypothesis**.

The null hypotheses for a sample survey may be worded using either form below:

H_0: The proportion, p, of successes in the population from which the sample was taken is equal to the hypothesized, or standard, value, p_0.

H_0: $p = p_0$, where p is the proportion of successes in the population from which the sample was taken.

The null hypothesis for a study involving the probability of a success would be worded this way:

H_0: The probability, p, of a success on any one trial is equal to the hypothesized, or standard, value, p_0.

The **alternative hypothesis** has three forms, depending on what you want to establish. The following are the three forms for sample surveys.

H_a: $p \neq p_0$ The proportion of successes, p, in the population is not equal to the hypothesized value p_0.

H_a: $p > p_0$ The proportion of successes, p, in the population is greater than the hypothesized value p_0.

H_a: $p < p_0$ The proportion of successes, p, in the population is less than the hypothesized value p_0.

The first form defines a **two-sided test**. The second and third define **one-sided tests.**

Let's formalize the Chernobyl study

Null hypothesis

$H_0: p = 0.02$

Standard: Radiation did not cause mutations

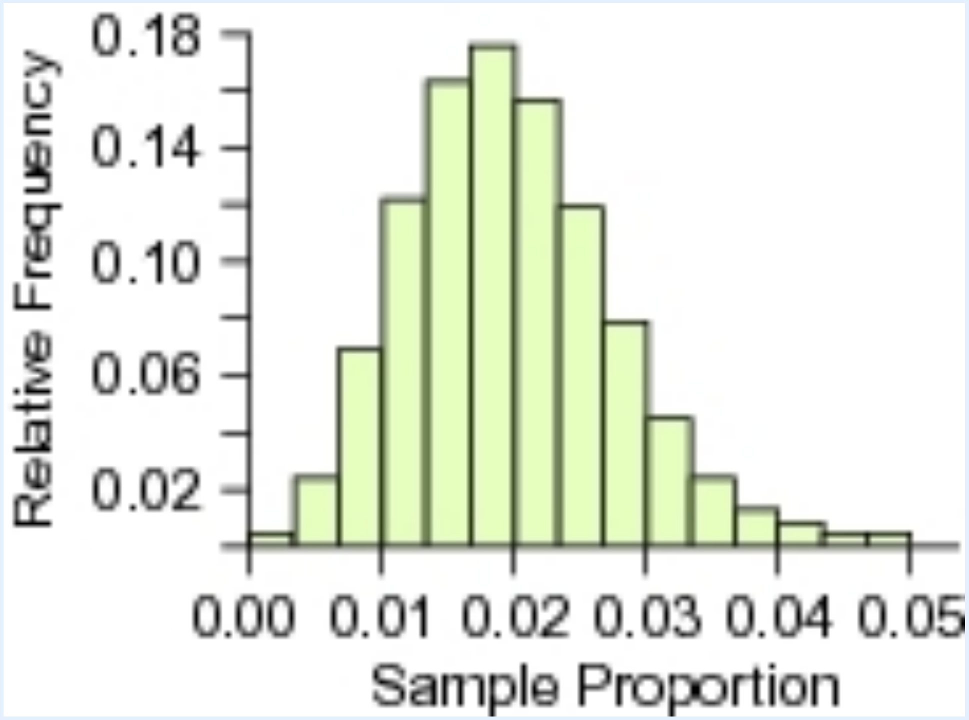
Alternative hypothesis

$H_a: p > 0.02$

To provide evidence that mutations were caused by radioactive leaks

Let's formalize the Chernobyl study

We do have enough evidence to prove that the alternative hypothesis is correct



Let's formalize the astrology study

Null hypothesis

H_0: $p = 0.333$

Standard: Given 3 astrological charts of which only one is correct, a person will pick it at random

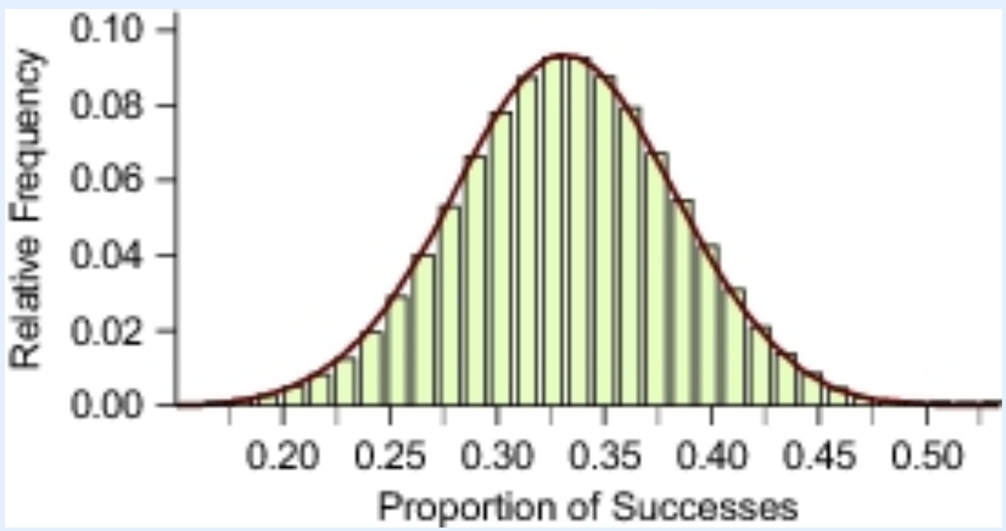
Alternative hypothesis

H_a: $p > 0.333$

To provide evidence that the correct astrological charts are not picked at random

Let's formalize the astrology study

We don't have enough evidence to prove that the alternative hypothesis is correct



DISCUSSION
Hypotheses

D10. For each of these situations, give the value of the standard p_0, say whether the situation calls for a one-sided or two-sided test, write the null and alternative hypotheses, and give the value of the sample proportion, $\hat{p}$.

a. You want to see if people can identify the gourmet coffee from three cups of coffee containing the gourmet coffee, ordinary coffee, and instant coffee. You give 100 randomly selected people a taste of each (in random order) and 52 people correctly choose the gourmet coffee.

b. You suspect gender discrimination in hiring in your local police department. Forty percent of the applicants are women. A random sample of employees finds that only 15% are women.

DISCUSSION
Hypotheses

D10. For each of these situations, give the value of the standard p_0, say whether the situation calls for a one-sided or two-sided test, write the null and alternative hypotheses, and give the value of the sample proportion, $\hat{p}$.

a. You want to see if people can identify the gourmet coffee from three cups of coffee containing the gourmet coffee, ordinary coffee, and instant coffee. You give 100 randomly selected people a taste of each (in random order) and 52 people correctly choose the gourmet coffee.

a) p0 = 1/3

One sided - we are asking only if the gourmet coffee can be identified or not

Null hypoth. H0 p = 1/3 Alternative hypoth. Ha p > 1/3

Our sampling $\hat{p}$ = 52/100 = 0.52

b. You suspect gender discrimination in hiring in your local police department. Forty percent of the applicants are women. A random sample of employees finds that only 15% are women.

a) $p0 = 0.4$ not 0.5 because only 40% of women are applying!

Two sided -
if our $p > 0.4$ then potential discrimination towards men if $p < 0.4$ towards women

Null hypoth. H0 $p = 0.4$
Alternative hypoth. Ha p different 0.4

Our sampling $\hat{p} = 0.15$

These were ‘easy’ cases

It was easy to see that the Chernobyl values were very different from each other and the astrology ones very similar.

How to do this in a systematic way?

A test statistic

Let's calculate the test statistic z

$$z = \frac{statistic - parameter}{std.error} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

This will tell us how many standard errors we are away from the null hypothesis value

A test statistic for natal charts

Let's calculate the test statistic z

$$z = \frac{statistic - parameter}{standard\ error} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = ?$$

How many standard errors we are away from the null hypothesis value in the case of the astrology study?

A test statistic for natal charts

Let's calculate the test statistic z

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.337 - 0.333}{\sqrt{\frac{0.333(1 - 0.333)}{83}}} = 0.08$$

We are only 0.08 std. errors away from the null hypothesis value

The P value

Is the probability of finding a result
that is way more extreme than we would
find by using the
null hypothesis alone

Assume null hypothesis p0

1) Make sure our sampling distribution is normal and that

$np_0 > 10$ And that $n(1 - p_0) > 10$

2) Calculate the z statistic (z score)

3) Use the table for Standard Normal probabilities (page 759) to find the area that falls outside the z score depending on your alternative hypothesis

A concrete example

We know that about 60% of students are math-anxious
What about on our own campus? We study 100 students.

Null hypothesis $p = 0.6$

We find from our survey that our test statistic is $z = 1.84$.
We believe the proportion is higher than 0.6.

Our alternate hypothesis is that $p > 0.6$

A concrete example

1) Check 100* (0.6) = 60 and 100 *(0.4) = 40
they are both bigger than 10 so we can assume
That the sampling distribution is normal.

$$np_0 > 10 \qquad\qquad n(1 - p_0) > 10$$

We believe we are MORE scared

This is a one sided test.
Alternative hypothesis $p > 0.6$

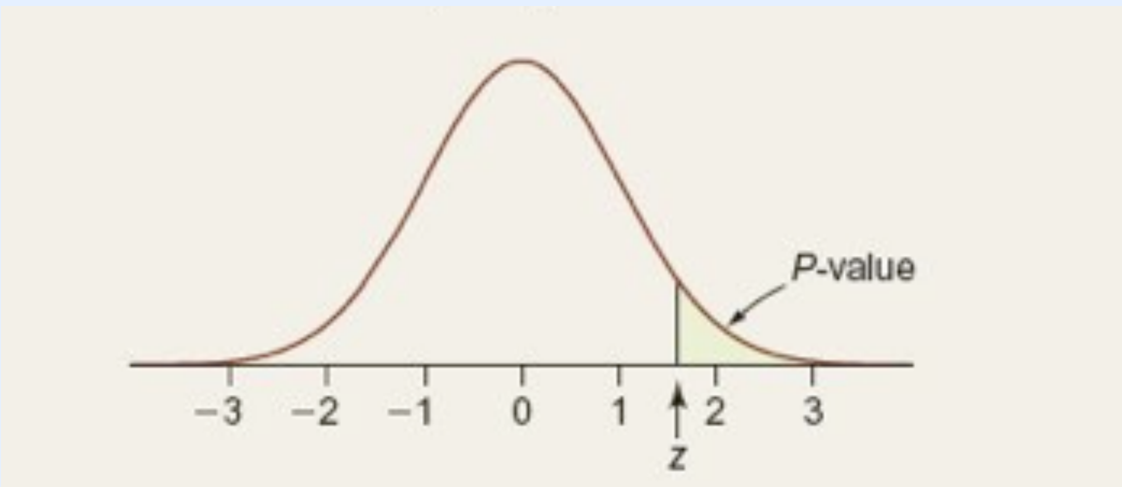
Evidence is against the null hypothesis
and in favor of the alternative hypothesis if
our measurements give
$\hat{p} > 0.6$

2) They gave us $z=1.84$

We believe we are MORE scared

3) Since we are asking whether on our campus the Proportion is larger, we need to calculate The area above the given z-score.

From page 759, we find the area below z=1.84 to be 0.9671



z = 1.84

This means the area above is
1 - 0.9671 = 0.0329

We believe we are MORE scared

The P value is 0.0329 and this means that IF the true value of $p = 0.6$ and we take a sample distribution, then the probability of finding $z > 1.84$ is equal to 0.0329

About 3%

Now, we DID measure $z=1.84$.
What does this tell us about the null hypothesis?

We believe we are LESS scared

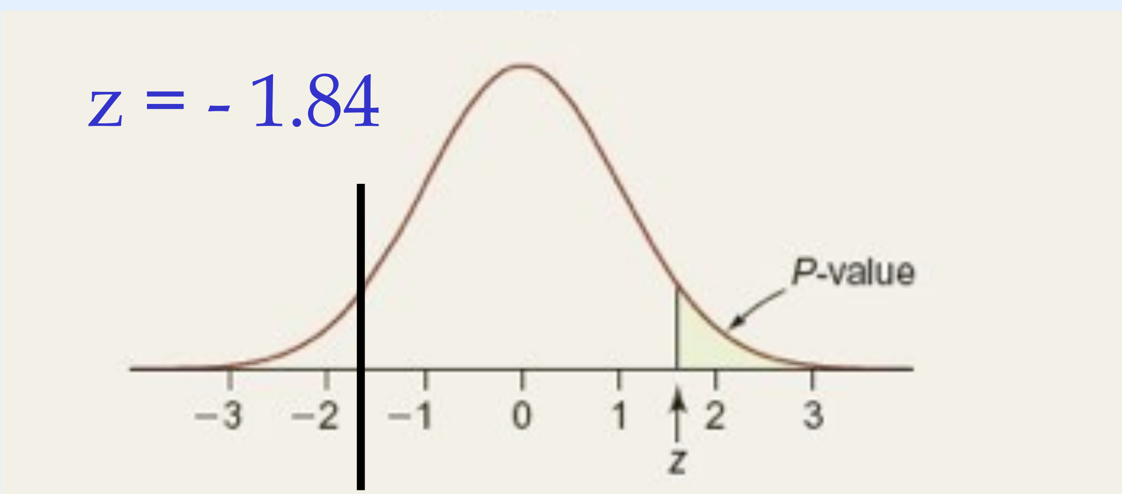
Now our test statistic $z = -1.84$ and we believe that the proportion in our school is lower than 0.6.

Still one-sided. The alternative hypothesis is $p < 0.6$.

We believe we are LESS scared

From page 759
The area BELOW z=-1.84 is 0.0329

We need to look at the area below because we are
Making the assumption that our campus has a LOWER
Value of p.



The P value is still
0.0329

We believe we are LESS scared

The P value is 0.0329 and this means that IF the true value of $p = 0.6$ and we take a sample distribution, then the probability of finding $z < -1.84$ is equal to 0.0329

About 3%

Now, we DID measure $z = -1.84$.
What does this tell us about the null hypothesis?

The P value

The *P*-Value for a Test of Significance

The ***P*-value** for a test is the probability of seeing a result that is as extreme as or more extreme than the result you got from your sample *if the null hypothesis is true.*

The *P*-value measures the strength of the evidence against the null hypothesis. The closer the *P*-value is to 0, the stronger the evidence against the null hypothesis (and in favor of the alternative hypothesis). The closer the *P*-value is to 1, the weaker the evidence against the null hypothesis.

Another example yet

We don't have a hypothesis, but we found that $z = 1.84$

The alternative hypothesis is that p is different than 0.6, neither larger nor smaller, but different. It is a two sided test.

We assume here, that we want to be outside of the interval [-z, z], that is EITHER above 1.84 or below -1.84.

So the P value is 2* 0.03329 = 0.0658

Statistical significance

The closer the P value is to 0,
the more likely it is that the null hypothesis is violated.

We say that a sample proportion is
statistically significant IF the P value is less than 0.05

This gives us a cutoff for deciding when the null
hypothesis is acceptable or not.

Our sample is significant,
That is, it is telling us something significantly different
Than the null hypothesis.

Statistical Significance

A sample proportion is said to be statistically significant if the *P*-value places it in an outer tail of the sampling distribution constructed under the assumption that the null hypothesis is true. In many statistical studies, the term "statistically significant" is used only when the *P*-value is less than 0.05. Sometimes the *level of significance* will be 0.01 or 0.001, or another value.

Halloween treats

What if we gave kids toys instead of candy?

Kids could choose candy vs. little toys.

Out of 283 children, 148 (about 52.3%) chose candy.

Is this statistically significant?

n=283, the $\hat{p}$ we measured = 0.523

Halloween treats

If they did not care either way we would have p0=0.5.
This is the null hypothesis.
Half the kids would pick candy, the other half toys.

We can use our z statistic test because
283 * 0.5 and 283*(1 - 0.5) are both greater than 10.

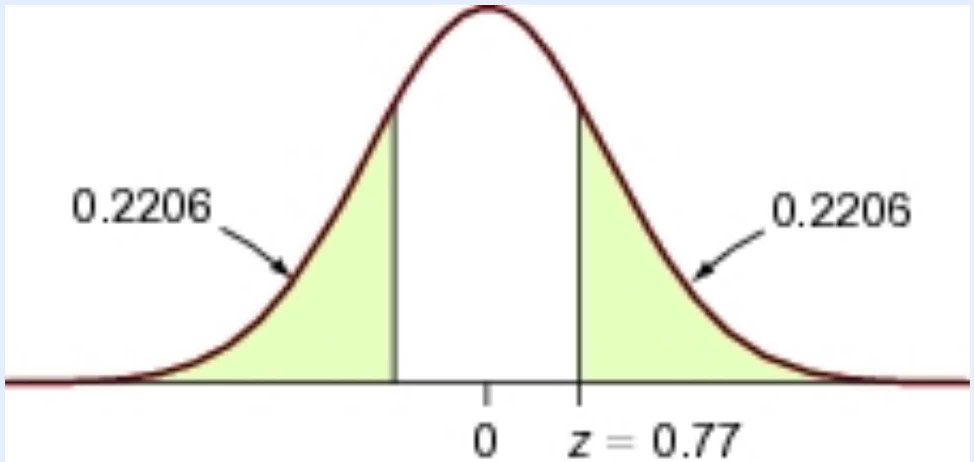
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.523 - 0.5}{\sqrt{\frac{0.5 * 0.5}{283}}} = 0.77$$

Halloween treats

This is a two sided test, we only want to know if our p is different than 0.5

That is, do kids largely prefer either candy or toys?

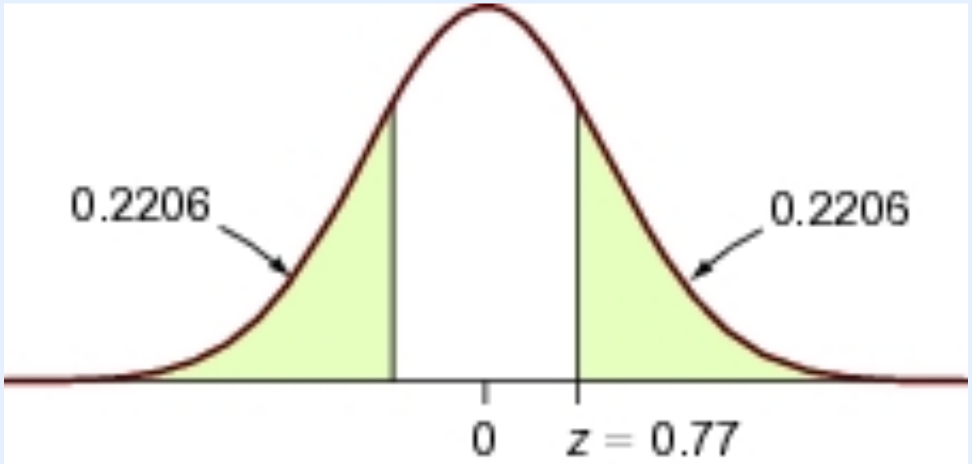
Halloween treats



The table at page 759 gives us a probability of being above 0.77 of 0.2206.

Because this is a two sided test, the P value is
2* 0.2206 = 0.4412

Halloween treats



The P value at 0.4412 is greater than 0.05, that means Our sample results are NOT statistically significant and getting 148 kids out of 283 kids to prefer candy is not so unlikely.

Homework

Page 393

P15, P16, P17, P19
E27, E28, E29, E30, E31, E32, E33, E34,E35,
E36, E37