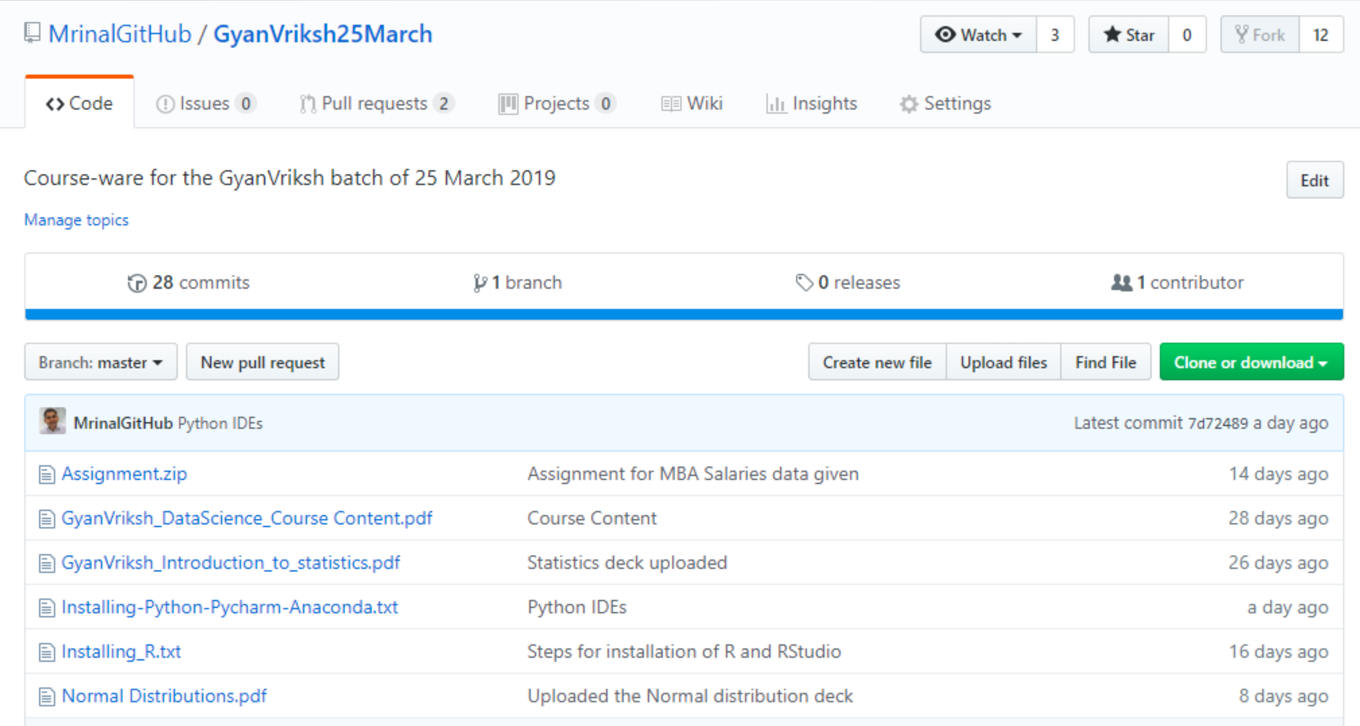


HYPOTHESIS TESTING

Managerial Decisions (revisited)

We will need 400 more person hours to finish this project.

We will acquire 8,000 new customers if I open a store in this area.

The retail market will grow by 50% in the next 5 years.

Our quality will not improve after the consulting project.

We will be able to rationalize the number of flights to 80% of the current level.

Our potential customers do not spend more than 60 minutes on the web every day.

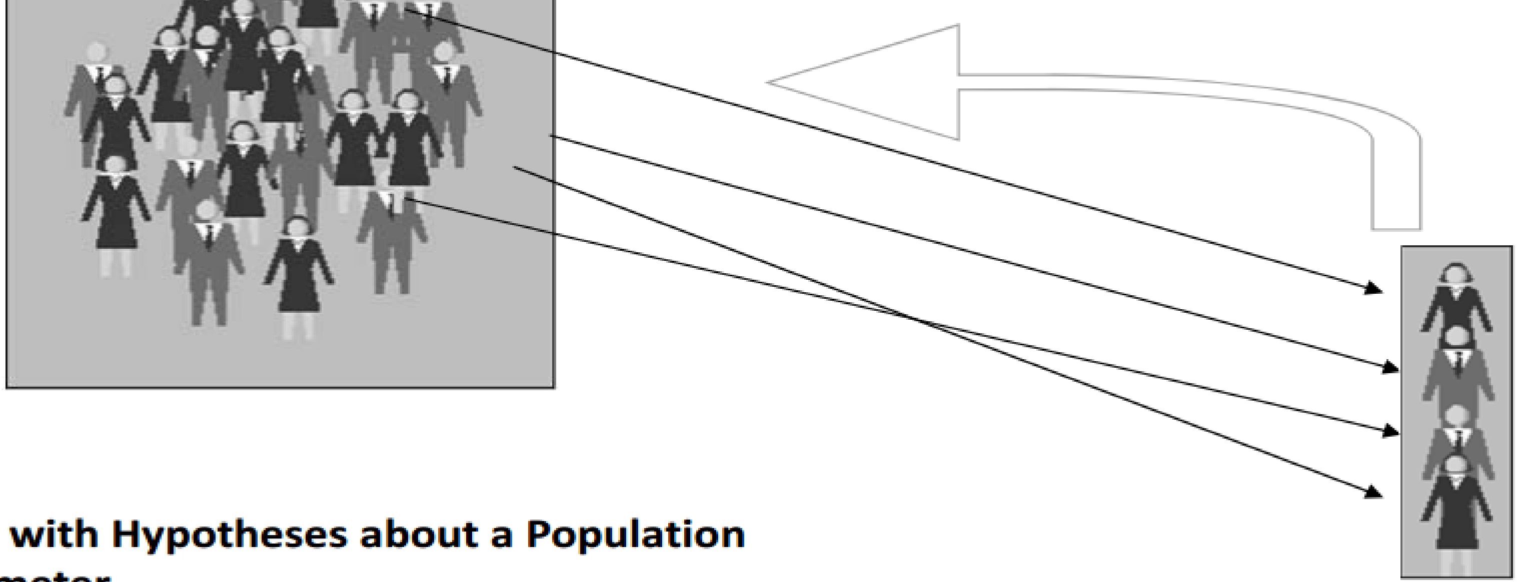
Less than 5% clients will default on their loans.

Learning Objectives

- How and when to formulate hypotheses about population parameters?
- How to quantify the strength of the evidence?
- What are Type I and Type II errors ?

Hypothesis Testing Process

3. **Reject/Do Not Reject Hypothesis**
   Is the sample information strongly inconsistent with the null hypothesis? If yes then the reject hypothesis.



1. **Start with Hypotheses about a Population Parameter**
   Parameter could be mean, proportion or something else.

2. **Collect Sample Information**
   Collect information from a randomly chosen sample and calculate the appropriate sample statistic.

The Testing Process

1. Begin by assuming that H_0 (typically status quo) is true
   - e.g. I believe that the spending will be less than or equal to \$120.
2. Quantify what is meant by "strong enough evidence" to reject H_0
   - e.g. Probability of finding a sample mean should be less than 0.05
3. Collect the evidence that would be used to test H_0
   - e.g. A pilot resulted in average spending of \$130 in a sample of 80 customers

Management

4. Calculate the probability of observing the given or stronger evidence
   - e.g. The maximum probability of getting a sample of \$130 or more under H_0 is 0.01

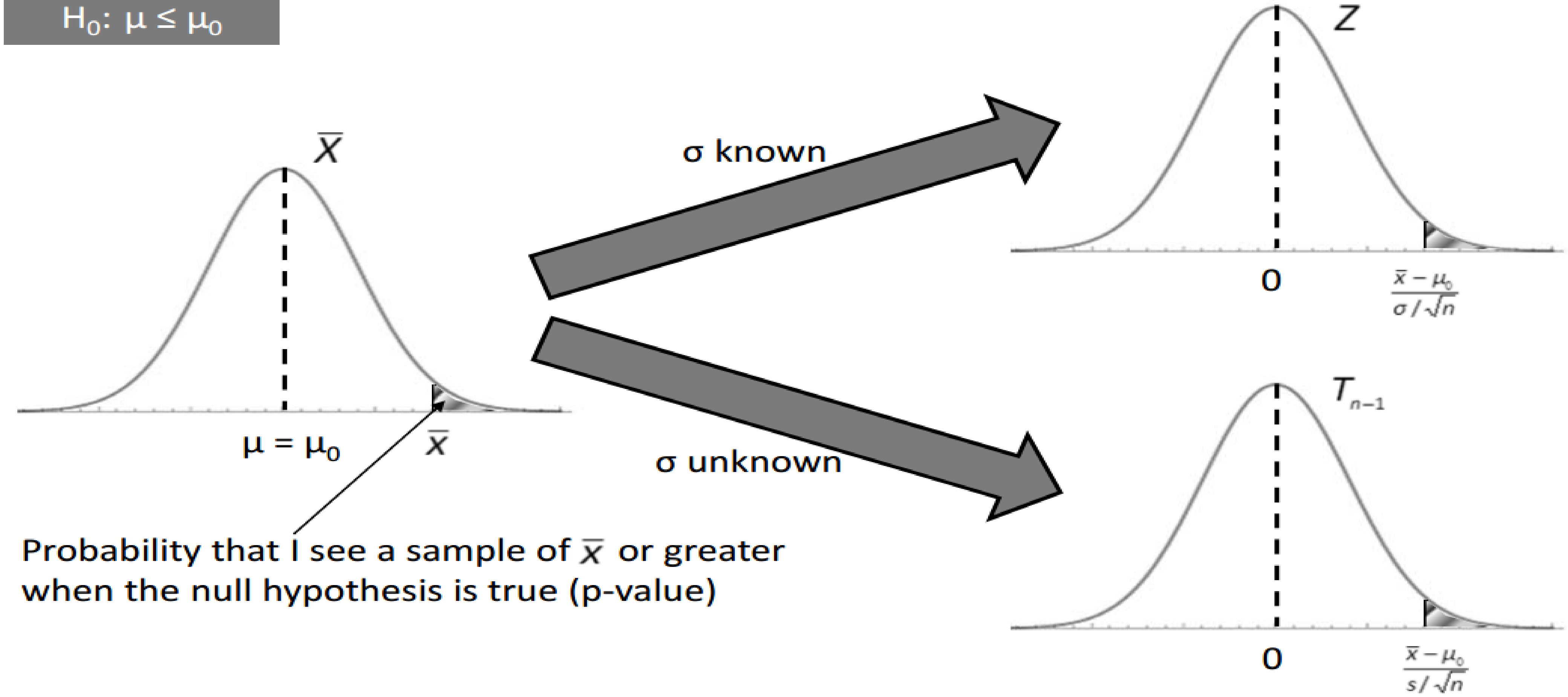
Statistics

5. Conclude and take appropriate action
   - e.g. The evidence is strong enough (0.01 < 0.05) to reject H_0; launch the card

Management

Calculating the probability of type-I error

$H_0: \mu \le \mu_0$



The definition:

The p-value is the probability of getting the observed value of the test statistic, or a value with even greater evidence against H_0, *if the null hypothesis is true.*

If we have a given significance level α, then:

$$\text{Reject } H_0 \text{ if } p\text{-value} \leq \alpha$$

(If p-value $\leq \alpha$, the evidence against H_0 is significant at the α level of significance.)

A company has stated that their straw machine makes straws that are 4 mm diameter. A worker believes the machine no longer makes straws of this size and samples 100 straws to perform a hypothesis test with 99% confidence.

Doctors believe that the average teen sleeps on average no longer than 10 hours per day. A researcher believes that teens on average sleep longer. Write Ho and Ha.

The school board claims that at least 60% of students bring a phone to school. A teacher believes this number is too high and randomly samples 25 students to test at a level of significance of 0.02. Write Ho and Ha.

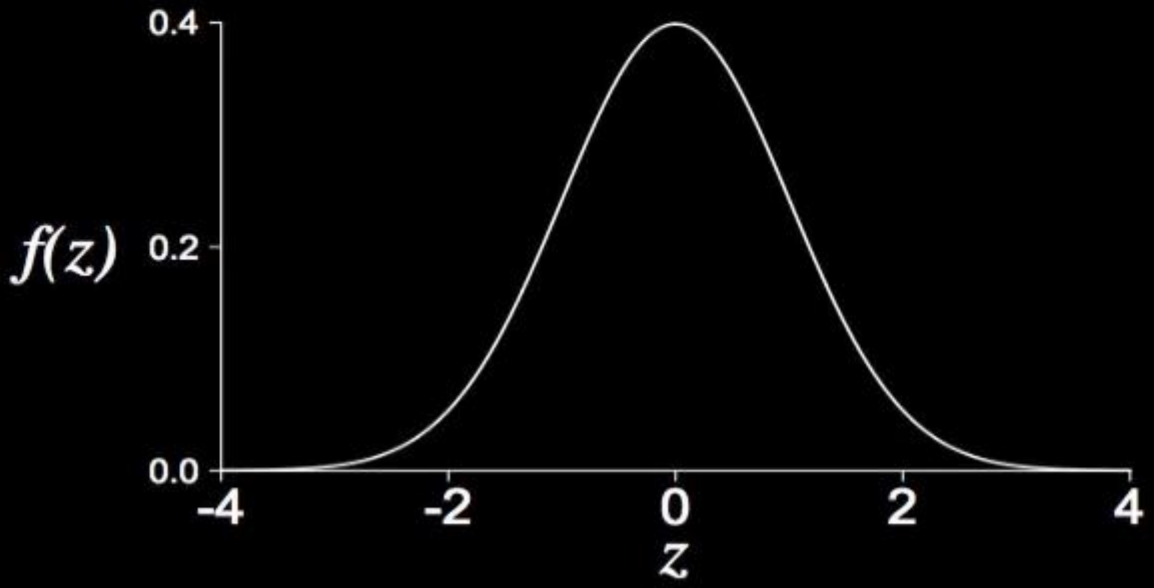
Example:

H_0: $\mu = \mu_0$

H_a: $\mu > \mu_0$

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Distribution of Z if H_0 is true:



We use p-values to make conclusions in significance testing. More specifically, we compare the p-value to a significance level α to make conclusions about our hypotheses.

If the p-value is lower than the significance level we chose, then we reject the null hypotheses H_0 in favor of the alternative hypothesis H_a. If the p-value is greater than or equal to the significance level, then we fail to reject the null hypothesis H_0—this doesn't mean we accept H_0. To summarize:

$$p\text{-value} < \alpha \Rightarrow \text{reject } H_0 \Rightarrow \text{accept } H_\mathrm{a}$$

$$p\text{-value} \geq \alpha \Rightarrow \text{fail to reject } H_0$$

Let's try a few examples where we use p-values to make conclusions.

A neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each to neurological stimulus, and recording its response time. The neurologist knows that the mean response time for rats not injected with the drug is 1.2 seconds. The mean of the 100 injected rats' response times is 1.05 seconds with a sample standard deviation of 0.5 seconds. Do you think that the drug has an effect on response time?

Alessandra designed an experiment where subjects tasted water from four different cups and attempted to identify which cup contained bottled water. Each subject was given three cups that contained regular tap water and one cup that contained bottled water (the order was randomized). She wanted to test if the subjects could do better than simply guessing when identifying the bottled water.

Her hypotheses were $H_0 : p = 0.25$ vs. $H_a : p > 0.25$ (where p is the true likelihood of these subjects identifying the bottled water).

The experiment showed that 20 of the 60 subjects correctly identified the bottle water. Alessandra calculated that the statistic $\hat{p} = \frac{20}{60} = 0.\bar{3}$ had an associated P-value of approximately 0.068.

QUESTION A (EXAMPLE 1)

What conclusion should be made using a significance level of $\alpha = 0.05$?

Choose 1 answer:

(A) Fail to reject H_0

(B) Reject H_0 and accept H_a

(C) Accept H_0

In context, what does this conclusion say?

Choose 1 answer:

- (A) The evidence suggests that these subjects can do better than guessing when identifying the bottled water.
- (B) We don't have enough evidence to say that these subjects can do better than guessing when identifying the bottled water.
- (C) The evidence suggests that these subjects were simply guessing when identifying the bottled water.

How would the conclusion have changed if Alessandra had instead used a significance level of $\alpha = 0.10$?

Choose 1 answer:

(A) She would have rejected H_a.

(B) She would have accepted H_0.

(C) She would have rejected H_0 and accepted H_a.

(D) She would have reached the same conclusion using either $\alpha = 0.05$ or $\alpha = 0.10$.

A certain bag of fertilizer advertises that it contains 7.25 kg, but the amounts these bags actually contain is normally distributed with a mean of 7.4 kg and a standard deviation of 0.15 kg.

The company installed new filling machines, and they wanted to perform a test to see if the mean amount in these bags had changed. Their hypotheses were $H_0: \mu = 7.4\text{ kg}$ vs. $H_{\text{a}}: \mu \neq 7.4\text{ kg}$ (where μ is the true mean weight of these bags filled by the new machines).

They took a random sample of 50 bags and observed a sample mean and standard deviation of $\bar{x} = 7.36\text{ kg}$ and $s_x = 0.12\text{ kg}$. They calculated that these results had a P-value of approximately 0.02.

What conclusion should be made using a significance level of $\alpha = 0.05$?

In context, what does this conclusion say?

Choose 1 answer:

- (A) The evidence suggests that these bags are being filled with a mean amount that is different than 7.4 kg.
- (B) We don't enough evidence to say that these bags are being filled with a mean amount that is different than 7.4 kg.
- (C) The evidence suggests that these bags are being filled with a mean amount of 7.4 kg.

Be cautious in your conclusion!

- First state your statistical conclusion from the hypothesis test
  - Reject or Fail to reject H_0 at a significance level of α

Reject the null hypothesis	"Fail to reject" the null hypothesis
• You have strong enough evidence to reject the null and go with the alternate	• You are saying that you do not have strong enough evidence to reject the null
• Does not mean that the alternate hypothesis is true → you could have committed a Type I error	• Does not mean that the null hypothesis is true → you could have committed a Type II error
• Take the action associated with the rejection of null hypothesis	• Continue with the status quo

Summary of Session

- Hypothesis is an assumption about a population parameter that is subject to a test and rejection based on evidence
- Hypothesis test is applicable when the manager has specific position on a population parameter which needs to be rejected in order to take action
- Given the uncertainty, we can never reject or not reject a hypothesis with certainty
  - Type I error – Null hypothesis is true but you reject it based on “unusual” sample
  - Type II errors – Null hypothesis is not true but you accept it based on “usual” sample
- A manager typically targets a certain type-I error called level of significance
- If the calculated probability of a given sample is less than the level of significance under the null hypothesis, the manager rejects his null hypothesis and makes necessary change