Hypothesis and Product Testing

Application of Statistical Concepts

INSTRUCTOR VERSION

Description

Hypotheses are central in decision making processes. A hypothesis is an assumption or concession made for the sake of argument. It is an interpretation of a practical situation or condition taken as the ground for action. It is an assumptive statement made that can be tested through some logical or empirical means. A hypothesis is an essential part of the validation process in the scientific method and relies on significance tests to answer questions.

Material and product testing relies on a set of rules and standards to make sure the best products are provided to the consumer. Consumers expect that products are what they say they are. There are generally two forms of testing done on products: manufacturers testing to make sure products meet standards; and consumer testing to make sure products are representative of what they are identified as being. The process is validated through statistical tests of significance which are robust.

The goal of this lesson is to understand the fundamentals of hypothesis testing and see how it is applied in product testing.

Objectives

Students will be able to:

- Understand the concept of hypothesis testing as it applies to the process validation, reliability, and repeatability
- Explore the process of product testing and how it works with the scientific method.
- Demonstrate the process of product testing

North Dakota State Standards

9-10.2.6	Design and conduct a guided investigation
9-10.2.7	Maintain clear and accurate records of scientific investigations
9-10.2.8	Analyze data found in tables, charts, and graphs to formulate conclusions
9-10.6.3	Explain how emerging technologies may impact society and the environment
11-12.1.2	Identify structure, organization, and dynamics of components within a system
11-12.8.1	Identify the criteria that scientific explanations must meet to be considered valid

Next Generation Science Standards

HS-ETS1-2	Design a solution to a complex real-world problem by breaking it down into smaller more manageable problems that can be solved through engineering.
HS-ETS1-3	Evaluate a solution to a complex real-world problem based on prioritized criteria and trade-offs that account for a range of constraints, including cost, safety, reliability, and aesthetics as well as possible, social, cultural, and environmental impacts.
RST.11-12.7-	Integrate and evaluate multiple sources of information presented in diverse formats and media(e.g., quantitative data, video, multimedia in order to address a question or solve a problem

Schedule

11:00-11:20	*General Organization and Cultural Connection*
11:20-11:40	Activity 1: *Basketball Performance Stats*
11:40-12:00	*PowerPoint Presentation*
12:00-12:40	Lunch
12:40-01:10	Activity 2: *Fundamentals of Hypothesis testing*
01:10-01:40	Activity 3A: *Manufacturer performance Testing and validation: Rubber band stretching*
01:40-02:00	Activity 3B: *Manufacturer quality Testing and validation: Seedless fruit standards*
02:00-02:30	Activity 4A: Consumer product testing and validation: *comparative performance testing*
02:30-02:50	Activity 4B: *Consumer product testing and validation: product claim testing*
02:50-03:00	Wrap-up activity and Reflection questions

Cultural Connection:

Terminology to Note:

Assumption	Two-tailed test
Hypothesis	P-value
Type I and Type II errors	Significance level
Testing	Error term
T-test	Coefficient of determination
Z-test	Sampling
Experimental Design	Population mean
Comparative test	Sample mean
Quality standards	Variance
Design criteria	Degrees of freedom
One-tailed test	Sample size

Reflection Questions:

1. **Why do we test products?**
2. **How has testing impacted how we report statistical facts?**
3. **Sports analytics involves the use of statistical modeling and numbers. How do you make sure the information you get is accurate and reflects the true nature of the observation?**

Hypothesis Testing

Introduction

A hypothesis is an **educated guess** about something in the world around you. It should be testable, either by experiment or observation. It can really be *anything at all* as long as you can put it to the test.
If you are going to propose a hypothesis, it's customary to write a statement. Your statement will define the relationship that exists between 2 variables (independent variable and the dependent variable).

Examples of testable statements will include:

- If I (decrease the amount of water given to herbs) then (the herbs will increase in size).
- If I (give patients counseling in addition to medication) then (their overall depression scale will decrease).
- If I (give exams at noon instead of 7) then (student test scores will improve).
- If I (look in this certain location) then (I am more likely to find new species).

Can you think of any other observations or occurrences you can put in to a hypothesis statement?

A good hypothesis statement should:

- Include an "if" and "then" statement
- Include both the independent and dependent variables.
- Be testable by experiment, survey or other scientifically sound technique.
- Be based on information in prior research (either yours or someone else's).
- Have design criteria (for engineering or programming projects).

Hypothesis Testing

Hypothesis testing in statistics is a way for you to test the results of a survey or experiment to see if you have meaningful results. It improves the validity and reliability of the results and, if done right, increases the chances that other researchers will accept your work as true or factual because it is repeatable. Therefore testing ensures: Reliability, Validity, and Repeatability! Hypothesis tests are formulated to the null hypothesis, which is the opposite of your observation. Essentially, you are trying to prove that the opposite of your observation is unlikely to have occurred. In other words, it basically involves testing by figuring out the odds that your results have happened by chance. If your results may have happened by chance, the experiment won't be repeatable and so has little use.

Hypothesis testing has easy steps that apply no matter how complex the test appears to be. All you need to do is:

1. Figure out your null hypothesis; this is always the accepted fact or the status quo.
2. State your null hypothesis, It is written in mathematical terms
3. State your alternate hypothesis; this is derived from your experiment or observation
4. Choose what kind of test you need to perform,
5. Either support or reject the null hypothesis.

Other aspects important to the test process include:

1. The Z-test (one sample) or the T-test (2 sample, unbalanced data, unknown variance) are usually used to determine significance.
2. The test can be a one-tailed or two-tailed test.
3. The degree or level of significance is often measured against an alpha level. If you are not given an alpha level, use 5% (0.05).
4. Test results can be affected by TYPE I or TYPE II errors.
5. In statistics, we always make one of two decisions. We either "reject the null hypothesis" or we "fail to reject the null hypothesis."

Activity 1: Basketball Performance Stats

Introduction

Who is a better shot, a right or left handed basketball player? Which side of the basket, left or right side, are shooters most likely to score the most 3-point shots? These are some examples of exciting questions we can answer by collecting data!
In this activity, we will be discussing the fundamentals of decision making and data analysis in sports. We will use a scenario to exemplify the importance of making decisions in sports based off of sound data and analysis. Students will be expected to use their critical thinking abilities to answer a series of questions.

Materials

Basketball (or similar ball for shooting)
Basket or a bucket (take shooters to a court if available nearby)
Data sheet and pencil
Timer

Procedure

- The class will collect data between shooters who shoot 3-point scores on the left and right side of the basket, irrespective of whether they are right or left handed, age, height, and gender.
- Split the class into 2 groups. Group A will shoot on the left side and Group B will shoot on the right side.
- Have each group identify their 5 top shooters. The shooters will have a 1 minute period to practice.
- Each shooter will have 5 tries a making their 3-point shot and less than 2 minutes to complete the 5 shots.
- Shots must be made at the standard 3-point distance. Shooters have a choice of spot along the arc to take the shot.
- Each group will record the number of shots made and lost by each player per group.
- There should be a total of 25 shots per group (50 total shots)
- Analyze the data and see if you can draw conclusions on the abilities of the shooters in each group.
- Answer the following questions.

Questions Answers will vary and the students should be able formulate an opinion

1. What was the outcome of the trial?
2. Will you consider the trials valid as a sample to make important decisions? Why or why not?
3. How can you improve the significance of the conclusions you just made from your data?
4. How can you validate the significance of the conclusions you just made from your data?
5. Suggest an effective method to improve on this trial design. What will you change, and what will you not change?

Activity 2: *Fundamentals of Hypothesis testing*

Introduction

In this activity, we will be discussing real events tied to statistics and hypothesis concepts. Students will be expected to use their knowledge of statistics and hypothesis testing to answer a series of questions.

How to write a good hypothesis using "if… then… because" statements

An "If… then…because" statement in a hypothesis tells the readers what you believe will happen in an investigation when something is changed, so you can see the effect of the change.

- **IF**…tells the readers what will be changed. This is the manipulated (independent) variable in the investigation.
- **THEN**… tells the reader what will happen because of the change (manipulated variable) described in the If… statement. This is the responding (dependent) variable in the investigation.
- **BECAUSE**… tells the reader how you know this will occur. It should be based on something you have experienced, or perhaps something you infer.

Examples:

- **If** 7th graders and 8th graders complete the same math problems, **then** the 8th graders will have more answers correct, **because** they have studied math for one year longer than the 7th graders.
- **If** dry bread and moist bread are left in bags for two weeks, **then** the moist bread will grow mold more quickly than the dry bread, **because** mold is a living organism, and organisms need water to survive.
- **If** some students eat breakfast before school and others do not, **then** the ones who do eat breakfast will have better grades in their morning classes, **because** their brains have more energy to think.

Now, let's try this together. To warm up, identify the three types of variables below. Then use the variables to make a good hypothesis.

Melissa raises crickets at her pet store that she sells for reptile food. She thinks that crickets chirp more often when the temperature gets warmer. She decides to conduct an experiment to prove her theory.

a. Manipulated variable Temperature.
b. Responding variable Crickets.
c. Controlled variable Crickets chirping.

Hypothesis: **If** temperatures get warmer in the pet store (manipulated variable) **then** crickets will chirp more often (responding variable), **because** changes in temperatures affect crickets and their ability to chirp.

Can you figure out the following?

Jenna likes to work with her friend Joe on 7th grade science class labs. However, she notices that she tends to get lower grades when she does work with Joe, because she and Joe like to talk about basketball and not science. She's decided to investigate if her science lab grades are higher when she works with Joe or if they're higher when she works with someone else. She has set goals to become a better student and scientist, so could you please help her write a hypothesis for her study?

a. Manipulated variable Lab partners.
b. Responding variable Grades.
c. Controlled variable Student science grades.

Hypothesis: **If** Jenna works with another lab partner (manipulated variable) **then** her lab science grades will be higher (responding variable), **because** she wants to talk about science to improve her grades.

The cooler the temperature in a lake, the more oxygen the water holds. Daniel notices that he catches more fish in a lake that is cooler than 55 degrees. He wants to conduct a study so he can catch the most fish possible this year. He's having trouble writing a hypothesis. Please help him.

Hypothesis: **If** lake temperatures are cooler than 55 degrees (manipulated variable) **then** Daniel will catch more fish in the lake (responding variable), **because** the cooler the temperature in a lake, the more oxygen the water holds.

Kasey lives in Moab, Utah. She likes to mountain bike for miles and miles until she can't bike anymore. She thinks that she can bike further when she drinks more than a liter of water before her bike ride. Please assist her in developing a hypothesis, so she can make the best of her future bike rides.

Hypothesis: **If** Kasey drinks more than a liter of water before her bike ride (manipulated variable) **then** she will cover more miles on her mountain bike rides (responding variable), **because** taking water before makes her ride longer.

Mr. Montanari has noticed that there is a wide range of grades that students get on tests, even though they are all in the same class. He wonders whether students, who study for 20 minutes per night, every night, get higher scores on tests or not. Mr. Montanari wants as many kids to be successful as possible. Please help him write a hypothesis about this problem.

Hypothesis: **If** all students study for 20 minutes per night every night (manipulated variable) **then** students in the same class will get similar grades (responding variable), **because** students who study for tests are successful.

Hypothesis Scenarios

Read and answer the following questions based on scenarios presented

1. An engineer measured the Brinell hardness of 25 pieces of ductile iron that were subcritically annealed. The engineer hypothesized that the mean Brinell hardness of *all* such ductile iron pieces is greater than 170.
    a. State the null and alternate hypothesis using statistical notation?

       $H_0: \mu = 170$
       $H_A: \mu > 170$
    b. Is this a left tailed or a right tailed test?
       Right one tailed test
2. The engineer entered his data into Minitab and requested that the "one-sample *t*-test" be conducted for the above hypotheses. The test statistic ***t** is 1.22**, and the ***P*-value is 0.117**. If the engineer set his **significance level α at 0.05** and used the critical value approach to conduct his hypothesis test, he would reject the null hypothesis if his test statistic ***t** were greater than 1.7109**:
    a. Does the engineer reject the null hypothesis?
       Since the engineer's test statistic, $t^* = 1.22$, is not greater than 1.7109, the engineer fails to reject the null hypothesis. There is insufficient evidence, at the $\alpha = 0.05$ level, to conclude that the mean Brinell hardness of all such ductile iron pieces is greater than 170.
3. A researcher thinks that if knee surgery patients go to physical therapy twice a week (instead of 3 times), their recovery period will be longer. Average recovery times for knee surgery patients is 8.2 weeks.
    a. State the null and alternate hypothesis using statistical notation?
       $H_0 \; \mu \leq 8.2$
       $H_A: \mu > 8.2$
4. A biologist was interested in determining whether sunflower seedlings treated with an extract from *Vinca minor* roots resulted in a lower average height of sunflower seedlings than the standard height of 15.7 cm. The biologist treated a random sample of 33 seedlings with the extract and subsequently obtained the heights of the plant.
    a. State the null and alternate hypothesis using statistical notation?
       $H_0: \mu = 15.7$
       $H_A: \mu < 15.7$
    b. Is this a left tailed or a right tailed test?
       Left one tailed test
5. The biologist entered her data into Minitab and requested that the "one-sample *t*-test" be conducted for the above hypotheses. The test statistic ***t** is -4.60**, and the ***P*-value, 0.0001**. If the biologist set her significance **level α at 0.05** and used the critical value approach to conduct her hypothesis test, she would reject the null hypothesis if her test statistic ***t** were less than -1.6939**:
    a. Does the biologist reject the null hypothesis?
       Since the biologist's test statistic, $t^* = -4.60$, is less than -1.6939, the biologist rejects the null hypothesis. There is sufficient evidence, at the $\alpha = 0.05$ level, to conclude that the mean height of all such sunflower seedlings is less than 15.7 cm.
6. A manufacturer claims that the thickness of the spearmint gum it produces is 7.5 one-hundredths of an inch. A quality control specialist regularly checks this claim. On one production run, he took a random sample of 10 pieces of gum and measured their thickness.
    a. State the null and alternate hypothesis using statistical notation?
       $H_0: \mu = 7.5$
       $H_A: \mu \neq 7.5$
    b. Is this a one-tailed or a two-tailed test?
       Two tailed test because the specialist simply tests for difference in the claim
7. The quality control specialist entered his data into Minitab and requested that the "sample *t*-test" be conducted for the above hypotheses. The test statistic ***t** is 1.54,** and the ***P*-value is 0.158**. If the quality control specialist sets his

significance level α at 0.05 and used the critical value approach to conduct his hypothesis test, he would reject the null hypothesis if his test statistic ***t**** **were less than -2.2622 or greater than 2.2622.**

a. Does the quality control specialist reject the null hypothesis?
   Since the quality control specialist's test statistic, $t^* = 1.54$, is not less than -2.2622 nor greater than 2.2622, the quality control specialist fails to reject the null hypothesis. There is insufficient evidence, at the $\alpha = 0.05$ level, to conclude that the mean thickness of all of the manufacturer's spearmint gum differs from 7.5 one-hundredths of an inch.

Activity 3: *Manufacturer performance Testing and validation*

Introduction:
In this activity, we will be putting on the hat of a product manufacturer. Our aim is to make sure that the products meet the best specification possible in a variety of situations. Students will test products to see if they meet product specifications? Manufacturers want to minimize product failure before a specified significant limit. Before products are released to the market, they are tested multiples times until any defects that can adversely affect performance are corrected. This process falls under the Research and Development Phase of a product (R&D) and manufacturers spend significant amounts of money during this phase of making the product.
It is worth noting that no product is 100% failure proof. Manufacturers are willing to accept a minimum risk on their products failing. This risk is usually defined by the statistical test and its alpha level or confidence level (Usually 5% and 1%). The statistical risk varies from one industry to the next and from one product to the next. The goal is usually to minimize errors that can increase the risk of product failure.
For this activity we are going to test 2 products from different industries: Rubber bands and seedless fruits.

State the Test Hypothesis to use for each test:
H_0: $\mu_0 = \mu_1$
H_A: $\mu_0 \neq \mu_1$

Hypothesis Test Procedure

Step 1: State the null hypothesis:
Step 2: State the alternate hypothesis:
Step 3: State your alpha level. We'll use 0.05 for all our tests. As this is a two-tailed test, split the alpha into two (0.05/2=0.025)
Step 4: Find the z-score associated with your alpha level. You're looking for the area in *one tail only*. A z-score for 0.75(1-0.025=0.975) is 1.96. As this is a two-tailed test, you would also be considering the left tail (z=1.96)

$$Z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

Step 5: Find the test statistic using this formula:
Step 6: If Step 5 is less than -1.96 or greater than 1.96 (Step 3), reject the null hypothesis.

Activity 3A: ***Manufacturer performance Testing and validation: Rubber band stretching***

Materials (per group)
1 package of rubber bands of similar type (at least 90 bands per bag)
Meter stick or ruler
Beaker – 500 mL
Hotplate or hot water
Softwood for designing the testing mechanism
Spreadsheet
Tongs

Procedure
1. Each group will obtain the material listed above.
2. The group will design a testing mechanism to use for measuring the elasticity of rubber bands.
3. The bag of rubber bands will be split into 3 groups of 20 bands and each band will be tested:
    i. 20 bands will be tested at room temperature;
    ii. 20 bands will be tested after the bands have been exposed to heat in boiling water
    iii. 20 bands will be tested after the bands have been exposed to cold and frozen
4. The group will find a way to stretch a rubber band until it breaks and be able to record results.
5. Stretch a rubber band at room temperature to breaking. Repeat for other 19 bands. Record how long it was stretched on a spreadsheet.
6. Heat a rubber band in a beaker of boiling water for 30 seconds, then stretch to breaking. Repeat for other 19 bands. Record how long it was stretched on a spreadsheet.
7. Cool a rubber band using ice for 30 seconds and stretch to breaking. Repeat for other 19 bands. Record how long it was stretched on a spreadsheet.
8. Each group will have 3 sets a data with 20 (**n**) data points: for room, hot, and cold temperatures.
9. Get the mean or average for the 3 sets of data ($\bar{x}$),
10. Assume a standard deviation of 5 (σ), and an ultimate elongation average of 700% (μ_0).
    Example: a band of length 1.25 inches, will stretch on average to 8.57 inches or 7 times its length.
11. Test the 3 hypotheses of difference at the alpha level of 0.05 (95% CL)

The results for this test will vary based on data collected but should all trend in the same direction…

Activity 3B: ***Manufacturer quality Testing and validation: Seedless fruit standards***

Materials:
Plastic knife and Paper plates
1 bag per group – seedless fruits (grapes or mandarins will work)
Spreadsheet

Procedure
1. Obtain a bag of seedless fruits, a paper plate and a plastic knife.
2. Place a fruit on the paper plate and cut open and see if there are seeds in the grape.
3. Each group will repeat until there are no more grapes in the bag. Total number of grapes will be (**n**)
4. Record the total number of fruits and the number of fruits which contained seeds.
5. Get the average number of seeds collected for the data ($\bar{x}$),
6. Assume a standard deviation of 4(σ),
7. Assume the producer has a maximum tolerance average of 6 seeds per bag (μ_0).
8. Test the hypotheses of difference at the alpha level of 0.05 (95% CL)

The results for this test will vary based on data collected but should all trend in the same direction…

Questions:

1. Which product, rubber band or seedless fruit, met the specifications of the manufacturer?
   Results will vary and interpretation will be made after test results
2. State the null and the alternate hypothesis using statistical notation for the rubber band test?
   H_{01}: $\mu_0 = \mu_{normal}$
   H_{A1}: $\mu_0 \neq \mu_{normal}$
   H_{02}: $\mu_0 = \mu_{cold}$
   H_{A2}: $\mu_0 \neq \mu_{cold}$
   H_{03}: $\mu_0 = \mu_{hot}$
   H_{A3}: $\mu_0 \neq \mu_{hot}$
3. State the null and the alternate hypothesis using statistical notation for the seedless fruit test?
   H_0: $\mu = 6$
   H_A: $\mu \neq 6$
4. How will you improve on this test?
   Collect more data or improve sample size
5. If you are the quality control compliance office, what will be your recommendations based on the tests for:
   a. Rubber band: Will vary based on results and right interpretation
   b. Seedless fruits: Will vary based on results and right interpretation

Activity 4: *Consumer product testing and validation*

Introduction

In this activity, we will be putting on the hat of consumer advocate group. Our aim is to make sure that the products reflect the statements made about them by the manufacturer. Students will test products to see if they meet product statements. Students will test whether or not the label on a product is correct and compare 2 brands of a product. Students will also measure how significant this statement is by doing a significance test.

State the Test Hypothesis to use for each test:

2 sample comparison tests (Paper towel tests)

H_0: $\mu_1 = \mu_2$

H_A: $\mu_1 \neq \mu_2$

1 sample comparison tests (Candy test)

H_0: $\mu_0 = \mu_A$

H_A: $\mu_0 \neq \mu_A$

Activity 4A: *Consumer product testing and validation: comparative performance testing*

Materials

Paper towels – 2 brands
Graduated cylinders – 100 ml
Weights – grams, or pennies
Stopwatch
Spreadsheet

Procedure A:

1. Each group will obtain a roll of each brand of paper towel.
2. They will tear off 1 sheet and see how much weight the towel can hold until tearing.
3. The group will record the mass needed.
4. This will be done 20 times with each brand of paper towel.

Procedure B:

1. With same brands of paper towels, groups will test the absorption of each.
2. Obtain a 100 mL Graduated Cylinder and 1 sheet of one brand of paper towel. Fill cylinder with water to 100 mL.
3. Twist paper towel into roughly a cylinder shape which will fit into Graduated Cylinder.
4. Place into cylinder for 3 seconds.
5. Remove and record amount of water left in the Graduated Cylinder. Repeat this procedure 20 times for each brand.

Hypothesis Test Procedure:

1. Copy your data into the online calculator in the following link:
   http://www.physics.csbsju.edu/stats/t-test_bulk_form.html
1. Data for brand 1 should by copied and pasted in the group A box. Data should be separated by commas. Put data for brand 2 in the group B box and separate by commas as well.
2. Once the datasets have been put in, click on **Calculate Now** to run the test and get your results!
3. Perform the same test for data on paper towel strength and paper towel absorption rate.

The results for this test will vary based on data collected but should all trend in the same direction…

Activity 4B: *Consumer product testing and validation: product claim testing*

Materials

Bite sized Snickers or small bags of M & M's– 1 large bag per group
Scale
Spreadsheet

Procedure:

1. Each group will obtain 1 bag of bite size Snickers.
2. Record the mass as stated on the Snickers wrapper.
3. Remove a Snickers from wrapper and find and record the mass using a scale.
4. Repeat for the entire bag of Snickers. Total number sampled will be (**n**)
5. Get the average number of masses for the data,
6. Assume a standard deviation of 3 (**σ**),
7. Get the average mass listed on the wrapping of the small bags for population mean (**μ_0**).
8. Test the hypotheses of difference at the alpha level of 0.05 (95% CL)
9. Feel free to consume the Snickers upon completion of procedure.

Hypothesis Test Procedure

Step 1: State the null hypothesis:
Step 2: State the alternate hypothesis:
Step 3: State your alpha level. We'll use 0.05 for all our tests. As this is a two-tailed test, split the alpha into two (0.05/2=0.025)
Step 4: Find the z-score associated with your alpha level. You're looking for the area in *one tail only*. A z-score for 0.75(1-0.025=0.975) is 1.96. As this is a two-tailed test, you would also be considering the left tail (z=1.96)

$$Z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}}$$

Step 5: Find the test statistic using this formula:

Step 6: If Step 5 is less than -1.96 or greater than 1.96 (Step 3), reject the null hypothesis.

The results for this test will vary based on data collected but should all trend in the same direction…

Questions:

1. Which brand of paper towel was strongest? What factors do you feel accounted for brand being stronger?
   Test results will vary.
   You can list factors like material used, whether tissue is a single or double ply etc.

2. Which brand of paper towel absorbed the most water? What factors do you feel accounted for this observation?
   Test results will vary.
   You can list factors like material used, whether tissue is a single or double ply etc.

3. Was the mass on the Snickers wrapper the same as the mass obtained from the scale? What factors do you feel accounted for any differences observed?
   Test results will vary.