UNIVERSITY OF KWAZULU-NATAL

INYUVESI YAKWAZULU-NATALI

PROBABILITY

(25 JUL 2025)

S.O Oduse

Discipline of Public Health Medicine
School of Nursing and Public Health
University of KwaZulu-Natal

EDGEWOOD CAMPUS | HOWARD COLLEGE CAMPUS | NELSON R MANDELA SCHOOL OF MEDICINE | PIETERMARITZBURG CAMPUS | WESTVILLE CAMPUS

UKZN INSPIRING GREATNESS

UNIVERSITY OF KWAZULU-NATAL

INYUVESI YAKWAZULU-NATALI

Probability

EDGEWOOD CAMPUS | HOWARD COLLEGE CAMPUS | NELSON R MANDELA SCHOOL OF MEDICINE | PIETERMARITZBURG CAMPUS | WESTVILLE CAMPUS

INTRODUCTION

- Imagine you’re sitting in a crowded waiting room during flu season. Someone nearby sneezes loudly. Will you catch the flu? Your chance of getting sick depends on factors like your vaccination status, how close you are sitting, and how long you’re exposed — all these clues help you estimate your risk.

INTRODUCTION

- **Probability quantifies uncertainty and helps us make decisions when outcomes are not certain. In public health, understanding probability is essential to model risk, interpret data, and guide evidence-based action.**

Basic Concepts in Probability

- **Experiment**: A process or action that yields a set of results (e.g., blood pressure measurement)
- **Outcome**: A possible result of an experiment (e.g., high blood pressure).
- **Sample Space (S):** The set of all possible outcomes (e.g., {normal, elevated, high} BP). **Event (A):** Any subset of the sample space (e.g., high blood pressure).

Basic Concepts in Probability

- **Random Variable:** A variable that takes on values determined by chance.
- **Probability of an event (P(E)):** Number of favorable outcomes divided by total number of possible outcomes. - Example: If 3 out of 10 people have malaria, P(malaria) = 3/10 = 0.3

Types of Probability

- Theoretical (Classical) Probability
- Empirical (Experimental) Probability
- Subjective Probability

Theoretical (Classical) Probability

- Based on assumptions and known models.
- $P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total number of outcomes}}$
- **Example:** Probability of rolling a 3 on a fair six-sided die:
- $P(3) = \frac{1}{6} = 0.167$

Empirical (Experimental) Probability

- Based on actual data and experimentation.
- $P(A) = \frac{\text{Number of times event } A \text{ occurred}}{\text{Total number of trials}}$
- **Example:** If 5 out of 100 blood samples test positive for malaria:
- $P(\text{Positive}) = \frac{5}{100} = 0.05$

Subjective Probability

- Based on personal judgment or expert opinion.
- No formal calculation.
- Example: An epidemiologist estimates a 70% chance of a measles outbreak based on past experience and current trends.

Types of Events in Probability

- Simple Event
- Compound Event
- Mutually Exclusive Events
- Independent Events
- Dependent Events
- Complementary Events
- Exhaustive Events

Types of Events in Probability

- Simple Event: An event with a single outcome (e.g., P(HIV positive) = 0.05).
- Compound Event: Involves more than one outcome (e.g., P(HIV positive and male)).
- Mutually Exclusive Events: Events that cannot happen at the same time (e.g., male vs. female).
- Independent Events: One event does not affect the probability of the other (e.g., result of test A does not affect test B).

Types of Events in Probability

- Dependent Events: The outcome of one event affects the probability of another (e.g., exposure to a virus increases disease risk).
- Complementary Events: Two events whose probabilities add to 1 (e.g., infected vs. not infected).
- Exhaustive Events: Together, the events cover all possible outcomes. Example: Getting heads or tails on a coin.

Properties of Probability

- The value of any probability lies between 0 and 1 i.e. $0 \leq P(A) \leq 1$
- The sum of probabilities of all outcomes in a sample space is 1
- The probability of a certain event (the entire sample space) is 1 i.e. $P(S)=1$ (certainty)
- The probability of an impossible event is 0 i.e. $P(\emptyset)=0$ (impossible event)

Rules of Probability

- **Addition Rule**
- For mutually exclusive A and B: $P(A \cup B) = P(A) + P(B)$
- Example: P(male) = 0.45, P(female) = 0.55
- P(male or female) = 1.0
- For non-mutually exclusive A and B: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Example: P(diabetes) = 0.25, P(hypertension) = 0.40, P(both) = 0.10
- P(either) = 0.25 + 0.40 - 0.10 = 0.55

Rules of Probability

- **Multiplication Rule**
- For independent A and B:$P(A \cap B) = P(A) \times P(B)$
- Example: P(flu) = 0.1 P(two people get flu) =
- =0.1X 0.1 = 0.01
- For dependent A and B:$P(A \cap B) = P(A) \times P(B|A)$
- Example: P(testing positive) depends on whether the person is infected.
- **Complement Rule**
- $P(A^c) = 1 - P(A)$
- **Example:** If probability of infection is 0.2:
- $P(\text{No infection}) = 1 - P(\text{Infection}) = 1 - 0.2 = 0.8$

Examples-Basic Probability

- **Exercise 1.1:**
  200 screened for TB, 30 positive. $P(\text{Positive})$?
  Solution:
- $P(\text{Positive}) = \frac{\text{Number Positive}}{\text{Total}} = \frac{30}{200} = 0.15$
- **Exercise 1.2:**
  300 babies, 15 low birth weight. $P(\text{LBW})$?
- $P(\text{LBW}) = \frac{\text{Number with LBW}}{\text{Total}} = \frac{15}{300} = 0.05$

Examples-Basic Probability

- **Exercise 1.3:**
  500 vaccinated, 40 develop fever. $P(\text{Fever})$?
- $P(\text{Fever}) = \frac{40}{500} = 0.08$
- **Exercise 1.4:**
  1,000 people, 5 have typhoid. $P(\text{Typhoid})$?
- $P(\text{Typhoid}) = \frac{5}{1000} = 0.005$

Examples-Mutually Exclusive Events

- **Exercise 2.1:**
  20% diabetes, 15% hypertension, 5% both.
  $P(\text{Diabetes or HTN})$?
- $P(D \cup H) = P(D) + P(H) - P(D \cap H) = 0.20 + 0.15 - 0.05 = 0.30$
- **Exercise 2.2:**
  15% smoke, 10% drink, 2% both. Probability either?
- $P(S \cup A) = P(S) + P(A) - P(S \cap A) = 0.15 + 0.10 - 0.02 = 0.23$

Examples-Mutually Exclusive Events

- **Exercise 2.3:**
  18% HIV, 12% HepB, 3% both. $P($HIV or HepB$)$?
- $P(H \cup B) = P(H) + P(B) - P(H \cap B) = 0.18 + 0.12 - 0.03 = 0.27$
- **Exercise 2.4:**
  40% overweight, 25% hypertensive, 8% both. $P($OW or HTN$)$?
- $P(OW \cup H) = P(OW) + P(H) - P(OW \cap H) = 0.40 + 0.25 - 0.08 = 0.57$

Examples-Independent Events

- **Exercise 3.1:**
  $P(\text{male}) = 0.51$. Both among 2 births?
- $P(\text{Both male}) = P(\text{male}) \times P(\text{male}) = 0.51 \times 0.51 = 0.2601$
- **Exercise 3.2:**
  Probability of negative tests: 0.8 (disease A), 0.9 (B). Both negative?
- $P(\text{Both negative}) = 0.8 \times 0.9 = 0.72$

Examples-Independent Events

- **Exercise 3.3:**
  40% have immunity. Probability both in a sample of two?
- $P(\text{Both immune}) = 0.4 \times 0.4 = 0.16$
- **Exercise 3.4:**
  $P(\text{rain today}) = 0.3$, $P(\text{rain tomorrow}) = 0.2$. Both days?
- $P(\text{Rain both days}) = 0.3 \times 0.2 = 0.06$

Examples-Conditional Probability (Dependent Events)

- **Exercise 4.1:**
  10% malaria prevalence; test sensitivity 95%.
  $P(\text{Test}^{+}|\text{Malaria})$?
- $P(\text{Test}^{+}|\text{Malaria}) = \text{Sensitivity} = 0.95$
- **Exercise 4.2:**
  2% HIV positive, test detects 99% of time.
  $P(\text{Test}^{+}|\text{HIV})$?
- $P(\text{Test}^{+}|\text{HIV}) = 0.99$

Examples-Conditional Probability (Dependent Events)

- **Exercise 4.3:**
  5% TB, test sensitivity 90%.
- $P(\text{Test}^{+}|\text{TB}) = 0.90$
- **Exercise 4.4:**
  15% anemic women, 80% report fatigue.
- $P(\text{Fatigue}|\text{Anemia}) = 0.80$

Examples-Complementary Probability

- **Exercise 5.1:**
  60% vaccinated. Not vaccinated?
- $P(\text{Not vaccinated}) = 1 - P(\text{Vaccinated}) = 1 - 0.6 = 0.4$
- **Exercise 5.2:**
  85% virus-free. Contaminated?
- $P(\text{Contaminated}) = 1 - 0.85 = 0.15$

Examples-Complementary Probability

- **Exercise 5.3:**
  10% genetic marker. Not have marker?
- $P(\text{No marker}) = 1 - 0.10 = 0.90$
- **Exercise 5.4:**
  $P(\text{black}) = 0.35$. Not black?
- $P(\text{Not black}) = 1 - 0.35 = 0.65$

UNIVERSITY OF KWAZULU-NATAL
INYUVESI YAKWAZULU-NATALI

Confidence Interval

Discipline of Public Health Medicine
School of Nursing and Public Health
University of KwaZulu-Natal

EDGEWOOD CAMPUS | HOWARD COLLEGE CAMPUS | NELSON R MANDELA SCHOOL OF MEDICINE | PIETERMARITZBURG CAMPUS | WESTVILLE CAMPUS

UKZN INSPIRING GREATNESS

Confidence interval

A ***confidence interval*** is an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data. (*Definition taken from Valerie J. Easton and John H. McColl's* *Statistics Glossary v1.1*

Estimation

- In statistics, **estimation** refers to the process by which one makes inferences about a population, based on information obtained from a sample.
- In other words estimation is the process by which sample data are used to indicate the value of an unknown quantity in a population

Population and Sample

- A population consists of all elements- individuals, items or objects whose characteristics are being studied
- A portion of the population selected for study is called a sample

Point vs. Interval estimates

- Statisticians use sample statistics to estimate population parameters. For example, sample means are used to estimate population means; sample proportions, to estimate population proportions.
- An estimate of a population parameter may be expressed in two ways:
- **Point estimate .** A point estimate of a population parameter is a single value of a statistic. For example, the sample mean x is a point estimate of the population mean μ. Similarly, the sample proportion *p* is a point estimate of the population proportion *P*.
- **Interval estimate .** An interval estimate is defined by two numbers, between which a population parameter is said to lie. For example, $a < x < b$ is an interval estimate of the population mean μ. It indicates that the population mean is greater than *a* but less than *b*.

Confidence Intervals

Statisticians use a **confidence interval** to express the **precision** and **uncertainty** associated with a particular **sampling method**. A confidence interval consists of three parts.

- A confidence level.
- A statistic.
- A margin of error.

- The confidence level describes the uncertainty of a sampling method. The statistic and the margin of error define an interval estimate that describes the precision of the method. The interval estimate of a confidence interval is defined by the *sample statistic* $\pm$ *margin of error*.

Confidence interval continued

- For example, suppose we compute an interval estimate of a population parameter. We might describe this interval estimate as a 95% confidence interval. **This means that if we used the same sampling method to select different samples and compute different interval estimates, the true population parameter would fall within a range defined by the *sample statistic* ± *margin of error* 95% of the time.**

Confidence interval continued

- Confidence intervals are preferred to point estimates, because confidence intervals indicate (a) the precision of the estimate and (b) the uncertainty of the estimate.

Confidence level

- The probability part of a confidence interval is called a **confidence level**. The confidence level describes the likelihood that a particular sampling method will produce a confidence interval that includes the true population parameter.
- Here is how to interpret a confidence level. Suppose we collected all possible samples from a given population, and computed confidence intervals for each sample. Some confidence intervals would include the true population parameter; others would not. A 95% confidence level means that 95% of the intervals contain the true population parameter; a 90% confidence level means that 90% of the intervals contain the population parameter; and so on.

Margin of Error

- In a confidence interval, the range of values above and below the sample statistic is called the **margin of error**
- For example, suppose the local newspaper conducts an election survey and reports that the independent candidate will receive 30% of the vote. The newspaper states that the survey had a 5% margin of error and a confidence level of 95%. These findings result in the following confidence interval: There is a 95% likelihood that the independent candidate will receive between 25% and 35% of the vote.

What is the Standard Error?

- The standard error is an estimate of the standard deviation of a statistic.
- The standard error is important because it is used to compute other measures, like confidence intervals and margins of error.

Standard Deviation of Sample Estimates

- Statisticians use sample statistics to estimate population parameters. Naturally, the value of a statistic may vary from one sample to the next.
- The variability of a statistic is measured by its standard deviation.

Standard Error of Sample Estimates

- Sadly, the values of population parameters are often unknown, making it impossible to compute the standard deviation of a parameter. When this occurs, use the standard error.
- The standard error is computed from known sample statistics, and it provides an unbiased estimate of the standard deviation of the parameter.

How to Compute the Margin of Error

- The margin of error can be defined by either of the following equations.
- Margin of error = Critical value x Standard deviation of the parameter

  Margin of error = Critical value x Standard error of the statistic
- If you know the standard deviation of the parameter, use the first equation to compute the margin of error. Otherwise, use the second equation.

How to Interpret Confidence Intervals

- Suppose that a 90% confidence interval states that the population mean is greater than 100 and less than 200. How would you interpret this statement?
- Some people think this means there is a 90% chance that the population mean falls between 100 and 200. This is incorrect. Like any population parameter, the population mean is a constant, not a random variable. It does not change. The probability that a constant falls within any given range is always 0.00 or 1.00.

How to Interpret Confidence Intervals

- The confidence level describes the uncertainty associated with a *sampling method*. Suppose we used the same sampling method to select different samples and to compute a different interval estimate for each sample. Some interval estimates would include the true population parameter and some would not. A 90% confidence level means that we would expect 90% of the interval estimates to include the population parameter; A 95% confidence level means that 95% of the intervals would include the parameter; and so on.

How to Construct a Confidence Interval

There are four steps to constructing a confidence interval:

- Identify a sample statistic. Choose the statistic (e.g, sample mean, sample proportion) that you will use to estimate a population parameter.
- Select a confidence level. As we noted in the previous section, the confidence level describes the uncertainty of a sampling method. Often, researchers choose 90%, 95%, or 99% confidence levels; but any percentage can be used.

How to Construct a Confidence Interval

- Find the margin of error. If you are working on a homework problem or a test question, the margin of error may be given. Often, however, you will need to compute the margin of error, based on one of the following equations.

- Margin of error = Critical value * Standard deviation of statistic
  Margin of error = Critical value * Standard error of statistic

- Specify the confidence interval. The uncertainty is denoted by the confidence level. And the range of the confidence interval is defined by the following equation.
- Confidence interval = sample statistic $\pm$ Margin of error

Confidence level	Z-value
90%	1.645
95%	1.96
99%	2.58

Sampling Variation

100 patients with Carcinoma of the lung treated with new drug

Mean survival time =27.5 with a standard deviation of 25 months. Calculate a 95% confidence interval survival time for the population that has taken the drug

Sample n = 100 people

x= 27.5 months

s = 25 months

$$\mu = \bar{x} \pm z\sigma_{\bar{x}}$$

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{25}{10}$$

$$\sigma_{\bar{x}} = 2.5$$

$$\mu = \bar{x} \pm z\sigma_{\bar{x}}$$

$$\mu = 27.5 \pm 1.96\sigma_{\bar{x}}$$

$$\mu = 27.5 \pm 1.96(2.5)$$

$$\mu = 27.5 \pm 4.9$$

$$\mu = (27.5 - 4.9, 27.5 + 4.9)$$

$$\mu = (22.6, 32.4)$$

The mean systolic blood pressure (SBP) of a sample of 250 women is 130 mmHg. The population standard deviation is 20 mmHg. Find the 90% confidence interval.

HYPOTHESIS TESTING AND ONE SAMPLE T-TEST

Outline

- Hypothesis Testing

- One –Sample t-test

Hypothesis Testing

Hypothesis Testing

- Make an assumption / Establish a hypothesis
- Gather information relating to your assumption.
- Use statistical techniques to either reject or not reject the assumption, based on collected information.

Why is it necessary to test hypotheses?

- Draw conclusions about a population parameter

- Minimize likelihood of making decision attributed to chance occurrence only.

Example

Data of 100 underweight infants, weighing less than 1500 grams, born in two teaching hospitals in KZN. Interested to know if head circumference of babies born to mothers that were diagnosed with toxaemia is the same as that of mothers that were not diagnosed with toxemia.

Recorded information: Head circumference, height, gestational age, birth weight, mother’s age, mother’s diagnosis of toxaemia.

Two possibilities in real life

– Head circumferences are the same

– Head circumferences are not the same

Two possible "statistics" decisions

– Head circumferences are the same

– Head circumferences are not the same

		Real Life	
Decision		HC are the same	HC Not the same
	HC are the same	correct	incorrect
	HC are not the same	incorrect	correct

Hypothesis Testing

- Minimize likelihood of making decision attributed to chance occurrence only.

Steps in Hypothesis Testing

- ✓ Establish H_0 and H_a
- ✓ Set the significance level α (usually 0.05)
- ✓ Choose the appropriate statistical test
- Calculate the appropriate test statistic
- A test statistic is a random variable that is calculated from sample data and it measures the degree of agreement between a sample of data and the null hypothesis.
- Calculate the probability value (p value)
- ✓ Compare the p value with the significance level.

Steps in Hypothesis Testing (cont)

- ✓ Make a decision regarding H_0. If the p value is $< \alpha$ we reject H_0. If not we fail to reject the H_0.
- ✓ When the null hypothesis is rejected, the outcome is said to be "statistically significant"; when the null hypothesis is not rejected then the outcome is said be "not statistically significant."
- ✓ Draw a conclusion regarding your original research hypothesis based on your decision above.

Null Hypothesis

✓Establish H_0 and H_a

- **Null hypothesis (H_0)**: The hypothesis of no effect/difference/association
- Represents current “state of knowledge” (i.e. no conclusive research exists)

Alternative Hypothesis

Alternative hypothesis (H_a) : The hypothesis of effect/difference/association

Typically represents what you are trying to prove

Hypothesis Testing

- Testing both hypotheses at the same time
- Results allows us to “Reject” null hypothesis or “fail to reject” null hypothesis

Hypothesis Testing

		TRUTH	
		H_o True	*H_o False*
INFERENCE (decision)	*Do not reject H_o*	Correct	**Type II error (β)**
	Reject H_o	**Type I error (α)**	Correct Power (1-β)

Hypothesis Testing

✓ Set the significance level α (usually 0.05)

➢ Alpha α is the error rate that you are willing to accept.

An alpha of 0.05 means that you are willing to accept that there is a 5% possibility that your results are due to chance

Hypothesis Testing

✓ Choose the appropriate statistical test

- Types of variables
- Distribution of data
- Number of groups (one, two or more)
- If more than one group, are they dependent or not?

Statistical tests for continuous data

Number of groups	Dependent / Independent	Statistical test
One	N/A	One- Sample test
Two	Independent	Two-independent samples t-test
Two	Dependent	Paired-samples t-test
Three	Independent	One-way Analysis of Variance

Assume normal distribution

Non-parametric tests

Number of groups	**Dependent / Independent**	**Statistical test**
One	N/A	
Two	Independent	Mann Whitney U test / Wilcoxon ranksum test
Two	Dependent	Wilcoxon signrank test
Three	Independent	Kruskal Wallis test

For skewed data

Hypothesis Testing

- Calculate the appropriate test statistic

  - Depends on the statistical technique used
  - Standard formula for each test

Hypothesis Testing

- Calculate the probability value (p value)

- **<u>p-value</u>**
  - the probability the observed results occurred by chance

    - Stats software calculates the exact p-values

    - May be estimated from statistics tables

Hypothesis Testing

- ✓Compare the p value with the significance level α.

- ✓Make a decision regarding $H_{0:}$ If the p value is < α we reject H_0. If not we fail to reject the $H_{0.}$

One- sample T-test

Statistical Tests - Examples

9 children with congenital heart disease

Mean age to start walking = 12.8 ± 2.4 m

General population walk at 11.4 m

Does the age to start walking of children with CHD differ from that of children in the general population

Statistical Tests - Examples

- ✓ Establish H_0 and H_a

  H_0: $\mu_{CHD} = \mu_{GP}$

  H_a: $\mu_{CHD} \neq \mu_{GP}$

- ✓ Set the significance level α (usually 0.05)

Statistical Tests - Examples

✓ Choose the appropriate statistical test

Type of variable (age) – continuous
Distribution of age – normal
Number of groups – one

Appropriate test statistic: one-sample ttest

Statistical Tests - Examples

- Calculate the appropriate test statistic

$$t_{n-1} = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$t_{n-1} = \frac{12.8 - 11.4}{2.4/\sqrt{9}}$$

$$t_{n-1} = 1.75$$

Statistical Tests - Examples

- Calculate the probability value (p value)

$p=0.118$

✓ Compare the p value with the significance level α.

Statistical Tests - Examples

- ✓ Make a decision regarding $H_{0:}$ If the p value is < α we reject H_0. If not we fail to reject the $H_{0.}$

- ➢ Since p>α, we fail to reject the null hypothesis

Statistical Tests - Examples

- Draw a conclusion regarding your original research hypothesis based on your decision above.

Age to start walking for children that have CHD is not different from the age to start walking of the children in the general population.

Hypothesis Testing Using Statistics Tables

- Check the critical value associated with the test statistics in the table
- Compare the critical value with the calculated test statistics
- Decision rule: Reject the null hypothesis if the calculated value if greater than the critical value.

Hypothesis Testing Using Statistics Tables

Compare the critical value with the calculated test statistics.

$$t_{n-1} = 1.75$$

At 8 df the critical values are $\pm$ 2.31

Statistical Tests - Examples

Decision rule: Reject the null hypothesis if the calculated value is greater than the critical value.

Questions?

P-values for one sample T test for the population mean, μ

Null Hypothesis	$H_0\colon \mu = \mu_0$		
Alternative Hypothesis	$H_1\colon \mu < \mu_0$ (left-tailed test)	$H_1\colon \mu > \mu_0$ (right-tailed test)	$H_1\colon \mu \neq \mu_0$ (two-tailed test)
Test statistic	$t_0 = \dfrac{\bar{x} - \mu_0}{S/\sqrt{n}}$ with $n - 1$ dof		
P-value	$P(t \leq t_0)$	$P(t \geq t_0)$	$2 \times P(t \geq \lvert t_o \rvert)$

Example 3:

A public health official is concerned that a factory filling '500 gram' containers of nutritional supplements may be overfilling them, which could impact consumer safety and compliance with regulatory standards. A random sample of 20 of these containers is tested, revealing a mean weight of 501.8 grams with a variance of 30. At the 5% level of significance, determine if there is enough evidence to support the official's concern about the potential overfilling of these containers

$n = 20$, $\bar{x} = 501.8$, $S^2 = 30$ (sample variance), $\alpha = 0.05$

Referring to being calculated from the sample of 20.

Solution:

$H_0: \mu = 500$

$H_1: \mu > 500$ (right-tailed test)

$n = 20,\quad \bar{x} = 501.8,\quad S^2 = 30,\quad S = \sqrt{30}$

$H_0: \mu = 500 \longrightarrow \mu_0 = 500$

$H_1: \mu > 500$

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$= \frac{501.8 - 500}{\sqrt{\frac{30}{20}}}$$

$$= 1.470$$

Always round t_0 off to 3 decimal places

Since the test statistic follows a t-distribution, the t-tables must be used…

Using the t-table:

$H_0: \mu = 500 \qquad t_0 = 1.470$

$H_1: \mu > 500 \qquad df = 19$

For a right-tailed test:

P-value $= P(t \geq t_0)$

$\qquad = P(t \geq 1.470)$

$\therefore$ $0.05 <$ P-value < 0.1

Even without finding the exact P-value, we can still obtain our decision:

Thus, P-value > 0.05

Percentage points for the Student's t-distribution

df	α						
	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
[illegible]	[illegible]	[illegible]	[illegible]	2.650	3.012	3.852	4.221
[illegible]	[illegible]	[illegible]	[illegible]	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	$\boxed{1.328}$	$\boxed{1.729}$	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850

$1.328 < t_0 = 1.470 < 1.729$

$\therefore$ Since P-value $> \alpha = 0.05$, H_0 is NOT rejected at a 5% level of significance.

$\therefore$ Conclusion: There is insufficient evidence to conclude that the factory is over filling the 500g containers