DOKUZ EYLÜL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

THE CALCULATION AND INTERPRETATION OF THE ENTROPIES OF SURVEY RESULT IN EDUCATION

by

Uğur ATEŞ

August, 2015

İZMİR

THE CALCULATION AND INTERPRETATION OF THE ENTROPIES OF SURVEY RESULT IN EDUCATION

A Thesis Submitted to the Graduate School of Natural and Applied Sciences of Dokuz Eylül University In partial Fulfillment of the Requirements for the Degree of Master of Science in Statistics Program

by

Uğur ATEŞ

August, 2015

İZMİR

M.Sc. THESIS EXAMINATION RESULT FORM

We have read the thesis entitled **"THE CALCULATION AND INTERPRETATION OF THE ENTROPIES OF SURVEY RESULT IN EDUCATION"** completed by **UĞUR ATEŞ** under supervision of **ASSIST. PROF. DR. TUĞBA YILDIZ** and we certify that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Sciences.

Assist. Prof. Dr. TUĞBA YILDIZ

Supervisor

Assoc. Prof. Dr. İlhan KARAKILIÇ

(Jury Member)

Assoc. Prof. Dr. Özlem EGE ORUÇ

(Jury Member)

Prof.Dr. Ayşe OKUR

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my supervisor Assist. Prof. Dr. Tuğba YILDIZ, for her support and invaluable guidance throughout the study. I also wish to thank the jury members for their constructive criticism.

I would like to acknowledge several people who kindly provided me with valuable information and materials, without which this study would not have been possible.

I am indebted to Assoc. Prof. Dr. Özlem Ege ORUÇ for helping me understand the mathematical base of statistics which was crucial for this study.

I also thank Assoc. Prof. Dr. Sedat ÇAPAR and Assoc. Prof. Dr. Emel KURUOĞLU for helping me learn some statistical software which I used frequently in this study.

Thanks are extended to Assist. Prof. Dr. Hatice Gonca Usta SAYIN of Department of Educational Measurement and Evaluation in Cumhuriyet University for being generous with her time and share her knowledge and experiences with me.

I also extend gratitude to Abdulkadir GÜLLÜ, my true and eternal friend, who helped me with the translation of some parts in this study.

I would like to thank my parents, Eser and Ufuk ATEŞ for their endless love and support in every phase of my life.

Finally, special thanks are for my lovely little sister, Esin, for being not just a sister but also a close friend.

Uğur ATEŞ

THE CALCULATION AND INTERPRETATION OF THE ENTROPIES OF SURVEY RESULT IN EDUCATION

ABSTRACT

In this thesis, the application of entropy to measurement and evaluation which is one of the most important fields of education, is studied.

In the study, a test that is approved in terms of reliability, validity and applicability is administered. The test is applied to 31 chosen freshmen out of 180 with simple random sampling from Vezirköprü Vocational and Technical Anatolian High School. Some descriptive statistics on the scores obtained from subscales of the test; Mathematics, Grammar, Social Studies and Science, are calculated and probability distribution tables are constructed. Based on these tables, Shannon entropy and joint entropy values are computed. According to the results, information obtained from the test can be attained with fewer questions.

Keywords: Shannon entropy, joint entropy, measurement and evaluation in education, descriptive evaluation

EĞİTİMDE ÖLÇME SONUÇLARININ ENTROPİLERİNİN HESAPLANMASI VE YORUMLANMASI

ÖZ

Bu tezde entropi kavramının eğitimin en önemli alanlarından biri olan ölçme ve değerlendirmede uygulaması çalışılmıştır.

Çalışmada güvenirliği, geçerliliği ve kullanışlılığı onaylanmış bir test kullanılmıştır. Bu test Vezirköprü Mesleki ve Teknik Anadolu Lisesine yeni başlayan 180 öğrenci içinden basit rastgele örnekleme ile seçilen 31 öğrenciye uygulanmıştır. Testin alt ölçekleri olan Matematik, Dil Bilgisi, Sosyal Bilgiler ve Fen Bilgisi ile elde edilen puanlara ait bazı tanılayıcı istatistikler hesaplanmış ve olasılık dağılım tabloları oluşturulmuştur. Bu tablolara dayanarak Shannon entropi ve bileşik entropi değerleri hesaplanmıştır. Elde edilen sonuçlara göre bu testin verdiği bilgiye daha az soru ile ulaşılabilmektedir.

Anahtar kelimeler: Shannon entropi, bileşik entropi, eğitimde ölçme ve değerlendirme, tanılayıcı değerlendirme

CONTENTS

Pages

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

The history of the word entropy can be traced back to 1865 when the German physicist Rudolf Clausius tried to give a new name to irreversible heat loss, what he previously called “equivalent-value”. The word “entropy” was chosen because in Greek, “entropein” means “content transformative” or “transformation content” (Laidler, 1995). Since then entropy has played an important role in thermodynamics. Being defined as the sum of heat supplied divided by temperature, it is central to the Second Law of Thermodynamics (Clausius, 1865). It also helps measure the amount of order and disorder and/or chaos. Entropy can be defined and measured in many other fields than the thermodynamics. For instance, in classical physics, entropy is defined as the quantity of energy incapable of physical movements. Von Neumann used the density matrix to extend the notion of entropy to quantum mechanics. The entropy of a random variable measures uncertainty in probability theory. The term by itself in this context usually refers to the Shannon entropy which is given by Shannon. This entropy quantifies, in the sense of an expected value, the information contained in a message. Entropy quantifies the exponential complexity of a dynamical system, that is, the average flow of information per unit of time in the theory of dynamical systems (Rongxi Zhou, Ru Cai & Guanqun Tong, 2013). In sociology, entropy is the natural decay of structures (Liu, Liu & Wang, 2011)

Examinations are used to collect quantitative information about items in a population such as classrooms. Developing an examination is a very important duty especially for a teacher. Because the exam has to be reliable, valid and applicable. In this case reliability is the degree to which an assessment tool produces stable and consistent results. Validity refers to how well a test measures what it is purported to measure. While reliability is necessary, it alone is not sufficient. For a test to be reliable, it also needs to be valid. Applicable means applied and evaluated easy and in a short period of time. An examination that measures students’ knowledge and abilities is dealt with in this study. By how many questions the intended information would be reached using this scale with the calculated entropy values are investigated.

The organization of this thesis is as follows: In chapter one, a general introduction is presented. From a general point of view, information theory, entropy and measurement in education is given. In chapter two, fundamental concepts of measurement and evaluation in education is explained. In chapter three, information theory, entropy and its properties are discussed as the theoretical background. In chapter four, the application of information theory on measurement in education which is the main topic of the study is given.

CHAPTER TWO
SOME CONCEPTS OF MEASUREMENT AND EVALUATION IN EDUCATION

Measurement generally means to determine a characteristics' amount. In other words, measurement is the assignment of number to objects or events (Pedhazur & Schmelkin, 1991). Educational measurement refers to the use of educational assessments and the analysis of data such as scores obtained from educational assessments to infer the abilities and proficiencies of students. The approaches overlap with those in psychometrics. Educational measurement is the assigning of numbers to traits such as achievement, interest, attitudes, aptitudes, intelligence and performance. Beyond its general definition, refers to the set of procedures and the principles for how to use the procedures in educational tests and assessments. Some of the basic principles of measurement in educational evaluations would be raw scores, percentile ranks, derived scores, standard scores, etc. (Overton, 2011).

Evaluation: Procedures used to determine whether the subject (in example student) meets preset criteria, such as qualifying for special education services. This uses assessment to make a determination of qualification in accordance with predetermined criteria.

Assessment: The process of gathering information to monitor progress and make educational decisions if necessary. An assessment may include a test, but also includes methods such as observations, interviews, behavior monitoring, etc. To many teachers and students, "assessment" simply means giving students tests and assigning them grades. This conception of assessment is not only limited, but also limiting. It fails to take into account both the utility of assessment and its importance in the teaching/learning process. In the most general sense, assessment is the process of making a judgment or measurement of worth of an entity for example person, process, or program. Educational assessment involves gathering and evaluating data evolving from planned learning activities or programs. This form of assessment is often referred to as evaluation. Learner assessment represents a particular type of educational assessment normally conducted by teachers and designed to serve several related purpose (Brissenden and Slater, n.d.). These purposed include:

- Motivating and directing learning
- Providing feedback to student on their performance
- Providing feedback on instruction and/or the curriculum
- Ensuring standards of progression are met

Learner assessment is best conceived as a form of two-way communication in which feedback on the educational process or product is provided to its key stakeholders (McAlpine, 2002). Specifically, learner assessment involves communication to teachers (feedback on teaching), students (feedback on learning), curriculum designers (feedback on curriculum) and administrators (feedback on use of resources).

For teachers and curriculum/course designers, carefully constructed learner assessment techniques can help determining whether or not the stated goals are being achieved. According to Brissenden and Slater (n.d.), classroom assessment can help teachers answer the following specific questions:

- To what extent are my students achieving the stated goals?
- How should I allocate class time for the current topic?
- Can I teach this topic in a more efficient or effective way?
- What parts of this course/unit are my students finding most valuable?

First and foremost, assessment is important because it drives students learning (Brissenden and Slater, n.d.). Whether we like it or not, most students tend to focus their energies on the best or most expeditious way to pass their 'tests'. Based on this knowledge, we can use our assessment strategies to manipulate the kinds of learning that takes place. For example, assessment strategies that focus predominantly on recall of knowledge will likely promote superficial learning. On the other hand, if we choose assessment strategies that demand critical thinking or creative problem-solving, we are likely to realize a higher level of student performance or achievement. In addition, good assessment can help students become more effective self-directed learners (Angelo & Cross, 1993).

Bachman (1990), quoting Weiss (1972) defines evaluation as “the systematic gathering of information for the purpose of making decisions”. Lynch (2001) adds the fact that this decision or judgment is to be about individuals. In this conceptualization, both authors agree that evaluation is the superordinate term in relation to both measurement and testing. Assessment is sometimes used interchangeably for evaluation. The systematic information can take many forms, but these forms are either quantitative or qualitative. This is what distinguishes measures from qualitative descriptions.

There are three forms of assessment:

- Diagnostic: Pre testing students to see what they know before teaching the unit. Diagnostic assessment is used to diagnose the student’s strengths and weaknesses. This will help teachers identify what they need to do to help the student become stronger.
- Formative: Assessing students’ strengths and weaknesses, and providing feedback during the unit. Formative assessment is the ongoing analysis of a student’s needs - we must recognize when it needs to be learned, motivated, and provided with positive reinforcement in order for it to grow. Most formative assessment is informal. The feedback and response involves both teacher and student. Formative assessment has the greatest impact on learning and achievement. Formative assessment works best when the students understand what the learning objectives are.
- Summative: Testing the student’s knowledge at the end of teaching a unit. Summative assessment of the measuring of the student’s growth at a point. The measurement tells us how much the students have grown. It does not affect the growth of the students. Measures the progress a student has made and it is often used at the end of a unit of work.

CHAPTER THREE
SOME CONCEPTS OF INFORMATION THEORY AND ENTROPY

In the early 1940s it was thought to be impossible to send information at a positive rate with negligible probability of error. Shannon surprised the communication theory community by proving that the probability of error could be made nearly zero for all communication rates below channel capacity. The capacity can be computed simply from the noise characteristics of the channel. Shannon further argued that random processes such as music and speech have an irreducible complexity below which the signal cannot be compressed. This he named the entropy, in deference to the parallel use of this word in thermodynamics, and argued that if the entropy of the source is less than the capacity of the channel, asymptotically error-free communication can be achieved (Cover & Thomas, 2006).

Information theory is a branch of applied mathematics and electrical engineering involving the quantification of information. Historically, information theory was developed by Claude E. Shannon (1948) to find fundamental limits on compressing and reliably storing and communicating data. Since its inception it has broadened to find applications in many other areas, including statistical inference, natural language processing, cryptography, engineering, biology, medical science, sociology, and psychology. Figure 1.1 shows the areas that are in relation with information theory and some common areas of research. According to this figure, some of the common areas of research for information theory and statistics can be exemplified as hypothesis thesis and fisher informatics (Cover & Thomas, 2006).

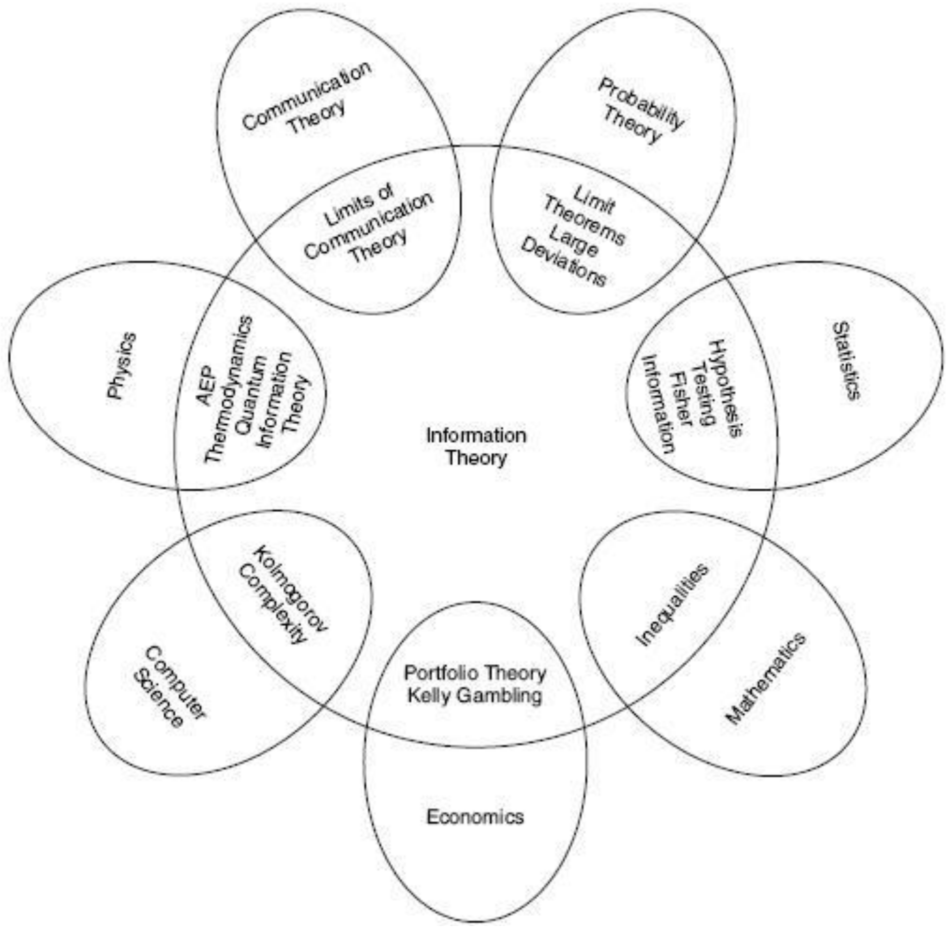

Figure 3.1 Areas related with information theory (Cover & Thomas, 2006)

A key measure of information in the theory is known as entropy, which is usually expressed by the average number of bits needed for storage or communication. Intuitively, entropy quantifies the uncertainty involved when encountering a random variable (Gündüz, 2010). This chapter describes Shannon entropy, joint entropy, relative entropy, conditional entropy and mutual information.

3.1 Entropy

Shannon argues that, the process of transmitting the message that has been produced in the source is one of a probabilistic one in information theory. Acquiring information about some particular event is valid only if there is indeterminacy on that event. The required information for a system's probable states accurately equals to the entropy of that system. By this approach entropy can be defined as the expected value of the states that an event can take (Dursun, 2009). Simply, entropy is a measure of uncertainty associated with a random variable. There are many kinds of entropy measures. Shannon entropy is widely used in information theory. It is explained in this part.

Let X be a discrete random variable which can take $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ respectively such as

$$\forall\ \mathrm{i} = 1,2,\ldots,\mathrm{n} \qquad p_i \geq 0 \ \text{ and } \ \sum_{i=1}^{n} p_i = 1 \tag{3.1}$$

The Shannon entropy value of discrete random variable X is calculated by the equation:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, log_2 p(x_i) = -\,\mathrm{E}[\log_2 x_i] = \ -\,\mathrm{E}(I_x\ ) \tag{3.2}$$

If X is continuous random variable with

$$\forall \mathrm{x} \in (-\infty, \infty) \qquad \mathrm{f(x)} \geq 0 \ \text{ and } \ \int_{-\infty}^{\infty} f(x) = 1 \tag{3.3}$$

The Shannon entropy is calculated by

$$H(X) = -\int f(x) log_2 f(x) dx = \ -E[log_2 f(x)] = -E(I_x) \tag{3.4}$$

For both cases the information that obtained from X is given in equation (3.5)

$$I_X = -\log_2 p_i \ \ or \ I_X = -\log_2 f(x) \tag{3.5}$$

The notation $\mathrm{E}(\mathrm{I}_\mathrm{X}\ )$ denotes the expected value of I_X. Entropy can also be called average uncertainty. The entropy function H(X) is maximum when $\mathrm{p}_\mathrm{i} = 1/\mathrm{n}$ for all i.

This makes intuitive sense, because uncertainty is greatest when all outcomes are equally likely. The entropy graph of an event with two possible equal results ($p_1 = p_2 = 0.5$) is given in Figure 3.2 (Cover & Thomas, 2006).

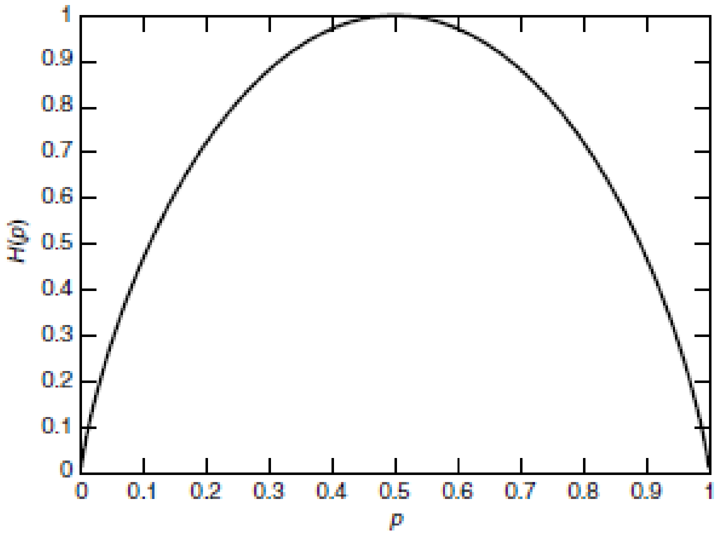

Figure 3.2 Entropy values for probabilities

The below inferences can be made about entropy according to the occurrence probabilities of different states;

- No information occurs with a state, having an occurrence probability 1.

$$\exists i, \qquad p(x_i) = 1 \Rightarrow E(X) = 0 \tag{3.6}$$

- The entropy value of a system increases as the probable states of that systems increase.
- Occurrence of a state with lower occurrence probability accumulates more amount of information than the one with higher probability.
- As entropy of a system increases, the estimation or knowing of the results beforehand gets more difficult. The power of estimation will decrease since indeterminacy increases (Dursun, 2009).

3.2 Joint, Conditional and Relative Entropy

In information theory, joint entropy is a measure of the uncertainty associated with a set of variables.

In information theory, the conditional entropy quantifies the amount of information needed to describe the outcome of a random variable Y given that the value of another random variable X is known.

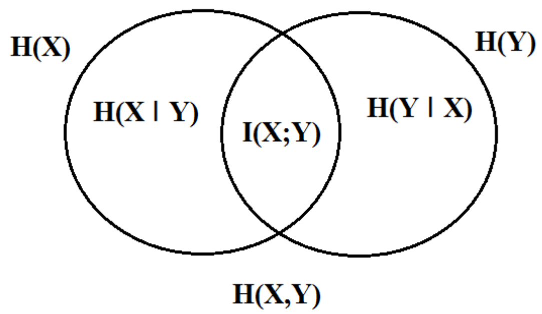

Figure 3.3 Venn diagram for various information measures

Figure 3.3 shows that the relationship between entropy, conditional entropy and mutual information. The set on the left is marginal entropy of X. Similarly the right set is marginal entropy of Y. The union of both sets is the joint entropy, H(X,Y). The sets H(X I Y) and H(Y I X) are conditional entropies. The intersection of this set is the mutual information, I(X;Y).

In information theory, the relative entropy, namely the Kullback-Leibler divergence is a non-symmetric measure of the difference between two probability distributions P and Q. Specifically, the Kullback-Leibler divergence of Q from P, is a measure of the information lost when Q is used to approximate P. It measures the expected number of extra bits required to code samples from P when using a code optimized for Q, rather than using the true code optimized for P.

3.2.1 Joint Entropy

X and Y be two discrete random variables taking values $\{x_1,\ldots..,x_n\}$ and $\{y_1,\ldots\ldots,y_m\}$ respectively. If $P(X=x_i, Y=y_j)$ denote the joint probability mass function of X and Y. The joint entropy of these random variables is obtained by;

$$H(X,Y) = -\sum_j \sum_i P\left(X = x_i, Y = y_j\right) \log P\left(X = x_i, Y = y_j\right) \tag{3.7}$$

If X and Y are independent, then the joint entropy equals to the sum of the marginal entropies of each random variable and the formula is given by;

$$H(X,Y) = H(X) + H(Y) \tag{3.8}$$

Joint entropy is also called the common information measure.

3.2.2 Conditional Entropy

X and Y are random variables that have joint probability distributions. When the values of the random variable Y are given, the measurement of the uncertainty in the random variable X is the conditional entropy of X dependent on Y. Knowing Y always decreases the uncertainty of X. It is denoted by H(X|Y) and calculated as follows (Cover & Thomas, 2006).

$$H(X|\mathrm{Y}) = -\sum\sum p\left(x_i, y_j\right) \log\left(p\left(x_i \middle| y_j\right)\right) \tag{3.9}$$

If the variables X and Y are independent, by using the chain rule, the relationship between conditional and joint entropy can be written as follows.

$$H(X,Y) = H(X) + H(Y|\mathrm{X}) \tag{3.10}$$

3.2.3 Relative Entropy

The Kullback-Leibler divergence is a non-commutative measure of the divergence between two probability distributions p and q (Kullback, 1987). Kullback-Leibler is also sometimes called the information gain about X if p is used instead of q. It is also called the relative entropy in using q in the place of p. The relative entropy is an appropriate measure of the similarity of the underlying distribution. It is denoted by $D(p \parallel q)$ and is given below,

$$D(p \parallel q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} \tag{3.11}$$

The properties of the relative entropy equation make it non-negative, non-symmetric and it is zero if both distributions are equivalent namely p=q. The smaller the relative entropy is the more similar the distribution of the two variables and vice versa. (Kullback, 1987; Leutenneger, 2000) Note that $D(p \parallel q) \neq D(q \parallel p)$ in general.

3.3 Concept of Information and Properties

Information is a sophisticated concept. Therefore, it is hard to ensure a broad definition. Researchers come to common grounds in this point and develop mathematical terms for information systems analysis. Information is storable, visible, transferable, re-obtainable, observable and interpretable.

In this part of study, the concept of information used in the framework of information theory and its properties will be briefly explained. Information is the processed state of data and facts related to objects, event or people.
The value of;

$$I(x_i) = -\log_2(p_i) = \log_2 \frac{1}{p_i} \tag{3.12}$$

calculated for the $\{x_1, x_2, \ldots, x_i\}$ values of the discrete random variable X in the state of i = 1,2,..,n is called the information content of x_i. The information value of the random variable X is calculated given below.

$$I(x_i) = \sum_i p_i I(x_i) \qquad (3.13)$$

This value is the weighted average of the information contents of the values that X has taken, and the probability of taking these values; and at the same time it is called entropy. The information content that the random variable takes is only dependent on the random variable's probability of the taking that value. As lower this probability is so bigger is the information content (Dursun, 2009).

There are four basic axioms for information:

- Information is a value that is not negative.

$$I(p) \geq 0 \qquad (3.14)$$

- The information value of an accurate event is zero.

$$I(1) = 0 \qquad (3.15)$$

- For two independent event, the information is obtained from observations equals to the sum of two information.

$$I(p_1 * p_2) = I(p_1) + I(p_2) \qquad (3.16)$$

- I(p) is monotonous and constant.

3.4 Mutual Information

The mutual information between X and Y is obtained by as follows,

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

$$= H(X) - H(X|\mathrm{Y}) = \mathrm{H(Y)} - \mathrm{H(Y|X)} \qquad (3.17)$$

or

$$I(X;Y) = \sum_{x,y} P(X = x, Y = y) \log \frac{P(X = x, Y = y)}{P(X = x)P(Y = y)} \qquad (3.18)$$

The mutual information I(X;Y) of two random variables X and Y is the Kullback-Leibler divergence between their joint distribution and the product of their marginal distributions. By the definition the mutual information provides some measure of the dependence between the variables.

Some important properties of mutual information are given below:

- If X and Y are independent, I(X;Y)=0 if not, H(X|Y)=0 and I(X;Y)=H(X)
- Symmetry, I(X;Y)=I(Y;X)
- I(X;Y)≥0
- I(X;X)=H(X)
- Chain rule: $I(X_1,X_2,\ldots,X_n;Y)=\sum I(X_i;Y|X_1,\ldots,X_{i-1})$

3.5 Relationship Between Entropy and Information

The definition of information is done by taking the definition of entropy, which measures the randomness in a system, as a model. Therefore, information and entropy can be considered as intertwined concepts. Answering the question yes or no which is assumed to be that is uncertainty, there is uncertainty for answers. Here the question carries an information value. If the answer is known accurately, asking the question will be unnecessary. Increase of information causes entropy to decrease by decreasing uncertainty. Thus, minimum uncertainty is obtained by maximum information.

$$I(X;Y) = \log(P(X|\mathrm{Y})) + \mathrm{I(X)} = \mathrm{H(X)} - \mathrm{H(X|\,Y)} \qquad (3.19)$$

$$H(X,Y) = H(X) + H(Y|\,\mathrm{X}) = \mathrm{H(Y)} + \mathrm{H(X|\,Y)} \qquad (3.20)$$

The relations between the conditional entropies and joint of the random variables X and Y can be defined as equation above (Cover & Thomas, 2006).

CHAPTER FOUR
APPLICATIONS

It has always been a challenge for a teacher who has just started working in a school to determine the readiness of students. The teacher should be aware of the readiness of the school as a whole, the classes separately and the students individually. With the start of the term, the teacher designates the subjects to be covered and the activities to be done in accordance with the level of the students in a short period of time. Two decisive factors that show how difficult this is to do are that the school is highly populated and that the subjects are numerous. Regarding such a situation for a teacher who starts working in a crowded school, it is obvious that a practical approach to determine the level is needed.

In this study, a test of 100 questions, equally divided into four subjects; Mathematics, Grammar, Social Studies and Science, is administered to 31 students chosen with simple random sampling from 180 ninth grade students studying at Vezirkopru Vocational and Technical Anatolian High School. Each question is graded with 4 points. The results are entitled as “poor” for points from 0 to 20, “mediocre” from 20 to 40, “good” from 40 to 60, “great” from 60 to 80 and “wonderful” from 80 to 100. According to the entropy values gathered from scores, conclusions have been reached on the optimum numbers of questions for each subject in order to measure the level of other students that are assumed to have similar qualifications. The goal of this application is to determine the level of readiness of the students based on entropy values in a practical way.

The questions representing the lessons in the survey and the probability distribution tables are created separately for each lesson mentioned, utilizing the frequency values calculated in accordance with the scores of the questions for each lesson. With the use of these probability distribution tables, Shannon entropy values are computed for Maths, Grammar, Social Studies and Science. While examining the kind of entropy values each lesson had with residence, joint probability distribution tables are constructed separately from the frequencies obtained from Residence -

Maths, Residence - Grammar, Residence - Social Studies and Residence - Science scores. The joint entropy values of all lessons and residence are calculated separately from the joint probability distribution tables. Mutual information values are computed separately for each lesson and residence, utilizing the same joint probability distribution tables. All values are clarified within the scope of the information theory.

The Shannon entropy values are calculated by using the probability distributions constructed for the random variables of Maths, Grammar, Social Studies and Science. The frequencies, probabilities and entropy values of these random variables are shown in Table 4.1.

Table 4.1 Frequency and probability table for the maths, grammar, social studies and science random variables

	Maths		**Grammar**		**Social Studies**		**Science**	
	f	**P**	**f**	**P**	**f**	**P**	**f**	**P**
0-20	10	0.323	6	0.194	7	0.226	17	0.548
20-40	10	0.323	7	0.226	5	0.161	10	0.323
40-60	9	0.290	7	0.226	11	0.355	3	0.097
60-80	2	0.065	7	0.226	6	0.194	1	0.032
80-100	0	0.000	4	0.129	2	0.065	0	0.000
TOTAL	31	1	31	1	31	1	31	1
ENTROPY	**1.826**		**2.294**		**2.153**		**1.488**	

In addition, the bar graph of the entropy values for these random variables is presented in Figure 4.1.

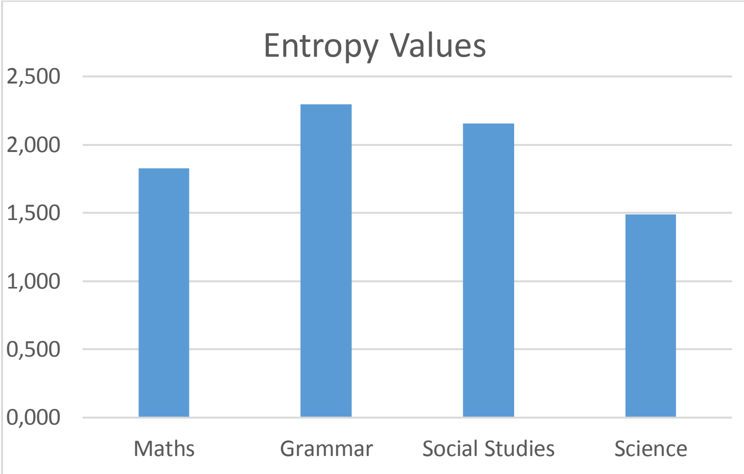

Figure 4.1 Entropy values for maths, grammar, social studies and science random variables

In here entropy value 1.826 is calculated by the formula (3.2) as follows:

$$H(X) = -\sum_{i=1}^{4} p(x_i)\, log_2 p(x_i)$$

$$= -(0.323 \log_2 0.323 + 0.323 \log_2 0.323 + 0.29 \log_2 0.29 + 0.065 \log_2 0.065)$$

$$= 1.826\ bits \tag{4.1}$$

The entropy value 1.826 of Maths indicates that it is enough to ask 2 questions for Maths to have information about students' readiness. Likewise, the entropy values found for Grammar 2.294, Social Studies 2.153 and Science 1.488 also indicate that it would be enough to ask 3 questions for Grammar, 3 questions for Social Studies and 2 questions for Science.

19 out of 31 students undertaking the test are living in small villages and 12 of them are living in the county. First, these two groups are examined individually and probability distribution tables are constructed, and then entropy values are calculated. Interpretations are done according to the values. Table 4.2 and Table 4.3 show probability distribution and entropy values of Residence - Maths, Residence - Grammar, Residence - Social Studies and Residence - Science.

Table 4.2 Probability distribution for village and maths, grammar, social studies and science random variables

VILLAGE	Maths		Grammar		Social Studies		Science	
	f	P	f	P	f	P	f	P
0-20	10	0.526	4	0.211	4	0.211	9	0.474
20-40	9	0.4737	4	0.211	4	0.211	8	0.421
40-60	0	0	6	0.316	6	0.316	2	0.105
60-80	0	0	4	0.211	4	0.211	0	0
80-100	0	0	1	0.053	1	0.053	0	0
TOTAL	19	1	19	1	19	1	19	1
ENTROPY	**0.998**		**2.168**		**2.168**		**1.472**	

In addition, entropy values for village and Maths, Grammar, Social Studies and Science random variables are presented with bar graph in Figure 4.2.

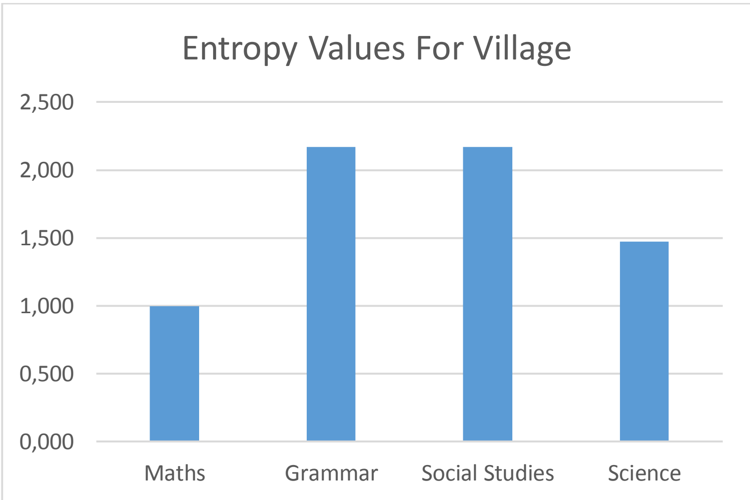

Figure 4.2 Entropy values for village and maths, grammar, social studies and science random variables

The entropy values calculated from the scores of students residing in villages seen in Table 4.2 are as follows: 0.998 for Maths, 2.168 for Grammar, 2.168 for Social Studies and 1.472 for Science. According to these entropy values, one can safely assume that 1 question for Maths, 3 for Grammar, 3 for Social Studies and 2 for

Science, are adequate to gather information about the general readiness of students residing in village.

Table 4.3 Probability distribution for county and maths, grammar, social studies and science random variables

COUNTY	Maths		Grammar		Social Studies		Science	
	f	P	f	p	f	P	f	P
0-20	0	0	2	0.167	3	0.25	8	0.667
20-40	1	0.083	3	0.250	1	0.083	2	0.167
40-60	9	0.75	1	0.083	5	0.417	1	0.083
60-80	2	0.167	3	0.25	2	0.167	1	0.083
80-100	0	0	3	0.25	1	0.083	0	0
TOTAL	12	1	12	1	12	1	12	1
ENTROPY	**1.041**		**1.730**		**2.055**		**1.418**	

In addition, entropy values for county and Maths, Grammar, Social Studies and Science random variables are presented with bar graph in Figure 4.3.

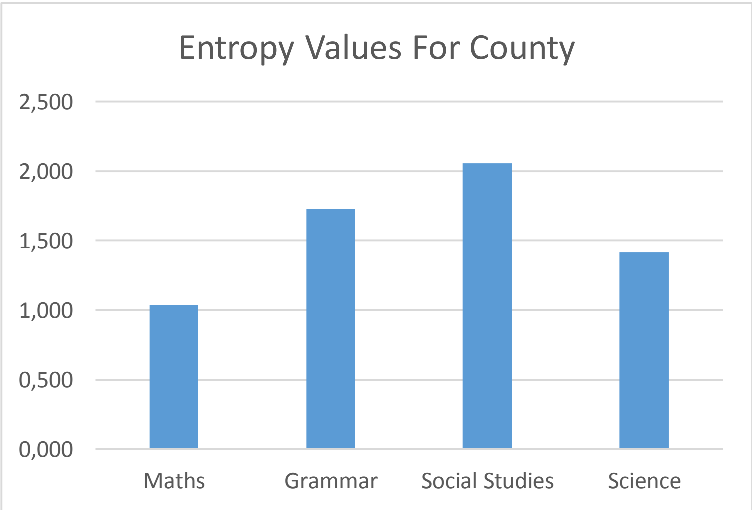

Figure 4.3 Entropy values for county and maths, grammar, social studies and science random variables

The entropy values calculated from the scores of students residing in counties seen in Table 4.3 are as follows: 1.041 for Maths, 1.73 for Grammar, 2.055 for Social Studies and 1.418 for Science. According to these entropy values, one can safely assume that, 2 questions for Maths, 2 for Grammar, 3 for Social Studies and 2 for Science, are adequate to gather information about the general readiness of students residing in county.

To investigate what kind of entropy values the variables of Mathematics, Grammar, Social Studies and Science considering residence, joint probability distribution tables are constructed separately from the frequencies obtained from Residence - Mathematics, Residence - Grammar, Residence - Social Studies and Residence - Science scores. Table 4.4 gives the joint probability distribution functions for all subscales and residence.

Table 4.4 Joint probability distributions of all subscales and residence

Maths

	0-20	20-40	40-60	60-80	80-100	TOTAL
Village	0.323	0.290	0.000	0.000	0.000	0.613
County	0.000	0.032	0.290	0.065	0.000	0.387
TOTAL	0.323	0.323	0.290	0.065	0.000	1.000

Grammar

	0-20	20-40	40-60	60-80	80-100	TOTAL
Village	0.129	0.129	0.194	0.129	0.032	0.613
County	0.065	0.097	0.032	0.097	0.097	0.387
TOTAL	0.194	0.226	0.226	0.226	0.129	1.000

Social Studies

	0-20	20-40	40-60	60-80	80-100	TOTAL
Village	0.129	0.129	0.194	0.129	0.032	0.613
County	0.097	0.032	0.161	0.065	0.032	0.387
TOTAL	0.226	0.161	0.355	0.194	0.065	1.000

Science

	0-20	20-40	40-60	60-80	80-100	TOTAL
Village	0.290	0.258	0.065	0.000	0.000	0.613
County	0.258	0.065	0.032	0.032	0.000	0.387
TOTAL	0.548	0.323	0.097	0.032	0.000	1.000

According to these joint probability distributions, joint entropy values are calculated as Table 4.5.

Table 4.5 Joint entropy values for all subscales and residence

Variables	Joint Entropy
Maths	1.977
Grammar	3.155
Social Studies	3.087
Science	2.356

In addition, the bar graph of joint entropy values is presented in Figure 4.4.

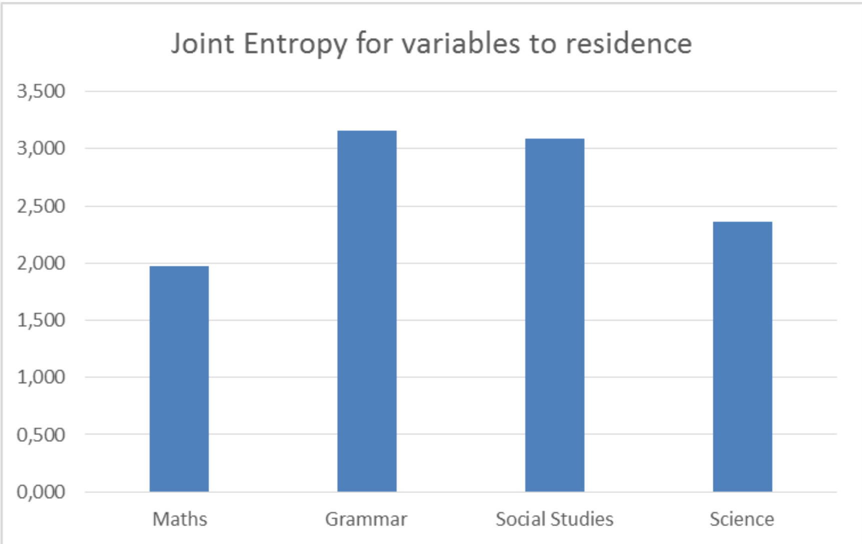

Figure 4.4 Joint entropy values for all subscales and residence

The result in the joint entropy seen in Table 4.5 H(X, Y) = 1.977 with X = Residence and Y = Maths means that on average it would require 2 questions to guess the level of both variables. The similar result is 4 questions for Grammar, 4 questions for Social Studies and 3 questions for Science.

Mutual information values are computed using the joint probability distribution calculated for residence and all subscales. The entropy value is found as 0.963 bits for residence and mutual information values calculated for Residence – Maths, Residence – Grammar, Residence – Social Studies and Residence – Science. Mutual information values are given for all variables in Table 4.6.

I(Maths, Residence) = H(Maths) + H(Residence) – H(Maths, Residence)
I(Maths, Residence) = 1.826 + 0.963 – 1.977 = 0.812 bits.

I(Grammar, Residence) = H(Grammar) + H(Residence) – H(Grammar, Residence)
I(Grammar, Residence) = 2.294 + 0.963 – 3.155 = 0.102 bits.

I(Social Studies, Residence) = H(Social Studies) + H(Residence) – H(Social Studies, Residence)
I(Social Studies, Residence) = 2.153 + 0.963 – 3.087 = 0.029 bits.

I(Science, Residence) = H(Science) + H(Residence) – H(Science, Residence)
I(Science, Residence) = 1.488 + 0.963 – 2.356 = 0.095 bits.

Table 4.6 Mutual information for variables and residence

Variables	Mutual Information
Maths	0.812
Grammar	0.102
Social Studies	0.029
Science	0.095

The interpretation of the calculated mutual information values seen in Table 4.6 is as follows. The mutual information value calculated for the Maths-Residence is 0.812 bits. Then the variables Maths and Residence seem to have some information in common. Knowing the residence of any student provides information about the level of the students' Maths score. The information values calculated for the Grammar-Residence, Social Studies-Residence and Science-Residence are 0.102, 0.029 and 0.095 bits, respectively. As a result there seem to have a little information in common between these variables.

The mutual information of two random variables is a quantity that measures the mutual dependence of two variables in probability theory and information theory. If X and Y are independent, then knowing X does not give any information about Y and the other way around, so their mutual information is zero.

CHAPTER FIVE
CONCLUSIONS

Information theory is an important subject that comprises lots of concepts including entropy. According to information theory, every random variable might be considered as information and entropy can be calculated and interpreted for every probability distribution.

Basic statistical subjects are frequently applied in measurement and evaluation in education. In this study, the statistical subject entropy is applied in survey results in education. The scores of students are converted to random variables accordingly and probability functions are constructed. Then, the entropies for these functions are computed and interpreted.

The interpretation of the calculated entropy values seen in Table 4.1, Table 4.2, Table 4.3, Table 4.5 and Table 4.6 are given in detail after each table. As in Table 4.1, it is sufficient to ask 10 questions so as to be informed about readiness of students who have just started Vezirkopru Vocational and Technical Anatolian High School. The entropy values calculated from the scores of students residing in villages and county are seen in Table 4.2 and Table 4.3 respectively. According to these entropy values, one can safely assume that 9 questions in total, are adequate to gather information about the general readiness of students residing in both village and county. Although for the students in both groups 9 questions in total are sufficient, questions needed for each lesson vary. As it is seen in the joint entropy values in Table 4.5, it is enough to ask 13 questions in total. Thus, asking 13 questions is a healthier way to determine students' readiness. Finally from the Table 4.6, the mutual information measures tell us that residence plays an important role for students' scores, mostly in Maths.

Consequently, with the acquired entropy values it has been concluded that the same result may be obtained by measuring student levels with the fewest questions possible for each lesson. The joint and separate entropy values have been calculated based on residence of students. Thus, it is evident that a teacher can reach a conclusion in a more practical way using fewer questions while doing descriptive evaluation.

REFERENCES

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers.* (2nd Ed.). San Francisco, CA: Jossey-Bass.

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Brissenden, G., & Slater, T. (n.d.). Retrieved February 17, 2015, from http://www.flaguide.org

Clausius, R. (1865). Ueber verschiedene für die Anwendung bequeme formen der Hauptgleichungen der mechanischen Wärmetheorie. *Annalen der Physic, 201* (7), Retrieved December 12, 2014, from http://onlinelibrary.wiley.com/doi/10.1002/andp.18652010702/abstract;jsessionid=60045DD9513137DCE1189C9F07C54B7B.f02t03.

Cover, T. M., & Thomas, J.A. (2006). *Elements of information theory* (2nd Ed.). John Wiley & Sons: New Jersey.

Dursun, Ö. (2009). *An application of information theory for dna structure*. M.Sc. Thesis, Dokuz Eylül University, İzmir.

Gündüz, A. (2010). *Entropy applications for customer satisfaction survey in information theory*. M.Sc. Thesis, Dokuz Eylül University, İzmir.

Kullback, S. (1987). The Kullback-Leibler distance. *The American Statistician*, 41 (4). Retrieved December 15, 2014, from http://www.jstor.org/discover/10.2307/2684769?sid=21105868342363&uid=2&uid=3739192&uid=4.

Laidler, K. J. (1995). Thermodynamics. *In The World of Physical Chemistry* (1st Ed.) (156-240). New York: Oxford University Press.

Leutenneger, A.L. (2000). Relative Entropy. M. Tompa, (Ed.), *Lecture Notes Biological Sequence Analysis*. (39-43). Washington: University of Washington.

Liu, Y., Liu, C. J., & Wang, D. H. (2011). Understanding atmospheric behavior in terms of entropy: a review of applications of the second law of thermodynamics to meteorology. *Entropy*, *13*, (1) 211–240.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing*, 18 (4), 351–372.

Mcalpine, M. (2002). *Principles of assessment*. Retrieved December 18, 2013, from http://www.caacentre.ac.uk/dldocs/Bluepaper1.pdf

Overton, T. (2011). *Assessing learners with special needs: an applied approach* (7th ed.). New Jersey: Prentice Hall PTR.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: an integrated approach* (1st Ed.) (15-29). New Jersey: Lawrence Erlbaum Associates.

Weiss, C. H. (1972). *Evaluation research: methods for assessing program effectiveness.* New Jersey: Prentice-Hall.

Zhou, R., Cai, R., & Tong, G. (2013). Applications of entropy in finance: a review. *Entropy*, 15, 4909-4931

APPENDICES

Appendix 1 Students' residences and scores

Student Number	Residence	Mathematics	Grammar	Social Studies	Science
1	Village	38	55	44	28
2	Village	13	55	70	12
3	Village	21	52	46	28
4	Village	23	66	31	44
5	Village	5	66	46	18
6	Village	13	54	49	5
7	Village	5	56	11	12
8	County	54	42	58	4
9	Village	22	61	34	21
10	County	35	89	65	17
11	Village	34	18	17	40
12	County	55	70	49	50
13	County	65	15	58	16
14	Village	32	27	12	45
15	Village	31	21	69	23
16	Village	24	62	37	3
17	Village	18	86	69	1
18	Village	12	34	6	15
19	County	63	14	55	21
20	Village	36	44	59	35
21	County	44	21	84	11
22	County	45	83	32	13
23	County	44	66	77	35
24	Village	18	19	55	11
25	Village	13	28	87	32
26	County	46	26	1	18
27	County	53	86	19	14
28	Village	7	15	34	21
29	County	57	34	11	7
30	County	43	66	49	65
31	Village	5	20	65	13