Spring 2024

The Distribution Of The Significance Level

Paul O. Monnu

THE DISTRIBUTION OF THE SIGNIFICANCE LEVEL

by

PAUL O. MONNU

(Under the Direction of Charles W .Champ)

ABSTRACT

Reporting the p-value is customary when conducting a test of hypothesis or significance. We give a definition of p-value. For the F-test in a one-way ANOVA and the t-tests for population means, we defined the significance level and its observed value and derived the sampling distribution. The t-test and the F-test are not without controversy. Specifically, we demonstrate that as sample size increases, the expectation of the T statistic in the t-test increases to infinity. The F statistic in the F-test is equivalent in this regard. Nonetheless, we demonstrate that the variance of these two test statistics is solely dependent on the total effect size in both scenarios.

INDEX WORDS: noncentral t-distribution, noncentral F-distribution, observed significance level, p-value

THE DISTRIBUTION OF THE SIGNIFICANCE LEVEL

by

PAUL O. MONNU

B.S, University of Lagos, Nigeria, 2016

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial

Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

THE DISTRIBUTION OF THE SIGNIFICANCE LEVEL

by

PAUL O. MONNU

Major Professor: Charles W .Champ
Committee: Andrew V. Sills
Zheni Utic

Electronic Version Approved:
May 2024

DEDICATION

To my beautiful daughter, Gianna Mojisola Monnu, who arrived into this world with the dawn of a new chapter in my life just days before the culmination of this academic journey. Your arrival has filled my heart with joy and my life with purpose. May this thesis signify my academic achievements and stand as a testament to the love and hope you have brought into my life. Here's to the beginning of our shared adventures and the endless possibilities. With all my love, now and forever.

ACKNOWLEDGMENTS

First and foremost, I extend my deepest gratitude to Dr. Charles Champ, whose expert guidance and unwavering support from the initial to the final phase shaped this thesis. His insights and dedication were crucial in bringing this work to fruition.

I am immensely thankful to my wife, Oyindamola, whose patience and unwavering support sustained me through the challenges of this academic journey. Her understanding and love provided the strength I needed to persevere.

My heartfelt thanks go to my friends, who provided both mental and psychological support throughout this process. Their encouragement and presence were invaluable.

Lastly, I give thanks to God Almighty, who saw me through this challenging but rewarding journey, making the completion of this thesis possible.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS 3

LIST OF TABLES 6

CHAPTER

1 INTRODUCTION 1

1.1 Introduction 1

1.2 What Is Statistics? 2

1.3 Ordering the Sample Space 4

1.4 Statistical Tests of Significance 5

1.5 Significance Level 7

1.6 Clinical, Practical, and Statistical Significance 8

1.7 Conclusion 10

2 MODEL, SAMPLING METHOD, AND SOME DISTRIBUTIONAL RESULTS 11

2.1 Introduction 11

2.2 Sampling Method 11

2.3 Random Sample 12

2.4 Family of Normal Distributions 14

2.5 Family of Noncentral t-Distributions 15

2.6 Family of Noncentral Chi Square Distributions 19

2.7 Family of Noncentral F-Distributions 23

2.8 Statistics Useful in Comparing Two Population Means 24

2.9 Conclusion 30

3 TEST OF SIGNIFICANCE AND TEST OF HYPOTHESIS 32

3.1 Introduction 32

3.2 Likelihood Ratio Test 32

3.3 The t-Test 35

3.4 Conclusion 41

4 ONE-WAY ANALYSIS OF VARIANCE (ANOVA) 42

4.1 Introduction 42

4.2 An Example of a One-Way Analysis of Variance 42

4.3 Conclusion 50

5 CONFIDENCE INTERVALS 51

5.1 Introduction 51

5.2 Conclusion 53

6 CONCLUSION 54

6.1 General Conclusions 54

6.2 Areas for Further Research 54

LIST OF TABLES

Table Page

3.1 Hypotheses and Significance Level 38

4.1 Power of the Test 49

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Our interest is to examine "null hypothesis significance testing procedures (NHSTP)" and the surrounding controversy. We will examine the test of significance as well as a test of hypothesis. Rozeboom (1960) states

> In this paper, I wish to examine a dogma of inferential procedure which, for psychologists at least has attained the status of religious conviction. The dogma to be scrutinized is the 'null-hypothesis significance test' orthodoxy that passing statistical judgment on a scientific hypothesis by means of experimental observation is a decision procedure wherein one rejects or accepts a null hypothesis according to whether or not the value of a sample statistic yielded by an experiment falls within a certain predetermined 'rejection region' of its possible values.

In a test of significance or a test of hypothesis, one assumes the null hypothesis and then, using a predetermined method, attempts to answer the question about how strongly the evidence in the data is against the null hypothesis. By assuming the null hypothesis is true, if the data fails to reject the null hypothesis, then one logically cannot conclude that the null hypothesis is true. In null hypothesis testing, there are only two decisions "fail to reject the null hypothesis" or "reject the null hypothesis." Demidenko (2016) states "[t]here is a growing frustration with the concept of the p-value. Besides having an ambiguous interpretation, the p-value can be made as small as desired by increasing the sample size n." We agree with Kuffner and Walker (2016) and, as we will explain, the p-value has a clear meaning. We will also give meaning to "significance" in test of significance and test of hypothesis based on a ordering of the sample space selected by the researcher. The ordering of the sample space is often pushed to the background when discussing tests of significance and test of hypothesis. The significance level is the statistic under a given model that Kuffner and Walker (2016) refer to as a "bijection" mapping of the data into the interval $(0, 1)$. Beji (1985) discusses the p-value as a random variable. We prefer to refer to the transformation of the data into the interval $(0, 1)$ as the significance level (SL). The significance level is a statistic and has a sampling distribution. The ordering on the sampling space allows one to interpret the p-value as a probability. Numerically, it is equal to the observed value of the significance level (OSL) but requires one to imagine taking a second sample of size n from the population under the assumption that the null hypothesis is true.

1.2 What Is Statistics?

As stated in Fisher (1958, p. 1), ", the science of statistics is essentially a branch of Applied Mathematics and may be regarded as mathematics applied to observational data." He goes on to state "as other mathematical studies, the same formula is equally relevant to widely different groups of subject-matter. Consequently the unity of the different applications had usually been overlooked, the more naturally because the development of the underlying mathematical theory had been much neglected."

Researchers study a process(es). Fisher (1958, p. 33) states "when a large number of individuals are measured in respect of physical dimensions, weight, color, density, etc., it is possible to describe with some accuracy the *population* of which our experience may be regarded as a sample. By this means, it may be possible to distinguish it from other populations differing in their genetic origin or in environmental circumstances." Further, Fisher (1958, p. 41) states the following about population and frequency distributions. "The idea of an infinite **population** distributed in a **frequency distribution** in respect of one or more characters is fundamental to all statistical work. From a limited experience, for example, of individuals of a species, or of the weather of a locality, we may obtain some idea of the infinite hypothetical population from which our sample is drawn, and so of the probable nature of future samples to which our conclusions are to be applied." He alludes to a process (genetic origin or environmental circumstances), describes a population as an infinite collection, and does not make a distinction between the population as a collection of individuals or as the collection of the measurements on these individuals. As Fisher described, we define the population as all the individuals the process has or could have generated. It is our belief that Fisher's use of "infinite," he meant "uncountable infinite." In statistics, a population is uncountable. On each individual in a population, a measurement X or a vector of measurements $\mathbf{X}$ is to be taken.

The collection of individuals the process has generated is a sample of size N. The integer N maybe known or unknown; its value is determined by the process. The size N of the representative sample can be thought of as a random variable. We will refer to this sample as the *representative sample* as it is representative of what the process can generate. This sample is often claimed by statisticians as to be the population. For variety of reasons, it is not feasible for the researcher to measure each individual in the representative sample. Consequently, the researcher using a sampling method selected a sample of size n, where $1 \leq n \leq N$. We will refer to this sample as the *researcher's sample*. The researcher in some cases fixes the value of n. There are cases in which the researcher allows the data to select n resulting in n being a random variable. On each of these n individuals, the researcher will obtain a measurement X or a $p \times 1$ vector of measurements $\mathbf{X}$. We denote these n measurements by $X_1, \ldots, X_n$ ($\mathbf{X}_1, \ldots, \mathbf{X}_n$). As the representative sample is representative of the population, it is the desire of the researcher to choose a sampling method that results in a sample representative of the representative sample. There does not exist a sampling method that has this property. At best the researcher can hope that the sampling method used gives him/her a good chance of selecting a researcher's sample that is representative of the representative sample and hence representative of the population. We define the sample space $\mathbf{S}$ as the collection of all possible researcher's samples.

1.3 Ordering the Sample Space

We first examine what is meant by partial and total orderings. The Cartesian product, denoted by $\mathbf{A} \times \mathbf{B}$, of set $\mathbf{A}$ with set $\mathbf{B}$ is the collection of all ordered pairs (a, b) with $a \in \mathbf{A}$ and $b \in \mathbf{B}$. Any sub-collection $\mathbf{R}$ of $\mathbf{A} \times \mathbf{B}$ is known as a "relation" from $\mathbf{A}$ to $\mathbf{B}$. If $\mathbf{A} = \mathbf{B}$, then we say the relation is on set $\mathbf{A}$. A relation $\mathbf{R}$ on a set $\mathbf{A}$ is said to have the *reflexive property* if for all $x \in \mathbf{A}$ the ordered pair (x, x) is in $\mathbf{R}$. If for each ordered pair (x, y) in $\mathbf{R}$ the ordered pair (y, x) is also in $\mathbf{R}$, then we say the relation $\mathbf{R}$ has the *symmetric property*. A relation $\mathbf{R}$ has the transitive property if the ordered pair (x, y) is in $\mathbf{R}$ and the ordered pair (y, z) is in $\mathbf{R}$, then the ordered pair (x, z) is in $\mathbf{R}$. The relation $\mathbf{R}$ is said to be *antisymmetric* if (x, y) and (y, x) in $\mathbf{R}$ implies that $x = y$.

A relation $\mathbf{R}$ is said to be a *partial ordering* on a set $\mathbf{A}$ if it is reflexive, transitive, and anti-symmetric. It is a *total ordering* if it also has the trichotomy property, that is, for all $x, y \in \mathbf{A}$ either (x, y) or (y, x) is in $\mathbf{R}$. It will be convenient to write "$x \succeq_{\mathbf{R}} y$" to mean (x, y) is in $\mathbf{R}$ for $x, y \in \mathbf{A}$. Typically, the ordering is placed on the sample space $\mathbf{S}$ is a total ordering. In the case the test statistic has a discrete distribution, then the ordering is a partial ordering.

Kempthorne and Folks (1971) state that at least a partial ordering should be placed on the sample space in a test of significance or test of hypothesis. An ordering on the sample space $\mathbf{S}$ is typically related to an ordering that is placed on the test statistic. If the ordering placed on the statistic is partial resp. (total) ordering, then the ordering on the sample space is a partial resp. (total) ordering.

1.4 Statistical Tests of Significance

DeGroot (1975) outlined a test of hypothesis. Suppose that a researcher has an interest in the distribution of measurement X (which may be a vector of measurements) on the individual in a population where a possible vector of parameters θ associated with the distribution of X characterize the process that is generating individuals. The collection of possible values of θ is denoted by $\mathbf{\Omega}$ which is often referred to as the parameter space. To gain information about the process parameter θ, the researcher will take a sample of n individuals from the collection of individuals the process has generated. We denote the X measurements on these n individuals by $(X_1, \ldots, X_n)$ and treat the sample as an n-tuple of real numbers. The collection of all possible n-tuples will be referred to as the sample space $\mathbf{S}$. A researcher believes (hypothesizes) that $\theta \in \mathbf{\Omega}_a \subset \mathbf{\Omega}$. Opposite in truth value to the researcher's hypothesis is the hypothesis $\theta \in \mathbf{\Omega}_0$, where $\mathbf{\Omega}_0 \cup \mathbf{\Omega}_a = \mathbf{\Omega}$ and $\mathbf{\Omega}_0 \cap \mathbf{\Omega}_a = \emptyset$ or $\mathbf{\Omega}_0 = \mathbf{\Omega} - \mathbf{\Omega}_a$. A test of hypothesis assumes the hypothesis $\theta \in \mathbf{\Omega}_0$, opposite in truth value to the researcher's hypothesis, is true. This hypothesis is often referred to as the null hypothesis and labeled as H_0. The researcher's hypothesis is often referred to as the alternative to the null hypothesis and labeled as H_a. For convenience, we write $H_0 : \theta \in \mathbf{\Omega}_0$ and $H_a : \theta \in \mathbf{\Omega}_a$. An ordering is then placed on the sample space by the researcher such that for any two samples, it can be determined which is at least as contradictory as the other to the null hypothesis. A decision rule for the test rejects the null hypothesis in favor of the alternative hypothesis if $(X_1, \ldots, X_n) \in \mathbf{C} \subset \mathbf{S}$, where $\mathbf{C}$ is selected by the researcher otherwise, based on the data, one fails to reject the null hypothesis. DeGroot (1975) defined the power function π by

$$\pi(\theta) = P\left[(X_1, \ldots, X_n) \in \mathbf{C} \,|\theta\right],$$

for $\theta \in \mathbf{\Omega}$. The size α of the test or the level of significance of the test is

$$\alpha = \max_{\theta \in \mathbf{\Omega}_0} \pi(\theta).$$

The probability of a Type II error β is $\beta = \beta(\theta) = 1 - P\left[(X_1, \ldots, X_n) \in \mathbf{C} \,|\theta \in \mathbf{\Omega}_a\right]$. DeGroot (1975) states “In many problems, a statistician will specify an upper bound $0 < \alpha_0 < 1$ and will consider only tests for which

$$\max_{\theta \in \mathbf{\Omega}_0} \pi(\theta) \leq \alpha_0$$

for every value of $\theta \in \mathbf{\Omega}_0$.” among all procedures for which $\alpha \leq \alpha_0$, δ is a minimum.” The value α_0 is referred to as the size of the test or level of significance of the test. Once the value of α_0 is specified by the researcher, a test procedure is chosen to minimize the Type II error $\beta(\theta_a)$ for a particular value θ_a of θ that makes the alternative hypothesis true. Note we are using the vector θ of parameters while DeGroot (1975) uses the single parameter θ.

A test of significance (a null hypothesis but no alternative hypothesis, see Fisher (1958)) and a test of hypothesis (both a null and an alternative hypothesis) are based on an ordering of the sample space which the researcher provides. An ordering gives meaning to “significance.” The ordering is selected by the researcher. The ordering allows one to compare two samples with respect to evidence in the samples against the null hypothesis.

One may have seen the phrase “as extreme or more extreme” in an introductory statistic course. This phrase implies an ordering on the sample space. One could now replace “extreme” in this phrase with “significant” giving significance meaning.

Beji (1985) stated that “[t]ests of hypotheses are usually used either as inference procedures or as decision procedures.” We consider both tests of significance and tests of hypothesis which are used by researchers to make inferences and decisions about a process. Rozeboom (1960) discusses the “fallacy” of null hypothesis significance tests. In his paper, the author mentions that induction is a special case of statistical inference. Scientists use inductive statistics to make evidence-based decisions based on empirical/experimental results based on probability theory. The statistical induction principle is used to make decisions using experimental results. Wasserstein and Lazar (2016) discusses the American Statistical Association’s stance on tests of significance and tests of hypothesis. Wilkinson (1999) discusses guidelines and explanations for statistical methods in psychology journals.

1.5 Significance Level

Kempthorne and Folks (1971) give the following definition of significance level (SL) on page 222.

“Definition 9.1. Let the possible data sets under a probability model M_0 be denoted by $\{D_i\}$. A test of significance consists of: (1) the arrangements of the possible data sets, D_i, as a partially ordered set; if D_i does not occur in the partial ordering after D_j, we write $D_i \gg D_j$; and (2) attaching to the observed data sets D_o the number

$$SL\left(D_o;M_0\right) = \sum_{D \gg D_o} P\left(D, M_0\right)$$

which is called the significance level of the data set D_o with regard to the model M_0 for the partial ordering chosen."

This definition of SL seems to suggest that the SL is a probability. We define the significance level (SL) as a transformation of the measurements on the individuals in the sample into the interval $(0, 1)$. Under our general definition of the significance level (SL), the SL is a statistic. Their definition of SL could be viewed as the p-value. We refer to the observed value of the significance level as the observed significance level (OSL).

In the case of the one-sample t-test, we can observe the the SL. Our interest is to examine the distribution of the SL for a variety of statistical tests of significance and hypothesis. Although the support of the distribution of the SL is the interval $(0, 1)$, it is not a probability, nor is its observed value, the statistic the OSL. Based on the given ordering of the sample space, the conditional probability of selecting another sample from the sample space that is at least as contradictory to the null hypothesis, assuming the null hypothesis is true than the sample that has been observed is referred to as the p-value. The p-value is what Kempthorne and Folks (1971) define as the significance level. The p-value is a probability. In many tests of hypotheses, the p-value is numerically equal to the OSL. While in these cases, the OSL and the p-value have the same magnitude, they clearly do not have the same meaning. In research studies, the p-value should be abandoned and replaced with the statistic of the observed significance level (OSL). Note that the p-value as defined, is based on a sample that will never be observed.

1.6 Clinical, Practical, and Statistical Significance

The American Psychological Association Dictionary of psychology states that "practical significance is the extent to which a study result has meaningful applications in real-world settings. An experimental result may lack statistical significance or show a small effect size and yet potentially be important nonetheless. For example, consider a study showing that the consumption of baby aspirin helps prevent heart attacks. Even if the effect is small, the finding may be of practical significance if it saves lives over time. practical significance is also called substantive significance. See also clinical significance; psychological significance." The dictionary also defines the effect size as "any of various measures of the magnitude or meaningfulness of a relationship between two variables. For example, Cohen's d shows the number of standard deviation units between two means. See Cohen (1997). Often, effect sizes are interpreted as indicating the practical significance of a research finding. Additionally, in meta-analyses they allow for the computation of summary statistics that apply to all the studies considered as a whole. Further, the dictionary defines "statistical significance as the degree to which a research outcome cannot reasonably be attributed to the operation of chance or random factors." It is determined during significance testing and given by a critical p value, which is the probability of obtaining the observed data if the null hypothesis (i.e., of no significant relationship between variables) were true. Significance generally is a function of sample size - the larger the sample, the less likely it is that one's findings will have occurred by chance. Determining statistical significance, practical significance, and effect size falls to the researcher. Peterson (2008) argued that clinical significance and practical significance are not the same thing. A common method of calculating clinical significance is given in Jacobson and Truax (1984,1991). Harrington, et al. (2019) discuss statistical reporting in *The New England Journal of Medicine*. Kazdin (1999) discussed the meanings and measurement of clinical significance. Wilkerson (1999) discussed guidelines and gave explanations of statistical methods in psychology journals.

1.7 Conclusion

Our intent is to examine the sampling distribution of the significance level (SL) for various tests of hypotheses. The significance level is related to the level of significance or size of the test. Beji (1985) showed how to obtain the distribution of the SL. We extend his results by looking at the non-central distribution of the significance level.

CHAPTER 2

MODEL, SAMPLING METHOD, AND SOME DISTRIBUTIONAL RESULTS

2.1 Introduction

Models are used for the distribution of a measurement X or a vector of measurements $\mathbf{X}$. The most commonly used model is the (multivariate) Normal distribution. Further, it is typically assumed that the researcher's sample is a random sample. The researcher should argue that the sample is a random sample based on the sampling method used. We will examine sampling method, random sample, the family of Normal distributions, and the distribution of families of statistics.

2.2 Sampling Method

Researhers study a process(es). The collection of all individuals the process has (actual) or could have (conceptual) generated is the population. The collection of conceptual individuals is an uncountable collection and hence the population is an uncountable collection. We will refer to the collection of individuals the process has generated as the representative sample of size N. Using a sampling method the researcher will select a subcollection of size n from the representative sample with $1 \leq n \leq N$. This sample is referred to as the researcher's sample. The integer $N \geq 1$ is determined by the process. The value of N may be known but in general should be considered as a random variable. The researcher's sample size n is selected by the researcher and is considered as fixed in what is to follow. Ideally, we contend that the researcher is interested in using a sampling method that results in a researcher's sample that is "representative" of the representative sample and hence representative of the population. It is not practical to have such a hope. Sampling methods are used that give the researcher a good chance of selecting a sample that is representative of the representative sample and hence the population. One such sample method is simple random sampling. This method gives each sample of size n the same chance of being selected from the representative sample of size N. Thus each sample has one in $\binom{N}{n}$ chance of being selected.

2.3 Random Sample

A random sample is defined in terms of the X ($\mathbf{X}$) measurements $X_1, \ldots, X_n$ ($\mathbf{X}_1, \ldots, \mathbf{X}_n$) on the n individuals in the researcher's sample. The n measurements are stochastically independent and have the same distribution. Mathematically, when X is a continuous random variable, we have

$$f_{X_1,\ldots,X_n}\left(x_1, \ldots, x_n\right) = \prod\nolimits_{i=1}^{n} f_X\left(x_i\right),$$

where $f_{X_1,\ldots,X_n}\left(x_1, \ldots, x_n\right)$ is the joint probability density function of $X_1, \ldots, X_n$ and $f_X\left(x\right)$ is the probability density function that describes the distribuiton of X. When X is a discrete random variable, then

$$P\left(X_1 = x_1, \ldots, X_n = x_n\right) = \prod\nolimits_{i=1}^{n} P\left(X = x_i\right),$$

where $P\left(X_1 = x_1, \ldots, X_n = x_n\right)$ is the joint probability mass function of $X_1, \ldots, X_n$ and $P\left(X = x\right)$ is the probability mass function describing the distribution of X. We can similarly define the representative sample as a random sample. A simple random sampling method is a sampling method that gives every researcher's sample of size n to be taken from the representative sample of size N the same chance of being selected.

Theorem: If the representative sample $X_1, \ldots, X_N$ is a random sample with common distribution that of X and the researcher's sample is a simple random sample, then the researcher's sample is a random sample with the same common distribution.

Proof. Suppose the representative sample is a random sample and the researcher's sample is a simple random sample. Then,

$$f_{X_1,\ldots,X_N}\left(x_1,\ldots,x_N\left|N\right.\right)=\prod\nolimits_{i=1}^{N}f_X\left(x_i\left|N\right.\right)\text{ or}$$
$$P\left(X_1=x_1,\ldots,X_N=x_N\left|N\right.\right)=\prod\nolimits_{i=1}^{N}P\left(X=x_i\left|N\right.\right).$$

Suppose that the X values of the researcher's sample are denoted by $X_{i_1},\ldots,X_{i_n}$. Let

$$\mathbf{X}=\left\{X_1,\ldots,X_N\left|N\right.\right\},\ \mathbf{X}_i=\left\{X_{i_1},\ldots,X_{i_n}\left|N\right.\right\},\text{ and }\mathbf{Y}=\mathbf{X}-\mathbf{X}_i.$$

It follows that

$$\begin{aligned}
f_{X_{i_1},\ldots,X_{i_n}}\left(x_{i_1},\ldots,x_{i_n}\right)&=\int_{\mathbf{Y}}f_{X_1,\ldots,X_N}\left(x_1,\ldots,x_N\left|N\right.\right)d\mathbf{y}\\
&=\int_{\mathbf{Y}}\prod\nolimits_{i=1}^{N}f_X\left(x_i\left|N\right.\right)d\mathbf{y}\\
&=\prod\nolimits_{i=1}^{N}\int_{\mathbf{Y}}f_X\left(x_i\left|N\right.\right)d\mathbf{y}\\
&=\prod\nolimits_{j=1}^{n}f_X\left(x_{i_j}\right)\text{ or}\\
P\left(X_{i_1}=x_{i_1},\ldots,X_{i_n}=X_{i_n}\right)&=\sum\nolimits_{\mathbf{Y}}P\left(X_1=x_1,\ldots,X_N=x_N\left|N\right.\right)\\
&=\sum\nolimits_{\mathbf{Y}}\prod\nolimits_{i=1}^{N}P\left(X=x_i\left|N\right.\right)\\
&=\prod\nolimits_{i=1}^{N}\sum\nolimits_{\mathbf{Y}}P\left(X=x_i\left|N\right.\right)\\
&=\prod\nolimits_{j=1}^{n}P\left(X_{i_j}=x_{i_j}\right).
\end{aligned}$$

Hence, the random variables $X_{i_1},\ldots,X_{i_n}$ are independent and have a common distribution that of X and therefore the researcher's sample is a random sample.■

2.4 FAMILY OF NORMAL DISTRIBUTIONS

A non-degenerate member of the family of Normal is described by

$$f\left(x\,|\mu,\sigma\right)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

where μ is the population mean and σ is the population standard deviation. It is not difficult to show (1) that for all real numbers x, $f\left(x\,|\mu,\sigma\right)>0$ and

$$\int_{-\infty}^{\infty}f\left(x\,|\mu,\sigma\right)dx=1.$$

When used as a model, the function $f\left(x\,|\mu,\sigma\right)$ describes the distribution of a measurement X. Hence, the function $f\left(x\,|\mu,\sigma\right)$ is a probability density function. The cumulative distribution function is

$$F\left(x\,|\mu,\sigma\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}dy.$$

The random variable X has a Normal distribution with mean μ and standard deviation σ is expressed more compactly as $X\sim N\left(\mu,\sigma^2\right)$. Assuming the sample is a random sample with a common $N\left(\mu,\sigma^2\right)$ distribution, will be known as the independent Normal model. The transformation $Z=\left(X-\mu\right)/\sigma$ assuming $X\sim N\left(\mu,\sigma^2\right)$ is a random variable that has a Normal distribution with mean 0 and variance 1 (a standard Normal distribution).

A statistic is a number that characterizes the sample that can be observed once the measurements on the individuals in the sample are obtained. Two statistics that commonly appear in statistical inference are the sample mean $\overline{X}$ and sample variance S^2. For the researcher's sample with measurements $X_1,\ldots,X_n$, they are defined by

$$\overline{X}=\frac{1}{n}\sum\nolimits_{i=1}^{n}X_i\text{ and }S^2=\frac{1}{n-1}\sum\nolimits_{i=1}^{n}\left(X_i-\overline{X}\right)^2.$$

It can be shown under the assumption $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$ that $\overline{X}$ and S^2 are stochastically independent. Similar results hold for the mean and variance of the representative sample conditioned on the sample size N. A well known theorem that can be found in most books on mathematical statistics states the sample mean and sample variance are (stochastically) independent if the sample is a random sample with a common $N(\mu, \sigma^2)$ distribution. Further, one can show that if the sample is a random sample with a common $N(\mu, \sigma^2)$ distribution, then

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ and } \frac{(n-1)\, S^2}{\sigma^2} \sim \chi^2_{n-1,0},$$

where $\chi^2_{n-1,0}$ is a random variable that has a central Chi Square distribution with $n-1$ degrees of freedom. See Bain and Engelhardt (1992) for a discussion of these results and the family of central Chi Square distribuitons. We will assume in what follows that $X_1, \ldots, X_n$ iid $N(\mu, \sigma^2)$. We refer to this model as the independent Normal model.

2.5 Family of Noncentral t-Distributions

The members of the family of noncentral t-distributions are indexed by the parameters ν and θ. A random variable T that can be expresses as

$$T = \frac{Z + \theta}{\sqrt{W/\nu}},$$

where θ is a real number, $Z \sim N(0, 1)$, and $W \sim \chi^2_\nu$ (Chi Square random variable with ν degrees of freedom) which is independent of Z is said to have a noncentral t-distribution with ν degrees of freedom and noncentrality parameter θ. If $\theta = 0$, the distribution of T is referred to as a central t-distribution with ν degrees of freedom. The mean and variance of the distribution of T are

$$\mu_T = \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)}\theta \text{ and}$$
$$\sigma_T^2 = \left[\nu\left(1+\theta^2\right)-\theta^2\right]\left(\frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{\nu}{2}\right)}\right)^2.$$

For the case in which

$$T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}} = \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}+\frac{\mu-\mu_0}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/\left(n-1\right)}}$$
$$= \frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}+\sqrt{n}\frac{\mu-\mu_0}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/\left(n-1\right)}} = \frac{Z+\sqrt{n}\delta}{\sqrt{W/\left(n-1\right)}},$$

we have under our model that $Z = \left(\overline{X}-\mu\right)/\left(\sigma/\sqrt{n}\right) \sim N\left(0,1\right)$, $W = \left(n-1\right)S^2/\sigma^2 \sim \chi^2_{n-1}$, $\delta = \left(\mu-\mu_0\right)/\sigma$, and $\theta = \sqrt{n}\delta$ with Z and W independent. Hence, the statistic T has a noncentral t-distribution with $n-1$ degrees of freedom and noncentrality parameter $\sqrt{n}\delta$. The mean μ_T and the variance σ_T^2 of the distribution of T are

$$\mu_T = \frac{\sqrt{n\left(n-1\right)}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)}\delta \text{ and}$$
$$\sigma_T^2 = \frac{\left(n-1\right)\left(1+n\delta^2\right)}{n-3} - \frac{n-1}{2}\frac{\Gamma^2\left(\frac{n-2}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)}n\delta^2.$$

One can show that

$$\lim_{n\to\infty}\mu_T = \infty \text{ and } \lim_{n\to\infty}\sigma_T^2 = \frac{1}{2}\delta^2+1.$$

The statistic T can be expressed as

$$T = \frac{Z+\sqrt{n}\delta}{\sqrt{W/\left(n-1\right)}},$$

where $Z \sim N(0,1)$ and $W \sim \chi^2_{n-1}$ are independent under our model. It follows that

$$\begin{aligned}
E(T) &= \sqrt{n-1}E\left(Z+\sqrt{n}\delta\right)E\left(W^{-1/2}\right)\\
&= \sqrt{n-1}\sqrt{n}\delta\int_0^\infty w^{-1/2}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}w^{(n-1)/2-1}e^{-w/2}dw\\
&= \sqrt{n-1}\sqrt{n}\delta\frac{\Gamma\left(\frac{n-2}{2}\right)2^{(n-2)/2}}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}\int_0^\infty\frac{1}{\Gamma\left(\frac{n-2}{2}\right)2^{(n-2)/2}}w^{(n-2)/2-1}e^{-w/2}dw\\
&= \frac{\sqrt{n-1}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)}\sqrt{n}\delta.
\end{aligned}$$

Further, we have

$$\begin{aligned}
E\left(T^2\right) &= (n-1)E\left(Z^2+2\sqrt{n}\delta Z+n\delta^2\right)E\left(W^{-1}\right)\\
&= (n-1)\left(1+n\delta^2\right)\int_0^\infty w^{-1}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}w^{(n-1)/2-1}e^{-w/2}dw\\
&= (n-1)\left(1+n\delta^2\right)\frac{\Gamma\left(\frac{n-3}{2}\right)2^{(n-3)/2}}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}\int_0^\infty\frac{1}{\Gamma\left(\frac{n-3}{2}\right)2^{(n-3)/2}}w^{(n-3)/2-1}e^{-w/2}dw\\
&= \frac{(n-1)\left(1+n\delta^2\right)}{n-3}.
\end{aligned}$$

Hence,

$$V(T) = \frac{(n-1)\left(1+n\delta^2\right)}{n-3} - \frac{n-1}{2}\frac{\Gamma^2\left(\frac{n-2}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)}n\delta^2.$$

The mean and variance of the distribution of T under the alternative hypothesis are

$$\begin{aligned}
E(T) &= \frac{\sqrt{n-1}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)}\sqrt{n}\delta \text{ and}\\
V(T) &= \frac{(n-1)\left(1+n\delta^2\right)}{n-3} - \frac{n-1}{2}\frac{\Gamma^2\left(\frac{n-2}{2}\right)}{\Gamma^2\left(\frac{n-1}{2}\right)}n\delta^2.
\end{aligned}$$

Proof. We make the one-to-one transformation

$$t = \frac{z+\theta}{\sqrt{w/\nu}} \text{ and } u = w/\nu.$$

The inverse of this transformation is

$$z = tu^{1/2} - \theta \text{ and } w = \nu u$$

with Jacobian $\nu u^{1/2}$. The joint probability density function of T and U is

$$\begin{aligned}
f_{T,U}(t,u) &= f_Z\left(tu^{1/2}-\theta\right) f_W(\nu u)\,\nu u^{1/2} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\left(tu^{1/2}-\theta\right)^2/2} \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}} (\nu u)^{\nu/2-1} e^{-\nu u/2} \nu u^{1/2} \\
&= \frac{\nu^{\nu/2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)2^{(\nu+1)/2}} e^{-\left(t^2u-2\theta tu^{1/2}+\theta^2\right)/2} u^{(\nu+1)/2-1} e^{-\nu u/2} \\
&= \frac{\nu^{\nu/2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)2^{(\nu+1)/2}} u^{(\nu+1)/2-1} e^{-\left(\nu+t^2\right)u/2} e^{\theta tu^{1/2}/2} e^{-\theta^2/2}.
\end{aligned}$$

Observe that

$$e^{\theta tu^{1/2}/2} = \textstyle\sum_{j=0}^{\infty} \dfrac{(\theta t)^j u^{j/2}}{2^j j!} = 1 + \textstyle\sum_{j=1}^{\infty} \dfrac{(\theta t)^j u^{j/2}}{2^j j!}.$$

Thus, the joint distribution of T and U is

$$\begin{aligned}
f_{T,U}(t,u) &= \frac{\nu^{\nu/2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)2^{(\nu+1)/2}} u^{(\nu+1)/2-1} e^{-\left(\nu+t^2\right)u/2} \\
&\quad \times \left(1 + \textstyle\sum_{j=1}^{\infty} \dfrac{(\theta t)^j u^{j/2}}{2^j j!}\right) e^{-\theta^2/2} \\
&= \frac{e^{-\theta^2/2}\nu^{\nu/2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)2^{(\nu+1)/2}\left(\nu+t^2\right)^{(\nu+j+1)/2}} \\
&\quad \times \left(\left(\nu+t^2\right)u\right)^{(\nu+1)/2-1} e^{-\left(\nu+t^2\right)u/2}\left(\nu+t^2\right) \\
&\quad + \textstyle\sum_{j=1}^{\infty} \dfrac{e^{-\theta^2/2}(\theta t)^j \nu^{\nu/2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)2^{(\nu+j+1)/2}\left(\nu+t^2\right)^{(\nu+j+1)/2} j!} \\
&\quad \times \left(\left(\nu+t^2\right)u\right)^{(\nu+j+1)/2-1} e^{-\left(\nu+t^2\right)u/2}\left(\nu+t^2\right).
\end{aligned}$$

To determine the probability density function $f_T(t\,|\nu,\theta)$ describing the marginal distribution of T, we integrate $f_{T,U}(t,u)$ with respect to u over the interval $(0,\infty)$. It follows that

$$
\begin{aligned}
f_T\left(t\,|\nu,\theta\right) &= \int_0^\infty f_{T,U}\left(t,u\right)du \\
&= \frac{e^{-\theta^2/2}\nu^{\nu/2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)2^{(\nu+1)/2}\left(\nu+t^2\right)^{(\nu+j+1)/2}} \\
&\times \int_0^\infty \left(\left(\nu+t^2\right)u\right)^{(\nu+1)/2-1}e^{-\left(\nu+t^2\right)u/2}\left(\nu+t^2\right)du \\
&+ \textstyle\sum_{j=1}^\infty \frac{e^{-\theta^2/2}\left(\theta t\right)^j\nu^{\nu/2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)2^{(\nu+j+1)/2}\left(\nu+t^2\right)^{(\nu+j+1)/2}j!} \\
&\times \int_0^\infty \left(\left(\nu+t^2\right)u\right)^{(\nu+j+1)/2-1}e^{-\left(\nu+t^2\right)u/2}\left(\nu+t^2\right)du.
\end{aligned}
$$

Noting that

$$
\int_0^\infty \left(\left(\nu+t^2\right)u\right)^{(\nu+j+1)/2-1}e^{-\left(\nu+t^2\right)u/2}\left(\nu+t^2\right)du = \Gamma\left(\frac{\nu+j+1}{2}\right)2^{(\nu+j+1)/2}.
$$

We can express $f_T\left(t\,|\nu,\theta\right)$ as

$$
\begin{aligned}
f_T\left(t\,|\nu,\theta\right) &= \frac{e^{-\theta^2/2}\nu^{\nu/2}\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)\left(\nu+t^2\right)^{(\nu+1)/2}} + \textstyle\sum_{j=1}^\infty \frac{e^{-\theta^2/2}\left(\theta t\right)^j\nu^{\nu/2}\Gamma\left(\frac{\nu+j+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)\left(\nu+t^2\right)^{(\nu+j+1)/2}j!} \\
&= \frac{e^{-\theta^2/2}\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{t^2}{\nu}\right)^{(\nu+1)/2}} + \textstyle\sum_{j=1}^\infty \frac{e^{-\theta^2/2}\left(\theta t\right)^j\Gamma\left(\frac{\nu+j+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)\left(1+\frac{t^2}{\nu}\right)^{(\nu+j+1)/2}j!}.
\end{aligned}
$$

2.6 Family of Noncentral Chi Square Distributions

The members of the family of noncentral Chi Square distributions are indexed by the parameters ν and θ. The probability density function $f_W\left(w\,|\nu\right)$ of a random variable W that has a central Chi Square distribution with ν degrees of freedom can be expressed as

$$
f_W\left(w\,|\nu\right) = \frac{1}{\Gamma\left(\frac{\nu}{2}\right)2^{\nu/2}}w^{\nu/2-1}e^{-w/2}I_{(0,\infty)}\left(w\right),
$$

where $I_{(0,\infty)}\left(w\right) = 1$ if $w \in (0,\infty)$ and 0 otherwise. It is well known that if $Z_1,\ldots,Z_k$ are independent with a common $N\left(0,1\right)$ distribution, then

$$
\textstyle\sum_{i=1}^k Z_i^2 \sim \chi_k^2,
$$

where χ_k^2 is a random variable that has a Chi Square distribution with k degrees of freedom. For any nonzero real number θ, consider the random variable W defined as

$$W = (Z_1 + \theta)^2 + \sum_{i=2}^{k} Z_i^2,$$

where $Z_1, \ldots, Z_k$ are independent with a common $N(0,1)$ distribution. More generally, we define W as

$$W = \sum_{i=1}^{k} (Z_i^* + \tau_i)^2 = (\mathbf{Z}^* + \tau)^{\mathbf{T}} (\mathbf{Z}^* + \tau),$$

where $\tau_1, \ldots, \tau_k$ are real numbers, $\mathbf{Z}^* = [Z_1^*, \ldots, Z_k^*]^{\mathbf{T}}$, $\tau = [\tau_1, \ldots, \tau_k]^{\mathbf{T}}$, and $Z_1^*, \ldots, Z_k^*$ iid $N(0,1)$. Consider the $k \times k$ matrix $\mathbf{B}$ defined by

$$\mathbf{B} = \frac{1}{\theta(\tau_1 + \theta)} (\tau + \theta\mathbf{e}_1)(\tau + \theta\mathbf{e}_1)^{\mathbf{T}} - \mathbf{I},$$

where τ_1 is the first component of the vector τ, $\theta^2 = \tau^{\mathbf{T}}\tau$, and $\mathbf{e}_1$ is a $k \times 1$ vector with first coordinate one and the remaining coordinates zero. One can show that

$$\mathbf{B}\tau = \theta\mathbf{e}_1.$$

Thus, we have

$$\begin{aligned} W &= (\mathbf{Z}^* + \tau)^{\mathbf{T}} \left(\mathbf{B}^{\mathbf{T}}\mathbf{B}\right) (\mathbf{Z}^* + \tau) = (\mathbf{B}\mathbf{Z}^* + \mathbf{B}\tau)^{\mathbf{T}} (\mathbf{B}\mathbf{Z}^* + \mathbf{B}\tau) \\ &= (\mathbf{Z} + \theta\mathbf{e}_1)^{\mathbf{T}} (\mathbf{Z} + \theta\mathbf{e}_1) = (Z_1 + \theta)^2 + \sum_{i=2}^{k} Z_i^2 \\ &= U_1 + U_2, \end{aligned}$$

where $\mathbf{Z} = \mathbf{B}\mathbf{Z}^* = [Z_1, \ldots, Z_k]^{\mathbf{T}} \sim N_k(\mathbf{0}, \mathbf{I})$, $U_1 = (Z_1 + \theta)^2$, and $U_2 = \sum_{i=2}^{k} Z_i^2$. It can be shown that

$$U_2 = \sum_{i=2}^{k} Z_i^2 \sim \chi_{k-1}^2.$$

The distribution of $U_1 = (Z_1 + \theta)^2$ is determined as follows.

$$\begin{aligned} F_{U_1}(q) = F_{(Z_1+\theta)^2}(q) &= P\left[(Z_1+\theta)^2 \le q\right] I_{(0,\infty)}(q) \\ &= P\left[-\sqrt{q} - \theta \le Z_1 \le \sqrt{q} - \theta\right] I_{(0,\infty)}(q) \\ &= \left[\Phi\left(\sqrt{q} - \theta\right) - \Phi\left(-\sqrt{q} - \theta\right)\right] I_{(0,\infty)}(q), \end{aligned}$$

where $\Phi(z)$ is the cumulative distribution function of a standard Normal distribution. It follows that

$$\begin{aligned} f_{(Z_1+\theta)^2}(q) &= \frac{d}{dq} F_{(Z_1+\theta)^2}(q) \\ &= \left[\frac{1}{2} q^{-1/2} \phi\left(\sqrt{q} - \theta\right) + \frac{1}{2} q^{-1/2} \phi\left(-\sqrt{q} - \theta\right)\right] I_{(0,\infty)}(q) \\ &= \frac{1}{2} q^{-1/2} \left[\phi\left(\sqrt{q} - \theta\right) + \phi\left(-\sqrt{q} - \theta\right)\right] I_{(0,\infty)}(q), \end{aligned}$$

where $\phi(z)$ is the probability density function of a standard Normal distribution. We now have

$$\begin{aligned} f_{(Z_1+\theta)^2}(q) &= \frac{1}{2} q^{-1/2} \left[\frac{1}{\sqrt{2\pi}} e^{-\left(\sqrt{q}-\theta\right)^2/2} + \frac{1}{\sqrt{2\pi}} e^{-\left(-\sqrt{q}-\theta\right)^2/2}\right] I_{(0,\infty)}(q) \\ &= \frac{1}{2\sqrt{2\pi}} q^{-1/2} \left[e^{-\frac{1}{2}\left(q - 2\theta\sqrt{q} + \theta^2\right)} + e^{-\frac{1}{2}\left(q + 2\theta\sqrt{q} + \theta^2\right)}\right] I_{(0,\infty)}(q) \\ &= \frac{e^{-\theta^2/2}}{2\sqrt{2\pi}} q^{-1/2} e^{-q/2} \left(e^{\theta\sqrt{q}} + e^{-\theta\sqrt{q}}\right) I_{(0,\infty)}(q). \end{aligned}$$

Observe that

$$\begin{aligned} e^{\theta\sqrt{q}} + e^{-\theta\sqrt{q}} &= \sum_{j=0}^{\infty} \frac{\theta^j q^{j/2}}{j!} + \sum_{j=0}^{\infty} \frac{(-1)^j \theta^j q^{j/2}}{j!} \\ &= 2 \sum_{j=0}^{\infty} \frac{\theta^{2j} q^j}{(2j)!}. \end{aligned}$$

It follows that

$$\begin{aligned} f_{U_1}(q) &= \frac{e^{-\theta^2/2}}{2\sqrt{2\pi}} q^{-1/2} e^{-q/2} \left(2 \sum\nolimits_{j=0}^{\infty} \frac{\theta^{2j} q^j}{(2j)!} \right) I_{(0,\infty)}(q) \\ &= e^{-\theta^2/2} \sum\nolimits_{j=0}^{\infty} \frac{\theta^{2j}}{\Gamma\left(\frac{1}{2}\right) 2^{1/2} (2j)!} q^{(2j+1)/2-1} e^{-q/2} I_{(0,\infty)}(q) \\ &= e^{-\theta^2/2} \sum\nolimits_{j=0}^{\infty} \frac{\theta^{2j} \Gamma\left(\frac{2j+1}{2}\right) 2^j}{\Gamma\left(\frac{1}{2}\right)(2j)!} \frac{1}{\Gamma\left(\frac{2j+1}{2}\right) 2^{(2j+1)/2}} q^{(2j+1)/2-1} e^{-q/2} I_{(0,\infty)}(q) \\ &= e^{-\theta^2/2} \sum\nolimits_{j=0}^{\infty} \frac{\theta^{2j} \Gamma\left(\frac{2j+1}{2}\right) 2^j}{\Gamma\left(\frac{1}{2}\right)(2j)!} f_{\chi^2_{2j+1}}(q) I_{(0,\infty)}(q). \end{aligned}$$

Consider the transformation

$$W = U_1 + U_2 \text{ and } U = U_2.$$

The inverse transformation is

$$U_1 = W - U \text{ and } U_2 = U$$

with Jacobian $J = 1$. It follows that the joint probability distribution function describing the joint distribution of W and U is

$$f_{W,U}(w,u) = f_{U_1}(w-u) f_{U_2}(u) |J| = f_{U_1}(w-u) f_{U_2}(u).$$

Hence,

$$\begin{aligned} f_{W,U}(w,u) &= e^{-\theta^2/2} \sum\nolimits_{j=0}^{\infty} \frac{\theta^{2j} \Gamma\left(\frac{2j+1}{2}\right) 2^j}{\Gamma\left(\frac{1}{2}\right)(2j)!} f_{\chi^2_{2j+1}}(w-u) \\ &\quad \times \frac{1}{\Gamma\left(\frac{k-1}{2}\right) 2^{(k-1)/2}} u^{(k-1)/2-1} e^{-u/2} I_{(0,\infty)}(u) I_{(0,\infty)}(w-u) \\ &= e^{-\theta^2/2} \sum\nolimits_{j=0}^{\infty} \frac{\theta^{2j} \Gamma\left(\frac{2j+1}{2}\right) 2^{2j+1}}{\Gamma\left(\frac{1}{2}\right)(2j)!} \frac{1}{\Gamma\left(\frac{2j+1}{2}\right) 2^{2j+1}} (w-u)^{2j+1-1} \\ &\quad \times \frac{1}{\Gamma\left(\frac{k-1}{2}\right) 2^{(k-1)/2}} u^{(k-1)/2-1} e^{-u/2} I_{(0,\infty)}(u) I_{(0,\infty)}(w-u). \end{aligned}$$

Therefore,

$$f_W(w) = \int_0^\infty f_{W,U}(w,u)\,du = e^{-\theta^2/2}\sum\nolimits_{j=0}^{\infty} \frac{\theta^{2j}\Gamma\left(\frac{2j+1}{2}\right)2^j}{\Gamma\left(\frac{1}{2}\right)(2j)!} f_{\chi^2_{2j+1}}(w-u)\, I_{(0,\infty)}(w-u)$$
$$\times \frac{1}{\Gamma\left(\frac{k-1}{2}\right)2^{(k-1)/2}} u^{(k-1)/2-1} e^{-u/2} I_{(0,\infty)}(u).$$

It can be shown that the probability density function describing the distribution of W is given by

$$f_{\chi^2_{p,\theta^2}}(w) = e^{-\theta^2/2} f_{\chi^2_p}(w) + e^{-\theta^2/2}\sum\nolimits_{k=1}^{\infty} \frac{\left(\theta^2\right)^k}{2^k k!} f_{\chi^2_{p+2k}}(w).$$

It then follows that the cumulative distribution function of the distribution of W is

$$F_{\chi^2_{k-1,\theta^2}}(y\,|\theta,n) = e^{-\theta^2/2} F_{\chi^2_{k-1}}(w) + e^{-\theta^2/2}\sum\nolimits_{k=1}^{\infty} \frac{\left(\theta^2\right)^k}{2^k k!} F_{\chi^2_{p+2k}}(w).$$

2.7 Family of Noncentral F-Distributions

The random variable

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu_1,\nu_2,\theta^2},$$

where F_{ν_1,ν_2,θ^2} has a noncentral F-distribution with ν_1 degrees of freedom in the numerator, ν_2 degrees of freedom in the denominator, and noncentrality parameter θ^2. Here $W_1 \sim \chi^2_{\nu_1,\theta^2}$ and $W_2 \sim \chi^2_{\nu_2}$are independent and θ is a real number constant. One can show that the probability density function describing the distribution of Q is

$$f_F\left(q\,\middle|\nu_1,\nu_2,\theta^2\right) = e^{-\theta^2/2} f_{F_{\nu_1,\nu_2}}(q)$$
$$+\, e^{-\theta^2/2}\sum\nolimits_{j=1}^{\infty} \frac{\left(\theta^2\right)^j}{2^j j!} f_{F_{\nu_1+2j,\nu_2}}\left(\frac{\nu_1}{\nu_1+2j}q\right)\frac{\nu_1}{\nu_1+2j}.$$

Further, we see that the cumulative distribution function $F_Q(q)$ describing the distribution of Q is given by

$$F_Q\left(q\left|\nu_1,\nu_2,\theta^2\right.\right)=e^{-\theta^2/2}F_{F_{\nu_1,\nu_2}}(q)$$
$$+\,e^{-\theta^2/2}\sum\nolimits_{j=1}^{\infty}\frac{\left(\theta^2\right)^j}{2^j j!}F_{F_{\nu_1+2j,\nu_2}}\left(\frac{\nu_1}{\nu_1+2j}q\right).$$

The mean and variance of a noncentral F-distribution are

$$\begin{aligned}
E\left(F_{\nu_1,\nu_2,\theta^2}\right)&=\int_0^\infty qf_Q\left(q\left|\nu_1,\nu_2,\theta^2\right.\right)dq\\
&=e^{-\theta^2/2}\int_0^\infty qf_{F_{\nu_1,\nu_2}}(q)\,dq\\
&+e^{-\theta^2/2}\sum\nolimits_{j=1}^{\infty}\frac{\left(\theta^2\right)^j}{2^j j!}\int_0^\infty qf_{F_{\nu_1+2j,\nu_2}}\left(\frac{\nu_1}{\nu_1+2j}q\right)\frac{\nu_1}{\nu_1+2j}dq\\
&=\frac{\nu_2\left(\theta^2+\nu_1\right)}{\nu_1\left(\nu_2-2\right)}\text{ and}\\
V\left(F_{\nu_1,\nu_2,\theta^2}\right)&=\frac{\nu_2^2\left[\left(2\nu_1+4+1\right)\theta^2+\nu_1\left(\nu_1+2\right)\right]}{\nu_1^2\left(\nu_2-4\right)\left(\nu_2-2\right)^2}.
\end{aligned}$$

2.8 Statistics Useful in Comparing Two Population Means

Suppose we have two independent samples $X_{i,1},\ldots,X_{i,n_1}$ iid $N\left(\mu_i,\sigma_i^2\right)$ for $i=1,2$. The statistic

$$T_1=\frac{\overline{X}_1-\overline{X}_2}{S_p\sqrt{1/n_1+1/n_2}},$$

with

$$S_p^2=\frac{\left(n_1-1\right)S_1^2+\left(n_2-1\right)S_2^2}{\left(n_1-1\right)+\left(n_2-1\right)}$$

is used to compare two population means μ_1 and μ_2 when the population variances σ_1^2 and σ_2^2 are assumed to be equal, where $\overline{X}_i$ and S_i^2 are the mean and variance of the sample of size n_i to be taken from population i, for $i = 1, 2$. If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, then

$$\frac{(n_i - 1)\, S_i^2}{\sigma_i^2} = \frac{(n_i - 1)\, S_i^2}{\sigma^2} \sim \chi^2_{n_i - 1}$$

for $i = 1, 2$. It follows that

$$\begin{aligned}\frac{(n_1 + n_2 - 2)\, S_p^2}{\sigma^2} &= \frac{(n_1 - 1)\, S_1^2}{\sigma^2} + \frac{(n_2 - 1)\, S_2^2}{\sigma^2} \\ &\sim \chi^2_{n_1 - 1} + \chi^2_{n_2 - 1} = \chi^2_{n_1 + n_2 - 2}.\end{aligned}$$

Next observe that

$$E\left(\overline{X}_1 - \overline{X}_2\right) = \mu_1 - \mu_2 \text{ and } V\left(\overline{X}_1 - \overline{X}_2\right) = \sigma^2\left(1/n_1 + 1/n_2\right).$$

Hence,

$$T_1 = \frac{\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sigma\sqrt{1/n_1 + 1/n_2}} + \frac{(\mu_1 - \mu_2)/\sigma}{\sqrt{1/n_1 + 1/n_2}}}{\sqrt{\frac{(n_1 + n_2 - 2) S_p^2}{\sigma^2} / \left(n_1 + n_2 - 2\right)}} = \frac{Z + \frac{\delta}{\sqrt{1/n_1 + 1/n_2}}}{\sqrt{\chi^2_{n_1 + n_2 - 2} / \left(n_1 + n_2 - 2\right)}},$$

where $\delta = (\mu_1 - \mu_2)/\sigma$. Hence, the statistic T_1 has a noncentral t-distribution with $n_1 + n_2 - 2$ degrees of freedom and noncentrality parameter $\delta/\sqrt{1/n_1 + 1/n_2}$.

The Behrens-Fisher statistic T_2 is used when comparing two population means μ_1 and μ_2 when the population variances σ_1^2 and σ_2^2 are not assumed to be equal. It is defined by

$$T_2 = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}},$$

where $\overline{X}_i$ and S_i^2 are the mean and variance of the sample of size n_i to be taken from Population i, for $i = 1, 2$. Its distribution was studied by Behrens (1929) and Fisher (1939).

Observe that we can write

$$T_2 = \frac{\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} + \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}}{\sqrt{\frac{S_1^2/n_1 + S_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}}}$$

$$= \frac{\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} + \frac{(\mu_1 - \mu_2)/\sigma_2}{\sqrt{\frac{\sigma_1^2}{\sigma_2^2}/n_1 + 1/n_2}}}{\sqrt{\frac{S_1^2/n_1 + S_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}}}$$

$$= \frac{\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} + \frac{\delta}{\sqrt{\lambda^2/n_1 + 1/n_2}}}{\sqrt{\frac{S_1^2/n_1 + S_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}}},$$

where $\delta = (\mu_1 - \mu_2)/\sigma_2$ and $\lambda^2 = \sigma_1^2/\sigma_2^2$. Welch (1938) proposed approximating the distribution of

$$\frac{S_1^2/n_1 + S_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}$$

as a Chi Square distribution that has been divided by its degrees of freedom ν. If the random variable W has a Chi Square distribution with ν degrees of freedom, then

$$V\left(\frac{S_1^2/n_1 + S_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right) \approx V(W/\nu) = V(W)/\nu^2 = 2/\nu.$$

Under our independent Normal model.

$$V\left(\frac{S_1^2/n_1 + S_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right)$$

$$= \frac{\frac{\sigma_1^4}{(n_1-1)^2} V\left(\frac{(n_1-1)S_1^2}{\sigma_1^2}\right)/n_1^2 + \frac{\sigma_2^4}{(n_2-1)^2} V\left(\frac{(n_2-1)S_2^2}{\sigma_2^2}\right)/n_2^2}{\left(\sigma_1^2/n_1 + \sigma_2^2/n_2\right)^2}$$

$$= \frac{2\frac{1}{n_1-1}\left(\frac{\sigma_1^2}{n_1}\right)^2 + 2\frac{1}{n_2-1}\left(\frac{\sigma_2^2}{n_2}\right)^2}{\left(\sigma_1^2/n_1 + \sigma_2^2/n_2\right)^2}.$$

Setting this last expression equal to $2/\nu$ and solving for ν, we have

$$\nu \approx \frac{\left(\sigma_1^2/n_1 + \sigma_2^2/n_2\right)^2}{\frac{1}{n_1-1}\left(\frac{\sigma_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{\sigma_2^2}{n_2}\right)^2}$$
$$= \frac{\left(\frac{\sigma_1^2}{\sigma_2^2}/n_1 + 1/n_2\right)^2}{\frac{1}{n_1-1}\left(\frac{\sigma_1^2/\sigma_2^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{1}{n_2}\right)^2}$$
$$= \frac{\left(\lambda^2/n_1 + 1/n_2\right)^2}{\frac{1}{n_1-1}\left(\frac{\lambda^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{1}{n_2}\right)^2},$$

where $\lambda^2 = \sigma_1^2/\sigma_2^2$. Since

$$\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N\left(0, 1\right),$$

then T_2 has approximately a noncentral t-distribution with ν degrees of freedom and noncentrality parameter

$$\theta = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{\left(\mu_1 - \mu_2\right)/\sigma_2}{\sqrt{\frac{\sigma_1^2}{\sigma_2^2}/n_1 + 1/n_2}} = \frac{\delta}{\sqrt{\lambda^2/n_1 + 1/n_2}},$$

where $\delta = \left(\mu_1 - \mu_2\right)/\sigma_2$. It follows that

$$T_2 = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$$

has an approximate central t-distribution with ν degrees of freedom. Welch (1938) proposed estimating the distribution T_2 by estimating the approximate degrees of freedom ν by

$$\widehat{\nu} = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}.$$

Using $\widehat{\nu}$ in constructing a confidence interval for $\mu_1 - \mu_2$ results in an estimated of the approximate $100\left(1-\alpha\right)\%$ confidence interval. Statistical tests about $\mu_1 - \mu_2$ result in estimates of the approximate distribution of the significance level, the observed significance level, and p-value.

The exact distribution of the Behrens-Fisher statistics was derived by Hu (2010) and is given in Champ and Hu (2022). The following theorem is useful in determining the distribution of the T_1 and T_2.

Theorem: If $Z \sim N(0,1)$, $W_1 \sim \chi^2_{\nu_1}$ and $W_2 \sim \chi^2_{\nu_2}$ are independent with $\theta, \xi, \nu > 0$ constants, then the pdf $f_T(t)$ and cdf $F_T(t)$ describing the distribution of

$$T = \frac{Z+\theta}{\sqrt{(\xi W_1 + W_2)/\nu}}$$

can be expressed, respectively, as

$$f_T(t) = \begin{cases} \xi^{\nu_2/2} f_{t_{\nu_1+\nu_2,\theta}}\left(\sqrt{\frac{\xi(\nu_1+\nu_2)}{\nu}}t\right)\sqrt{\frac{\xi(\nu_1+\nu_2)}{\nu}} \\ +\xi^{\nu_2/2}\sum_{k=1}^{\infty}\frac{(1-\xi)^k\Gamma\left(\frac{\nu_2+2k}{2}\right)}{\Gamma\left(\frac{\nu_2}{2}\right)k!} \\ \times f_{t_{\nu_1+\nu_2+2k,\theta}}\left(\sqrt{\frac{\xi(\nu_1+\nu_2+2k)}{\nu}}t\right)\sqrt{\frac{\xi(\nu_1+\nu_2+2k)}{\nu}}, \\ \text{if } 0<\xi\le 1; \\ \xi^{-\nu_1/2} f_{t_{\nu_1+\nu_2,\theta}}\left(\sqrt{\frac{\nu_1+\nu_2}{\nu}}t\right)\sqrt{\frac{\nu_1+\nu_2}{\nu}} \\ +\xi^{-\nu_1/2}\sum_{k=1}^{\infty}\frac{\left(1-\xi^{-1}\right)^k\Gamma\left(\frac{\nu_1+2k}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)k!} \\ \times f_{t_{\nu_1+\nu_2+2k,\theta}}\left(\sqrt{\frac{\nu_1+\nu_2+2k}{\nu}}t\right)\sqrt{\frac{\nu_1+\nu_2+2k}{\nu}}, \\ \xi>1. \end{cases}$$

and

$$F_T(t) = \begin{cases} \xi^{\nu_2/2} F_{t_{\nu_1+\nu_2,\theta}}\left(\sqrt{\frac{\xi(\nu_1+\nu_2)}{\nu}}t\right) \\ +\xi^{\nu_2/2}\sum_{k=1}^{\infty}\frac{(1-\xi)^k\Gamma\left(\frac{\nu_2+2k}{2}\right)}{\Gamma\left(\frac{\nu_2}{2}\right)k!} \\ \times F_{t_{\nu_1+\nu_2+2k,\theta}}\left(\sqrt{\frac{\xi(\nu_1+\nu_2+2k)}{\nu}}t\right), \\ \text{if } 0 < \xi \leq 1; \\ \xi^{-\nu_1/2} F_{t_{\nu_1+\nu_2,\theta}}\left(\sqrt{\frac{\nu_1+\nu_2}{\nu}}t\right) \\ +\xi^{-\nu_1/2}\sum_{k=1}^{\infty}\frac{\left(1-\xi^{-1}\right)^k\Gamma\left(\frac{\nu_1+2k}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)k!} \\ \times F_{t_{\nu_1+\nu_2+2k,\theta}}\left(\sqrt{\frac{\nu_1+\nu_2+2k}{\nu}}t\right), \\ \text{if } \xi > 1. \end{cases}$$

where $f_{t_{\nu,\theta}}$ and $F_{t_{\nu,\theta}}$ are, respectively, the pdf and cdf of a noncentral t-distribution with ν degrees of freedom.

Corollary 1 to the Theorem**.** If we set

$$\theta = \frac{(\mu_1-\mu_2)/\sigma_2}{\sqrt{\lambda^2/n_1+1/n_2}}, \xi = \lambda^2\frac{n_2(n_2-1)}{n_1(n_1-1)}, \lambda^2 = \sigma_1^2/\sigma_2^2,$$

$$\nu = n_2(n_2-1)\left(\lambda^2/n_1+1/n_2\right), \nu_1 = n_1-1, \nu_2 = n_2-1,$$

$$Z = \frac{\left(\overline{X}_1-\overline{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}, W_1 = \frac{(n_1-1)S_1^2}{\sigma_1^2}, \text{ and}$$

$$W_2 = \frac{(n_2-1)S_2^2}{\sigma_2^2},$$

then $f_{T_2}(t)$ and $F_{T_2}(t)$ are the pdf and cdf, respectively, that describe the distribution of the Behrens-Fisher statistic T_2. Further, setting $\theta = 0$, results in the distribution of

$$T_{2,0} = \frac{\left(\overline{X}_1-\overline{X}_2\right)-(\mu_1-\mu_2)}{\sqrt{S_1^2/n_1+S_2^2/n_2}}.$$

Note that the distribution of T_2 depends on the ratio of the variances $\lambda^2 = \sigma_1^2/\sigma_2^2$.

Corollary 2 of the Theorem**.** If we set

$$\begin{aligned}
\theta &= \frac{(\mu_1 - \mu_2)/\sigma_2}{\sqrt{\lambda^2/n_1 + 1/n_2}}, \xi = \lambda^2, \\
\lambda^2 &= \sigma_1^2/\sigma_2^2, \\
\nu &= \frac{(n_1 + n_2 - 2)\left(\lambda^2/n_1 + 1/n_2\right)}{1/n_1 + 1/n_2}, \\
\nu_1 &= n_1 - 1, \nu_2 = n_2 - 1, \\
Z &= \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}, \\
W_1 &= \frac{(n_1 - 1)\, S_1^2}{\sigma_1^2}, \text{ and} \\
W_2 &= \frac{(n_2 - 1)\, S_2^2}{\sigma_2^2},
\end{aligned}$$

then $f_{T_1}(t)$ and $F_{T_1}(t)$ are the pdf and cdf, respectively, that describe the distribution of the statistic T_1. Further, setting $\theta = 0$, results in the distribution of

$$T_{1,0} = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(1/n_1 + 1/n_2\right)}}.$$

Welch (1938) recommends estimating the approximate distribution of the statistic T_2 as a noncentral t-distribution with $\widehat{\nu}$ degrees of freedom and noncentrality parameter θ, where

$$\widehat{\nu} = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2} \text{ and } \theta = \frac{\delta}{\sqrt{\lambda^2/n_1 + 1/n_2}}$$

with $\delta = (\mu_1 - \mu_2)/\sigma_2$ and $\lambda^2 = \sigma_1^2/\sigma_2^2$.

2.9 CONCLUSION

Noncentral t and F-distributions were derived and discussed. Also, distribution that are used in comparing two population means were discussed. We have introduced the model, the independent Normal model, that will assumed in what follows.

CHAPTER 3

TEST OF SIGNIFICANCE AND TEST OF HYPOTHESIS

3.1 INTRODUCTION

Tests of significance and hypothesis are based on an ordering of the sample space. We will discuss the ordering of the sample space. The likelihood ratio is used to order the sample space. The significance level (SL) is a statistic that transforms the data into a number in the interval $(0, 1)$. Our interest is to study the distribution of the SL. Its observed value is the observed significance level (OSL). The observed value of a statistic is a statistic. Based on the ordering of the sample space and assuming a second sample is to be taken, the probability of observing an OSL that is "as extreme or more extreme" evidence against the null hypothesis is referred to as the p-value. The magnitude of the p-value is the same as the OSL. Their interpretations are quite different, one is a probability and the other is a statistic.

3.2 LIKELIHOOD RATIO TEST

On each individual in a population, a measurement X characterizes the individual. A measurement has a distribution. For the cases in which the measurement X is a real number, the distribution of X can be described by a probability density and cumulative distribution functions $f_X(x|\theta)$ and $F_X(x|\theta)$ respectively, where θ is a vector of values called parameters the characterize the distribution of X. Following DeGroot (1975) and others, we let $\mathbf{\Omega}$ be the collection of all possible values of θ ($\theta \in \mathbf{\Omega}$). Suppose a researcher has a belief (hypothesis) that $\theta \in \mathbf{\Omega} - \mathbf{\Omega}_0$, where $\mathbf{\Omega}_0 \subset \mathbf{\Omega}$. It is typical to refer to the researcher's hypothesis as the "alternative" hypothesis $H_a : \theta \in \mathbf{\Omega} - \mathbf{\Omega}_0$ discussed a test of hypothesis in which the null (H_0) and alternative (H_a) hypotheses are

$$H_0 : \theta \in \mathbf{\Omega}_0 \text{ and } H_a : \theta \in \mathbf{\Omega} - \mathbf{\Omega}_0,$$

where $\mathbf{\Omega}_0 \subset \mathbf{\Omega}$.

Assume that the information about the parameter θ will be available from a random sample of n X measurements

$$\mathbf{X} = [X_1, \ldots, X_n]^{\mathbf{T}}.$$

The generalized likelihood ratio is defined by

$$u(\mathbf{X}) = \frac{\sup_{\theta \in \mathbf{\Omega}_0} f_{\mathbf{X}}(\mathbf{X} \,|\theta)}{\sup_{\theta \in \mathbf{\Omega}} f_{\mathbf{X}}(\mathbf{X} \,|\theta)} = \frac{f_{\mathbf{X}}\left(\mathbf{X} \left| \widehat{\theta}_0 \right.\right)}{f_{\mathbf{X}}\left(\mathbf{X} \left| \widehat{\theta} \right.\right)},$$

where $\widehat{\theta}_0$ is the maximum likelihood estimator of θ assuming the null hypothesis is true and $\widehat{\theta}$ is the maximum likelihood estimate of θ. The generalized likelihood ratio test reject H_0 in favor of H_a if $u(\mathbf{X}) \leq c$, where c is chosen to provide a size α test. The level of significance or the size of the test is

$$\alpha = \sup_{\theta \in \mathbf{\Omega}_0} P[u(\mathbf{X}) \leq c \,|\theta] = \sup_{\theta \in \mathbf{\Omega}_0} F_{u(\mathbf{X})}(c \,|\theta).$$

See Bain and Engelhardt (1992) page 418. It can be shown that $0 < u(\mathbf{X}) \leq 1$.

The likelihood ratio is used to order the sample space relative to the hypotheses. A sample $\mathbf{X}$ provides as much or more evidence against H_0 than the sample $\mathbf{X}^*$ if $u(\mathbf{X}) \leq u(\mathbf{X}^*)$. We say that the sample $\mathbf{X}$ is as extreme or more extreme against H_0 than the sample $\mathbf{X}^*$ if $u(\mathbf{X}) \leq u(\mathbf{X}^*)$.

There are two types of errors one can make using a test of hypothesis. The first is called a Type I error if using the data one using the data rejects a true null hypothesis. The probability of a Type I error is the size of the test α. The second kind of error is called a Type II error. A Type II error occurs if the test fails to reject a false null hypothesis. The probability of a Type II error is often denoted by β. A Type II error occurs when the alternative hypothesis is true. Thus, the probability of a Type II error is a function of $\theta \in \mathbf{\Omega} - \mathbf{\Omega}_0$. We denote this by $\beta(\theta)$. It is not difficult to see that $\beta(\theta) = 1 - F_{u(\mathbf{X})}(c\,|\theta)$. The power of the test is defined as $power(\theta) = 1 - \beta(\theta)$. Little guidance is given in the literature for selecting α and β. The Food and Drug Administration (FDA) usually require $\alpha = 0.05$ and $\beta = 0.20$ according to Dr. Karl Peace.

The significance level for the likelihood ratio test is

$$SL = \sup_{\theta \in \mathbf{\Omega}_0} F_{U(\mathbf{X})}\left(u(\mathbf{X})\,|\theta\right) \sim F_{SL}\left(q\,|\theta\right)$$

Note that $u(\mathbf{X})$ and SL are transformations of the sample measurements $\mathbf{X}$ into the interval $(0, 1)$. They are random variables since they depend on the sample values. The observed significance level (OSL) is

$$OSL = \sup_{\theta \in \mathbf{\Omega}_0} F_{u(\mathbf{X})}\left(u(\mathbf{x})\,|\theta\right)$$

The observed values of $u(\mathbf{x})$ and OSL are transformation of the observed sample measurements $\mathbf{x}$ into the interval $(0, 1)$, where $u(\mathbf{x})$ and OSL are the observed value of a random variable $u(\mathbf{X})$ and SL, respectively. Equivalently, the likelihood ratio test reject H_0 in favor of H_a if the $OSL \leq \alpha$.

Let $\mathbf{X}^*$ be the n measurements to be taken on a future sample. The p-value is defined by

$$p\text{-value } = \sup_{\theta \in \mathbf{\Omega}_0} P\left[u(\mathbf{X}^*) \leq u(\mathbf{x})\,|u(\mathbf{X}) = u(\mathbf{x}), \theta\right].$$

The OSL is numerically equal to the p-value. However, note that OSL is the observed value of a random variable. Further, note that the p-value is a conditional probability not based on a sample that will be observed. The decision to reject or fail to reject H_0 should be based on the OSL not the p-value.

The likelihood ratio test under the assumption the sample is a random sample with a common $N(\mu, \sigma^2)$ distribution results in a t-test of the hypothesis $H_0 : \mu = \mu_0$ with $H_a : \mu \neq \mu_0$. See Bain and Engelhardt (1992). The decision rule is to reject H_0 in favor of H_a if the observed value of $|T| \geq t_{n-1,\alpha/2}$, where α is the size of the test and $T = \left(\overline{X} - \mu_0\right) / \left(S/\sqrt{n}\right)$.

3.3 THE t-TEST

Concato and Hartigan (2016) state "[a] threshold probability value of '$p \leq 0.05$' is commonly used in clinical investigations to indicate statistical significance. To allow clinicians to better understand evidence generated by research studies, this review defines the p value, summarizes the historical origins of the p value approach to hypothesis testing, describes various applications of $p \leq 0.05$ in the context of clinical research and discusses the emergence of $p \leq 5 \times 10^{-8}$ and other values as thresholds for genomic statistical analyses."

Suppose the null and alternative hypotheses of interest are $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$. These hypotheses are equivalent to the hypotheses

$$
\begin{aligned}
&H_0 : \mu - \mu_0 \leq 0 \text{ versus } H_a : \mu - \mu_0 > 0 \text{ or} \\
&H_0 : \frac{\mu - \mu_0}{\sigma} \leq 0 \text{ versus } H_a : \frac{\mu - \mu_0}{\sigma} > 0 \text{ or} \\
&H_0 : \delta \leq 0 \text{ versus } H_a : \delta > 0,
\end{aligned}
$$

where $\delta = (\mu - \mu_0)/\sigma$, the effect size, is the number of population standard deviations μ differs from μ_0.

The data the researcher will have available to make a decision about the process is a random sample $X_1, \ldots, X_n$ with a common $N(\mu, \sigma^2)$ distribution. We denote the mean and standard deviation of the sample by $\overline{X}_n$ and S_n, respectively. The test statistic we will use is

$$\begin{aligned} T = T_{n-1,\sqrt{n}\delta} &= \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} = \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} + \sqrt{n}\frac{\mu-\mu_0}{\sigma}}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2}/(n-1)}} \\ &= \frac{\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} + \sqrt{n}\delta}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2}/(n-1)}} = \frac{Z + \sqrt{n}\delta}{\sqrt{W/(n-1)}} \sim t_{n-1,\sqrt{n}\delta}, \end{aligned}$$

where $Z \sim N(0,1)$ and $W \sim \chi^2_{n-1}$ are independent under our model with $t_{n-1,\sqrt{n}\delta}$ a random variable that has a non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter $\sqrt{n}\delta$. Recall that $\delta = (\mu - \mu_0)/\sigma$ which is the effect size. The mean μ_T and variance σ_T^2 of the distribution of T are given by

$$\begin{aligned} \mu_T &= \frac{\sqrt{n-1}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)}\sqrt{n}\delta \text{ and} \\ \sigma_T^2 &= \frac{(n-1)(1+n\delta^2)}{n-3} - \left(\frac{\sqrt{n-1}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)}\right)^2 n\delta^2. \end{aligned}$$

These results are easily derived. For a fixed value of δ, one can show that

$$\lim_{n\to\infty} \mu_T = \infty \text{ and } \lim_{n\to\infty} \sigma_T^2 = \frac{1}{2}\delta^2 + 1.$$

Hence, the mean of the distribution of T under the alternative hypothesis increases to infinity as n increases with finite variance that depends on the effect size δ. This implies that the observed significance level and the p-value decrease as n increases.

The significance level for testing $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$ is defined as

$$SL_{n,\delta} = \max_{\delta \leq 0}\left(1 - F_{t_{n-1,\sqrt{n}\delta}}\left(T_{n-1,\sqrt{n}\delta}\right)\right) = 1 - F_{t_{n-1,0}}\left(T_{n-1,\sqrt{n}\delta}\right),$$

The distribution of the $SL_{n,\delta}$ is determined as follows by finding the cumulative distribution function that describes its distribution.

$$\begin{aligned}
F_{SL_{n,\delta}}(q) &= P\left(1 - F_{t_{n-1,0}}\left(T_{n-1,\sqrt{n}\delta}\right) \leq q\right) \\
&= P\left(F_{t_{n-1,0}}\left(T_{n-1,\sqrt{n}\delta}\right) \geq 1-q\right) \\
&= 1 - P\left(F_{t_{n-1,0}}\left(T_{n-1,\sqrt{n}\delta}\right) < 1-q\right) \\
&= 1 - P\left[T_{n-1,\sqrt{n}\delta} < F^{-1}_{t_{n-1,0}}(1-q)\right] \\
&= 1 - F_{t_{n-1,\sqrt{n}\delta}}\left(F^{-1}_{t_{n-1,0}}(1-q)\right).
\end{aligned}$$

Now suppose that the researcher is interested in a test that would detect a change in the mean with $\delta \geq \delta_a = (\mu_a - \mu_0)/\sigma$, where μ_a is a value of μ that makes the alternative hypothesis true that is provided by the researcher. The value of μ_a is the minimum value of μ that the researcher believes is either “clinically” or “practically” significant. We will show that an unbiased estimator $\widehat{\delta}_a$ of δ_a is

$$\widehat{\delta}_a = \frac{\mu_a - \mu_0}{S/c_4}$$

with S the standard deviation of the sample and

$$c_4 = \frac{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{n-1}\Gamma\left(\frac{n-2}{2}\right)}.$$

Under the independent Normal model, S/c_4 is an unbiased estimator of σ. The expectation of $\widehat{\delta}_a$ is

$$
\begin{aligned}
E\left(\widehat{\delta}_a\right) &= E\left(\frac{\mu_a - \mu_0}{\frac{\sigma}{\sqrt{n-1}} W^{1/2}/c_4}\right) \\
&= \sqrt{n-1}c_4\left(\frac{\mu_a - \mu_0}{\sigma}\right) E\left(W^{-1/2}\right) \\
&= \sqrt{n-1}c_4\left(\frac{\mu_a - \mu_0}{\sigma}\right) \\
&\times \int_0^\infty w^{-1/2} \frac{1}{\Gamma\left(\frac{n-1}{2}\right) 2^{(n-1)/2}} w^{(n-1)/2-1} e^{-w/2} dw \\
&= \sqrt{n-1}c_4\left(\frac{\mu_a - \mu_0}{\sigma}\right) \frac{\Gamma\left(\frac{n-2}{2}\right) 2^{(n-2)/2}}{\Gamma\left(\frac{n-1}{2}\right) 2^{(n-1)/2}} \\
&\times \int_0^\infty \frac{1}{\Gamma\left(\frac{n-2}{2}\right) 2^{(n-2)/2}} w^{(n-2)/2-1} e^{-w/2} dw \\
&= c_4\left(\frac{\mu_a - \mu_0}{\sigma}\right) \frac{\sqrt{n-1}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)} \\
&= c_4\left(\frac{\mu_a - \mu_0}{\sigma}\right) c_4^{-1} = \frac{\mu_a - \mu_0}{\sigma} \\
&= \delta_a,
\end{aligned}
$$

where $W = (n-1)\, S^2/\sigma^2 \sim \chi^2_{n-1}$. Hence, the statistic $\widehat{\delta}_a$ is an unbiased estimator of δ_a.

When using a t-test, the following Table 3.4.1 gives the possible null and alternative hypotheses as well as the SL.

Table 3.1: Hypotheses and Significance Level

$H_0 : \mu \leq \mu_0$	$H_a : \mu > \mu_0$	$SL_{n,\delta} = 1 - F_{t_{n-1,0}}\left(T_{n-1,\sqrt{n}\delta}\right)$
$H_0 : \mu = \mu_0$	$H_a : \mu \neq \mu_0$	$SL_{n,\delta} = 2\left[1 - F_{t_{n-1,0}}\left(\left\|T_{n-1,\sqrt{n}\delta}\right\|\right)\right]$
$H_0 : \mu \geq \mu_0$	$H_a : \mu < \mu_0$	$SL_{n,\delta} = F_{t_{n-1,0}}\left(T_{n-1,\sqrt{n}\delta}\right)$

The distribution of T^2 is a non-central F-distribution with 1 degree of freedom in the numerator, $n-1$ degree of freedom in the denominator, and non-centrality parameter $n\delta^2$.

We have

$$T^2 = \frac{\left(Z + \sqrt{n}\delta\right)^2}{W/\left(n-1\right)} \sim F_{1,n-1,\sqrt{n}\delta}.$$

The mean $E\left(T^2\right)$ of T^2 is determined as follows.

$$\begin{aligned}
E\left(T^2\right) &= \left(n-1\right) E\left[\left(Z + \sqrt{n}\delta\right)^2\right] E\left(W^{-1}\right) \\
&= \left(n-1\right) E\left(Z^2 + 2\sqrt{n}\delta Z + n\delta^2\right) E\left(W^{-1}\right) \\
&= \left(n-1\right)\left(1 + n\delta^2\right) \int_0^\infty w^{-1} \frac{1}{\Gamma\left(\frac{n-1}{2}\right) 2^{(n-1)/2}} w^{(n-1)/2-1} e^{-w/2} dw \\
&= \left(n-1\right)\left(1 + n\delta^2\right) \int_0^\infty \frac{1}{\Gamma\left(\frac{n-1}{2}\right) 2^{(n-1)/2}} w^{(n-3)/2-1} e^{-w/2} dw \\
&= \frac{\left(n-1\right)\left(1+n\delta^2\right)\Gamma\left(\frac{n-3}{2}\right) 2^{(n-3)/2}}{\Gamma\left(\frac{n-1}{2}\right) 2^{(n-1)/2}} \int_0^\infty \frac{1}{\Gamma\left(\frac{n-3}{2}\right) 2^{(n-3)/2}} w^{(n-3)/2-1} e^{-w/2} dw \\
&= \frac{\left(n-1\right)\left(1+n\delta^2\right)\left(\frac{n-1}{2} - 1\right)\Gamma\left(\frac{n-1}{2} - 1\right) 2^{(n-3)/2}}{\left(\frac{n-1}{2} - 1\right)\Gamma\left(\frac{n-1}{2}\right) 2^{(n-1)/2}} \\
&= \frac{\left(n-1\right)\left(1+n\delta^2\right)}{\left(n-3\right)}.
\end{aligned}$$

Now, suppose the researcher can provide a value μ_p for μ that makes the alternative hypothesis true that is of practical significance. What sample size n is needed to detect this mean μ_p that is of practical significance? Associated with the desired value of n is the power of the test $1 - \beta_p$. For this test, we require that n is be a solution to the equation

$$\begin{aligned}
1 - P\left(\frac{\overline{X}_n - \mu_p}{S_n/\sqrt{n}} \leq t_{n-1,0,\alpha} \left|\mu > \mu_p\right.\right) &\geq 1 - \beta_p \text{ or} \\
1 - F_{t_{n-1,\sqrt{n}\delta_p}}\left(t_{n-1,0,\alpha}\right) &\geq 1 - \beta_p \text{ or} \\
F_{t_{n-1,\sqrt{n}\delta_p}}\left(t_{n-1,0,\alpha}\right) &\leq \beta_p,
\end{aligned}$$

where $\delta_p = \left(\mu - \mu_p\right)/\sigma$. This equation can be solved for n.

The question is what size of the test is needed to detect this change in the mean? A critical value for this test is $t_{n-1,0,\alpha_0}$ for a test of size α_0, where $t_{n-1,0,\alpha_0}$ is the $100\left(1-\alpha_0\right)$th percentile of a central t-distribution with $n-1$ degrees of freedom. Let $t_{n-1,\sqrt{n}\delta,\alpha}$ be the $100\,(1-\alpha)$th percentile of a non-central t-distribution with $n-1$ degrees of freedom and non-centrality parameter $\sqrt{n}\delta$. Let's set $t_{n-1,0,\alpha_0} = t_{n-1,\sqrt{n}\delta_a,\alpha}$ and solve for $\alpha_0 = \alpha_0\,(\alpha, n, \delta_a)$.

$$\delta = \frac{\mu-\mu_0}{\sigma} \text{ or } \mu = \mu_0 + \delta\sigma.$$

The significance level (SL) for this test is

$$SL = \max_{\mu\le\mu_0}\left\{1 - F_{t_{n-1,\sqrt{n}\delta}}\left(T_{n-1,\sqrt{n}\delta}\right)\right\} = 1 - F_{t_{n-1,0}}\left(T_{n-1,\sqrt{n}\delta}\right),$$

where $T_{n-1,\sqrt{n}\delta} \sim t_{n-1,\sqrt{n}\delta}$. The cumulative distribution function that describes the sampling distribution of the statistic SL is found as follows with the indicator function $I_{(0,1)}\,(q) = 1$ if $q \in (0,1)$ and zero otherwise.

$$\begin{aligned}
F_{SL}\,(q) &= P\left[1 - F_{t_{n-1,0}}\,(T) \le q\right] I_{(0,1)}\,(q) \\
&= P\left[F_{t_{n-1,0}}\,(T) \ge 1-q\right] I_{(0,1)}\,(q) \\
&= P\left[T \ge F^{-1}_{t_{n-1,0}}\,(1-q)\right] I_{(0,1)}\,(q) \\
&= 1 - P\left[T < F^{-1}_{t_{n-1,0}}\,(1-q)\right] I_{(0,1)}\,(q) \\
&= 1 - F_{t_{n-1,\sqrt{n}\delta}}\left(F^{-1}_{t_{n-1,0}}\,(1-q)\right) I_{(0,1)}\,(q)\,.
\end{aligned}$$

The probability density function describing the sampling distribution of SL is

$$\begin{aligned}
f_{SL}\,(q) = \frac{d}{dq}F_{SL}\,(q) &= -\frac{d}{dq}F^{-1}_{t_{n-1,\sqrt{n}\delta}}\left(F^{-1}_{t_{n-1,0}}\,(1-q)\right)\frac{d}{dq}F^{-1}_{t_{n-1,0}}\,(1-q)\,I_{(0,1)}\,(q) \\
&= \frac{f_{t_{n-1,\sqrt{n}(\mu-\mu_0)/\sigma}}\left(F^{-1}_{t_{n-1,0}}\,(1-q)\right)}{f_{t_{n-1,0}}\left(F^{-1}_{t_{n-1,0}}\,(1-q)\right)} I_{(0,1)}\,(q)\,.
\end{aligned}$$

If we assume the null hypothesis holds, then

$$f_{SL}\,(q) = \frac{f_{t_{n-1,0}}\left(F^{-1}_{t_{n-1,0}}\,(1-q)\right)}{f_{t_{n-1,0}}\left(F^{-1}_{t_{n-1,0}}\,(1-q)\right)} I_{(0,1)}\,(q) = I_{(0,1)}\,(q)\,.$$

Hence, the significance level has a Uniform distribution on the interval $(0,1)$.

3.4 Conclusion

As the sample size n increases, we have shown that the expectation of T increases to infinity while the variance depends only on the overall effect size. An unbiased estimator of the overall effect size was given. We derived the distribution of the significance level for the t-test. If the null hypothesis is true, we have shown that the distribution of the significance level is a Uniform distribution on the interval $(0, 1)$.

CHAPTER 4

ONE-WAY ANALYSIS OF VARIANCE (ANOVA)

4.1 Introduction

Analysis of variance (ANOVA) is used to analysis the data of many statistically designed experiments. We give an example of a one-way analysis of variance.

4.2 An Example of a One-Way Analysis of Variance

The model in a one-way analysis of variance (ANOVA) has the response variable Y defined by

$$Y_{ijk} = \mu + \tau_i + \epsilon_{ij},$$

where μ is an overall mean and τ_i is the ith treatment effect for $i = 1, \ldots, a$ and $j = 1, \ldots, n$. It is further assumed that

$$\sum\nolimits_{i=1}^{a} \tau_i = 0.$$

See Scheffe (1959) for further discussion of the one-way analysis of variance.

The method used to analyze the data begins by examining the statistic

$$SST = \sum\nolimits_{i=1}^{a} \sum\nolimits_{j=1}^{n} \left(Y_{ij} - \overline{Y}_{..}\right)^2$$

that is a measure of the variability in the responses, where

$$\overline{Y}_{..} = \frac{1}{na} \sum\nolimits_{i=1}^{a} \sum\nolimits_{j=1}^{n} Y_{ij}.$$

The statistic SST is referred to as sum of squares total. This variability can be partitioned as $SST = SSE + SSTR$, where

$$SSE = \sum\nolimits_{i=1}^{a} \sum\nolimits_{j=1}^{n} \left(Y_{ij} - \overline{Y}_{i.}\right)^2 \text{ and } SSTR = \sum\nolimits_{i=1}^{a} \sum\nolimits_{j=1}^{n} \left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2,$$

where

$$\overline{Y}_{i.} = \frac{1}{n}\sum\nolimits_{j=1}^{n} Y_{ij}$$

for $i = 1, \ldots, a$. The statistics SSE and $SSTR$ are commonly referred to as the sum of squares error and the sum of squares treatment, respectively.

Observe that we can write the $SSTR$ using vector notation as

$$\begin{aligned} SSTR &= \sum\nolimits_{i=1}^{a}\sum\nolimits_{j=1}^{n}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 = n\sum\nolimits_{i=1}^{a}\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)^2 \\ &= n\begin{bmatrix} \overline{Y}_{1.} - \overline{Y}_{..} \\ \overline{Y}_{2.} - \overline{Y}_{..} \\ \vdots \\ \overline{Y}_{a.} - \overline{Y}_{..} \end{bmatrix}^{\mathbf{T}} \begin{bmatrix} \overline{Y}_{1.} - \overline{Y}_{..} \\ \overline{Y}_{2.} - \overline{Y}_{..} \\ \vdots \\ \overline{Y}_{a.} - \overline{Y}_{..} \end{bmatrix}. \end{aligned}$$

It is not difficult to show that

$$SSTR = \overline{\mathbf{Y}}^{\mathbf{T}}\left(\mathbf{I} - \frac{1}{a}\mathbf{J}\right)\overline{\mathbf{Y}},$$

where $\mathbf{I}$ is a $a \times a$ identity matrix, $\mathbf{J}$ is a $a \times a$ matrix of ones, and $\overline{\mathbf{Y}} = \left[\overline{Y}_{1.}, \ldots, \overline{Y}_{k.}\right]^{\mathbf{T}}$. As one can see, this is the quadratic form of $SSTR$. It can be shown the $a \times a$ matrix $\mathbf{I} - \frac{1}{a}\mathbf{J}$ has $a - 1$ eigenvalues that are ones and one eigenvalue that is zero. Further, one can show that the matrix can be expressed as

$$\mathbf{I} - \frac{1}{a}\mathbf{J} = \mathbf{V}\mathbf{H}\mathbf{V}^{\mathbf{T}},$$

where the $a \times a$ matrix $\mathbf{H}$ has ones in the first $a - 1$ diagonal components and zeroes elsewhere, and the first $a-1$ columns of the $a \times a$ matrix $\mathbf{V}$ are the normalized eigenvectors associated with the $a - 1$ eigenvalues of 1 and the a column is the normalized eigenvector associated with the eigenvalue of zero. It is not difficult to show that $\mathbf{H}$ is idempotent.

We can now write

$$SSTR = \overline{\mathbf{Y}}^{\mathbf{T}}\mathbf{V}\mathbf{H}\mathbf{V}^{\mathbf{T}}\overline{\mathbf{Y}} = \left[\mathbf{H}\mathbf{V}^{\mathbf{T}}\overline{\mathbf{Y}}\right]^{\mathbf{T}}\left[\mathbf{H}\mathbf{V}^{\mathbf{T}}\overline{\mathbf{Y}}\right].$$

Further, we can express $\overline{\mathbf{Y}}$ as

$$\overline{\mathbf{Y}} = \sigma\begin{bmatrix} \frac{\overline{Y}_{1.}-\mu_1}{\sigma/\sqrt{n}} + \sqrt{n}\frac{\mu_1}{\sigma} \\ \frac{\overline{Y}_{2.}-\mu_2}{\sigma/\sqrt{n}} + \sqrt{n}\frac{\mu_2}{\sigma} \\ \vdots \\ \frac{\overline{Y}_{k.}-\mu_k}{\sigma/\sqrt{n}} + \sqrt{n}\frac{\mu_k}{\sigma} \end{bmatrix} = \sigma\begin{bmatrix} Z_1 + \sqrt{n}\delta_1 \\ Z_2 + \sqrt{n}\delta_2 \\ \vdots \\ Z_k + \sqrt{n}\delta_k \end{bmatrix} = \sigma\left(\mathbf{Z} + \sqrt{n}\delta\right),$$

where $Z_i = \left(\overline{Y}_{i.} - \mu_i\right)/\left(\sigma/\sqrt{n}\right) \sim N(0,1)$, $\delta_i = \mu_i/\sigma$, $\mathbf{Z} = [Z_{1.}, \ldots, Z_{k.}]^{\mathbf{T}}$ and $\delta = [\delta_{1.}, \ldots, \delta_{k.}]^{\mathbf{T}}$. It is not difficult to see that Z_i's are independent since the $\overline{Y}_{i.}$'s are independent. It follows that

$$\begin{aligned} SSTR &= \sigma^2\left[\mathbf{H}\mathbf{V}^{\mathbf{T}}\left(\mathbf{Z} + \sqrt{n}\delta\right)\right]^{\mathbf{T}}\left[\mathbf{H}\mathbf{V}^{\mathbf{T}}\left(\mathbf{Z} + \sqrt{n}\delta\right)\right] \\ &= \sigma^2\left[\left(\mathbf{H}\mathbf{V}^{\mathbf{T}}\mathbf{Z} + \sqrt{n}\mathbf{H}\mathbf{V}^{\mathbf{T}}\delta\right)\right]^{\mathbf{T}}\left[\left(\mathbf{H}\mathbf{V}^{\mathbf{T}}\mathbf{Z} + \sqrt{n}\mathbf{H}\mathbf{V}^{\mathbf{T}}\delta\right)\right]. \end{aligned}$$

The random vector

$$\mathbf{Z}^* = \mathbf{H}\mathbf{V}^{\mathbf{T}}\mathbf{Z} = \left[Z^*_{1.}, \ldots, Z^*_{a-1.}, 0\right]^{\mathbf{T}} \sim N_a(0, \mathbf{H}).$$

Hence, the first $a-1$ components of $\mathbf{Z}^*$ are independent standard Normal random variables.

Letting

$$\tau = \sqrt{n}\mathbf{H}\mathbf{V}^{\mathbf{T}}\delta,$$

we have

$$SSTR = \sigma^2\left(\mathbf{Z}^* + \tau\right)^{\mathbf{T}}\left(\mathbf{Z}^* + \tau\right) = \sigma^2\sum_{i=1}^{a-1}\left(Z^*_i + \tau_i\right)^2.$$

There is an orthongonal matrix $\mathbf{B}$ such that $\mathbf{B}\tau = \sqrt{\tau^{\mathbf{T}}\tau}\mathbf{e}_1 = d\mathbf{e}_1$, where $d^2 = \tau^{\mathbf{T}}\tau$ and $\mathbf{e}_1$ is a $a \times 1$ vector with first component 1 and the other $a-1$ components zeros. It follows that

$$\mathbf{B}\left(\mathbf{Z}^{*}+\tau\right)=\mathbf{B}\mathbf{Z}^{*}+\mathbf{B}\tau=\mathbf{Z}^{**}+\sqrt{\tau^{\mathbf{T}}\tau}\mathbf{e}_{1}=\mathbf{Z}^{**}+d\mathbf{e}_{1}.$$

It is not difficult to show that $\mathbf{Z}^{**}\sim N_{a}\left(\mathbf{0},\mathbf{I}\right)$. Hence,

$$SSTR=\sigma^{2}\left(\mathbf{Z}^{**}+d\mathbf{e}_{1}\right)^{\mathbf{T}}\left(\mathbf{Z}^{**}+d\mathbf{e}_{1}\right)=\sigma^{2}\left[\left(Z_{i}^{**}+\sqrt{n}d\right)^{2}+\sum\nolimits_{i=2}^{a-1}\left(Z_{i}^{**}\right)^{2}\right].$$

Therefore,

$$SSTR\sim\sigma^{2}\chi_{a-1,nd^{2}}^{2}\text{ or }\frac{SSTR}{\sigma^{2}}\sim\chi_{a-1,nd^{2}}^{2},$$

where

$$\begin{aligned}d^{2}&=\sum\nolimits_{i=1}^{a-1}\tau_{i}^{2}=\left(\mathbf{H}\mathbf{V}^{\mathbf{T}}\sqrt{n}\delta\right)^{\mathbf{T}}\left(\mathbf{H}\mathbf{V}^{\mathbf{T}}\sqrt{n}\delta\right)\\&=\left(\sqrt{n}\delta\right)^{\mathbf{T}}\left(\mathbf{I}-\frac{1}{na}\mathbf{J}\right)\left(\sqrt{n}\delta\right)\\&=n\sum\nolimits_{i=1}^{a}\left(\delta_{i}-\overline{\delta}\right)^{2}=n\sum\nolimits_{i=1}^{a}\left(\frac{\mu_{i}-\overline{\mu}}{\sigma}\right)^{2},\end{aligned}$$

with

$$\overline{\delta}=\frac{1}{a}\sum\nolimits_{i=1}^{a}\delta_{i}=\frac{1}{a}\sum\nolimits_{i=1}^{a}\mu_{i}/\sigma=\overline{\mu}/\sigma.$$

The mean square due to treatment $MSTR$ is defined by $MSTR=SSTR/\left(a-1\right)$, where $a-1$ is the degrees of freedom associated with $SSTR$.

Further, we observe that

$$\begin{aligned}SSE&=\sum\nolimits_{i=1}^{a}\sum\nolimits_{j=1}^{n}\left(Y_{ij}-\overline{Y}_{i.}\right)^{2}=\sum\nolimits_{i=1}^{a}\left(n-1\right)\frac{1}{n-1}\sum\nolimits_{j=1}^{n}\left(Y_{ij}-\overline{Y}_{i.}\right)^{2}\\&=\sum\nolimits_{i=1}^{a}(n-1)\,S_{i}^{2}=\sigma^{2}\sum\nolimits_{i=1}^{a}\frac{\left(n-1\right)S_{i}^{2}}{\sigma^{2}}.\end{aligned}$$

Thus,

$$\frac{SSE}{\sigma^{2}}\sim\sum\nolimits_{i=1}^{a}\chi_{n-1}^{2}\sim\chi_{na-a}^{2}$$

The mean square due to error MSE is defined by $MSE = SSE/\left(na - a\right)$, where $na - a$ is the degrees of freedom associated with SSE.

The null and alternative hypotheses that are typically tested are

$H_0 : \tau_1 = \ldots = \tau_a = 0$ and

H_a : at least one of the τ_i's not equal to zero.

A size α F-test rejects H_0 if the observed value of

$$F = \frac{MSTR}{MSE} \geq F_{a-1,na-a,\alpha},$$

where $F_{a-1,na-a,\alpha}$ is the $100\left(1 - \alpha\right)$th percentile of a central ($d = 0$) F-distribution. Note that F has a noncentral F-distribution with $a - 1$ numerator degrees of freedom, $na - a$ denominator degrees of freedom, and noncentrality parameter d^2. We express this as

$$F \sim F_{a-1,na-a,d^2},$$

where $F_{a-1,na-a,d^2}$ is a random variable having a non-central F-distribution with $a - 1$ numerator degrees of freedom, $na - a$ denominator degrees of freedom, and non-centrality parameter d^2.

The expectation of F is determined as follows.

$$
\begin{aligned}
E(F) &= E\left(F_{a-1,na-a,nd^2}\right) = E\left(\frac{MSTR}{MSE}\right) \\
&= E\left(\frac{SSTR/(a-1)}{SSE/(na-a)}\right) = \frac{na-a}{a-1}E\left(\frac{SSTR}{SSE}\right) \\
&= \frac{(n-1)\,a}{a-1}E\left(\frac{SSTR}{SSE}\right) = \frac{(n-1)\,a}{a-1}E\left(\frac{\frac{SSTR}{\sigma^2}}{\frac{SSE}{\sigma^2}}\right) \\
&= \frac{(n-1)\,a}{a-1}E\left(\frac{\chi^2_{a-1,d^2}}{\chi^2_{na-a}}\right) \\
&= \frac{(n-1)\,a}{a-1}E\left(\chi^2_{a-1,d^2}\right)E\left[\left(\chi^2_{na-a}\right)^{-1}\right] \\
&= \frac{(n-1)\,a}{a-1}E\left(a-1+nd^2\right) \\
&\times \int_0^\infty w^{-1}\frac{1}{\Gamma\left(\frac{na-a}{2}\right)2^{(na-a)/2}}w^{(na-a)/2-1}e^{-w/2}dw \\
&= \frac{(n-1)\,a\,(a-1+nd^2)\,\Gamma\left(\frac{na-a-2}{2}\right)2^{(na-a-2)/2}}{(a-1)\,\Gamma\left(\frac{na-a}{2}\right)2^{(na-a-2)/2+1}} \\
&= \frac{(n-1)\,a\,(a-1+nd^2)\,\Gamma\left(\frac{na-a}{2}-1\right)}{2\,(a-1)\left(\frac{na-a}{2}-1\right)\Gamma\left(\frac{na-a}{2}-1\right)} \\
&= \frac{(n-1)\,a\,(a-1+nd^2)}{(a-1)\,(na-a-2)} \\
&= \frac{a\,(a-1+nd^2)}{a-1}\frac{1-1/n}{a-\frac{a-2}{n}}.
\end{aligned}
$$

We see that as n goes to infinity

$$
\lim_{n\to\infty}E(F) = \lim_{n\to\infty}\frac{(n-1)\,a\,(a-1+d^2)}{2\,(a-1)\left(\frac{na-a-2}{2}\right)} = \frac{(a-1+d^2)}{a-1} = \infty.
$$

Further, we have

$$
\begin{aligned}
V(F) &= 2\frac{\left(a-1+d^2\right)^2+\left(a-1+2d^2\right)(na-a-2)}{(na-a-2)^2\,(na-a-4)}\left(\frac{na-a}{a-1}\right)^2 \\
&= 2\frac{\left(a-1+d^2\right)^2+\left(a-1+2d^2\right)(a\,(n-1)-a-2)}{(a\,(n-1)-a-2)^2\,(a\,(n-1)-a-4)}\left(\frac{a\,(n-2)}{a-1}\right)^2.
\end{aligned}
$$

It is of interest to examine

$$\lim_{n\to\infty} V(F) = \frac{2\left(a-1+2d^2\right)}{(a-1)^2}.$$

As in the case of the t-test, the mean of the distribution of the test statistic F increases to infinity while the variance only depends on the overall effect size d.

Determining the sample size n requires a level of significance α, a probability of a Type II error β, and a value d_a^2 of d^2 that is of statistically significant, of practical significance, or of clinical significance. The power of the test is defined as $1-\beta$. The minimal sample size n is the least value of n that is the solution to the equation

$$1 - F_{F_{a-1,na-a,nd^2}}\left(F_{a-1,na-a,\alpha}\right) \geq 1-\beta \text{ or}$$
$$F_{F_{a-1,na-a,nd^2}}\left(F_{a-1,na-a,\alpha}\right) \leq \beta$$

subject to

$$1 - F_{F_{a-1,na-a,0}}\left(F_{a-1,na-a,\alpha}\right) = \alpha$$

with cumulative distribution function

$$\begin{aligned} F_Q\left(q\left|\nu_1,\nu_2,d^2\right.\right) &= e^{-d^2/2}F_{F_{a-1,na-a,0}}(q) \\ &\quad + e^{-d^2/2}\sum\nolimits_{j=1}^{\infty}\frac{\left(d^2\right)^j}{2^j j!}F_{F_{a-1+2j,na-a}}\left(\frac{a-1}{a-1+2j}q\right). \end{aligned}$$

For example, suppose $\alpha = 0.05$ and $1-\beta = 0.80$ ($\beta = 0.20$) with $a = 4$ and $d_a = 2$. For $n = 10$, we have for $n = 10$ the critical value of the test is

$$F_{4-1,10(4)-4,0.05} = \text{FInv}(0.95; 4-1, 10\,(4)-4) = 2.86627,$$

rounded to five decimal places. The power of the test for $n = 10$ and $d_a = 2$ is 0.80252, rounded to five decimal places. The following table give the power of the test for various sample sizes obtained through trial and error.

Table 4.1: Power of the Test

n	Power of the Test
6	0.79856
7	0.80000
8	0.80106
9	0.80188

The values in the table have been rounded to five decimal places. We see that a sample of size $n = 7$ gives us the first sample size in which the power is at least 80%.

The significance level (SL) for this test is defined by

$$\begin{aligned} SL &= \max_{H_0:\delta=0} \left\{1 - F_{F_{a-1,na-a,\sqrt{n}\delta}} \left(F_{a-1,na-a,\sqrt{n}\delta}\right)\right\} \\ &= 1 - F_{F_{a-1,na-a,0}} \left(F_{a-1,na-a,\sqrt{n}\delta}\right). \end{aligned}$$

The cumulative distribution function describing the distribution of the statistic SL is determined as follows.

$$\begin{aligned} F_{SL}(q) &= P\left[1 - F_{F_{a-1,na-a,0}} \left(F_{a-1,na-a,\delta}\right) \leq q\right] \\ &= P\left[F_{F_{a-1,na-a,0}} \left(F_{a-1,na-a,\delta}\right) \geq 1 - q\right] \\ &= P\left[F_{a-1,na-a,\delta} \geq F^{-1}_{F_{a-1,na-a,0}} (1-q)\right] \\ &= 1 - F_{F_{a-1,na-a,\delta}} \left(F^{-1}_{F_{a-1,na-a,0}} (1-q)\right). \end{aligned}$$

The probability density function describing the sampling distribution of SL is

$$\begin{aligned}
f_{SL}(q) &= \frac{d}{dq} F_{SL}(q) = -\frac{d}{dq} F^{-1}_{F_{a-1,na-a,\delta}} \left(F^{-1}_{F_{a-1,na-a,0}}(1-q) \right) \\
&\times \frac{d}{dq} F^{-1}_{F_{a-1,na-a,0}}(1-q)\, I_{(0,1)}(q) \\
&= \frac{f_{F_{a-1,na-a,\delta}} \left(F^{-1}_{F_{a-1,na-a,0}}(1-q) \right)}{f_{F_{a-1,na-a,0}} \left(F^{-1}_{F_{a-1,na-a,0}}(1-q) \right)} I_{(0,1)}(q) .
\end{aligned}$$

If the overall effect size $\delta = 0$, then

$$f_{SL}(q) = \frac{f_{F_{a-1,na-a,0}} \left(F^{-1}_{F_{a-1,na-a,0}}(1-q) \right)}{f_{F_{a-1,na-a,0}} \left(F^{-1}_{F_{a-1,na-a,0}}(1-q) \right)} I_{(0,1)}(q) = I_{(0,1)}(q) .$$

In this case, the distribution of the significane level is a Uniform distribution on the interval $(0, 1)$.

4.3 Conclusion

The mean of the distribution of the F test statistic increases as n increases while the variance is finite only depending on the overall effect size δ. We defined the significance level and derived its distribution. When the null hypothesis is true, the distribution of the significance level is a Uniform distribution on the interval $(0, 1)$.

CHAPTER 5

CONFIDENCE INTERVALS

5.1 Introduction

Confidence intervals are recommended in the literature to estimate the parameter of interest. For the population mean μ, the t-interval is commonly used provided the sample is a "large" size sample. It has the form

$$\left(\overline{x} - t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}, \overline{x} + t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}\right),$$

where $\overline{x}$ and s are the respective mean and standard deviation of the observed measurements $x_1, \ldots, x_n$. This interval is the observed value of the random interval

$$\left(\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right)$$

that has a $100\,(1-\alpha)\,\%$ chance of containing the population mean μ. Hence, we say that we are $100\,(1-\alpha)\,\%$ confident that μ is in the observed interval.

The margin of error (MoE) for the t-interval is

$$MoE = t_{n-1,\alpha/2}\frac{s}{\sqrt{n}}.$$

A researcher may have an interest in this margin of error to be no more than a certain amount B. Our interest is to solve the inequality

$$t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} \leq B \text{ or } n \geq \frac{t_{n-1,\alpha/2}^2 s^2}{B^2}.$$

However, since the data has yet to be collected, we do not know the value of s^2. Another approach is to obtain from the researcher a bound B^* on the expected value of the margin of error before the data is collected. That is,

$$E\left(t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right) \leq B^*.$$

As we will see the expected value of the margin of error is a function of σ. We see that

$$E\left(t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right)=\frac{t_{n-1,\alpha/2}}{\sqrt{n}}E\left(S\right).$$

Observe that under the independent Normal model that

$$W=\frac{\left(n-1\right)S^2}{\sigma^2}\sim\chi^2_{n-1},$$

where χ^2_{n-1} is a random variable having (central) Chi Square distribution with $n-1$ degrees of freedom. Next observe that

$$S=\left(\frac{\sigma^2}{n-1}\frac{\left(n-1\right)S^2}{\sigma^2}\right)^{1/2}=\frac{\sigma}{\sqrt{n-1}}W^{1/2}.$$

Hence,

$$\begin{aligned}E\left(S\right)&=\frac{\sigma}{\sqrt{n-1}}E\left(W^{1/2}\right)=\frac{\sigma}{\sqrt{n-1}}\int_0^\infty w^{1/2}f_W\left(w\right)dw\\&=\frac{\sigma}{\sqrt{n-1}}\int_0^\infty w^{1/2}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}w^{(n-1)/2-1}e^{-w/2}dw\\&=\frac{\Gamma\left(\frac{n-2}{2}\right)2^{(n-2)/2}}{\sqrt{n-1}\Gamma\left(\frac{n-1}{2}\right)2^{(n-2)/2-1/2}}\sigma\\&=\frac{\sqrt{2}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{n-1}\Gamma\left(\frac{n-1}{2}\right)}\sigma.\end{aligned}$$

The following function of n

$$c_4=\frac{\sqrt{2}\Gamma\left(\frac{n-2}{2}\right)}{\sqrt{n-1}\Gamma\left(\frac{n-1}{2}\right)}$$

under the independent Normal model is the unbiased constant for the sample standard deviation S, that is, $E\left(S/c_4\right)=\sigma$.

There are two ways to obtain estimates of σ. The first is uses an estimated of σ found in the literature. The second is to take a preliminary sample of size $n_0<n$ and the standard

deviation of this sample s_{n_0} divided by the unbiased constant c_4, s/c_4, to estimate σ. Both methods will result in estimates of the sample size n of the form

$$\frac{t_{n-1,\alpha/2}}{\sqrt{n}}\widehat{\sigma} \leq B^{*}\text{or } n \geq \left\lceil \left(\frac{t_{n-1,\alpha/2}\widehat{\sigma}}{B^{*}}\right)^{2} \right\rceil .$$

5.2 Conclusion

Because of the observed significance level of the t-test decreases as n increases, we recommend using confidence intervals to summarize the data. As the sample size n increases, a confidence interval becomes a more precise interval estimate of the population mean μ.

CHAPTER 6

CONCLUSION

6.1 General Conclusions

We have investigated the meaning of the statistic, which we referred to as the significance level, and derived its distribution for the t-test and the F-test. In both cases, if the null hypothesis is true, then the significance level has a Uniform distribution on the interval $(0, 1)$. We have shown that the observed significance level in both the t-test and F-test becomes smaller as the sample size n increases.

6.2 Areas for Further Research

There are many other applications of tests of significance and tests of hypothesis. For example, there are a variety of non-parametric tests and tests for factorial experiments, as well as other tests of significance and hypothesis. We are interested in studying the distribution of the significance level for each of these tests. Further, we are interested in studying the results found in Beji (1985).

REFERENCES

Bain, L.J. and Engelhardt, M. (1992), Introduction to Probability and Mathematical Statistics, PWS-KENT: Boston.

Behrens, W.V. (1929), "Ein Beitrag zur Fehlerberechung Beiwenigen Beobachtungen," Landwirtsch, Jahrbiicher, 68,807-837.

American Psychological Association. (2018). Practical significance. *In APA Dictionary of Psychology*. Retrieved from https://dictionary.apa.org/practical-significance

Beji, M.A. (1985), "Evidential Approach to Hypothesis Testing," PhD Dissertation, Department of Statistics, University of Missouri-Columbia.

Champ, C.W. and Hu, F. (2024), "An Exact Expression of the Behrens-Fisher Distribution with Applications," accepted for publication in *Applied Mathematical Analysis and Computation 1st SGMC*, Statesboro, GA, USA, April 2-3, 2021.

Cohen J. (1997), "The Earth Is Round ($p < 0.05$)," *The American Psychologist* **49**(**12**), 997–1003.

Concato, J. and Hartigan, J.A. (2016), "P Values: From Suggestion to Superstition," *Journal of Investigative Medicine* **64**, 1166-1171.

DeGroot, M. H. (1975), *Probability and Statistics*, Addison-Wesley: Reading, MA.

Demidenko, E. (2016), "The p-Value You Can't Buy," *The American Statistician* **77**(**1**), 33-38.

Fisher, R. A. (1939), "The Comparison of Samples With Possibly Unequal Variances," *Annals of Eugenics* **9**,174-180.

Fisher, R.A. (1958), *Statistical Methods for Research Workers*, Hafner Publishing Company Inc.: New York.

Harrington,D., D'Agostino, Sr., R.B., Gatsonis, C., Hogan, J.W., Hunter, D.J., Normand, S.T., Drazen, J.M., Hamel M.B. (2019), "New Guidelines for Statistical Reporting in the Journal," *The New England Journal of Medicine* **381**(**3**), 285-286.

Hu, F. (2022), "Methods for Comparing Two Means with Application in Adaptive Clinical Trials," Master's Thesis, Georgia Southern University, Summer 2010.

Jacobson, N.S. Follette W.C., and Revenstorf D. (1984), "Psychotherapy Outcome Research: Methods for Reporting Variability and Evaluating Clinical Significance," *Behavior Therapy* **15**(**4**), 336–52.

Jacobson, N.S. and Truax, P. (1991), "Clinical Significance: a Statistical Approach to Defining Meaningful Change in Psychotherapy Research," *Journal of Consulting and Clin-*

ical Psychology **59**(**1**): 12–19.

Kempthorne, O. and Folks, L. (1971), *Probability, Statistics, and Data Analysis*, The Iowa State University Press: Ames, Iowa.

Kuffner, T.A. and Walker, S.G. (2016), "Why Are p-Values Controversial?," *The American Statistician* **73**(**1**), 1-3.

Kazdin, A.E. (1999), "The Meanings and Measurement of Clinical Significance," *Journal of Consulting and Clinical Psychology* **67**(**3**), 332–339.

Peterson, L.S. (2008), " "Clinical" Significance and "Practical" Significance are NOT the Same Things," a paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 7, 2008, pp. 1-17.

Rozeboom, W.W. (1960), "The Fallacy of the Null-Hypothesis Significance Test," *Psychological Bulletin* **57**(**5**), 416-428.

Scheffe, H. (1959), *The Analysis of Variance*, Wiley: New York.

Wasserstein, R.L. and Lazar, N.A. (2016), "The ASA's Statement on p-Values: Context, Process, and Purpose," *The American Statistician* **70**(**2**), 129–133.

Welch, B.L. (1938), "The Significance of the Difference Between Two Means When the Population Variances Are Unequal," *Biometrika* **29**, 350–362.

Wilkinson L. (1999), “Statistical Methods in Psychology Journals: Guidelines and Explanations,” *American Psychologist* **54**(**8**), 594–604.