

Statistics

End-Term Report

Ronak Gupta

23B2148

Mentor: Ashutosh Gandhe

Statistics is the art of learning from data. It is concerned with the collection of data, its subsequent description, and its analysis, which often leads to the drawing of conclusions.

Statistics is of two types :- descriptive and inferential statistics

Descriptive Statistics is concerned with the description and summarization of data.

Inferential Statistics is concerned with the drawing of conclusions

Central Tendency

Mean :-

The *sample mean*, designated by \bar{x} , is defined by

$$\bar{x} = \sum_{i=1}^n x_i/n$$

Median :-

Order the values of a data set of size n from smallest to largest. If n is odd, the *sample median* is the value in position $(n+1)/2$; if n is even, it is the average of the values in positions $n/2$ and $n/2 + 1$.

Mode :-

Another statistic that has been used to indicate the central tendency of a data set is the *sample mode*, defined to be the value that occurs with the greatest frequency. If no single value occurs most frequently, then all the values that occur at the highest frequency are called *modal values*.

Variance:-

The *sample variance*, call it s^2 , of the data set x_1, \dots, x_n is defined by

$$s^2 = \sum_{i=1}^n (x_i - \bar{x})^2 / (n - 1)$$

We divide by $n-1$ not n in case of sample as it corrects the bias and makes the sample variance closer to population variance.

We can use this identity to further simplify the variance calculation.

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

The computation of the sample variance can also be eased by noting that if

$$y_i = a + bx_i, \quad i = 1, \dots, n$$

then $\bar{y} = a + b\bar{x}$, and so

$$\sum_{i=1}^n (y_i - \bar{y})^2 = b^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

That is, if s_y^2 and s_x^2 are the respective sample variances, then

$$s_y^2 = b^2 s_x^2$$

Standard Deviation :-

The quantity s , defined by

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 / (n-1)}$$

Sample Percentile :-

The *sample 100p percentile* is that data value such that at least $100p$ percent of the data are less than or equal to it and at least $100(1-p)$ percent are greater than or equal to it. If two data values satisfy this condition, then the sample $100p$ percentile is the arithmetic average of these two values.

To determine the sample $100p$ percentile of a data set of size n , we need to determine the data values such that

1. At least np of the values are less than or equal to it.
2. At least $n(1-p)$ of the values are greater than or equal to it.

The sample 25 percentile is called the *first quartile*; the sample 50 percentile is called the sample median or the *second quartile*; the sample 75 percentile is called the *third quartile*.

Chebyshev Inequality :-

Let \bar{x} and s be the sample mean and sample standard deviation of the data set consisting of the data x_1, \dots, x_n , where $s > 0$. Let

$$S_k = \{i, 1 \leq i \leq n : |x_i - \bar{x}| < ks\}$$

and let $|S_k|$ be the number of elements in the set S_k . Then, for any $k \geq 1$,

$$\frac{|S_k|}{n} \geq 1 - \frac{n-1}{nk^2} > 1 - \frac{1}{k^2}$$

But there is a problem, as Chebyshev's inequality holds universally, it might be expected for given data that the actual percentage of the data values that lie within the interval from $\bar{x} - ks$ to $\bar{x} + ks$ might be quite a bit larger than the bound given by the inequality.

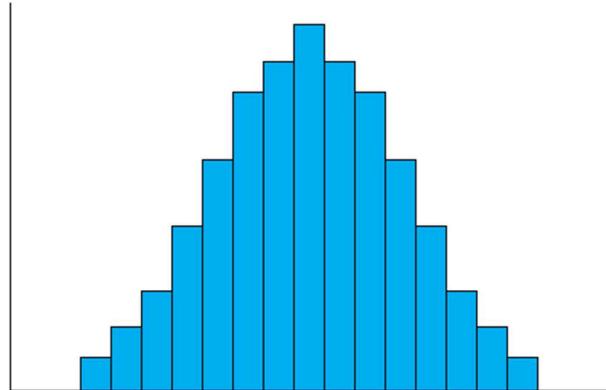
So we often use one-sided Chebyshev Inequality

Let \bar{x} and s be the sample mean and sample standard deviation of the data set consisting of the data x_1, \dots, x_n . Suppose $s > 0$, and let $N(k) = \text{number of } i : x_i - \bar{x} \geq ks$. Then, for any $k > 0$,

$$\frac{N(k)}{n} \leq \frac{1}{1+k^2}$$

Normal Data Sets

Normal Data Sets often reach their peaks at the sample median and then decrease on both sides of this point in a bell-shaped symmetric fashion.



A data set whose histogram has two local peaks is said to be *bimodal*.

Normal Data Sets follow the Empirical Rule

The Empirical Rule

If a data set is approximately normal with sample mean \bar{x} and sample standard deviation s , then the following statements are true.

1. Approximately 68 percent of the observations lie within

$$\bar{x} \pm s$$

2. Approximately 95 percent of the observations lie within

$$\bar{x} \pm 2s$$

3. Approximately 99.7 percent of the observations lie within

$$\bar{x} \pm 3s$$

Correlation coefficient

Consider the data pairs (x_i, y_i) , $i = 1, \dots, n$. and let s_x and s_y denote, respectively, the sample standard deviations of the x values and the y values. The *sample correlation coefficient*, call it r , of the data pairs (x_i, y_i) , $i = 1, \dots, n$ is defined by

$$\begin{aligned} r &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \end{aligned}$$

When $r > 0$ we say that the sample data pairs are *positively correlated*, and when $r < 0$ we say that they are *negatively correlated*.

The following are properties of the sample correlation coefficient.

Properties of r

1. $-1 \leq r \leq 1$
2. If for constants a and b , with $b > 0$,

$$y_i = a + bx_i, \quad i = 1, \dots, n$$

then $r = 1$.

3. If for constants a and b , with $b < 0$,

$$y_i = a + bx_i, \quad i = 1, \dots, n$$

then $r = -1$.

4. If r is the sample correlation coefficient for the data pairs x_i, y_i , $i = 1, \dots, n$ then it is also the sample correlation coefficient for the data pairs

$$a + bx_i, \quad c + dy_i, \quad i = 1, \dots, n$$

provided that b and d are both positive or both negative.

Axioms of Probability

AXIOM 1

$$0 \leq P(E) \leq 1$$

AXIOM 2

$$P(S) = 1$$

AXIOM 3

For any sequence of mutually exclusive events E_1, E_2, \dots (that is, events for which $E_i E_j = \emptyset$ when $i \neq j$),

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i), \quad n = 1, 2, \dots, \infty$$

We call $P(E)$ the probability of the event E .

Generalized Basic Principle of Counting

If r experiments that are to be performed are such that the first one may result in any of n_1 possible outcomes, and if for each of these n_1 possible outcomes there are n_2 possible outcomes of the second experiment, and if for each of the possible outcomes of the first two experiments there are n_3 possible outcomes of the third experiment, and if, \dots , then there are a total of $n_1 \cdot n_2 \cdots n_r$ possible outcomes of the r experiments.

Conditional Probability

In conditional probability if the event F occurs, then in order for E to occur it is necessary that the actual occurrence be a point in both E and F ; that is, it must be in EF .

We calc the conditional probability $P(E|F)$ as

$$P(E|F) = \frac{P(EF)}{P(F)}$$

Bayes' Formula

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)] \end{aligned}$$

Probability of the event E is a weighted average of the conditional probability of E given that F has occurred and the conditional probability of E given that F has not occurred, with each conditional probability being given as much weight as the event it is conditioned on has of occurring.

Suppose there are n events

$$\begin{aligned} P(E) &= \sum_{i=1}^n P(EF_i) & P(F_j|E) &= \frac{P(EF_j)}{P(E)} \\ &= \sum_{i=1}^n P(E|F_i)P(F_i) & &= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^n P(E|F_i)P(F_i)} \end{aligned}$$

Independent Events :-

Independent events are those events whose occurrence is not dependent on any other event.

$$P(EF) = P(E)P(F)$$

Of course we may also extend the definition of independence to more than three events. The events E_1, E_2, \dots, E_n are said to be independent if for every subset $E_{1'}, E_{2'}, \dots, E_{r'}, r \leq n$, of these events

$$P(E_1'E_2' \cdots E_{r'}') = P(E_1')P(E_2') \cdots P(E_{r'}')$$

Random Variables

These quantities of interest that are determined by the result of the experiment are known as random variables. If X is the random variable then

$$1 = P(S) = P\left(\bigcup_{i=2}^{12}\{X = i\}\right) = \sum_{i=2}^{12} P\{X = i\}$$

Random Variable are divided into two parts:- *Discrete* and *Continuous*

Random variables whose set of possible values can be written either as a finite sequence x_1, \dots, x_n , or as an infinite sequence x_1, \dots are said to be **discrete**.

Random variables that take on a continuum of possible values are said to be **continuous**.

The *cumulative distribution function*, or more simply the *distribution function*, F of the random variable X is defined for any real number x by

$$F(x) = P\{X \leq x\}$$

For a discrete random variable X, we define the probability mass function $p(a)$ of X by

$$p(a) = P\{X = a\}$$

The probability mass function $p(a)$ is positive for at most a countable number of values of a . That is, if X must assume one of the values x_1, x_2, \dots , then

$$\begin{aligned} p(x_i) &> 0, & i = 1, 2, \dots \\ p(x) &= 0, & \text{all other values of } x \end{aligned}$$

Since X must take on one of the values x_i , we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

EXAMPLE 4.2a Consider a random variable X that is equal to 1, 2, or 3. If we know that

$$p(1) = \frac{1}{2} \quad \text{and} \quad p(2) = \frac{1}{3}$$

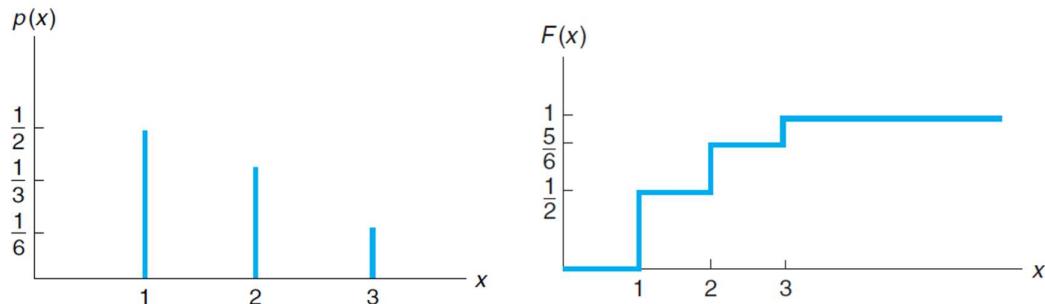
then it follows (since $p(1) + p(2) + p(3) = 1$) that

$$p(3) = \frac{1}{6}$$

A graph of $p(x)$ is presented in Figure 4.1. ■

The cumulative distribution function F can be expressed in terms of $p(x)$ by

$$F(a) = \sum_{\text{all } x \leq a} p(x)$$



We say that X is a continuous random variable if there exists a nonnegative function $f(x)$, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{X \in B\} = \int_B f(x) dx$$

The function $f(x)$ is called the *probability density function* of the random variable X .

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x) dx$$

$$P\{a \leq X \leq b\} = \int_a^b f(x) dx$$

If we let $a = b$ in the above, then

$$P\{X = a\} = \int_a^a f(x) dx = 0$$

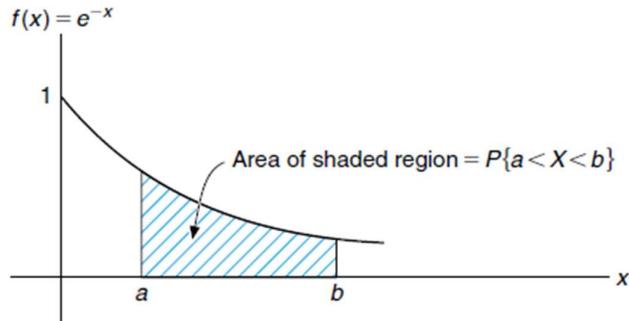
this equation states that the probability that a continuous random variable will assume any particular value is zero.

The relationship between the cumulative distribution $F(\cdot)$ and the probability density $f(\cdot)$ is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^a f(x) dx$$

Differentiating both sides yields

$$\frac{d}{da} F(a) = f(a)$$



Jointly Distributed Random Variables

To specify the relationship between two random variables, we define the joint cumulative probability distribution function of X and Y by

$$F(x, y) = P\{X \leq x, Y \leq y\}$$

distribution function of X — call it FX — can be obtained from the joint distribution function F of X and Y as follows:

$$\begin{aligned} FX(x) &= P\{X \leq x\} \\ &= P\{X \leq x, Y < \infty\} \\ &= F(x, \infty) \end{aligned}$$

Similarly, the cumulative distribution function of Y is given by

$$F_Y(y) = F(\infty, y)$$

If X and Y are both discrete random variables them joint probability mass function of X and Y , $p(x_i, y_j)$, by

$$p(x_i, y_j) = P\{X = x_i, Y = y_j\}$$

$$\begin{aligned} P\{X = x_i\} &= P\left(\bigcup_j \{X = x_i, Y = y_j\}\right) \\ &= \sum_j P\{X = x_i, Y = y_j\} \\ &= \sum_j p(x_i, y_j) \end{aligned}$$

X and Y are jointly continuous if there exists a function $f(x, y)$ defined for all real x and y, having the property that for every set C of pairs of real numbers (that is, C is a set in the two-dimensional plane).

$$P\{(X, Y) \in C\} = \iint_{(x,y) \in C} f(x, y) dx dy$$

The function $f(x, y)$ is called the joint probability density function of X and Y . If A and B are any sets of real numbers, then by defining $C = \{(x, y) : x \in A, y \in B\}$

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x, y) dx dy$$

If X and Y are jointly continuous, they are individually continuous, and their probability density functions can be obtained as follows:

$$\begin{aligned} P\{X \in A\} &= P\{X \in A, Y \in (-\infty, \infty)\} \\ &= \int_A \int_{-\infty}^{\infty} f(x, y) dy dx \\ &= \int_A f_X(x) dx \end{aligned}$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Independent Random Variables

The random variables X and Y are said to be independent if for any two sets of real numbers A and B

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

In other words, X and Y are independent if, for all A and B, the events EA = {X ∈ A} and FB = {Y ∈ B} are independent.

When X and Y are discrete independent random variables,

$$p(x, y) = p_X(x)p_Y(y)$$

In the jointly continuous case, the condition of independence is equivalent to $f(x, y) = f_X(x)f_Y(y)$ for all x, y.

Conditional Distributions

if X and Y are discrete random variables, it is natural to define the conditional probability mass function of X given that Y = y, by

$$\begin{aligned} p_{X|Y}(x|y) &= P\{X = x | Y = y\} \\ &= \frac{P\{X = x, Y = y\}}{P\{Y = y\}} \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

for all values of y such that $p_Y(y) > 0$.

If X and Y have a joint probability density function $f(x, y)$, then the conditional probability density function of X, given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

if X and Y are jointly continuous, then, for any set A,

$$P\{X \in A | Y = y\} = \int_A f_{X|Y}(x|y) dx$$

Expectation

It is the weighted average of the possible values that random variable can take,

For discrete random variables

$$\sum_{i=1}^n x_i p(x_i) = E[X]$$

For continuous random variables

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Properties of expected value

- (a) If X is a discrete random variable with probability mass function $p(x)$, then for any real-valued function g ,

$$E[g(X)] = \sum_x g(x)p(x)$$

- (b) If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g ,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x) dx$$

If a and b are constants, then

$$E[aX + b] = aE[X] + b$$

The expected value of a random variable X , $E[X]$, is also referred to as the mean or the first moment of X . The quantity $E[X^n]$, $n \geq 1$, is called the n th moment of X .

$$E[X^n] = \begin{cases} \sum_x x^n p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^n f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

Expected Value of Sums of Random Variables

If X and Y are random variables and g is a function of two variables

$$\begin{aligned} E[g(X, Y)] &= \sum_y \sum_x g(x, y) p(x, y) && \text{in the discrete case} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy && \text{in the continuous case} \end{aligned}$$

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n]$$

Variance

If X is a random variable with mean μ , then the *variance* of X , denoted by $\text{Var}(X)$, is defined by

$$\text{Var}(X) = E[(X - \mu)^2]$$

Another form is

$$\text{Var}(X) = E[X^2] - (E[X])^2$$

Covariance and Variance of Sums of Random Variables

$$\begin{aligned} \text{Var}(X + X) &= \text{Var}(2X) \\ &= 2^2 \text{Var}(X) \\ &= 4 \text{Var}(X) \\ &\neq \text{Var}(X) + \text{Var}(X) \end{aligned}$$

There is, however, an important case in which the variance of a sum of random variables is equal to the sum of the variances; and this is when the random variables are independent.

The *covariance* of two random variables X and Y , written $\text{Cov}(X, Y)$, is defined by

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

where μ_x and μ_y are the means of X and Y , respectively.

- $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(aX, Y) = a \text{Cov}(X, Y)$
- $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$ (Covariance, like expectation, possesses an additive property.)

$$\text{Cov} \left(\sum_{i=1}^n X_i, Y \right) = \sum_{i=1}^n \text{Cov}(X_i, Y)$$

$$\text{Cov} \left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$$

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \text{Cov}(X_i, X_j)$$

If X and Y are independent random variables, then

$$\text{Cov}(X, Y) = 0$$

and so for independent X_1, \dots, X_n ,

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i)$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Special Random Variables

Bernoulli and Binomial Random Variables

A random variable X is said to be a Bernoulli random variable if its probability mass function is given by

$$P\{X = 0\} = 1 - p$$

$$P\{X = 1\} = p$$

Where X=1 is when outcome is success and X=0 when outcome is failure.
Its expected value is

$$E[X] = 1 \cdot P\{X = 1\} + 0 \cdot P\{X = 0\} = p$$

If there are n independent trials, each with probability of success and failure is p and 1-p respectively. If X represents the number of successes that occur in n trials then X is said to be a binomial random variable with parameters (n,p). The probability mass function of a binomial random variable with parameters n and p is given by

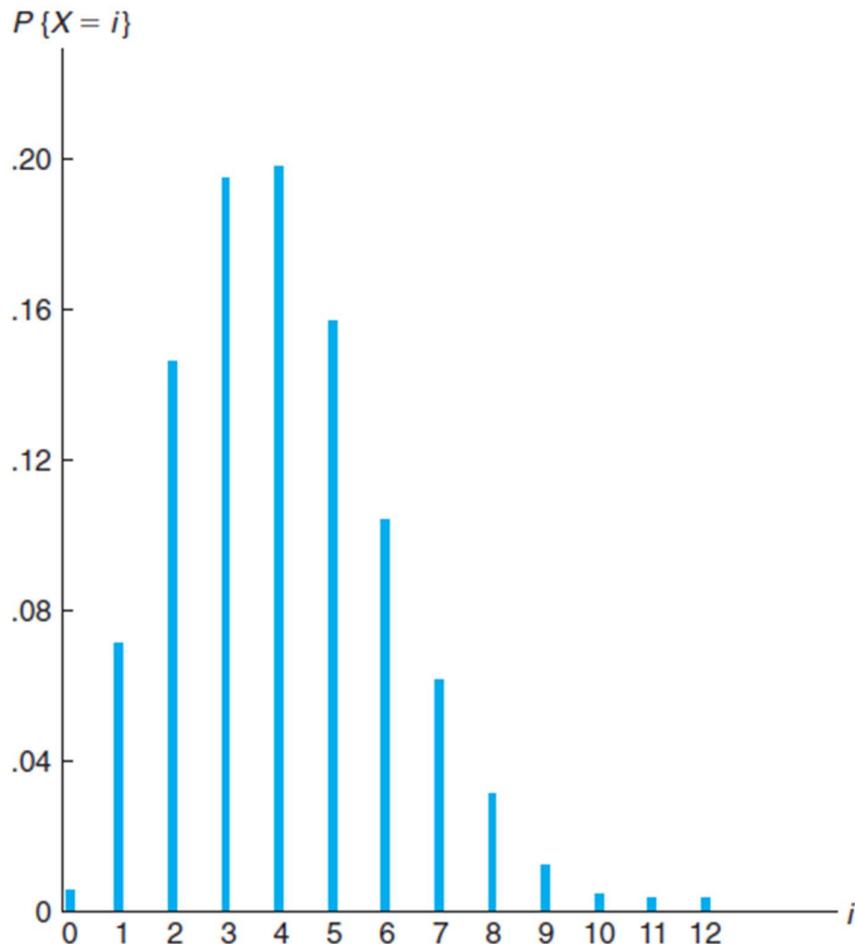
$$P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0, 1, \dots, n$$

The Poisson Random Variable

A random variable X, taking on one of the values 0, 1, 2, . . . , is said to be a Poisson random variable with parameter λ , $\lambda > 0$, if its probability mass function is given by

$$P\{X = i\} = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0, 1, \dots$$

$$\sum_{i=0}^{\infty} p(i) = e^{-\lambda} \sum_{i=0}^{\infty} \lambda^i / i! = e^{-\lambda} e^{\lambda} = 1$$



Poisson probability mass function with $\lambda = 4$.

Both the mean and the variance of a Poisson random variable are equal to the parameter λ .

The Poisson random variable has a wide range of applications in a variety of areas because it may be used as an approximation for a binomial random variable with parameters (n, p) when n is large and p is small.

The Uniform Random Variable

A random variable X is said to be uniformly distributed over the interval $[\alpha, \beta]$ if its probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

The Normal Random Variable

A random variable is said to be normally distributed with parameters μ and σ^2 , and we write $X \sim \mathcal{N}(\mu, \sigma^2)$, if its density is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty^*$$

$$\begin{aligned} E[X] &= \mu \\ \text{Var}(X) &= \sigma^2 \end{aligned}$$

It follows from the foregoing that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma}$$

is a normal random variable with mean 0 and variance 1. Such a random variable Z is said to have a *standard*, or *unit*, normal distribution.

The Exponential Random Variable

A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

The key property of an exponential random variable is that it is memoryless, where we

say that a nonnegative random variable X is *memoryless* if

$$P\{X > s + t \mid X > t\} = P\{X > s\} \text{ for all } s, t \geq 0$$

The Chi - Square Distribution

If Z_1, Z_2, \dots, Z_n are independent standard normal random variables, then X , defined by

$$X = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

is said to have a *chi-square distribution with n degrees of freedom*.

The t Distribution

If Z and χ_n^2 are independent random variables, with Z having a standard normal distribution and χ_n^2 having a chi-square distribution with n degrees of freedom, then the random variable T_n defined by

$$T_n = \frac{Z}{\sqrt{\chi_n^2/n}}$$

is said to have a t-distribution with n degrees of freedom.

The Central Limit Theorem

Let X_1, X_2, \dots, X_n be a sequence of independent and identically distributed random variables each having mean μ and variance σ^2 . Then for n large, the distribution of

$$X_1 + \dots + X_n$$

is approximately normal with mean $n\mu$ and variance $n\sigma^2$.

It follows from the central limit theorem that

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$$

is approximately a standard normal random variable; thus, for n large,

$$P\left\{\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} < x\right\} \approx P\{Z < x\}$$

where Z is a standard normal random variable.

Approximate Distribution of the Sample Mean

$$\bar{X} = \sum_{i=1}^n X_i/n$$

How Large a Sample Is Needed

A general rule of thumb is that one can be confident of the normal approximation whenever the sample size n is at least 30. In most cases, the normal approximation is valid for much smaller sample sizes.

Joint Distribution of \bar{X} and S^2

X and S^2 are independent with $(n - 1) S^2/\sigma^2$ having a chi-square distribution with $n - 1$ degrees of freedom.

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2$$

that

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n (x_i - \mu)^2 - n(\bar{x} - \mu)^2$$

Now, if X_1, \dots, X_n is a sample from a normal population having mean μ and variance σ^2 , then we obtain from the preceding identity that

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$

or, equivalently,

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} + \left[\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \right]^2$$

If X_1, \dots, X_n is a sample from a normal population having mean μ and variance σ^2 , then \bar{X} and S^2 are independent random variables, with \bar{X} being normal with mean μ and variance σ^2/n and $(n - 1)S^2/\sigma^2$ being chi-square with $n - 1$ degrees of freedom.

MAXIMUM LIKELIHOOD ESTIMATORS

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= f_{X_1}(x_1)f_{X_2}(x_2) \cdots f_{X_n}(x_n) \\ &= \frac{1}{\theta} e^{-x_1/\theta} \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta}, \quad 0 < x_i < \infty, i = 1, \dots, n \\ &= \frac{1}{\theta^n} \exp \left\{ - \sum_{i=1}^n x_i / \theta \right\}, \quad 0 < x_i < \infty, i = 1, \dots, n \end{aligned}$$

$f(x_1, \dots, x_n | \theta)$ is often referred to as the *likelihood* function of θ .

In determining the maximizing value of θ , it is often useful to use the fact that $f(x_1, \dots, x_n | \theta)$ and $\log[f(x_1, \dots, x_n | \theta)]$ have their maximum at the same value of θ . Hence, we may also obtain $\hat{\theta}$ by maximizing $\log[f(x_1, \dots, x_n | \theta)]$.

Maximum Likelihood Estimator of a Bernoulli Parameter

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n}$$

Hence, the maximum likelihood estimator of the unknown mean of a Bernoulli distribution is given by

$$d(X_1, \dots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$$

Maximum Likelihood Estimator of a Poisson Parameter

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$$

and so the maximum likelihood estimator is given by

$$d(X_1, \dots, X_n) = \frac{\sum_{i=1}^n X_i}{n}$$

Maximum Likelihood Estimator in a Normal Population

Hence, the maximum likelihood estimators of μ and σ are given, respectively, by

$$\bar{X} \quad \text{and} \quad \left[\sum_{i=1}^n (X_i - \bar{X})^2 / n \right]^{1/2}$$

It should be noted that the maximum likelihood estimator of the standard deviation σ differs from the sample standard deviation

$$S = \left[\sum_{i=1}^n (X_i - \bar{X})^2 / (n - 1) \right]^{1/2}$$

Estimating the Mean of a Uniform Distribution

$$\hat{\theta} = \max(X_1, X_2, \dots, X_n)$$

INTERVAL ESTIMATES

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n} \frac{(\bar{X} - \mu)}{\sigma}$$

$$P \left\{ -1.96 < \sqrt{n} \frac{(\bar{X} - \mu)}{\sigma} < 1.96 \right\} = .95$$

or, equivalently,

$$P \left\{ -1.96 \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96 \frac{\sigma}{\sqrt{n}} \right\} = .95$$

Multiplying through by -1 yields the equivalent statement

$$P \left\{ -1.96 \frac{\sigma}{\sqrt{n}} < \mu - \bar{X} < 1.96 \frac{\sigma}{\sqrt{n}} \right\} = .95$$

or, equivalently,

$$P \left\{ \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right\} = .95$$

That is, 95 percent of the time the value of the sample average \bar{X} will be such that the distance between it and the mean μ will be less than $1.96 \sigma/\sqrt{n}$.

Prediction Intervals

$$\frac{X_{n+1} - \bar{X}_n}{\sigma \sqrt{1 + 1/n}}$$

is a standard normal random variable.

Because this is independent of $S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/(n-1)$, it follows from the same argument used to establish Corollary 6.5.2, that replacing σ by its estimator S_n in the preceding expression will yield a t-random variable with $n-1$ degrees of freedom. That is,

$$\frac{X_{n+1} - \bar{X}_n}{S_n \sqrt{1 + 1/n}}$$

is a t-random variable with $n-1$ degrees of freedom. Hence, for any $\alpha \in (0, 1/2)$,

$$P \left\{ -t_{\alpha/2, n-1} < \frac{X_{n+1} - \bar{X}_n}{S_n \sqrt{1 + 1/n}} < t_{\alpha/2, n-1} \right\} = 1 - \alpha$$

which is equivalent to

$$P \left\{ \bar{X}_n - t_{\alpha/2, n-1} S_n \sqrt{1 + 1/n} < X_{n+1} < \bar{X}_n + t_{\alpha/2, n-1} S_n \sqrt{1 + 1/n} \right\}$$

Hence, if the observed values of \bar{X}_n and S_n are, respectively, \bar{x}_n and s_n , then we can predict, with $100(1-\alpha)$ percent confidence, that X_{n+1} will lie between $\bar{x}_n - t_{\alpha/2,n-1} s_n \sqrt{1 + 1/n}$ and $\bar{x}_n + t_{\alpha/2,n-1} s_n \sqrt{1 + 1/n}$. That is, with $100(1 - \alpha)$ percent confidence, we can predict that

$$X_{n+1} \in \left(\bar{x}_n - t_{\alpha/2,n-1} s_n \sqrt{1 + 1/n}, \bar{x}_n + t_{\alpha/2,n-1} s_n \sqrt{1 + 1/n} \right)$$

Confidence Intervals for the Variance of a Normal Distribution

If X_1, \dots, X_n is a sample from a normal distribution having unknown parameters μ and σ^2 , then we can construct a confidence interval for σ^2 by using the fact that

$$(n-1) \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Hence,

$$P \left\{ \chi_{1-\alpha/2,n-1}^2 \leq (n-1) \frac{S^2}{\sigma^2} \leq \chi_{\alpha/2,n-1}^2 \right\} = 1 - \alpha$$

or, equivalently,

$$P \left\{ \frac{(n-1)S^2}{\chi_{\alpha/2,n-1}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2,n-1}^2} \right\} = 1 - \alpha$$

Hence when $S^2 = s^2$, a $100(1 - \alpha)$ percent confidence interval for σ^2 is

$$\left(\frac{(n-1)s^2}{\chi_{\alpha/2,n-1}^2}, \frac{(n-1)s^2}{\chi_{1-\alpha/2,n-1}^2} \right)$$

APPROXIMATE CONFIDENCE INTERVAL FOR THE MEAN OF A BERNOULLI RANDOM VARIABLE

If we let X denote the number of the n items that meet the standards, then X is a binomial random variable with parameters n and p . Thus, when n is large, it follows by the normal approximation to the binomial that X is approximately normally distributed with mean np and variance $np(1-p)$.

$$\frac{X - np}{\sqrt{np(1-p)}} \sim \mathcal{N}(0, 1)$$

$$P\{\hat{p} - z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}\} \approx 1 - \alpha$$

which yields an approximate $100(1 - \alpha)$ percent confidence interval for p .

HYPOTHESIS TESTING

SIGNIFICANCE LEVELS

Consider a population having distribution F_θ , where θ is unknown, and suppose we want to test a specific hypothesis about θ . We shall denote this hypothesis by H_0 and call it the *null hypothesis*. For example, if F_θ is a normal distribution function with mean θ and variance equal to 1, then two possible null hypotheses about θ are

- (a) $H_0 : \theta = 1$
- (b) $H_0 : \theta \leq 1$

A hypothesis that, when true, completely specifies the population distribution is called a *simple hypothesis*; one that does not is called a *composite hypothesis*.

TESTS CONCERNING THE MEAN OF A NORMAL POPULATION

Case of Known Variance

Suppose that X_1, \dots, X_n is a sample of size n from a normal distribution having an unknown mean μ and a known variance σ^2 and suppose we are interested in testing the null hypothesis

$$H_0 : \mu = \mu_0$$

against the alternative hypothesis

$$H_1 : \mu \neq \mu_0$$

where μ_0 is some specified constant.

Since $\bar{X} = \sum_{i=1}^n X_i/n$ is a natural point estimator of μ , it seems reasonable to accept H_0 if \bar{X} is not too far from μ_0 . That is, the critical region of the test would be of the form

$$C = \{X_1, \dots, X_n : |\bar{X} - \mu_0| > c\} \quad (8.3.1)$$

for some suitably chosen value c .

If we desire that the test has significance level α , then we must determine the critical value c in Equation 8.3.1 that will make the type I error equal to α . That is, c must be such that

$$P_{\mu_0}\{|\bar{X} - \mu_0| > c\} = \alpha \quad (8.3.2)$$

where Z is a standard normal random variable. However, we know that

$$P\{Z > z_{\alpha/2}\} = \alpha/2$$

and so

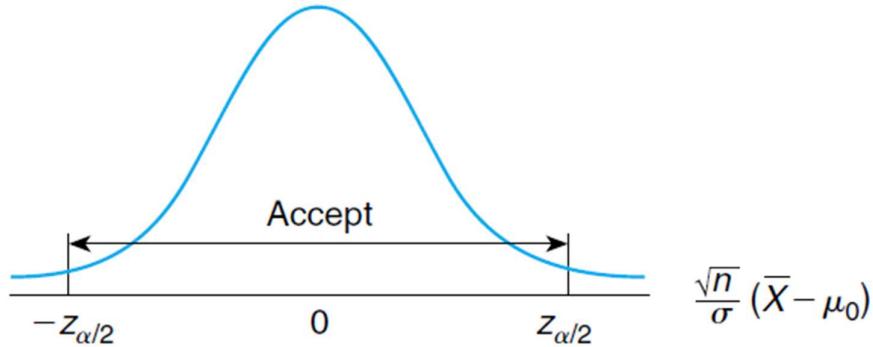
$$\frac{c\sqrt{n}}{\sigma} = z_{\alpha/2}$$

or

$$c = \frac{z_{\alpha/2}\sigma}{\sqrt{n}}$$

Thus, the significance level α test is to reject H_0 if $|\bar{X} - \mu_0| > z_{\alpha/2}\sigma/\sqrt{n}$ and accept otherwise; or, equivalently, to

$$\begin{aligned} \text{reject } H_0 &\quad \text{if } \frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| > z_{\alpha/2} \\ \text{accept } H_0 &\quad \text{if } \frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| \leq z_{\alpha/2} \end{aligned}$$



ONE-SIDED TESTS

In testing the null hypothesis that $\mu = \mu_0$, we have chosen a test that calls for rejection when \bar{X} is far from μ_0 . That is, a very small value of \bar{X} or a very large value appears to make it unlikely that μ (which \bar{X} is estimating) could equal μ_0 . However, what happens when the only alternative to μ being equal to μ_0 is for μ to be greater than μ_0 ? That is, what happens when the alternative hypothesis to $H_0 : \mu = \mu_0$ is $H_1 : \mu > \mu_0$? Clearly, in this latter case we would not want to reject H_0 when \bar{X} is small (since a small \bar{X} is more likely when H_0 is true than when H_1 is true). Thus, in testing

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0$$

we should reject H_0 when \bar{X} , the point estimate of μ_0 , is much greater than μ_0 . That is, the critical region should be of the following form:

$$C = \{(X_1, \dots, X_n) : \bar{X} - \mu_0 > c\}$$

Since the probability of rejection should equal α when H_0 is true (that is, when $\mu = \mu_0$), we require that c be such that

$$P\left\{Z > \frac{c\sqrt{n}}{\sigma}\right\} = \alpha$$

when Z is a standard normal. But since

$$P\{Z > z_\alpha\} = \alpha$$

we see that

$$c = \frac{z_\alpha \sigma}{\sqrt{n}}$$

$$\begin{aligned} \text{accept } H_0 &\quad \text{if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) \leq z_\alpha \\ \text{reject } H_0 &\quad \text{if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) > z_\alpha \end{aligned}$$

This is called a *one-sided* critical region (since it calls for rejection only when \bar{X} is large). Correspondingly, the hypothesis testing problem

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu > \mu_0 \end{aligned}$$

is called a one-sided testing problem (in contrast to the *two-sided* problem that results when the alternative hypothesis is $H_1 : \mu \neq \mu_0$).

Case of Unknown Variance: The *t*-Test

when $\mu = \mu_0$, a *t*-distribution with $n - 1$ degrees of freedom. Hence,

$$P_{\mu_0} \left\{ -t_{\alpha/2, n-1} \leq \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \leq t_{\alpha/2, n-1} \right\} = 1 - \alpha$$

TESTS CONCERNING THE MEAN OF A NORMAL POPULATION

Case of Known Variances

Suppose that X_1, \dots, X_n and Y_1, \dots, Y_m are independent samples from normal populations having unknown means μ_x and μ_y but known variances σ_x^2 and σ_y^2 . Let us consider the problem of testing the hypothesis

$$H_0 : \mu_x = \mu_y$$

versus the alternative

$$H_1 : \mu_x \neq \mu_y$$

Since \bar{X} is an estimate of μ_x and \bar{Y} of μ_y , it follows that $\bar{X} - \bar{Y}$ can be used to estimate $\mu_x - \mu_y$. Hence, because the null hypothesis can be written as $H_0 : \mu_x - \mu_y = 0$, it seems reasonable to reject it when $\bar{X} - \bar{Y}$ is far from zero. That is, the form of the test should be to

$$\begin{aligned} &\text{reject } H_0 \quad \text{if} \quad |\bar{X} - \bar{Y}| > c \\ &\text{accept } H_0 \quad \text{if} \quad |\bar{X} - \bar{Y}| \leq c \end{aligned}$$

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right)$$

which implies that

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim \mathcal{N}(0, 1)$$

Hence, when H_0 is true (and so $\mu_x - \mu_y = 0$), it follows that

$$\begin{aligned} &(\bar{X} - \bar{Y}) \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \\ &P_{H_0} \left\{ -z_{\alpha/2} \leq \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \leq z_{\alpha/2} \right\} = 1 - \alpha \end{aligned}$$

$$\text{accept } H_0 \quad \text{if} \quad \frac{|\bar{X} - \bar{Y}|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \leq z_{\alpha/2}$$

$$\text{reject } H_0 \quad \text{if} \quad \frac{|\bar{X} - \bar{Y}|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \geq z_{\alpha/2}$$

Case of Unknown Variances

Suppose again that X_1, \dots, X_n and Y_1, \dots, Y_m are independent samples from normal populations having respective parameters (μ_x, σ_x^2) and (μ_y, σ_y^2) , but now suppose that all four parameters are unknown. We will once again consider a test of

$$H_0 : \mu_x = \mu_y \quad \text{versus} \quad H_1 : \mu_x \neq \mu_y$$

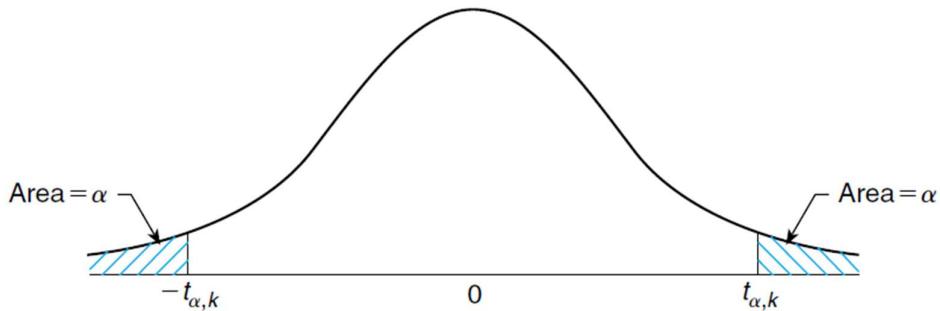
$$\sigma^2 = \sigma_x^2 = \sigma_y^2$$

As before, we would like to reject H_0 when $\bar{X} - \bar{Y}$ is “far” from zero. To determine how far from zero it needs to be, let

$$S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$S_y^2 = \frac{\sum_{i=1}^m (Y_i - \bar{Y})^2}{m-1}$$

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{S_p^2(1/n + 1/m)}} \sim t_{n+m-2}$$



where S_p^2 , the *pooled* estimator of the common variance σ^2 , is given by

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

Hence, when H_0 is true, and so $\mu_x - \mu_y = 0$, the statistic

$$T \equiv \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2(1/n + 1/m)}}$$

has a t -distribution with $n+m-2$ degrees of freedom. From this, it follows that we can test the hypothesis that $\mu_x = \mu_y$ as follows:

$$\begin{aligned} \text{accept } H_0 & \text{ if } |T| \leq t_{\alpha/2, n+m-2} \\ \text{reject } H_0 & \text{ if } |T| > t_{\alpha/2, n+m-2} \end{aligned}$$

Alternatively, the test can be run by determining the p -value. If T is observed to equal v , then the resulting p -value of the test of H_0 against H_1 is given by

$$\begin{aligned} p\text{-value} &= P\{|T_{n+m-2}| \geq |v|\} \\ &= 2P\{T_{n+m-2} \geq |v|\} \end{aligned}$$

where T_{n+m-2} is a t -random variable having $n+m-2$ degrees of freedom.

If we are interested in testing the one-sided hypothesis

$$H_0 : \mu_x \leq \mu_y \quad \text{versus} \quad H_1 : \mu_x > \mu_y$$

then H_0 will be rejected at large values of T . Thus the significance level α test is to

$$\begin{aligned} \text{reject } H_0 & \text{ if } T \geq t_{\alpha, n+m-2} \\ \text{not reject } H_0 & \text{ otherwise} \end{aligned}$$

Testing for the Equality of Variances of Two Normal Populations

Let X_1, \dots, X_n and Y_1, \dots, Y_m denote independent samples from two normal populations having respective (unknown) parameters μ_x, σ_x^2 and μ_y, σ_y^2 and consider a test of

$$H_0 : \sigma_x^2 = \sigma_y^2 \quad \text{versus} \quad H_1 : \sigma_x^2 \neq \sigma_y^2$$

If we let

$$\begin{aligned} S_x^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \\ S_y^2 &= \frac{\sum_{i=1}^m (Y_i - \bar{Y})^2}{m-1} \end{aligned}$$

denote the sample variances, then as shown in Section 6.5, $(n-1)S_x^2/\sigma_x^2$ and $(m-1)S_y^2/\sigma_y^2$ are independent chi-square random variables with $n - 1$ and $m - 1$ degrees of freedom, respectively. Therefore, $(S_x^2/\sigma_x^2)/(S_y^2/\sigma_y^2)$ has an F -distribution with parameters $n - 1$ and $m - 1$. Hence, when H_0 is true

$$S_x^2/S_y^2 \sim F_{n-1, m-1}$$

and so

$$P_{H_0}\{F_{1-\alpha/2, n-1, m-1} \leq S_x^2/S_y^2 \leq F_{\alpha/2, n-1, m-1}\} = 1 - \alpha$$

Thus, a significance level α test of H_0 against H_1 is to

$$\begin{aligned} \text{accept } H_0 & \quad \text{if } F_{1-\alpha/2, n-1, m-1} < S_x^2/S_y^2 < F_{\alpha/2, n-1, m-1} \\ \text{reject } H_0 & \quad \text{otherwise} \end{aligned}$$

The preceding test can be effected by first determining the value of the test statistic S_x^2/S_y^2 , say its value is v , and then computing $P\{F_{n-1, m-1} \leq v\}$ where $F_{n-1, m-1}$ is an F -random variable with parameters $n - 1$, $m - 1$. If this probability is either less than $\alpha/2$ (which occurs when S_x^2 is significantly less than S_y^2) or greater than $1 - \alpha/2$ (which occurs when S_x^2 is significantly greater than S_y^2), then the hypothesis is rejected. In other words, the p -value of the test data is

$$p\text{-value} = 2 \min(P\{F_{n-1, m-1} < v\}, 1 - P\{F_{n-1, m-1} < v\})$$

The test now calls for rejection whenever the significance level α is at least as large as the p -value.

"... a hypothesis test tells us whether the observed data are consistent with the null hypothesis, and a confidence interval tells us which hypotheses are consistent with the data."

Thank You