Tests of Hypotheses

Mya Mya Soe[1]

Abstract

Testing procedures are used in a wide range of industrial processes to monitor quality and in scientific research to help direct a line of inquiry. With any of these tests there is always the risk of an error occurring. In this paper we will discuss about these errors and how to compare different tests.

Keywords: Null hypothesis, Alternative hypothesis, Critical region, Significance level, Type I and type II error

Introduction

The two principle areas of statistical inference are the areas of estimation of parameters and of tests of statistical hypothesis. In this paper some aspects of statistical hypothesis and tests of statistical hypothesis will be considered. The testing is based on random samples, so the result ("yes" or "no") is not definite, but it can be considered a random variable. Traditionally a null hypothesis (denoted by H_0) and alternative hypothesis (denoted by H_1) are presented.

1.1 Definition. A ***statistical hypothesis*** is an assertion about the distribution of one or more random variables. If the statistical hypothesis completely specifies the distribution, it is called a simple statistical hypothesis; if it does not, it is called a composite statistical hypothesis.

1.2 Definition. A ***test*** of a statistical hypothesis is a rule which, when the experimental sample values have been obtained, leads to a decision to accept or to reject the hypothesis under consideration.

1.3 Definition. Let C be that subset of the sample space which, in accordance with a prescribed test, leads to the rejection of the hypothesis under consideration. Then C is called the ***crtical region*** of the test.

1.4 Definition. The ***power function*** of a test of a statistical hypothesis H_0 against an alternative hypothesis H_1 is that function defined for all distributions under consideration, which yields the probability that the sample point falls in the critical region C of the test, that is, a function that yields the probability of rejecting the hypothesis is under consideration. The value of the power function at a parameter point is called the ***power*** of the test at that point.

1.5 Definition. Let H_0 denote a hypothesis that is to the tested against an alternative hypothesis H_1 in accordance with a prescribed test. The ***significance level*** of the test is the maximum value of the power function of the test when H_0 is true.

[1] Lecturer, Department of Mathematics, Dagon University

1.6 Theorem. Let $k > 0$, be a constant and W be a critical region of size α such that

$$W = \left\{ x \in S : \frac{f(x,\theta_1)}{f(x,\theta_0)} > k \right\}$$

$$= \left\{ x \in S : \frac{L_1}{L_0} > k \right\}$$

and $$\bar{W} = \left\{ x \in S : \frac{L_1}{L_0} \le k \right\}$$

where L_0 and L_1 are the likelihood functions of the sample observations $x = (x_1, x_2, \ldots, x_n)$ under H_0 and H_1 respectively. Then W is the most powerful critical region of the test hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$.

We are given

$$P(x \in W | H_0) = \int_w L_0 dx = \alpha .$$

The power of the region is

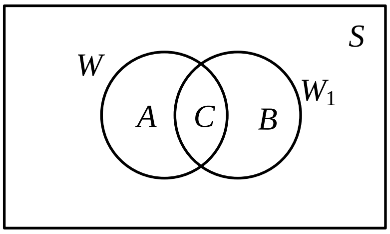
$$P(x \in W | H_1) = \int_w L_1 \, dx = 1 - \beta, \text{ (say)}$$

Let W_1 be another critical region of size $\alpha_1 \le \alpha$ and power $1 - \beta$, so that we have

$$P(x \in W_1 | H_0) = \int_{w_1} L_0 \, dx = \alpha_1$$

and $$P(x \in W_1 | H_1) = \int_{w_1} L_1 \, dx = 1 - \beta_1.$$

Now we have proven that $1 - \beta \ge 1 - \beta_1$.



Let $W = A \cup C$ and $W_1 = B \cup C$

(C may be empty, i.e., W and W_1 may be disjoint). If $\alpha_1 \le \alpha$, we have

$$\int_{w_1} L_0\, dx \le \int_{w} L_0\, dx$$

$$\Rightarrow \quad \int_{B \cup C} L_0\, dx \le \int_{A \cup C} L_0\, dx$$

$$\Rightarrow \quad \int_{B} L_0\, dx \le \int_{A} L_0\, dx$$

$$\Rightarrow \quad \int_{A} L_0\, dx \ge \int_{B} L_0\, dx$$

Since $A \subset W$,

$$\Rightarrow \quad \int_{A} L_1\, dx > k\int_{A} L_0\, dx \ge \int_{B} L_0\, dx$$

$$\Rightarrow \quad \frac{L_1}{L_0} \le k \quad \forall x \in \bar{W}$$

$$\Rightarrow \quad \int_{\bar{W}} L_1\, dx \le k \int_{\bar{W}} L_0\, dx\,.$$

This result also holds for any subset of $\bar{W}$, say $\bar{W} \cap W_1 = B$. Hence

$$\int_{B} L_1\, dx \le k\int_{B} L_0\, dx \le \int_{A} L_1\, dx\,.$$

Adding $\int_{c} L_1\, dx$ to both sides, we get

$$\int_{W_1} L_1\, dx \le \int_{W} L_1\, dx$$

$$\Rightarrow \quad 1-\beta \ge 1-\beta_1\,.$$

1.7 Theorem. Every most powerful or uniformly most powerful critical region is necessarily unbiased.

(i) If W be an most powerful critical region (MPCR) of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, then it is necessarily unbiased.

(ii) Similarly if W be uniformly most powerful critical region (UMPCR) of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, then it is also unbiased.

Since W is an MPCR of size α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$, by Neyman-Pearson Lemma, we have; for $\forall\ k > 0$,

$W = \{x : L\ (x,\ \theta_1) \geq kL\ (x,\ \theta_0) = \{x : L_1 \geq kL_0\}$

and $W' = \{x : L(x,\ \theta_1) < kL(x,\ \theta_0)\} = \{x : L_1 < kL_0\}$,

where k is determined so that the size of the test is α. i.e.,

$$P_{\theta_0}(W) = P\{x \in W \mid H_0\} = \int_W L_0\, dx = \alpha$$

To prove that W is unbiased, we have to show that:

Power of $W \geq \alpha$ i.e., $P_{\theta_1}(W) \geq \alpha$

we have:

$$P_{\theta_1}(W) = \int_W L_1\, dx \geq k\int_w L_0\, dx = k\alpha \quad [\because \text{ on } W,\ L_1 \geq kL_0]$$

i.e., $P_{\theta_1}(W) \geq k\alpha,\ \forall\ k > 0$.

$$\begin{aligned}
1 - P_{\theta_1}(W) &= 1 - P(x \in W \mid H) = P(x \in W' \mid H_1) \\
&= \int_{W'} L_1\, dx \\
&< k\int_{W'} L_0\, dx \\
&= kP(x : x \in W' \mid H_0) \\
&= k[1 - P(x : x \in W \mid H_0)] \\
&= k(1-\alpha)
\end{aligned}$$

i.e., $1 - P_{\theta_1}(W) < k(1-\alpha),\ \forall k > 0$

Case (i) $k \geq 1$ if $k \geq 1$, we get

$$P_{\theta_1}(W) \geq k\alpha \geq \alpha$$

$\Rightarrow$ W is unbiased CR.

Case (ii) $0 < k < 1$. If $0 < k < 1$, then we get:

$$1 - P_{\theta_1}(W) < 1 - \alpha$$

$\Rightarrow$ $P_{\theta_1}(W) > \alpha$

$\Rightarrow$ W is unbiased critical region. Hence most powerful critical region is unbiased.

If W is UMPCR of size α then also the above proof holds if for θ_1 we write θ such that $\theta \in \Theta_1$. So we have

$$P_\theta(W) > \alpha, \quad \forall\ \theta \in \Theta_1$$

$\Rightarrow$ W is unbiased critical region.

2. Hypothesis Testing

A hypothesis is a statement made about the value of a population parameter that we wish to test by collecting evidence in the form of a sample. Procedures which enable us to decide whether to accept or reject hypothesis or to determine whether observed samples differ significantly from expected result are called tested of hypotheses, tests of significance, or rule of decision.

2.1 Type I and Type II errors

In a hypothesis testing problem, there are four possible outcomes; these are described in Table 1.

Table 1. Four possible outcomes

	H_0 is true	H_0 is false
accept H_0	correct	Type II error
reject H_0	Type I error	correct

Two of the four possible outcomes result in a correct decision, while the other two result in an error. The two errors have means:

Type I error = Rejecting H_0 when it's true

Type II error = Accepting H_0 when it's false.

2.2 One and two tailed tests

If the hypothesis test is about population parameter θ, then we test a null hypothesis H_0 which specifies a particular value for θ, against an alternative hypothesis H_1. It is this alternative hypothesis which will indicate whether the test is one-tailed or two-tailed.

A one-tailed test looks either for an increase in the value of a parameter or for a decrease in the value of parameter. If the null hypothesis is of the form $H_0 : \theta = m$ (for some number m), then a one-tailed test is used when the alternative hypothesis of the form $H_1 : \theta > m$, (a definite increase in θ), or when it is of the form $H_1 : \theta < m$, (a definite decrease in θ). A one-tailed test will have a single part to the critical region and one critical value.

2.3 Example. Over a long period of time it has been found that in Link, the ratio of non-vegetarian to vegetarian meals is 2 to 1. In Mr Poem 1, in a random sample of 10 people ordering meals, 1 ordered vegetarian meals. Using a 5% level of significance, test whether or not the proportion of people eating vegetarian meals in Mr Poem's restaurant is different from that of Link.

The proportion of people eating vegetarian meals at Link's is $\frac{1}{3}$ and let p be the proportion of people at Mr Poem 1's that order a vegetarian meal. Let X be the number of people in the sample who are eating vegetarian meal.

The test will be two tailed as we are testing if they are different.

$$H_0 : p = \frac{1}{3} \qquad\qquad H_1 : p \neq \frac{1}{3}$$

if H_0 is true $X \sim B\ (10, \frac{1}{3})$

$$P(X \leq 1) \;=\; P(X = 0) + P(X = 1)$$

$$= 0.104 > 0.025$$

There is insufficient evidence to reject H_0.

There is no evidence that proportion of vegetarian meals at Mr Poem's restaurant is different from that of Link.

2.4 Example. Over a long period of time, Thanda found that the bus taking her to school was late of a rate of 2.5 times per month. In the month following the start of the new summer bus schedules, Thanda finds that her bus is late 6 times. Assuming that the number of times the bus is late each month has a Poisson distribution, test at the 2% level of significance, whether or not the new schedules have changed the frequency with which the bus is late.

Let the random variable X be the number of times the bus is late in a month.

$$H_0 : \lambda = 2.5 \qquad\qquad H_1 : \lambda \neq 2.5$$

Assume H_0 so that $X \sim P_0\ (2.5)$.

Significance level 2%, so significance level in each tail is 1%.

$$P(X \geq 6) \;=\; 1 - \mathrm{P}(X \leq 5)$$

$$= 0.0420 > 0.01$$

There is insufficient evidence at the 2% level to reject H_0, so conclude that the new schedules have not changed the frequency with which the bus is late.

2.5 Example. Accidents used to occur at a certain road at the rate of 6 per month. The residents decided to construct slowdown for traffic. In the month after the slowdown were constructed there was only one accident. Test, at the 5% level of significance, whether there is evidence that the lights have reduced the rate of accidents.

Let the random variable X be the number of accidents in a month.

$$H_0 : \lambda = 6 \qquad\qquad H_1 : \lambda < 6.$$

Assume H_0, so that $X \sim P_0\ (6)$.

Significance level 5%. $P(X \leq 1) = 0.0174 < 0.05$.

Therefore, there is sufficient evidence at the 5% level to reject H_0 and conclude that slowdown have reduced the number of accidents.

2.6 Example. Consider four students were playing a simple game of cards. The game was one of chance so the probability of any particular person winning should have been $\frac{1}{4}$. After playing a number of games Mg Aye complained that Mg Hla must have been cheating as the kept winning. Their teacher quickly intervented and decided to carry out a proper investigation and carefully watched the next 20 games. We discuss a critical region for one-tail test using a

5% level of significance. First state the hypotheses. If Mg Hla is cheating then you would expect the proportion of games he wins to be more than $\frac{1}{4}$.

$$H_0 : p = \frac{1}{4} \qquad H_1 : p > \frac{1}{4}$$

Let X be the number of games Mg Hla wins out of the next 20.

So $X \sim B\,(20, \frac{1}{4})$.

Reject H_0 if $X \geq c$ where $P\,(X \geq c) < 0.05$.

$P(X \leq 8) = 0.9591$ so $P\,(X \geq 9)=0.0409$.

$P(X \leq 7) = 0.8982$ so $P\,(X \geq 8)=0.1018$.

So the critical region is $X \geq 9$.

So if Mg Hla won 9 or more games there would be evidence to suggest that he was cheating.

P(type I error)= P(Rejecting H_0 when H_0 is true)

$= P(X \geq 9 \mid X\sim B(20,0.25))$

$= 0.0409$

P(type II error)= P(Accept H_0 when H_0 is false)

$= P(X \leq 8 \mid H_0$ is false)

Given that $p = 0.35$

P(type II error)= $P(X \leq 8 \mid X\sim B(20,0.35))$

$= 0.7624$

2.7 Example. We consider another form of the alternative hypothesis that occurs when a chain either up or down. In such cases a two-tailed test is used. If a coin is tossed 20 times and a head is obtained on 7 occasions. We discuss whether or not the coin is biased and the probability of a type I error for this test. If the coin is biased and that this bias causes the tail to appear 3 times for each head that appear then we also discuss the probability of a type II error for the test.

This is a test for the proportion of a binomial distribution, and since we are testing to see if the coin is biased in either direction, a two-tailed test has to be used. The critical will be in two parts.

The hypothesis is

$$H_0 : p = 0.5,\ H_1 : p \neq 0.5$$

Let X be the number of heads in 20 tosses of the coin.

Assuming H_0 is true then $X \sim B\,(20, 0.5)$.

Then two-tailed test at the 5% significance and we require values c_1 and c_2 so that

$P(X \leq c_1) \leq 0.025$ and $P(X \geq c_2) \leq 0.025$

$P(X \leq c_2 - 1) \geq 0.975$

$P(X \leq 6) = 0.0577 \quad , \quad P(X \leq 5) = 0.0207$

The value of $c_1 = 5$.

$P(X \geq 14) = 0.0577 \quad , \quad P(X \geq 15) = 0.0207$.

So the value of $c_2 = 15$.

Thus the critical region for X is $X \leq 5$ or $X \geq 5$.

Thus efficient evidence to reject H_0.

The coin is not biased.

Type I error occurs when we reject H_0, and bias occurs when $X \leq 5$ and $X \geq 15$.

$P(\text{Type I error}) = P(X \leq 5 \mid p = 0.5) + P(X \geq \mid p = 0.5)$

$= 0.0414$

A type II error occurs when we do not have efficient evidence to reject H_0 when H_1 is true. We do not have evidence to reject H_0 is $X \geq 6$ and $X \leq 14$.

$P(\text{type II error}) = P(6 \leq X \leq 14 \mid p = 0.25)$

$= P(X \leq 14 \mid p = 0.25) - P(X \leq 5 \mid p = 0.25)$
$= 0.3873$.

Conclusion

For any fixed significance level, an increase in the sample size will cause a decrease in the probability of type II error.

For any fixed sample size, a decrease in the probability of type I error will cause an increase in probability of type II error. Conversely, an increase in the probability of type I error will cause a decrease in probability of type II error.

To decrease both the probability of type I error and probability of type II error, increase the sample size.

Acknowledgements

I am most deeply indebted to Dr. Daw Nu Nu Yee, Pro-Rector of Dagon University and Dr. Daw Nay Thwe Kyi, Pro-Rector of Dagon University for their permission to write this research and for their constant encouragement to complete it.

I would like to extend my deepest thanks to the Professor Dr. Daw Hla Hla Kyi, Head of Department of Mathematics, Professor Dr. Daw Hla Hla Yi, Department of Mathematics, for the time and energy, they sacrificed for my research and for their suggestion without which nothing true and worthwhile is possible.

Special thanks are also due to all those who supported me in writing this research.

References

S.C. Gupta & V.K. Kapoor, ***Fundamentials of Mathematical Statistics***, Educational Publishers, Tenth Revised edition, Sultan Chand & Sons, New Delhi,2002.

R.V. Hogg & E.A. Tanis, ***Probalitily and Statistics Inference***, Third Edition, Macmillan Publishing Co. Inc., 1983.

R.V. Hogg & A.T. Craig, ***Introduction to Mathematical Statistics***, Fourth Edition, Macmillan Publishing Co. Inc., 1978.

R. Spiegel, ***Theory and Problems of Statistics***, Mcgraw. Hill boot Co.lit, New York, 1961.