CHAPTER 7.

NONPARAMETRIC STATISTICS

Introduction

Most of the hypothesis-testing procedures discussed in previous chapters are based on the assumption that the random samples are selected from normal populations. Fortunately, most of these tests are still reliable when the sample size is large. Traditionally, these testing procedures have been referred to as parametric methods. In this chapter, there is a number of alternative test procedures, called nonparametric or distribution-free methods, that often assume no knowledge whatsoever about the distributions of the underlying populations, except perhaps that they are continuous. If a parametric and a nonparametric test are both applicable to the same set of data, we should carry out the more efficient parametric technique.

SIGN TEST

The sign test is used to test hypotheses on a population median. In the case of many of the nonparametric procedures, the mean is replaced by the median as the pertinent location parameter under test. The population counterpart, denoted by $\tilde{\mu}$, has an analogous definition. Given a random variable X, $\tilde{\mu}$ is defined such that $P(X > \tilde{\mu}) \leq 0.5$ and $P(X < \tilde{\mu}) \leq 0.5$. In the continuous case,

$$P(X < \tilde{\mu}) = P(X > \tilde{\mu}) = 0.5$$

Of course, if the distribution is symmetric, the population mean and median are equal. In testing the null hypothesis H_0 that $\tilde{\mu} = \tilde{\mu}_0$ against an appropriate alternative, on the basis of a random sample of size n, it is replaced each sample value exceeding $\tilde{\mu}_0$ with a plus sign and each sample value less than $\tilde{\mu}_0$ with a minus sign. If the null hypothesis is true and the population is symmetric, the sum of the plus signs should be approximately equal to the sum of the minus signs. When one sign appears more frequently than it should be based on chance alone, so it can reject the hypothesis that the population median $\tilde{\mu}$ is equal to $\tilde{\mu}_0$. The types of hypothesis are for $n \leq 10$ and $np = 5$:

Null hypothesis	Alternative Hypothesis	Critical Region (using p-value)
$H_0: \tilde{\mu} = \tilde{\mu}_0$	$H_1: \tilde{\mu} < \tilde{\mu}_0$	$P = P(X \leq x \mid p = 0.5) = \sum_{i=0}^{x} b\left(x; n; \frac{1}{2}\right) < \alpha$
	$H_1: \tilde{\mu} > \tilde{\mu}_0$	$P = P(X \geq x \mid p = 0.5) = 1 - \sum_{i=0}^{x-1} b\left(x; n; \frac{1}{2}\right) < \alpha$
	$H_1: \tilde{\mu} \neq \tilde{\mu}_0$	$P = 2P(X \leq x \mid p = 0.5) = \sum_{i=0}^{x} b\left(x; n; \frac{1}{2}\right) < \alpha$
$H_0: \tilde{\mu}_1 - \tilde{\mu}_2 = 0$	$H_1: \tilde{\mu}_1 - \tilde{\mu}_2 < 0$	$P = P(X \leq x \mid p = 0.5) = \sum_{i=0}^{x} b\left(x; n; \frac{1}{2}\right) < \alpha$
	$H_1: \tilde{\mu}_1 - \tilde{\mu}_2 > 0$	$P = P(X \geq x \mid p = 0.5) = 1 - \sum_{i=0}^{x-1} b\left(x; n; \frac{1}{2}\right) < \alpha$
	$H_1: \tilde{\mu}_1 - \tilde{\mu}_2 \neq 0$	$P = 2P(X \leq x \mid p = 0.5) = \sum_{i=0}^{x} b\left(x; n; \frac{1}{2}\right) < \alpha$

$\sum_{i=0}^{x} b\left(x; n; \frac{1}{2}\right) \rightarrow$ see in Walpole in appendix A table A1.

Whenever $n > 10$, binomial probabilities with $p = 0.5$ can be approximated from the normal curve, since $np = nq > 5$, the statistical test is:

$$Z_{score} = \frac{x \pm 0.5 - np}{\sqrt{n.p.q}}$$

And the critical region is:

$P(Z < Z_{score}) < \alpha \rightarrow " < "$

$P(Z > Z_{score}) > \alpha \rightarrow " > "$

$2P(Z < |Z_{score}|) < \alpha \rightarrow " \neq "$

SIGNED-RANK TEST

The Sign test utilizes only the plus and minus signs of the differences between the observations and $\tilde{\mu}_0$ in the one-sample case, or the plus and minus signs of the differences between the pairs of observations in the paired-sample case; it does not take into consideration the magnitudes of these differences. A test utilizing both direction and

magnitude, proposed in 1945 by Frank Wilcoxon, is now commonly referred to as the Wilcoxon signed-rank test. It is also called the Wilcoxon Test.

The various test procedures for both the single- and paired-sample cases are summarized in table below:

H_0	H_1	Compute
$\tilde{\mu} = \tilde{\mu}_0$	$\tilde{\mu} < \tilde{\mu}_0$	w_+
	$\tilde{\mu} > \tilde{\mu}_0$	w_-
	$\tilde{\mu} \neq \tilde{\mu}_0$	w
$\tilde{\mu}_1 = \tilde{\mu}_2$	$\tilde{\mu}_1 < \tilde{\mu}_2$	w_+
	$\tilde{\mu}_1 > \tilde{\mu}_2$	w_-
	$\tilde{\mu}_1 \neq \tilde{\mu}_2$	w

It is not difficult to show that whenever n < 5 and the level of significance does not exceed 0.05 for a one-tailed test or 0.10 for a two-tailed test, all possible values of $w+, w-, or\ w$ will lead to the acceptance of the null hypothesis. However, when $5 \leq n \leq 30$, Table A.16 shows approximate critical values of $w +\ and\ w -$ for levels of significance equal to 0.01, 0.025, and 0.05 for a one-tailed test and critical values of W for levels of significance equal to 0.02, 0.05, and 0.10 for a two-tailed test. The null hypothesis is rejected if the computed value $w+, w-, or\ w$ is less than to the appropriate tabled value (Table A.16 in Walpole).