Discussion section 9

Math 181B

1 Review

1.1 The Sign test

The Sign test is a non-parametric test whose null hypothesis is about checking if the median of a distribution is equal to some specific value. The idea is simple: The number of samples greater than the median should have a binomial distribution.

More specifically, let $y_1, y_2, \ldots, y_n$ be a random sample of size n from any continuous distribution having median $\tilde{\mu}$, where $n \geq 10$. Let k denote the number of y_i 's greater than $\tilde{\mu}_0$, and let $z = \frac{k-n/2}{\sqrt{n/4}}$.

- To test $H_0 : \tilde{\mu} = \tilde{\mu}_0$ versus $H_1 : \tilde{\mu} > \tilde{\mu}_0$ at the α level of significance, reject H_0 if $z \geq z_\alpha$.
- To test $H_0 : \tilde{\mu} = \tilde{\mu}_0$ versus $H_1 : \tilde{\mu} < \tilde{\mu}_0$ at the α level of significance, reject H_0 if $z \leq -z_\alpha$.
- To test $H_0 : \tilde{\mu} = \tilde{\mu}_0$ versus $H_1 : \tilde{\mu} \neq \tilde{\mu}_0$ at the α level of significance, reject H_0 if z is either (1) $\leq -z_{\alpha/2}$ or (2) $\geq z_{\alpha/2}$.

When $n < 10$, the normal approximation to binomial is not appropriate. Therefore, we need to use the exact binomial distribution to compute the p-value and then make a conclusion about the hypothesis test.

1.2 Wilcoxon Tests

1.2.1 Wilcoxon Signed Rank Test

Wilcoxon test is another non-parametric test that can be used to test location (mean of one sample, mean of two samples), etc. If we define the Wilcoxon signed rank statistic, w as

$$w = \sum_{i=1}^{n} r_i z_i;$$

where

$$z_i = \begin{cases} 0 & \text{if } y_i - \mu_0 < 0 \\ 1 & \text{if } y_i - \mu_0 > 0 \end{cases},$$

and r_i are ranks of $|y_1 - \mu_0|, |y_2 - \mu_0|, \ldots, |y_n - \mu_0|$, where μ_0 is the mean under the null hypothesis.

If $y_1, y_2, \ldots, y_n$ be a set of independent observations drawn, respectively, from the continuous and symmetric (but not necessarily identical) pdfs $f_{Y_i}(y), i = 1, 2, \ldots, n$. Suppose that each of the $f_{Y_i}(y)$'s has the same mean μ. Under the null hypothesis $H_0 : \mu = \mu_0$, $p_W(w)$ can be calculated as

$$p_W(w) = P(W = w) = \left(\frac{1}{2^n}\right) \cdot c(w)$$

where $c(w)$ is the coefficient of e^{wt} in the expansion of

$$\prod_{i=1}^{n} \left(1 + e^{it}\right).$$

When the sample size is large $n > 12$, we can get a normal approximation, i.e.,

$$\frac{W - [n(n+1)]/4}{\sqrt{[n(n+1)(2n+1)]/24}}$$

can be adequately approximated by a normal distribution.

Therefore, we can accomplish the hypothesis test based on the exact probability mass function $p_W(w)$ or the normal distribution. For example, when the sample size is large, let

$$z = \frac{w - [n(n+1)]/4}{\sqrt{[n(n+1)(2n+1)]/24}}$$

- To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ at the α level of significance, reject H_0 if $z \geq z_\alpha$.
- To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ at the α level of significance, reject H_0 if $z \leq -z_\alpha$.
- To test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ at the α level of significance, reject H_0 if z is either (1) $\leq -z_{\alpha/2}$ or (2) $\geq z_{\alpha/2}$.

1.2.2 Wilcoxon Rank Sum Test

Wilcoxon Rank Sum Test can be used to test $H_0 := \mu_x = \mu_y$ for two continuous distributions which have the same shape and the same standard deviation, but they may differ with respect to location.

More specifically, let $x_1, x_2, \ldots, x_n$ and $y_{n+1}, y_{n+2}, \ldots, y_{n+m}$ be two independent random samples from $f_X(x)$ and $f_Y(y)$, respectively, where the two pdfs are the same except for a possible shift in location. Let r_i denote the rank of the ith observation in the combined sample (where the smallest observation is assigned a rank of 1 and the largest observation, a rank of $n + m$). Let

$$w' = \sum_{i=1}^{n+m} r_i z_i$$

where z_i is 1 if the i-th observation comes from $f_X(x)$ and 0 , otherwise. Then

$$E\left(W'\right) = \frac{n(n+m+1)}{2}$$
$$\operatorname{Var}\left(W'\right) = \frac{nm(n+m+1)}{12}$$

and

$$\frac{W' - n(n+m+1)/2}{\sqrt{nm(n+m+1)/12}}$$

is approximately a standard normal distribution if $n > 10$ and $m > 10$.