Comparing Parametric and Non-parametric Tests

Rahim Jutha

March 27 2020

1 Introduction

To effectively make inference on data, we use hypothesis testing to measure and evaluate claims about population parameters. Hypothesis tests classify as either parametric or non-parametric. According to Sheskin (2011), the "distinction is generally predicated on the number and severity of assumptions regarding the population". Sheskin (2011) suggests that a reason for preferring a parametric test is they generally have greater power than non-parametric tests. This means that parametric tests are more likely to reject a false null hypothesis. However, if one of the underlying assumptions is violated a non-parametric test will yield a more reliable analysis (Sheskin, 2011). To highlight the differences of parametric and non-parametric tests, as well as when one is better than the other, I will compare the Student's t-Test and the Wilcoxon-Signed-Rank Test.

2 Hypothesis Testing and P-values

For clarification, a hypothesis test starts with an assumption about a population parameter called the null hypothesis, denoted by H_0. We define another hypothesis, called the alternative hypothesis which is the opposite of what is stated in H_0. The alternative hypothesis is denoted by H_a. The procedure involves calculating a test statistic using data from a random sample. Then assuming H_0 is true, how well does the statistic calculated with the sample follow H_0. Evans (2010) describes the procedure as "measuring how surprising the observed [statistic] is when we assume H_0 to be true" (Evans, 2010). If the statistic calculated is surprising under H_0, then there is evidence against H_0. We evaluate this by calculating a p-value.

A p-value is a probability, ranging from 0 to 1. Evans (2010) notes that small p-values indicate to us that a surprising event has occurred if the null hypothesis Ho was true. However, a large p-value is not evidence that Ho is true. Furthermore, a p-value is not the probability that the null hypothesis is true (Evans, 2010). The p-value gives us a way to evaluate claims, like a null hypothesis on population parameters.

3 The Student's t-Test

Student's t-Test, also known as the One-Sample t-Test, was introduced by William Sealy Gosset under the pen name Student (1908). It is a parametric test, meaning it relies on assumptions about the population. The assumption the test make is that our sample comes from normal population. The Student's t-Test is used for testing a statistical hypothesis about the mean μ of a normal population whose population variance σ^2 is unknown (Evans, 2010). In the test, the null hypothesis is defined as $H_0 : \mu = \mu_0$ and the alternate hypothesis can take the forms, $H_a : \mu \neq \mu_0$, $\mu \leq \mu_0$, or $\mu \geq \mu_0$.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{1}$$

The goal of a Student's t-Test is to determine the mean of the population μ using the T-statistic. Suppose we have a large random sample $x_1, \ldots, x_n$ of size n from normal population with unknown mean μ and unknown variance σ^2, where $\sigma > 0$. We use the sample mean, denoted to calculate a Z-statistic or z-score. The Z-Statistic is given by the formula

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}, \tag{2}$$

which is known to have a $N(0, 1)$ distribution (see Evans, 2010). By the Central Limit Theorem Evans (2010), when the sample size is large the sample mean's distribution approaches a normal distribution which means we can apply the test even though the population cannot be assumed to be normal. Also, since the sample size is large we can substitute the sample variance S^2, given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \overline{X})^2, \tag{3}$$

for σ^2 into equation 2 and it becomes the T-Statistic

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}, \tag{4}$$

also known as Student's t-variable (Kalpić et al., 2011). It can be shown that if H_0 is true, then T has a Student's t-distribution with degrees of freedom $n - 1$ (see Evans, 2010). The t-distribution is centered at 0, with regions of low probability are given by the tails. We use this test statistic to access the hypothesis "computing the probability of observing a value as far or farther away from 0 as" equation 4 (Evans, 2010). By this the p-value for the $H_a : \mu \neq \mu_0$ is given by

$$P_{(\mu_0,\sigma^2)}\left(|T| \geq \left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right|\right) = 2\left[1 - \theta\left(\left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right|, n-1\right)\right], \tag{5}$$

where $\theta(., n-1)$ is the distribution function of the $t(n-1)$ distribution. With this p-value we now have evidence against H_0 whenever it is small. This procedure is the Student's t-test.

4 The Wilcoxon-Signed-Rank Test

The Wilcoxon-Signed-Rank Test was proposed in 1945 by Frank Wilcoxon (1945). It is a non-parametric test for the one-sample location problem and can be used to test the hypothesis that the median, denoted $\tilde{\mu}$, of a symmetrical distribution equals a given constant $\tilde{\mu}_0$ (Rey and Neuhäuser, 2011). The median, which is the midpoint, or the 50^{th} percentile, of a distribution helps us to make inference its symmetry. The assumption in this test is that the population distribution is symmetric, meaning the median of the distribution is equal to its mean. Since all normal distributions are symmetric, assuming symmetry is weaker than assuming normality(Devore, 2015). In this test, the null hypothesis is $H_0 : \tilde{\mu} = \tilde{\mu}_0$ and the alternate hypothesis is $H_a : \tilde{\mu} \neq \tilde{\mu}_0$, $\tilde{\mu} \leq \tilde{\mu}_0$, or $\mu \geq \mu_0$.

The goal of the Wilcoxon-Signed-Rank Test is to make claims about the symmetry of the population's distribution. Suppose we have a random sample $x_1, \ldots, x_n$ of size n from a continuous and symmetric distribution with mean μ and median $\tilde{\mu}$, where $\mu = \tilde{\mu}$. To get the test statistic for this hypothesis test we need to convert our sample into ranked data. The first step to get our ranked data is to subtract our μ_0 from each observation to scale them. The next step is to assign the rank to each observation by ordering them by absolute magnitudes, ignoring sign, with the smallest getting rank 1, the second smallest rank 2, and so on (Devore, 2015). Now apply the sign of each observation to each corresponding rank to obtain the signed ranks. Finally we can calculate our test statistic S_+, which is given by

$$S_+ = \text{the sum of the ranks associated with positive } \left(x_i - \mu_0\right)' \text{'s.} \tag{6}$$

Now that we have our test statistic S_+ we can calculate a p-value. Devore (2015) explains, intuitively large magnitudes of S_+ signify large clusters of same-signed ranks at larger rank values, which indicate non-symmetry. Under the null hypothesis, we expect S_+ to be close to zero since each rank has an equal chance to be positive or negative. Thus we have a way to get the distribution of S_+ under H_0, which leads to our p-value formula. The p-value for the $H_a : \tilde{\mu} \neq \tilde{\mu}_0$ is given by

$$2P_{(s_+)}\left(S_+ \geq \max\left\{s_+, n(n+1)/2 - s_+\right\}\right), \tag{7}$$

where S_+ is the distribution of the test statistic under H_0, s_+ is our observed test statistic, $\max\{\}$ picks the greater argument, and $n(n+1)$ represents the maximum value s_+ can take (the case when every rank in our sample is positive) (Devore, 2015). This gives us a way to calculate an exact p-value, however this method is complex so a continuity correction is commonly used when performing a Wilcoxon-Signed-Rank Test. When $n \geq 20$, it can be shown that S_+ has an approximately normal distribution with mean μ_{S_+} and variance σ_{S_+}, given by

$$\mu_{S_+} = \frac{n(n+1)}{4}, \quad \sigma^2_{S_+} = \frac{n(n+1)(2n+1)}{24}, \tag{8}$$

when H_0 is true (Devore, 2015). This result comes from noting that when H_0 is true (the symmetric distribution is centered at μ_0), then each rank is just as likely to receive a + sign as it is to receive a − sign (Devore, 2015). Thus S_+ is now given by,

$$S_+ = W_1 + W_2 + W_3 + \ldots + W_n, \tag{9}$$

where

$$W_1 = \begin{cases} 1 & \text{with probability .5} \\ 0 & \text{with probability .5} \end{cases} \cdots W_n = \begin{cases} n & \text{with probability .5} \\ 0 & \text{with probability .5} \end{cases} . \tag{10}$$

S_+ is then a sum of random variables, and when H_0 is true, the W_i's can be shown to be independent and shown to be normal. (see Devore, 2015). These results gives the large-sample test statistic Z_0 which is a z-score for a $H_a : \tilde{\mu} \neq \tilde{\mu}_0$ test is given by

$$Z_0 = \frac{\max\{s_+, n(n+1)/2 - s_+\} - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}. \tag{11}$$

With this value we can calculate a p-value using the the normal distribution given by

$$P(|Z| \geq Z_0) \approx 2\left[1 - \Phi(Z_0)\right], \tag{12}$$

where $\Phi(Z_0)$ denotes the $N(0,1)$ distribution function. With this p-value we now have evidence against H_0 whenever it is small. This procedure is the Wilcoxon-Signed-Rank Test.

5 Similarities and Differences

The Student's t-test and the Wilcoxon-Signed-Rank Test both have their own pros and cons when it comes to deciding which one to use. The first difference between the tests is the type of data present. Sheskin (2011) suggests that it is common knowledge that ratio and interval data, which a parametric test like the t-test can be used on, contain more information, whereas non-parametric tests like the signed-rank test is commonly associated with ranked categorical data. Another reason using a parametric test over a non-parametric test is they have greater power (Sheskin, 2011). This means that the t-test will usually compute a smaller p-value than the signed-rank test, which means it is more likely to reject a false null hypothesis. However if one of the assumptions of a parametric, like normality in the t-test, has been violated, a non-parametric test will provide a more reliable analysis (Sheskin, 2011). As a result, the parametric Student's t-test and the non-parametric Wilcoxon-Signed-Rank Test make inference on the same statistic, but both have their own unique situations as to when one is better than the other and both are commonly used to make analysis.

References

Devore, J. (2015). *Probability and Statistics for Engineering and the Sciences.* Cengage Learning.

Evans, M. M. J. (2010). *Probability and statistics : the science of uncertainty / Michael J. Evans and Jeffrey S. Rosenthal.* W.H. Freeman and Co., 2nd ed. edition.

Kalpić, D., Hlupić, N., and Lovrić, M. (2011). *Student's t-Tests*, pages 1559–1563. Springer Berlin Heidelberg, Berlin, Heidelberg.

Rey, D. and Neuhäuser, M. (2011). *Wilcoxon-Signed-Rank Test*, pages 1658–1659. Springer Berlin Heidelberg, Berlin, Heidelberg.

Sheskin, D. J. (2011). *Parametric Versus Nonparametric Tests*, pages 1051–1052. Springer Berlin Heidelberg, Berlin, Heidelberg.

Student (1908). The probable error of a mean. *Biometrika*, 6(1):1–25.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.