Statistical Hypothesis Testing

Part 2

Statistical Hypothesis

A hypothesis is a statement about a population.

In statistical modelling it is assumed that the phenomenon of interest is described by a RV X whose distribution depends on some parameter θ that is unknown i.e.

$$X \sim f(x, \theta)$$

$\theta:$ parameter of interest

The hypothesis takes the form of a prediction that the parameter θ takes a particular numerical value or falls in a certain range of values.

Test on a statistical hypothesis for a population parameter θ – general procedure

- **Set** the null hypothesis about a population parameter, i.e. $H_0: \theta = \theta_0$
  and the alternative hypothesis that contains alternative parameter values from the value/values in H_0, i.e. $H_1: \theta > \theta_0$
- **Calculate** the test statistics on the sample data, i.e $T(x_1, \ldots, x_n; \theta_0) = t_{obs}$
  (It is supposed that a random sample is available)
- **Calculate** the p-value under the assumption that H_0 is true
  i.e. $p - value = P_{\mathbf{H_0}}(T(x_1, \ldots, x_n; \theta_0) > t_{obs})$, assuming that H_0 is rejected for large values
- **Take a decision**: the smaller the p-value is, the stronger the evidence against H_0 and in favour of H_1
  Reject the NULL hypothesis if **p-value<α** where α is fixed in advance (level of the test)

Test on the mean of a not Normal population

X ~ *f(x)* with $E(X) = \mu$ and $Var(X) = \sigma^2$

$x_1, \ldots, x_n$: random sample with $n \gg 0$

$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$

Test statistics: $Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ (same previous case)

- If H_0 is true and $n \gg 0$, Z is **approximately** normal with 0 mean and unit variance
- $p - value = 1 - \Phi(z_{obs})$, z_{obs}: observed value on the sample
- Small p-values suggest a value of the true μ larger than μ_0.

In practice if $p - value < 0.05$ H_0 is rejected

- $H_1 : \mu > \mu_0$ and $H_1 : \mu \neq \mu_0$ are handled in a similar manner

Test on the proportion (*probability of success*)

Suppose that X takes only two values, i.e. success and failure

Conventionally we assume that $X = 1$ (i.e. success) and $X = 0$ (i.e. failure)

For instance, if we are interested in the health status of a patient

$X = 1$: indicates that the patient experienced disease progression

$X = 0$: indicates that the patient did not experience disease progression

This random variable is called *Bernoulli* random variable

Notation : $X \sim Ber(\theta)$ where θ is a parameter $0 \leq \theta \leq 1$

If the population is composed by a finite number of units, θ is the relative frequency of 1

Test on the probability of success

Bernoulli random variable: $X \sim Ber(\theta) \quad 0 \leq \theta \leq 1$

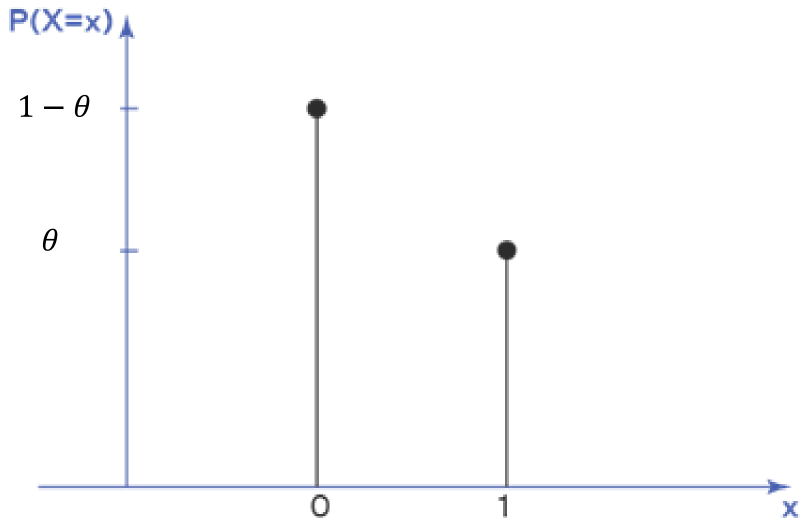
$X = \{0,1\}$

Probability function $\mathrm{P}(X = 1) = \theta$ and

$\mathrm{P}(X = 0) = 1 - P(X = 1) = 1 - \theta$

i.e.

$$\mathrm{P}(X = x) = \begin{cases} \theta & x = 1 \\ 1 - \theta & x = 0 \end{cases}$$



- Expectation $\mathrm{E}(X) = \mathrm{P}(X = 1) = \theta$
- Variance $\mathrm{Var}(X) = \theta(1 - \theta)$

Example

The National Center for Health Statistics reported that about 43 in 100 U.S. adults were obese in 2017–2018 ("adult" were defined as age 20 and over).

So, the probability of being obese in the US can be assumed to be as big as 0.43 (number of the favourable cases divided by all possible cases).

If we take an American at random from the US population, we have a 43% chance to get an obese person (the success).

So if X is the RV that identifies an obese person i.e. $X = \begin{cases} 1 & \text{the person is obese} \\ 0 & \text{otherwise} \end{cases}$

this is a Bernoulli RV with $\theta = 0.43$.

… … A reasonable mechanism to model obesity occurrence in the US … …

Test for proportions (probability of success) for a binary variable

We want to test $H_0\!: \theta = \theta_0; \quad H_1\!: \theta \neq \theta_0$

Test statistics: $Z = \frac{\hat{\theta} - \theta_0}{S/\sqrt{n}} = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}}$ (same previous case)

Note that in this case $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \hat{\theta}(1 - \hat{\theta})$

- If H_0 is true, Z is **approximately** normal with 0 mean and unit variance
- $p-value = 1 - \Phi(z_{obs})$, z_{obs}: value of Z observed on the sample
- Small p-values suggest a value of the true θ larger than θ_0.

In practice: if $p - value < 0.05$ H_0 is rejected

- $H_1\!: \theta > \theta_0$ and $H_1\!: \theta < \theta_0$ are handled in a similar manner

Comparing the mean of two populations

Assume that the variable of interest is measured on two different groups (i.e. treated and not treated patients, male and female,….).

Let X be the variable in the first group.

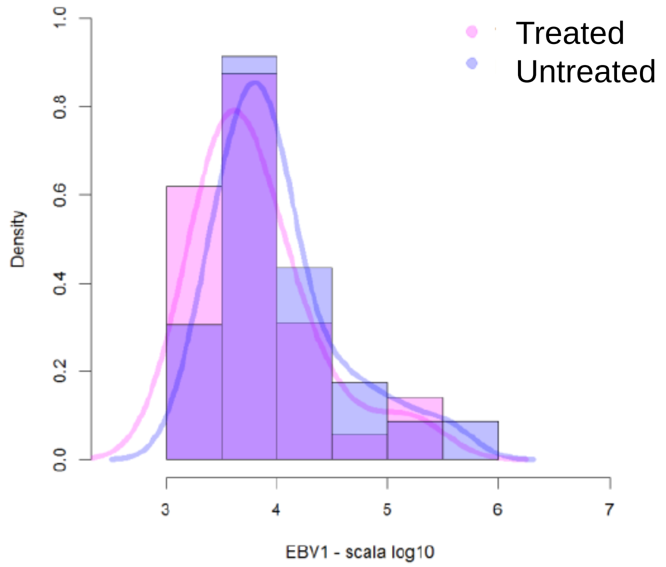
Let Y be the variable in the second group.

Example (Petrara et al. The Journal of Infectious Diseases, 2014)

Blood samples from 213 HIV-1–infected children, 140 of whom were receiving antiretroviral therapy (ART).

Nucleic acids were extracted and analyzed for quantification of Epstein-Barr Virus (EBV).

It is desired to test whether the Epstein-Barr Virus load differs between treated and untreated children.



Comparing the mean of two normal populations

Assume that $X \sim N(\mu_X, \sigma^2)$ and $X_1, \ldots, X_n$ be a random sample taken from X where $\mathrm{E}(X) = \mu_X$

Assume that $Y \sim N(\mu_Y, \sigma^2)$ and $Y_1, \ldots, Y_m$ be a random sample taken from Y where $\mathrm{E}(Y) = \mu_Y$

$$H_0\colon \mu_X = \mu_Y \qquad H_1\colon \mu_X \neq \mu_Y$$

Example: assume we are testing the effectiveness of a new drug compared to a standard treatment.

X and Y measure that outcome under the two regimes, respectively.

The null hypothesis states that there is no difference in effectiveness between the new drug and the standard treatment.

The alternative hypothesis states that there is a difference in effectiveness.

Note that in this case, H_1 does not specify whether the new drug is better or worse than the standard treatment. This is an example of a two-sided test.

Comparing the mean of two normal populations

Assume that $X \sim N(\mu_X, \sigma^2)$ and $X_1, \dots, X_n$ be a random sample taken from X where $\mathrm{E}(X) = \mu_X$

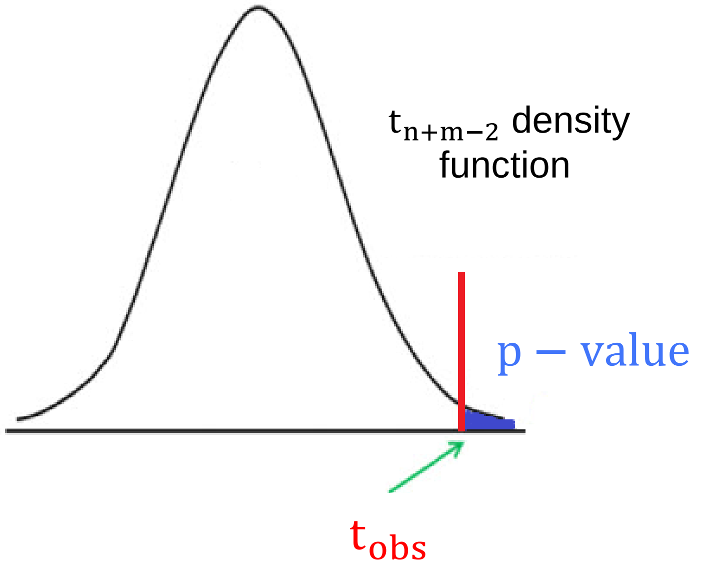
Assume that $Y \sim N(\mu_Y, \sigma^2)$ and $Y_1, \dots, Y_m$ be a random sample taken from Y where $\mathrm{E}(Y) = \mu_Y$

$$H_0\colon \mu_X = \mu_Y \qquad H_1\colon \mu_X > \mu_Y$$

Test Statistics $\mathrm{T} = \dfrac{\bar{\mathrm{X}} - \bar{\mathrm{Y}}}{\mathrm{S_P}\sqrt{\frac{1}{\mathrm{n}} + \frac{1}{\mathrm{m}}}} | \mathrm{H_0} \sim \mathrm{t_{n+m-2}}$ where $\mathrm{S_P^2} = \dfrac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$

Observed value of T $\mathrm{t_{obs}} = \dfrac{\bar{\mathrm{x}}_{\mathrm{oss}} - \bar{\mathrm{y}}_{\mathrm{oss}}}{s_{P\,oss}\sqrt{\frac{1}{n} + \frac{1}{m}}}$

$\mathrm{p-value} = \mathrm{P}(\mathrm{T} > \mathrm{t_{obs}}\ | \mathrm{H_0}) = 1 - \mathrm{F}_{\mathrm{T}_{n+m-2}}(\mathrm{t_{obs}})$



NOTE: $Var(X) = Var(Y) = \sigma^2$ homoschedasticy condition (to be tested…)

Comparing the mean of two normal populations

Assume that $X \sim N(\mu_X, \sigma^2)$ and $X_1, \ldots, X_n$ be a random sample taken from X where $\mathrm{E}(X) = \mu_X$

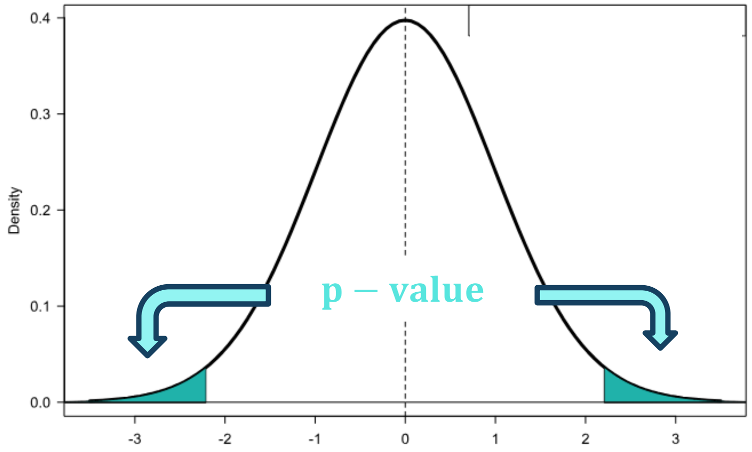
Assume that $Y \sim N(\mu_Y, \sigma^2)$ and $Y_1, \ldots, Y_m$ be a random sample taken from Y where $\mathrm{E}(Y) = \mu_Y$

$$H_0\colon \mu_X = \mu_Y \qquad H_1\colon \mu_X \neq \mu_Y$$

Test Statistics $\mathrm{T} = \dfrac{\bar{\mathrm{X}} - \bar{\mathrm{Y}}}{\mathrm{S_P}\sqrt{\frac{1}{\mathrm{n}} + \frac{1}{\mathrm{m}}}} | \mathrm{H_0} \sim \mathrm{t_{n+m-2}}$ where $\mathrm{S_P^2} = \dfrac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$

Observed value of T $\mathrm{t_{obs}} = \dfrac{\bar{\mathrm{x}}_{\mathrm{obs}} - \bar{\mathrm{y}}_{\mathrm{obs}}}{\mathrm{s}_{Pobs}\sqrt{\frac{1}{n} + \frac{1}{m}}}$

$\mathrm{p - value} = \mathrm{P}(|\mathrm{T}| > |\mathrm{t_{obs}}| \;|\mathrm{H_0}) = 2\left[1 - \mathrm{F}_{\mathrm{T}_{n+m-2}}(|\mathrm{t_{oss}}|)\right]$



$-|\mathrm{t_{obs}}|$ $|\mathrm{t_{obs}}|$

NOTE: $Var(X) = Var(Y) = \sigma^2$ homoschedasticy condition (to be tested…)

Comparing the proportion of two binary populations (cont'd)

Assume that $X \sim Ber(\theta_X)$ and $X_1, \ldots, X_n$ be a random sample taken from X and $\mathrm{P}(X = 1) = \theta_X$

Assume that $Y \sim Ber(\theta_Y)$ and $Y_1, \ldots, Y_m$ be a random sample taken from Y and $\mathrm{P}(Y = 1) = \theta_Y$

$$H_0: \theta_X = \theta_Y (= \theta) \qquad H_1: \theta_X \neq \theta_Y$$

Test Statistics $\mathrm{Z} = \dfrac{\hat{\theta}_X - \hat{\theta}_Y}{\sqrt{\hat{\theta}(1-\hat{\theta})\frac{n+m}{\mathrm{nm}}}} | \mathrm{H}_0 \sim N(0,1)$ (approximately for large samples)

$$\hat{\theta} = \frac{n\hat{\theta}_X + m\hat{\theta}_Y}{n+m} = \frac{\sum_{i=1}^n X_i + \sum_{j=1}^m Y_j}{n+m} \quad \hat{\theta}_X = \frac{1}{n}\sum_{i=1}^n X_i \ (= \overline{\mathrm{X}}) \quad \text{and} \quad \hat{\theta}_Y = \frac{1}{m}\sum_{j=1}^m Y_j \ (= \overline{\mathrm{Y}})$$

Observed value of Z $z_{obs} = \dfrac{\hat{\theta}_{X,obs} - \hat{\theta}_{Y,obs}}{\sqrt{\hat{\theta}_{obs}(1-\hat{\theta}_{obs})\frac{n+m}{\mathrm{nm}}}}$

$\mathbf{p - value} = \mathrm{P}_{\mathrm{H}_0}(|Z| > |z_{\mathrm{obs}}|) = 2[1 - \Phi(|z_{\mathrm{obs}}|)]$ (two-sided test)

Rationale: we expect a small value of Z if H_0 holds true

Comparing the proportion of two binary populations (cont'd)

Assume that $X \sim Ber(\theta_X)$ and $X_1, \dots, X_n$ be a random sample taken from X and $\mathrm{P}(X=1) = \theta_X$

Assume that $Y \sim Ber(\theta_Y)$ and $Y_1, \dots, Y_m$ be a random sample taken from Y and $\mathrm{P}(Y=1) = \theta_Y$

$H_0: \theta_X = \theta_Y$ $H_1: \theta_X > \theta_Y$ (one-sided test)

Test Statistics $\mathrm{Z} = \frac{\widehat{\theta}_X - \widehat{\theta}_Y}{\sqrt{\widehat{\theta}(1-\widehat{\theta})\frac{n+m}{\mathrm{nm}}}} | \mathrm{H}_0 \sim N(0,1)$ (approximately for large samples)

Observed value of Z $z_{obs} = \frac{\widehat{\theta}_{X,obs} - \widehat{\theta}_{Y,obs}}{\sqrt{\widehat{\theta}_{obs}(1-\widehat{\theta}_{obs})\frac{n+m}{\mathrm{nm}}}}$

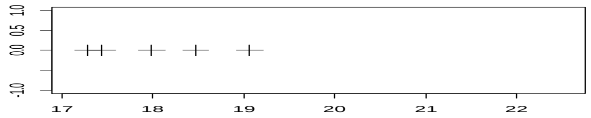
$\mathbf{p-value} = \mathrm{P}_{\mathrm{H}_0}(Z > \mathrm{z_{obs}}) = 1 - \Phi(\mathrm{z_{obs}})$

If $H_1: \theta_X < \theta_Y$ $\mathbf{p-value} = \mathrm{P}_{\mathrm{H}_0}(Z < \mathrm{z_{obs}}) = \Phi(\mathrm{z_{obs}})$

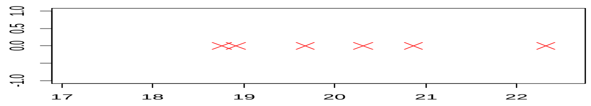
Comparing two general populations: the Mann-Whitney–Wilcoxon test

Assume that

$X \sim f_X$ and $X_1, \dots, X_n$ be a random sample from X with $\mathrm{Median}(\mathrm{X}) = Me_X$

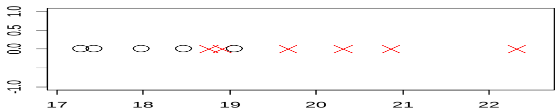


$Y \sim f_y$ and $Y_1, \dots, Y_m$ be a random sample from Y with $\mathrm{Median}(Y) = Me_Y$



$$H_0 \colon Me_X = Me_Y \qquad H_1 \colon Me_X < Me_Y$$

Indicate by



- $L_1, \dots L_{n+m} = X_1, \dots, X_n, Y_1, \dots, Y_m$ the union of the two samples
- $r_L(X_1), \dots, r_L(X_n)$ the ranks of $X_1, \dots, X_n$ in the sorted joined sample

  $L_{(1)}, \dots L_{(n+m)}$

Wilcoxon Test Statistics: $\mathrm{W} = \sum_{i=1}^{n} r_L(X_i)$ $\qquad \frac{\mathrm{n(n+1)}}{2} \leq \mathrm{W} \leq mn + \frac{n(n+1)}{2}$

Comparing two general populations
the Mann-Whitney–Wilcoxon test cont’d

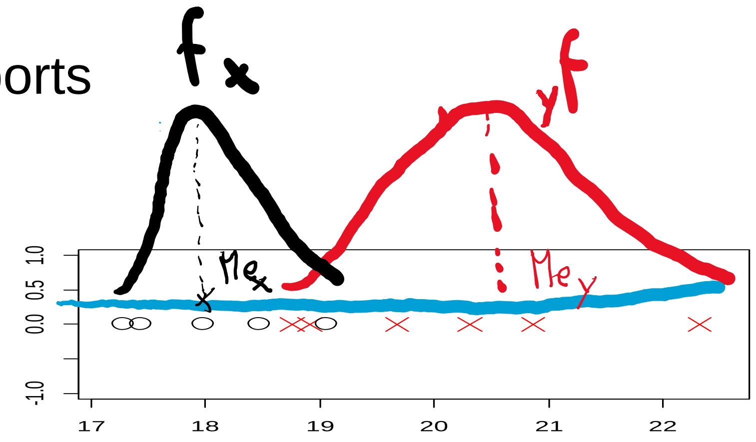
Hypothesis of interest $H_0\colon Me_X = Me_Y \quad H_1\colon Me_X < Me_Y$

Wilcoxon Test Statistics: $\mathrm{W} = \sum_{i=1}^{n} r_L(X_i) \qquad \frac{\mathrm{n(n+1)}}{2} \leq \mathrm{W} \leq mn + \frac{n(n+1)}{2}$

When $\mathrm{W} = \frac{\mathrm{n(n+1)}}{2}$ all the X_i precede all Y_i hence the test statistics supports

- the alternative hypothesis if it is close to 0 and

- the null hypothesis if it is close to 1.



$\mathbf{p - value} = \mathrm{P}_{\mathrm{H}_0}(\mathrm{W} < w_{\mathrm{obs}}) \approx \Phi(w_{\mathrm{obs}})$

p-value is calculated using the normal distribution if $n, m \gg 0$ (at least 7). For smaller sample sizes the normal approximation in inappropriate and p-values are tabulated

Comparing two general populations the Mann-Whitney–Wilcoxon test cont'd

Assume that $X \sim f_X$ and $X_1, \ldots, X_n$ be a random sample from X with $\mathrm{Median(X)} = Me_X$

Assume that $Y \sim f_y$ and $Y_1, \ldots, Y_m$ be a random sample from Y with $\mathrm{Median}(Y) = Me_Y$

$$H_0: Me_X = Me_Y \quad H_1: Me_X > Me_Y$$

or

$$H_0: Me_X = Me_Y \quad H_1: Me_X \neq Me_Y$$

In this case the test is derived similarly. It is necessary only to adjust the calculation of p-values according to the direction specified in the alternative hypothesis.

This test can be defined equivalently using a different test statistics due to Mann and Withney and this motivates the name of the test.

Tests for normality

Test for normality are statistical tests of whether a given sample of data is drawn from a normal distribution. They help in understanding whether a normal distribution adequately describes a set of data and are powerful statistical tools for detecting most departures of a set of data from normality.

Graphical diagnostics

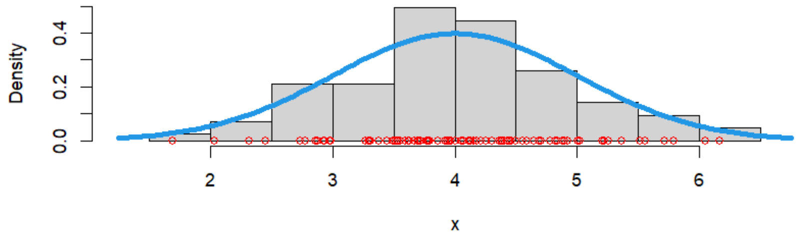
- Boxplot
- Histogram
- Normal probability (p-p) plot
- Quantile-quantile (q-q) plot

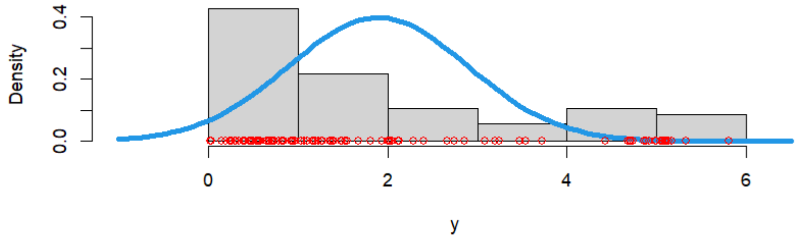
Formal statistical tests

- Shapiro–Wilk test
- Anderson–Darling test
- Cramér–von Mises criterion
- D'Agostino's K-squared test
- Kolmogorov–Smirnov test
- Lilliefors test
- Shapiro–Francia test
- …

Test for normality: Shapiro–Wilk test

The Shapiro–Wilk test is a statistical test of whether a given sample of data is drawn from a Normal distribution, and it is one the most powerful test to check data departures from normality.





Assume that $X_1, \ldots, X_n$ is a random sample drawn from X. We use the data to test

H_0: X is normally distributed $\quad$ H_1: X is not normally distributed

Test for normality: Shapiro–Wilk test cont'd

$X_1, \dots, X_n$: random sample drawn from X

$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$: order statistics

Test Statistics: $W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$

$a_1, \dots, a_n$: set of known coefficients

$0 < W \le 1$: values close enough to 1 support the normality hypothesis.

If the p-value is less than the chosen alpha level (i.e. 0.05), then H_0 is rejected and the data cannot be considered normally distributed.

If the p-value is greater than the chosen alpha level, then H_0 (that the data came from a normally distributed population) cannot be rejected