A List of Hypothesis Tests

Jean-Baka Domelevo Entfellner

February 2015

Test name	Purpose, validity and procedure	Null hypothesis (H_0)	Test statistic	Probability distribution of test statistic under H_0	Parametric? (and if yes, assumption)	R command
Testing the mean(s) of the distribution(s)						
one-sample Z-test	Test the **equality of the true mean to a given value** μ_0. Variance σ^2 is known.	$\mu = \mu_0$	$Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$	$Z \sim \mathcal{N}(0,1)$	yes (normality)	`asbio::one.sample.z(x, sigma=s, null.mu=mu0)`
one-sample t-test	Test the **equality of the true mean to a given value** μ_0. Variance σ^2 is unknown.	$\mu = \mu_0$	$T = \frac{\bar{x}-\mu}{s_{\mathbf{x}}/\sqrt{n}}$	$T \sim t_{(n-1)}$	yes (normality)	`t.test(x, mu=mu0)`
two-sample Z-test for *unpaired* samples	Test the **equality of the true means**. Variances are known but not necessarily equal.	$\mu_1 = \mu_2$	$Z = \frac{\overline{x_1}-\overline{x_2}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$Z \sim \mathcal{N}(0,1)$	yes (normality)	— (use `qnorm(...)`)

Test name	Purpose, validity and procedure	Null hypothesis (H_0)	Test statistic	Probability distribution of test statistic under H_0	Parametric? (and if yes, assumption)	R command
two-sample exact t-test for *unpaired* samples	Test the **equality of the true means**. Variances are unknown but supposed equal ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) and estimated through pooled estimator $s_p^2 = \frac{n_1-1}{n_1+n_2-2}\ s_{\mathbf{x}_1}^2 + \frac{n_2-1}{n_1+n_2-2}\ s_{\mathbf{x}_2}^2$.	$\mu_1 = \mu_2$	$T = \frac{\overline{x_1}-\overline{x_2}}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	$T \sim t_{(n_1+n_2-2)}$	yes (normality)	`t.test(x1, x2, var.equal=T)`
two-sample Welch's t-test for *unpaired* samples	Test the **equality of the true means**. Variances are unknown and not supposed equal. $\nu \in \mathbb{R}_+$ is approximated using the Welch–Satterthwaite equation.	$\mu_1 = \mu_2$	$T = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{s_{\mathbf{x}_1}^2}{n_1}+\frac{s_{\mathbf{x}_2}^2}{n_2}}}$	$T \sim t_\nu \quad (\nu \in \mathbb{R}_+)$	yes (normality)	`t.test(x1, x2, var.equal=F)`
paired-sample t-test	Test the **equality of the true means**. The $2n$ sample values are paired: dependency between x_i and y_i. $\mathbf{d}$ is the vector of all the $d_i = x_i - y_i$, $\bar{d}$ its mean and $s_{\mathbf{d}}^2 = \frac{1}{n-1}\ \sum_i (d_i - \bar{d})$.	$\mu_x = \mu_y$	$T = \frac{\bar{d}}{s_{\mathbf{d}}/\sqrt{n}}$	$T \sim t_{(n-1)}$	yes (normality)	`t.test(x, y, paired=T)`

Test name	Purpose, validity and procedure	Null hypothesis (H_0)	Test statistic	Probability distribution of test statistic under H_0	Parametric? (and if yes, assumption)	R command
Testing the medians of two distributions (a.k.a. non-parametric tests on the mean)						
Mann-Whitney U test for **unpaired samples** (a.k.a. Wilcoxon rank-sum test or Mann-Whitney-Wilcoxon test)	Test the **equality of the medians**. Rank all the values from $\mathbf{x_1}$ and $\mathbf{x_2}$ in one batch, and retain the vectors of ranks (values from 1 to $n_1 + n_2$). Let R_1 be the sum of the ranks from $\mathbf{x_1}$.	$\mathrm{med}(\mathbf{x_1}) = \mathrm{med}(\mathbf{x_2})$	$Z = \frac{U - \frac{n_1 n_2}{2}}{\sigma_U}$ where $U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$ and $\sigma_U = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$	$Z \underset{\text{approx.}}{\sim} \mathcal{N}(0,1)$	no	`wilcox.test(x1, x2, paired=F)`
paired-sample sign test	Test the **equality of the medians**. From $\mathbf{x}$ and $\mathbf{y}$ we build the vector of the signs of the differences: $w_i = \mathrm{sgn}(x_i - y_i)$. W is the number of pairs for which $x_i > y_i$. $m \leq n$ is the number of pairs with a non-zero difference.	$\mathrm{med}(\mathbf{x}) = \mathrm{med}(\mathbf{y})$	$W = \lvert \{w_i,\ w_i = +1\} \rvert$	$W \sim \mathcal{B}(m, p = 0.5)$	no	`binom.test(W, m)`

Test name	Purpose, validity and procedure	Null hypothesis (H_0)	Test statistic	Probability distribution of test statistic under H_0	Parametric? (and if yes, assumption)	R command
Wilcoxon signed-rank test for **paired samples**	Test the **equality of the medians**. From $\mathbf{x}$ and $\mathbf{y}$ we build the vector of the signs of the differences: $w_i = \mathrm{sgn}(x_i - y_i)$. Order (by ascending absolute values) all m non-zero differences to calculate their ranks r_i. W is the absolute value of the sum of signed ranks.	$\mathrm{med}(\mathbf{x}) = \mathrm{med}(\mathbf{y})$	$Z = \frac{W-0.5}{\sigma_W}$ where $W = \lvert \sum_{i=1}^{m} w_i r_i \rvert$ and $\sigma_W = \sqrt{\frac{m(m+1)(2m+1)}{6}}$	$Z \underset{\text{approx.}}{\sim} \mathcal{N}(0,1)$	no. Only assumption: the distribution of the differences in each independent sample is assumed to be symmetric around the median.	`wilcox.test(x, y, paired=T)`

Testing the variance(s) of the distribution(s)

Test name	Purpose, validity and procedure	Null hypothesis (H_0)	Test statistic	Probability distribution of test statistic under H_0	Parametric? (and if yes, assumption)	R command
one-sample χ^2-test for the variance	Test the **equality of the true variance to a given value σ_0^2**.	$\sigma = \sigma_0$	$X = \frac{(n-1)s_{\mathbf{x}}^2}{\sigma^2}$	$X \sim \chi^2_{(n-1)}$	yes (normality)	— (use `qchisq(...)`)
two-sample Fisher-Snedecor F-test for homeoscedasticity	Test the **equality of the true variances**. Samples are supposed independent.	$\sigma_1 = \sigma_2$	$F = \frac{s_{\mathbf{x}_1}^2}{s_{\mathbf{x}_2}^2}$	$F \sim F(n_1 - 1, n_2 - 1)$	yes (normality)	`var.test(x1,x2)`

Testing the normality of a sample

Test name	Purpose, validity and procedure	Null hypothesis (H_0)	Test statistic	Probability distribution of test statistic under H_0	Parametric? (and if yes, assumption)	R command
Shapiro-Wilk test	Test whether a sample originates from a normal distribution	$\mathbf{x}$ comes from $X \sim \mathcal{N}(\dots)$			no	`shapiro.test(x)`

Test name	Purpose, validity and procedure	Null hypothesis (H_0)	Test statistic	Probability distribution of test statistic under H_0	Parametric? (and if yes, assumption)	R command
Kolmogorov-Smirnov test	Test whether two samples originate from the same distribution. One can give a formal distribution instead of the second sample.	$\exists \mathcal{D},\ X \sim \mathcal{D}$ and $Y \sim \mathcal{D}$			no	`ks.test(x, "pnorm", mean=mean(x), sd=sd(x))`