

2-sample t-test

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ Estimates the population mean μ . $S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$ Estimates the variance σ^2 . The Hypothesis test is: $H_0: \mu_1 = \mu_2$; $H_a: \mu_1 \neq \mu_2$, or $<$ (left-tailed) or $>$ (right-tailed).

Known Variances use the z test; $Z = (\bar{y}_1 - \bar{y}_2) / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ (two population means), or $Z = (\bar{y} - \mu_0) / \frac{\sigma}{\sqrt{n}}$ (one). $100(1-\alpha\%)$ C.I. for $\mu_1 - \mu_2$: $(\bar{y}_1 - \bar{y}_2) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$. C.R. where $H_a: \neq \implies Z_{\alpha/2}$. $< \implies -Z_{\alpha}$. $> \implies Z_{\alpha}$

Unknown Variances ($\sigma_1^2 = \sigma_2^2$) $t = (\bar{y}_1 - \bar{y}_2) / S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, $d.f = v = n_1 + n_2 - 2$. $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$ ($S_p = \sqrt{S_p^2}$) $100(1-\alpha)\%$ C.I. for $\mu_1 - \mu_2$: $(\bar{y}_1 - \bar{y}_2) \pm t_{\alpha/2, v} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ (two means), or $\bar{y} \pm t_{\alpha, n-1} \frac{\sigma}{\sqrt{n}}$ (one mean).

Unknown Variances ($\sigma_1^2 \neq \sigma_2^2$) $t = (\bar{y}_1 - \bar{y}_2) / \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$, $d.f = v = \frac{((S_1^2/n_1) + (S_2^2/n_2))^2}{(\frac{S_1^2/n_1}{n_1-1} + \frac{S_2^2/n_2}{n_2-1})}$. (round down v to nearest int.) C.R. where $H_a: \neq \implies t_{\alpha/2, v}$. $H_a: < \implies -t_{\alpha, v}$. $H_a: > \implies t_{\alpha, v}$

Tests on Variances The Hypothesis test is: $H_0: \sigma_1^2 = \sigma_2^2$; $H_a: \sigma_1^2 \neq \sigma_2^2$, or $<$ (left-tailed) or $>$ (right-tailed).

two-tailed(\neq): $F = \frac{S_1^2}{S_2^2} > F_{\alpha/2, n_1-1, n_2-1}$, **left-tail($<$):** $F = \frac{S_2^2}{S_1^2} > F_{\alpha, n_2-1, n_1-1}$, **right-tail ($>$):** $F = \frac{S_1^2}{S_2^2} > F_{\alpha, n_1-1, n_2-1}$.

Paired t-test (2 measurements from each exp unit)

Advantage: less variability than a design with 2 separate samples, test is more sensitive, ie more powerful.

Hypothesis: $H_0: \mu_d = 0$. $H_a: \mu_d \neq 0$. Where $\mu_d = \mu_2 - \mu_1$ (before - after). Test statistic: $t = \frac{\bar{d}}{S_d/\sqrt{n}}$ Where $\bar{d} = \frac{1}{n} \sum_{j=1}^n d_j$ is the sample mean. And $S_d^2 = \frac{1}{n-1} \sum_{j=1}^n (d_j - \bar{d})^2$. (note: use $S_d = \sqrt{S_d^2}$) Reject H_0 if $|t| > t_{\alpha/2, n-1}$.

ANOVA

The overall mean μ is $\frac{1}{a} \sum_{i=1}^a \mu_i$. Note that $\sum_{i=1}^a \tau_i = 0$.

LS estimates: $\hat{\mu} = \bar{y}_{..}$, $\hat{\mu}_i = \bar{y}_{i.}$, $\hat{\tau} = \bar{y}_{i.} - \bar{y}_{..}$, to estimate common population variances σ^2 , use $S_p^2 = \frac{\sum_{i=1}^a (n_i-1)s_i^2}{\sum_{i=1}^a (n_i-1)}$

Anova table (one-way layout)

Source	df	SS	MS	F	Note: N is the total number of observations, Pooled sample variance $S_p^2 = \frac{SSE}{N-a} = MSE$. Reject H_0 if $F > F_{\alpha, a-1, N-a}$, or p-value $< \alpha$.
Treatments	$a - 1$	$SSTr$	$MSTr = \frac{SSTr}{a-1}$	$F = \frac{MSTr}{MSE}$	
Error	$N - a$	SSE	$MSE = \frac{SSE}{N-a}$		
Total	$N - 1$	SST			

Hypotheses: $H_0: \mu_1 = \mu_2 = \dots = \mu_a$. H_a : Not all μ_i are equal. Or $H_0: \tau_1 = \tau_2 = \dots = \tau_a = 0$. H_a : Not all $\tau_i = 0$.

$100(1-\alpha)\%$ C.I. for μ_i : $\bar{y}_{i.} \pm t_{\alpha/2, N-a} \sqrt{\frac{MSE}{n}}$

To check model assumptions use a normal probability plot and normality test (Anderson-Darling or some other), and a plot of residuals vs. predicted value (residual plot). The residuals should look like they are from a normal distribution if the model assumptions were met. If the model assumptions were met, the residual plot should look like a random scatter. if the plot shows a 'funnel' shape, this is an indication of non-constant variance (heteroscedasticity). The remedy is typically a transformation of the response.

Normal probability plot

To estimate the standard deviation σ , note that the slope $\approx \frac{1}{\sigma}$, only with z-score. Or you can estimate the distance from 16% to 50% as on standard deviation. To estimate the mean μ look at the 50th percentile. Hypotheses: H_0 : The data are drawn from a normal distribution. H_a : The data are not drawn from normal distribution. Large p-value means normal.

Standard to Percentile:	Standard Deviations	-2	-1	0	1	2
	Percentile	2.27%	15.88%	50%	84.23%	97.72%

Note that $\Phi(\text{standard deviation}) = \text{percentile}$, and $\Phi^{-1}(\text{percentile}) = \text{standard deviation}$.

Contrasts

$\Gamma = \sum_{i=1}^a c_i \mu_i$, where c_i 's are constants such that $\sum_{i=1}^a c_i = 0$. To estimate Γ use $\hat{\Gamma} = \sum_{i=1}^a c_i \bar{y}_i$.

Hypotheses: $H_0: \sum_{i=1}^a c_i \mu_i = 0$. $H_a: \sum_{i=1}^a c_i \mu_i \neq 0$. (or $<$ or $>$).

Test statistic: $t = (\sum_{i=1}^a c_i \bar{y}_i) / (\sqrt{\frac{MSE}{n} \sum_{i=1}^a c_i^2})$. Reject H_0 if t in C.R. of $t_{\alpha, N-a}$ (or $\alpha/2$ for 2-tail) or if p-value $< \alpha$.

Note: $F = (\sum_{i=1}^a c_i \bar{y}_i)^2 / (\frac{MSE}{n} \sum_{i=1}^a c_i^2)$ can be used to test 2-tailed. Reject H_0 if $F > F_{\alpha, 1, N-a}$ or if p-value $< \alpha$.

Pairwise comparisons

Note: LSD is most powerful, but worst at guarding against Type I errors. Tukey's is most balanced for equal sample sizes. And Bonferroni's is least likely to make a Type I error.

Fishers LSD For every pair of treatments test the hypothesis at α . $H_0: \mu_i = \mu_j$. $H_a: \mu_i \neq \mu_j$.

Test statistic $t = |\bar{y}_i - \bar{y}_j| / \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$. Reject H_0 if $|t| > t_{\alpha/2, N-a}$.

Therefore, The 2 treatment means are significantly different if $|\bar{y}_i - \bar{y}_j| > t_{\alpha/2, N-a} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$.

Bonferroni's Method Denote the FWE by α . Identical to Fisher's LSD except that $|\bar{y}_i - \bar{y}_j| > t_{\alpha/2k, N-a} \sqrt{MSE \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$.

Where $k = \binom{a}{2}$.

Tukey's Method Means are significantly different at $FWE\alpha$ when: $|\bar{y}_i - \bar{y}_j| > q_{\alpha, a, N-a} \sqrt{\frac{MSE}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$.

100(1- α)% Simultaneous C.I.'s for all possible treatment differences $\mu_i - \mu_j$ using Tukey's Method: $\bar{y}_i - \bar{y}_j \pm q_{\alpha, a, N-a} \sqrt{\frac{MSE}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$

Graphical results Note that a line under the treatments indicates that they are not significantly different.

(2)	(1)	(4)	(3)	(5)	Treatments should be ordered from smallest sample mean to largest sample mean. This graph indicates that treatment 2 and 1 are not significantly different, and 4 and 3, and 5 are not significantly different.
<hr/>		<hr/>			

Conclusions

Rejection Methods Reject H_0 if p-value $\leq \alpha$. Reject H_0 if the test statistic falls within the critical region.

100(1- α)% **C.I** "We are 100(1- α)% confident that the true {test} is between {upper bound} and {lower bound}."

Reject H_0 "There is enough statistical evidence to support (the statement of H_a)."

Fail to reject H_0 "There is not enough statistical evidence to support (the statement of H_a).\" or \"The evidence from the data is consistent with (the statement of H_0)\""

Test of equality of variances

Hypothesis: $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$. H_a : at least one σ_i^2 is different.

Bartlett's test (use if data is normal) assumes normality for each treatment. Very sensitive to non-normality.

Modified Levene's test (use if data \approx normal) robust for non-normality. For data from a normal dist, not as powerful as Bartlett's.