

One- and Two-Sample tests

Leif E. Peterson

May 18, 2006

Repetitive Sampling of Means

Assume an infinitely large sample of patients having a continuous normally distributed variable with known population mean μ and population standard deviation σ , from which a sample of n patients are drawn. Calculate the sample mean \bar{x} and sample standard deviation s . Replace the data for the n patients, and then draw another random sample of size n , calculate \bar{x} and repeat multiple times.

Does $\bar{x} = \mu$?

The grand mean of the multiple sample means (\bar{x}) is the population mean μ , and its standard deviation is the standard error of the mean σ/\sqrt{n} . Given this, how we can calculate the probability that that a single sample mean \bar{x} would obtain from random sampling from a population with parameters μ and σ ?

Z-score

This can be done by calculating a z-score in the form

$$Z = \frac{\bar{x} - \mu}{\sigma_x / \sqrt{n}} \quad (1)$$

This applies to a sampling distribution of \bar{x} 's for which the standard deviation is σ_x / \sqrt{n} rather than σ .

t-distribution

A requirement for testing the significance of \bar{x} with the standard normal Z distribution is that σ is known. It is more likely, however, that σ is unknown, and therefore we must use a t -test like

$$t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}} \quad (2)$$

Tabled Critical Values of t

t -tables list probabilities in columns, $n - 1$ degrees of freedom in rows, and values of t in the cells of the table. For $n \geq 30$ the t -distribution approaches the z -distribution.

Estimating the population mean from \bar{x}

The 95% confidence interval around \bar{x} is based on the relationship

$$\mu = \bar{x} \pm t(s/\sqrt{n}) \quad (3)$$

Example: 17 patients with mean hemoglobin concentration of 15 g/dl, and $s = 1$. What is the 95% confidence interval about the mean?

$$\begin{aligned} \mu &= 15 \pm 2.12(1/\sqrt{17}) \\ &= 15 \pm 0.5142g/dl \\ &= 14.48\text{to}15.51g/dl \end{aligned} \quad (4)$$

One Sample t -test

Let \bar{x} and s be the sample mean and standard deviation from a study sample. Suppose we wish to test whether the mean is significantly different from a reference mean, μ_0 . In this case the null hypothesis for a two-tailed test ($\bar{x} < \mu_0$ or $\bar{x} > \mu_0$) is $H_0 : \mu = \mu_0$ and the alternative hypothesis is $H_a : \mu \neq \mu_0$. The test statistic is

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}} \quad (5)$$

Again, look up the p-value corresponding to the calculated t_{n-1} statistic for $n - 1$ degrees of freedom. This is a two tailed test since $\mu \neq \mu_0$ in the alternative hypothesis implies that $\mu < \mu_0$ or $\mu > \mu_0$. If we wanted to test that $\bar{x} < \mu_0$, the null hypothesis for a one-tailed test would be $H_o : \mu \geq \mu_0$ with an alternative hypothesis $H_a : \mu < \mu_0$. Analogously, if the goal was to test that $\bar{x} > \mu_0$, the null hypothesis for a one-tailed test is $H_o : \mu \leq \mu_0$ and the alternative hypothesis is $H_a : \mu > \mu_0$. The criterion for determining whether a test is one-tailed or two-tailed is based on the possible equality/inequality combination in the alternative hypothesis.

t-test for Independent Samples ($\sigma_1^2 \neq \sigma_2^2$)

Assume that two samples of size n_1 and n_2 were drawn from two infinitely large populations. Calculate the sample means \bar{x}_1 and \bar{x}_2 , replace the sample data, resample, recalculate means, resample, etc. The mean differences $d = \bar{x}_1 - \bar{x}_2$ obtained from multiple samplings would be normally distributed with variance σ^2 .

A typical question is: is the difference ($d = \bar{x}_1 - \bar{x}_2$) between two sample means significant. The hypotheses are as follows:

$$\begin{aligned} H_0 : \delta &= 0 \\ H_a : \delta &\neq 0 \end{aligned} \tag{6}$$

where δ is the hypothesized difference in means.

t-test for Independent Samples ($\sigma_1^2 \neq \sigma_2^2$)

$$\begin{aligned} t_{1-\alpha/2; \nu=n_1+n_2-2} &= \frac{d - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \\ &= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \end{aligned} \quad (7)$$

T-Test		
Test var:	BXGGTOT	
Grouping var:	RECUR	
	1	2
Sample Size	1078	213
x_bar, y_bar	5.93	6.55
x_bar - y_bar	-0.62	
PSE	0.07	
Test Statistic	-8.19	
P-value	0.0000	
(Unequal variance assumption)		

t-test for Independent Samples ($\sigma_1^2 = \sigma_2^2$)

When the variances for the two samples are assumed to be equal, we must weight each sample variance by the sample size to derive the pooled sample variance as

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (8)$$

Substituting into our t -test, we now have

$$t_{1-\alpha/2; \nu=n_1+n_2-2} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (9)$$

T-Test		
Test var:	BXGGTOT	
Grouping var:	RECUR	
	1	2
Sample Size	1078	213
x_bar, y_bar	5.93	6.55
x_bar - y_bar	-0.62	
PSE	0.08	
Test Statistic	-8.39	
P-value	0	
(Equal variance assumption)		

T-Test		
Test var:	PREPSA	
Grouping var:	RECUR	
	1	2
Sample Size	1078	213
x_bar, y_bar	8.45	17.31
x_bar - y_bar	-8.86	
PSE	0.76	
Test Statistic	-12.31	
P-value	0.0000	
(Equal variance assumption)		

T-Test		
Test var:	PREPSA	
Grouping var:	RECUR	
	1	2
Sample Size	1078	213
x_bar, y_bar	8.45	17.31
x_bar - y_bar	-8.86	
PSE	0.66	
Test Statistic	-7.78	
P-value	0.0000	
(Unequal variance assumption)		

t-test for Paired Samples Confounding occurs when the variables not being tested are independently related to the outcome (e.g., age, gender) and independent of exposure. Matching is used to make subjects in different exposure groups more similar in terms of extraneous variables that are independently associated with the outcome. The goal of matching is therefore to achieve differences between outcome between the exposure groups that are more likely to be due to exposure rather than confounding variables. The overall effect of matching may reduce the standard deviation of differences rather than the differences themselves. Since standard deviation is in the denominator of the t -test, the reduced standard deviation increases the test statistic. This reduction in variance cause an increase in statistical efficiency, which is the power to detect a difference for a fixed sample size.

t-test for Paired Samples

For n paired samples, the t -test is

$$\begin{aligned} t_{1-\alpha/2; \nu=n-1} &= \frac{\bar{d} - \delta}{s_d / \sqrt{n}} \\ &= \frac{\bar{d}}{s_d / \sqrt{n}} \end{aligned} \quad (10)$$

where \bar{d} is the mean of the n pairwise differences $d_i = x_{i1} - x_{i2}$ and s_d is the standard deviation of differences given as

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}} \quad (11)$$

T-Test		
	BXGGTOT	PREPSA
Sample Size	1291	1291
x_bar, y_bar	6.03	9.91
x_bar - y_bar	-3.88	
PSE	0.29	
Test Statistic	-13.93	
P-value	0.0000	
(Paired test)		

One-Sample Test for Proportion

Frequently we are interested in testing whether a proportion is significantly different from a reference proportion. Let n represent the number of patients in the sample size, p the proportion of patients possessing the outcome of interest out of the total sample (sample proportion), and P the unknown proportion in the population in which sampling was performed. For small n , that is, $np < 5$, inferences about p can be made using the binomial probability distribution. Whereas when $np \geq 5$, the normal approximation to the binomial distribution holds.

One-sample Proportion Test

The sample mean proportion is P and the standard deviation is

$$s.e.(p) = \sqrt{\frac{PQ}{n}} \quad (12)$$

where $Q = 1 - P$. A two-tailed test ($H_o : p = P_0; H_a : p \neq P_0$) to determine if p is equal to a reference proportion P_0 is

$$z = \frac{|p - P_0| - 1/(2n)}{\sqrt{\frac{P_0Q_0}{n}}} \quad (13)$$

where $1/(2n)$ is a continuity correction to closer approximate normal curve probabilities with binomial probabilities.

Two-sample Test for Proportions

It is often of interest to test if two proportions from different populations are the same.

A two-tailed test ($H_o : p_1 = p_2; H_a : p_1 \neq p_2$) to determine if $p_1 = p_2$ is

$$z = \frac{|p_1 - p_2| - 1/(2n) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}{\sqrt{\bar{p}\bar{q} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (14)$$

where $\bar{p} = n_1/(n_1 + n_2)$.

Two-way Contingency Tables - Counts

Characteristic A	Characteristic B		Total
	Present	Absent	
Characteristic Present	n_{11}	n_{12}	$n_{1.}$
Absent	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

Two-way Contingency Table - Proportions

Characteristic A	Characteristic B		Total
	Present	Absent	
Characteristic Present	p_{11}	p_{12}	$p_{1.}$
Absent	p_{21}	p_{22}	$p_{2.}$
Total	$p_{.1}$	$p_{.2}$	$p_{..}$

Chi-square test of row-column independence

$$\chi^2 = n_{..} \sum_{i=1}^2 \sum_{j=1}^2 \frac{\left(|p_{ij} - p_{i.}p_{.j}| - 1/(2n_{..}) \right)^2}{p_{i.}p_{.j}} \quad (15)$$

with d.f. equal to $(r - 1)(c - 1)$

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (16)$$

Chi-square test

Chi-Square test				
		DRE		
		1	2	Total
RECUR	1	548	530	1078
	2	45	168	213
	Total	593	698	
	Chi-square (d.f.)	63.21(1)		
	P-value	0.0000		

Chi-Square test				
		RECUR		
		1	2	Total
LN	1	1063	156	1219
	2	15	57	72
	Total	1078	213	
	Chi-square (d.f.)	217.37(1)		
	P-value	0.0000		

Randomization Tests

For each gene, calculate the observed t -statistic t_i based on the raw data for groups A and B. Next use B permutations, where the number of permutations required is

$$B = \binom{n_1 + n_2}{\min\{n_1, n_2\}} \quad (17)$$

Randomization tests (cont'd)

During each permutation, calculate the t -statistic, $t_i^{(b)}$, for permutation b ($b = 1, 2, \dots, B$) after randomly shuffling the group labels for A and B across both groups. The empirical p -value after B iterations is

$$p = \frac{\#\{b : |t_i^{(b)}| \geq |t_i|\}}{B} \quad (18)$$

Iter	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	pse
1	1	2	3	4	5	6	7	8	22	23	24	25	26	27	28	29	1.22
2	24	26	25	29	23	3	8	6	4	22	27	7	28	5	1	2	5.52
3	24	7	6	29	28	1	23	27	3	2	4	25	5	8	22	26	5.50
4	24	29	5	25	7	28	6	4	2	23	3	8	22	1	26	27	5.72
5	25	24	6	27	1	3	26	23	7	29	4	28	5	22	8	2	5.66
6	26	22	23	24	4	27	2	25	29	3	8	28	7	5	6	1	5.30
7	7	25	4	22	8	29	23	3	1	6	24	26	5	27	2	28	5.74
8	26	22	28	2	5	24	4	25	3	23	27	7	1	8	29	6	5.64
9	1	8	26	5	3	28	23	22	29	27	24	7	2	6	25	4	5.74
10	3	4	22	27	5	25	7	1	24	23	6	28	29	8	2	26	5.48
11	26	3	6	25	22	4	2	28	1	5	8	29	7	27	23	24	5.74
12	23	28	2	26	27	5	7	1	22	6	4	8	24	29	25	3	5.74
13	27	25	28	24	2	22	8	1	23	4	6	5	29	7	3	26	5.63
14	6	22	25	27	3	5	2	28	7	24	8	29	1	4	26	23	5.74
15	22	8	26	23	28	7	4	27	29	5	2	3	6	25	1	24	5.50
16	2	27	6	5	29	8	25	1	22	3	4	7	23	28	24	26	5.63
17	22	8	24	3	2	1	27	6	7	26	23	28	4	29	5	25	5.45
18	23	26	4	22	7	8	5	24	1	25	29	27	6	3	2	28	5.74
19	6	23	8	22	1	27	4	25	7	26	3	24	2	29	5	28	5.74
20	5	26	24	29	27	23	7	28	8	25	3	22	4	2	6	1	4.72

Adjustments to P-value for Multiple Tests

The Sidak adjustment to the p -value for multiple tests is

$$\alpha^* = 1 - (1 - \alpha)^{1/\#\text{tests}} \quad (19)$$

and the Bonferroni adjustment is

$$\alpha^* = \alpha / \#\text{tests} \quad (20)$$