Comparing Treatments When the Response is either Success or Failure

When responses are binary (such as success/failure), the ANOVA assumption of normal distributions for all treatment responses does not hold. A common approach for testing equality of all treatments is a chi-squared test. The example below illustrates this approach.

Example 1. Suppose four treatments for headache relief are studied. One hundred migraine headache sufferers are randomly divided into four groups of size 25. Group 1 receives the medication Ablex, group 2 receives Breezex, group 3 receives Calmex, and group 4 receive Driftex. Use the results in the table below to test that all 4 treatments have an identical probability of success.

Headache Treatment Results

	Success	Failure	Total
Treatment 1 (Ablex)	13	12	25
Treatment 2 (Breezex)	18	7	25
Treatment 3 (Calmex)	10	15	25
Treatment 4 (Driftex)	9	16	25
Totals	50	50	100

Defining parameters. Let p_i denote the probability of headache relief from treatment i.

Hypotheses. $H_0 : p_1 = p_2 = p_3 = p_4$ ($=$ some common p)
H_a : not all p_i are equal

Test statistic derivation. Under H_0, the number of successes for all treatments have independent identical binomial distributions with parameters $n = 25$ and p (unknown).

The unknown constant p can be estimated by $\widehat{p} = \frac{total\,number\,of\,successes}{total\,number\,of\,observations} = \frac{50}{100} = .5.$

Letting X_i denote the number of successes for treatment i, we have $X_i \overset{\bullet}{\sim}$ binomial(25, .5) with parameters mean $\mu = 25(.5) = 12.5$ and standard deviation $\sigma = \sqrt{25(.5)(.5)} = \sqrt{6.25}.$

Therefore, by the central limit theorem, $Z_i = \frac{X_i - 12.5}{\sqrt{6.25}} \overset{\bullet}{\sim} N(0,1)$ for $i = 1, 2, 3, 4.$

Since the squared of a standard normal random variable has a chi-squared distribution, we have $Z_i^2 = \left(\frac{X_i - 12.5}{\sqrt{6.25}}\right)^2 \overset{\bullet}{\sim} \chi^2(1).$

Thus $T = \sum_{i=1}^{4} Z_i^2$ has an approximate χ^2 distribution with 3 degrees of freedom (losing 1 degree of freedom by using estimate $\widehat{p}$).

Note that, under $H_0,$ we expect T to be in the ballpark of its degrees of freedom, that is, about 3. Therefore a large value for T provides evidence for the alternative hypothesis.

Observed test statistic value and p-value.

For our data, the value for T is given by the following calculations:

$$t = \frac{(13-12.5)^2+(18-12.5)^2+(10-12.5)^2+(9-12.5)^2}{6.25} = 7.84$$

The approximate p-value is $P(T \geq 7.84 | T \sim \chi^2(3)) = .0494.$

Conclusion. Using $\alpha = .05$, we have sufficient evidence to conclude that **not** all treatments have identical success probability.

Multiple Comparisons. Since we rejected $H_0,$ we investigate to determine where are the significant difference or differences.

Bonferroni. Using a family wide 95% confidence level, the Bonferroni approach implies that we use 99.167% confidence levels for all 6 pair-wise differences.

Lower Bound for D	D	Upper Bound	Significant Differences
-.5544	$p_1 - p_2$	.15445	No
-.2492	$p_1 - p_3$	.48922	No
-.2056	$p_1 - p_4$	.52559	No
-.0307	$p_2 - p_3$	.67066	No
.01317	$p_2 - p_4$	.70683	Yes
-.3219	$p_3 - p_4$	.40191	No

Alternate (but equivalent) form for the test statistic. The test statistic $T = \sum_{i=1}^{r} Z_i^2$ used above in example 1 is algebraically equivalent to $T = \sum_k \frac{(O_k - E_k)^2}{E_k}$ where the sum runs over all 8 cells in the body of the table, where O_k denotes the count in cell k, and where E_k denotes the expected count under the null hypothesis. Here $E_k = n_i\widehat{p} = 25(.5)$ when cell k is a success cell, and $E_k = n_i(1-\widehat{p}) = 25(.5)$ when cell k is a failure cell.

Note that $X_i = O_i$ for success cells and $n_i - X_i = O_i$ for failure cells.
Therefore $(O_i - E_i)^2 = (X_i - n_i\widehat{p})^2$ for success cells and
$(O_i - E_i)^2 = (n_i - X_i - n_i(1-\widehat{p})^2 = (n_i\widehat{p} - X_i)^2 = (X_i - n_i\widehat{p})^2$ for failure cells.

For the general case, with table

	Success	Failure	Sample sizes
Treatment 1	$O_{11} = X_1$	$O_{12} = n_1 - X_1$	n_1
Treatment 2	$O_{21} = X_2$	$O_{22} = n_2 - X_2$	n_2
...	...	...	...
Treatment r	$O_{r1} = X_r$	$O_{r2} = n_r - X_r$	n_r
Total	S	$n - S$	n

and $\widehat{p} = S/n$ and $(1 - \widehat{p}) = (n - S)/n$, the derivation goes as follows:

$$T = \sum_{i=1}^{r} Z_i^2$$

$$= \sum_{i=1}^{r} \left(\frac{X_i - n_i\widehat{p}}{\sqrt{n_i\widehat{p}(1-\widehat{p})}} \right)^2$$

$$= \sum_{i=1}^{r} \frac{(X_i - n_i\widehat{p})^2}{n_i\widehat{p}(1-\widehat{p})}$$

$$= \frac{1}{(1-\widehat{p})} \sum_{i=1}^{r} \frac{(X_i - n_i\widehat{p})^2}{n_i\widehat{p}}$$

$$= \left(1 + \frac{\widehat{p}}{(1-\widehat{p})}\right) \sum_{i=1}^{r} \frac{(X_i - n_i\widehat{p})^2}{n_i\widehat{p}}$$

$$= \sum_{i=1}^{r} \frac{(X_i - n_i\widehat{p})^2}{n_i\widehat{p}} + \sum_{i=1}^{r} \frac{(X_i - n_i\widehat{p})^2}{n_i(1-\widehat{p})}$$

$$= \sum_{i=1}^{r} \frac{(X_i - n_i\widehat{p})^2}{n_i\widehat{p}} + \sum_{i=1}^{r} \frac{(n_i - X_i - n_i(1-\widehat{p}))^2}{n_i(1-\widehat{p})}$$

$$= \sum_{k} \frac{(O_k - E_k)^2}{E_k}$$

since $E_k = \begin{cases} n_i S/n = n_i\widehat{p} & \text{for success cells} \\ n_i(n-S)/n = n_i(1-\widehat{p}) & \text{for failure cells.} \end{cases}$.

Small sample test for comparing population proportions

Example. Suppose two treatments (Cloudnine and Floataway) for relief of migraine headaches are studied. Suppose 10 migraine sufferers are randomly assigned into 2 groups of five. One group of subjects receive treatment Cloudnine and the other group receives Floataway. The null hypothesis is that the probability of relief is identical for both treatments, that is, $p_c = p_f = p$ (unknown). Given the following results

	Success	Failure
Cloudnine	1	4
Floataway	4	1

what is the probability of results as contradictory (to the null) as those above if the null were true (if the probabilities were identical)? A chi-squared test is not appropriate here since the sample sizes are small. An “exact” test is possible using binomial probabilities.

If H_0 were true, an estimate of the common probability (p) for relief is given by $\widehat{p} = .5$. Below are 2×2 tables with proportions that differ as much as that observed in the data:

$$\text{Tables} \quad \begin{array}{|c|c|}\hline 1 & 4 \\ \hline 4 & 1 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 4 & 1 \\ \hline 1 & 4 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 5 & 0 \\ \hline 0 & 5 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 0 & 5 \\ \hline 5 & 0 \\ \hline\end{array}$$

$$[\text{Prob:} \quad 2\binom{5}{1}\binom{5}{4}(.5)^{10} \quad + \quad 2\binom{5}{0}\binom{5}{5}(.5)^{10}]$$

$$\text{Tables} \quad \begin{array}{|c|c|}\hline 0 & 5 \\ \hline 3 & 2 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 3 & 2 \\ \hline 0 & 5 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 1 & 4 \\ \hline 5 & 0 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 5 & 0 \\ \hline 1 & 4 \\ \hline\end{array}$$

$$[\text{Prob:} \quad 2\binom{5}{0}\binom{5}{3}(.5)^{10} \quad + \quad 2\binom{5}{1}\binom{5}{5}(.5)^{10}]$$

$$\text{Tables} \quad \begin{array}{|c|c|}\hline 2 & 3 \\ \hline 5 & 0 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 5 & 0 \\ \hline 2 & 3 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 0 & 5 \\ \hline 4 & 1 \\ \hline\end{array} \qquad \begin{array}{|c|c|}\hline 4 & 1 \\ \hline 0 & 5 \\ \hline\end{array}$$

$$\text{Prob:} \quad 2\binom{5}{2}\binom{5}{5}(.5)^{10} \quad + \quad 2\binom{5}{0}\binom{5}{4}(.5)^{10}$$

Therefore an “exact” (two-sided) p-value is given by

$$p\text{-value} = \frac{25+1+10+5+10+5}{512} = .109375$$

We note that a two-proportions normal approximation Z-test gives p_z-value $= .05778$ (as does a χ^2 test for homogeneity).

A Minitab (exec) macro for a variation of the above test is given below. It simulates binomial observations (with a random value for p coming from a beta distribution) and compares the observed absolute difference in sample proportions to the absolute difference simulated under the null hypothesis.

```
noecho
#########################################################
#
#        Name:   2prop2sim.mtb
#        Date:  October 11, 2009
#        Author:  Dennis Walsh
#
#        Purpose: This Minitab macro will approximate the p-value
#         for testing the equality of two population proportions.
#
#
#        Hypotheses:      Ho: p1=p2 versus Ha: p1 not= p2.
#
#        Method:  We assume two independent random samples of
#   sizes n1 and n2 with number of successes x1 and x2 respectively. Using
#   a Bayesian approach with the observed data, we assume a
#   beta(x1+x2+1, n1+n2-x1-x2+1) distribution for the unknown common null
#   proportion p. We then generate a value for p from this beta distribution. Next
#   we use that random p to generate an observation s1 from a binomial(n1,p)
#   distribution and an observation s2 from a binomial(n2,p) distribution.
#   Exexuting the macro 10,000 times will provide 10,000 random pairs (s1,s2).
#   Counting the number of times |s1/n1-s2/n2| is greater than or equal to
#   |x1/n1-x2/n2| provides a proportion that is used for the p-value of the test.
#
#        Set up:  Prior to invoking the exec command in Minitab, store
#  the constants x1, n1, x2, n2, 0, and 0 in k101 through k106.
#
#        Results:  After running the macro 10,000 times, view the
#  estimated p-value by printing constant k114. Clean up afterwards by
#  erasing c101-c104 and k101-k104.
#
############################################################
noecho
let k108=abso(k101/k102-k103/k104)
let k109=k101+k103+1
let k110=k102+k104-k109+2
random 1 c101;
beta k109 k110.
let k111=c101(1)
random 1 c102;
binomial k102 k111.
random 1 c104;
binomial k104 k111.
let k112=abso(c102(1)/k102-c104(1)/k104)
let k113=round(.5*(sign(k112-k108)+1))
let k105=k113+k105
let k106=1+k106
let k114=k105/k106
name k114 "estimated p_value"
end
```