BIO 682
Nonparametric Statistics
Spring 2010

Steve Shuster

http://www4.nau.edu/shustercourses/BIO682/index.htm

Lecture 5

Williams' Correction

a. Divide G value by q (see S&R p. 699)

$$q = 1 + (a^2 - 1)/6nv$$

(where $v = a - 1$).

b. In the previous example,

$q = 1 + (9 - 1)/1200 = 1.007$

3. $G_{adj} = 106.5/1.007 = 105.75$

Williams' Correction: Result

1. Note that the adjusted value is *smaller.*

a. i.e., more likely to accept Ho (i.e., is more conservative).

2. Williams correction applies only for situations in which $n < 200$.

a. This is *most of the time*.

Williams' Correction: Result

3. The reason is that with large sample sizes X^2 approaches normality.

c. Williams' Correction doesn't change G much (note that n is in denominator)

$$q = 1 + (a^2 - 1)/6nv$$

When To Use *G*-tests

1. Usually determined by sample size and the magnitude of frequencies.

a. Like X^2, G-tests can't be used when smallest *expected* f_i is < 5.

b. But there are some exceptions.

Advantages of *G*-tests

1. When smallest *expected* $f_i > 10$ (e.g., $f_{i\text{-}hat}$), G-tests give a good approximation of exact multinomial probability.

a. It is as if the probability of observed counts among classes was calculated *exactly*.

2. G-tests have similar interpretations to X^2 except they have the advantage of additivity (more on this with heterogeneity tests).

Further Advantages of *G*-tests

1. For $a > 5$ and $f_{i\text{-}hat} > 3$, G is *better* than X^2
   a. where a = number of classes.
   b. $f_{i\text{-}hat}$ = expected frequency of smallest cell.
   c. This is true because under these conditions *G*-tests simultaneously minimize Type I and Type II errors better than X^2.

When Assumptions Are Violated

1. Exact tests are better then *G*-tests when:
   1. $a > 5$ and $f_{i\text{-}hat} < 3$, or when
   2. $a < 5$ and $f_{i\text{-}hat} < 5$.
2. This can be a problem if one wishes to do heterogeneity tests.
   a. Thus, small $f_{i\text{-}hat}$ can be avoided by lumping classes.

Sex ratio in 6115 sibships of 12 in Saxony. The fourth column gives the expected frequencies, assuming a binomial distribution. These were first computed in Table 5.4 but are here given to five decimal place precision to give sufficient accuracy to the computation of *G*.

(1) ♂♂	(2) ♀♀	(3) f		(4) $\hat{f}$		(5) Deviation from expectation
12	0	7	52	2.34727	28.42973	+
11	1	45		26.08246		
10	2	181		132.83570		+
9	3	478		410.01256		+
8	4	829		854.24665		−
7	5	1112		1265.63031		−
6	6	1343		1367.27936		−
5	7	1033		1085.21070		−
4	8	670		628.05501		+
3	9	286		258.47513		+
2	10	104		71.80317		+
1	11	24	27	12.08884	13.02168	+
0	12	3		0.93284		
		6115 = n		6115.00000		

Sex ratio in 6115 sibships of 12 in Saxony. The fourth column gives the expected frequencies, assuming a binomial distribution. These were first computed in Table 5.4 but are here given to five decimal place precision to give sufficient accuracy to the computation of G.

(1) ♂♂	(2) ♀♀	(3) f		(4) $\hat{f}$		(5) Deviation from expectation
12	0	7	52	2.34727	28.42973	+
11	1	45		26.08246		
10	2	181		132.83570		+
9	3	478		410.01256		+
8	4	829		854.24665		−
7	5	1112		1265.63031		−
6	6	1343		1367.27936		−
5	7	1033		1085.21070		−
4	8	670		628.05501		+
3	9	286		258.47513		+
2	10	104		71.80317		+
1	11	24	27	12.08884	13.02168	+
0	12	3		0.93284		
		6115 = n		6115.00000		

Calculating *G*

$$G = 2\sum^{a} f_i \ln\left(\frac{f_i}{\hat{f}_i}\right) = 2\left[52 \ln\left(\frac{52}{28.42973}\right) + 181 \ln\left(\frac{181}{132.83570}\right) + \cdots + 27 \ln\left(\frac{27}{13.02168}\right)\right]$$

$$= 94.87155$$

Williams' Correction

$$q = 1 + \frac{(a^2 - 1)}{6n\nu}$$

$$= 1 + \frac{(11^2 - 1)}{6(6115)(9)} = 1.000,363,4$$

$$G_{\text{adj}} = \frac{G}{q} = \frac{94.87155}{1.000,363,4} = 94.83709$$

$$G_{\text{adj}} = 94.83709 > \chi^2_{.001[9]} = 27.877$$

The null hypothesis—that the sample data follow a binomial distribution—is therefore rejected decisively.

The Result of Pooling

1. Pooling creates larger $f_{i\text{-}hat}$; this can help.
   2. But, may lose information
   a. The decision is up to the experimenter.
2. For small $f_{i\text{-}hat}$, G may too often reject Ho
   a. This is *without* a correction.
   b. Different authors prefer different tests.

Correction for Continuity

1. This is done by adding and subtracting .5 to observed values ($f_i \pm .5$) to decrease value of G or X^2.
2. S&R consider this procedure likely to make tests too conservative.
3. They recommend Williams correction for n < 25.

Degrees of Freedom

1. Usually is (a - 1) for goodness of fit.

a. This is used when hypothesis is *extrinsic* to data

1. e.g., if there is some *external* hypothesis against which the data are to be tested.
2. Example: genetic data; chance.

Degrees of Freedom

2. when parameters are estimated from the data themselves, the hypothesis is *intrinsic*.

a. Rule of thumb: (a-1) – (the number of parameters estimated).

b. The number of additional estimated parameters depends on the distribution used to test the hypothesis.

Degrees of Freedom

Distribution	Parameters estimated from sample	df
Binomial	$\hat{p}$	$a - 2$
Normal	μ, σ	$a - 3$
Poisson	μ	$a - 2$

When the parameters for such distributions are estimated from hypotheses *extrinsic* to the sampled data, the degrees of freedom are uniformly $a - 1$.

Heterogeneity Tests

1. Test is useful when replicated tests are performed:
   a. e.g., Genetic analyses.
   b. Replicated analyses *of any kind*.
2. Most meaningful with *G*-test due to additivity of *G*-values.
   a. X^2 test is not appropriate.
   b. Neither are exact probability tests.
3. S&R go into detail to demonstrate calculation of G_H
   a. This is unnecessary if individual G_i values are calculated.

Recall That,

$$G = 2\sum^{a} f_i \ln\left(\frac{f_i}{\hat{f}_i}\right)$$

where: a = # of classes (k in other notation).

f_i = observed number of counts in the i-th class.

$f_{i\text{-hat}}$ = expected number of counts in the i-th class; = p_i(N)

with (a – 1) degrees of freedom.

Example:

Replicated tests of a genetic hypothesis:
3:1 phenotype ratio.

a. $p_1 = .75; f_{1\text{-hat}} = p_1(100) = 75$
b. $p_2 = .25; f_{2\text{-hat}} = p_2(100) = 25$

Case	f_1	f'_1	f_2	f'_2	N	G_i	P
1	75	75	25	25	100	0.00	ns
2	81	73.5	17	24.5	98	1.66	ns
3	94	87.8	23	28.2	117	1.72	ns
Σ	250	236.3	65	78.7	315	3.38	

Step 1

Compare observed and expected frequencies to calculate individual values of G_i.

a. $p_1 = .75; f_{1\text{-hat}} = p_1(100) = 75$
b. $p_2 = .25; f_{2\text{-hat}} = p_2(100) = 25$

$$G = 2\sum^{a} f_i \ln\left(\frac{f_i}{\hat{f}_i}\right)$$

Case	f_1	f'_1	f_2	f'_2	N	G_i	P
1	75	75	25	25	100	0.00	ns
2	81	73.5	17	24.5	98	1.66	ns
3	94	87.8	23	28.2	117	1.72	ns
Σ	250	236.3	65	78.7	315	3.38	

Step 2

Case	f_1	f'_1	f_2	f'_2	N	G_i	P
1	75	75	25	25	100	0.00	ns
2	81	73.5	17	24.5	98	1.66	ns
3	94	87.8	23	28.2	117	1.72	ns
Σ	250	236.3	65	78.7	315	3.38	

Sum G_i to get G_T

$G_T = 3.38$, df = $b(a\text{-}1)$; b = # tests; a = # classes per test;
= 3(2-1) = 3

$X^2_{[.05,3]} = 7.82$, ns

Step 3

Case	f_1	f'_1	f_2	f'_2	N	G_i	P
1	75	75	25	25	100	0.00	ns
2	81	73.5	17	24.5	98	1.66	ns
3	94	87.8	23	28.2	117	1.72	ns
Σ	250	236.3	65	78.7	315	3.38	

Add f_i for all classes to calculate G_P;

$G_P = 2[250 \ln (250/236.5) + 65 \ln (65/78.7)]$

$= 1.65$, df $= (a-1) = 1$

$X^2[.05] = 1.84$, ns

Step 4

Case	f_1	f'_1	f_2	f'_2	N	G_i	P
1	75	75	25	25	100	0.00	ns
2	81	73.5	17	24.5	98	1.66	ns
3	94	87.8	23	28.2	117	1.72	ns
Σ	250	236.3	65	78.7	315	3.38	

Calculate G_H as,

$G_T - G_P = G_H$ because $G_T = G_P + G_H$

Step 4

Case	f_1	f'_1	f_2	f'_2	N	G_i	P
1	75	75	25	25	100	0.00	ns
2	81	73.5	17	24.5	98	1.66	ns
3	94	87.8	23	28.2	117	1.72	ns
Σ	250	236.3	65	78.7	315	3.38	

$G_H = G_T - G_P = 3.38 - 1.65 = 1.72$

df $= (a-1)(b-1) = (1)(2) = 2$

$X^2_{[.05,2]} = 5.99$, ns

Step 5: Interpretation

Case	f_1	f'_1	f_2	f'_2	N	G_i	P
1	75	75	25	25	100	0.00	ns
2	81	73.5	17	24.5	98	1.66	ns
3	94	87.8	23	28.2	117	1.72	ns
Σ	250	236.3	65	78.7	315	3.38	

G_i : *all non-significant;* $G_T = 3.38$: *non-significant;* G_H
$G_P = 1.65$: *non-significant*, $G_H = 1.72$: *non-significant*.

There is no evidence of heterogeneity in any aspect of this test.

Another Example: A

	Progenies	♀	♂	*n*		*df*	*G*
A	1	59	41	100		1	3.25773 *ns*
	2	58	42	100		1	2.57104 *ns*
	3	57	42	99		1	2.28150 *ns*
	4	58	40	98		1	3.32497 *ns*
					Total	4	11.43523 $P < 0.025$
	Σ	232	165	397	Pooled	1	11.36160 $P < 0.001$
					Heterogeneity	3	0.07363 *ns*

1. Non-significant G_i and G_H
2. Significant G_T and G_P
3. Indicates that deviations in G_i are individually not large enough to lead to significance.
   a. Summed, however, there is a significant bias.
   b. Also they are consistent in their direction of bias.
   c. There may be another, more appropriate hypothesis.

Another Example: B

B	1	59	41	100		1	3.25773 *ns*
	2	42	58	100		1	2.57104 *ns*
	3	57	42	99		1	2.28150 *ns*
	4	40	58	98		1	3.32497 *ns*
					Total	4	11.43523 $P < 0.025$
	Σ	198	199	397	Pooled	1	0.00252 *ns*
					Heterogeneity	3	11.43272 $P < 0.01$

1. Non-significant G_i, G_P.
2. Significant G_T, G_H
3. Indicates that deviations in G_i are not different from H_o, nor is pooled sample.
   a. ***But*** significant G_H and inspection show that deviations exist and simply *cancel* in calculation of G_P.
   b. G_T is equivalent to Ex. A, but conclusion is different.

Another Example: C

C	1	59	41	100		1	3.25773 *ns*
	2	58	42	100		1	2.57104 *ns*
	3	72	26	98		1	22.46416 $P < 0.001$
	4	73	26	99		1	23.23751 $P < 0.001$
					Total	4	51.53044 $P < 0.001$
	$\sum$	262	135	397	Pooled	1	41.35016 $P < 0.001$
					Heterogeneity	3	10.18027 $P < 0.025$

1. Significant nearly *everything*.

a. Two G_i are significant, two are not.

b. Despite trend toward females, the samples *are* heterogeneous

Another Example: D

D	1	52	48	100		1	0.16004 *ns*
	2	36	64	100		1	7.94580 $P < 0.005$
	3	52	47	99		1	0.25263 *ns*
	4	52	46	98		1	0.36758 *ns*
					Total	4	8.72605 *ns*
	$\sum$	192	205	397	Pooled	1	0.42577 *ns*
					Heterogeneity	3	8.30028 $P < 0.05$

1. Non-significant G_T, G_P, all G_i except one is non-significant (but note their magnitudes).

2. However, there is *still* significant G_H

3. Despite non-signicant values in most individual tests, the single significant G_i is enough to make the entire set of tests heterogeneous.

Other Notes on G_H Tests

1. Since Williams' correction changes the distribution of *G*-values,

a. These corrections are *not appropriate* in heterogeneity tests.

2. Occasionally, G_T is significant, but none of the other indices are.

a. This indicates a generally poor fit of the data.

b. Suggests another hypothesis (or hypotheses) might be better.

More Notes

3. Pin-pointing the source of the heterogeneity:
   a. It is possible to use a post-hoc test (STP; simultaneous test procedure).

4. This involves calculating G_T and adding G_i values stepwise (low to high) until significance is reached.
   a. Procedure is outlined in S&R.

Comparisons of Distributions

1. Break observed and expected distributions into intervals and compare intervals using X^2 or G test.
   a. Sometimes it is possible to make approximations depending on *shape of the distribution*.
   b. Contagious distributions are mostly contained in the first few classes.
   c. It is possible to ignore other classes because their combined contribution to X^2 is small.

2. Generally best to use Kolomogorov-Smirnov test if entire distribution is tested.

Tests of 2 Independent Samples

1. 2x2 tests
   a. These test the hypothesis that two factors have nothing to do with each other.
      1. Thus they are designed to test *independence*
      2. If H_o is *rejected*, it indicates that two samples *do influence* each other

2x2 Tests: Examples

1. The proportion of hybrid or non-hybrid plants attacked by insect A (presence or absence?).
2. Response of operated and non-operated frogs to prey (strike or non-strike?).
3. Occurrence of electromorphs at two loci (segregation or linkage?).
4. In all cases:
   a. Pay attention to the marginal totals.
   b. These let you know which test to use.

2x2 Tests: X^2 test

1. The older method, often replaced by *G*-test, but still useful for figuring out expected frequencies:
   a. Standard Method:
      1. Set up 2x2 table
         a. Say, 100 randomly selected plants (*Wt* and *Hy*), sampled for the presence or absence of insect A.

2x2 Tests: X^2 test

		Plant type		
		Hy	WT	
Insect	-	A	B	A+B
A	+	C	D	C+D
		A+C	B+D	A+B+C+D

2x2 Tests: X^2 test

		Plant type		
		Hy	WT	
Insect	-	4	14	18
A	+	32	50	82
		36	64	100

Calculate expected values as

1. $A_{exp} = [(A+B)(A+C)]/A+B+C+D$
2. A = (18)(36)/100 = 6.48
3. B = (18)(64)/100 = 11.52
4. C = (36)(82)/100 = 29.52
5. D = (64)(82)/100 = 52.48

		Plant type		
		Hy	WT	
Insect	-	A	B	A+B
A	+	C	D	C+D
		A+C	B+D	A+B+C+D

		Plant type		
		Hy	WT	
Insect	-	4	14	18
A	+	32	50	82
		36	64	100

Calculating X^2

		Plant type		
		Hy	WT	
Insect	-	4	14	18
A	+	32	50	82
		36	64	100

$$X^2 = \sum^{r} \sum^{c} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$ with df = (r-1)(c-1)

= 4.52, $P < 0.05$

2x2 Tests: X^2 test

1. Clearly, This is a bit cumbersome
2. Also, when using X^2, it is necessary to apply Yates' Correction for continuity.

a. Reduction or augmentation of observed values by .5 depending on whether they are larger than or smaller than observed.

Calculate expected values as

1. A_{exp} = [(A+B)(A+C)]/A+B+C+D
2. A = (18)(36)/100 = 6.48
3. B = (18)(64)/100 = 11.52
4. C = (36)(82)/100 = 29.52
5. D = (64)(82)/100 = 52.48

		Plant type		
		Hy	WT	
Insect	-	A	B	A+B
A	+	C	D	C+D
		A+C	B+D	A+B+C+D

			Plant type			
			Hy		WT	
Insect	-	+	4	-	14	18
A	+	-	32	+	50	82
			36		64	100

2x2 Tests: X^2 test

1. Clearly, This is ***really*** cumbersome
2. S&R consider this likely to lead to an unnecessarily conservative result.