Hair – Eye Color Data:
Comparing Different Exact Tests

September 21, 2016

In this lesson, we examine a data set of people (students in a large class) who report their hair color and eye color. Our objective is to see if there is some kind of association between the two variables and to try to characterize that association. The original data are shown in Table 1. Also shown are the expected counts under the independence assumption and the standardized Pearson residuals ("z-values"). We have arranged the rows and columns from what I somewhat subjectively have determined is darker to lighter, so they are essentially ordinal variables.

Following typical practice, we test for independence. If we cannot reject the null hypothesis, then there is probably not much point in going further - we don't have strong evidence for an association between hair color and eye color. Any association we think we may see in the data could be due to chance.

The observed value of the the χ^2 test statistic is 138.2925. The degrees of freedom is $\nu = (4-1)(4-1) = 9$. The p-value based on the asymptotic χ^2_9 distribution is $< 2.2 \times 10^{-16}$, essentially 0. There is strong evidence for some kind of association. All of the expected cell counts are quite large (> 5) so there should be no difficulty with the χ^2 approximation of the null distribution. Here is the copy-and-paste from the R-console:

```
> # have already entered the data into a 4 by 4 matrix:
> haireye
      black brown red blond
brown    68   119  26     7
hazel    15    54  14    10
green     5    29  14    16
blue     20    84  17    94
> chisq.test(haireye)

Pearson's Chi-squared test

data:  haireye
X-squared = 138.29, df = 9, p-value < 2.2e-16
```

Looking at the z-values in Table 1 which are outside the range of ± 2, we see that they tend to be positive near the main diagonal and negative away from the diagonal, suggesting a monotone association between the ordinal variables. There are various ways of testing for association including "monotone" association for ordinal variables. We list five approaches here:

X^2 , Pearson's χ^2 for non-ordinal association.

R^2 , Estimated Renyi's Maximal Correlation squared for non-ordinal association.

$\hat{\rho}$, for "linear association" (with some power for monotone association) between the identity scores for the ordinal values.

S , for linear association between the rank scores.

M^2 , the estimated monotone maximal correlation squared.

We would like to compare how these various test statistics perform. There is no known asymptotic null distribution for R^2 and M^2 so we will have to use permutation methods to get estimated p-values from a monte-carlo sample of permutations. Here is a function to perform such monte-carlo approximations for all 5 test statistics. Here is an R function to do this:

```
Exact5 = function(ctab,nmc=10000){
# function to perform 5 tests of association in 2 way table using monte carlo a
# For monotone associations, 1-sided tests of positive association are performe
# rho and spear as defined below.
# INPUTS:
# ctab: I by J matrix of counts (contingency table)
# nmc: number of monte carlo trials to estimate p-values
# OUTPUTS: p-values corresponding to 5 test statistics:
# X2: Pearson chi-squared
# R2: Renyi maximal correlation squared (as estimated by ACE algorithm)
# rho: correlation of identity scores
# spear: Spearman correlation (correlation of rank scores)
# M2: maximal monotone Renyi correlation
##########################################
# convert tabled data to X-Y pairs
```

```
I = nrow(ctab)
J = ncol(ctab)
N = sum(ctab)
X = rep(NA,N)
Y = X
n = 0
for(i in 1:I){ for(j in 1:J){
m = n+ctab[i,j]
X[(n+1):m] = i
Y[(n+1):m] = j
n = m
}}
############################################
# compute test statistics for real data
treal = Teststat5(X,Y)
# accumulator of counts of trand >= treal
pval = rep(0,5)
names(pval) = c("X2","R2","rho","spear","M2")
for(n in 1:nmc){
Xrand = sample(X)
trand = Teststat5(Xrand,Y)
pval = pval + (trand >= treal)
}
pval = pval/nmc
return(pval)
}
```

Here is the R function that computes all 5 test statistics on the X, Y data:

```
Teststat5 = function(X,Y){
# function to compute 5 test stat values for function Exact5
# Note: must install acepack package for ace() function
X2 = chisq.test(as.factor(X),as.factor(Y))$statistic
R2 = ace(as.matrix(X),Y,cat=c(0,1))$rsq
rho = cor(X,Y)
spear = cor(rank(X),rank(Y))
M2 = ace(as.matrix(X),Y,mon=c(0,1))$rsq
```

```
return(c(X2,R2,rho,spear,M2))
}
```

When it was tried on the original haireye data, the results are

```
> Exact5(haireye)
   X2    R2   rho spear    M2
    0     0     0     0     0
```

(That took about a minute of wall time for 10,000 monte carlo trials). Every test shows significance. Can we estimate the monte carlo sampling error? There is a simple way of constructing confidence intervals for binomial probabilities based on the Poisson approximation when the number of successes is small (like 0). When p is small but n is large, Binomial(n, p) is approximately Poisson(λ) with $\lambda = np$. Thinking of a Poisson process with intensity λ, the number of events in the interval $[0, 1]$ has a Poisson(λ) distribution. The probability of no events (approximating the binomial probability of no successes) is $e^{-\lambda}$, so for testing $H_0(\lambda_0)\lambda >= \lambda_0$ vs. $H_1 : \lambda < \lambda_0$ at 0.05 level, we cannot reject $H_0(\lambda_0)$ if $e^{-\lambda_0} \geq .05$, i.e., $\lambda_0 \leq -\log(.05)$. Going back to the binomial setting, if we observe 0 successes, then we get an approximate upper 95% confidence limit of $-log(.05)/n$, which for this case ($n = 10000$) gives 0.0002995732, approximately 0.0003. So we are pretty sure all the true randomization p-values are < 0.0003.

To get something more interesting, we take a smaller sample of $N = 50$ observations and redo the tests:

```
> length(X)
[1] 592
> subs = sample(1:592,size=50)
> X1 = X[subs]
> Y1 = Y[subs]
> table(X1,Y1)
   Y1
X1   1  2  3  4
  1  5  8  3  0
  2  0  7  1  1
  3  0  0  1  0
  4  3 10  2  9
```

```
> # asymptotic approximations wouldn't work well here - too many zeroes
> Exact5(table(X1,Y1))
    X2     R2    rho  spear     M2
0.0981 0.0630 0.0015 0.0022 0.0050
There were 50 or more warnings (use warnings() to see the first 50)
> warnings() # NEVER IGNORE WARNINGS!!!!!!!!!
Warning messages:
1: In chisq.test(as.factor(X), as.factor(Y)) :
  Chi-squared approximation may be incorrect
....................
50: In chisq.test(as.factor(X), as.factor(Y)) :
  Chi-squared approximation may be incorrect
> # DUH! THAT'S WHY WE ARE DOING EXACT TESTS!
```

This is just a single data set, of course. We are extremely confident that the null hypothesis of independence is false in the full data set (of 592 observations). It will be harder to reject the null hypothesis in the smaller data set of 50 observations. The tests based on $X2$ and $R2$ are omnibus - they aren't keyed into particular alternatives. They don't show significance at the 0.05 level. The tests based on rho and $spear$ are specifically geared to finding positive association between ordinal variables, and they do show significance. This is reasonable since we did see trends of positive association in the original table. The $M2$ test is a two sided test of association (because we can't constrain the score functions to be monotone increasing in the `ace` function, only monotone), and the p-value is around twice as big as other two test stats sensitive to positive association. All three tests which are sensitive to the monotone association alternatives do give significant results, though.

There is a lot more potential research on the issues raised in this lesson. One could compare powers (since the tests are exact, we know the level is fixed) at select alternatives in a monte-carlo study. One could consider other test stats (like Wilks' G^2, various Wald tests, correlations using other scoring methods, etc.).

		Hair Color				
Eye Color	Value	Black	Brown	Red	Blond	Row Sums
Brown	Obs.	68	119	26	7	220
	Exp.	40.14	106.28	26.39	47.20	
	z-value	6.13	2.16	-0.10	-8.33	
Hazel	Obs.	15	54	14	10	93
	Exp.	16.97	44.93	11.15	19.95	
	z-value	-0.58	2.05	0.99	-2.74	
Green	Obs.	5	29	14	16	64
	Exp.	11.68	30.92	7.68	13.73	
	z-value	-2.29	-0.51	2.57	0.73	
Blue	Obs.	20	84	17	94	215
	Exp.	39.22	103.87	25.79	46.12	
	z-value	-4.25	-3.40	-2.31	9.97	
Column Sums		108	286	71	127	592

Table 1: The Hair/Eye Color data. For each combination of Hair and Eye color, we show the observed number of subjects, the expected number of subjects, and the z-value.