Hypothesis Tests

We have seen a couple of hypothesis test so far. Here we will quickly review them and introduce a couple more.

Fisher's Exact test – another name for the test that compares the observed counts in a 2 by 2 table to those one would expect under the hypergeometric.

Example: See the Pepsi/Pepsi One example.

χ^2 **test of Independence** – We have seen this test used with an I by J table of counts. The independence assumption says that the probability $p_{i,j}$ of a subject falling into cell (i, j) is the product of π_i, the chance they are in row i, and γ_j, the chance they are in column j.

Example: See the dove/hawk example.

χ^2 **test of Homogeneity** – Here we sample units from I groups independently. Each unit falls into one of J categories. The assumption of homogeneity says that the chance a unit is of type j if π_j, $j = 1, \ldots, J$ regardless of which population being sampled, i.e. the populations are homogeneous with respect to the factor under scrutiny.

Example: Wright and Stearns county are neighboring counties. It was found that 15 of the 27 houses sampled in Stearns had radon levels in excess of 4 picoCuries/liter, and 9 of the 14 houses sampled in Wright county exceeded this level.

		Threshold		
		Below	Above	Total
County	Stearns	12	15	27
	Wright	5	9	14
	Total	17	24	41

Here

$I =$

$J =$

$n_1 =$

$n_2 =$

$\pi_1 =$ unknown

$\hat{\pi}_1 =$

Degrees of freedom $= 41 -$ ______ $-$ ________.

χ^2 **goodness-of-fit test** –

Example Addresses are sampled at random from a phone book. The leading digit of the street number is recorded. What distribution do you think these values should follow?

The table below contains the observed counts:

Digit	1	2	3	4	5	6	7	8	9
Counts	21	22	14	12	7	8	5	2	0

z **test** – The z here refers to the normal distribution. A z-test refers to the case when the test statistic follows a standard normal distribution.

Example (completely made up) IQs are normalized to have a mean of 100 and an SD of 15. A random sample of 50 Cal students is taken and it is found that their average IQ is 115. Is this difference just due to chance error?

Two sample z test – This is a special case of the z test when the statistic is a difference of two independent normals. Typically, we are testing the hypothesis that the means of two populations are the same. We don't know the means, and so we use the fact that the difference of the means should be 0 to construct a hypothesis test.

Example Revisit the Stearns/Wright county example above. We could do a two-sample hypothesis test instead of the chi-square test. With a null hypothesis of $\pi_x = \pi_y$. Then $\hat{\pi}_x = .556$, $\hat{\pi}_y = .643$, and the estimate of the variance of $\hat{\pi} - \hat{\pi}_y$ is

$$Var(\hat{\pi}_x) + Var(\hat{\pi}_y) = .556(1 - .556)/27 + .643(1 - .643)/14 = .16^2$$

So the test statistic is

$$(0.556 - 0.643)/0.16 = -0.54$$

The p-value is then 0.59. Note that this is the same p-value as for the χ^2 test. The two tests are identical in this case, and both are an approximation to the exact hypergeometric test.