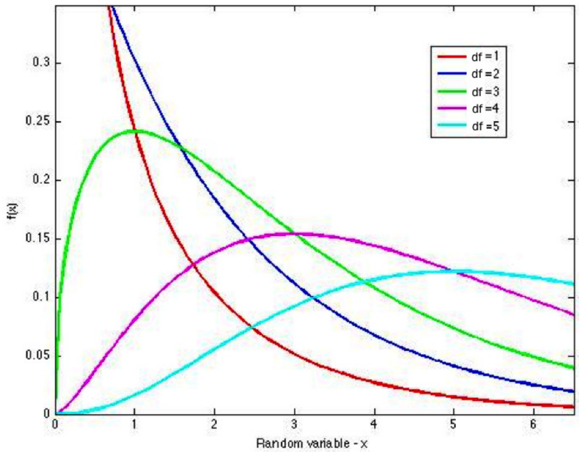
The Girl Scout Cookie Sales Goodness-of-Fit Test

Mark Harbison. Sacramento City College. harbism@scc.losrios.edu

To test how well a list of observed counts "fits" an expected distribution (by categories), compare $\chi^2 = \sum\left(\frac{(\text{obs}-\exp)^2}{\exp}\right)$ to a value on the following table:

	Right Tail Area $\chi^2_{sub...}$				
	0.100	0.050	0.025	0.010	0.005
d.f.	Critical Values (χ^2_c)				
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.60
3	6.251	7.815	9.348	11.34	12.84
4	7.779	9.488	11.14	13.28	14.86
5	9.236	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.95
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.72	26.76
⋮					



distributic
has a total area of 100% under the curve, so that area = probability).

For example, on Mar. 3, 2007 (from 3 to 6 pm) we *observed* a certain local girl scout troop sell { 71, 59, 29, 23, 20, and 35 } boxes of {Thin Mints, Samoas, P.B. Patties, P.B. Sandwich, Shortbread, and "Other"} flavors (respectively), where "Other" = either Cafe or All Abouts or Little Brownies flavors. n = 237 total boxes.

According to www.gs.org with that total, we *expect* to sell 237 * { 25%, 19%, 13%, 11%, 9% and 23% } of those flavors (respectively)
= { 59.25 , 45.03 , 30.81 , 26.07 , 21.33 , 54.51 } boxes.

So $\chi^2 = \frac{(71-59.25)^2}{59.25}+\frac{(59-45.03)^2}{45.03}+\frac{(29-30.81)^2}{30.81}+\frac{(23-26.07)^2}{26.07}+\frac{(20-21.33)^2}{21.33}+\frac{(35-54.51)^2}{54.51} \cong 14.20$

which is more than the 11.07 from the table OR 14.20 < 15.09 from the table. If we use "Area = 0.05" then we can safely conclude that at least 1 flavor of G.S. cookie is sold locally with a different frequency than expected.

We do not need to specify which particular flavor had different sales.

Or (since 14.20 15.09) at the Area = 0.01 "level of significance" there is NOT enough evidence to conclude that there is any difference (by flavor) in local sales than expected.

This method can also apply to other situations like "is the genetic distribution the same as Mendel expected?" or "are local types of traffic accidents the same as expected statewide?"

Technically, we should meet 2 conditions before using this test:

- the observed counts are based on a random sample, and
- the sample size is large enough that each expected count is sufficiently large (most texts use 5 as a minimum allowed exp).

Optionally, the table values can be verified with these formulas

(the area at $\chi^2 = x$) = $\int_0^x \frac{1}{\Gamma(r/2) * 2^{r/2}} w^{r/2-1} e^{-w/2} dw$ with d.f. = r and

$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y}\, dy$ for real $t > 0$ which simplifies to $\Gamma(n) = (n-1)!$ for natural n • 1.

Bonus example: on Feb 25, 2011 (5-7 pm), the troop sold the following # of boxes (of the same flavors) { 41, 45, 14, 12, 8, 22 } (n = 142).