

13.1 Test for Goodness of Fit

color	observed count O	^{H₀:} hypothesized proportion	expected count E	component $\frac{(O - E)^2}{E}$
white	144	$\frac{1}{6}$	$1021/6$	4.024
pink	216	$\frac{1}{6}$	$1021/6$	12.345
orange	184	$\frac{1}{6}$	$1021/6$	1.125
yellow	187	$\frac{1}{6}$	$1021/6$	1.665
green	88	$\frac{1}{6}$	$1021/6$	39.675
purple	202	$\frac{1}{6}$	$1021/6$	5.955
	<u>+ 1021</u>		<u>+ 1021</u>	<u>64.788</u>

Table C χ^2 critical values

df	Tail probability <i>p</i>										
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51
6	7.84	8.56	9.46	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.96	37.70
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.96	40.79
18	21.60	22.76	24.16	25.99	28.87	31.53	32.36	34.81	37.16	39.42	42.31
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18
25	29.34	30.68	32.28	34.38	37.65	40.66	41.57	44.31	46.93	49.44	52.62
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40
50	56.33	58.16	60.36	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4

P POP: Smarties

H₀: Colors are evenly distributed

H_a: " " not " "

A Random sample is assumed here.

IT - Dun 1021 smarties are < 10% of all smarties

AL EN - no ^{expected} counts are 0 & no more than 1/5 of them are under 5. In fact the expected is over 170.

N

Chi-square test for goodness of fit.

$$T \sum \frac{(O - E)^2}{E} = \frac{(144 - 170.17)^2}{170.17} + \frac{(216 - 170.17)^2}{170.17} + \dots$$

$$= 64.788$$

O

$$p\text{value} = 1.24 \times 10^{-12}$$

M

Since our p value is very low (< .05), we'll reject H₀.

S

We have strong evidence that the colors of smarties are not evenly distributed.

Tom's might buy a small gas station and wants to plan for gasoline sales throughout the year. He asks the seller for sales data. A SRS shows 18,448 gallons of gas sold last year. Tom believes sales would be fairly consistent from month to month, in fact, essentially uniform. How could we test this belief?

Q: If sales **ARE** uniform and total 18,448 gallons, what would we expect of these sales?

A: $\frac{18448 \text{ gallons}}{12 \text{ months}} = 1537.3 \text{ gallons/month}$

We would expect sales of 1537.333 gallons of gas per month.

If sales really **ARE** uniformly distributed, how unusual are these data?

Month	Gallons
January	1,610
February	1,581
March	1,649
April	1,590
May	1,540
June	1,397
July	1,416
August	1,350
September	1,495
October	1,564
November	1,601
December	1,655
	18,448

pg. 732

A **goodness of fit test** helps us decide whether a population has some distribution categories.

Suppose a hypothesized distribution has n outcome categories. To test the hypothesis

H_0 : the actual population proportions are equal to the hypothesized proportions

first calculate the chi-square test statistic:
$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

which has approximately a chi-square distribution with $(n-1)$ degrees of freedom.

For a test of H_0 against the alternative hypothesis,

H_a : the actual population proportions are *different* from the hypothesized proportions,
the P-value is $P(\chi^2 \geq X^2)$

Conditions:

- Random
- Independent (Ten % or less)
- Large enough n that all individual expected counts are at least 1 & no more than 1/5 of expected counts are under 5

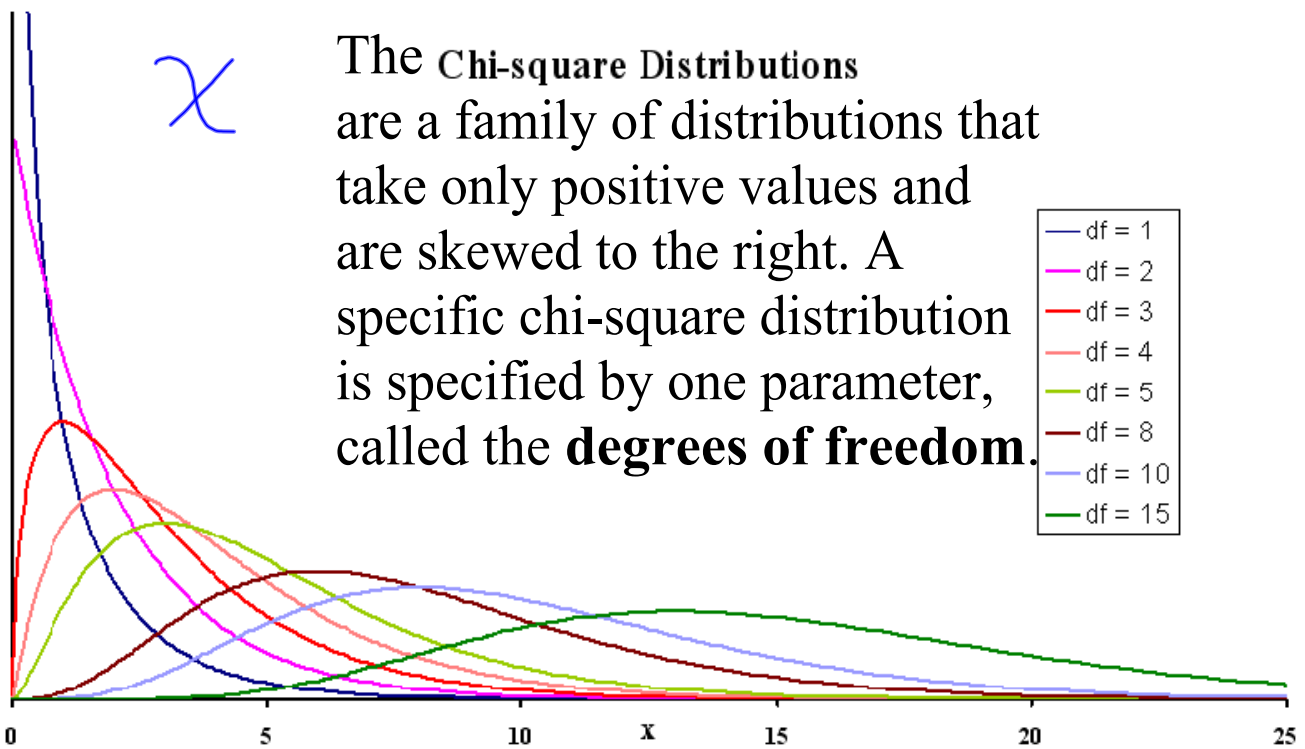
Step 1: ID the pop., parameter referenced, H_0 and H_a :

Step 2: Choose the appropriate inference procedure and verify the conditions:

Step 3: If conditions met, carry out inference procedure:

Month	Observed	Expected
January	1610	1537.333
February	1581	1537.333
March	1649	1537.333
April	1590	1537.333
May	1540	1537.333
June	1397	1537.333
July	1416	1537.333
August	1350	1537.333
September	1495	1537.333
October	1564	1537.333
November	1601	1537.333
December	1655	1537.333
	18448	18448

Step 4: Interpret the results in the context of the problem.



Lion #39: Stalking Role Distribution

A	B	C	D	E	F	G
0	9	15	56	54	10	1