STATISTICAL TESTING

1. **Box Model:** Every legitimate test of significance involves a box model. The test gets at the question of whether an observed difference is real, or just a chance variation. A real difference is one that says something about the box, and doesn't just reflect a fluke of sampling.

2. **Null Hypothesis:** The null hypothesis says that an observed difference just reflects chance variation. The alternate hypothesis says that the observed difference is real.

3. **Test Statistic:** A test statistic measures the difference between the data and what is expected if the null hypothesis is true. You must be able to compute the probability values for the test statistic; that is, the test statistics must follow the normal curve (z-test), or a t-curve (t-test), or a chi-squared curve (χ^2 - test), or some other known probability variable.

4. **P-Value:** The probability value or the observed significance level of a test is the chance or probability of getting the test statistic as extreme as or more extreme than the observed one. The probability is computed on the basis that the null hypothesis is correct. The smaller this probability is, the stronger the evidence against the null hypothesis.

The Chi-Square Test:

1. A die is rolled 60 times. The following results were observed. Is the die fair?

face	1	2	3	4	5	6
frequency	4	6	17	16	8	9

The Chi-Squared (χ^2) statistic can be used to test the hypothesis that data were generated according to a particular chance model.

$$\chi^2 = \text{sum of } \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

The Chi-Square Test:

1. A die is rolled 60 times. The following results were observed. Is the die fair?

face	1	2	3	4	5	6
frequency	4	6	17	16	8	9
expected	10	10	10	10	10	10

Null hypothesis: die is fair

$$\chi^2 = \frac{(4-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(17-10)^2}{10} + \frac{(16-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(9-10)^2}{10}$$

$= 14.2$, $df = 5$

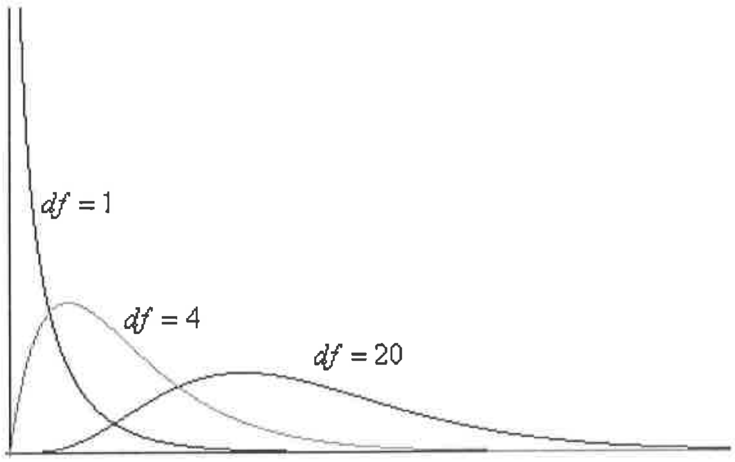
p-value $\approx 2\%$, reject the null.

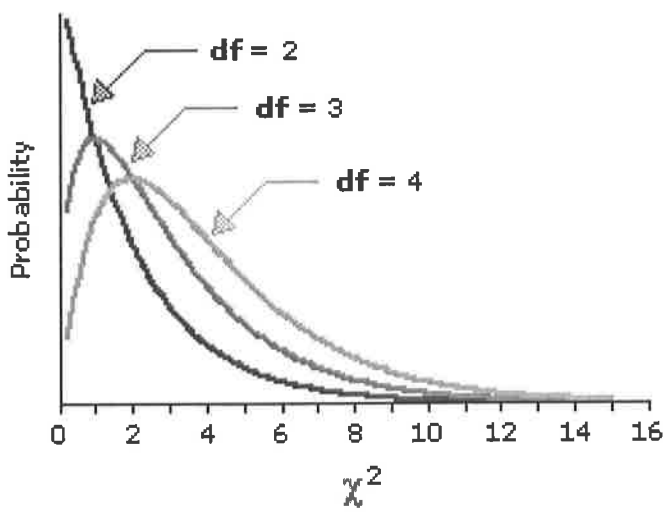
The die is probably biased.

The Chi-Squared (χ^2) statistic can be used to test the hypothesis that data were generated according to a particular chance model.

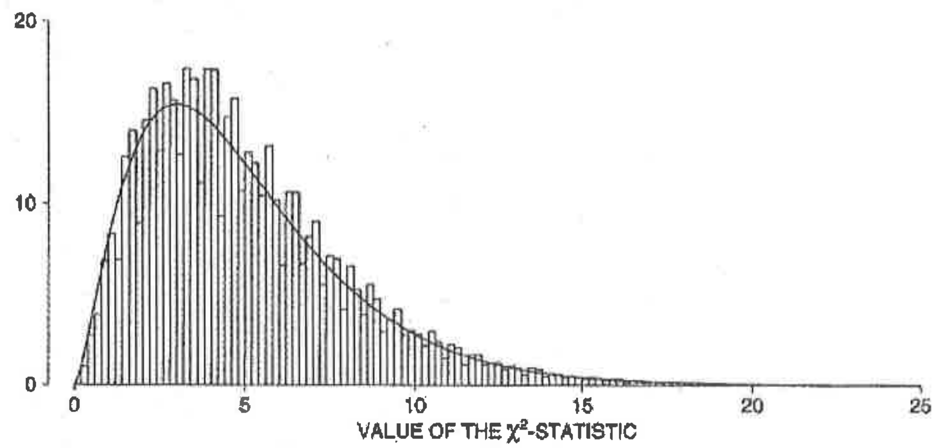
$$\chi^2 = \text{sum of } \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

Chi Squared Curves:

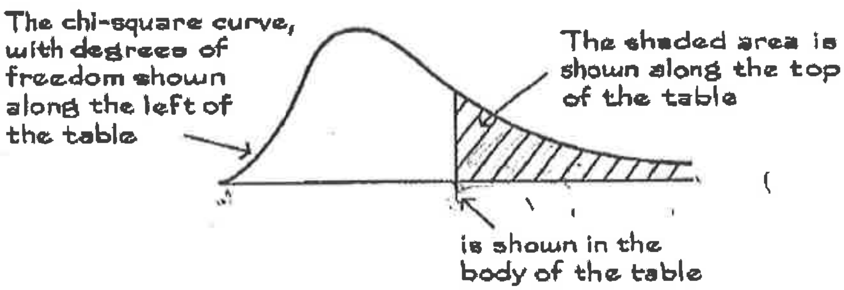




Rolling a fair die – the probability histogram and the χ^2 curve with 5 degrees of freedom



A CHI-SQUARE TABLE



Degrees of freedom	99%	95%	90%	70%	50%	30%	10%	5%	1%
1	0.00016	0.0039	0.016	0.15	0.46	1.07	2.71	3.84	6.64
2	0.020	0.10	0.21	0.71	1.39	2.41	4.60	5.99	9.21
3	0.12	0.35	0.58	1.42	2.37	3.67	6.25	7.82	11.34
4	0.30	0.71	1.06	2.20	3.36	4.88	7.78	9.49	13.28
5	0.55	1.14	1.61	3.00	4.35	6.06	9.24	11.07	15.09
6	0.87	1.64	2.20	3.83	5.35	7.23	10.65	12.59	16.81
7	1.24	2.17	2.83	4.67	6.35	8.38	12.02	14.07	18.48
8	1.65	2.73	3.49	5.53	7.34	9.52	13.36	15.51	20.09
9	2.09	3.33	4.17	6.39	8.34	10.66	14.68	16.92	21.67
10	2.56	3.94	4.86	7.27	9.34	11.78	15.99	18.31	23.21
11	3.05	4.58	5.58	8.15	10.34	12.90	17.28	19.68	24.73
12	3.57	5.23	6.30	9.03	11.34	14.01	18.55	21.03	26.22
13	4.11	5.89	7.04	9.93	12.34	15.12	19.81	22.36	27.69
14	4.66	6.57	7.79	10.82	13.34	16.22	21.06	23.69	29.14
15	5.23	7.26	8.55	11.72	14.34	17.32	22.31	25.00	30.58
16	5.81	7.96	9.31	12.62	15.34	18.42	23.54	26.30	32.00
17	6.41	8.67	10.09	13.53	16.34	19.51	24.77	27.59	33.41
18	7.00	9.39	10.87	14.44	17.34	20.60	25.99	28.87	34.81
19	7.63	10.12	11.65	15.35	18.34	21.69	27.20	30.14	36.19
20	8.26	10.85	12.44	16.27	19.34	22.78	28.41	31.41	37.57

Source: Adapted from p. 112 of Sir R. A. Fisher, *Statistical Methods for Research Workers* (Edinburgh: Oliver & Boyd. 1958).

3. One study of grand juries in Alameda County, California, compared the demographic characteristics of jurors with the general population to see if the jury panels were representative. The results for 86 jurors are given below. Were these 86 jurors like a random sample from the population of Alameda County? (The county age distribution is known from Public Health Department data.)

Age	County %	Number Jurors	
21 - 40	42	15	
41 - 50	23	19	
51 – 60	16	19	
61 and up	19	33	

4. One study of grand juries in Alameda County, California, compared the demographic characteristics of jurors with the general population to see if the jury panels were representative. The results for 86 jurors are given below. Were these 86 jurors like a random sample from the population of Alameda County? (The county age distribution is known from Public Health Department data.)

Age	County %	Number Jurors	expected
21 - 40	42	15	36
41 - 50	23	19	20
51 – 60	16	19	14
61 and up	19	33	16

86

Null: Chance variation only, fair distribution

χ^2 = sum of

$$\frac{(15-36)^2}{36} + \frac{(19-20)^2}{20} + \frac{(19-14)^2}{14} + \frac{(33-16)^2}{16}$$

$= 32.14$ $\qquad df = 3$

p-value ≈ zero.

Reject!!

2. Genetic theory shows that when two pink snapdragons are crossed, we should expect ½ the offspring to be pink, ¼ to be red and ¼ to be white. Suppose we learn of an experiment in which 98 are red, 101 are white, and 201 are pink. Is there evidence that this research is less than honest?

2. Genetic theory shows that when two pink snapdragons are crossed, we should expect ½ the offspring to be pink, ¼ to be red and ¼ to be white. Suppose we learn of an experiment in which 98 are red, 101 are white, and 201 are pink. Is there evidence that this research is less than honest?

	pink	red	white
observed	201	98	101
expected	200	100	100

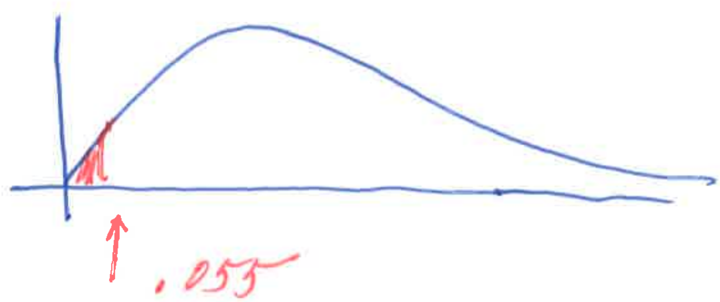
Null: Honest research

Alternate: The experimental data is to close to the expected frequencies

$$\chi^2 = \text{sum of } \frac{(\text{observed frequency} - \text{expected } f)^2}{\text{expected frequency}}$$

$$= \frac{1}{200} + \frac{4}{100} + \frac{1}{100} = \frac{11}{200} = .055$$

df = 2



p-value ≈ 3 %

We reject the null, data was likely fudged.

4. A die is rolled 60 times. The following results were observed. Is the die fair?

	face	1	2	3	4	5	6
observed	frequency	9	10	8	10	11	12
expected freq		10	10	10	10	10	10

$$\chi^2 = \frac{1}{10} + \frac{0}{10} + \frac{4}{10} + \frac{0}{10} + \frac{1}{10} + \frac{4}{10} = 1$$

df = 5, p-value ≈ 4%

Testing Independence

Sometimes, we have two variables of interest and we know how many times each combination of values occurs in our sample. If the sample is a simple random sample from the population, we can test:

null hypothesis: the two variables are independent in the population

alt hypothesis: the two variables are NOT independent in the population

In this case, we get expected counts by assuming independence and we compute the test statistic in the usual way, except that we use

$$df = (\text{number of rows} - 1) \times (\text{number of cols} - 1).$$

5. To test the effectiveness of a new drug designed to reduce mental depression, a pharmaceutical company selects 500 subjects and randomly selects a treatment group of size 300 with the remaining 200 assigned to the control group. The results are tabulated below. Is the drug effective?

	Felt Better	No Effect	Worse	Totals
New Drug	50	240	10	300
Placebo	25	160	15	200
Totals	75	400	25	500

5. To test the effectiveness of a new drug designed to reduce mental depression, a pharmaceutical company selects 500 subjects and randomly selects a treatment group of size 300 with the remaining 200 assigned to the control group. The results are tabulated below. Is the drug effective?

	Felt Better	No Effect	Worse	Totals
New Drug	50 (45)	240 (240)	10 (15)	300
Placebo	25 (30)	160 (160)	15 (10)	200
Totals	75	400	25	500

Null: variables are independent, no difference

How do we compute the expected frequencies?

$\frac{75}{500} \cdot 300 = 45$ This is the expected number who report they felt better.

$\frac{75}{500} \times 200 = 30$

$\frac{400}{500} \times 300 = 240$

$\frac{25}{500} \times 300 = 15$

χ^2, df = 2 = $(2-1) \times (3-1) = 2$

$$\chi^2 = \frac{(50-45)^2}{45} + \frac{0^2}{240} + \frac{(10-15)^2}{15} + \frac{(25-30)^2}{30} + \frac{0^2}{160} + \frac{(15-10)^2}{10}$$

$= 5.55$

5.55

p-value $\approx 6\%$

.06

6. Are you more likely to be left handed if you are female? Are women less likely to be ambidextrous? HANES took a probability sample 2237 Americans and the results are summarized below. What do you conclude?

	Men	Women	Totals
Right Handed	934	1070	2004
Left Handed	113	92	205
Ambidextrous	20	8	28
Totals	1067	1170	2237

6. Are you more likely to be left handed if you are female? Are women less likely to be ambidextrous? HANES took a probability sample 2237 Americans and the results are summarized below. What do you conclude?

	Men	Women	Totals
Right Handed	934 (956)	1070 (1048	2004
Left Handed	113 (98)	92 (107)	205
Ambidextrous	20 (13)	8 (15)	28
Totals	1067	1170	2237

M 10.5% left

W 7.8%

Null: no difference

How do we compute the expected frequencies?

$\frac{1067}{2237} \times 2004 \doteq 956$ ← expected number of right-handed men

↑ proportion of men

↑ total right-handed

$\frac{1067}{2237} \times 205 = 98$ ← expected number of left-handed men

↑ proportion of men

↑ total left-handed

df = 2, $\chi^2 = 11.8$, p-value = .0027

7. A simple random sample of 200 Utah schoolchildren is implemented and each child is asked whether or not they like math. There are 102 boys, of whom 41 like math, and 98 girls, of whom 29 like math. Is liking-math independent of gender for Utah schoolchildren?

7. A simple random sample of 200 Utah schoolchildren is implemented and each child is asked whether or not they like math. There are 102 boys, of whom 41 like math, and 98 girls, of whom 29 like math. Is liking-math independent of gender for Utah schoolchildren?

	like math	don't	
Girls	29 (34) (34.3)	69 (64) (63.7)	98
Boys	41 (36) (35.7)	61 (66) (66.3)	102
	70	130	

$$\frac{70}{200} \times 98 = 34.3 \approx 34$$

$$\frac{70}{200} \times 102 = 35.7 \approx 36$$

$$\frac{130}{200} \times 98 = 63.7 \approx 64$$

$$\frac{130}{200} \times 102 = 66.3 \approx 66$$

$$\chi^2 = \frac{(29-34)^2}{34} + \frac{(69-64)^2}{64} + \frac{(41-36)^2}{36} + \frac{(61-66)^2}{66}$$

$$= 2.199 \quad , \quad df = (r-1)(c-1) = 1$$

p-value $\approx 12\%$

The M&Ms website no longer posts the proportions of colors found in bags of M&Ms. The following is their response to the inquiry from the fall of 2008.

Our color blends were selected by conducting consumer preference tests, which indicate the assortment of colors that pleased the greatest number of people and created the most attractive overall effect.

On average, our mix of colors for M&M'S CHOCOLATE CANDIES is:

M&M'S MILK CHOCOLATE: 24% cyan blue, 20% orange, 16% green, 14% bright yellow, 13% red, 13% brown.

M&M'S PEANUT: 23% cyan blue, 23% orange, 15% green, 15% bright yellow, 12% red, 12% brown.

M&M'S KIDS MINIS: 25% cyan blue, 25% orange, 12% green, 13% bright yellow, 12% red, 13% brown.

M&M'S DARK: 17% cyan blue, 16% orange, 16% green, 17% bright yellow, 17% red, 17% brown.

M&M'S PEANUT BUTTER and ALMOND: 20% cyan blue, 20% orange, 20% green, 20% bright yellow, 10% red, 10% brown.

Each large production batch is blended to those ratios and mixed thoroughly. However, since the individual packages are filled by weight on high-speed equipment, and not by count, it is possible to have an unusual color distribution.

	24%	20%	16%	14%	13%	13%	
	B	O	G	Y	R	Br	
observed	15	10	12	5	5	8	55
expected	13.2	11	8.8	7.7	7.15	7.15	

Null: Hershey's claim is correct, the %'s are OK.

$$\chi^2 = \text{sum of } \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$= \frac{(15-13.2)^2}{13.2} + \frac{(10-11)^2}{11} + \frac{(12-8.8)^2}{8.8}$$

$$+ \frac{(5-7.7)^2}{7.7} + \frac{(5-7.15)^2}{7.15} + \frac{(8-7.15)^2}{7.15}$$

$$= 3.194 \qquad df = 5$$

p-value 67%