Research Methods (HDFS 3390), Alan Reifman, Texas Tech University

Intro to Statistics

Education scholars David Berliner and Bruce Biddle, in their 1995 book, *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools*, make the following argument (p. 316):

...we cannot understand why a person who dislikes mathematics and does not want to work in a science field should be forced to take calculus... If we had to nominate a topic in mathematics that is needed today by all informed citizens, it would be statistics.

Similarly, Arthur Benjamin, the "Mathemagician," endorses the study of statistics and probability as the most useful aspect of mathematics for most people

For Quantitative/Numerical Variables (Ordinal or Ratio)

EXAMPLES

- **No. of Years of Education**
- **No. of Facebook Friends**
- **Agreement (from 1 = Strongly Disagree to 7 = Strongly Agree) that the President of the U.S. is doing a good job**

Descriptive Statistics

Measures of Central (Typical) Tendency

MEAN (average): Add up all scores and divide by number of people (most familiar).

MEDIAN: Score that same number of people fall above and fall below.

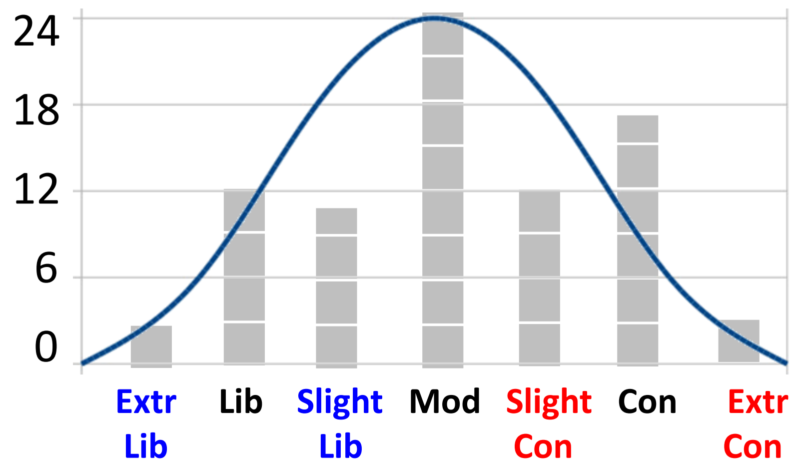
MODE (primarily for nominal variables): Most frequently occurring score.

Mean, Median, and Mode are the same in a normal, bell-shaped distribution (illustrated on this website, scroll down to heading "Normal Distribution")

Even though the Bell Curve is a well-known statistical concept, many (if not most) variables do not follow it.

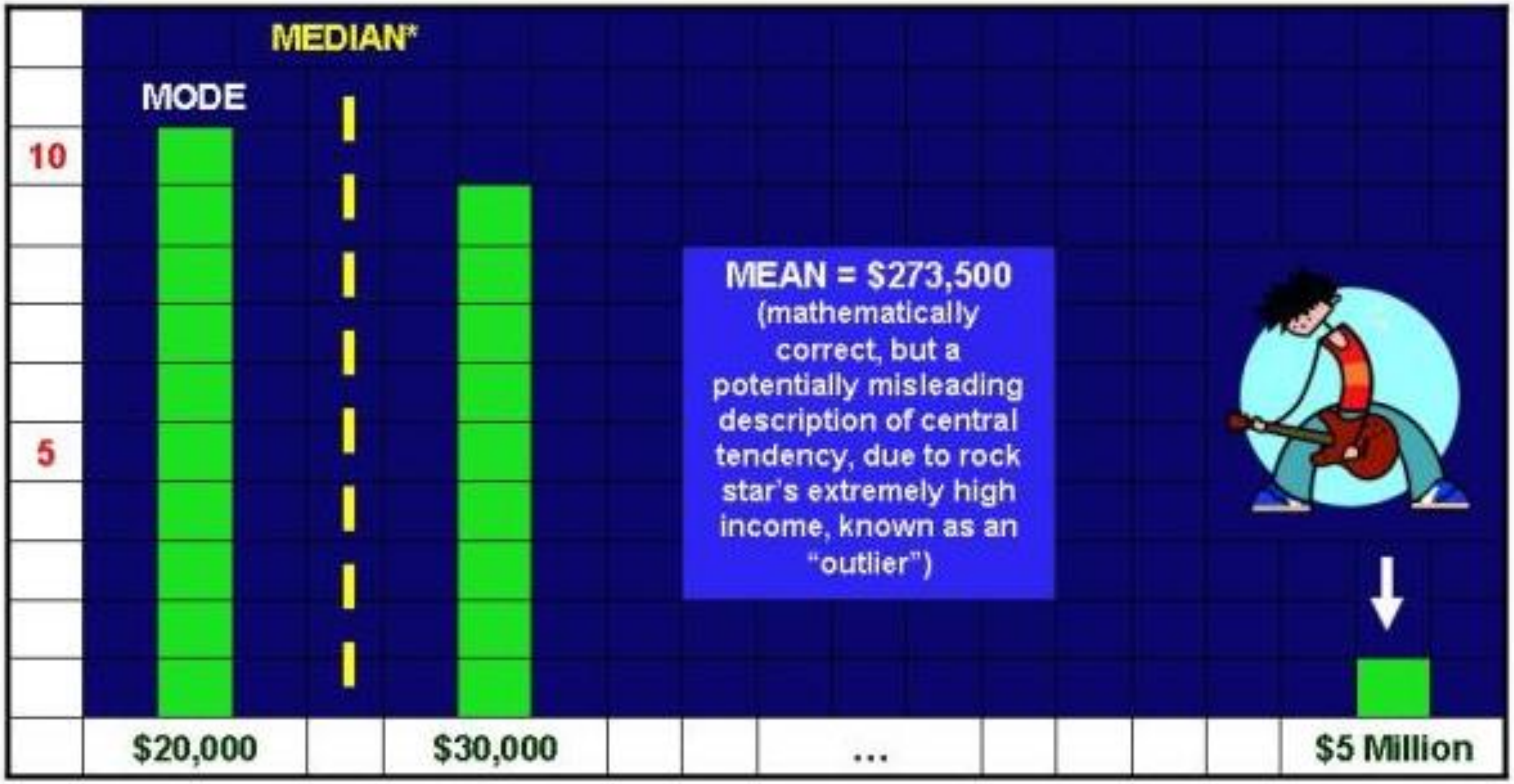
Examples of Variables That Tend to Be...	
Normally Distributed	**Non-Normally Distributed**
Americans' Political Ideology (see next slide) **People's Height**	**People's Numbers of Sexual Partners** (Fig. 2) **People's Alcohol Consumption**
Students in my graduate statistics class should also see here.	

2016 – "Relatively" Normal Distribution



	'94	'96	'98	'00	'02	'04	'08	'12	'16
Extremely Liberal	1 (21)	1 (20)	2 (25)	2 (14)	2 (26)	2 (28)	3 (64)	3 (169)	3 (110)
Liberal	6 (109)	7 (116)	7 (89)	9 (76)	12 (175)	9 (105)	10 (227)	10 (603)	12 (436)
Slightly Liberal	7 (130)	10 (175)	9 (121)	9 (77)	9 (127)	8 (94)	9 (205)	11 (631)	11 (404)
Moderate, Middle of Road	26 (465)	24 (411)	29 (365)	23 (194)	22 (329)	25 (304)	22 (506)	32 (1859)	24 (863)
Slightly Conservative	14 (249)	15 (256)	15 (196)	12 (98)	10 (153)	12 (144)	12 (278)	14 (842)	12 (426)
Conservative	19 (333)	15 (253)	13 (160)	15 (130)	21 (307)	16 (196)	17 (390)	18 (1047)	17 (615)
Extremely Conservative	3 (54)	3 (49)	2 (31)	3 (28)	4 (53)	3 (36)	3 (71)	4 (250)	3 (125)
DK, Haven't Thought	24 (423)	25 (433)	23 (293)	27 (231)	22 (321)	25 (304)	25 (578)	8 (470)	18 (650)
N	1784	1713	1280	848	1491	1211	2319	5871	3629

Distribution of Annual Income from
a Small High School Class 5 Years After Graduation
(total = 20 people; each little green square = one person)



*Anywhere between $20,000 and $30,000 would meet definition of median (10 above and 10 below), but $25,000 is probably most convenient to use as median.

The Mean

Lyrics by Alan Reifman

(May be sung to the tune of "Jolene," Dolly Parton)

The mean, the mean, the mean, the mean,
You add up all the scores, and divide by *N*,
The mean, the mean, the mean, the mean,
The average you've computed, time and again,

You typically make sense to heed,
But with outliers, you mislead,
Because all exact values, it must know,

The median, won't show much swing,
How far away, is not its thing,
Just so you've half above, and half below,

What comes up most, is called the mode,
For nominal, the only road,
Sometimes, more than one max, is in your frame,

When data shapes, follow the bell,
There's only one thing, left to tell,
Mean, median, and mode, are all the same,

The mean, the mean, the mean, the mean,
A useful stat to know, but it's not all,
The mean, the mean, the mean, the mean,
Into a trap, you do not want to fall...

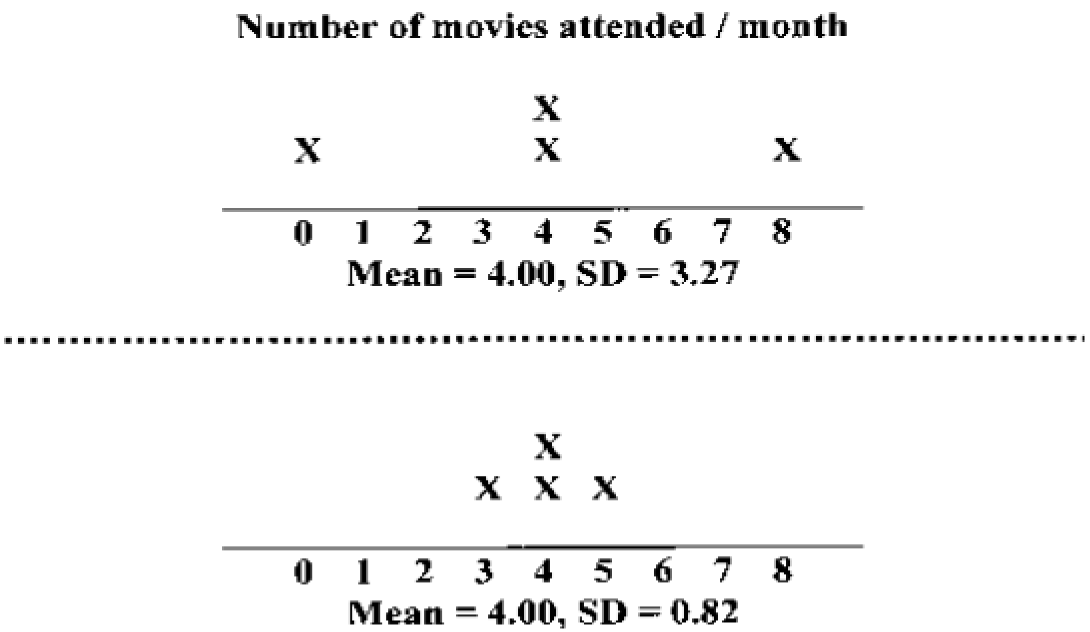
You typically make sense to heed,
But with outliers, you mislead,
Because all exact values, it must know,

The median, won't do the same,
How far away, is not its game,
Just so you've half above, and half below,

The mean, the mean, the mean, the mean,
A useful stat to know, but it's not all,
The mean, the mean, the mean, the mean,
Into a trap, you do not want to fall...

Standard Deviation

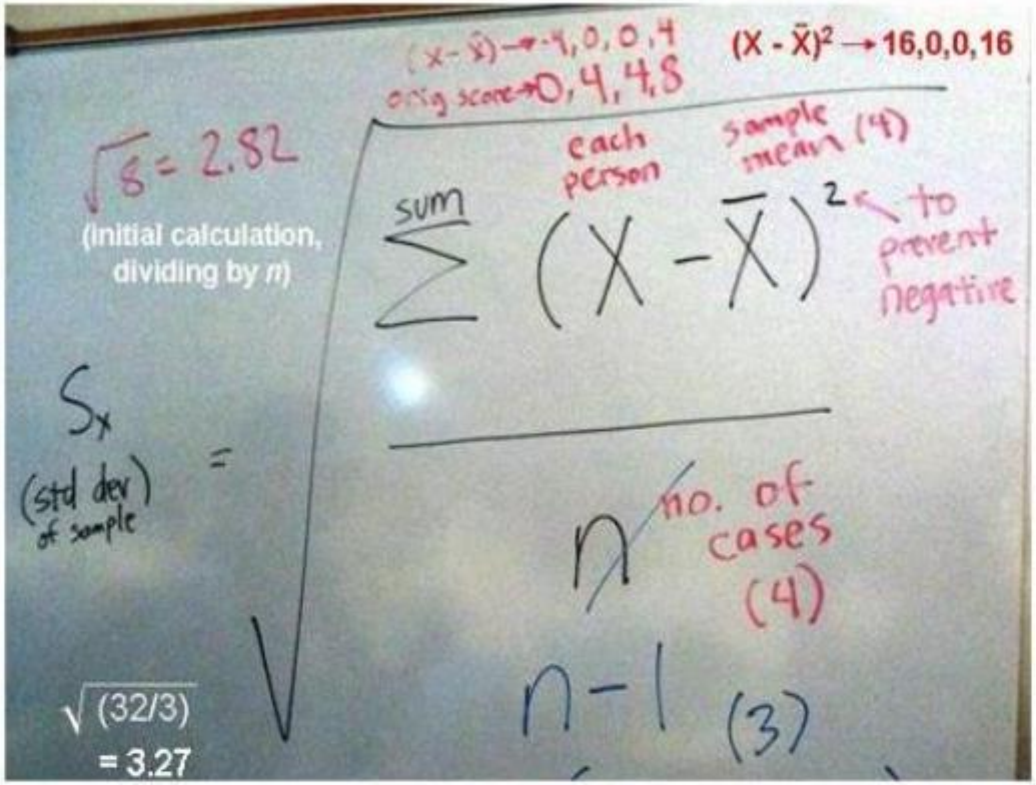
A Measure of "Spread"



Same means, but different amounts of spread

To get an idea of the amount of spread or dispersion in a distribution, it can be helpful to take the ratio of the SD to the mean. For example, an SD of 5 with a mean of 10 (ratio of .5) suggests more spread than an SD of 5 with mean of 1,000 (.005). This ratio is also known as the relative standard deviation.

For graduate students, here's the calculation of the top SD (3.27).
(Thanks to HB for the photo!)



Explanation of why we [divide by "*n* - 1"]() rather than just "*n*."

Concrete example of why the standard deviation can sometimes be important. Ian Ayres writes in his book Super Crunchers, which is about "number-crunching" statistical analyses, as follows:

***When I taught at Stanford Law School, professors were required to award grades that had a 3.2 mean. ...[S]tudents would ask if a professor was a "spreader" [wide range, high SD] or "clumper" [narrow range, low SD]. Good students would want to avoid clumpers so that they would have a better chance at getting an A, while bad students hated the spreaders who handed out more As but also more Fs* (p. 201; segments in *red* inserted by Dr. Reifman).**

Dr. Reifman wonders how this grading requirement is enforced. Perhaps non-compliant professors are made to teach 8:00 a.m. courses, stripped of their parking privileges, or made to dress up as the Stanford tree mascot!

t-test ("t for two")

Compares two means, e.g., experimental vs. control group; men vs. women

t is based on the following (actual formula a bit more involved):

$$\frac{\text{Mean}_{\text{Group 1}} - \text{Mean}_{\text{Group2}}}{\text{Spread (SD's) of the groups' data points}}$$

t is increased (difference more likely to be significant) when:

- **Two groups' means are very different.**
- **Spread (SD's) are small.**

EXAMPLE: Students are randomly assigned to one of two groups (each group with 10 students), an experimental group (E) that receives new math teaching techniques or a control group (C) that receives the "usual" math instruction (independent variable = instruction type). At the end of the term, all students are given a 25-point test (dependent variable). The black rows (below) represent scores on the test, with the red and blue blocks indicating how many students got each score.

Scenario 1

$Mean_C = 12$ $SD_C = 0.82$	$t = 16.36, p < .000001$ (Result SIGNIFICANT: Extremely unlikely to obtain such dramatic difference between E&C means by chance IF the null hypothesis of no mean difference in the population were true; REJECT null hypothesis)	$Mean_E = 18$ $SD_E = 0.82$

						Clear gap between groups								
				C						E				
			C	C	C		↓		E	E	E			
			C	C	C				E	E	E			
			C	C	C				E	E	E			
8	9	10	11	12	13	14	15	16	17	18	19	20	21	22

Mean, standard deviation, and *t* statistics can be obtained from this site.

Scenario 2

$Mean_C = 12$ $SD_C = 2.45$	$t = 5.48, p < .00005$ (also SIGNIFICANT, but not quite as dramatically as above)	$Mean_E = 18$ $SD_E = 2.45$

						Groups Overlap (difference not quite as clear as above)								
			C	C	C		C		E	E	E			
C	C		C	C	C	E	E	C	E	E	E		E	E
8	9	10	11	12	13	14	15	16	17	18	19	20	21	22

The respective means for the two groups are the same in both scenarios, but the standard deviations (spreads) are different, thus making the t-test results different.

Statistical significance

Just because the means of two groups on a DV are different, it doesn't guarantee that the difference is authentic, substantial, or appreciable. The difference could be due to chance. As Westfall and Henning (2013) state, "When a difference is not easy to explain by chance alone, it is called a *statistically significant* difference" (p. 401).

Statistical significance (determined through various statistical tests) tells us if difference is so large, it is extremely unlikely to have resulted from chance. A common standard in the social sciences is whether a finding would come up by chance 5 times out of 100 or less often (p <.05).

Study of "[College Cuteness](),” conducted by two male students at the University of Chicago in the early 2000s, compared the mean attractiveness of female students at different universities. The *t* test is used to compare one school to another. I apologize in advance for the unbalanced nature of the study (only women are evaluated on attractiveness, not men), which many would consider sexist. However, the study presents significance testing in a very down-to-earth manner.

College Cuteness

10 schools. 350 girls. 105,000 votes.

[Home] [What We Did] [Schools] [**Results**] [FAQ] [Contact]

[What We Did]

It is already a fairly well-known fact that students at the University of Chicago are not very good looking. There's a saying that goes like this: “The University of Chicago: Where the squirrels are cuter than the girls.” People constantly complain that it is almost impossible to find someone attractive here, which is funny since they often are not very attractive themselves. Tucker Max, a well known internet writer and University of Chicago alumnus, even wrote an article while he still went to school at UChicago entitled “Everyone Here is Ugly.”

Based on this, we wanted to see if there was any actual evidence showing that our school is in fact less attractive than most. We also figured that other schools would be interested in seeing how they compared when it came to this often-discussed topic.

After some thinking, it occurred to us that we could use popular internet services to collect the data we needed.

First we used thefacebook, an online college directory where virtually every student is registered, to gather pictures of girls from the class of 2008 at 10 different schools. The schools were chosen in order to provide as much diversity for the study as possible: we chose small liberal arts colleges as well as big universities from around the country, with varying levels of academic competitiveness in each case (see the schools section for the list). 35 girls were chosen at each school using a random number generator. The pictures were filed numerically, with no names recorded for anyone involved. If it wasn't clear in a facebook picture who the actual girl was, the picture was discarded and a new random picture was chosen.

Next we registered each picture on HOTorNOT.com; a website where people can rate anonymous pictures on a scale of 1 to 10 based on attractiveness. We waited until each picture had at least 300 votes (which took up to a month in some cases), and then recorded the (mean) average score. We figured that after 300 votes the picture's rating was an accurate score, especially since HOTorNOT themselves says that most picture's scores don't change much after as few as 50 votes.

After that we analyzed the sample data from the 35 girls at each school using a graphing calculator, and put our findings on this website. You can go to our results section to see what we found out.

p (probability of chance finding) < .05 is the common cut-off for "significance"

Rank	School	Average
1st	Albion College	7.60
2nd	Michigan State University	7.55
3rd	University of Oregon	7.52
4th	Vanderbilt University	7.49
5th	University of Michigan	7.11
6th	Princeton University	7.06
7th	University of California, Los Angeles	6.78
8th	Stanford University	6.75
9th	University of Puget Sound	6.35
10th	University of Chicago	6.07

Comparison	P-Value
Albion > UCLA	**0.0324**
Albion > Puget Sound	**0.0014**
Albion > Vanderbilt	0.3950
Albion > Michigan	**0.0979**
Albion > MSU	0.4486
Albion > UChicago	**0.00005**
Albion > Oregon	0.4214
Albion > Princeton	**0.0811**
Albion > Stanford	**0.0155**

Puget Sound < Michigan	**0.0296**
Puget Sound < MSU	**0.0024**
Puget Sound > UChicago	0.2368
Puget Sound < Oregon	**0.0032**

We did tests on the data collected to try to determine how statistically significant our results were. The p-value, or probability value, is a way of measuring that statistical significance. The p-value for each comparison of two schools gives the probability of getting results that are as extreme or more extreme than the ones we got in this study if the two schools are really equal in terms of attractiveness. In other words, **smaller p-values mean the results for that comparison are more significant** because there is a lower probability that we could have gotten results like these if the schools are really equal.

Hypothesis Testing

We learned earlier that a hypothesis is "a relatively specific prediction of how two or more variables should be related;" today we look at scientific hypothesis testing more formally.

Important new concept: Null Hypothesis (H_o): Statement that there is no relationship, no difference, no effect, etc. Treatment will have no effect, experimental and control groups will not differ on DV, etc.

Researcher does not necessarily believe null hypothesis; it's just a standard procedure.

Three Steps of Scientific Hypothesis Testing:

1. **State the null hypothesis (H_o)**
2. **Do the study.**
3. **If the groups significantly differ with $p < .05$, REJECT the null hypothesis.**
   (If the difference is not significant, null hypothesis must be kept alive.)

Analogy to jurors in a criminal trial:

- **Start with "null hypothesis" --- Defendant presumed innocent.**
- **Jurors receive evidence (witnesses, physical evidence).**
- **If evidence overwhelmingly indicative of guilt ("beyond a reasonable doubt"), REJECT null hypothesis of innocence and vote "guilty."**

Chi-Square (χ^2) Test

(Usually done with two nominal/categorical variables)

2 X 2	Democrat	Republican
Men	Obs. (Exp.)	Obs. (Exp.)
Women	Obs. (Exp.)	Obs. (Exp.)

Tests whether men's proportions of Democratic and Republican identification match women's.

2 X 2 most common, but table dimensions can take on other values

3 (Live) X 5 (Major)	Arts & Sciences	Human Sciences	Engineering	Business Admin.	Other
Dorm					
Apt					
Parents					

Observed frequencies are the actual number of people appearing in each cell (e.g., no. of male Democrats). Expected frequencies represent the null hypothesis, H_o (e.g., what <u>frequencies would be expected</u> if there were no male-female differences).

Exp. Freq. = (Total for Row X Total for Column)/ Grand Total

Chi-Square (χ^2) Test
Data from Spring 2022 HDFS 3390 Project

Compares the actual observed number of people in a cell with the number expected under the null hypothesis (H_o: men and women don't differ in vax rate)

Vaccinated?

		Yes*	No	
Male	Observed	29	19	48
	Expect	30.9333	17.0667	
	Chi-Sq	0.1208	0.2190	
Female		87	45	132
		85.0667	46.9333	
		0.0439	0.0796	
		116	64	180

Men 60% Yes/40% No

Women 66% Yes/34% No

DF	Chi Square	P
1	0.4634	0.4960

(Rows -1)(Columns – 1)

Expected Freq for each cell =

$$\frac{\text{(Column Total x Row Total)}}{\text{Grand Total}}$$

Chi-Square for each cell =

$$\frac{(\text{Obs - Exp})^2}{\text{Expected}}$$

Compute Expected Count & Chi-Square for each cell, then add up all chi-squares

* "Yes" includes those who received one dose of a two-dose regimen.

http://srjcstaff.santarosa.edu/~gsturr/statCalc/chiTest.html

Chi-Square

Lyrics by Alan Reifman

(May be sung to the tune of "Hey Jude," Lennon/McCartney)

Chi-square, yes you are there,
So we can test, association,
Of nominal, variables, that you've got,
Give it a shot, and see what happens,

Chi-square, yes you compare,
Each cell's observed counts, with the expected ones,
The null hypothesis, determines your E's,
If you will, please, apply the formula,

Cell differences, each one you square, so that they're all,
Displayed in a, positive direction,
Each cell's squared difference, divide by its expected,
Summing these yields, the overall chi-square,
Nah nah nah, nah nah, nah nah nah nah,

Chi-square, columns and rows,
Give the table's, degrees of freedom,
Remember to use these, when you test the,
Significance, of all your findings,

So is your test, significant, or is it not?
You must consult, the critical value,
So go ahead, with your df, can you reject,
The null that counts, are independent?
Nah nah nah, nah nah, nah nah nah nah,

Chi-square, yes you compare,
Each cell's observed counts, with the expected ones,
The null hypothesis, determines your E's,
If you will, please, have the computer...
Run it, run it, run it, run it, run it, run it, yeah...

O minus E, difference squared, divided by E, for each cell,
Then, add the cell-based chi-squares, into a sum, that will tell...

O minus E, difference squared, divided by E, for each cell,
Then, add the cell-based chi-squares, into a sum, that will tell...

O minus E, difference squared, divided by E, for each cell,
Then, add the cell-based chi-squares, into a sum, that will tell...

O minus E, difference squared, divided by E, for each cell,
Then, add the cell-based chi-squares, into a sum, that will tell...

O minus E, difference squared, divided by E, for each cell,
Then, add the cell-based chi-squares, into a sum, that will tell...

O minus E, difference squared, divided by E, for each cell,
Then, add the cell-based chi-squares, into a sum, that will tell...

Websites Useful for Statistics

Survey Documentation and Analysis (SDA) project (University of California-Berkeley)
Nice website for conducting statistical analyses on real data (surveys in the public domain)

Statistical Thinking (University of Baltimore)
Has a "ton" of information on virtually any statistical topic that you'd encounter

Stat Pages -- Compendium of virtually any kind of statistical calculator one would need

Brief checklist for assessing one's proficiency in basic statistics

CAUSEweb: Consortium for the Advancement of Undergraduate Statistics Education
collection of fun resources (songs, jokes, etc.).