Lecture 3

- Inference for a Single Proportion:

✓test of significance for a single proportion
✓Chi-square test of goodness-of –fit
✓Ch-square test for independence

Recall: Population Proportion

- Let p be the proportion of "successes" in a population. A random sample of size n is selected, and X *is the count of successes* in the sample.
- Suppose n is small relative to the population size, so that X can be regarded as a binomial random variable with

$$\mu = np \qquad \text{and} \qquad \sigma = \sqrt{np(1-p)}$$

Recall: Population Proportion

- We use the *sample proportion* $\hat{p} = X/n$ as an estimator of the population proportion p.
- $\hat{p}$ is an unbiased estimator of p, *with* mean and SD:

$$p \qquad \text{and} \qquad \sqrt{\frac{p(1-p)}{n}}$$

- When n is large, $\hat{p}$ is approximately normal. Thus

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

is approximately standard normal.

Classical Confidence Interval for a Population Proportion

- The *standard error* of $\hat{p}$ is

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- An *approximate* level C confidence interval for p:

$$\hat{p} \pm z^* SE(\hat{p}) = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $P(Z \geq z^*) = (1 - C)/2$.

Example:

A news program constructs a call-in poll about a proposed city ban on handguns. 2372 people call in to the show. Of these, 1921 oppose the ban.

Construct a 95% confidence interval for the population proportion of people who oppose the ban.

What are the possible problems with the study design?

Solution:

- **Note:** Since p is a proportion, if you ever get an upper value of > 1 or lower <0, replace by 1 and 0 (respectively).

SAS

```
data fraction;
input ban $ count;
cards;
yes  451
no   1921
;
run;
proc freq order=freq;
weight count;
tables ban/ binomial alpha=0.01;
run;
```

```
The FREQ Procedure

                              Cumulative    Cumulative
ban    Frequency     Percent     Frequency      Percent

no        1921        80.99        1921        80.99
yes        451        19.01        2372       100.00


Binomial Proportion for ban = no

Proportion              0.8099
ASE                     0.0081
99% Lower Conf Limit    0.7891
99% Upper Conf Limit    0.8306

Exact Conf Limits
99% Lower Conf Limit    0.7883
99% Upper Conf Limit    0.8302
```

Testing for a single population proportion

- When **n** is **large**, $\hat{p}$ is approximately normal, so

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

is approximately standard normal.

We may test H_0: $p = p_0$ against one of these:

- H_a: $p > p_0$
- H_a: $p < p_0$
- H_a: $p \neq p_0$

Large-sample Significance Test for a Population Proportion

- The null hypothesis – H_0: $p = p_0$
- The test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Alternative Hypothesis	P-value
H_a: $p > p_0$	$P(Z \geq z)$
H_a: $p < p_0$	$P(Z \leq z)$
H_a: $p \neq p_0$	$2P(Z \geq \mid z \mid)$

Large-sample Significance Test for a Population Proportion

- How big does the sample size need to be?
- The general rule of thumb to use here, as before for approximation of binomial distribution by normal distribution, is

$$np_0 \geq 10, \quad n(1-p_0) \geq 10$$

Example:

- A claim is made that only 34% of all college students have part-time jobs. You are a little skeptical of this result and decide to conduct an experiment to show that more students work. You get a sample of 100 college students and find that 47 of these students have part-time jobs.
- Conduct a hypothesis test with $\alpha = 0.05$ to determine whether more than 34% of college students have part-time jobs.

Solution

SAS

```
data work;
input work $ count;
cards;
yes  47
no   53
;
run;
proc freq;
weight count;
tables work/ binomial (p=0.34 level='yes');
run;
```

```
Binomial Proportion
for work = yes

Proportion                 0.4700
ASE                        0.0499
95% Lower Conf Limit       0.3722
95% Upper Conf Limit       0.5678

Exact Conf Limits
95% Lower Conf Limit       0.3694
95% Upper Conf Limit       0.5724

Test of H0: Proportion = 0.34

ASE under H0           0.0474
Z                      2.7443
One-sided Pr >  Z      0.0030
Two-sided Pr > |Z|     0.0061
```

- Does proportion of people with higher education (Master or above) in American population exceeds 10 % ?
- We will use the data set individuals.dat

SAS

```
data idividuals;
infile 'c:/users/mbogdan/ECMI/data/individuals.dat';
input id age edu gen income class;
proc freq;
tables edu/ binomial  (p=0.10 level=6);
run;
```

```
Binomial Proportion for edu = 6

Proportion             0.1002
ASE                    0.0013
95% Lower Conf Limit      0.0977
95% Upper Conf Limit      0.1027

Exact Conf Limits
95% Lower Conf Limit      0.0977
95% Upper Conf Limit      0.1027

Test of H0: Proportion = 0.1

ASE under H0           0.0013
Z                      0.1565
One-sided Pr >  Z      0.4378
Two-sided Pr > |Z|     0.8756
```

Chi-square test for goodness of fit

- categorical data; a random sample of size n
- have hypothesised values for the population proportions π for each category;
- these are specified in or implied by the problem
- an approximate test which works when sample size is large

Simplest case: two categories

- Example:
- There are two homozygous lines of Drosophila, one with red eyes, and one with purple eyes. It has been suggested that there is a single gene responsible for this phenotype, with the red eye trait dominant over the purple eye trait. If that is true we expect a cross of these two lines to produce F2 progeny in the ratio 3 red : 1 purple. We want to test the hypothesis that red is (autosomal) dominant. To do this we perform the cross of red-eyed and purple-eyed flies with several parents from the two lines and obtain **43** flies in the F2 generation, with **29 red-eyed** flies and **14 purple-eyed** flies.

Categories:

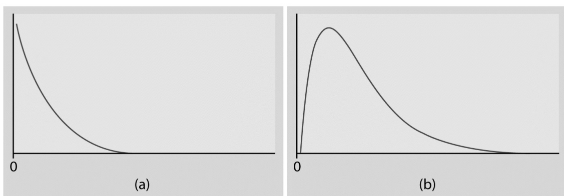
- Red eyes; hypothesised proportion π = 3/(3+1) = 0.75
- "expected" number: E1 = (43)(0.75) = 32.25
- Purple eyes; hypothesised proportion 1 – π = 1/(3+1) = 0.25
- "expected" number: E2 = (43)(0.25) = 10.75
- Is the red-eye trait dominant over purple?

- Let π be the probability that an F2 fly has red eyes
- H0: $\pi = 0.75$; the F2 progeny are in a 3:1 ratio of red to purple-eyed flies
- HA: $\pi \neq 0.75$; the F2 progeny are not in a 3:1 ratio

Chi-square goodness of fit test

- $X^2 = \Sigma$(observed - expected)^2 / expected = $\Sigma(O-E)^2/E$
- Under HO X^2 has a chi-square distribution with df = #categories - 1 = 1.
- Test at level $\alpha = 0.05$; Critical value = 3.84

Chi-square distributions with df=2 and 4:



- **P-value** for chi-square test is: $P(\chi^2 \geq X^2)$
- This is always the right tail of the distribution.

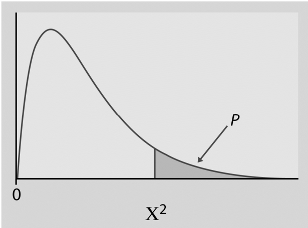


TABLE F χ^2 distribution critical values

	Tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

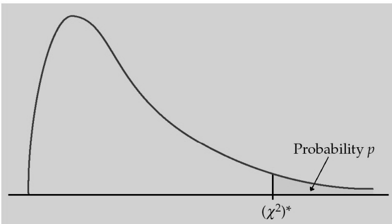


Table entry for p is the critical value $(\chi^2)^*$ with probability p lying to its right.

Solution

SAS

- **data** flies;
- input eyes $ count;
- cards;
- purple 14
- red 29
- ;
- **run**;
- **proc freq**;
- weight count;
- tables eyes/ chisq testp=(**0.25 0.75**);
- **run**;

eyes	Frequency	Percent	Test Percent	Cumulative Frequency	Cumulative Percent
purple	14	32.56	25.00	14	32.56
red	29	67.44	75.00	43	100.00

- Chi-Square Test
- for Specified Proportions
- Chi-Square 1.3101
- DF 1
- Pr > ChiSq 0.2524

```
proc freq;
weight count;
tables eyes/ binomial  (p=0.25);
run;
```

- Test of H0: Proportion = 0.25

ASE under H0	0.0660
Z	1.1446
One-sided Pr > Z	0.1262
Two-sided Pr > |Z|	0.2524

More than 2 Categories

- Example:
- In the sweet pea, the allele for purple flower colour (P) is dominant to the allele for red flowers (p), and the allele for long pollen grains (L) is dominant to the allele for round pollen grains (l). We have P1 parents homozygous for the dominant alleles (PPLL) and P2 parents homozygous for the recessive alleles (ppll). The F1 generation are all PpLl and have purple flowers and long pollen grains. The F1's are crossed to give an F2 generation. It is thought that the genes controlling these two traits are 20 cM apart. If that were true then the F2 offspring proportions should follow the ratio **66:9:9:16**

- **66%** purple/long : PPLL or PpLL or PPLl or PpLl,
- **9%** purple/round : PPll or Ppll,
- **9%** red/long = ppLL or ppLl,
- **16%** red/round = ppll
- 381 F2 offspring are collected, and we observe
- **284 purple/long**
- **21 purple/round**
- **21 red/long**
- **55 red/round**
- **Are these genes 20 cM apart?**

- Let $\pi1$, $\pi2$, $\pi3$, $\pi4$ be the probabilities of purple/long, purple/round, red/long, red/round offspring, respectively, resulting from this F2.
- H0: $\pi1=0.66$, $\pi2 = 0.09$, $\pi3=0.09$, $\pi4=0.16$; the category probabilities are those predicted by a 20cM genetic distance
- HA: the category probabilities are different from those predicted by a 20cM genetic distance
- Use a chi-square goodness-of-fit test with
- df = #categories - 1 = 4 - 1 = 3
- $X^2 = \Sigma(O-E)^2/E$ has a χ^2_3 distribution under H0.

Solution

- Test at level $\alpha = 0.05$; critical value for χ^2_3 is 7.81. Will reject H0 if $X^2 > 7.81$

```
data peas;
input  colour $ shape $ count;
cards;
purple long 284
purple round 21
red long 21
red round 55
;
run;

data peas; set peas;
if ((colour eq 'purple')*(shape eq 'long')) then cs='pl';
if ((colour eq 'purple')*(shape eq 'round')) then cs='pr';
if ((colour eq 'red')*(shape eq 'long')) then cs='rl';
if ((colour eq 'red')*(shape eq 'round'))  then cs='rr';
run;

proc freq data=peas;
weight count;
tables cs/ chisq testp=(0.66 0.09 0.09 0.16);
run;
```

- The FREQ Procedure

cs	Frequency	Percent	Test Percent	Cumulative Frequency	Cumulative Percent
pl	284	74.54	66.00	284	74.54
pr	21	5.51	9.00	305	80.05
rl	21	5.51	9.00	326	85.56
rr	55	14.44	16.00	381	100.00

- Chi-Square Test
- for Specified Proportions

Chi-Square	15.0953
DF	3
Pr > ChiSq	0.0017

Test of independence Example:

- Do men and women participate in sport for the same reasons?
- 67 males and 67 females examined. Results:

- HSC-HM female 14
- HSC-HM male 31
- HSC-LM female 7
- HSC-LM male 18
- LSC-HM female 21
- LSC-HM male 5
- LSC-LM female 25
- LSC-LM male 13

Legend: HSC (LSC)-high (low) **social comparison**;
HM (LM)-high (low) **mastery**

Table of goal by sex

goal / sex; Frequency; Percent

goal	female	male	Total
HSC-HM	14 10.45	31	
HSC-LM	7	18	
LSC-HM	21	5	
LSC-LM	25	13	
Total			134

Complete “Percent”—i.e. give the JOINT distribution of “goal” and “sex”.
“goal”-**column variable** (often *response*), “sex”-**row variable** (often *explanatory*)

goal / **sex**; Frequency; Expected; Percent; Row Pct; Col Pct

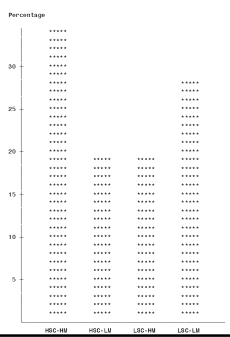
goal	female	male	Total
HSC-HM	14 10.45	31 23.13	45 33.58
HSC-LM	7 5.22	18 13.43	
LSC-HM	21 15.67	5 3.73	
LSC-LM	25 18.66	13 9.70	
Total			134 100.00

Find the MARGINAL distribution of **goal.**
Find the MARGINAL distribution of **sex**.

MARGINAL distribution of **goal** in this study

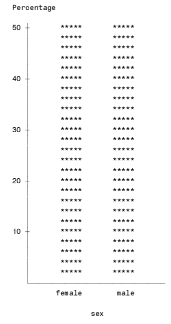
The FREQ Procedure

goal	Frequency	Percent	Cumulative Frequency	Cumulative Percent
HSC-HM	45	33.58	45	33.58
HSC-LM	25	18.66	70	52.24
LSC-HM	26	19.40	96	71.64
LSC-LM	38	28.36	134	100.00



MARGINAL distribution of **sex** in this study.

sex	Frequency	Percent	Cumulative Frequency	Cumulative Percent
female	67	50.00	67	50.00
male	67	50.00	134	100.00

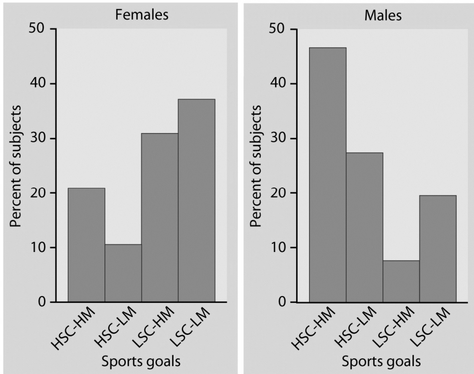


goal sex

Frequency Percent Col Pct	female	male	Total
HSC-HM	14 10.45 20.90	31 23.13	45 33.58
HSC-LM	7 5.22	18 13.43	25 18.66
LSC-HM	21 15.67	5 3.73	26 19.40
LSC-LM	25 18.66	13 9.70	38 28.36
Total	67 50.00	67 50.00	134 100.00

Find the distribution of goals among females—i.e. **conditional distribution** for "sex"=female.
Do the same for men.
Question: what percent of women have LSC-HM attitude?

Conditional distributions for males and females.



The final result: goal sex

Frequency Percent Row Pct Col Pct	female	male	Total
HSC-HM	14 10.45 31.11 20.90	31 23.13 68.89 46.27	45 33.58
HSC-LM	7 5.22 28.00 10.45	18 13.43 72.00 26.87	25 18.66
LSC-HM	21 15.67 80.77 31.34	5 3.73 19.23 7.46	26 19.40
LSC-LM	25 18.66 65.79 37.31	13 9.70 34.21 19.40	38 28.36
Total	67 50.00	67 50.00	134 100.00

TWO-WAY table with marginal and conditional distributions.

proc freq
see SAS file: 9-1.sas

```
proc freq data=sport;
  tables goal*sex;
  weight count;
run;
```

Simpsons's paradox:

- An association or comparison that holds for all of several groups can reverse direction when the data are combined to form a single group.
- This can be due to a lurking variable.

Example :

- Here are the numbers of flights on time and delayed for 2 airlines at 5 airports. Overall on-time %s for each airline are often reported in the news. Lurking variables can make such reports misleading.

	Alaska Airlines			America West		
	On time	**Delayed**	**Total**	**On Time**	**Delayed**	**Total**
L.A.	497	62	559	694	117	811
Phoenix	221	12	233	4840	415	5255
San Diego	212	20	232	383	65	448
San Francisco	503	102	605	320	129	449
Seattle	1841	305	2146	201	61	262
Total		501	3775		787	7225

a) Find the % of delayed flights for Alaska Airlines at each of the 5 airports, and then do the same for America West. (Note: these are not joint probabilities.)

	Alaska Airlines	**America West**
L.A.		
Phoenix		
San Diego		
San Francisco		
Seattle		

b) What % of all Alaska Airlines flights were delayed? What % of all America West flights were delayed? These are the numbers usually reported.

c) America West does worse at every one of the 5 airports, yet does better overall. That sounds impossible. Explain carefully, referring to the data, how this can happen.

Perils of aggregation

- This example was essentially a Three-Way Table with variables: airline, timing, airport.
- Such tables are often reported as several two-way tables. Think a book, rather than a page.
- Adding entries from such elementary tables ("pages") to get the overall summary (for the "book") is aggregation and leads to ignoring the third variable (here: airport).
- This may lead to false general conclusions.

Inference for Two-Way tables

Hypothesis testing with 2-way tables

- $\mathbf{H_0}$: there is no association between the row and column variables (they are independent)
- $\mathbf{H_a}$: there is an association between the row and column variables

- To test the hypotheses, compare observed cell counts with **expected** cell counts.
- **Expected**=calculated under the assumption that the null hypothesis is true.

expected count = $\dfrac{\text{row total} \times \text{column total}}{\text{n}}$

Here n = total # of observations for the table.

```
goal      sex

Frequency|
Expected |
Percent  |
Row Pct  |
Col Pct  |female  |male    |  Total
---------+--------+--------+
HSC-HM   |     14 |     31 |     45
         |   22.5 |        |
         |  10.45 |  23.13 |  33.58
         |  31.11 |  68.89 |
         |  20.90 |  46.27 |
---------+--------+--------+
HSC-LM   |      7 |     18 |     25
         |        |        |
         |   5.22 |  13.43 |  18.66
         |  28.00 |  72.00 |
         |  10.45 |  26.87 |
---------+--------+--------+
LSC-HM   |     21 |      5 |     26
         |        |        |
         |  15.67 |   3.73 |  19.40
         |  80.77 |  19.23 |
         |  31.34 |   7.46 |
---------+--------+--------+
LSC-LM   |     25 |     13 |     38
         |        |        |
         |  18.66 |   9.70 |  28.36
         |  65.79 |  34.21 |
         |  37.31 |  19.40 |
---------+--------+--------+
Total          67       67      134
            50.00    50.00   100.00
```

Calculate EXPECTED counts.

proc freq
see SAS file: <u>9-1.sas</u>

```
proc freq data=sport;
  tables goal*sex / expected ;
  weight count;
run;
```

Test statistic: Chi Square Test Statistic

$$X^2 = \sum \frac{(\text{observed count - expected count})^2}{\text{expected count}}$$

χ^2 distribution

- The X^2 test statistic has an approximately chi-square distribution.
- To use the chi-square table, you need the degrees of freedom:

  (r-1)(c-1)=(#rows-1)(#colums-1).

- Our example has (4-1)(2-1)=3 df.

Finale: Do men and women participate in sport for the same reasons?

Frequency Expected Percent Row Pct Col Pct	female	male	Total
HSC-HM	14	31	45
	22.5	22.5	
	10.45	23.13	33.58
	31.11	68.89	
	20.90	46.27	
HSC-LM	7	18	25
	12.5	12.5	
	5.22	13.43	18.66
	28.00	72.00	
	10.45	26.87	
LSC-HM	21	5	26
	13	13	
	15.67	3.73	19.40
	80.77	19.23	
	31.34	7.46	
LSC-LM	25	13	38
	19	19	
	18.66	9.70	28.36
	65.79	34.21	
	37.31	19.40	
Total	67	67	134
	50.00	50.00	100.00

Solution:

- Recall: $X^2 = \sum \frac{(\text{observed count - expected count})^2}{\text{expected count}}$

proc freq
see SAS file: 9-1.sas

```
proc freq data=sport;
  tables goal*sex / expected chisq;
  weight count;
run;
```

The FREQ Procedure (output):

Statistics for Table of goal by sex

Statistic	DF	Value	Prob
Chi-Square	3	24.8978	<.0001
Likelihood Ratio Chi-Square	3	26.0362	<.0001
Mantel-Haenszel Chi-Square	1	16.2249	<.0001
Phi Coefficient		0.4311	
Contingency Coefficient		0.3958	
Cramer's V		0.4311	

Sample Size = 134

Rules for using the test:

- The chi-square test becomes more accurate as the cell counts increase and for tables larger than 2x2.
- For tables larger than 2x2: use the chi-square test whenever:
- ✓ the average of the expected counts is 5 or more
- ✓ the smallest expected count is 1 or more
- ✓ <20% of cells have expected counts of less than 5.
- For 2x2 tables: use chi-square test

whenever all 4 expected cell counts are 5 or more

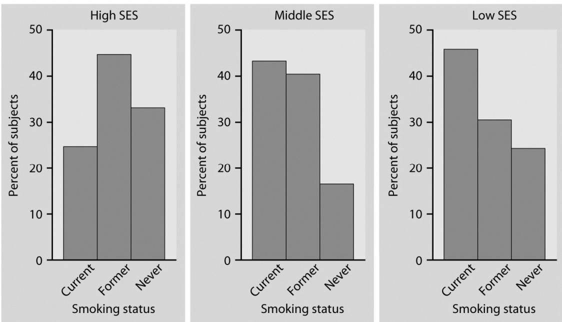
If the asymptotic assumptions are not met use the exact chisq statement in proc freq.

Example:

- 356 volunteers classified according socioeconimic status (SES) and smoking habits.
- Is smoking associated with SES?

smoking SES

Frequency Percent Row Pct Col Pct	high	low	middle	Total
current	**51** 14.33 43.97 24.17	**43** 12.08 37.07 46.24	**22** 6.18 18.97 42.31	116 32.58
former	**92** 25.84 65.25 43.60	**28** 7.87 19.86 30.11	**21** 5.90 14.89 40.38	141 39.61
never	**68** 19.10 68.69 32.23	**22** 6.18 22.22 23.66	**9** 2.53 9.09 17.31	99 27.81
Total	211 59.27	93 26.12	52 14.61	356 100.00



Smoking is associated to SES:

smoking SES

Frequency Expected Percent Row Pct Col Pct	high	low	middle	Total
current	51 68.753 14.33 43.97 24.17	43 30.303 12.08 37.07 46.24	22 16.944 6.18 18.97 42.31	116 32.58
former	92 83.57 25.84 65.25 43.60	28 36.834 7.87 19.86 30.11	21 20.596 5.90 14.89 40.38	141 39.61
never	68 58.677 19.10 68.69 32.23	22 25.862 6.18 22.22 23.66	9 14.461 2.53 9.09 17.31	99 27.81
Total	211 59.27	93 26.12	52 14.61	356 100.00

Statistics for Table of smoking by SES

Statistic	DF	Value	Prob
Chi-Square	4	18.5097	0.0010
Likelihood Ratio Chi-Square	4	18.6635	0.0009
Mantel-Haenszel Chi-Square	1	12.2003	0.0005
Phi Coefficient		0.2280	
Contingency Coefficient		0.2223	
Cramer's V		0.1612	

Sample Size = 356

Example (Aspirin study):

- 21,996 male American physicians.
- Half of these took aspirin.
- After 3 years, 139 of those who took aspirin and 239 of those who took placebo had had heart attacks.
- Determine whether there is an association of aspirin with heart attacks.

fate treatment

Frequency Expected Percent Row Pct Col Pct	aspirin	placebo	Total
heart_at	**139** 189 0.63 36.77 1.26	**239** 189 1.09 63.23 2.17	378 1.72
no_heart	**10859** 10809 49.37 50.23 98.74	**10759** 10809 48.91 49.77 97.83	21618 98.28
Total	10998 50.00	10998 50.00	21996 100.00

Statistics for Table of fate by treatment

Statistic	DF	Value	Prob
Chi-Square	**1**	**26.9176**	**<.0001**
Likelihood Ratio Chi-Square	1	27.2352	<.0001
Continuity Adj. Chi-Square	1	26.3819	<.0001
Mantel-Haenszel Chi-Square	1	26.9164	<.0001
Phi Coefficient		-0.0350	
Contingency Coefficient		0.0350	
Cramer's V		-0.0350	

Fisher's Exact Test

Cell (1,1) Frequency (F)	139
Left-sided Pr <= F	**1.203E-07**
Right-sided Pr >= F	1.0000
Table Probability (P)	5.228E-08
Two-sided Pr <= P	2.407E-07

Sample Size = 21996

Conclusion: Aspirin reduces chance of heart attack (P<.0001).