Chapter Sixteen

Statistical Testing Of Differences And Relationships

LEARNING OBJECTIVES
1. Learn how to evaluate differences and changes.
2. Understand the concept of hypothesis development and how to test hypotheses.
3. Be familiar with several of the more common statistical tests of goodness of fit, hypotheses about one mean, hypotheses about two means, and hypotheses about proportions.
4. Learn the hypotheses about one mean.
5. Learn the hypotheses about two means.
6. Learn the hypotheses about proportions.
7. Learn about analysis of variance.
8. Understand the P values and significance testing.

Statistical Significance

- Exam Score

  Female=120

  Male=110

  Female higher than male?

Statistical Significance

Mathematical Differences:

By definition, if numbers are not exactly the same, they are different. This does not, however, mean that the difference is either important or statistically significant.

Statistical Significance:

A difference that is large enough that it is not likely to have occurred because of **chance or sampling error**.

Statistical Significance

Important Differences for Managers:

One must be able to distinguish between mathematical differences and statistically significant differences in using the data analysis in managerial decision making.

Hypothesis:

Assumption or theory that a researcher or manager makes about some characteristic of the population under study.

Goodness of Fit

Chi-Square Test:

Test of the goodness of fit between the OBSERVED distribution and the EXPECTED distribution of a variable.

Marketing researchers often need to determine whether there is any association between two or more variables.

Types of Hypothesis Tests

About One Mean:

Z Test:

Hypothesis test used for a single mean if the sample is large enough and drawn at random. Usually for samples of about 30 and above.

t Test:

Hypothesis test used for a single mean if the sample is too small to use the Z test. Usually for samples below 30.

About Two Means:

Hypothesis testing that tests the differences *between* groups of data.

Z Test - Example

Z Test

One of the most common goals of marketing research studies is to make some inference about the population mean. If the sample size is large enough ($n \geq 30$), the appropriate test statistic for testing a hypothesis about a single mean is the Z **test**. For small samples ($n <$ 30) the t test with $n - 1$ degrees of freedom (where n = sample size) should be used.

Video Connection, a Dallas video store chain, recently completed a survey of 200 consumers in its market area. One of the questions was "Compared to other video stores in the area, would you say Video Connection is much better than average, somewhat better than average, average, somewhat worse than average, or much worse than average?" Responses were coded as follows:

Response	Code
Much better	5
Somewhat better	4
Average	3
Somewhat worse	2
Much worse	1

The mean rating of Video Connection is 3.4. The sample standard deviation is 1.9. How can the management of Video Connection be confident that its video stores' mean rating is significantly higher than 3 (average in the rating scale)? The Z test for hypotheses about one mean is the appropriate test in this situation. The steps in the procedure follow.

Key Terms & Definitions

Z Test – Example

Continued

1. Specify the null and alternative hypotheses.
   - Null hypothesis H_0: $M \leq 3$ (M = response on rating scale)
   - Alternative hypothesis H_a: $M > 3$
2. Specify the level of sampling error (α) allowed. For $\alpha = .05$ the table value of Z(critical) $= 1.64$. (See Exhibit 3 in Appendix 2 for d.f. $= \infty$, .05 significance, one-tail. The table for t is used because $t = Z$ for samples greater than 30.) Management's need to be very confident that the mean rating is significantly higher than 3 is interpreted to mean that the chance of being wrong because of sampling error should be no more than .05 (an $\alpha = .05$).
3. Determine the sample standard deviation (S), which is given as $S = 1.90$.
4. Calculate the estimated standard error of the mean, using the formula

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

where $S_{\bar{X}}$ = estimated standard error of the mean

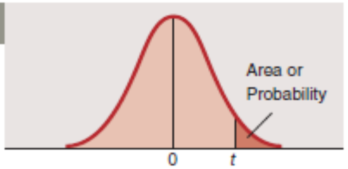
In this case,

$$S_{\bar{X}} = \frac{1.9}{\sqrt{200}} = 0.13$$

Key Terms & Definitions

EXHIBIT 3 *t*-Distribution

Entries in the table give *t*-values for an area or probability in the upper tail of the *t*-distribution. For example, with 10 degrees of freedom and a .05 area in the upper tail, $t_{.05} = 1.812$.



	Area in Upper Tail				
Degrees of Freedom	**.10**	**.05**	**.025**	**.01**	**.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Z Test – Example

Continued

5. Calculate the test statistic:

$$Z = \frac{(\text{Sample mean}) - \begin{pmatrix}\text{Population mean specified} \\ \text{under the null hypothesis}\end{pmatrix}}{\text{Estimated standard error of the mean}}$$

$$= \frac{3.4 - 3}{0.13} = 3.07$$

6. State the result. *The null hypothesis can be rejected* because the calculated Z value (3.07) is larger than the critical Z value (1.64). Management of Video Connection can infer with 95 percent confidence that its video stores' mean rating is significantly higher than 3. Further discussion of Z test application is provided in the feature above.

Hypothesis About Two Means

Continued

- Marketers are frequently interested in testing differences between groups. In the following example of testing the differences between two means, the samples are independent.
- The management of a convenience store chain is interested in **differences between the store visit rates of men and women**.

➢ Mean for Male=11.49

➢ Mean for Female=8.51

- Believing that men visit convenience stores more often than women, management collected data on convenience store visits from 1,000 randomly selected consumers. Testing this hypothesis involves the following steps:

Hypothesis About Two Means

Continued

EXHIBIT 16.4 Data for 32 Test of Two Independent Samples

Visits to Convenience Store by Males				Visits to Convenience Stores by Females			
Number X_m	Frequency f_m	Percent	Cumulative Percent	Number X_f	Frequency f_f	Percent	Cumulative Percent
2	2	4.4	4.4	2	5	7.0	7.0
3	5	11.1	15.6	3	4	5.6	12.7
5	7	15.6	31.1	4	7	9.9	22.5
6	2	4.4	35.6	5	10	14.1	36.6
7	1	2.2	37.8	6	6	8.5	45.1
8	2	4.4	42.2	7	3	4.2	49.3
9	1	2.2	44.4	8	6	8.5	57.7
10	7	15.6	60.0	9	2	2.8	60.6
12	3	6.7	66.7	10	13	18.3	78.9
15	5	11.1	77.8	12	4	5.6	84.5
20	6	13.3	91.1	15	3	4.2	88.7
23	1	2.2	93.3	16	2	2.8	91.5
25	1	2.2	95.6	20	4	5.6	97.2
30	1	2.2	97.8	21	1	1.4	98.6
40	1	2.2	100.0	25	1	1.4	100.0

$n_m = 45$ $\qquad$ $n_f = 71$

Mean number of visits by males, $\overline{X}_m = \dfrac{\sum X_m f_m}{45} = 11.5$

Mean number of visits by females, $\overline{X}_f = \dfrac{\sum X_f f_f}{71} = 8.5$

Hypothesis About Two Means - Example

1. Specify the null and alternative hypotheses.
   - Null hypothesis H_0: $M_m - M_f \leq 0$; the mean visit rate of men (M_m) is the same as or less than the mean visit rate of women (M_f).
   - Alternative hypothesis H_a: $M_m - M_f > 0$; the mean visit rate of men (M_m) is higher than the mean visit rate of women (M_f).

   The observed difference in the two means (Exhibit 16.4) is 11.49 – 8.51 = 2.98.
2. Set the level of sampling error (α). The managers decided that the acceptable level of sampling error for this test is $\alpha = .05$. For $\alpha = .05$ the table value of Z(critical) = 1.64. (See Exhibit 3 in Appendix 3 for d.f. = ∞, .05 significance, one-tail. The table for t is used because $t = Z$ for samples greater than 30.)
3. Calculate the estimated standard error of the differences between the two means as follows:

$$S_{X_{m-f}} = \sqrt{\frac{S_m^2}{n_m} + \frac{S_f^2}{n_f}}$$

where

S_m = estimated standard deviation of population m (men)

S_f = estimated standard deviation of population f (women)

n_m = sample size for sample m

n_f = sample size for sample f

Hypothesis About Two Means

Continued

Therefore,

$$S_{\bar{X}_{m-f}} = \sqrt{\frac{(8.16)^2}{45} + \frac{(5.23)^2}{71}} = 1.37$$

Note that this formula is for those cases in which the two samples have unequal variances. A separate formula is used when the two samples have equal variances. When this test is run in SAS and many other statistical packages, two *t* values are provided—one for each variance assumption.

4. Calculate the test statistic *Z* as follows:

$$Z = \frac{\begin{pmatrix}\text{Difference between means}\\ \text{of first and second sample}\end{pmatrix} - \begin{pmatrix}\text{Difference between means}\\ \text{under the null hypothesis}\end{pmatrix}}{\text{Standard error of the differences between the two means}}$$

$$= \frac{(11.49 - 8.51) - 0}{1.37} = 2.18$$

5. State the result. The calculated value of *Z* (2.18) is larger than the critical value (1.64), so *the null hypothesis is rejected*. Management can conclude with 95 percent confidence ($1 - \alpha = .95$) that, on average, men visit convenience stores more often than do women.

Hypothesis About Two Means

Continued

A child psychologist observed 8-year-old children behind a one-way mirror to determine how long they would play with a toy medical kit. The company that designed the toy was attempting to determine whether to give the kit a masculine or feminine orientation. The length of time (in minutes) the children played with the kits are shown below. Calculate the value of *Z* and recommend to management whether the kit should have a male or female orientation.

Sex	Time played with toy medical kit												
Boys	31	12	41	34	63	7	67	67	25	73	36	41	15
Girls	26	38	20	32	16	45	9	9	16	26	81	20	5

Hypothesis About Two Means
Continued

STEP 1: Hypotheses

Ho: The difference in the mean length of time is equal to zero.

Ha: The difference in the mean length of time is non-zero.

Hypothesis About Two Means
Continued

STEP 2: Decide on Significance Level and Look up Appropriate Table Value

Assume an alpha of 0.05.

Since the sample size is less than 30, it is appropriate to use a t-test.

This is also a two tailed test, hence, the value from the t table for n=12 and alpha = 0.05 is 2.179

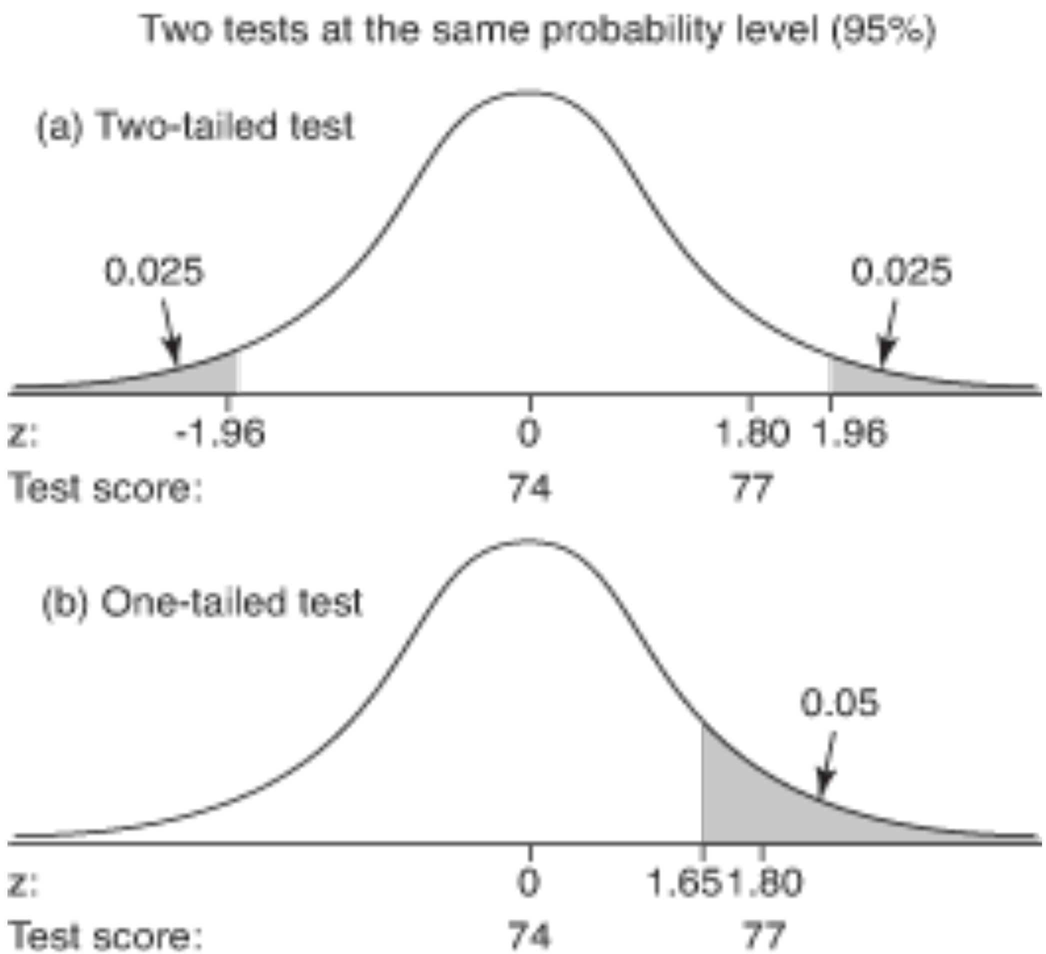
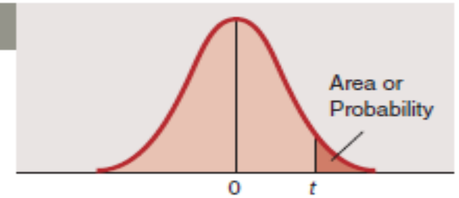

Two tests at the same probability level (95%)
(a) Two-tailed test
0.025
0.025
z:
-1.96
0
1.80
1.96
Test score:
74
77
(b) One-tailed test
0.05
z:
0
1.65
1.80
Test score:
74
77

EXHIBIT 3 *t*-Distribution

Entries in the table give t-values for an area or probability in the upper tail of the t-distribution. For example, with 10 degrees of freedom and a .05 area in the upper tail, $t_{.05} = 1.812$.



	Area in Upper Tail				
Degrees of Freedom	**.10**	**.05**	**.025**	**.01**	**.005**
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Hypothesis About Two Means

Continued

STEP 3: Calculate the Pooled Standard Error for the two samples

The mean of the boys is 39.38. The mean of the girls is 26.83. The variance for the boys was 494.09 and the variance for the girls was 404.59. Therefore, the pooled standard error is 8.31 (square root of 494.09/13 + 404.59/13).

Sex	Time played with toy medical kit												
Boys	31	12	41	34	63	7	67	67	25	73	36	41	15
Girls	26	38	20	32	16	45	9	9	16	26	81	20	5
Observ.	1	2	3	4	5	6	7	8	9	10	11	12	13

Avg.	S	S^2	S^2/n
39.38	22.23	494.09	38.01
26.38	20.11	404.59	31.12

$$S_{x_{b-g}} = \sqrt{\frac{S_b^2}{n_b} + \frac{S_g^2}{n_g}} = \sqrt{\frac{494.09}{13} + \frac{404.59}{13}} = 8.31$$

Hypothesis About Two Means
Continued

STEP 4: Calculate the test statistic

The observed difference in the means is 13. Therefore the t-calculated is equal to 13/8.31 or 1.56

STEP 5: Compare Results and State Conclusion

Since the t calculated value of 1.56 is less than the t-table value of 2.179, we fail to reject the null hypothesis. No significant difference in the attention span of boys and girls was found. No recommendation can be made to management, except to perhaps take a larger sample.

Analysis of Variance (ANOVA)

ANOVA:

Test for the differences among the means of two or more independent samples.

When the goal is to test the differences among the means of two or more independent samples, analysis of variance (ANOVA) is an appropriate statistical tool. Although it can be used to test differences between two means, ANOVA is more commonly used for hypothesis tests regarding the differences among the means of several (C) independent groups (where C > or = 3). It is a statistical technique that permits the researcher to determine whether the variability among or across the C sample means is greater than expected because of sampling error.

F Test

F Test:

Test of the probability that a particular calculated value could have been due to chance.

$$MSA = \frac{\text{Sum of squares among groups (SSA)}}{\text{Degrees of freedom (d.f.)}}$$

where Degrees of freedom = number of groups (C) − 1

$$F = \frac{MSA}{MSE} = \frac{860}{232.14} = 3.70$$

$$MSE = \frac{\text{Sum of squares within groups (SSE)}}{\text{Degrees of freedom (d.f.)}}$$

p Values and Significance Testing

p Value:

Exact probability of getting a computed test statistic that is due to chance. The smaller the *p* value, the smaller the probability that the observed result occurred by chance.

Chi-square Goodness of Fit test for Frequencies

- A statistical test to determine whether some observed pattern of frequencies corresponds to an expected pattern.

  One way table—Test of one factor

  Two way table---Test of two factors

Chi Square Example – Test of the effectiveness of three special deals

- A retail electronics chain needs to test the effectiveness of three special deals
- Each deal offered for a month
- Measure the effect of each deal on the number of customers visiting a test store during the time deal is on

Chi-square Goodness of Fit test for Frequencies

- Question: Is there a significant difference between the number of customers visiting the store under each deal?
  - Ho: number of customers equal across deals
  - Ha: significant difference in the no of customers visiting the store under various deals
- Level of significance = .05

Deal	Month	Customers per Month	Expected
1	April	11,700	11,860
2	May	12,100	11,860
3	June	11,780	11,860
Total		**35,580**	**35,580**

Chi-Square Goodness-of-Fit Test for Frequencies

– A statistical test to determine whether some observed pattern of frequencies corresponds to an expected pattern

$$\chi^2 = \sum_{i=1}^{k} \frac{[O_i - E_i]^2}{E_i}$$

– O_i is the observed number of cases falling in the ith category,
– E_i is the expected number of cases falling in the ith category, and
– k is the number of categories

Calculate Chi-Square value

$$= \frac{(11{,}700 - 11{,}860)^2}{11860} + \frac{(12{,}100 - 11{,}860)^2}{11860} + \frac{(11{,}780 - 11{,}860)^2}{11860}$$

= 7.6

$$\chi^2 = \sum_{i=1}^{k} \frac{[O_i - E_i]^2}{E_i}$$

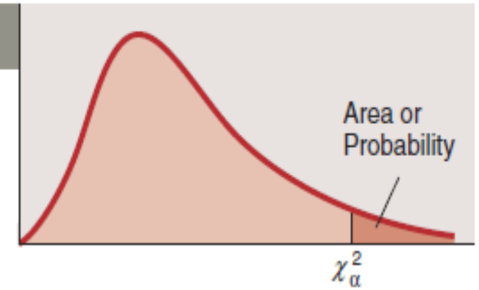
- df = k-1=3-1=2,
- Significance level= 0.05 , table value= 5.99.
- Conclusion: Because observed value is higher than the table value

(7.6 > 5.99), we reject the null.

Deal	Month	Customers per Month	Expected
1	April	11,700	11,860
2	May	12,100	11,860
3	June	11,780	11,860
Total		**35,580**	**35,580**

EXHIBIT 4 Chi-Square Distribution

Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, χ^2_α 23.2093.



Degrees of Freedom	Area in Upper Tail									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.0000393	.000157	.000982	.000393	.015709	2.70554	3.84146	5.02389	6.63490	7.87944
2	.0100251	.0201007	.0506356	.102587	.210720	4.60517	5.99147	7.37776	9.21034	10.5966
3	.0717212	.114832	2.15795	.351846	.584375	6.25139	7.81473	9.34840	11.3449	12.8381
4	.206990	.297110	.484419	.710721	1.063623	7.77944	9.48773	11.1433	13.2767	14.8602
5	.411740	.554300	.831211	1.145476	1.61031	9.23635	11.0705	12.8325	15.0863	16.7496
6	.675727	.872085	1.237347	1.63539	2.20413	10.6446	12.5916	14.4494	16.8119	18.5476
7	.989265	1.239043	1.68987	2.16735	2.83311	12.0170	14.0671	16.0128	18.4753	20.2777
8	1.344419	1.646482	2.17973	2.73264	3.48954	13.3616	15.5073	17.5346	20.0902	21.9550
9	1.734926	2.087912	2.70039	3.32511	4.16816	14.6837	16.9190	19.0228	21.6660	23.5893
10	2.15585	2.55821	3.24697	3.94030	4.86518	15.9871	18.3070	20.4831	23.2093	25.1882
11	2.60321	3.05347	3.81575	4.57481	5.57779	17.2750	19.6751	21.9200	24.7250	26.7569
12	3.07382	3.57056	4.40379	5.22603	6.30380	18.5494	21.0261	23.3367	26.2170	28.2995

Univariate versus Multivariate Analysis

- The ***chi-square test*** provides a method for testing the association between the row and column variables in a two-way table.
- H_0 assumes that there is no association between the variables (in other words, one variable does not vary according to the other variable),
- H_a claims that some association does exist.
  - The alternative hypothesis does not specify the *type* of association, so close attention to the data is required to interpret the information provided by the test.

Analyses Involving Categorical Measures

- Two-Way Cross Tabulations
  - A multivariate technique used for studying the relationship between two or more categorical variables (i.e., nominal- or ordinal-level)
  - Cross tabs consider the joint distribution of sample elements across variables
  - It is the most used multivariate data analysis technique

Example

- **Your university library is concerned about student desires for library hours on Sunday morning (9:00 a.m. – 12:00 p.m.).**
- **It has undertaken a random sample of 1,600 undergraduate students (one-half men, one half women) in each of four status levels (i.e. 400 freshmen, 400 sophomores, 400 juniors, 400 seniors.)**
- **If the percentage of students preferring Sunday morning hours are those shown below, what conclusions can the library reach?**

Observed Results

	Sen	Jun	Soph	Fresh	TOTAL
Women	70	53	39	26	188
Men	30	48	31	27	136
TOTAL	100	101	70	53	324

Step 1: Hypotheses

- Ho: There is no relationship between gender and class year.
- Ha: There is a significant relationship between gender and class year.
- Note again:
  - The null hypothesis H_0 assumes that there is no association between the variables while the alternative hypothesis H_a claims that some association does exist.

Step 2: Determine Expected Results

	Sen	Jun	Soph	Fresh	TOTAL
Women	70	53	39	26	188
Men	30	48	31	27	136
TOTAL	100	101	70	53	324

	Sen	Jun	Soph	Fresh	TOTAL
Women	58.02	58.60	40.62	30.75	188
Men	41.98	42.40	29.38	22.25	136
TOTAL	100	101	70	53	324

**Cell values calculated by multiplying each row margin total by each column margin total and dividing by the grand total, e.g., cell 1(woman x sen)=58.02=188*100/324=58.02.*

STEP 3: Calculate Chi-Square Value

	Sen	Jun	Soph	Fresh	TOTAL
Women	2.47	0.54	0.06	0.73	
Men	3.42	0.74	0.09	1.02	
TOTAL	$\chi^2 = 2.47 + 0.54 + 0.06 + 0.73 + 3.42 + ... + 1.02 = 9.07$				

Cell values calculated by taking the squared difference between the expected and observed and dividing expected,
e.g. cell 1(woman x sen)=(70-58.02)^2/58.02=2.47

$$\chi^2 = \sum_{i=1}^{k} \frac{[O_i - E_i]^2}{E_i}$$

STEP 4: Find Chi Square Table Value based on Significance Level and Degrees of Freedom

- Degrees of freedom = (r-1) * (k-1) = (2-1) * (4-1) = 3
- The tabulated chi square, 3 df, .05 alpha = 7.814

STEP 5: Compare Results and State Conclusion

- The calculated chi square = 9.06
- Conclusion: Calculated chi-square > Table chi-square
- We can state that, of those who prefer Sunday morning hours, gender and class year are statistically dependent. That is, there is a significant difference in how gender and class status will affect the use of the library on Sunday mornings. We can expect more Senior women than any other group.