Goodness of Fit

- χ^2-Goodness of Fit Hypothesis Test
- Test for Independence
- Test for Homogeneity

Observed vs. Expected

A craps player suspects that the casino is using weighted dice. A die throw is observed 300 times and the outcomes are shown below.

	1	2	3	4	5	6
Observed	45	50	58	40	53	54
Expected	50	50	50	50	50	50

Observed vs. Expected

A craps player suspects that the casino is using weighted dice. A die throw is observed 300 times and the outcomes are shown below.

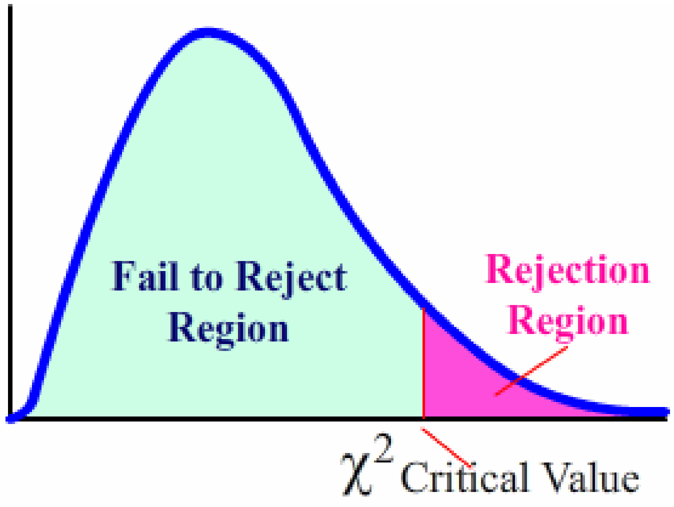
	1	2	3	4	5	6
Observed	45	50	58	40	53	54
Expected	50	50	50	50	50	50

$$\chi^2 = \sum \frac{(O-E)^2}{E} \qquad \frac{(45-50)^2}{50} = 0.5...$$

χ^2 Hypothesis Test

$H_0: p_1 = p_2 = p_3 = p_4 = p_5 = p_6$

$H_1:$ At least two of the proportions differ from each other



Requirements: Data were randomly selected and all of the expected counts are at least 5.

TI 83+/TI 84 Calculator

TI 83+/Older 84

- Download the app at www.aw.com/triola or from another calculator.
- Enter observed and expected into L1 and L2
- Press PRGM -> GF

Newer TI 84

- STAT -> TESTS -> χ^2GOF-Test
- Enter observed and expected into L1 and L2

Hypothesis Test

A craps player suspects that the casino is using weighted dice. A die throw is observed 300 times and the outcomes are shown below.

	1	**2**	**3**	**4**	**5**	**6**
Observed	45	50	58	40	53	54
Expected	50	50	50	50	50	50

$H_0: p_1 = p_2 = p_3 = p_4 = p_5 = p_6$

$H_1:$ At least two of the proportions differ from each other

$\alpha = 0.05$

P-Value = 0.51

There is insufficient evidence to make a conclusion about the die having any one number more likely to occur than any other.

Goodness of Fit

South Lake Tahoe is 62% White, 23% Hispanic, 7% Asian, and 8% Other. A survey of 350 LTCC students found that 245 were White, 55 were Hispanic, 36 were Asian, and 14 were Other. What can be concluded at the 0.05 level of significance?

Goodness of Fit

Bedford's Law States that the leading digits of numbers follows this distribution

Digit	1	2	3	4	5	6	7	8	9
Percent	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6

The IRS suspects that a business is making up numbers in its tax return. They look at the 348 leading digits of all the numbers from the return and come up with the following frequency table. What can be concluded at the 0.05 level of significance?

Digit	1	2	3	4	5	6	7	8	9
Frequency	48	52	35	40	34	50	47	20	22

Contingency Tables

A contingency table is a table in which frequencies correspond to two variables.

	< High School	High School Grad	College Grad	Post College Grad	**Total**
Democrat	62	240	81	35	**418**
Republican	56	209	131	59	**455**
Independent	25	48	42	12	**127**
Total	**143**	**497**	**254**	**106**	**1000**

Independence

Recall that A and B are independent if

$$P(A \text{ and } B) = P(A)P(B)$$

	< High School	HS Grad	College Grad	Post Coll. Grad	Total
Democrat	62	240	81	35	418
Republican	56	209	131	59	455
Independent	25	48	42	12	127
Total	143	497	254	106	1000

$$P(\text{Dem}) = \frac{418}{1000},\ P(\text{College Grad}) = \frac{254}{1000},\ P(D \text{ and } CG) = \frac{81}{1000}$$

$$\text{Expected Probability} = \frac{418}{1000}\frac{254}{1000} = \frac{\frac{418 \times 254}{1000}}{1000} \qquad \frac{418 \times 254}{1000} \stackrel{?}{=} 81$$

χ^2

$$\text{Expected } (E) = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Independence

Recall that A and B are independent if

$$P(A \text{ and } B) = PA)P(B)$$

	< High School	HS Grad	College Grad	Post Coll. Grad	Total
Democrat	62	240	81	35	**418**
Republican	56	209	131	59	**455**
Independent	25	48	42	12	**127**
Total	**143**	**497**	**254**	**106**	**1000**

H_0 : Political affiliation and Education are Independent

H_1 : Political affiliation and Education are Dependent

X2 = 29.33, P-Value = 0.0000525

Reject H0. There is sufficient evidence to conclude that political affiliation and Education are Dependent

2nd x-1 (MATRIX)

EDIT -> 3x4 ENTER

Put in Data

Stats ->Tests -> X2-Test

Observed: A

Calculate

Test for Independence

The contingency table below shows the results of a survey on the sport athletes play and the color of their car. What can be concluded at the 0.05 level?

	Black	White	Red	Green	Blue
Baseball	34	45	52	23	45
Football	45	39	51	30	55
Basketball	18	20	24	15	22
Soccer	38	43	50	30	40

Test for Homogeneity

A Test for Homogeneity is used when we have **two samples** from **two different populations** and we want to see if they have the same distributions as each other.

This differs from a Goodness of Fit test in that a Goodness of Fit test involves a **single sample's** distribution that is being **compared** to a **known population** distribution

Test for Homogeneity

Do men and women have the same grade distribution at LTCC?

	A	B	C	D	F
Men	51	34	40	12	10
Women	72	45	40	15	17

$\chi^2 - \text{Test}$

H_0: The grade distribution is the same for men and women

H_1: The grade distribution for men differs from the distribution for women.

$\chi^2 = 2.047$

$\text{P-Value} = 0.727$

Conclusion: There is insufficient evidence to make a conclusion about the grade distributions being different for men and women.

Test for Homogeneity

Day and night students were asked if they agreed with the policy of giving low income students priority registration. Is there a difference between day and night students in how they agree with this policy? Use a 5% level of significance

	Strongly Agree	Agree	Disagree	Strongly Disagree
Day	24	31	23	18
Night	7	15	18	22