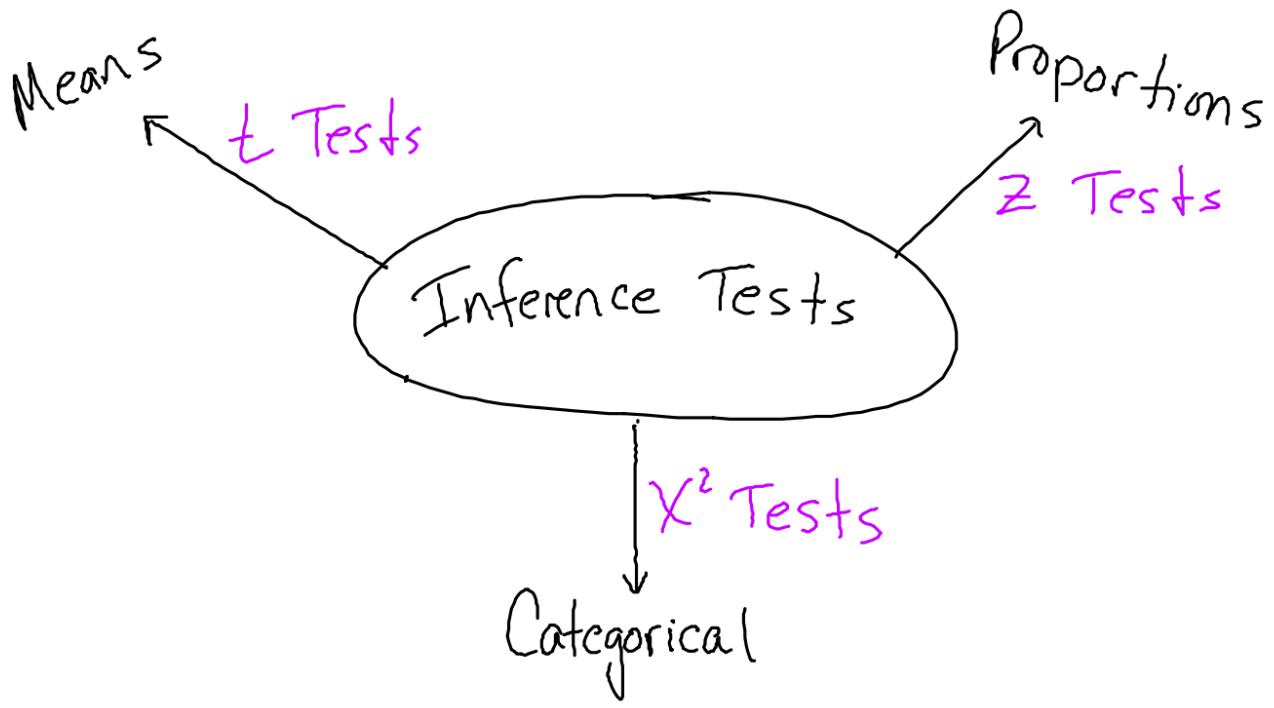
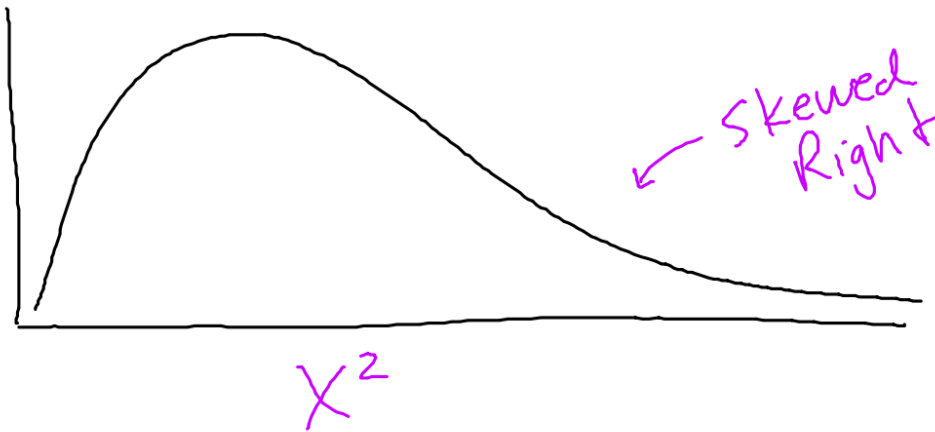


Sec 13.1



Chi-Square (χ^2) Distribution



~~P~~ No need to state parameter

H Hypotheses (in words)

H_0 : There is no difference ...

H_a : There is a difference ...

A Assumptions

1) SRS

2) Counts (not %)

3) Every expected count ≥ 1 and 80% ≥ 5

T Perform Test

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

* degrees of freedom varies

S State conclusion in context

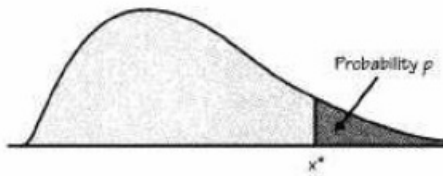


Table entry for p is the critical value x^* with probability p lying to its right.

TABLE E Chi-square distribution critical values

df	p											
	.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

CHI-SQUARE GOODNESS OF FIT TEST (GOF)

This test is used to determine if observed counts are equal to a hypothesized distribution.

A researcher believes the Mars Company is misleading the public on its color distribution of M&Ms. He wants to compare the color distribution from a random sample of M&Ms to the Mars Company's expected values:

	Brown	Red	Yellow	Green	Orange	Blue	Purple
Sample	4	4	16	10	8	4	4
Expected	4 (5)	4 (10)	16 (10)	10 (5)	8 (5)	4 (5)	4 (10)

= 50

H STATE NULL AND ALTERNATIVE HYPOTHESES:

H_0 : Color distribution of MMs is same as company claims

H_a : Color distribution of MMs is different than company claims

A DETERMINE THAT CONDITIONS FOR TEST ARE ACCEPTABLE:

- SRS - Unknown but random sample used
- ✓ Counts (not percents)
- Every expected count ≥ 1 and $80\% \geq 5$ See Above $\hat{\pi}_j \geq 5$

T PERFORM TEST:

a) Calculate Chi-Square statistic:

$$X^2 = \sum (O_i - E_i)^2 / E_i = \frac{(4-5)^2}{5} + \frac{(4-10)^2}{10} + \dots + \frac{(4-10)^2}{10} = 18$$

b) Determine Degrees of Freedom = Number of Categories - 1 = 7 - 1 = 6

c) Determine P-Value

i) Using Table E

At 5%, critical $\chi^2 = 12.59$, Since $18 > 12.59$, $p < .05$

ii) Using calculator:

DISTR $\rightarrow \chi^2 \text{cdf}(18, 100, 6) = .006$
 χ^2 Big df

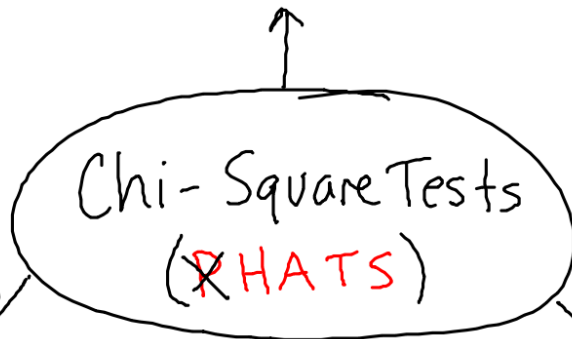
S STATE CONCLUSION IN CONTEXT

There is strong evidence ($p < .05$) to reject H_0 and conclude the distribution of MM colors is different than what the company claims

χ^2 GOF

Sec 13.2

Goodness of Fit (GOF)



2 (or more)
Samples

An arrow points from the left side of the central oval to this text.

Homogeneity of
Populations

Text written in purple, positioned below the arrow from the left side of the central oval.

Same
Sample

An arrow points from the right side of the central oval to this text.

Association / Independence

Text written in purple, positioned below the arrow from the right side of the central oval.

CHI-SQUARE TEST FOR HOMOGENEITY OF POPULATIONS

This test is used to determine if a single categorical variable has the same distribution in 2 (or more) distinct populations from 2 (or more) samples.

To determine if there was an association between race and opinions about schools, researchers surveyed 3 randomly selected groups of parents and asked them "Are high schools in your state doing an excellent, good, fair or poor job or don't you know enough to say?".

	Black Parents	Hispanic Parents	White Parents	TOTAL
Excellent	12	34	22	68
Good	69	55	81	205
Fair	75	61	60	196
Poor	24	24	24	72
Don't Know	22	28	14	64
TOTAL	202	202	201	605

DETERMINE EXPECTED COUNTS:

$$\text{Expected Count} = (\text{Row Total})(\text{Column Total}) / \text{Sample Size}$$

	Black Parents		Hispanic Parents		White Parents	
	Actual	Expected	Actual	Expected	Actual	Expected
Excellent	12	22.7	34	22.7	22	22.6
Good	69	68.5	55	68.5	81	69.1
Fair	75	65.4	61	65.4	60	65.1
Poor	24	24.0	24	24.0	24	23.9
Don't Know	22	21.4	28	21.4	14	21.3

H STATE NULL AND ALTERNATIVE HYPOTHESES

H_0 : There is no relationship between race and opinions about schools

H_a : There is a relationship between race and opinions about school

A DETERMINE THAT CONDITIONS FOR TEST ARE ACCEPTABLE:

- ✓ SRS- Random Sample
- ✓ Counts (not percents)
- ✓ Every expected count ≥ 1 and $80\% \geq 5$

> See Above $\hat{\alpha}_j$

T PERFORM TEST USING...

FORMULA/TABLE E:

a) Chi-Square Statistic: $\chi^2 = \sum (O_i - E_i)^2 / E_i =$

$$\chi^2 = \frac{(12-22.7)^2}{22.7} + \frac{(69-68.5)^2}{68.5} + \dots + \frac{(14-21.3)^2}{21.3} = 22.43$$

b) Degrees of Freedom = $(r-1)(c-1) = (5-1)(3-1) = 8$
Number of rows in table Number of columns in table

c) P-Value

i) Table E $\rightarrow p < .005$

ii) $\chi^2_{cdf}(22.43, 100, 8) \rightarrow p = .004$

CALCULATOR:

a) Store observed counts in a [R,C] matrix:

MATRIX \rightarrow EDIT \rightarrow [A] \rightarrow 5 x 3 \rightarrow Enter Counts \rightarrow QUIT

b) Perform χ^2 Test:

TESTS \rightarrow χ^2 -Test \rightarrow $\chi^2 = 22.43$
 \rightarrow $p = .004$ } Expected Counts in [B]

S STATE CONCLUSION IN CONTEXT:

There is significant evidence to reject H_0 ($p = .004$) and conclude that there is a relationship between race and opinions about schools

CHI SQUARE TEST OF ASSOCIATION/INDEPENDENCE

This test is used to determine whether there is a significant association between 2 categorical variables from the same sample.

To determine if there was a relationship between smoking status and socioeconomic levels, researchers categorized 356 federal male employees:

ACTUAL COUNTS	Socioeconomic Level		
	High	Middle	Low
Current Smoker	51	22	43
Former Smoker	92	21	28
Never Smoked	68	9	22
	211	52	93

116
141
99

356

DETERMINE EXPECTED COUNTS:

Expected Count = (Row Total)(Column Total)/ Sample Size

EXPECTED COUNTS	Socioeconomic Level		
	High	Middle	Low
Current Smoker	68.75	16.94	30.30
Former Smoker	83.57	20.60	36.83
Never Smoked	58.68	14.46	25.86

H STATE NULL AND ALTERNATIVE HYPOTHESES

H₀: There is no association between smoking and SES for federal male employees

H_a: There is an association between smoking and SES for federal male employees

A DETERMINE THAT CONDITIONS FOR TEST ARE ACCEPTABLE:

- ? • SRS - unknown; results may be invalid
- ✓ • Counts (not percents)
- ✓ • Every expected count ≥ 1 and $80\% \geq 5$

} See Above

T PERFORM TEST USING...

FORMULA/TABLE E:

a) Chi-Square Statistic: $X^2 = \sum (O_i - E_i)^2 / E_i =$

$$X^2 = \frac{(51 - 68.75)^2}{68.75} + \dots + \frac{(22 - 25.86)^2}{25.86} = 18.51$$

b) Degrees of Freedom = $(r - 1)(c - 1) = (3 - 1)(3 - 1) = 4$
Number of rows in table Number of columns in table

c) P-Value

i) Table E $p < .001$ ($p < .05$)

ii) X^2 cdf $(18.51, 100, 4) = .00098$

CALCULATOR:

a) Store observed counts in a [R,C] matrix:

MATRIX → EDIT [A] → 3 × 3 → Enter Counts → QUIT

b) Perform X^2 Test:

STAT → TESTS → X^2 -Test → $X^2 = 18.51$
→ p-value = .00098

S STATE CONCLUSION IN CONTEXT:

There is strong evidence to reject H_0 ($p = .00098$) and conclude an association exists between smoking and SES for federal male employees