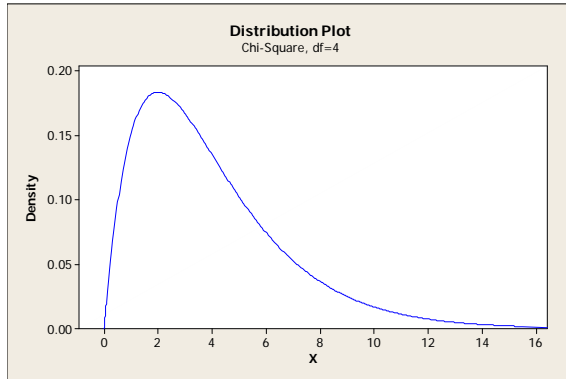


Chapter 26: Comparing Counts

1. The χ^2 model.

a) Draw a picture of the χ^2 model (either by hand or with the probability distribution plot in Minitab).



I used 4 degrees of freedom to get that graph.

b) Describe the skewness of the χ^2 distribution.

It is right skew.

c) Our χ^2 tests are always which tail?

Right

d) What are the two features that are important to keep in mind when sketching one by hand? (Alternatively, how is it different from the normal distribution or the t distribution?)

1. It is skew to the right.
2. It touches the x-axis at 0, passing through the point (0,0)

2. The null/alternatives.

a) For a test of homogeneity, write a sample (or skeleton) for null/alternative hypotheses.

H_0 : the distribution of _____ is the **same** across all levels of _____

H_A : the distributions are different.

b) For a test of independence, write a sample (or skeleton) for null/alternative hypotheses.

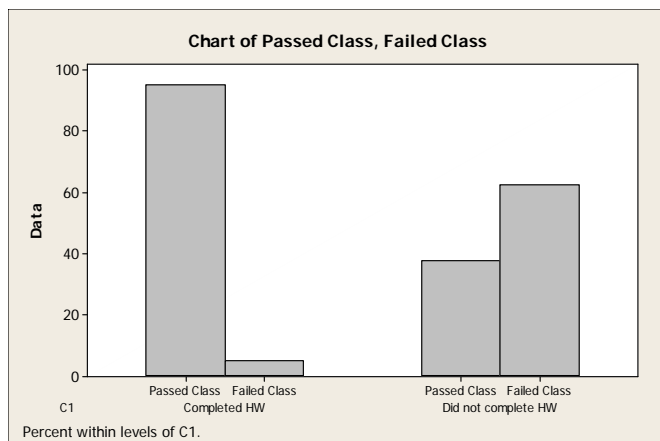
H_0 : _____ and _____ are **independent**

H_A : they are dependent.

3. A statistics teacher kept track of who passed/failed her class and whether or not they completed the homework.

	Passed Class	Failed Class
Completed HW	20	1
Did not complete HW	3	5

a) Provide a graphical representation of the data.



NOTE: for this question, it is best to have the HW as the outermost category because we are interested in the distributions of passing/not for each level of completing homework.

We can see that the distribution of passing a class appears to be different whether or not the student completed the homework. We will next perform the hypothesis test to determine if this difference is statistically significant.

b) Is there evidence that the distribution of passing the class or not is different for the group who completed their homework compared to the group who did not complete their homework?

i) Name the test: **Test of Homogeneity**

ii) Use a 0.01 level of significance to perform a hypothesis test.

Hypotheses:

H_0 : The distribution of passing the class is the same across all levels of completing homework.

H_A : They are different

Check Conditions:

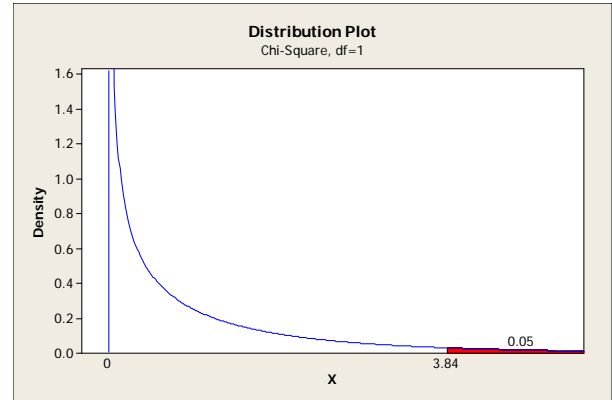
* Counted Data - Met

* Random – UNKNOWN

* Expected counts – NOT MET (as shown in the Minitab output below). We will still perform the test, but we acknowledge our results are unreliable.

Model:

We plot the Chi-Square distribution with one degree of freedom.



Mechanics:

Chi-Square Test: Passed Class, Failed Class

Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts

	Passed Class	Failed Class	Total
1	20	1	21
	16.66	4.34	
	0.672	2.575	
2	3	5	8
	6.34	1.66	
	1.763	6.759	
Total	23	6	29

Chi-Sq = 11.769, DF = 1, P-Value = 0.001

2 cells with expected counts less than 5.

Conclusions:

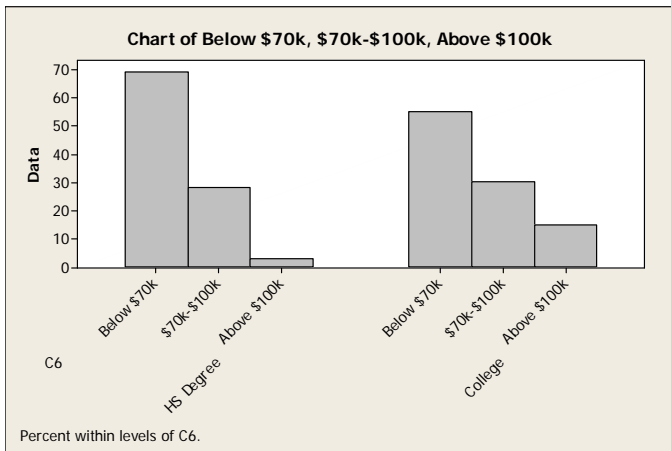
Reject H_0 .

Using the methods, we find that the distributions are different; however, our results are unreliable because we didn't pass the expected count condition.

4. Below are survey results from 100 professionals (income level and degree level).

	Below \$70k	\$70k-\$100k	Above \$100k
HS Degree	247	101	10
College	362	199	98

a) Provide a graphical representation of the data.



I used degree as the outermost category because we want to know if income level depends on the degree.

Since the distribution of income is not quite exactly the same for each level of degree, the variables appear to be slightly dependent. We will use a hypothesis test to determine if they are dependent.

b) Is there evidence that income level is dependent on level of degree?

i) Name the test: **Test of Independence**

ii) Perform a hypothesis test. Show all steps of the hypothesis test.

Hypotheses:

H_0 : Income level and degree level are independent

H_A : They are dependent

Check Conditions:

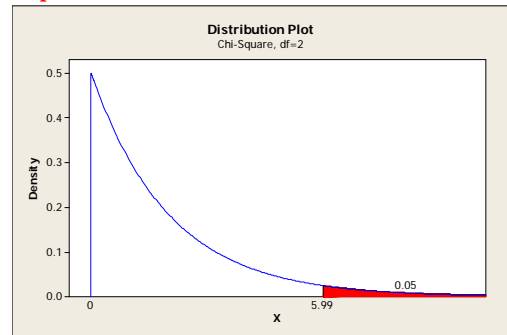
* Counted Data - Met

* Random – UNKNOWN

* Expected counts – met (as shown in the Minitab output below).

Model:

We plot the Chi-Square distribution with two degrees of freedom.



Mechanics:

Chi-Square Test: Below \$70k, \$70k-\$100k, Above \$100k

Expected counts are printed below observed counts

Chi-Square contributions are printed below expected counts

	Below \$70k	\$70k-\$100k	Above \$100k	Total
1	247	101	10	358
	214.38	105.60	38.02	
	4.964	0.201	20.648	
2	362	199	98	659
	394.62	194.40	69.98	
	2.697	0.109	11.217	
Total	609	300	108	1017

Chi-Sq = 39.836, DF = 2, P-Value = 0.000

Conclusions:

Reject H_0 .

There is significant evidence that income level is dependent on education level.