Test of Goodness of Fit

Lecture 42
Sections 14.1 - 14.3

Robb T. Koether

Hampden-Sydney College

Mon, Apr 19, 2010

Outline

Outline

The Hypotheses

Example (Goodness-of-Fit Test)

- I would like to test whether a die is fair.
- I rolled the die 90 times and obtained the following results

Number	Expected Counts	Observed Counts
1	15	20
2	15	20
3	15	14
4	15	8
5	15	14
6	15	14

The Hypotheses

Example (Goodness-of-Fit Test)

- I would like to test whether a die is fair.
- I rolled the die 90 times and obtained the following results

Number	Expected Counts	Observed Counts
1	15	20
2	15	20
3	15	14
4	15	8
5	15	14
6	15	14

- Does the die appear to be fair?

The Hypotheses

- To answer this question, we must return to the same question that we have asked many times already.
- Is the difference between what we observed and what we expected to observe small enough that we can reasonably attribute it to randomness?

The Hypotheses

- The null hypothesis specifies the probability (or proportion) for each category.
- The alternative hypothesis simply states that H_0 is false.

The Hypotheses

Example (Steps 1 and 2)

(1) $H_0 : p_1 = 1/6, p_2 = 1/6, p_3 = 1/6, p_4 = 1/6, p_5 = 1/6,$
$p_6 = 1/6.$

H_1 : H_0 is false.

(2) $\alpha = 0.05.$

Expected Counts

Definition (Observed and Expected Counts)

The observed counts are the counts that were actually observed in the sample. The expected counts are the counts that one would expect to observe if the null hypothesis were true.

Expected Counts

- To find the expected counts, we apply the hypothetical proportions to the sample size.
- For example, the hypothetical proportion for rolling a 1 is $1/6$, so we compute $1/6$ of 90:

$$\frac{1}{6} \times 90 = 15.$$

- Do *not* round the values off to whole numbers.

Outline

The Test Statistic

- Make a chart showing both the observed counts and the expected counts (in parentheses).

Number	1	2	3	4	5	6
Observed (Expected)	20 (15)	20 (15)	14 (15)	8 (15)	14 (15)	14 (15)

The Test Statistic

- Denote the observed counts by O and the expected counts by E.
- Define the chi-square (χ^2) statistic to be

$$\chi^2 = \sum_{\text{all cells}} \frac{(O - E)^2}{E}.$$

The Hypotheses

Example (Step 3)

(3) $$\chi^2 = \sum_{\text{all cells}} \frac{(O - E)^2}{E}.$$

The Value of the Test Statistic

- Clearly, if *all* of the deviations $O - E$ are small, then χ^2 will be small.
- But if *even a few* of the deviations $O - E$ are large, then χ^2 will be large.

The Value of the Test Statistic

Example (Step 4)

(4)

Number	1	2	3	4	5	6
Observed (Expected)	20 (15)	20 (15)	14 (15)	8 (15)	14 (15)	14 (15)

$$\chi^2 = \frac{(20-15)^2}{15} + \frac{(20-15)^2}{15} + \frac{(14-15)^2}{15} + \frac{(8-15)^2}{15} + \frac{(14-15)^2}{15} + \frac{(14-15)^2}{15}$$
$$= \frac{102}{15}$$
$$= 6.8.$$

Compute the *p*-Value

- The *p*-value for this example is the likelihood of observing a χ^2 value as large at 6.8.
- To find that value, we need to know something about the distribution of χ^2.

Outline

Chi-Square Degrees of Freedom

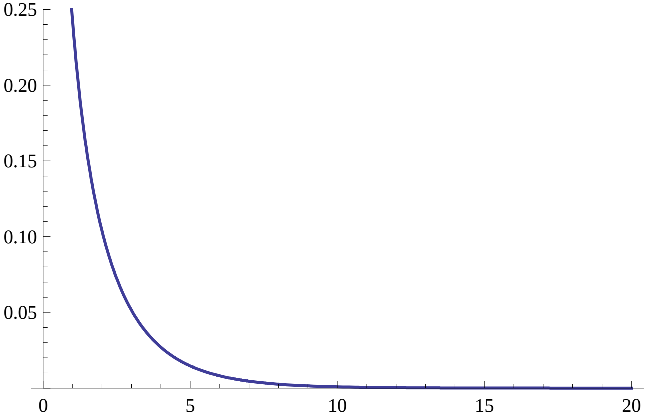
Definition (χ^2 degrees of freedom)

In a goodness-of-fit test, the number of degrees of freedom is one less than the number of cells.

- The χ^2 distribution has an associated degrees of freedom, just like the t distribution.
- Each χ^2 distribution has a slightly different shape, depending on the number of degrees of freedom.
- For example, we let χ^2_5 denote the chi-square statistic with 5 degrees of freedom.

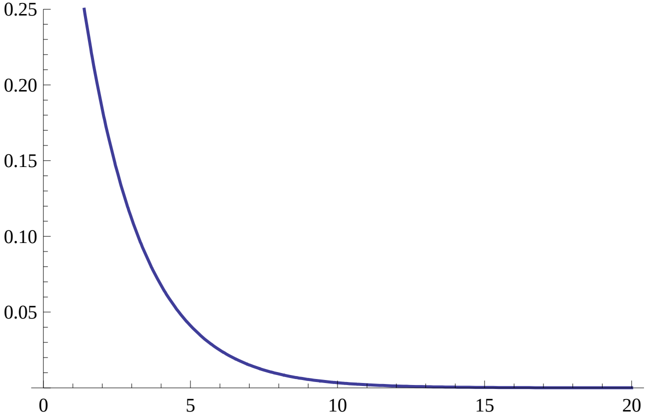
Chi-Square Degrees of Freedom

- The Graph of χ^2_1.



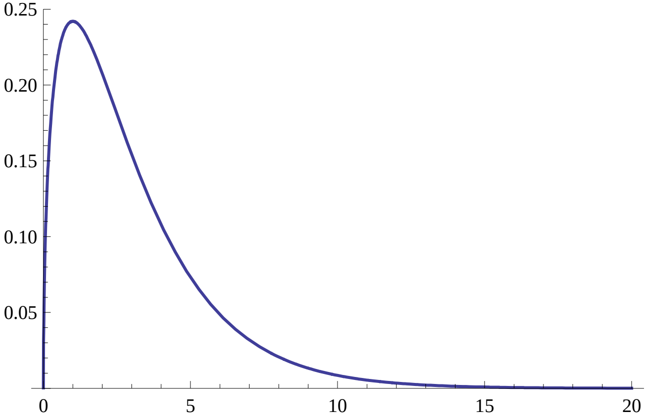
Chi-Square Degrees of Freedom

- The Graph of χ^2_2.



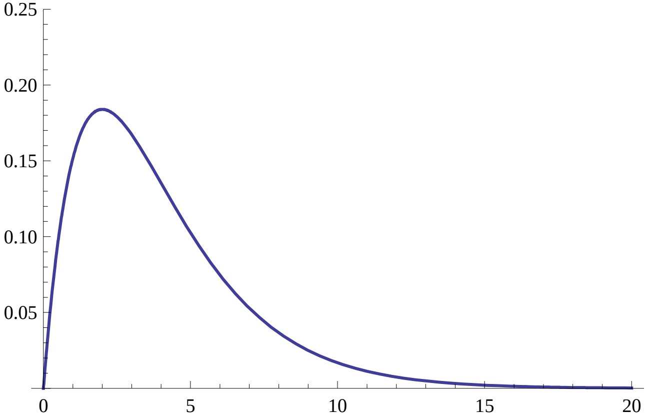
Chi-Square Degrees of Freedom

- The Graph of χ^2_3.



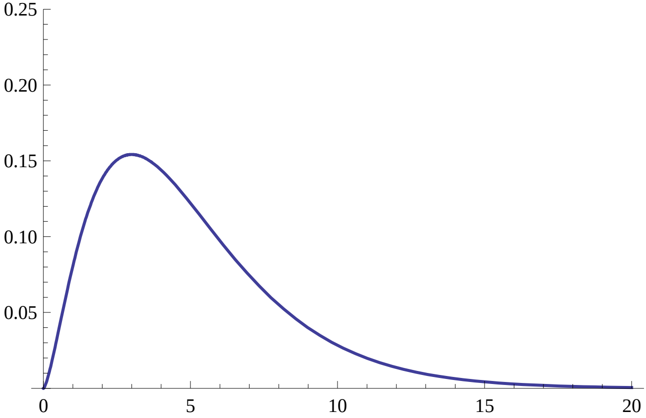
Chi-Square Degrees of Freedom

- The Graph of χ^2_4.



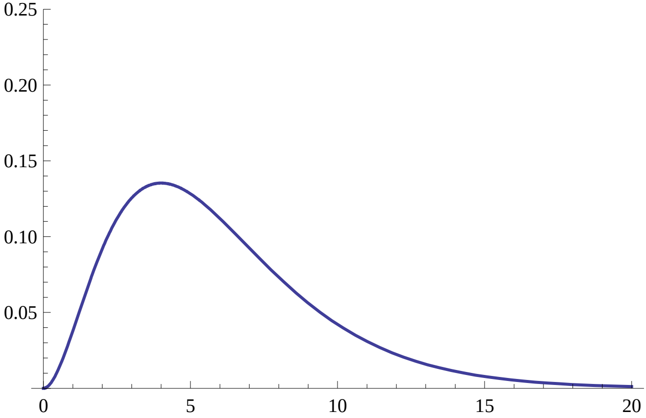
Chi-Square Degrees of Freedom

- The Graph of χ^2_5.



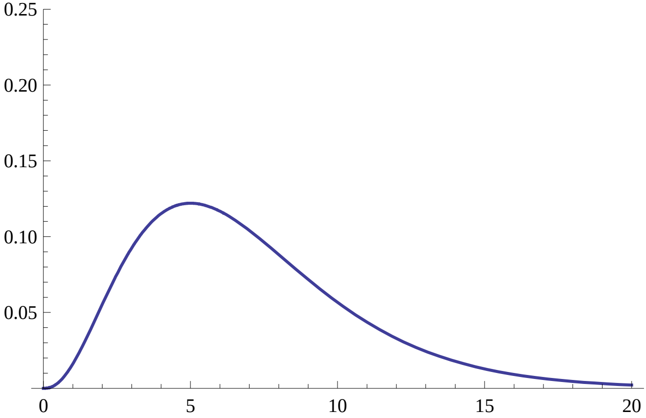
Chi-Square Degrees of Freedom

- The Graph of χ^2_6.



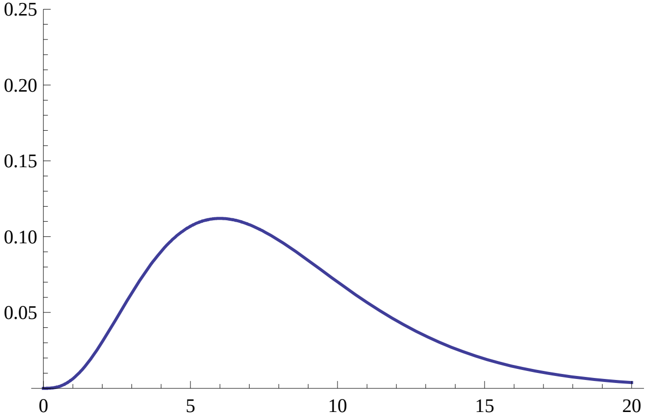
Chi-Square Degrees of Freedom

- The Graph of χ^2_7.



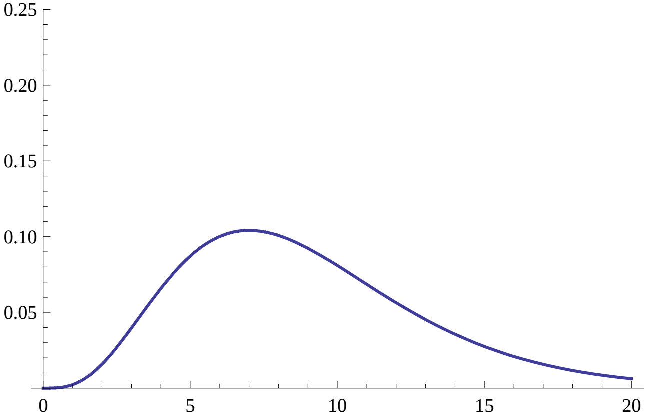
Chi-Square Degrees of Freedom

- The Graph of χ^2_8.



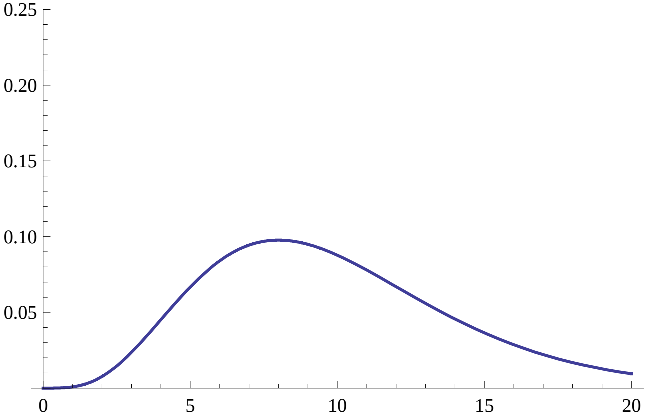
Chi-Square Degrees of Freedom

- The Graph of χ^2_9.



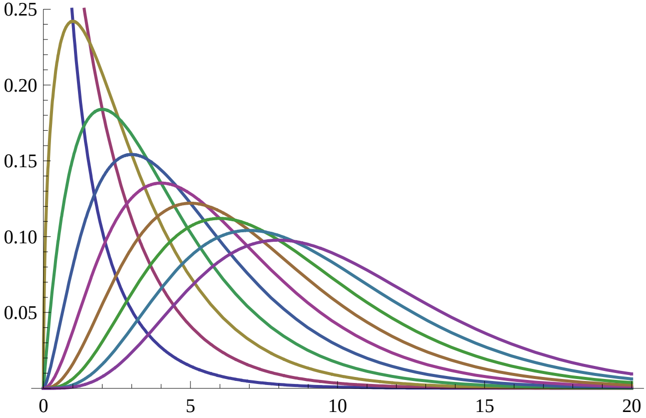
Chi-Square Degrees of Freedom

- The Graph of χ^2_{10}.



Chi-Square Degrees of Freedom

- The Graphs of $\chi^2_1, \chi^2_2, \ldots, \chi^2_{10}$.

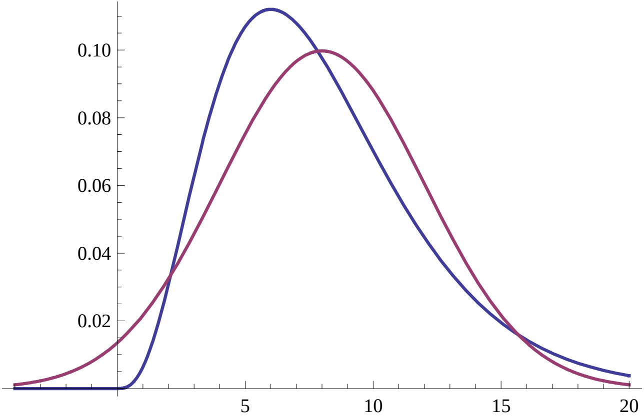


Properties of χ^2

- The chi-square distribution with *df* degrees of freedom has the following properties.
  - $\chi^2 \geq 0$.
  - It is unimodal.
  - It is skewed right (not symmetric!)
  - $\mu_{\chi^2} = df$.
  - $\sigma_{\chi^2} = \sqrt{2df}$.
  - If *df* is large, then χ^2_{df} is approximately normal with mean *df* and standard deviation $\sqrt{2df}$.

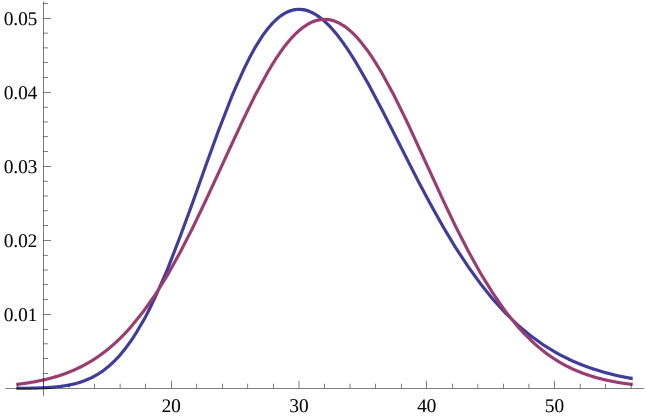
Chi-Square vs. Normal

- The graph of χ^2_8 vs. $N(8, 4)$



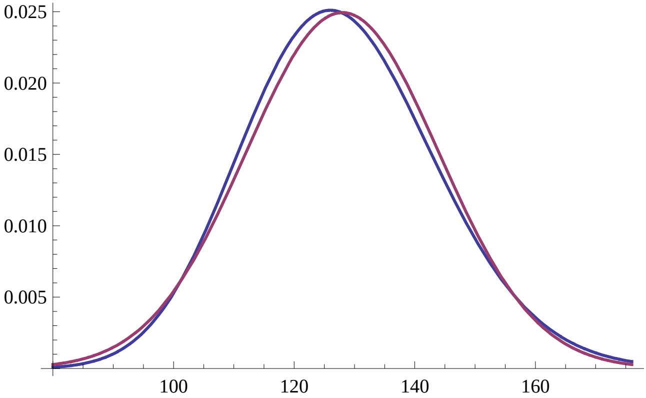
Chi-Square vs. Normal

- The graph of χ^2_{32} vs. $N(32, 8)$



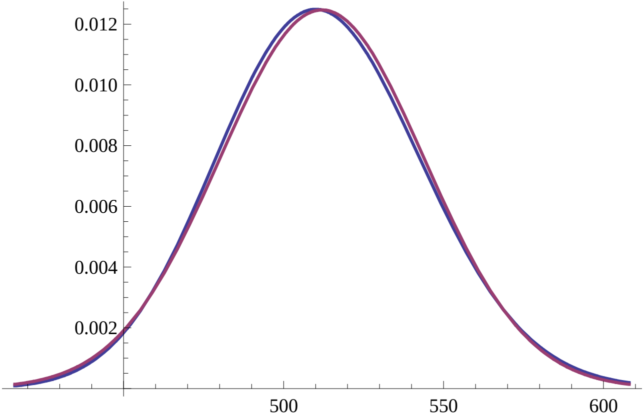
Chi-Square vs. Normal

- The graph of χ^2_{128} vs. $N(128, 16)$



Chi-Square vs. Normal

- The graph of χ^2_{512} vs. $N(512, 32)$



TI-83 - Chi-Square Probabilities

TI-83 Chi-square Probabilities

- Press `2nd DISTR`.
- Select χ^2`cdf`.
- Enter the lower endpoint, the upper endpoint, and the degrees of freedom.
- Press `ENTER`. The probability appears in the display.

TI-83 - Chi-Square Probabilities

Practice

- Find $P(\chi^2_3 > 6)$.
- Find $P(20 < \chi^2_{25} < 30)$.
- Find $P(\chi^2_6 < 10)$.
- Find the probability that χ^2_{512} is within one standard deviation of its mean.
- Find $P(\chi^2_5 > 6.8)$.

The Goodness-of-Fit Test

- In our example, we found $\chi^2 = 6.8$.
- There are 6 categories (1 - 6), so there are 5 degrees of freedom.

The Goodness-of-Fit Test

Example (Steps 5, 6, and 7)

(5) p-value $= \chi^2$`cdf(6.8,E99,5)` $= 0.2359$.

(6) Accept H_0.

(7) The die is fair.

Outline

Goodness-of-Fit Test on the TI-83

- Be careful when using the TI-83!
- There is a function called χ^2-`Test`, but it does *not* perform the goodness-of-fit test.
- Some TI-84s have a `GOF-Test` function.
- The `GOF-Test` function does perform the goodness-of-fit test.

Goodness-of-Fit Test on the TI-83

TI-83 Goodness-of-fit test

- Put the observed counts in list $\mathtt{L}_1$.
- Put the hypothetical proportions in list $\mathtt{L}_2$.
- Multiply $\mathtt{L}_2$ by the sample size and store as $\mathtt{L}_2$. These are the expected counts.
- Calculate $\mathtt{(L_1-L_2)^2/L_2}$ (either all at once or step by step).
- Go to `LIST > MATH` and select `sum` (item #5).
- Enter `Ans` and press `ENTER`. The value of χ^2 appears.
- Then use χ^2`cdf` to find the *p*-value.

Outline

Assignment

Homework

- Read Sections 14.1 - 14.2, pages 921 - 928.
- Let's Do It! 14.1.
- Exercises 1 - 5, page 928.