Class Worksheet

May 5 and 7

Math 125 *Kovitz* 2025

The χ^2-Test **(Chapter 28)**

The χ^2-test is used to check whether a box model for classification involving more than two categories is appropriate in view of certain observed data. This is an approximation to the actual probabilities and may be trusted when each expected frequency in the table is 5 or more.

The χ^2-statistic is obtained by evaluating

$$\chi^2 \;=\; \text{sum of} \;\; \frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}}$$

There is one term in the sum for each line in the table listing observed and expected frequencies.

This statistic measures the distance between observed and expected frequencies.

For the χ^2-test, P is approximately equal to the area to the right of the observed value for the χ^2-statistic, under the χ^2-curve with the appropriate number of degrees of freedom. When the model is fully specified (no parameters to estimate),

$$\text{degrees of freedom} \;=\; \text{number of terms in } \chi^2 \;-\; \text{one.}$$

- The χ^2-test says whether the data are like the result of drawing at random from a box whose contents are given.
- The z-test says whether the data are like the result of drawing at random from a box whose average is given.

The ingredients of the χ^2-test are the basic data, chance model, frequency table, χ^2-statistic, degrees of freedom, and the observed significance level.

With independent experiments, the results can be pooled by adding up the separate χ^2-statistics; the degrees of freedom add up too.

A Closer Look at Tests of Significance **(Chapter 29)**

Was the result significant?

Investigators should summarize the data, say what test was used, and report the P-value, instead of just comparing P to 5% or 1%.

Data Snooping

Data-snooping makes P-values hard to interpret.

One-tailed versus two-tailed tests.

The alternative hypothesis in each case.

Was the result important?

The P-value of a test depends on the sample size. With a large sample, even a small difference can be "statistically significant," that is, hard to explain by the luck of the draw. This doesn't necessarily make it important. Conversely, an important difference may not be statistically significant if the sample is too small.

The role of the model

To make sense out of a test of significance, a box model is needed.

If a test of significance is based on data for the whole population, watch out.

If a test of significance is based on a sample of convenience, watch out.

Does the difference prove the point?

A test of significance does not check the design of the study.

Problems to think about

(Chapter 28)

Find the area under the χ^2-curve with 5 degrees of freedom to the right of

(a) 1.14 (b) 6.06 (c) 11.07

One day, upon tossing a single die 60 times, I got:

12 ones, 21 twos, 8 threes, 6 fours, 3 fives, and 10 sixes.

Compute χ^2 and find P for this experiment.

Another day, upon tossing the same single die 600 times, I got:

93 ones, 123 twos, 94 threes, 108 fours, 101 fives, and 81 sixes.

Compute χ^2 and find P for this experiment.

Now, compute the pooled χ^2 using the combined degrees of freedom, and find the pooled P-value.

Is the die biased, based on the combined evidence?

(Chapter 29)

To test whether the number generators used in State Lotteries are truly random, eighty-five investigators run the χ^2-test on the first digits of the results from various games across the United States. Five of these investigators get significant values of P.

Comment on this result. How can this be explained unless some of the games are not random?

One year, 7,200 students in Boston took the math SAT. The average score was 421, and the SD was 100. In the same year, 385 students in Wellesley took the same math SAT. They averaged 511, and the SD was 97.

Did the students in Wellesley do significantly better than the students in Boston? If appropriate, make a two-sample z-test. If this isn't appropriate, explain why not.