List of Formulas, Procedures and Boxes for Quiz 5, May 7

(Quiz 5 will be on Wednesday, December 5, and covers material from Chapter 26: sections 1 to 5.)

Math 125 *Kovitz* Spring 2025

From Text

Boxes on pages 477, 479, 480, and 481.

Section 26.4 on page 482.

Formulas

Finding Areas under the Normal Curve

An area from minus a value to plus the same value is read off from the Normal Table; other areas are found by making a sketch and expressing the desired area in terms of areas that may be found by using the Table.

The Expected Value for the Sum of the Draws

The expected value for the sum of draws made at random with replacement from a box equals

$$(\text{number of draws}) \times (\text{average of box}).$$

The Standard Error for the Sum of the Draws

A sum is likely to be around its expected value, but to be off by a chance error similar in size to the standard error.

When drawing at random with replacement from a box of numbered tickets, the standard error for the sum of the draws is

$$\sqrt{\text{number of draws}} \times (\text{SD of box}).$$

A Shortcut for the SD of a Box with Only Two Kinds of Tickets

When the tickets in the box show only two different numbers, the SD of the box equals

$$\left(\begin{matrix}\text{bigger}\\ \text{number}\end{matrix} - \begin{matrix}\text{smaller}\\ \text{number}\end{matrix}\right) \times \sqrt{\begin{matrix}\text{fraction with}\\ \text{bigger number}\end{matrix} \times \begin{matrix}\text{fraction with}\\ \text{smaller number}\end{matrix}}$$

Classifying and Counting

If you have to classify and count the draws, put 0's and 1's on the tickets. Mark 1 on the tickets that count for you, 0 on the others.

Then find the average of the box and the expected value, and the SD (using the shortcut) and the standard error. Now convert any given values of the count to standard units and use the normal curve to approximate any chance being sought.

Various Standard Errors

For a given box model there are several SEs, each showing the likely size of a certain chance error. The corresponding formulas are:

$$
\begin{aligned}
\text{SE for sum} &= \sqrt{\text{number of draws}} \times \text{SD of box} \\
\text{SE for average} &= \frac{\text{SE for sum}}{\text{number of draws}} = \frac{\text{SD of box}}{\sqrt{\text{number of draws}}} \\
\text{SE for count} &= \text{SE for sum, from a 0–1 box} \\
\text{SE for percent} &= \frac{\text{SE for count}}{\text{number of draws}} \times 100\% = \frac{\text{SD of zero-one box}}{\sqrt{\text{number of draws}}} \times 100\%
\end{aligned}
$$

The SE for the sum is basic; the other formulas all come from that one. These formulas apply to draws made at random with replacement from a box.

Tests of Significance: The Null and the Alternative

To make a test of significance, the null hypothesis has to be formulated as a statement about a box model. Usually, the alternative does too.

- The *null hypothesis* says that an observed difference just reflects chance variation.
- The *alternative hypothesis* says that the observed difference is real.

The null hypothesis expresses the idea that an observed difference is due to chance. To make a test of signficance, the null hypothesis has to be set up as a box model for the data. The alternative hypothesis is another statement about the box; it says that the difference is real.

Testing Hypotheses: The Observed Significance Level

A test statistic is used to measure the difference between the data and what is expected on the null hypothesis.

z says how many SEs away an observed value is from its expected value, where the expected value is calculated using the null hypothesis. The formula is

$$z = \frac{\text{observed value} - \text{expected value}}{\text{standard error}}.$$

The observed significance level—often called the P-value—is the chance of getting a test statistic as extreme as or more extreme than the observed one. The chance is computed on the basis that the null hypothesis is right. The smaller the chance is, the stronger the evidence against the null.

A small value of P indicates that an explanation saying that this is chance variation is unreasonable. We cannot accept the model stated in the null.

A large value of P could very well be due to chance variation and we accept the null as a reasonable model.

The P-value of a test is the chance of getting a big test statistic—assuming the null hypothesis to be right. P is not the chance of the null hypothesis being right.

Making a Test of Significance

Based on some available data, the investigator has to—

- translate the null hypothesis into a box model for the data;
- define a test statisic to measure the difference between the data and what is expected on the null hypothesis;
- compute the observed significance level P.

The choice of test statistic depends on the model and the hypothesis being considered.

Making the Decision

Many statisticians have a dividing line that indicates how small the observed significance level has to be before an investigator should reject the null hypothesis.

- If P is less than 5%, the result is called *statistically significant* and the null hypothesis is rejected.
- If P is greater than 5%, we accept the null and state that chance error is a reasoanable explanation for this result.

Tests of Significance Using Zero-one Boxes (Chapter 26, section 5)

If a problem involves classifying and counting, the z-test can be used. The box will contain 0's and 1's.

Once the null hypothesis has been translated into a box model, it is easy to test using

$$z = \frac{\text{observed} - \text{expected}}{\text{SE}}$$